Researchers develop guidelines for large-scale sequence-based complex trait association studies

September 27 2016

Precision medicine, which utilizes genetic and molecular techniques to individually tailor treatments and preventative measures for chronic diseases, has become a major national project, with President Obama launching the Precision Medicine Initiative in 2015. In a study published today in the *American Journal of Human Genetics*, a team of researchers from Baylor College of Medicine and partner institutions, which include the University of Washington and the Broad Institute, detail their findings from the National Heart Lung and Blood Institute Exome Sequencing Project (ESP) and explain how these results can contribute to the advancement of precision medicine and the implications for the future of whole exome data analysis.

The goals of the study were to identify associations between a number of heart, lung and blood-related complex traits with rare variants and to advance the goals of precision medicine. To accomplish these goals, the research team studied samples from 7,034 individuals who were identified based on two selection strategies, sampling the extremes of quantitative traits and the selection of individuals who were relatively young at the onset of disease.

"This study is exciting because it helped us identify new links and insight into why certain people are more prone to problems like heart disease and changed our understanding of the human genome," said Dr. Deborah Nickerson, professor of genome sciences at the University of

Washington. "We were able to develop a new way of releasing the data anonymously so it can be shared by researchers and institutions across the country."

Of the 7,034 people sequenced, two-thirds (4,405) of the group were European American, and one-third (2,954) were African American, with the remaining 35 individuals identifying with another ancestry.

Twelve primary clinical disease-related traits were examined in the study, including acute lung injury, asthma, chronic obstructive pulmonary disease, early onset myocardial infarction, ischemic stroke, type 2 diabetes with obesity as a co-morbidity, and pulmonary arterial hypertension-systemic sclerosis, along with several quantitative cardiovascular risk factors. Fifty-nine secondary traits also were analyzed.

The team performed an association analysis on the data for the 71 different traits, running the tests separately for the European American and African American groups, providing an element of diversity in the sample set and the opportunity to study the differences in the allelic architecture of rare variants between the groups.

When the ESP project was originally designed in 2009, deep sequencing of large data sets was very costly and much more restricted in its capability. The initial findings from the study were translated and analyzed and led the research team to conclude that the findings needed to be tested by a larger sample size because the complex traits were actually less common than initially theorized.

"I've always had an interest in analyzing rare variants," said Suzanne Leal, professor of molecular and human genetics at Baylor College of Medicine. "Although many methods are available to analyze rare variants, including our Combined Multivariate Collapsing method, which

was the first rare variant association method published in 2008, we were somewhat limited by the lack of software to analyze large data sets, so it is exciting to now have developed association tools that we can apply to a large data set."

One of the more surprising results of the study was the discovery of the relationship between APOC3 and triglyceride levels. The rare variants present in APOC3 reduce triglyceride levels, which impacts the risk for coronary heart disease.

Researchers noted that, while many common genetic variants are shared across the globe, most rare variants are unique to closely related populations. This suggests that the complexity observed in many traits is partly due to the recent explosion of population growth, so association testing for rare variants must be done on a much larger scale than those done for common variants.

This research shows that, although rare variants are involved in complex traits, more research needs to be done to find the associations and to elucidate what the pathogenic susceptibility looks like.

The ESP helped to establish best practices to turn terabytes of raw sequence data into innovative genetic discoveries for complex traits and diseases and sparked the development of statistical methods and software to tackle both extreme trait sampling and rare variant association testing.

As sample sizes continue to increase with decreasing sequencing costs, the process of data generation will not pose a major technical barrier for future sequencing-based studies of rare variant associations, and the ESP model of data sharing and rapid analysis of large scale sequence data can be emulated.

"Future studies will focus on exome data and whole genome [sequence data](#), and it is thrilling to demonstrate that this type of large data analysis is feasible. Sequenced data is hiding important findings and associations that could have a huge impact on public health, and it's exhilarating to play a part in solving these puzzles," concluded Leal.

"This project identified many new relationships between inherited genetic differences and the risk for developing diseases such as heart disease and stroke, and this research helped to identify techniques that are "best practices" for large-scale genetic projects for human health. The experiences that we share in this paper help set the stage for President Obama's recently announced "Precision Medicine Initiative" and what might be expected from any genetics related research," said Dr. Paul Auer, first author and assistant professor of biostatistics at the Joseph J. Zilber School of Public Health at the University of Wisconsin-Milwaukee.

  More information: Paul L. Auer et al. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project, *The American Journal of Human Genetics* (2016). [DOI: 10.1016/j.ajhg.2016.08.012](#)

Provided by Baylor College of Medicine