

BIG DATA

## All of Us

### DNA-based medicine needs more diversity to avoid harmful bias. One big research project is fixing that

By *Stephanie Devaney*

**WHEN THE RACE TO** sequence the first human genome was rushing toward the finish line about 20 years ago, I remember feeling mesmerized by what was about to happen. It was the dawn of a new century, and it seemed we were on the cusp of unlocking the meaning behind the blueprint of life, DNA. Once we could line up all 3.1 billion base pairs of the molecule in our genome, I thought—I was an undergraduate student at the time, dazzled by science—we would understand everything there is to know about human health and disease.

What I didn't know was that those first decades of genetic medicine would leave a lot of people behind. So I was taken aback several years later, in 2009, just after I got my doctorate in molecular genetics, when researchers at Duke University

reported that 96 percent of the genomic data we had gathered came from people of European ancestry. This was not the result of small numbers: they calculated the percentage using the more than 1.7 million individual genome samples analyzed at the time, but the samples were lacking diversity. Over the next few years things did not get much better, and as recently as four years ago genomic databases were still way out of balance, with more representation of Europeans and less of everyone else.

This inequity, if it is not fixed, will turn into tremendous health inequality. Today more and more people are getting answers about the underlying causes of their diseases because of medicine's ability to mine their genomes. There are hundreds of drugs that contain genetic information in their labeling because gene variants affect how bodies process these drugs, and knowing the variants that patients have helps doctors set the most beneficial dose for their patients. Moreover, today improved knowledge about the genomic drivers of different cancers has paid dividends in how physicians diagnose and treat many tumors. Yet people who are not white and not male have different sets of genes that do not always fit into these treatment regimens.

For example, African-Americans and Latinos have the highest rate of asthma in the U.S., but studies show that common drugs used in inhalers do not help them as well as they help whites. Asians who take the antiseizure drug carbamazepine have a higher risk of a severe, sometimes fatal, reaction. Nobody developing these drugs, or prescribing them when they first came into use, anticipated these problems. If DNA is one important factor in our quest for more effective medical treatment, we need to address the lack of diversity in genetic data.

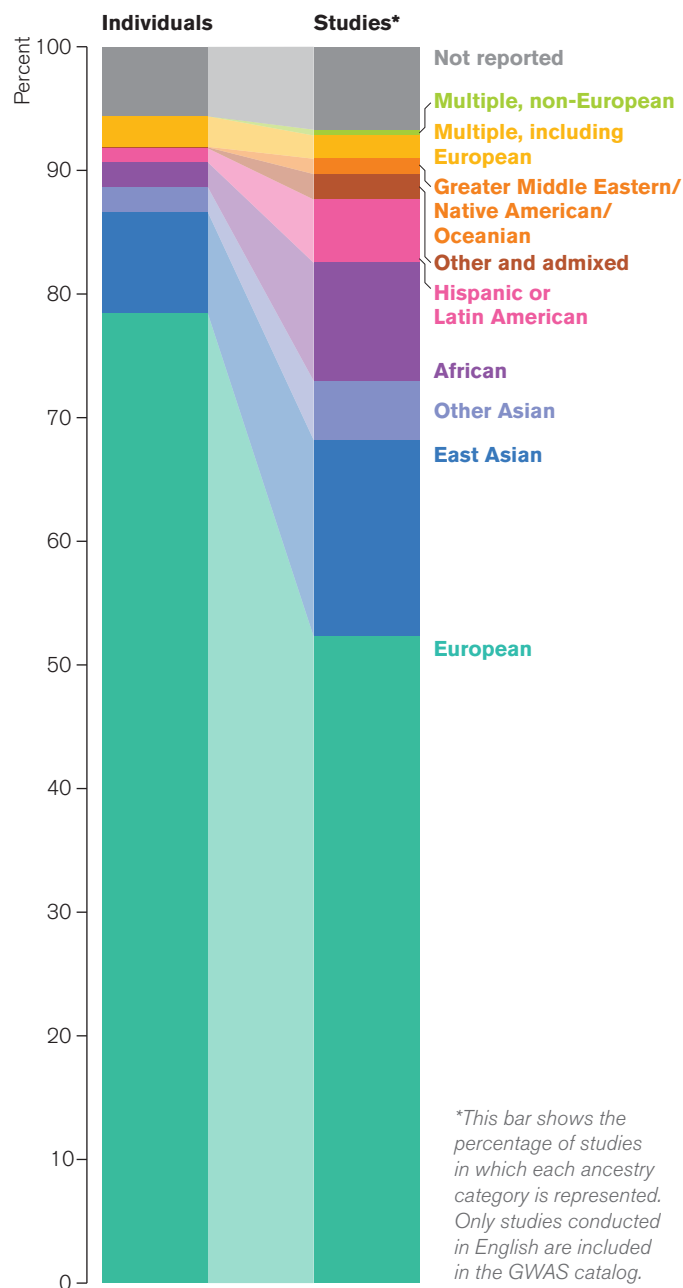
That is where the *All of Us* Research Program, where I work, hopes to help. Set up by the National Institutes of Health and launched in 2018, we are asking a million or more people from all backgrounds to join us as partners in research, not as human subjects, and share all kinds of health information over the course of their lives. Already we have more than 250,000 participants. More than 51 percent belong to racial and ethnic minorities, more than 10 percent are sexual and gender minorities, and overall more than 80 percent represent a group that has been historically underrepresented in research data sets.

People can join All of Us by going to our program Web site ([www.joinallofus.org](http://www.joinallofus.org)) and clicking "Join Now." After agreeing to participate, respondents can offer us their medical records, answer a variety of surveys about their health

## Biased Gene Studies

To link genes to disease risk and other traits, hundreds of genome-wide association studies (GWAS) have looked at the DNA of thousands of different people as of 2018. But in terms of racial background, these people are not so different. Taking all the projects together, 78 percent of the people in them are white Europeans, whereas just 2 percent are African and 1 percent are Hispanic or Latin American. The studies themselves also predominantly focused on Europeans and rarely on other populations. So gene variants that appear in non-European people and may be linked to illness rarely show up in this research. The scarcity makes it hard to analyze and understand the significance of the variants.

### Racial Backgrounds in Published Gene-Association Studies



and lifestyle, and participate in other activities such as syncing their fitness tracker data to our program. We also have hundreds of enrollment sites at local hospitals and health centers across the country where participants can provide samples of blood and urine to help researchers study their DNA. Our hope is for people to stick with us for 10 years or more because, as the program grows, we will regularly add new ways for them to learn about themselves and contribute to research.

### THE MOMENT IS RIGHT

**A LOT OF THIS PARTICIPANT-RESEARCHER** collaboration is linked to advances in technology. Sequencing that first human genome had a \$1-billion price tag. Today such a sequence costs less than \$1,000 and can take less than 24 hours to complete. It is also easier to integrate this information with other crucial medical data. Health care organizations have been turning their patients' paper-based medical records into electronic versions. As of 2017, 96 percent of all U.S. hospitals and 80 percent of all office-based doctors are using a certified electronic health record system. New apps on smartphones and other digital health technologies such as smart watches collect data from nearly anywhere and directly from a person. These trends all make it easier to store, share and mine large data sets for answers to questions about disease causes and effects. Such trends also raise big and disturbing issues about privacy, making it important for projects such as ours to have both strong security and full transparency to all our participants.

And it is crucial to treat these people as partners. The actions of past medical researchers have earned much distrust in minority communities, after causing harm in the Tuskegee Syphilis Study, where researchers misled African-American men with syphilis and never gave them adequate treatment, and with the widespread use of HeLa cells, which were taken from a patient named Henrietta Lacks without her knowledge or permission. People wanted to see research go forward but *with* them rather than about them. To overcome this kind of distrust, All of Us is using a new model for research, one that invites input from participants as well as researchers with science degrees. Participants serve on the program's advisory and governing bodies, working groups, and task forces. We have also partnered with local health care organizations, hospitals, and community groups to advise us and help find people to participate. Community engagement is not familiar ground for large medical research projects, and we are still learning the best ways to do it.

Some studies have provided us with blueprints for developing long-term relationships like the ones we hope to have, studies that have changed medicine for the better. The Framingham Heart Study, for example, started in 1948 with 5,209 men and women, largely white, from one town in Massachusetts. With a 99 percent retention rate, the study continues to this day. As participants share data year after

## A Better Balance

year, researchers can see how their heart health changes over time. The risk factors for heart disease identified by the Framingham study—such as high blood pressure, high cholesterol, smoking and obesity—are so ingrained in our collective consciousness and our approach to health care that they feel like common sense.

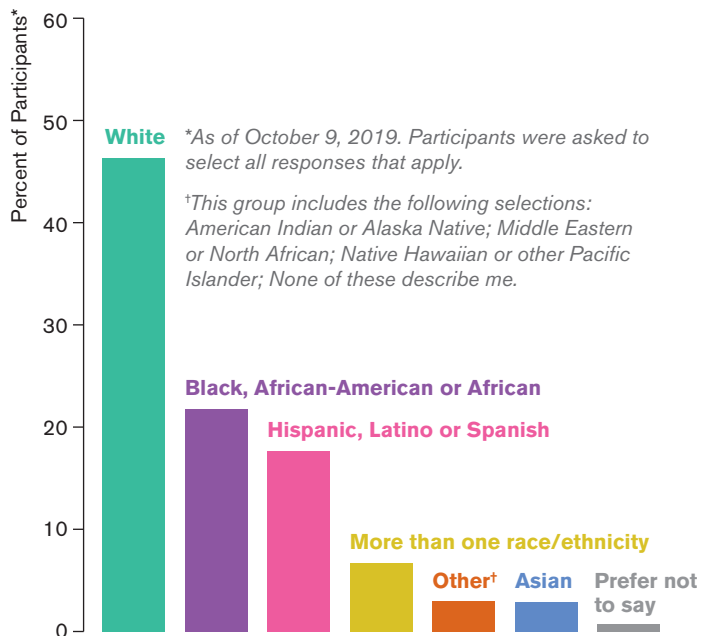
### GOING FURTHER

**THIS KIND OF MEDICAL DISCOVERY** is what we envision for All of Us, but we want to take it further, with participants who are not all white and who represent diversity in many dimensions, not just traditional race labels that, in reality, encompass a lot of different backgrounds. If we're going to get at the root causes of health and disease, this means understanding the differences and similarities among us all. For example, sickle cell disease occurs when someone inherits two mutated genes for the oxygen-carrying protein hemoglobin. It affects 100,000 African-Americans and more than 20 million people around the world. In contrast, sickle cell trait—meaning just one of these genes is mutated—actually gives people an advantage in surviving malaria, which makes evolutionary sense if your ancestors came from areas such as Africa where malaria is prevalent. New studies, however, have found that sickle cell trait might not be as benign as doctors used to believe, because it may increase the risk for kidney disease. Some African-Americans are more susceptible to this risk and some less. There's clearly more to learn about why this might be the case and about how different DNA variants might interact to affect the health of people with sickle cell trait. The DNA information from more than a million All of Us participants could help researchers learn much more about complex traits like this.

We do have to start with some of the broad-brush categories to recruit enough people to start recognizing the more fine-grained groups among them. Currently we are exceeding our goal of overrepresenting groups that have been historically underrepresented in research. For instance, African-Americans make up about 13 percent of the U.S. population but just 3 percent of the samples previously used in genome studies. In All of Us, 21.5 percent of participants so far are African-American. Similarly, Hispanics constitute about 18 percent of the U.S. population but in 2016 made up less than 1 percent of the data in our genomic databases. Today 17.6 percent of All of Us participants are Hispanic.

That diversity will help us discover more about how DNA affects health across different communities, but the molecule will not be our sole focus. Many factors beyond our genes are at play when it comes to disease. We know that where you were born, what you eat, the stress you feel, and other clinical and biological factors affect health, but we still don't understand by how much. For example, when we think about some of the most common chronic diseases that afflict our population—high blood pressure is one ex-

A new precision medicine project, All of Us, has much larger populations of groups that have been historically underrepresented in genetics research. The project, sponsored by the U.S. National Institutes of Health, began recruiting participants in 2018. More than 250,000 people enrolled by October 2019, and just over 20 percent are black, African-American or African. About 18 percent are Latino, Hispanic or Spanish. Nearly 3 percent are Asian, and 6.7 percent are of mixed races. Slightly less than half of the people are white. The project's goal is to get DNA and other health information from more than one million people.



ample—many of them disproportionately affect the most socially and economically disadvantaged people in our country. And from what we can tell at the moment, the determinants are not simply their race or ethnicity. Risks also include family structure, socioeconomic status, stressors such as trauma, sex and gender inequality, availability of nutrient-rich foods, access to health care, and many other factors that we can capture in the All of Us data set.

Within the next several years, we should be able to compare this rich set of information with participants' DNA. When we do so, scientists such as myself, the All of Us participants and all of you will start to get a clearer picture of the roles that biology and environment play in disease development, and—most important of all—what we can do about it.

Molecular geneticist **Stephanie Devaney** is deputy director of the All of Us research program at the National Institutes of Health. She was the staff lead for the White House Precision Medicine initiative.