

Health records and genetic data from more than 100,000 Californians power medical research

June 19 2015

By volunteering to mail saliva to researchers working with their health care provider, thousands of people in California have helped build one of the nation's most powerful medical research tools. The researchers have now published the first reports describing these volunteers' genetic characteristics, how their self-reported ethnicity relates to genetic ancestry, and details of the innovative methods that allowed them to complete DNA analysis within 14 months. The articles are published in the journal *GENETICS*.

"This is an incredible treasure trove of data. The information collected during medical care is much more comprehensive than the isolated measurements we would make in a traditional research study," says project co-principal investigator Neil Risch, University of California, San Francisco (UCSF). "By linking these clinical records with genomic data from each person, we now have the power to track down many genetic and environmental contributions to disease."

The data have already been used to investigate many diseases. For example, researchers have pinpointed genetic variants linked to prostate cancer, allergies, glaucoma, macular degeneration, diabetes, high cholesterol, and many more. "No matter which disease we've looked at, we found genetic variants that influence it. And the beauty of this dataset is that it covers countless diseases and traits, and the medical records are constantly being updated as the cohort grows older," says Risch.

The Genetic Epidemiology Research on Adult Health and Aging (GERA) resource was created in 2009 by a collaboration between the Kaiser Permanente Northern California Research Program on Genes, Environment, and Health (RPGEH) and the Institute for Human Genetics at UCSF.

The RPGEH is an ongoing study of more than 200,000 members of the Kaiser Permanente Medical Care Plan who have consented to share data from their electronic medical records with researchers, along with answers to survey questions on their behavior and background. The records include clinical, pharmacy, and laboratory test information. Participants also contributed saliva samples, and more than 100,000 of these samples were selected for genetic analyses performed at UCSF. These participants form the GERA cohort.

Because the average age of participants in the GERA cohort is 63, the GERA research team is focusing their efforts on aging-related diseases.

The data have been available to researchers through an application and review process managed by the RPGEH. Last year, the genetic data were also made available to the research community via the NIH program dbGaP. "The goal was to create a resource that many research groups could mine for genetic insights into a broad range of diseases," says Catherine Schaefer, GERA co-principal investigator and executive director of the RPGEH at Kaiser Permanente.

The new publications present crucial details of the methodology used to create the comprehensive genetic information on participants - including the length of their telomeres—chromosome caps that influence age-related diseases, as well as their genetic ancestry.

In one article, the team describes how they were able to process more than 100,000 samples—characterizing 70 billion genetic variants—all

within the two years dictated by their funding. "In 2009 this was a huge task, it hadn't been done this fast before," says co-author Pui-Yan Kwok, of the Institute for Human Genetics, UCSF. "The assays ran 24/7, so we had to develop new processes for analyzing data in real time to alert us to any problems as soon as they happened. We also had to boost the analysis quality to make best use of the data."

Because approximately 20% of people in the study were from minority groups, the researchers improved the analysis by developing four separate ethnicity-specific gene analysis arrays, or "gene chips." Each chip was tailored to the genetic variants common in either non-Hispanic whites, African Americans, East Asians, or Latinos.

Part of the reason for ensuring the study group was ethnically diverse was to redress the traditional overrepresentation of people with European ancestry in genomic studies. One of the new articles presents a detailed [genetic ancestry](#) study, including the relationship between the genetic results and self-reported ethnicity in the cohort.

Participants indicated their race or ethnicity by selecting from as many as twenty three different race/ethnicity/nationality categories in a questionnaire. Across all possible combinations of these categories, over 50 different race/ethnicity identities were represented in the study.

"We were particularly interested in those who checked off more than one box," says Risch. "More and more people are identifying as multi-ethnic, which can pose some technical challenges for genomic studies. At the same time, it also presents opportunities for analyzing genetic and social contributions to disease differences between groups."

People who identified as multi-ethnic were younger on average than those who chose a single ethnicity, which likely reflects increasing intermarriage and social change. People who identified as a different

ethnicity than their genetic siblings also tended to report a multi-ethnic identity.

The researchers have also published in *GENETICS* the methods they developed for automated measurement of telomere lengths. Telomeres are protective bundles of DNA and protein that cap the ends of chromosomes. Telomere DNA tends to erode with age, which leaves the chromosomes vulnerable to damage, and some disease risks have been linked with shorter telomere length.

The telomere work was led by the UCSF research group of Elizabeth Blackburn, who was awarded a Nobel Prize in 2009 for the discovery of telomeres.

The very large volume of samples to be processed meant the team had to develop a high-throughput robotic system that completed the laboratory tests in four months.

"This is the largest telomere length database ever constructed from a single study population," says Blackburn. "At the start, some were skeptical that we could get reliable data from saliva. But we had a 96 percent success rate, and the results are in fact highly consistent with conclusions from studies of blood."

The analysis confirmed that telomere lengths tended to be longer in women than men and to decline with age. And a remarkable surprise emerged: for those over 75, older people tended to have longer, not shorter telomeres. This suggests that in people older than 75, longer telomeres are associated with longer life. The team is also examining correlations between telomere length and disease, as well as behavioral and environmental factors.

"This project is one of the earliest examples of precision medicine in the

US, an approach that takes into account differences in the genes, environment, and lifestyles of individuals and leverages large clinical datasets to identify these individual risk factors," says Schaefer. "The powerful GERA resource is just a taste of what is to come."

More information: *Genetics*. Early Online June 19, 2015. [DOI: 10.1534/genetics.115.178616](https://doi.org/10.1534/genetics.115.178616)

Provided by Genetics Society of America

Citation: Health records and genetic data from more than 100,000 Californians power medical research (2015, June 19) retrieved 12 May 2024 from <https://medicalxpress.com/news/2015-06-health-genetic-californians-power-medical.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.