

HEALTH

Big Precision Medicine Plan Raises Patient Privacy Concerns

White House's move to develop customized care prompts worries
about data security and informed consent

By Dina Fine Maron on February 3, 2015



Credit: Thinkstock/ImageSource

A new effort to create tailor-made medicine for patients around the U.S. is getting a boost from a \$215-million presidential initiative. It's an ambitious undertaking fraught with concerns about patient privacy, funding and how such data would be stored. But because it's such an innovative idea, there are few blueprints to work with.

The broad federal effort, first announced during Pres. Barack Obama's State of the Union address and then fleshed out with a few more details and a presidential East Room address last week, would create a personal health care information database of more than a million individuals. In addition to patient histories the endeavor would include genetic

data and information from devices like wearable health monitors, and the collection of bacteria, fungi and viruses in and on the body called the microbiome.

Armed with reams of such data scientists hope they could one day offer more personalized medical care, or precision medicine, that would differ from person to person based on their unique genetic makeups and other factors. The end result of the initiative, according to Obama, will be “delivering the right medicine at the right time every time to the right person.” Moreover, as the president envisions it, patients would also be able to access their own data.

Rather than start culling data from scratch, however, the effort aims to tap existing info on patients in clinical trials and incorporate it into the new massive effort. And that’s where it gets complicated, says Kristen McCaleb, program manager of the Genomic Medicine Initiative at the University of California, San Francisco.

Scientists often disagree on the importance or meaning of particular genetic variants for disease. When a sick patient agrees to get his DNA analyzed it triggers a string of decision-making. A doctor may tell the lab to only seek results about specific genes. And once the genome is sequenced, another expert makes a judgment call—ruling if a mutated gene identified by the sequencer is risky or not. Certain mutations, such as variants of the *BRCA1* gene linked to breast cancer, are clearly defined. The significance of many others, however, remains muddier, so two scientists looking at the same list of more than 30,000 genetic variants for each person may have varying opinions about whether or not those genetic mutations are strongly linked to disease or worth exploring further. That ambiguity, McCaleb says, could spell trouble for the president’s precision medicine initiative. “If they plan on incorporating all 30,000 variants coming from one million people, somebody better have a gigantic, honking-fast supercomputer capable of capturing all that raw data,” she says, because otherwise investigators would be relying on a series of

relatively subjective interpretations of that information, making it cumbersome to work with. “As excited as we are that Pres. Obama has made this a priority, there are a lot of logistics to be worked out here,” she says.

Robert Green, the director of a genome research program, G2P, at Brigham and Women’s Hospital in Boston, says that a raw data set from a single genome takes roughly 100 gigabytes of storage. So all that data will also pose a computational challenge. When his team collected 800 genomes for a large Alzheimer’s study, the only way they could practically share the data, other than sending it around on hard drives as they do now, he says, would be to put it on a giant server in the cloud and then researchers could log in to access the server remotely and use analytic tools to explore the massive data set.* “That’s the only way you could access 800 genomes, much less 10,000 or a million,” he says.

Naturally, this gives rise to privacy concerns. When information from one million people is brought together, it would make an attractive target for a hacker working to link the data back to individuals. Such a breach could rob both patients and their families of their privacy. Data for research are typically scrubbed of identifying factors like a patient’s name and birth date, but someone with enough information about an individual’s family tree may be able to connect some dots.

Such data privacy concerns already have a track record of scaring away a segment of potential research subjects. When people agree to be part of an academic study they sign a consent form that says they consent to have their data used in specific ways. Green, for example, heads up a whole genome-sequencing project geared toward incorporating genetic data into clinical medicine. To that end, his team has sequenced the genomes of more than 100 people who agreed to have their personal data shared with large government databases as well as Green’s own biobank. That’s good news for the White House’s precision medicine initiative, says Green, who would like his data sets to be folded

into the effort. But getting people to sign on after they learned all the ways their data could be used did prove challenging, he says. About 25 percent of research participants that bowed out during the consent process—when they were in the office and talking in person—cited fear of health insurance discrimination as the primary reason, he says.

Still other projects, like U.C. San Francisco's, would have to go through an entirely new consent process as well as the time-consuming and expensive effort of recontacting patients. Their patients, McCaleb says, did not sign up to be part of larger databases like this one. And exactly who would pay for the staff time to do that remains unclear. Moreover, with different data sources coming together—say U.C. San Francisco's genome sequencing alongside comprehensive patient histories from the long-standing Framingham Heart Study—different questions were asked and the data were organized quite differently, which, in turn, raises questions about the margin of error on the info when it's all mashed together, she says.

Francis Collins, director of the National Institutes of Health, says that a board will be formed to advise on issues such as privacy and data reliability and to decide who will oversee the initiative and its details. Federal agencies, if awarded the \$215 million outlined in the president's 2016 budget request, would be tasked with creating an easily accessible database with needed privacy protections and streamlining the regulatory approval process for the instruments that would help scientists find the data. Moreover, patient advocates and privacy experts will be at the table, Obama said in his public remarks on January 30. “They won't be on the sidelines, it won't be an afterthought” and we will protect patients in a responsible way, he said. Further details of the proposal, whenever they are released, could help patients decide how protected they should feel.

**Clarification (2/3/14): This sentence was edited after posting to more precisely describe how data from the large Alzheimer's study is currently shared.*

ABOUT THE AUTHOR(S)



Dina Fine Maron

Dina Fine Maron is an award-winning journalist and an editor at *Scientific American* covering medicine and health. She is based in Washington, D.C.

Credit: Nick Higgins

Scientific American is part of Springer Nature, which owns or has commercial relations with thousands of scientific publications (many of them can be found at www.springernature.com/us). Scientific American maintains a strict policy of editorial independence in reporting developments in science to our readers.

© 2018 SCIENTIFIC AMERICAN, A DIVISION OF SPRINGER NATURE AMERICA, INC.

ALL RIGHTS RESERVED.