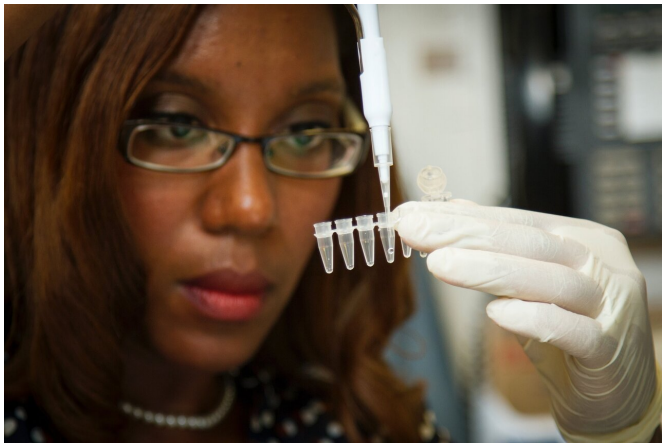
Using DNA sequencing data with electronic health records to find rare variants behind inherited diseases

12 January 2021, by Bob Yirka



Credit: Unsplash/CC0 Public Domain

A large team of researchers affiliated with a large number of institutions in the U.S. has used DNA sequencing data and information stored in electronic health record databases to search for rare genetic variants that may be behind some inheritable diseases. In their paper published in the journal *Nature Medicine*, the group describes how they analyzed DNA sequencing data and used it to help look for unknown gene variants behind some inherited diseases and what they found.

Prior research has shown that the presence of genetic variants can lead to inherited diseases such as diabetes type I. But prior research has also found that there are likely many more rare variants that are not yet known to medical science. In this new effort, the researchers sought to find such variants using health record databases of genetic information for very large numbers of people. In their effort, they used data from the Penn Medicine Biobank—a database created and used by the University of Pennsylvania Health System. For several years, students, faculty and even visitors to the campus have been encouraged to donate blood or tissue samples for use in research efforts.

The team used DNA sequencing information to search through data from 10,900 volunteers who had donated to the Biobank, looking for people who were carriers of genes that might be behind a host of genetic diseases. Genes that were found to have a minimum of 25 carriers of variants known to result in loss-of-function genetic ailments were then analyzed to determine if there were any associated phenotypes. This led to the discovery of 97 phenotype associations. The researchers then replicated 26 of their original findings on several other research databases—five of the phenotype associations were known to play a role in certain diseases. The other 21 represented new discoveries.

The researchers suggest their findings showcase the value of using electronic health record databases as tools for searching for, and finding, variants involved in inheritable diseases.

More information: Joseph Park et al. Exome-wide evaluation of rare coding variants using electronic health records identifies new gene–phenotype associations, *Nature Medicine* (2021). DOI: 10.1038/s41591-020-1133-8