2014 Research Report

NYU School of Medicine

Big Data
Big Challenges
Big Opportunities

2014 Research Report

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

550 First Avenue
New York, NY 10016
NYULMC.org

NYU School of Medicine
NYU Langone Medical Center

This NYU Langone Medical Center logo, made with living yeast engineered to produce different colors, was constructed with an acoustic ejection robot.

CREDIT: NICHOLAS PHILLIPS IN JEF BOEKE'S LABORATORY.

Credits

2014 Research Report of NYU Langone Medical Center Produced by the Office of Communications and Marketing

Senior Vice President: Kathy Lewis
Editor: Marjorie Shaffer
Writer: Bryn Nelson
Copy Editor: Mel Minter
Production: Sherry Zucker
Design: Pentagram Design, Inc.
Photography: page 2, John Abbott; pages 6, 7, Robert Glick; pages 16, 26, 33, 35, Andrew Neary; pages 5, 13, 14, 15, 18, 19, 20, 25, 27, 28, 29, 30, 31, 34, 37, 39, Jake Chessum
Printing: Allied Printing Services, Inc.

Contents

On the cover: We live in the age of information, of huge and complex datasets, exemplified by the roughly 3 billion letters of DNA that spell out the genetic code of our species. This genetic information is carried on 23 pairs of chromosomes. The first three chromosomes, vastly magnified, appear on the cover. On the back cover are the sex chromosomes, X and Y.

Endless Opportunities

This year, we celebrate the diverse community of researchers at NYU Langone Medical Center who are turning big data into big opportunities for medical breakthroughs.

When researchers launched the Human Genome Project in 1990, an unprecedented international effort to decode every gene in the human body, few could have predicted just how far biomedicine would advance in a quarter-century. Today, scientists have identified more than 1,800 genes that cause disease, and developed more than 2,000 tests for genetic conditions. But with that success have come new challenges. The evolution of everything from genomics to electronic medical records has produced a flood of raw data that flows faster every day. The future of biomedical research depends on our ability to store, analyze, share, and most important, transform the data deluge into information that can be used to advance basic research and improve human health.

NYU Langone is rising to the challenge. We have brought together a wealth of analytical expertise to help find and evaluate crucial clues amid the oceans of information. On the pages that follow, we invite you to read about how our innovative researchers are using big data to address critical issues. Among them: How medical claims might help detect undiagnosed diabetes; how shifting populations of microbial residents might help diagnose diseases; how a comprehensive portrait of gene activity may reveal the secrets of childhood cancer; how the careful study of protein-drug interactions might lead to better drugs; and how the precise mapping of altered proteins in the heart might warn of sudden cardiac death.

Big challenges? Of course.

But the opportunities are endless.

Sincerely,

ROBERT I. GROSSMAN, MD
The Saul J. Farber Dean and Chief Executive Officer

DAFNA BAR-SAGI, PHD
Vice Dean for Science and Chief Scientific Officer

Harnessing the Power of Big Data

The torrent of information is rushing all around us, pouring from the Internet and smartphones, from electronic medical records and digital images, from DNA sequencers, environmental sensors, satellites, and countless other sources. Even more remarkable, an estimated 90 percent of all information around the world has been created in only the past two years, thanks to ever-faster computer processors and a booming digital revolution that has ushered in a new era of "big data."

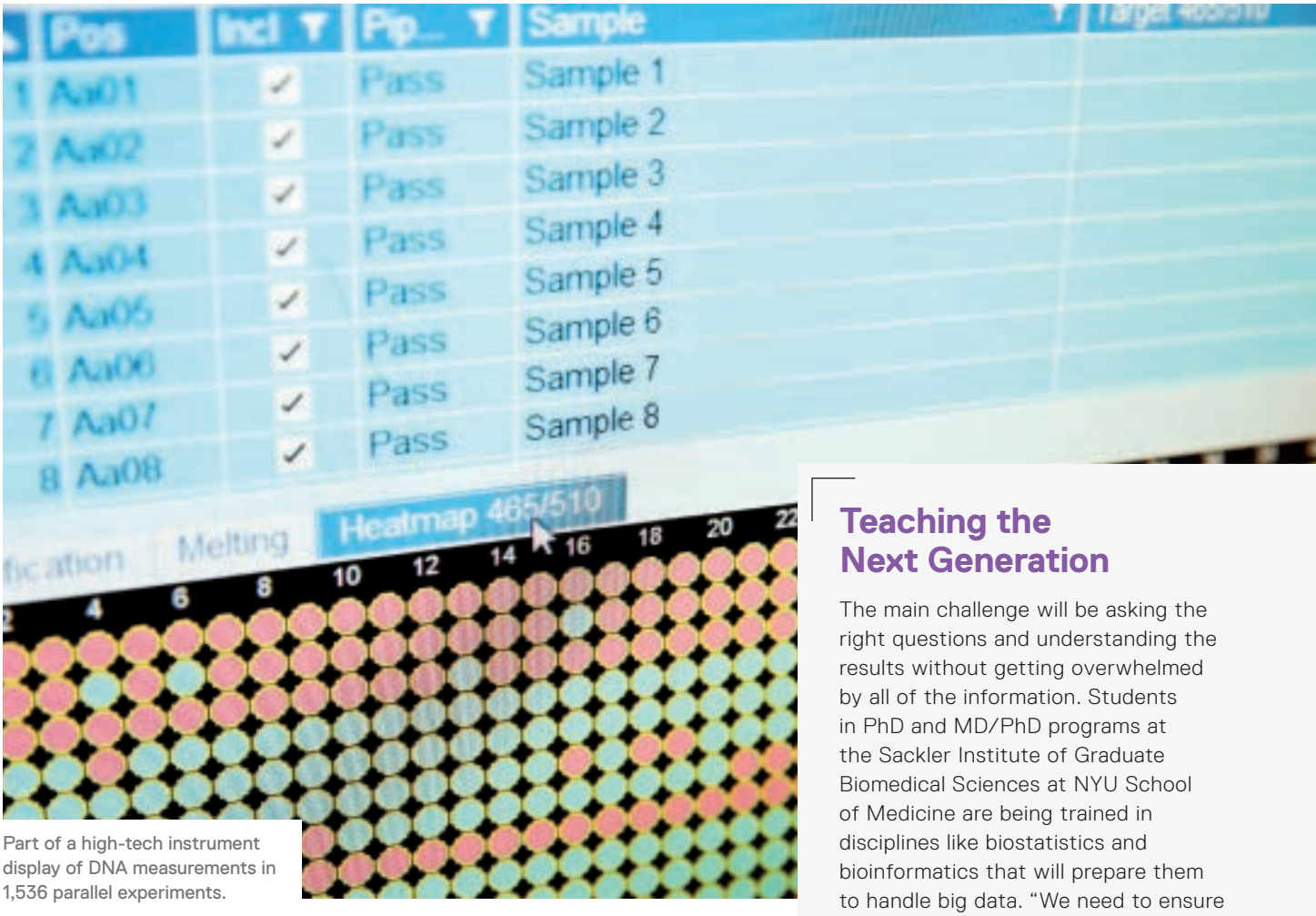
Big data, according to most definitions, refers to huge and complex sets of information that require computerized techniques for storage, sorting, and analysis. Perhaps nowhere is this boom more evident than in medicine and healthcare. The National Institutes of Health has taken notice and launched a massive funding initiative, called Big Data to Knowledge, that promises up to $24 million annually to help biomedical researchers extract meaning from the oceans of raw information generated daily.

"We are surrounded by a treasure trove of large datasets of all types, be it clinical data from our patients, genomic or protein data, data about our environment or ourselves, or data about the care that we're delivering," says Marc Triola, MD, associate dean for educational informatics and associate professor of medicine. "The challenges really come into play when we want to connect those dots and efficiently use all of that information to make smarter decisions in the lab and clinic. How we turn data into knowledge is the fundamental challenge."

The difficulty comes not just from the sheer volume, but also from the growing variety of information and the speed with which it is flowing and can be tapped. Personal health devices such as heart-monitoring electrocardiograms that link to a patient's smartphone can generate more clinical data in a week than a hospital typically produces for each patient in a year. In 1990, when researchers announced their intention to decipher all 3 billion letters of DNA within the human genome, the effort to identify our full catalog of genes required thousands of researchers and more than a decade of work. A quarter-century later, the technology and know-how have advanced so far that a small team of scientists can now sequence a person's entire genome within a few days.

NYU Langone Medical Center is part of a growing worldwide effort to develop advanced computational tools and methods to make sense of the information in ways that could transform medicine. By sifting data with computer-aided methods and developing models to see how they fit together, researchers can spot previously hidden patterns, make predictions about public health based on environmental factors, and accelerate translational medicine's journey from the bench to the bedside.

From research conducted here, composite pictures of data integrated from multiple sources are helping doctors determine which drug to give pediatric leukemia patients and assisting public health officials in deciding how to most effectively allocate resources to minimize cardiovascular disease in a minority community. By making the most of big data, "we could have a much richer understanding of the points of intervention for improving health and a much more nimble set of sensors that allows us to understand and intervene with influences on health," says Marc Gourevitch, MD, MPH, the Muriel G. and George W. Winger Professor of Population Health, chair of the Department of Population Health, and professor of medicine and psychiatry. "That's the big power of this."

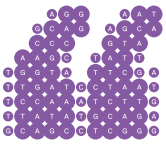


Part of a high-tech instrument display of DNA measurements in 1,536 parallel experiments.

Teaching the Next Generation

The main challenge will be asking the right questions and understanding the results without getting overwhelmed by all of the information. Students in PhD and MD/PhD programs at the Sackler Institute of Graduate Biomedical Sciences at NYU School of Medicine are being trained in disciplines like biostatistics and bioinformatics that will prepare them to handle big data. "We need to ensure that the new generation of scientists acquires the skills and quantitative tools to extract knowledge from large amounts of information," says Naoko Tanese, PhD, associate dean for biomedical sciences and director of the Sackler Institute.

Big data is also shaping the education of students in medical school. The Institute for Innovations in Medical Education, created in 2013, is teaching the next generation of physicians how to manage enormous projects. Future doctors and scientists will have so many large datasets at their fingertips, notes Dr. Marc Triola. "We know that big data is going to dramatically change every facet of our graduates' future lives," he says.



We are surrounded by a treasure trove of large datasets of all types, be it clinical data from our patients, genomic or protein data, data about our environment or ourselves, or data about the care that we're delivering," says Marc Triola, MD, associate dean for educational informatics.

Harnessing the Power of Big Data with Big Infrastructure

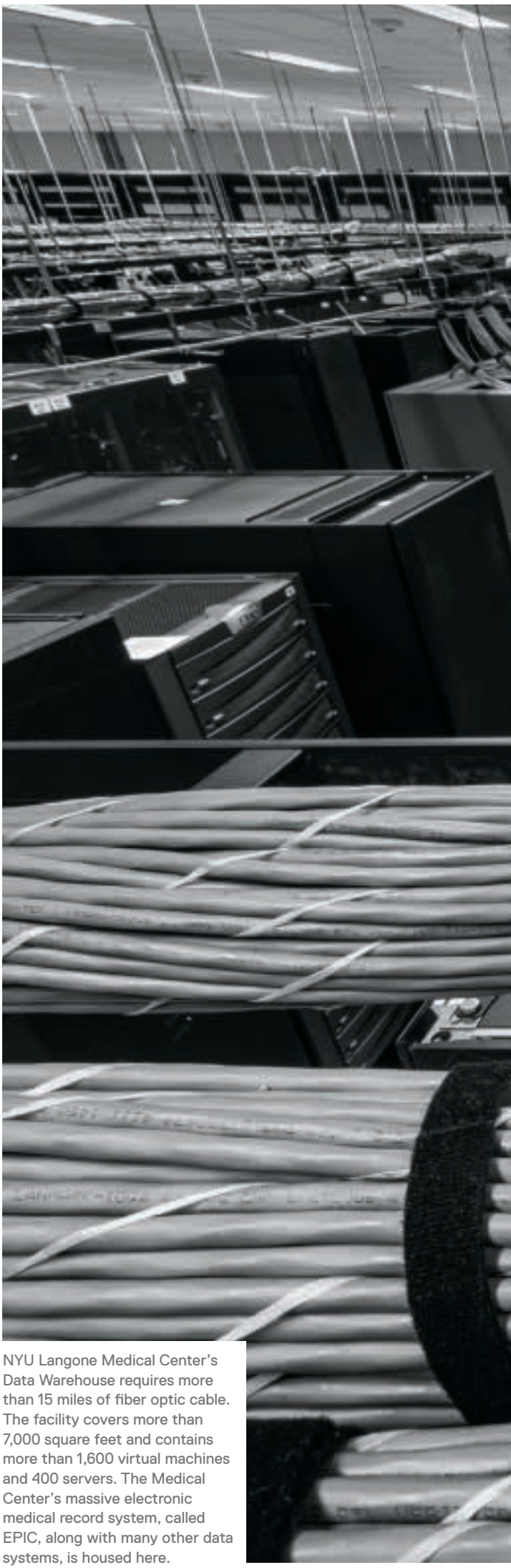
Navigating the vast expanse of big data in biomedicine can require something akin to a superhighway. In 2009, NYU Langone Medical Center became the first healthcare institution in the New York metropolitan region to begin construction of a massive electronic medical record system to track both inpatient and outpatient care.

The now-complete system, named Epic, has significantly bolstered NYU Langone's infrastructure by giving researchers unprecedented access to secure and de-identified patient information stored in a virtual warehouse. Epic also offers insights into providers' decisions and actions that may lead to breakthrough discoveries. Given the wealth of information that could be gleaned from the system, researchers have called it a potential gold mine.

One feature, known as MyChart, gives patients secure access to their own health information via their home computers, smartphones, and other devices. Meanwhile, clinicians and researchers can tap into Epic's campus-wide platform for a complete survey of every patient's past and present symptoms and diagnoses, lab results, scans, treatments, and prescriptions.

The Herculean effort to expand and solidify the Medical Center's information technology infrastructure led *Hospitals and Health Networks* magazine to designate NYU Langone as one of the country's most wired hospitals in 2013. Beyond overseeing the Epic system, the Information Technology Department has further aided the Medical Center's scientists with the launch of Research Navigator, a central portal that helps researchers manage their grants and clinical studies.

Worldwide, experts predict that demand for data storage capacity will soar by more than 30 percent every year. At NYU Langone, the High Performance Computing Facility, a data computing, storage, and consulting service run by the Center for Health Informatics and Bioinformatics, is keeping pace by significantly beefing up its storage. Its current capacity, equivalent to about 1 petabyte (or more than 1,000,000,000,000,000 bytes), is rapidly growing through the addition of computer servers and is enough to hold more than 2,000 years' worth of MP3 songs played continuously.



NYU Langone Medical Center's Data Warehouse requires more than 15 miles of fiber optic cable. The facility covers more than 7,000 square feet and contains more than 1,600 virtual machines and 400 servers. The Medical Center's massive electronic medical record system, called EPIC, along with many other data systems, is housed here.

How Complex Is It?

Although the human genome contains 3.2 billion letters of DNA, vast stretches of this genetic code are of unknown function and sometimes called "dark matter" by researchers. Recent estimates suggest that humans have roughly 20,000 genes, but of the 17,000 to 18,000 proteins tallied so far, some have been linked to DNA previously deemed "noncoding," while many potential genes don't seem to make any proteins. That's just the start of the complexity, however. DNA can be turned on and off in different cells through physical and chemical modifications to its structure, known as epigenetic changes. What's more, through an impressive variety of protein modifications, our true number of separate proteins may be orders of magnitude higher than what's been counted so far. Deciphering how all of these proteins interact—or fail to do so—in distinct cells throughout the human body remains one of medicine's biggest challenges.

Harnessing the Power of Big Data with Shared Access to Information

Our researchers are also part of national and international collaborations that are using big data analytics to design clinical trials and to advance neuroscience research. They are, for example, playing a major role in a program funded through the federal Patient-Centered Outcomes Research Institute to set up a national database of clinical data gathered from real-world settings, such as clinics. Participating medical centers in New York City are pooling their clinical data into what's known as the NYC Clinical Data Research Network.

The network, according to its leaders, is aiming to transform clinical research by making it possible to quickly conduct large studies based on data captured in electronic medical records. Instead of doing a clinical trial that compares two combinations of blood pressure drugs, for example, researchers could look to the pooled dataset to see how patients treated with those drugs fared over time.

Similarly, researchers at NYU Langone are collaborating on an international project called Neurodata without Borders: Neurophysiology, which aims to give neuroscientists greater access to big databases and an improved ability to share and combine their research results. Data-sharing pioneer Gyorgy Buzsaki, MD, PhD, the Biggs Professor of Neural Sciences, says the ambitious effort, if successful, will not only improve the quality of these datasets but also help researchers navigate through them to quickly find and access the most relevant data.

The Medical Center also joined the Clinical Proteomic Tumor Analysis Consortium. Launched by the National Cancer Institute, the consortium is a multicenter clinical collaboration dedicated to identifying the proteins linked to mutations in cancerous cells. Because researchers can now sequence the genomes of individual tumors, the associated catalog of altered proteins is growing rapidly. Identifying these proteins should help researchers better understand their potential roles in cancer progression and improve the ability to diagnose, treat, and prevent the disease.

Proteome

The full collection of proteins produced by the genes within a cell, a number that can vary according to the location of that cell within the body or to environmental conditions.

Genome

Our entire DNA catalog, consisting of 3.2 billion letters and an estimated 19,000 to 20,000 genes.

Epigenome

The composite group of physical and chemical modifications to the structure of DNA, which can inhibit or activate genes in different parts of the body.

The Language of Big Data

Big data is expanding our vocabulary. Researchers have introduced a menagerie of terms to describe the new bodies of information being discovered, assembled, and analyzed. New disciplines fueled by emerging technologies have sprung up, too. Altogether, a wider lens is opening across our molecular makeup, medical histories, social interactions, and other factors influencing health and disease.

Microbiome

The entire collection of microorganisms that live on and in our bodies, especially within the digestive tract.

Transcriptome

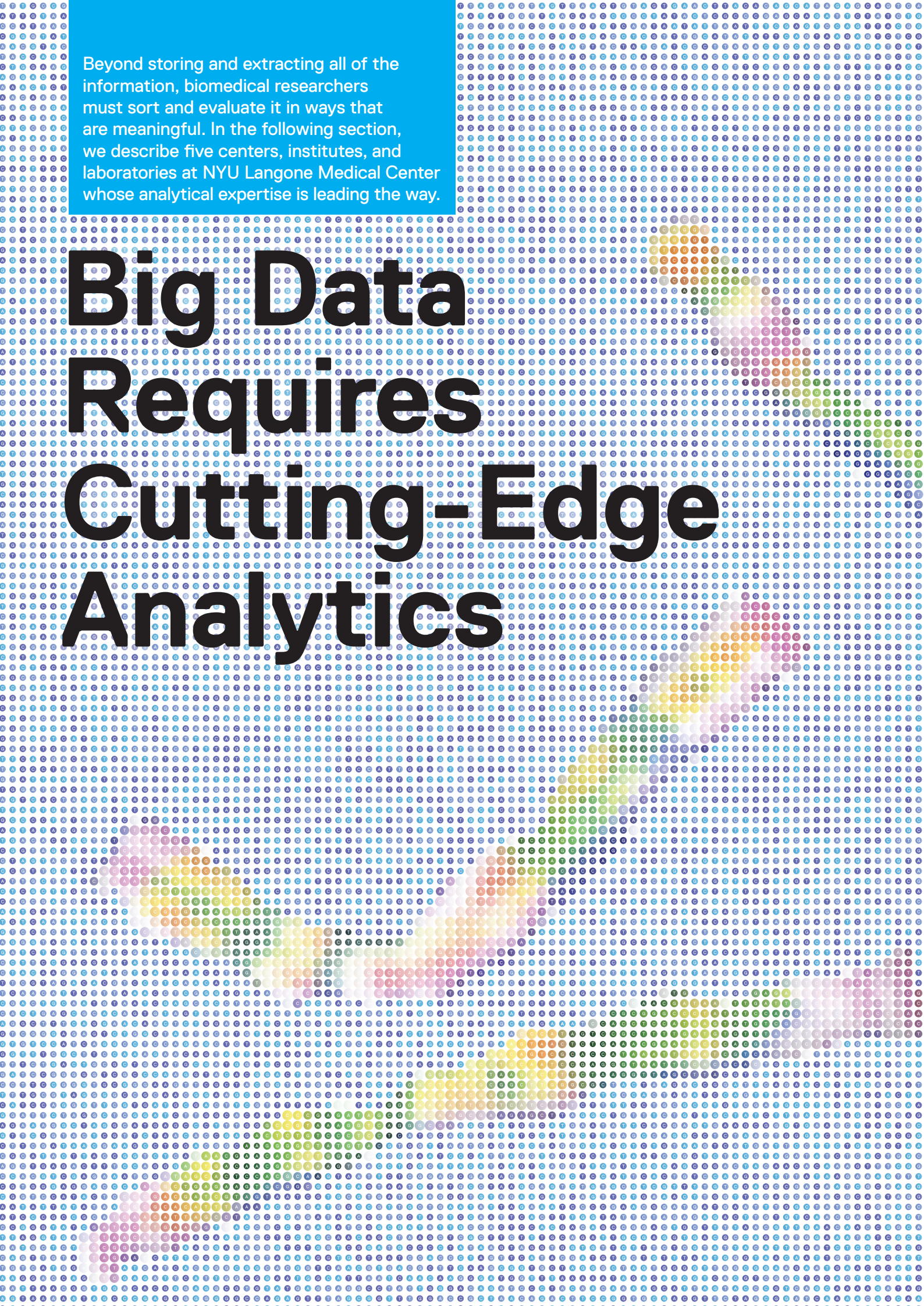
Our complete collection of RNA, a chemical cousin of DNA that provides the template for our genes. Measuring this RNA can help assess the relative activity level of every gene within specific parts of the body, like the brain.

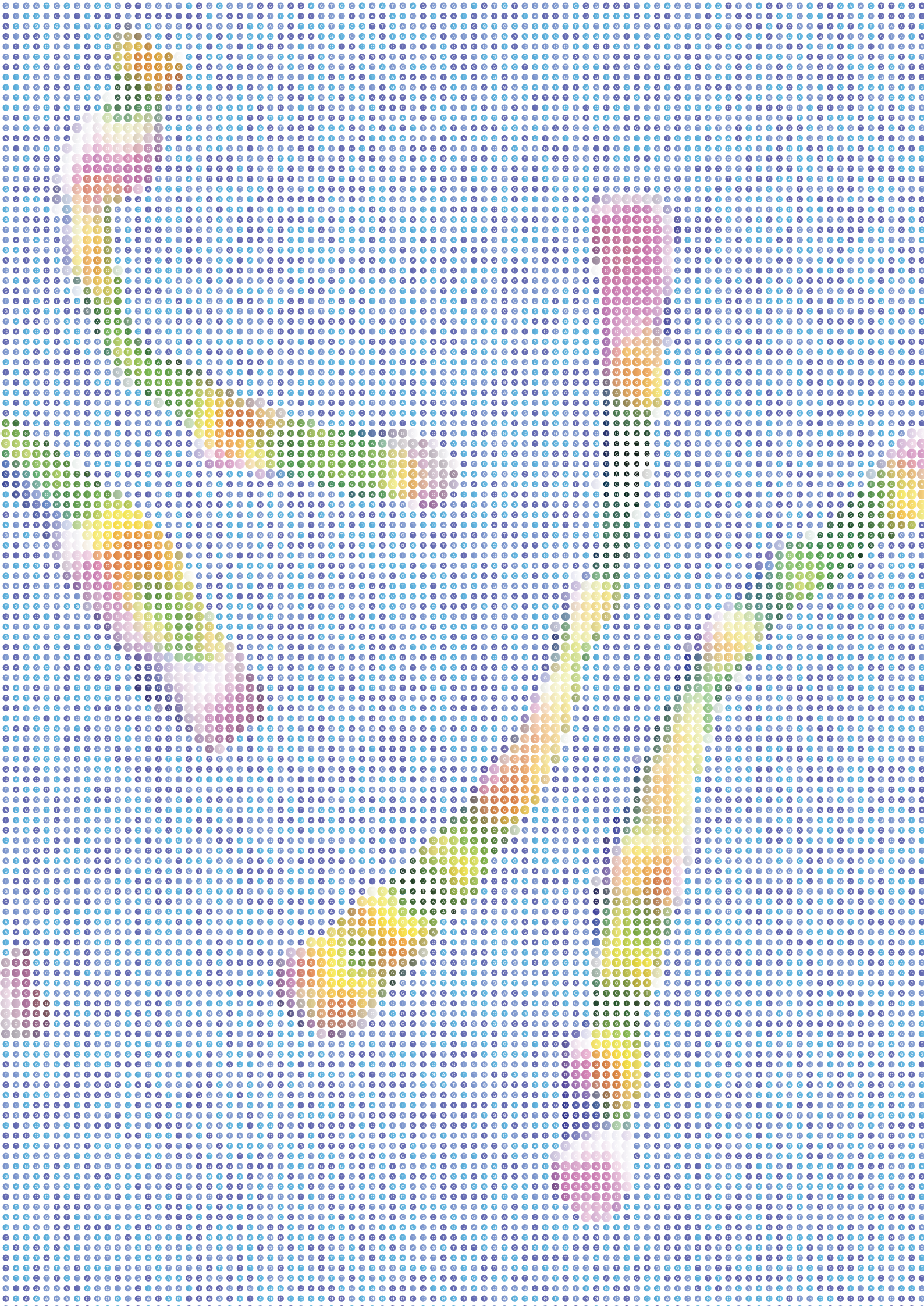
Bioinformatics

A scientific field in which researchers create and deploy computer-based tools and methods to help store, sort, and analyze biological data.

Beyond storing and extracting all of the information, biomedical researchers must sort and evaluate it in ways that are meaningful. In the following section, we describe five centers, institutes, and laboratories at NYU Langone Medical Center whose analytical expertise is leading the way.

Big Data Requires Cutting-Edge Analytics





Measuring Our Molecular Makeup

The Genome Technology Center

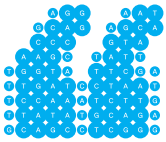
Fifteen years ago the first draft of the human genome was published and, in 2003, all 3.2 billion letters of DNA that spell out the genetic code of our species was finally completed at a cost of some $2.7 billion.

Next-generation sequencing technology, as it's known, has since advanced so far that a small team of scientists can now sequence a person's entire genome within a few days for a few thousand dollars.

Producing the data is only the start, however, says Adriana Heguy, PhD, director of the Genome Technology Center and professor of pathology. "Everything that we do with genomics is big data by definition, and it all needs to be mined by a big bioinformatics component," she says.

Beyond the lab's sophisticated sequencing machines, Dr. Heguy and colleagues are using other advanced equipment such as DNA and RNA microarrays, which can determine the relative activity of every gene in a specific sample to make unprecedented measurements of our genomic makeup. For one project, Dr. Heguy and her lab are sequencing the genome of rare pediatric brain tumors to determine how the tumors' genetic anomalies differ from normal cells and from tumor-linked mutations in adults. For another research effort, the center measured the relative activity of every gene in a specific part of the mouse brain by sequencing its full set of RNA. This transcriptome analysis is revealing how diet, age, and other conditions can determine which genes are on, off, or only partially active.

The center's tools will likely spur similar discoveries throughout NYU Langone. "We are optimizing some methods so that people can start looking at the entire transcriptome, but of a very limited amount of material—say, a few neurons in a fruit fly's brain," Dr. Heguy says. In selected cells, this remarkable precision could help researchers investigate the activity of every gene—and shine a new light on their roles.



Everything that we do with genomics is big data by definition, and it all needs to be mined by a big bioinformatics component," says Dr. Adriana Heguy.

ADRIANA HEGUY, PHD
Director of the Genome Technology Center and Professor of Pathology

Spotting Key Patterns among Proteins

The Proteomics Resource Center

Our genetic instruction manual may be exceedingly complicated. But it pales in comparison with the complexity of our full set of proteins, or the proteome. Each gene can be translated into many different proteins and each of these proteins can be extensively modified, resulting in multiple alternative forms. The ways in which these proteins all interact, in turn, present a dizzying number of possibilities and the potential for big data challenges.

Despite the challenge, Beatrix Ueberheide, PhD, director of the Proteomics Resource Center and assistant professor of biochemistry and molecular pharmacology, says deciphering the connections will be key to understanding biological mechanisms like disease pathways or the effects of drug interventions. The valuable information might help track the success of a chemotherapy regimen for breast cancer, for example, or warn of a cardiac condition that can lead to sudden death.

A chemist by training, Dr. Ueberheide specializes in a technique called mass spectrometry, which she likens to a "sensitive balance." It measures the mass of proteins and their fragments, enabling her to decipher their identity and abundance. The method, in fact, can characterize thousands of proteins in a particular mix in just two hours. Given the hundreds of samples in clinical studies, enormous datasets can accumulate, making bioinformatics data analysis essential for unraveling patterns.

In close collaboration with David Fenyo, PhD, assistant professor of biochemistry and molecular pharmacology, Dr. Ueberheide's center is conducting multiple studies to identify which proteins may be present, absent, or altered in a disease such as cancer, and to use these patterns as biomarkers to guide treatment. In other studies, the researchers are analyzing how our proteins interact under specific conditions, yielding an ever-growing dataset known as the interactome.

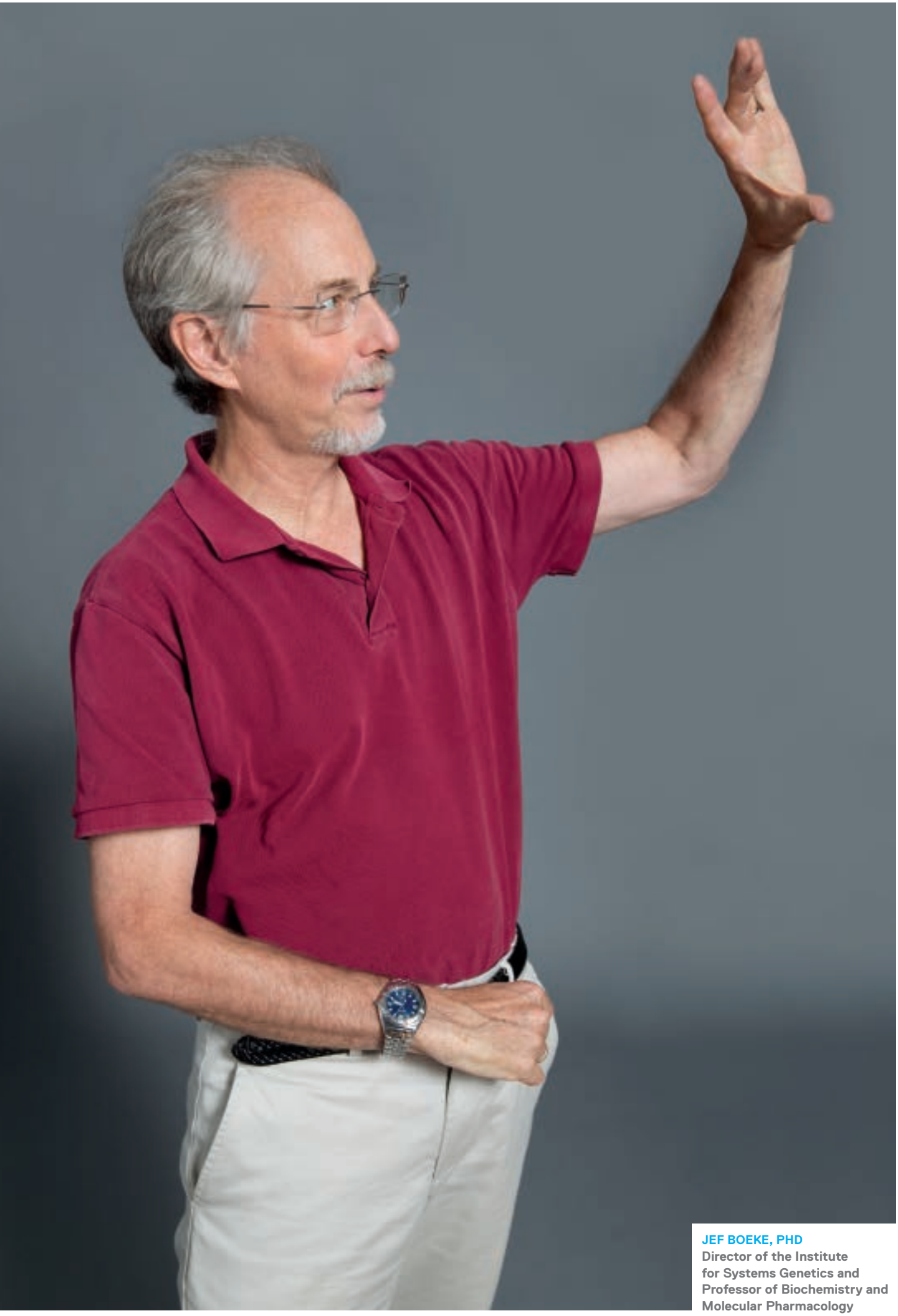
By knowing how our proteins link up, or why they fail to do so, researchers can better assess the risk for disease and manipulate specific proteins to prevent or enhance clinically relevant connections. "If you can actually see the changes in the proteome—the functional gene products—that gives you much better predictive power," Dr. Fenyo says.

BEATRIX UEBERHEIDE, PHD
Director of the Proteomics Resource Center and Assistant Professor of Biochemistry and Molecular Pharmacology

250,000–1 Million

Estimated number of proteins in the human proteome, based on the multiple variants of proteins encoded by our genes.

SOURCE: NATIONAL CANCER INSTITUTE



JEF BOEKE, PHD
Director of the Institute for Systems Genetics and Professor of Biochemistry and Molecular Pharmacology

Engineering New Solutions

Institute for Systems Genetics

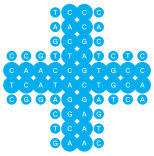
Jef Boeke, PhD, professor of biochemistry and molecular pharmacology, is no stranger to big data. In a seven-year effort, he and colleagues built an artificial chromosome in brewer's yeast—containing more than 270,000 letters of DNA—to demonstrate the potential of reprogramming organisms to produce drugs, biofuels, and other valuable compounds. This scientific discipline, known as synthetic biology, combines the tools of engineering and molecular biology to create carefully constructed biological systems with new features.

It's this kind of big potential that fueled the 2013 launch of the Institute for Systems Genetics, with Dr. Boeke at the helm. "It's the only institute in the world with the words 'systems genetics' in the title," he says. Systems genetics puts a high-tech twist on genome analysis, tapping data-processing techniques to map the vast networks that connect genes, proteins, and other molecules. Dr. Boeke's team will use the latest tools and insights from human genetics, systems biology, computational science, and biological engineering. "It's a major initiative to bring together this level of expertise under one roof," he says.

In the near future, the 3.2 billion letters of the human genome may be merely the starting point for highly complicated profiles of an individual's risk factors and well-being. On average, any two people differ by about 3 million DNA letters per genome—or roughly one variation every 1,000 letters. Most of these changes do not contribute to disease susceptibility. "But clearly, some of them do, and the challenge is to find the needles in the haystacks," Dr. Boeke says.

Genomic data can be immense just on its own. "But once you get into interpretation of images—for example, pictures of how gene expression changes over time inside individual cells—the analysis and storage problems just become orders of magnitude more complex," Dr. Boeke explains.

By using these tools of systems genetics, NYU Langone researchers may be able to interpret subtle patterns within the mounds of information. For patients with varying disease traits, this approach may point to differences that would otherwise remain hidden, convert big data into assessments of risk, and ultimately help clinicians design effective new treatments.



273,871

The number of DNA letters, or base pairs, that Dr. Boeke and colleagues pieced together to construct an artificial yeast chromosome called synIII.

The Clinical and Translational Science Institute

Which cardiovascular patients might benefit from which therapies? The Clinical and Translational Science Institute (CTSI), a partnership among New York University, NYU Langone Medical Center, and the New York City Health and Hospitals Corporation, is devoted to answering such questions.

"Big data analytics certainly has the potential to accelerate the pace of drug development, shortening the time it takes to bring effective treatments to the bedside," says Judith Hochman, MD, MA, the Harold Snyder Family Professor of Cardiology and senior associate dean for clinical sciences, who is leading a large multicenter trial comparing two treatment strategies for coronary heart disease.

Dr. Hochman and Bruce Cronstein, MD, the Dr. Paul R. Esserman Professor of Medicine and professor of pathology and pharmacology, are codirectors of the CTSI. Its particular strength, she says, is its ability to break down silos and work collaboratively with a broad range of departments, institutes, and centers, such as the Center for Health Informatics and Bioinformatics, and Research IT.

A new Research IT initiative being overseen by the CTSI, called DataCore, is helping to further close the gap between discovery and application. This suite of information technology tools and services helps researchers keep track of vast collections of clinical research data with electronic data capture systems to store and curate voluminous datasets. Other tools apply a sophisticated computer formula known as natural language processing to mine potentially valuable information contained within providers' notes.

Michael Cantor, MD, MA, director of Clinical Research Informatics, leads a team developing computer-based tools to manage diverse and often unwieldy datasets. Clinical information, he says, may be coming not only from doctors' notes but also from sensors and multiple other sources. Such data can be invaluable for predicting which diabetic patients may get better or worse, for example, but it may require plucking relevant clues from storehouses such as the Epic electronic medical record system and Medicare claims.

Quality control is a major concern. "When you have big data, you have many more data points, and you're introducing more potential for error," Dr. Cantor says. Sophisticated filters can separate good from bad, while other tools can find holes in a dataset that need to be filled.

"The potential is huge, from therapeutic development to comparative effectiveness research to identifying new subtypes of diseases," Dr. Cantor says, "but there's still a lot of work to be done."

JUDITH HOCHMAN, MD, MA
The Harold Snyder Family Professor of Cardiology, Senior Associate Dean for Clinical Sciences, and Codirector of the Clinical and Translational Science Institute

MICHAEL CANTOR, MD, MA
Director of Clinical Research Informatics and Associate Professor of Population Health and Medicine

Big data analytics certainly has the potential to accelerate the pace of drug development, shortening the time it takes to bring effective treatments to the bedside," says Dr. Judith Hochman.

CONSTANTIN ALIFERIS, MD, PHD
Director of the Center for Health Informatics and Bioinformatics and Associate Professor of Pathology

An Analytical Powerhouse

The Center for Health Informatics and Bioinformatics

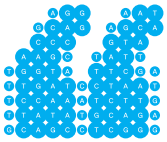
As technology such as high-speed sequencing machines began churning out millions of data points on molecules ranging from DNA to proteins, researchers faced a daunting challenge. How could they detect meaningful signals and patterns amid the rush of so much information?

To help make sense of this molecular barrage and the reams of medical and digital data that joined it, NYU Langone Medical Center created the Center for Health Informatics and Bioinformatics (CHIBI) in 2008. Steeped in math, computer science, and statistics, the center's health informatics and bioinformatics experts are helping researchers layer insights about our DNA, RNA, and protein composition onto medical and clinical records and other databases that capture information about behavior, environment, and even social interactions.

With the right analytical tools, a composite portrait of health and disease, whether of an individual or an entire population, can begin to emerge. "We are living in a world that is increasingly being linked digitally, and an incredible variety and volume of information is being stored online about anything that one can imagine," says Constantin Aliferis, MD, PhD, director of CHIBI and associate professor of pathology. "Combined with the capability to collect and analyze medical data and molecular data, for the first time this creates a complete picture that can go all the way from molecules to society."

CHIBI's faculty is developing sophisticated new algorithms, or math-based formulas, to separate the surge of data into powerful channels of information that can be used to build complicated simulations. These simulations allow medical researchers to make vital predictions about disease risks, for example.

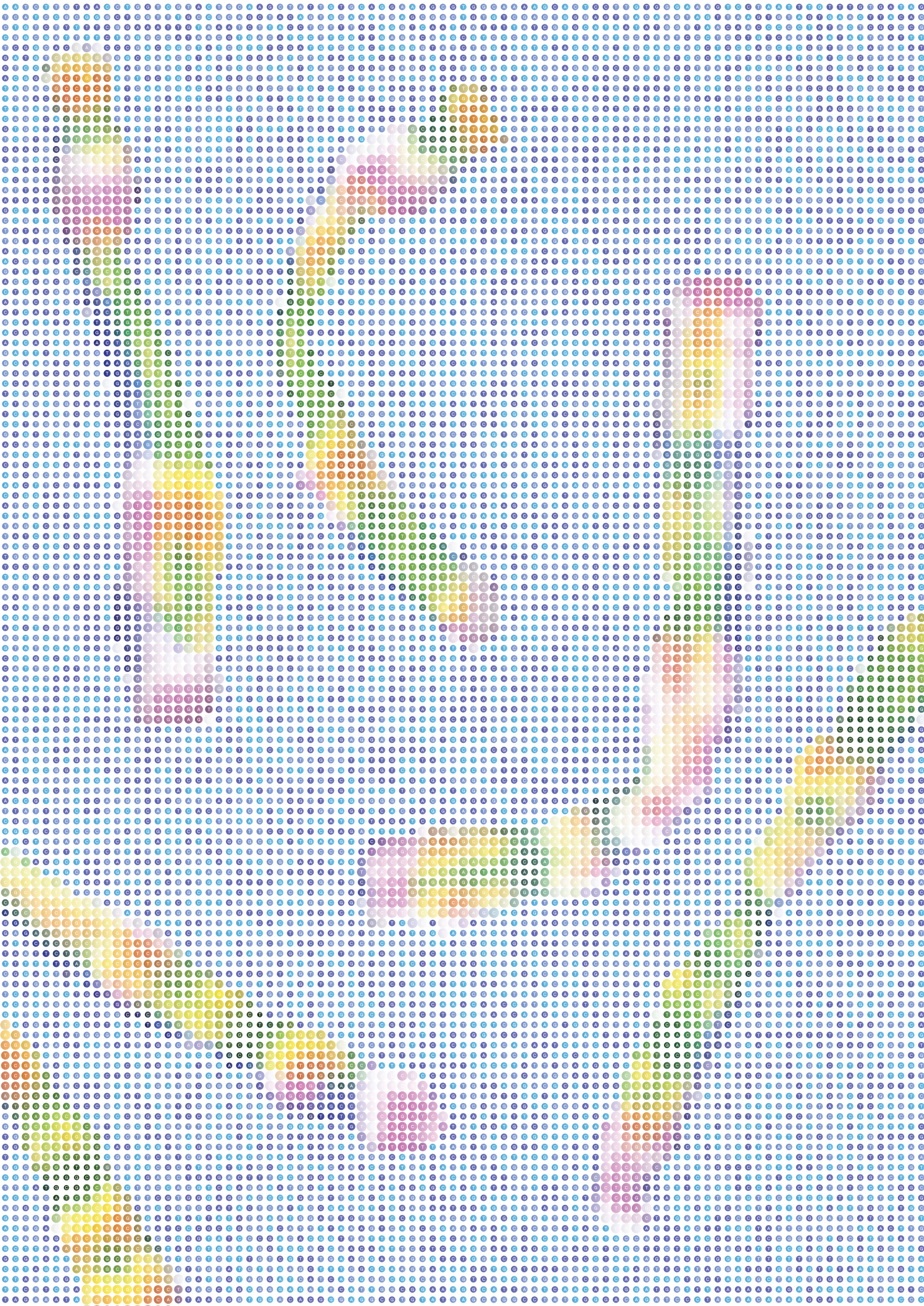
The algorithms created at CHIBI are being used by thousands of researchers around the world. With the increasing availability of these and other bioinformatics tools, Dr. Aliferis expects to see an "explosion of basic science and translational studies" that use big data to develop molecular signatures of disease, improve clinical studies, strengthen the cost-effectiveness of healthcare, and personalize patient care.



We are living in a world that is increasingly being linked digitally, and an incredible variety and volume of information is being stored online about anything that one can imagine," says Dr. Constantin Aliferis.

The ability to funnel trillions of data points into focused and well-informed inquiries could, in turn, help translate the unimaginably vast sea of information all around us into a transformative ideal of personalized care. In the section that follows, we highlight five examples of how researchers at NYU Langone Medical Center are using big data to ask weighty questions and make dramatic changes in medicine.

Featured Big Data Research



Learning to Read the Subtle Signs of Diabetes

> **ANN MARIE SCHMIDT, MD**
Professor of Endocrinology and Medicine

> **SAUL BLECKER, MD**
Assistant Professor of Population Health and Medicine

DAVID SONTAG, PHD
Assistant Professor of Computer Science at the Courant Institute of Mathematical Sciences

YANN LECUN, PHD
The Silver Professor of Computer Science, Neural Science, and Electrical and Computer Engineering at the Courant Institute of Mathematical Sciences

The numbers are sobering: roughly one-third of the estimated 25 million Americans with diabetes are unaware that they have the disease. Even with regular medical care, early signs of the disease may not be clear.

"The big picture is that the earlier a diagnosis of diabetes is made, the sooner you can work to reduce the high blood sugar, mitigate the acute complications, and start to prevent and slow down the long-term complications," says Ann Marie Schmidt, MD. For the vast majority of people with diabetes, she says, it's the long-term complications that eventually kill them.

The tools of big data are helping a team of collaborators from NYU's Courant Institute of Mathematical Sciences and NYU Langone Medical Center read diabetes' warning signs. The researchers are using a sophisticated computer-based mathematical formula to sift through millions of medical billing records in search of telling clues. Independence Blue Cross, the project's sponsor and a major insurer in the Philadelphia region, has provided extensive de-identified claims data for 4 million adult beneficiaries from 2006 to 2013.

"To put the size of the data in perspective, it's roughly the size of the nonimages part of Wikipedia," says project leader David Sontag, PhD, who has long been interested in using computer science to improve healthcare. He and Yann LeCun, PhD, are filtering the vast dataset through a tool called machine learning, in which a computer algorithm gets progressively better at analyzing the data and pointing out potential correlations.

Saul Blecker, MD, who has extensive experience working with large clinical datasets, is helping the team define the potential presence of diabetes or prediabetes through a variety of risk factors such as lab results, prescriptions, and billing codes. "The ultimate goal is to predict diabetes or to determine if the condition is present before it shows up in the data," Dr. Blecker says. By examining all of the data with a computer-based algorithm, he says, the machine learning method could uncover seemingly unrelated aspects of care that are actually connected, helping to improve diagnostic efforts. "I think that is the exciting piece of what we're doing," he says.

In July, the collaborators sent Independence Blue Cross their first round of predictions, and Dr. Sontag says the insurer's assessment of the model's accuracy and usefulness could help the team create progressively better versions. If it succeeds, the research could offer a blueprint for combing through billing and medical records to detect patients with other undiagnosed diseases. "The ability of big data and its analysis to really look far beyond the obvious can't be overstated," Dr. Schmidt says. "It's so important."

Dr. Schmidt and her colleagues have long studied the development of accelerated atherosclerosis, the leading cause of death in diabetics. A molecule known as RAGE, her group discovered, is abundant in arteries and other tissues at risk for diabetic complications and might contribute to damage by heightening a patient's inflammatory immune response. "For us, the ability to look at a large dataset in an unbiased way and try to pick out a RAGE-related phenomenon is worth its weight in gold," Dr. Schmidt says. Any signs of the molecule's involvement in damage to the arteries, nerves, kidneys, or other tissues could then be tested in mouse models of the disease.

"It's really an amazing opportunity," she says. "We're lucky that there is this spirit of cross-disciplinary collaboration, and that is a major strength at NYU."

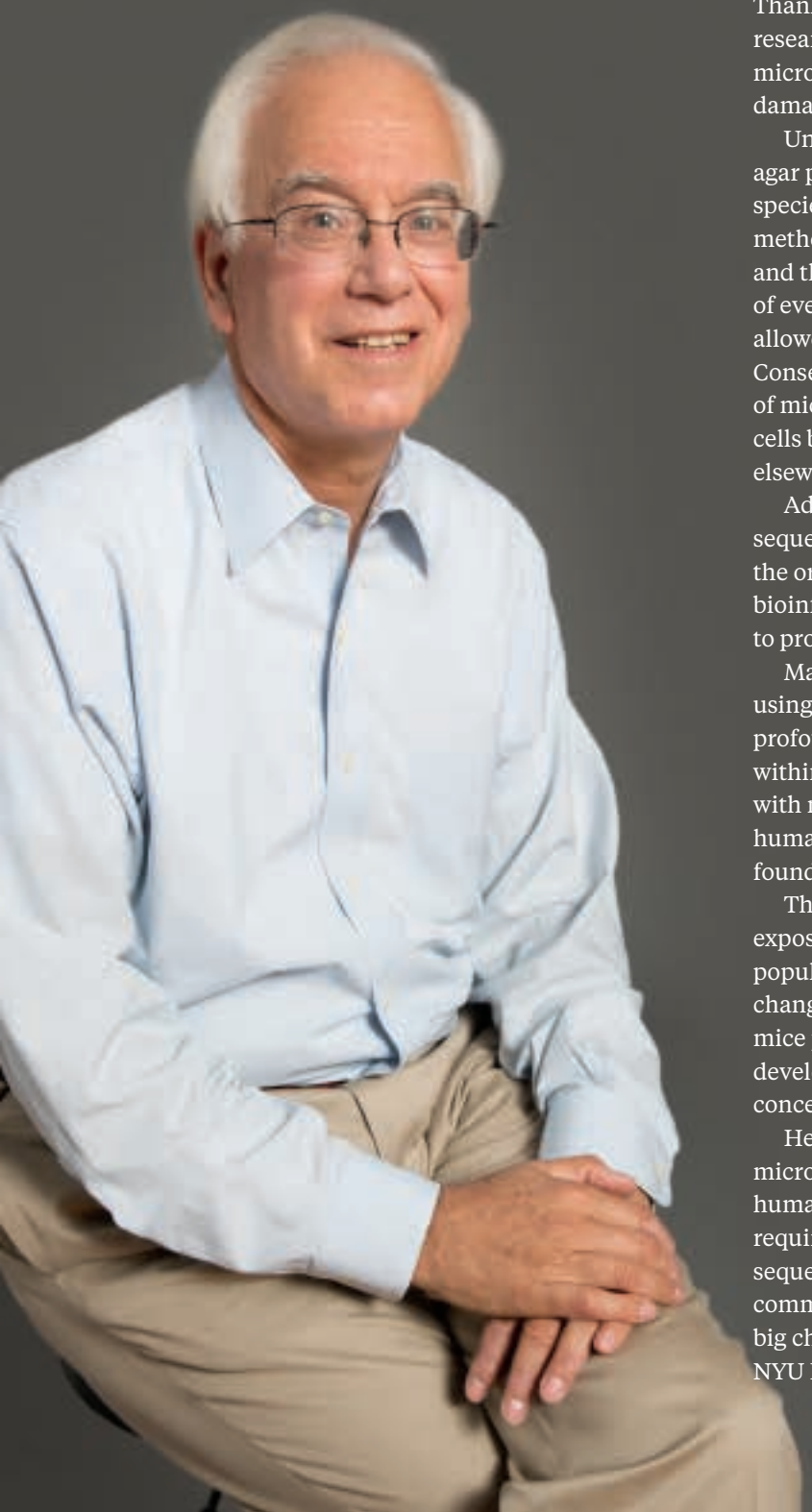
Nearly

80 Million

Estimated number of Americans with "prediabetes," meaning they have higher than normal blood glucose levels.

SOURCE: AMERICAN DIABETES ASSOCIATION

Linking Our Microbial Inhabitants to Health and Disease



In the Gut

Trillions of bacteria live on and in us, and we have co-evolved with these microorganisms for millennia. Thanks to an array of big data tools, an emerging field of research is now revealing that perturbing our resident microbial communities at critical time points can have damaging—and lasting—consequences.

Until recently, researchers had to culture bacteria on agar plates, a time-consuming process that limited the species they could identify. But a powerful molecular method that involves extracting DNA from the microbes and then sequencing the genes encoding a key component of every bacterium, called 16S ribosomal RNA, has allowed scientists to track all of our bacterial residents. Consequently, our microbiome—or the entire collection of microbes that share our bodies and outnumber our own cells by nearly 10 to 1—can be surveyed in the gut and elsewhere in unprecedented numbers.

Advances in what is known as next-generation sequencing have allowed researchers to rapidly determine the order of bacterial bits of DNA and RNA, and bioinformatics tools such as sequence aligners are helping to produce composite sequences of individual species.

Martin Blaser, MD, a pioneering scientist in the field, is using these tools to shed new light on how antibiotics can profoundly reshape the microbial communities that live within us. Quantifying microbial diversity is no easy task, with more than 1,500 species already associated with the human gut microbiome and roughly 500 species normally found within an individual's gut.

The animal research in Dr. Blaser's lab suggests that exposure to antibiotics early in life can alter the microbial population so markedly that it leads to obesity and changes in immune function. "Common treatments in mice produced alterations during that critical period of development," Dr. Blaser says. "That's why we're most concerned about early life antibiotics."

He now hopes to define the critical microbes and microbial pathways that must be present for normal human development. This big data endeavor will likely require combing through billions of bacterial DNA sequences and correlating variations in the microbial communities to changes in human development. It's a big challenge that calls for a collaborative effort with NYU Langone Medical Center's bioinformatics experts.

In the Lung

Researchers once thought that healthy people had sterile lungs. Thanks to research by Leopoldo Nicolas Segal, MD, and other scientists, the lung microbiome is now being eyed as a significant factor in respiratory health and disease.

Dr. Segal and colleagues are documenting microbial populations in normal lungs and changes within the lungs of patients who have conditions like chronic obstructive pulmonary disease, or COPD. Research has shown that 15 percent of people with a significant history of smoking will go on to develop COPD, but scientists don't know which factors determine who will be in that unfortunate group. Dr. Segal believes that differences in the lung microbiome may hold the key, and his collaborative effort with NYU Langone's Genome Technology Center is yielding some important clues.

"We do have evidence that when we do an intervention for COPD, there is a shift in the lung microbiome and a change in the inflammation patterns in the lung," he says. " This happens even in patients with very early disease."

Collecting bacterial DNA from the lungs to identify communities can be tricky, especially given the teeming microbes in the mouth and upper airway that can contaminate any sample. After sampling via a bronchoscope, a computer-based bioinformatics technique called source tracking is helping Dr. Segal's group and their collaborators in the Genome Technology Center pore over the dizzying amount of genetic data and discern which DNA samples derive from which niches in the body.

"That gives you terabytes of data out of a single patient," he says, "and if we're trying to make sense of a group of patients with a disease, that needs to be multiplied by several hundred patients."

So far, the hard work is paying off, as Dr. Segal and colleagues have associated unique microbial communities with increasing levels of lung inflammation. "The inflammation comes from our body's reaction, but there are functional aspects of the microbes that might be causing an increase in the inflammation," he says. "It gives you hope that we could find new targets for therapy."

< **MARTIN BLASER, MD**
The Muriel G. and George W. Singer Professor of Translational Medicine, Professor of Microbiology, and Director of the NYU Human Microbiome Program

LEOPOLDO NICOLAS SEGAL, MD
Assistant Professor of Medicine

In the Esophagus

Since 1975, the incidence of esophageal cancer in the United States has increased fivefold. This startling surge remains unexplained, but emerging research suggests that widespread use of antibiotics, and subsequent alterations in microbiota, may be to blame.

To investigate the phenomenon, pathologist Zhiheng Pei, MD, PhD, is focusing on changes within the bacterial community known as the esophageal microbiome. Like the lung, the esophagus was once thought to be devoid of microbes. But in 2000, Dr. Pei and colleagues began studying the microbes living in the esophagus of patients with gastroesophageal reflux disease and Barrett's esophagus, two cancer precursors. They discovered a distinct bacterial community and subsequently documented an increase in a specific type of microbe, known as gram-negative bacteria, in patients with the cancer precursors. "We see some changes that happen not only in the esophagus but also in the stomach and the oral cavity, or the foregut," Dr. Pei says.

His group is now pursuing a major study to determine whether and how these bacterial alterations may influence cancer. Changes in part of the esophagus might favor the excessive growth of bacteria that spur inflammation, for example, or that produce molecules promoting tumor formation.

"The microbiome has become such an interesting topic, and also a powerful tool to study the etiology of diseases that we don't know the cause of," he says. "This is different from the one-pathogen/one-disease concept, because we are dealing with a population." That means the researchers will have to pore over a "humongous" amount of DNA data from entire bacterial communities, he says. To piece together the microbiome of a patient's esophagus via a technique called shotgun metagenomic sequencing, the researchers may need to sequence up to 300 million fragments of DNA. In other words, they will sample every gene within every microorganism.

What's more, to track patients over time, the entire process may have to be repeated multiple times. Understanding the changes in these populations, though, may help doctors classify a patient's microbiome as normal or abnormal and devise ways to correct the trouble before it leads to disease.

< **ZHIHENG PEI, MD, PHD**
Associate Professor of Pathology and Medicine

> **DAN LITTMAN, MD, PHD**
The Helen L. and Martin S. Kimmel Professor of Molecular Immunology and a Howard Hughes Medical Institute Investigator

In the Immune System

Our gut-dwelling microbial residents have long been associated with helping us break down food and supplying nutrients. More unexpectedly, some of these normally benign, or commensal, microbes may also shape the development and function of specific parts of our immune system, according to pioneering research by Dan Littman, MD, PhD. "Basically, there are commensal bacteria that have a dialogue with our immune system," he says.

In mice, Dr. Littman has shown how the presence of specific gut microbes known as segmented filamentous bacteria, or SFB, control the fate of immune specialists known as T helper cells. In humans, he and his colleagues have discovered that a bacterial species called *Prevotella copri* exists in the gut flora of about three-fourths of patients with newly acquired rheumatoid arthritis, but only in about one-fifth of healthy volunteers. One hypothesis is that *Prevotella* may behave like SFB by activating specific T cells that promote inflammation and provoke autoimmune disease.

Understanding how these bacteria exert their surprising influence on our defense network could give researchers new insights into how to fortify the barricades. It may also help them prevent inadvertent autoimmune attacks from overactive immune cells that can lead to rheumatoid arthritis and other diseases. But teasing apart the influence of individual microbial components could present a considerable challenge.

For example, the SFB microbe alone contains 1.6 million letters of DNA and more than 1,400 genes, and it's just one of many bacteria that may play a role in immunity. Moreover, the sheer variability of separate bacterial communities living within individuals is nothing short of astounding. "If you go just a couple of inches away within the gut, you may have a completely different type of community," Dr. Littman says.

Tools that manipulate specific bits of the microbiome, such as molecular cloning methods that can remove a gene from one bacterium and put it into a different one so researchers can determine its effect, he says, may help unravel the complicated web of direct or indirect communication among the microbes and our own cells. Dr. Littman's lab is also raising mice in sterile conditions and then colonizing them with well-defined bacterial communities to determine how they interact with the animals. "These are complex ecological problems that are going to take us a lot of time and ingenuity to resolve, but it's an exciting challenge," he says.

300 Million

The number of DNA fragments that may need to be sequenced to piece together the microbiome of a single patient's esophagus, using a technique called shotgun metagenomic sequencing.

Building a Better Map to Block Leukemia's Escape Routes

JINHUA WANG, PHD
Associate Professor of Pediatrics and Associate Director of Biomedical Informatics

Imagine that a killer is on the loose and can use multiple pathways to evade capture. For thousands of children every year, that killer is acute lymphoblastic leukemia, or ALL. Improved chemotherapy and other treatments have boosted the cure rate to about 80 percent. For the remaining 20 percent of kids, however, the blood-borne cancer somehow escapes and triggers a relapse with a bleak prognosis.

William L. Carroll, MD, has devoted his career to this and other kinds of pediatric leukemia. "For the bulk of that time, we've developed new ways to treat the cancer without having any understanding of what was under the hood," he says. Two children of the same age might appear to share the same risk factors and type of disease. "Yet we couldn't predict how one was going to do versus the other."

The global effort to sequence the DNA that makes up the entire human genome, however, gave researchers a fresh infusion of genetic information. In addition, with the increased ability to identify the RNA molecules transcribed from that DNA, or the transcriptome, they began to understand the tremendous diversity that lay hidden just beneath the surface. As they learned, no two leukemia patients are alike.

In his quest to uncover leukemia's varied tactics, Dr. Carroll joined forces with Jinhua Wang, PhD, a trained physicist and participant in the 13-year Human Genome Project that assembled our complete genetic blueprint in 2003. (Today, the sequencing machines in NYU Langone Medical Center's Genome Technology Center can sequence three human genomes in just 11 days.) From 1998 to 2007, Dr. Wang worked at the Chinese Academy of Sciences in Beijing and St. Jude Children's Research Hospital in Memphis, Tennessee. During this time he helped piece together the genomes of influenza virus strains and a pathogenic form of the *Escherichia coli* bacterium. "I'm interested in the underlying mechanism, or machinery, of how these diseases work," he says.

Dr. Wang has since set his sights on another demanding goal, analyzing the sequence information from the genomes of cancerous cells to reveal how accumulated mutations may help drive ALL. At NYU Langone, his

(continued on page 32)

Finding the Best Therapeutic Fit for Breast Cancer

DAVID FENYO, PHD Associate Professor of Biochemistry and Molecular Pharmacology

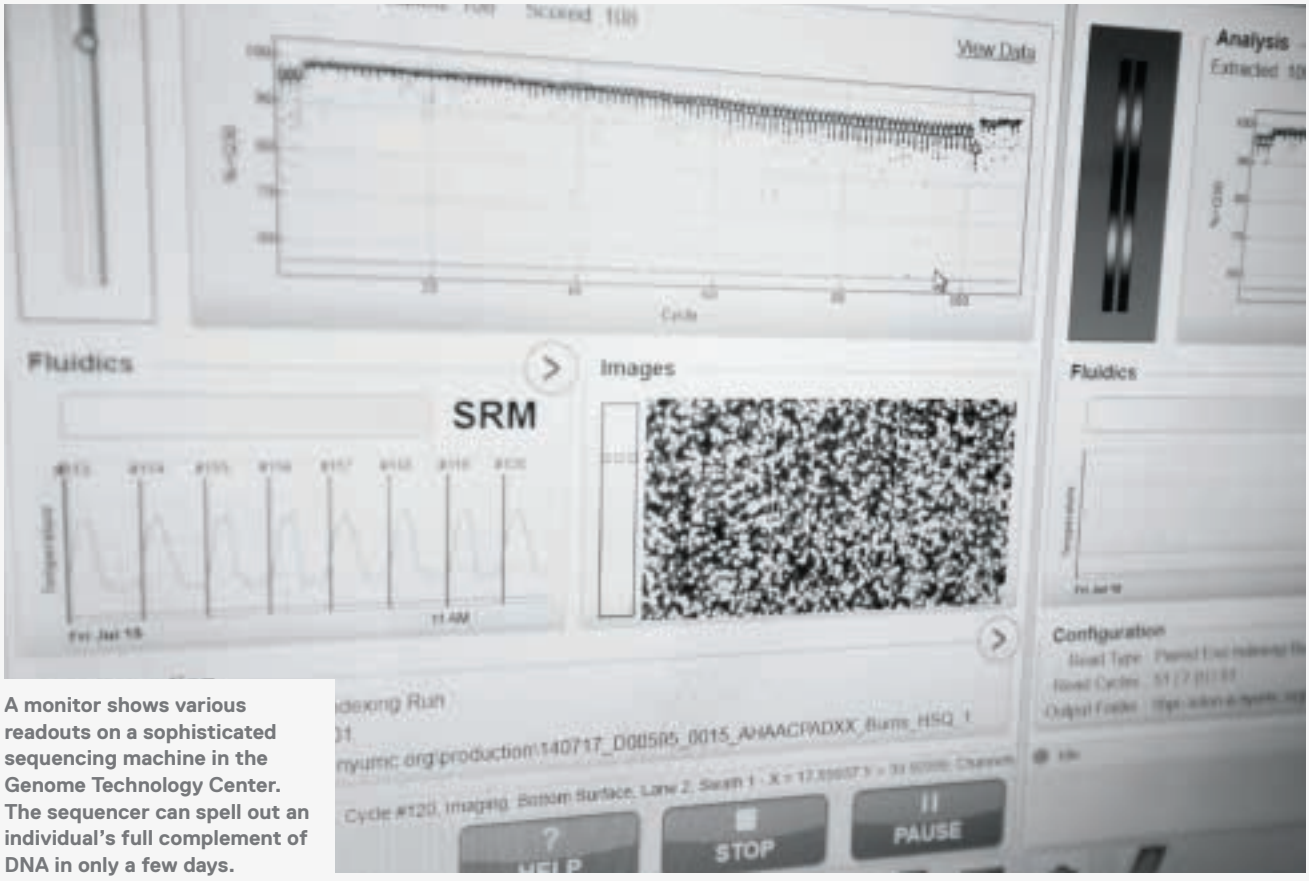
For some patients with "triple-negative" breast cancer, so called because three specific genes have been switched off in the tumors, the cancer responds to chemotherapy. For other patients with the same genetic profile, nothing seems to work.

To better understand what's behind the stark difference in prognosis, David Fenyo, PhD, has teamed up with researchers at the Broad Institute in Cambridge, Mass., and Washington University in St. Louis. The researchers are surveying tumor samples for a wide range of mutations and investigating how massive molecular datasets such as a tumor's genome (full set of DNA), transcriptome (sum total of RNA), and proteome (entire body of proteins) all change in response to different drugs. Beyond high-throughput sequencing, the effort requires a suite of sensitive tools such as microarrays to determine gene activity and mass spectrometry to identify subtle differences in proteins.

In addition, the collaborators are gathering information from patient tumors grafted into mice to learn how those tumors may react to specific kinds of chemotherapy. "What excites me is that you can collect different types of data," Dr. Fenyo says. "When you combine them, you can find out what really matters, what's driving a disease."

Merging everything into a coherent whole can be a daunting process. "You can easily drown in these large amounts of data," says Dr. Fenyo, a physicist by training. With his expertise in computational proteomics, which uses computer-based methods to identify, measure, and characterize proteins as a way of understanding their role in cellular processes, he is helping his colleagues home in on the best information. Combining tools like mass spectrometry for identifying proteins with computer models, and simulations that test how those proteins behave in a cell, for example, helps Dr. Fenyo sift through the sea of information.

Ultimately, the joint effort may reveal how to use altered proteins or other molecules as biomarkers that point to the right chemotherapy for triple-negative breast cancers or other specific tumors.



A monitor shows various readouts on a sophisticated sequencing machine in the Genome Technology Center. The sequencer can spell out an individual's full complement of DNA in only a few days.

80–100 Million

The average number of sequencing "reads" per sample of RNA, with each read consisting of about 50 to 75 base pairs. This enormous number allows researchers to pinpoint the locations in a genetic blueprint where the RNA may be plentiful or scarce, suggesting higher or lower gene activity, respectively.

(continued from page 30)

bioinformatics expertise has provided the perfect complement to Dr. Carroll's clinical experience, and the two have formed a fruitful collaboration. "The technology is now providing an opportunity for researchers to get as much information as possible to attack the disease," Dr. Wang says. "I feel like this is very exciting, and also it's very challenging."

As information in a cell flows from DNA to its chemical cousin, RNA, which in turn becomes the template for protein assembly, sophisticated methods such as next-generation sequencing, chromatin immunoprecipitation sequencing, and mass spectrometry are required to access the data stored within each kind of molecule. A growing suite of computer-based analytical tools, such as statistical packages with mathematical models to search for correlations in high-throughput genomic data, are helping researchers fit everything together. "We can collect all of this data to explore the whole universe of information related to ALL, and try to put it all into one picture," Dr. Wang says.

Dr. Carroll, Dr. Wang, and colleagues began assembling the profile of ALL with a painstaking analysis of bone marrow samples from 10 pediatric patients. The medical researchers pieced together a full sequence of the extracted RNA molecules—initially at the time of diagnosis and again when each patient relapsed. This transcriptome analysis, in collaboration with the NYU Langone Genome Technology Center, offered them a view of all active genes within the leukemia cells, while a computer-based analysis designed by Dr. Wang pointed out anomalies.

The team then began layering on additional information. How might physical or chemical modifications to the DNA impact that gene activity, for example? Are any critical genes deleted or amplified? What's the role of RNA that doesn't encode proteins but may regulate gene activity? "With each new technology, we've added another layer of complexity onto the original canvas, but this is the only way to really understand the circuitry in detail," Dr. Carroll says.

The emerging picture suggests that a very small subset of leukemia cells can harbor genetic changes that spare them from chemotherapy. Their survival advantage enables these cells to proliferate and eventually dominate, rendering the drug therapy useless.

The research also suggests that the leukemia cells may use a variety of methods to evade drugs. While these strategies may take advantage of multiple mutations, they seem to fall into a few specific pathways. By identifying these routes, researchers may be able to block the cancer's progression. "It's like the subway," Dr. Carroll says. "As long as you stop the train before it gets to its destination, you will inhibit it."

Bioinformatics, Dr. Wang says, can supply something akin to a multilayered Google map that shows not only the main subway routes but also surrounding structures and traffic patterns. With that additional context, researchers can get a more complete view of the most critical links and learn which combination of interventions might work best to block all forward progress.

So far, the researchers have implicated at least four dominant routes that can promote relapse among ALL patients. Melanoma cells exploit one of these same pathways as part of their own survival strategy, suggesting that a drug called a MEK inhibitor that blocks the melanoma escape route might also work well in thwarting ALL drug resistance. That strategy may require years to reach fruition, but the accumulating data have at least pointed researchers in the right direction. "We finally know the enemy," Dr. Carroll says.

WILLIAM L. CARROLL, MD
The Julie and Edward J. Minskoff Professor of Pediatrics and Director of the NYU Cancer Institute

Distilling Big Data into Better Drug Design

As a computational structural biologist, Timothy Cardozo, MD, PhD, has spent hours poring over three-dimensional models of proteins and their interactions with drugs and other molecules. Doing so, he says, can give real "Aha!" flashes of insight when the activity of a drug makes sense given its target in the human body.

For a massive new project aimed at improving drug design, the potential flashes could be dazzling indeed. Dr. Cardozo and his colleagues are using computer-based methods to match all known chemical compounds with their potential binding partners in the human proteome, or our entire collection of proteins.

The team has designed sophisticated computer formulas that use the structure of a compound to seek out its potential binding sites in protein nooks and crannies. "For every single drug, it's basically searching through every single place it could bind in proteins expressed by the human genome," Dr. Cardozo says. The better the fit, the higher the binding, or docking, score.

So far, the group has analyzed and rated more than 9 billion of these drug-protein interactions. Crucially, they are also assessing the activity of these drug targets in tissues where they may play a key role in disease. The researchers have dubbed this combined analysis of drug binding and receptor levels "HistoReceptomics."

One outcome of the research is a free, user-friendly website called drugable.org. A high HistoReceptomics score on the website suggests that a protein target has high affinity for a potential drug and is present in the right parts of the body, helping drug developers narrow their focus to interactions that merit a closer look.

"Our informatics design does two important things: it reduces the complexity of the big data, and it makes it more relevant to current knowledge in medicine," Dr. Cardozo says. More accurate matching of drugs to proteins, in turn, may aid in drug discovery.

Dr. Cardozo draws upon his years of expertise in analyzing structure-activity relationships. "Each one is like a case study," he says. From a structural analysis of the human immunodeficiency virus, for example, he helped design a vaccine candidate based on a protein fragment that mimics a vulnerable part of a key HIV protein. Based on that vaccine's structure, he realized he could also incorporate a second molecule mimicking cocaine, resulting in an experimental combination therapy.

A major goal of his new effort is to apply the enormous volume of structural information to one of medicine's most important numbers: one. "What excites me," he says, "is the potential to eventually personalize drugs for an individual's disease."

TIMOTHY CARDOZO, MD, PHD
Associate Professor of Biochemistry and Molecular Pharmacology

Pondering a Big Influx of Ethical Questions

Gene therapy. Cloning. Organ transplants. During his 35-year career as a bioethicist, Arthur Caplan, PhD, has weighed in on the ethical implications of a broad spectrum of medical issues. The new era of big data, he says, will ensure that he and his colleagues remain busy for the foreseeable future.

One ethical obstacle, he says, will be avoiding the perils of promising too much about what this burst of data can deliver. "We saw it with the mapping of the human genome: it took a good 15 years to start to generate tests and treatments from this gusher of genetic information," Dr. Caplan says.

Maintaining patient privacy and securing informed consent will also present challenges, particularly given the potential to discover unanticipated risk factors for diseases like cancer. Institutions such as NYU Langone, Dr. Caplan says, can help lead the way in establishing new standards such as emphasizing transparency, safeguarding patient information and involving patient advocacy groups early in the decision-making process about how to use datasets.

Bioethicists and others will need to wrestle over the right balance between risks and rewards, Dr. Caplan says. Even so, he says, the growing ability to mine big datasets may give researchers an unprecedented means to understand the basis of disease, help patients comply with medical interventions, and track a population's health. "I think there are some huge benefits and very exciting opportunities, and they have to be pursued because that's what the public and patients expect of academic health centers like NYU," he says.

ARTHUR CAPLAN, MD
Drs. William F. and Virginia Connolly Mitty Professor and head of the Division of Bioethics

100 Million

Estimated number of processor-hours needed by NYULMC and Google Inc.'s supercomputers to perform the largest of Dr. Cardozo's multiple protein-drug binding analyses.

Revealing the Warning Signs of Sudden Cardiac Death

> **MARIO DELMAR, MD, PHD**
Professor of Medicine and Cell Biology

ELI ROTHENBERG, PHD
Assistant Professor of Biochemistry and Molecular Pharmacology

> **MARINA CERRONE, MD**
Research Assistant Professor of Medicine

DAVID FENYO, PHD
Associate Professor of Biochemistry and Molecular Pharmacology

A 20-year-old soccer player in top form suddenly collapses on the field and dies. Why?

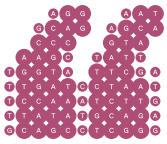
Medical sleuths have linked many tragic cases like this one to an inherited heart disorder known as arrhythmogenic cardiomyopathy, which causes a major electrical disturbance in the heart's normal rhythm. However, this notorious killer of athletes and young adults can strike with little warning, and frustrated clinicians have been unable to say with certainty who may be next.

A serendipitous meeting between cardiology researcher Mario Delmar, MD, PhD, and physicist Eli Rothenberg, PhD, has led to a major research collaboration that could be a game-changer. Their innovative approach to a data-rich microscopy method known as superresolution optical microscopy may help doctors chart out the true risk of individuals who carry genetic mutations linked to the heart disease.

The most common of these mutations affect a protein structure that glues our heart cells together, giving them their needed elasticity when they collectively contract and then relax during every heartbeat. "The heart has to be very flexible," says Dr. Delmar. "It stretches and contracts thousands of times each day throughout our life." For years, though, researchers puzzled over why a defect in this gluelike protein complex also caused electrical dysfunction in the heart.

When Dr. Delmar explained the research problem to Dr. Rothenberg at a gathering for new faculty, the researchers soon realized that their seemingly disparate interests might offer a synergistic solution. Dr. Rothenberg had developed a superresolution microscopy technique that uses fluorescent light and antibodies to overcome normal resolution limits and point out the center of individual molecules such as proteins. Whereas more traditional optical microscopy barely resolves images as small as 250 nanometers—about 1/30th the size of a red blood cell—Dr. Rothenberg's method can accurately resolve images as minuscule as 20 nanometers. "Instead of a blurred image, we get really refined details of where the proteins are and what the sizes of the protein clusters are," he says.

To chart the relative locations of multiple proteins, his technique takes movies of the cell, with each frame providing information about a subset of molecules. Condensing all of these frames into a single image can be a daunting feat, considering that a basic picture of a protein



The heart has to be very flexible," says Dr. Delmar. "It stretches and contracts thousands of times each day throughout our life."

complex may represent 2,000 separate movie frames, each taken 33 milliseconds apart. The visual data can quickly fill up a terabyte-size memory disk. Combined, however, it can provide a revealing three-dimensional reconstruction of how specific proteins connect.

Dr. Delmar compares the task of piecing together these multipart protein structures within heart cells to the task of properly assembling IKEA furniture. "The cardiac cell is extremely dependent not only on having the right catalog of proteins, but also on having them in exactly the right spot and having them arrive at exactly the right time," he says. With the cutting-edge microscopy technique—based on adaptations to a custom-built, single-molecule fluorescence microscope—the researchers were able to point out small deviations where the assembly process had gone awry due to misshapen, late-arriving, or otherwise errant proteins.

When certain proteins ended up more than 40 nanometers apart, for instance, they lost their ability to join up in a way that effectively transmitted the heart's electrical impulses. Or as Dr. Delmar says, everyone is mere nanometers from sudden death.

Through multiple rounds of experimentation and refinement, the researchers demonstrated the superresolution microscopy method's ability to zero in on proteins and have since applied for a patent. "We had these data that no one saw before us," Dr. Rothenberg says. "It's very exciting. I never thought I'd be doing something like this."

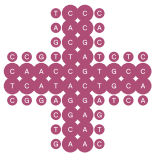
The project has required close collaboration with other experts as well, particularly given the challenge of combining thousands of images into each three-dimensional reconstruction. To help make sense of the enormous output of high-resolution microscopy data and to construct a meaningful composite map, the collaborators enlisted the analytical expertise of David Fenyo, PhD.

The joint effort has shown that three distinct protein complexes in a specialized region of heart cells work together to control how well the cells stick together, communicate, and pass along electrical impulses. Dr. Delmar calls these many interconnected proteins the connexome. "What you have is really a community of proteins that together achieves the function of holding the cells together and communicating among them," he says.

The group also works with Marina Cerrone, MD, who has long studied patients with heart arrhythmias and lends her clinical expertise. Despite the outpouring of information from the human genome and other sources, Dr. Cerrone says a major conundrum for clinicians has been a lack of knowledge about how specific genetic mutations may influence a patient's risk. "With the type of research that we are doing, we are hoping to at least put a little piece in the mosaic or puzzle to say that someone is more at risk than another," she says.

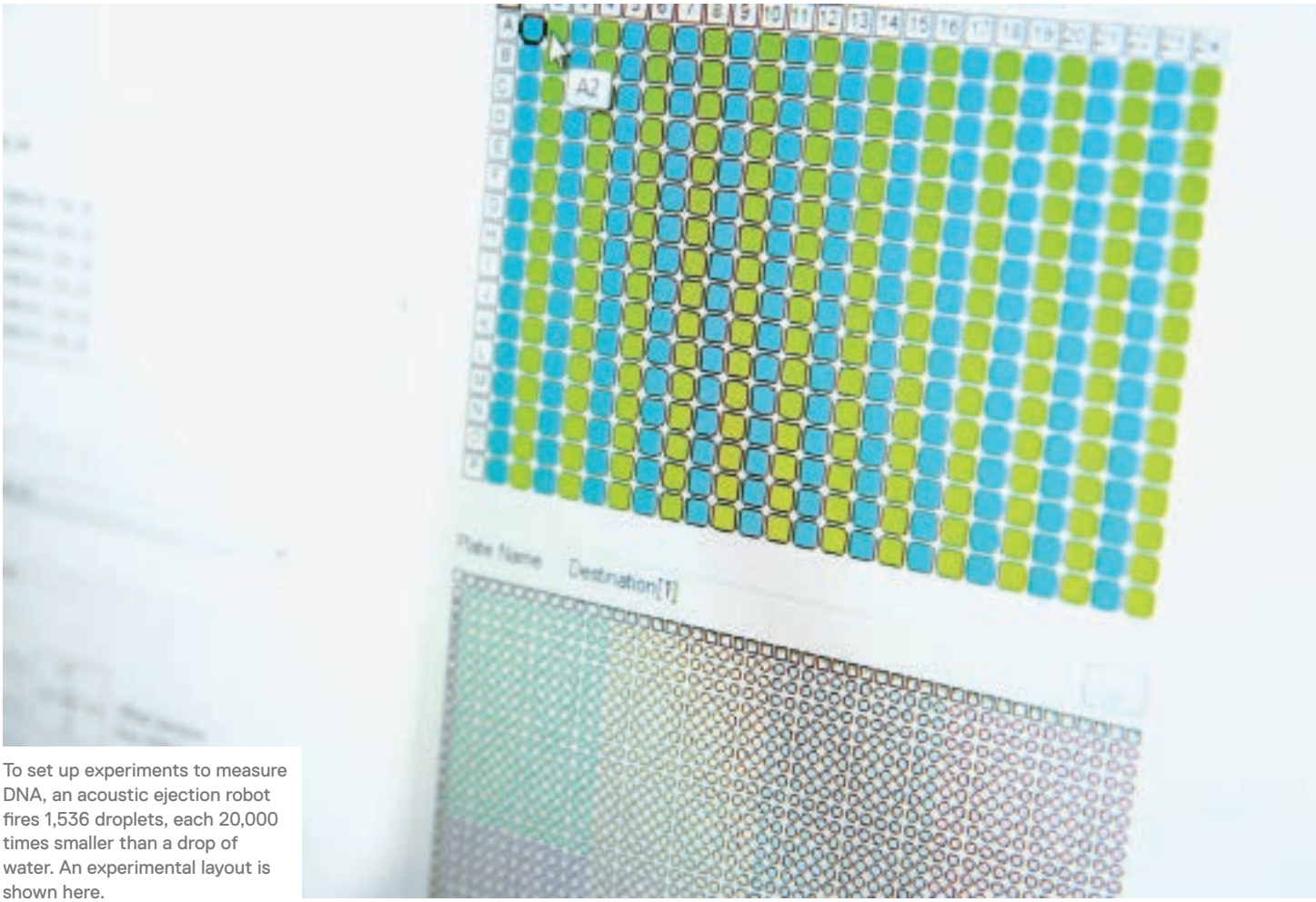
Based on their success, the collaborators are now embarking on an even more ambitious project. From blood samples of patients who may be at risk for arrhythmogenic cardiomyopathy, the researchers plan to create stem cells, or blank slates that can become a wide range of cell types. These cells can then be coaxed to develop into cardiac cells and assessed by the superresolution microscopy method for any signs of dysfunctional protein assembly. For this part of their project, they are working with Lei Bu, PhD, assistant professor of medicine, an expert in the generation of stem cell–derived cardiac cells.

"If I am inspecting a building because it is in structural danger, I don't want to find a wall already crumbled on the ground," Dr. Delmar says. "I want to find a crack on the wall." If this method reveals the telltale signs of danger in a patient's own cardiac cells, doctors could use it to test their efforts to stabilize or reinforce the compromised protein structure, potentially averting another needless tragedy.



10,000

The average number of heartbeats per day. When the heart skips a beat, stutters, or veers away from its steady expansion and contraction, the arrhythmias can be life threatening.



To set up experiments to measure DNA, an acoustic ejection robot fires 1,536 droplets, each 20,000 times smaller than a drop of water. An experimental layout is shown here.

Big Datasets Lead to Bold Questions

At NYU Langone Medical Center, a solid infrastructure, advanced analytics, and a collaborative approach are spurring enterprising questions about disease and health that would have been unthinkable just a few years ago.

How, for example, can medical billing claims help detect undiagnosed diabetes? How can shifting populations of microbes dwelling on and within us help predict diseases like obesity, rheumatoid arthritis, and chronic obstructive pulmonary disorder? What can a detailed portrait of gene activity teach us about childhood cancers and relapse? How can assessments of protein-drug interactions help researchers design better, safer medications? How can the precise locations of altered proteins in the heart warn of sudden death due to a cardiac arrhythmia?

"The trick is not to let big data just sit there but to be converting it to information and knowledge and better decisions that improve health," says R. Scott Braithwaite, MD, professor of population health and medicine, and director of the Division of Comparative Effectiveness and Decision Science.

"A promise of big data is to be able to offer more-finely-grained treatment that's better fitted to an individual's characteristics," says Dr. Braithwaite.

Fulfilling that promise means first using tools like bioinformatics to turn data into information and then harnessing advanced research tools like decision analysis to turn the information into carefully calibrated decisions based on the relative risks and benefits.

At NYU Langone, in other words, experts are using new methods to prioritize how to marshal the Medical Center's resources and analyze the information that may deliver the biggest bang for the buck.

Facts and Figures

Honors

Howard Hughes Medical Institute Investigators

07

Institute of Medicine Members

10

National Academy of Sciences Members

10

American Academy of Arts and Sciences Members

09

American Association for the Advancement of Science (AAAS) Fellows

17

Researchers & Faculty

Researchers (Includes Physician-Scientists)

400

New Faculty in Calendar Year 2013

32

Postdoctoral Fellows

417

Students

MD/PhD Students

75

PhD Students

296

PhD Recipients in 2014

34

Facilities & People

MD/PhD Students

220

Square Feet of Research Space

550,000

Countries Represented in Research Labs at NYULMC

55

Published Research

Original Papers, Commentaries, Reviews, and Other Material by Our Researchers That Appeared in the Science and Medical Literature in Calendar Year 2013

4,168

Publications That Had an Impact Factor of at Least 10

371

Funding

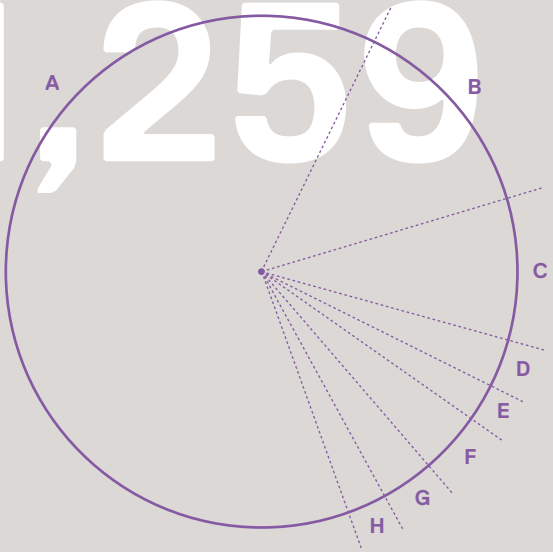
2013 Grant Revenue

$249,854,000

FY2013 Awards by Source

Total Number of Awards

1,259

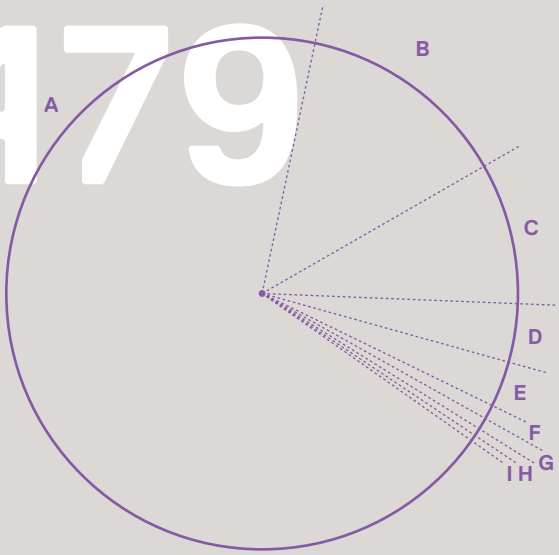


A	NIH	63%*
B	Nonfederal	13%
C	Federal: Non-NIH	9%
D	NIH: Subcontract	4%
E	Federal Non-NIH: Subcontract	3%
F	State/Local	3%
G	Industry/For Profit	2%
H	NonFederal: Subcontract	2%

FY2013 NIH Awards

Total Number of Awards

479



A	Research Projects	68.6%
B	Research Program Projects and Centers	13.5%
C	Cooperative Agreements	8.8%
D	Training Programs	3.6%
E	Research Career Programs	2.9%
F	Fellowship Programs	0.9%
G	Research-Related Programs	0.7%
H	Institutional Training & Director Program Projects	0.5%
I	Contract	0.3%

*PERCENTAGES BY DOLLAR AMOUNTS

Our Philanthropic Leadership

New Nonfederal Funding of $100,000 and Above*

A special thank-you to Fiona and Stanley Druckenmiller, Alexandra and Steven Cohen, Helen L. Kimmel, Leonard Litwin, Laura and Isaac Perlmutter, The Skirball Foundation, Joan and Joel Smilow, and Marica Vilcek and Jan Vilcek, MD, PhD, for their ongoing philanthropic investments in research.

Rita Allen Foundation

Alzheimer's Association

The Alzheimer's Drug Discovery Foundation

American Brain Foundation

American College of Phlebology

American Heart Association

Arnie's Place Foundation

Timur Artemyev

Avon Foundation for Women

Arnold and Mabel Beckman Foundation

Blavatnick Family Foundation

Brain and Behavior Research Foundation

Breast Cancer Alliance, Inc.

Nancy E. Bronstein

Steven & Alexandra Cohen Foundation

Sean and Susan Cullinan

Annette P. and Ian M. Cumming

Deutsche José Carreras Leukämie-Stiftung e.V.

The Dysautonomia Foundation, Inc.

Ellison Medical Foundation

The Enoch Foundation

F. Hoffman-La Roche Ltd.

William E. Flaherty

Samuel and Judith Florman

The Ralph S. French Charitable Foundation Trust in Memory of Ralph S. French and Louis and Herbert French

The Foundation Fighting Blindness, Inc.

Heffter Research Institute

Human Frontier Science Program

The Iacocca Family Foundation

The Irma T. Hirschl Trust

Hyundai Hope on Wheels

Kevin and Masha Keating Family Foundation

Sidney Kimmel Foundation for Cancer Research

The Knapp Family Foundation

Susan G. Komen for the Cure

Estate of Helen G. Koss

James D. and Marjorie Kuhn

Leon Levy Foundation

Lupus Research Institute

The Lustgarten Foundation

Making Headway Foundation, Inc.

Edward Mallinckrodt Jr. Foundation

Estate of Irwin Mandel

Estate of Estelle A. Manning

March of Dimes Foundation

Marc Jacobs International, L.L.C,

myFace

National Multiple Sclerosis Society

New York Stem Cell Foundation

Florence and Joseph P. Ritorto

Damon Runyon Cancer Research Foundation

William and Sylvia Silberstein Foundation

Laura Baudo and Robert F.X. Sillerman

Claudia and Kenneth Silverman

The Simons Foundation

Marica and Jan Vilcek

Whitehall Foundation, Inc.

Anonymous (3)

* NEW GIFTS AND PLEDGES MADE OR RECOMMENDED IN FISCAL YEAR 2013: SEPTEMBER 1, 2012, TO AUGUST 31, 2013.

Leadership

New York University

MARTIN LIPTON, ESQ
Chairman, Board of Trustees

JOHN SEXTON, PHD
President

ROBERT BERNE, PHD
Executive Vice President for Health

NYU Langone Medical Center

KENNETH G. LANGONE
Chairman, Board of Trustees

ROBERT I. GROSSMAN, MD
The Saul J. Farber Dean and Chief Executive Officer

Executive Leadership Team

STEVEN B. ABRAMSON, MD
Senior Vice President and Vice Dean for Education, Faculty, and Academic Affairs

DAFNA BAR-SAGI, PHD
Senior Vice President and Vice Dean for Science, Chief Scientific Officer

BERNARD A. BIRNBAUM, MD
Senior Vice President and Vice Dean, Chief of Hospital Operations

ANDREW W. BROTMAN, MD
Senior Vice President and Vice Dean for Clinical Affairs and Strategy, Chief Clinical Officer

MICHAEL T. BURKE
Senior Vice President and Vice Dean, Corporate Chief Financial Officer

ANNETTE JOHNSON, JD, PHD
Senior Vice President and Vice Dean, General Counsel

JOSEPH LHOTA
Senior Vice President and Vice Dean, Chief of Staff

VICKI MATCH SUNA, AIA
Senior Vice President and Vice Dean for Real Estate Development and Facilities

NADER MHERABI
Senior Vice President and Vice Dean, Chief Information Officer

NANCY SANCHEZ
Senior Vice President and Vice Dean for Human Resources and Organizational Development and Learning

RICHARD DONOGHUE
Senior Vice President for Strategic Planning and Business Development

KATHY LEWIS
Senior Vice President for Communications and Marketing

GRACE KO
Vice President for Development and Alumni Affairs

Science and Research Administration

LAURA AHLBORN
Vice President for Science Strategy

BRUCE CRONSTEIN, MD
Codirector, NYU-HHC Clinical and Translational Science Institute

ANNY FERNÁNDEZ
Senior Director for Financial Affairs and Administration

GREGG FROMELL, MD
Vice President for Science Operations and Transformation

JUDITH HOCHMAN, MD
Senior Associate Dean for Clinical Sciences and Codirector, NYU-HHC Clinical and Translational Science Institute

DAVID LEVY, PHD
Associate Dean for Collaborative Science

KEITH MICOLI, PHD
Director, Postdoctoral Program

ROBERT SCHNEIDER, PHD
Associate Dean for Therapeutics Alliances

NAOKO TANESE, PHD
Associate Dean for Biomedical Sciences and Director, Sackler Institute of Graduate Biomedical Sciences

ANGUS WILSON, PHD
Assistant Dean for Research Laboratory Operations and Facilities