

Researchers release analysis of largest, most diverse genetic data set

February 11 2021



Credit: CC0 Public Domain

Researchers at the University of Maryland School of Medicine (UMSOM) and their colleagues published a new analysis today in the journal *Nature* from genetic sequencing data of more than 53,000 individuals, primarily from minority populations. The early analysis, part of a large-scale program funded by the National Heart, Lung, and Blood Institute, examines one of the largest and most diverse data sets of high-

quality whole genome sequencing, which makes up a person's DNA. It provides new genetic insights into heart, lung, blood and sleep disorders and how these conditions impact people with diverse racial and ethnic backgrounds, who are often underrepresented in genetic studies.

The program, called Trans-Omics for Precision Medicine (TOPMed), seeks to understand the genetic variations that occur among individuals both in nuclear families and in populations from diverse ethnicities residing on different continents. The project's ultimate goal is to improve the diagnosis, treatment and prevention of the most common conditions that lead to disability or death.

"We have already identified some surprising new insights," said study corresponding author Timothy O'Connor, Ph.D., Associate Professor of Medicine & Endocrinology at the Institute for Genome Sciences (IGS) at UMSOM. For example, the team identified more than 400 million genetic variations, but 97 percent of them are extremely rare, occurring in less than 1 percent of the population. Gene variations or variants can occur by random chance when genes get recombined or mutate.

"Most of the time, these variants mean nothing," said Dr. O'Connor, "but they can provide a new understanding of mutational processes and recent human evolutionary history."

The TOPMed team includes more than 180 researchers from leading institutions in genomics worldwide who have been compiling huge datasets in systematic and defined ways to increase knowledge about diversity in [genetic studies](#). Since its launch in 2014, the TOPMed investigators have begun adding [whole genome sequencing](#) and "omics" analysis (which includes a study of genetic and molecular profiles like proteins) to research studies in order to better understand how variations affect different organ systems giving rise to disease in, for example, the heart and lungs.

In the new *Nature* paper, the researchers pointed out that the program "aims to identify causal genetic variants and how they interact with the environment, to characterize disease and its molecular subtypes, to understand differences in disease across diverse ancestries, and to establish a foundation for personalized disease prediction, prevention, diagnosis, and treatment." Braxton Mitchell, Ph.D., Professor of Medicine at UMSOM, and Jeffrey O'Connell, Ph.D., Associate Professor of Medicine at UMSOM, were co-authors on this paper.

TOPMed is the largest sequencing project to date and has identified over 400 million gene variants with an overarching mission of understanding global genetic diversity. Since joining the TOPMed program in 2016, UMSOM researchers have published valuable new insights on genetic diversity including sequencing data from the initial flagship paper on the first 53,831 TOPMed samples.

The increasing diversity of the population samples will help investigators learn more about how specific diseases impact different ethnic populations around the world. In addition, the group has established uniform standards for sequencing performed on a massive scale. The standards maximizes the integrity of the data as the large group of international researchers use uniform methods as they continue to add other "omics" methods for analysis such as the study of metabolic differences.

In addition to enabling detailed analysis of the combined genomic and health data for sequenced samples, TOPMed has enhanced the analyses of genotyped samples through a new reference panel that now includes over 97,000 individuals. The TOPMed imputation reference panel is publicly available for review and input of new genetic data by researchers.

The first stage of the data release in the *Nature* study demonstrated a

greater inclusion of a diversity of sampling, which will be invaluable to the international group to learn more about the diseases impacting these populations. Because of the vast sample sizes and the longitudinal scope of many of the population samples, the investigators were able to demonstrate that the rare variants represent recent and potentially deleterious changes that can impact protein function, gene expression or other biologically important elements.

"This is a major effort to rectify the underrepresentation of minority participants in genomic studies and tracks with a broader mission within the School of Medicine to increase diversity in clinical trials," said E. Albert Reece, MD, Ph.D., MBA, Executive Vice President for Medical Affairs, UM Baltimore, and the John Z. and Akiko K. Bowers Distinguished Professor and Dean, University of Maryland School of Medicine. "This will hopefully move the genomics field closer to extending personalized medicine for all patients."

Cashell Jaquish, Ph.D., an NHLBI program officer for TOPMed and a corresponding author on the *Nature* paper, agrees. "The NHLBI's TOPMed program is a huge resource for the scientific community. We didn't really know what genomic variation looked like in diverse groups until now. This new study represents truly historic findings and we look forward to continued research studies in this area as we move toward personalized medicine."

More information: Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program, *Nature* (2021). [DOI: 10.1038/s41586-021-03205-y](https://doi.org/10.1038/s41586-021-03205-y)

Provided by University of Maryland School of Medicine

Citation: Researchers release analysis of largest, most diverse genetic data set (2021, February 11) retrieved 25 April 2024 from <https://medicalxpress.com/news/2021-02-analysis-largest-diverse-genetic.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.