BIG DATA

THE NEW MATH IN PATIENT CARE

Precision Medicine Refines its Calculus with Emerging Analytical Tools and Innovative Study Designs

By Alla Katsnelson

It was an undergrown 7-year-old boy who kindled Karina Davidson's epiphany, some 25 years ago, about the value of truly personalized medicine. The child weighed just 37 pounds, within the lowest percentile range for boys his age; he was not only failing to thrive, but also intensely violent toward himself and others. Dr. Davidson—then an intern in clinical psychology at the University of Waterloo in Canada—and the rest of the boy's clinical team struggled with how to care for him. Might megadoses of Ritalin temper his behavior? And what drove his rages? The team brainstormed hypotheses: Poor sleep? A lack of structure in his environment? Insufficient calories? His interactions with his mother?

To find out, they designed a randomized, controlled trial—a stretch of Ritalin followed by a stretch of placebo—in which the boy was the sole subject and his own time-lapsed control. Dr. Davidson asked the practice's pharmacist to formulate a Ritalin lookalike placebo. She had a statistician from the university create randomization codes to blind the clinical staff, the boy, and his family to which treatment he was receiving. The team hand-recorded endless checklists to track his acting out against sleep, diet, social interactions, and other factors, then entered the observations into a computer for analysis. It was labor-intensive, but the approach worked. Ritalin, they learned, was not working. It suppressed his appetite and did nothing to reduce his violent and suicidal episodes. And his aggression seemed to be triggered by the dynamic with his mom. Based on these hard-won data, so specific to this one child, Dr. Davidson and her colleagues devised a treatment plan that helped him gain weight and rein in his emotional turbulence.

"The lesson that has stayed with me for a lifetime is that excellent clinical care should be informed by science, but it's best informed by the patient in front of you," says Dr. Davidson, director of Columbia's Center for Behavioral Cardiovascular Health. It is what she calls "precision therapeutics"—pinpointing the treatment that will help a specific person, rather than starting with what might benefit the average person and making guesses from there.

Her intern experience cemented Dr. Davidson's vision for excellence in the practice of medicine, but she has had to wait to see its potential begin to be realized. Only in the past few years have genome sequencing and other molecular techniques advanced enough to allow researchers to identify precise molecular differences between people and to begin to understand how those differences can be targeted therapeutically. Meanwhile, personal health tracking devices and mobile phone

Photographs by Jörg Meyer

Karina Davidson, PhD

apps can now seamlessly capture and digitize blood pressure, sleep patterns, mood changes, and more in real time—the exact information Dr. Davidson and her colleagues painstakingly took down with pen and paper. By matching this phenotypic data to genomic and other types of molecular data, researchers think they can begin to figure out the underlying cause for each individual's illness—be it acute, like a fast-moving cancer, or chronic, like high blood pressure.

"We can now measure so much in exquisite detail," says Muredach P. Reilly, MBBCh, who this year moved from the University of Pennsylvania to be director-designate of the Irving Institute for Clinical and Translational Research. "We have these incredible tools that allow us to look at each individual's genomic and physiological profile and to map them in health and disease."

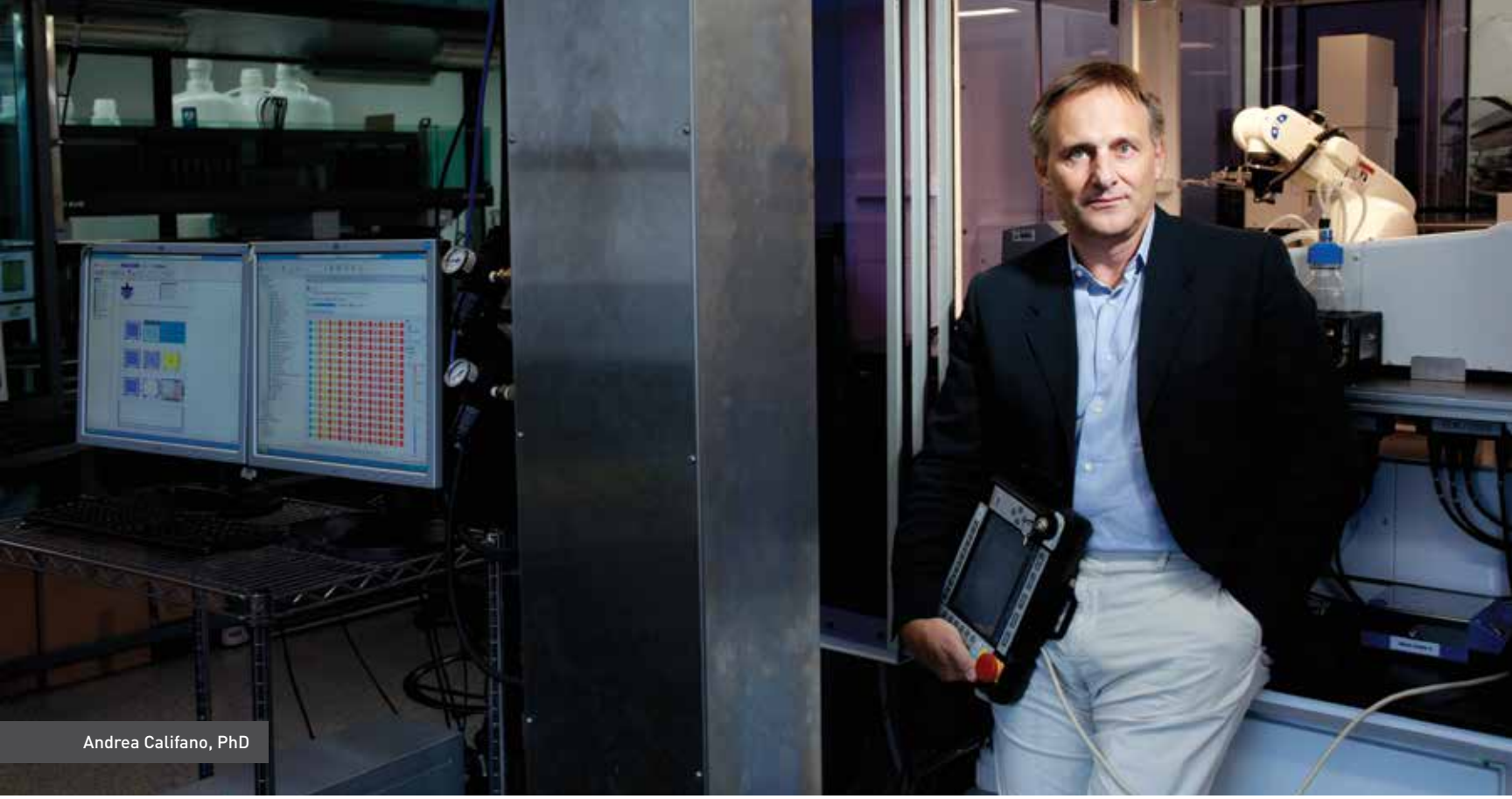
"For most chronic health problems we don't have a single treatment that works for everybody."

The ability to capture this vast quantity of data is transforming both research and medical care—and Columbia researchers and clinicians are at the forefront of that transformation. "The institutional vision for precision medicine at Columbia is very strong," says Dr. Reilly, who will succeed Henry N. Ginsberg, MD, as director of the Irving Institute for Clinical and Translational Research next year. "There's huge, top-down support here for developing precision medicine tools and big-data know-how for personalized care. And there's also a vibrant feeling on the ground of a commitment to collaboration among different fields of medicine—from bioinformatics, to genomics, to cancer, to behavioral health—to make ours the leading precision medicine program in the U.S."

The type of approach that Dr. Davidson takes is called an N of 1 study—a highly systematic and statistically rigorous examination of a single subject. Until recently, single-patient studies such as the one she ran for her young Canadian patient have been considered interesting and informative, perhaps, but anecdotal and ungeneralizable. Instead, for the past half-century, randomized controlled trials with a large N—the letter that symbolizes the number of patients—have been the gold standard in determining a treatment's efficacy. What these studies identify is essentially the treatment that works best for the average patient.

But as researchers increasingly come to understand the extent of inter-individual variability among patients, they are realizing that what helps the average patient—whoever he or she may be—often does little for their specific patient. "For most chronic health problems—pain, obesity, sleep issues, asthma, the list goes on and on—we don't have a single treatment that works for everybody," Dr. Davidson explains. Because first approximations are extremely useful, and work well for many people, randomized controlled trials will probably remain the bread and butter of clinical research for the indefinite future, she says. But alongside them, more personalized approaches are sorely needed.

Take blood pressure pills. There are many kinds, and overall they have similar efficacy. But for any one person some will work while others will not. A doctor might prescribe one and then, if the desired effect isn't achieved, add a second and even a third, never figuring out which is actually doing the job. "But if we standardly test those medications at the lowest dose in a randomized way, we might control people's symptoms



Andrea Califano, PhD

with the fewest drugs," she says. A clinician could identify the unique protocol that works best for one specific patient, marrying the promise of evidence-based approaches with a truly personalized touch.

Working with internist Ian Kronish, MD, Dr. Davidson is using N of 1 studies to help individuals improve their cardiovascular health. Recently she and her colleagues completed a study funded by the National Heart, Lung, and Blood Institute that explored the relationship between exercise and stress. The link, they found, was highly specific for every individual: For some, exercise seemed to drive down stress levels; for others, stress drove down exercise levels; and for others, still, there was no correlation at all. "All of those scenarios make sense, don't they?" Dr. Davidson says. "It's just that you have to find what's true for *you*—not for the average person." Unfortunately, that kind of individual variation washes out in large trials, which tend to find no association at all between stress and exercise.

Now Dr. Davidson's team is surveying physicians, nurses, and patients nationwide to identify conditions for which N of 1 trials would most benefit patients. Hypertension is a strong candidate because it is a costly disease for which patients often take multiple drugs for life. Pain is another possibility; being able to systematically test the efficacy of multiple interventions in one person could provide life-altering relief. The team is also surveying cancer survivors about their interest in using the approach to manage their depressive and fatigue symptoms. Whichever they tackle will be a vital step toward realizing Dr. Davidson's vision of 25 years. "It's an exciting time," she says, "because we could be the first in the country to offer such a service."

N of 1 studies come in an array of permutations. Andrea Califano, PhD, uses the approach to find better ways to treat rare and otherwise incurable cancers. The FDA has now approved a few dozen drugs that target particular genetic mutations in tumors—but it is unclear why many patients fail to respond and, even in the best cases, initial response is followed invariably by relapse. And although genome sequencing is becoming increasingly routine in cancer care, less than 11 percent of patients who are treated with a drug based on their genomic data experience a long-term benefit.

That is because mutated genes, pinpointed through sequencing, are not always the best therapeutic targets, says Dr. Califano. In some patients, genes associated with proteins that are hyperactive in the cancer cell do not carry mutations and vice versa—some mutations do not cause proteins to misbehave because other genes can compensate for the bad code. Instead, the clinically relevant players are "master regulators," effectively the puppeteers of the cell transcriptional state. Dr. Califano and his colleagues have used massive supercomputers to reconstruct the regulatory networks of more than 30 different cancer types. Using the RNA fingerprint of a patient's tumor—its gene expression profile—they can use these networks to predict both the master regulator proteins and the drugs that are likely to be most effective for that person's cancer. These drugs are then taken for a test run in mouse "avatars"—models created with the patient's own tumor tissue. In November, Dr. Califano was named a National Cancer Institute outstanding investigator and received a seven-year, $6.7 million grant to pursue the work. He has already piloted the approach in more than 40 patients and is testing it more widely in

PATIENT-ORIENTED RESEARCH

It can take decades for knowledge to go from bench to bedside. To spur the momentum of early-career investigators who have promising research programs, the Irving Scholars program annually selects a cohort of P&S assistant professors to receive stipends of $60,000 annually over three years along with a named professorship. This year's recipients:

- **Ali Jabbari, MD, PhD, Dermatology:** Patients suffer devastating psychosocial consequences of the hair loss triggered by alopecia areata, an autoimmune disease for which the FDA has yet to approve a treatment protocol. Clinical research by Dr. Jabbari and others suggests that a class of compounds known as JAK inhibitors has the potential to reverse alopecia areata symptoms. By integrating patient data with insights from mouse models of the disease, Dr. Jabbari seeks to identify the mechanisms by which JAK inhibitors spur hair regrowth.

- **Fay Kastrinos, MD, Medicine:** As a gastroenterologist with Columbia's Pancreas Center, Dr. Kastrinos oversees the clinical practice for the Muzzi Mirza Pancreatic Cancer Prevention and Genetics Program. By integrating data from the center's pancreatic cancer family registry with patient imaging and other clinical information, Dr. Kastrinos and her collaborators are developing a clinical prediction model to identify people at greatest risk for hereditary susceptibility to pancreatic cancer, the fourth leading cause of death for American men and women.

- **Krzysztof Kiryluk, MD, Medicine:** Dr. Kiryluk investigates IgA nephropathy, the most common form of primary glomerulonephritis worldwide. While up to 40 percent of patients with IgA nephropathy develop end-stage kidney failure within two decades of the diagnosis, the variable course of the disease has prevented doctors from offering patients a clear prognosis. Dr. Kiryluk and his team seek to home in on the genetic mechanisms of IgA nephropathy, enable noninvasive diagnosis, improve personalized prognostication, predict relapse and recurrence, and, ultimately, develop novel targeted therapeutics.

- **Joanna E. Steinglass, MD, Psychiatry:** Now an associate professor, Dr. Steinglass investigates the cognitive neuroscience of anorexia nervosa, a debilitating illness for which clinical interventions are only modestly effective. Her analyses combine clinical data with brain imaging to reveal the link between neural mechanisms that drive food choice and the behavioral disturbances manifested by people with the illness. She aims to leverage those insights to develop treatments to prevent relapse.

- **Nicholas Tatonetti, PhD, Biomedical Informatics:** Every day, electronic health records capture billions of clinical data points around the world. Dr. Tatonetti and his team use rigorous computational and mathematical methods to advance data science in the realm of "systems pharmacology." By integrating medical observations with systems and chemical biology models, they intend to pursue understanding of basic biology and human disease, explain drug effects, and predict adverse drug reactions.

several follow-up clinical trials. "The fact that you can computationally predict the most effective drugs and test them in a mouse model of the patient's tumor to find the ones that effectively kill it," he says, "is unprecedented."

Master regulator networks play a role, too, in alopecia areata, a genetically inherited condition in which deranged immune cells attack hair follicles. Associated with severe hair loss and currently very difficult to treat, the disease has been the primary topic of study by genetics and dermatology researcher Angela M. Christiano, PhD, since she was diagnosed with the condition more than 15 years ago.

In November 2015, Cell Systems published work by postdoc James C. Chen, PhD, with Dr. Christiano and her team to untangle how master regulator networks influence the pathology of alopecia areata. "We wanted to track the process of immune cells infiltrating an organ," says Dr. Christiano, who notes that unlike cancer, which typically involves a single cell type, a technical challenge associated with autoimmune disease is the interaction among tissue types—in this case, hair follicles and T cells.

Beyond insights into the mechanisms of pathology in autoimmunity, Dr. Christiano's work offers a through-the-looking-glass perspective on the quest for better ways to fight cancer. In alopecia areata, overzealous T cells destroy otherwise healthy hair follicles. In their quest for immunotherapy treatments, oncologists seek a tactic to spur T-cell function to boost patients' capacity to blast their own tumors. "Our hypothesis was that if we could do it in autoimmunity, we could find the drivers and extend it back to cancer," says Dr. Christiano. "For autoimmunity, we'd like to down-regulate or dampen the process of T cells attacking an organ, but for cancer you would try to find ways to enhance it."

Dr. Christiano credits Dr. Chen—a recipient of one of four precision medicine fellowships announced by the Irving Institute for Clinical and Translational Research in January 2016—for spurring the new line of inquiry for her group, which has used genome-wide association studies to identify pathways that could be targeted by a class of compounds known as JAK inhibitors to spur complete hair regrowth for 75 percent of clinical trial participants with alopecia areata. As a doctoral student with Dr. Califano, Dr. Chen investigated the role of master regulator networks in a particularly vicious form of brain tumor known as glioblastoma. Now he works with Dr. Christiano to extend those techniques to autoimmunity and understand what is going on at a molecular level for patients who do or do not respond to treatment with JAK inhibitors. "When a patient comes in with a new spot of hair loss from alopecia areata, it's not clear just by looking at them whether they'll get better, get worse, or stay the same," says Dr. Christiano. "For us, precision medicine means better diagnosis and being able to direct them toward or away from a JAK inhibitor or other new therapies on the horizon."

Data scientist Raul Rabadan, PhD, takes an *in silico* approach to patient-specific data. A theoretical physicist by training who turned his attention to biology about a decade ago, Dr. Rabadan

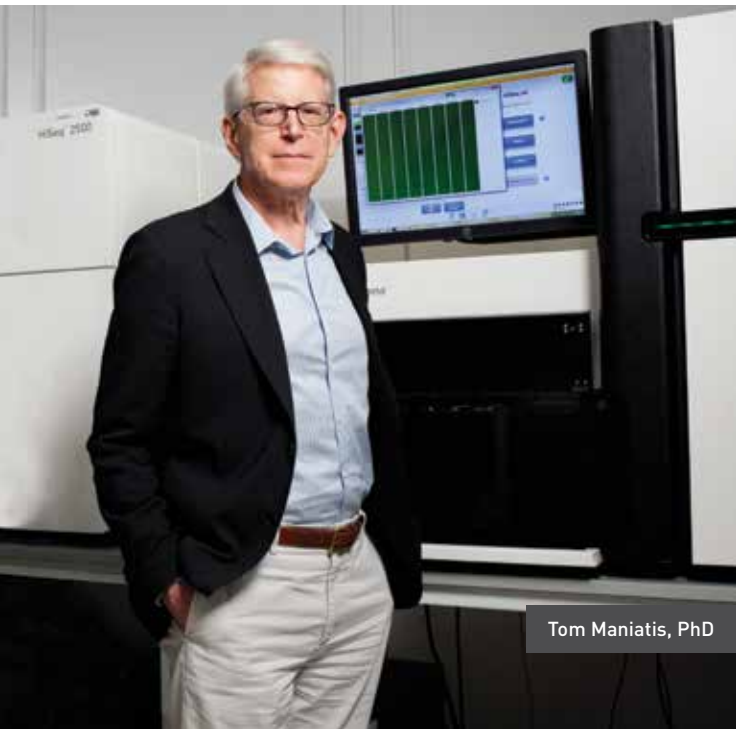
Angela M. Christiano, PhD

Raul Rabadan, PhD

uses genomics and a branch of mathematics called topology to map large genomic datasets. He is the director of an NCI center, the Center for Topology of Cancer Evolution and Heterogeneity, that connects mathematicians, physicists, biologists, and clinicians to study cancer using large-scale genomic data.

Dr. Rabadan is leveraging genomic and mathematical approaches to gain greater insight into the mechanisms by which tumor cells evolve in response to cancer-killing drugs. Mutations accrue over time, as well as in response to the therapies that a patient receives. No two patients will develop the same set of mutations, but a tumor's evolution may well affect the cancer's response to subsequent therapies. Some mutations will affect the tumor's growth, or its resistance to a particular drug protocol, while some will not have any effect at all. Last year, by using a mathematical approach to probe the genomes of patients with a particularly aggressive form of lymphoma, Dr. Rabadan's team was able to identify a novel pathway that drives the cancer in many cases. Because drugs targeting this pathway have already been approved by the FDA to treat other conditions, the study identified a new tactic available to oncologists treating patients in whom that target is mutated. Now Dr. Rabadan has focused his efforts on understanding how brain tumors evolve under therapy. A recent analysis reported in Nature Genetics in June was the largest ever study of brain tumors, conducted through an international collaboration led by his group. The group identified several genetic mechanisms of resistance that drive evolution of these tumors under standard therapy, providing important clues for novel treatments and showing how genomics can be used for precision medicine approaches to cancer.

> By following the trajectories of single cells, researchers seek to capture the biological phenomena that would otherwise be lost in the shuffle of complexity.



Tom Maniatis, PhD

Dr. Rabadan is also working with the group led by Tom Maniatis, PhD, director of Columbia's Precision Medicine Initiative, to refine and apply these computational tools to the study of stem cell differentiation at the single cell level. By following the trajectories of single cells—and the 20,000 or so genes they express—through space and time, his group seeks to capture the biological phenom-

ena that would otherwise be lost in the shuffle of complexity. This collaborative study has identified new genes involved in the differentiation of embryonic stem cells to produce motor neurons and made it possible to identify all of the genes expressed in individual cells at multiple stages of differentiation, a "systems" approach to understand complex cellular differentiation. In addition, this method should make it possible to identify neurodegenerative disease mechanisms using cells obtained from neurodegenerative disease patients, such as those with amyotrophic lateral sclerosis, known as Lou Gehrig's disease. In this case, patient fibroblasts are converted to induced pluripotent stem cells—iPS cells—which in turn are differentiated to produce motor neurons. Comparison with motor neurons produced from control cells may provide important insight into ALS disease mechanisms.

Research and interventions focused on individual patients have an unprecedented ability to pinpoint biological disease mechanisms and identify therapies that traditional clinical trials will miss. But scientists do not have to work at the individual level to contribute to the endeavor of precision medicine, says George Hripcsak, MD, a clinical informatics pioneer. His research gleans information about widely used treatments by examining select groups of patients with similar characteristics—an approach known as "stratified medicine."

Dr. Hripcsak is the co-PI of an international program called Observational Health Data Sciences and Informatics, or OHDSI, a voluntary network of 60 patient databases in 14 countries that so far total about 600 million patient records. Network investigators aim to enroll 1 billion patients by 2023. Using those records, any network participant can set up quick observational studies that analyze the patient data to address an important question about health care. "Observational studies don't prove causality, but randomized clinical trials often can't be generalized to *my* patient," says Dr. Hripcsak. By analyzing subsets of patients matched for shared features—response to a particular sequence of drugs, for example—clinicians may be able to tailor treatment more narrowly within large populations, such as people with diabetes, hypertension, heart disease, and other chronic conditions.

In the network's first major paper, published in June in the Proceedings of the National Academy of Sciences USA, Dr. Hripcsak and his colleagues provided proof of concept of that strategy with an analysis of the treatments received by patients with diabetes, hypertension, and depression around the world. The team set narrow parameters for which patients to include, whittling down the study sample to fewer than 1.8 million records. Then they queried the database for the sequence of treatments these select patients had received. Generally, patients with diabetes received the same first treatment around the world, but the researchers found huge variability for the other conditions.

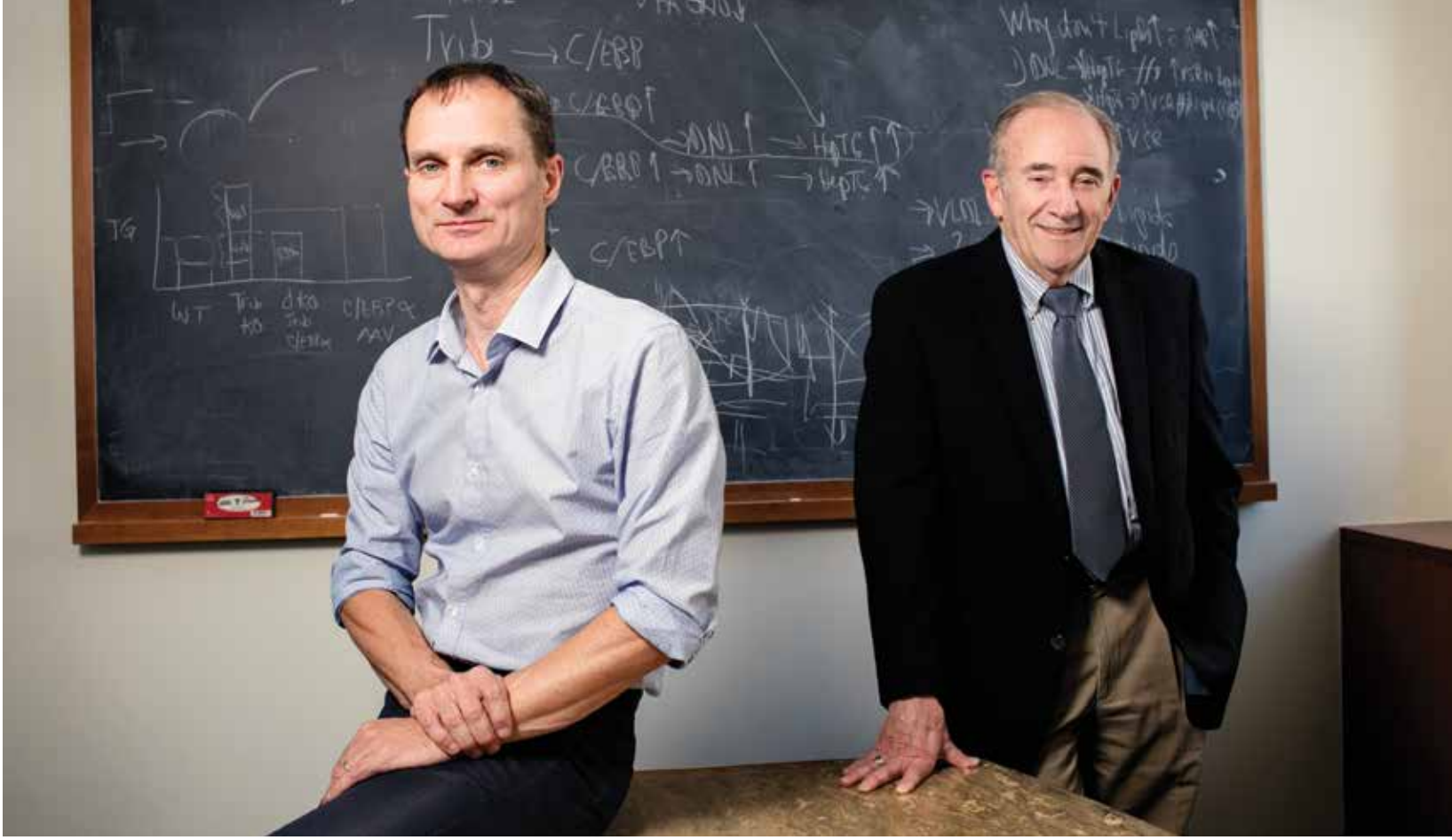
Knowing which therapies are actually used is crucial in conducting randomized trials for new therapies, says Dr. Hripcsak, and for integrating evidence-based approaches gleaned from randomized trials with a more personal approach. "It's hard to learn what's good treatment, or compare a prospective therapy to some standard care, when clinicians worldwide are all doing something different. If it turns out nobody uses the recommended therapy, then you end up doing very expensive randomized trials to prove that one unused drug should be replaced with another unused drug."

That study was something of a trial run for the network, but Dr. Hripcsak looks forward to addressing other questions using the network. His team's next study will examine side effects for all marketed medications. Of course, the massive analysis will not show whether any particular drugs cause the side effects, but the analysis will be able to flag the high and low fliers, he says. Meanwhile, independent of the OHDSI project, Dr. Hripcsak and colleagues are beginning to design patient-specific computational models for people with type 2 diabetes that predict in real time how their bodies will respond to certain food intake or activity levels. When programmed into a wearable device, these programs, tuned to the specifics of the wearers' bodies, will help diabetics know when and what to eat and when and how much to exercise to achieve their goals for disease management. "The availability of all these data—and also new algorithms and computing techniques—allows us to calculate things that even a few years ago just weren't feasible," he says.

Dr. Reilly, too, studies subsets of human subjects, but his goal is to hand-select subjects with specific genetic mutations in order to conduct deep-dive research into how genes function in health and disease. Genomic analysis has linked a multitude of genes to specific conditions, but of some 18,000 protein-coding genes in

Who's Who

- **Andrea Califano, PhD,** the Clyde'56 and Helen Wu Professor of Chemical Biology (in Biomedical Informatics and the Institute for Cancer Genetics), professor of biochemistry & molecular biophysics, and chair of the Department of Systems Biology
- **Angela M. Christiano, PhD,** the Richard and Mildred Rhodebeck Professor of Dermatology and professor of genetics & development
- **Karina Davidson, PhD,** professor of behavioral medicine in medicine and psychiatry, director of Columbia's Center for Behavioral Cardiovascular Health, and vice dean for organizational effectiveness
- **Henry Ginsberg, MD,** the Herbert and Florence Irving Professor of Medicine and director of the Irving Institute for Clinical and Translational Research
- **George Hripcsak, MD,** the Vivian Beaumont Allen Professor of Biomedical Informatics and chair of the Department of Biomedical Informatics
- **Ian Kronish, MD,** Florence Irving Assistant Professor of Medicine
- **Tom Maniatis, PhD,** the Isidore S. Edelman Professor of Biochemistry & Molecular Biophysics, chair of the Department of Biochemistry & Molecular Biophysics, and director of the Columbia University Precision Medicine Initiative
- **Raul Rabadan, PhD,** associate professor of systems biology and of biomedical informatics
- **Muredach P. Reilly, MBBCh,** director-designate of the Irving Institute for Clinical and Translational Research and the Florence and Herbert Irving Professor of Medicine



the human genome, the effect of about 15,000 is unknown. Dr. Reilly studies the genomics of cardiovascular disease, and you might call one focus of his studies "natural mutants," people who have been found, through gene sequencing, to have a mutation in a gene that has been implicated but not proved in a condition.

Dr. Reilly riffs that he works not on N of 1 studies, but on N of 10, or 20, or 30. "These are really intense studies of very selected people," he says, "to understand what a particular gene does." In many cases these people appear healthy and "normal," he explains, but subtle changes in factors like response to stressors or immune or metabolic changes can be uncovered by probing widely into their physiology. "If we think a gene alters the course of heart disease, it may affect multiple things, like exercise capacity, metabolic rates, or energy consumption." Maybe it affects a person's heart rate and blood pressure during exercise, but not at rest, for example. "We need to study the human very carefully to gain these insights into personal health and the role of specific genes in the human setting."

In genetic studies of people with coronary artery disease, for example, Dr. Reilly's group explored one gene, ADAMTS7, which codes for an enzyme that modulates matrix proteins in the blood vessel walls. The group is now looking for loss-of-function mutations of this gene in humans to understand specifically how that gene function contributes to cardiovascular effects. Following the trail of genes by characterizing them from the molecular to functional level can create possibilities for therapeutic targets, he says.

Another focus of Dr. Reilly's inquiry is immune response to stress, which can change in response to high-fat food and exercise and which in turn can alter the risk for other deleterious conditions, such as diabetes, stroke, and cancer. He and his colleagues set up a paradigm to inject tiny levels of inflammatory toxin into 300 healthy people, then intensely studied their subjects' responses, following their innate immune responses as well as gene transcription changes in multiple cell types like blood and adipose tissue. That work, published in March, led them to discover a new part of the genome that regulates whether a person develops a high or a low fever in response to an immune system stressor. "The area of the genome associated with the fever was not associated at all with a person's resting temperature," says Dr. Reilly. That finding points to a separate genetic dial controlling how a person regulates his or her response to infection and trauma. Understanding that is important because both high- and low-temperature response to infection and trauma predicts death.

Ultimately, the next frontier in medicine will involve analyzing multiple layers of data on sick and healthy people, to tie molecular differences within and among people to differences in physiology, behavior, and environment, says Dr. Reilly. "We spend a lot of time knocking out genes in mice, zebrafish, or *Drosophila* to create models of disease and those are incredibly important for understanding molecular, biochemical, and physiological mechanisms," he says. "But when it comes to understanding molecular mechanisms of the genome of a human, the human is the best model." ❖

Muredach P. Reilly, MBBCh, left, joined P&S as director-designate of the Irving Institute for Clinical and Translational Research and will become director next year when he succeeds Henry N. Ginsberg, MD, who has been director since 1995.

"When it comes to understanding molecular mechanisms of the genome of a human, the human is the best model."