our genes are vital determinants of your health, but they almost never act alone. Variants or mutations of specific genes might increase the risk of developing Alzheimer’s disease, say, or heart disease, but they interact with myriad variables: other seemingly unrelated genes, clinical events such as infections, and environmental and lifestyle dimensions, including stress, diet, and sleep.

The complexity of teasing out the interplay among all these factors is mind-boggling to understand and describe. In 2016, the National Institutes of Health launched an initiative known as All of Us to start untangling the cause and effect. Even in this era of big data, it is an ambitious undertaking: Track multiple contributors to health in 1 million Americans from all walks of life over the course of a decade, then crunch the numbers to pinpoint subtle interactions among them. “The results of this initiative will reveal more about our individual differences at all levels and will create the opportunity for new and more effective treatments,” says Tom Maniatis, PhD, director of Columbia’s Precision Medicine Initiative, a collaboration of all of Columbia University and NewYork-Presbyterian Hospital.

Last July, Columbia, in partnership with Weill Cornell, NewYork-Presbyterian, and Harlem Hospital Center, received a major grant to enroll 150,000 people of the 1 million planned nation-

Columbia Helps Lead Ambitious New Precision Medicine Project

By Alla Katsnelson

Illustrations by Joey Guidone
The grant—for $4 million initially but expected to total $46.5 million over five years—is significant for the University on many levels, says David Goldstein, PhD, director of Columbia’s Institute for Genomic Medicine and principal investigator for the grant. “Just having a leading role in the country’s signature precision medicine effort is momentous. By participating in this national initiative we can help optimize the way the underlying science of precision medicine develops.”

The project also will have benefits for patients at Columbia and in New York City in general, he says. “It is completely consistent with the institute’s and Columbia’s vision for bringing genomics as a starting point for precision medicine to the entirety of our patient population—it’s as simple as that.”

Boiled down to its essence, Columbia’s mandate from All of Us boasts a deceptively elegant simplicity: Enroll the requisite number of subjects from the New York institutions’ ethnically and economically diverse patient population, obtain their consent and basic information along with some biological samples, and send everything to the program’s national biobank at the Mayo Clinic in Roch-
ester, Minn. Scientists can then endlessly probe the resulting repository. The top-line idea, says Dr. Goldstein, is to combine these data with genomic sequencing to develop a method for state-of-the-art interpretation of everything each individual’s genome can reveal. “We don’t actually know how to do that interpretation right now,” he says. “I’m really excited about taking that on as a challenge as we define Columbia’s role in this project.”

Investigators also need to create a means for clinicians to provide results to participants, says Wendy Chung, MD, PhD, the Kennedy Family Professor of Pediatrics (in Medicine), who directs the University’s clinical genetics program and co-directs the molecular genetics diagnostics lab. “This is our first foray into something on this scale, a big deal for Columbia because we’ve never done anything on this scale,” she says. “We need to help every patient who walks through our doors understand that he or she has the opportunity to be a part of the cures of tomorrow.”

Ultimately, the idea is not just for clinicians and researchers to make discoveries and publish papers, but also to engage study participants with the research and help them learn about their health in actionable ways. For example, says Dr. Chung, a password-protected dashboard might allow All of Us participants to log in to review their basic health measures, like blood pressure or glucose levels, and assess other aspects of their health gleaned from their genomic and environmental and lifestyle data. “We have to realize that most of health is not happening at the doctor’s office. We want people to make daily decisions that are going to sustain their health.”

The NIH is still working out how data will be shared with participants, but Dr. Chung says communication must be done in a way that will benefit participants regardless of their economic or educational status. “Some of our underserved communities in New York don’t have
DiGeorge syndrome may be one of the most common genomic disorders you have never heard of. Affecting as many as one in 2,000 babies, the syndrome was first described nearly 60 years ago and spans a cluster of symptoms that can include heart defects, speech delay, intellectual disability, immunodeficiency, and low serum calcium. The assortment of seemingly unrelated health issues that children face can be dizzying for families. About a third of all patients also have congenital renal system malformations that can result in kidney failure.

Over the past six decades, researchers have traced DiGeorge syndrome to deletions within a segment of chromosome 22 that carries some 40 genes and have tied certain hallmarks of the disease, such as cardiac malformations, to specific genes within that stretch of DNA. But the deranged biochemical codes that yield other DiGeorge symptoms have eluded investigators.

In January, Simone Sanna-Cherchi, MD, the Paul Marks Scholar Assistant Professor of Medicine, and his colleagues identified the glitch behind DiGeorge-associated kidney malformations by conducting genomic analyses of 2,666 children born with congenital kidney and urinary tract abnormalities. The study, co-authored by Ali G. Gharavi, MD, the Jay Meltzer, MD, Professor of Nephrology and Hypertension (in Medicine), chief of nephrology, and a member of Columbia’s Institute for Genomic Medicine, led the team to a nine-gene region within the DiGeorge stretch on chromosome 22. To home in on the offending glitch, they tested all nine of the genes in both zebrafish and mice. Their results, published in the New England Journal of Medicine, revealed that errors in the gene CRKL are the main driver of kidney disease among children previously diagnosed with DiGeorge syndrome. Indeed, a full 1 percent of all kidney malformations—including those not previously associated with DiGeorge syndrome—could be traced back to CRKL.

Being able to trace kidney malfunction back to that particular stretch of chromosome 22 is life-changing, says Dr. Sanna-Cherchi, especially for children not previously diagnosed with DiGeorge syndrome, which is also associated with symptoms that emerge later in life, including schizophrenia and Parkinson’s disease. “So you can see how this knowledge would help patients understand their future diagnostic and disease management options.”

Dr. Sanna-Cherchi’s study, which mined a global database of pediatric kidney patients, is just one example of the crucial role such repositories are likely to play as the field of precision medicine develops. Investigators at Columbia expect such work will get a major boost as the NIH’s signature precision medicine study, All of Us, goes online. The massive database will integrate genomic, environmental, and other individual factors to allow researchers to connect the dots among variations that protect our health and those that contribute to illness.

For more than a decade, Columbia investigators have been proving the worth of smaller-scale and more defined databases like the one that led to Dr. Sanna-Cherchi’s DiGeorge discovery. Wendy Chung, MD, PhD, the Kennedy Family Professor of Pediatrics (in Medicine), has been a leader in SPARK, a project that aims to build a basic knowledge base about autism by collecting genomic, clinical, and environmental data on 50,000 individuals with the condition. Tom Maniatis, PhD, the Isidore S. Edelman Professor and Chair of Biochemistry & Molecular Biophysics at P&S and co-founder of the New York Genome Center, has worked extensively with the center’s ALS Consortium, which brings together multiple academic medical centers, clinicians, scientists, and industry partners to combine clinical and functional genomics with bioinformatics in the study of the mechanisms behind Lou Gehrig’s disease. Consortium members have pledged to complete whole genome sequencing and analysis of 3,200 clinical samples in a quest to enhance early diagnosis and effective drug discovery. “Columbia has a leading role in establishing this framework that will apply state-of-the-art clinical and functional genomics together with bioinformatics to study ALS disease mechanisms. This is our best hope for answering questions about this devastating illness,” says Dr. Maniatis.
access to the technology of Silicon Valley, and they don’t think about things the same way,” she stresses. “We have to make sure we are getting everyone to the table.”

Undoubtedly, the success of All of Us hinges on recruiting a diverse group of participants that represents the genetic and environmental experiences of the country. But it also relies on creation of a data infrastructure that integrates all participating institutions across the country. Developing that infrastructure is a project in its own right, and last year Columbia received a grant of $13.7 million to play a central role in launching the NIH’s Data and Research Support Center, a collaboration being led by Vanderbilt University.

So far, clinical data for All of Us consist of a collection of electronic health records. Eventually, data will include extensive surveys filled out by participants and even data from wearable technologies that track information such as sleep and exercise in real time, says George Hripcsak, MD, the Data and Research Support Center grant’s principal investigator at Columbia. Dr. Hripcsak has a lot of practice with enormous databases of health information: He is the principal co-investigator of an international program called OHDSI (Observational Health Data Sciences and Informatics), a network of 60 patient databases worldwide that contain hundreds of millions of patient records. His team developed a platform called OMOP (Observational Medical Outcomes Partnership) to standardize enormous reams of clinical data within OHDSI, and the NIH initiative will run on OMOP as well.

At first, says Dr. Hripcsak, the clinical and sociomedical data that OMOP stores will be used to probe the demographics of All of Us participants to confirm that they are statistically representative of the nation’s diversity. But like Dr. Goldstein, Dr. Hripcsak stresses that the magic will really kick in when those data can be combined with participants’ genome sequences. Imagine that 100 people share the same genomic quirk, but only a small fraction of them get the disease associated with it. “Then we can ask what factors seem to influence whether people who have that genomic characteristic get the disease or not,” Dr. Hripcsak says. “Is it your food, your environment, other diseases you have, or drugs you are taking?”

Researchers acknowledge that learning more about “All of Us” may result in more questions than answers at first, but they are convinced that only by looking at the bigger picture will the precision needed for individual health come into focus.

“We need to help every patient who walks through our doors understand that he or she has the opportunity to be a part of the cures of tomorrow.”