

Complex, large-scale genome analysis made easier

June 16 2015



The mSet algorithm by Oliver Stegle at EMBL-EBI makes large-scale, complex genome analyses easier. Credit: Spencer Phillips, EMBL-EBI

Researchers at EMBL-EBI have developed a new approach to studying the effect of multiple genetic variations on different traits. The new algorithm, published in *Nature Methods*, makes it possible to perform genetic analysis of up to 500,000 individuals - and many traits - at the same time.

The relationship between genes and specific traits is more complicated than simple one-to-one relationships between genes and diseases. Genome-wide association studies (GWAS) show that many genetic factors are at play for any given trait, but scientists are just beginning to explore how, specifically, genetic variations affect health and disease. Two major statistical challenges to finding these connections involve analysing associations between many different genetic variants and multiple traits, and making the best use of data from large cohorts that include hundreds of thousands of individuals.

"It is very challenging to identify genetic variants that underlie phenotypes, or traits, and usually we do this by analysing each phenotype and each variant one by one," explains Oliver Stegle, Research Group Leader at EMBL-EBI. "But the simple models we use to do this are too simplistic to uncover the complex dependencies between sets of genetic variants and disease phenotypes."

Complex models that let you look at the combined action of many different variants have, until now, involved so much computation that it would take a year to run a single complex query.

"The breakthrough here is that we've made it possible to perform an integrative analysis involving many variants and phenotypes at the same speed as current approaches," says Oliver.

The researchers tested their [algorithm](#) on data from two studies from public repositories, and compared the results with existing state-of-the-

art tools. Their study of four lipid-related traits (LDL and HDL cholesterol levels, C-reactive protein, triglycerides) proved that the new method is substantially faster, and can explain a larger proportion of these traits in terms of the genetics that drive them.

"We wanted to be able to look at these questions from both directions," says Oliver. "On the one hand, we want to look at all the variants in a single gene that may be involved in the regulation of one particular lipid trait. On the other, we want to look at the combined effect across larger sets of lipid levels, for example to find out something about lipid regulation in general."

Using the new method, GWAS researchers can explore several variants of a gene at once while comparing them with several related phenotypes. This makes it much easier to pinpoint which genes - or locations on genes - are involved in a particular function, such as lipid regulation.

"What's important about this work is that it improves statistical power and provides the tools people need to analyse multiple traits in very large cohorts," says Oliver. "Our algorithm can be used to study up to half a million individuals - that hasn't been possible until now."

"Currently, people are either using multiple variant methods on one phenotype, or multiple phenotype methods but looking at just one variant at a time. Oliver's new scheme is a real advance because it lets you do both at the same time, and is scalable to be used on the very large cohorts we are starting to see in initiatives like the UK BioBank," says Ewan Birney, Associate Director at EMBL-EBI.

The new algorithm provides much-needed methods for genomics, making large-scale, complex analysis a manageable and practical endeavour.

"Our method, which we call mSet, provides a principled approach to testing for statistical relationships between multiple genetic variants and groups of traits. These methods will help researchers determine which specific aspects of our biology are inherited, and uncover new insights into the genetics behind our countless biological processes."

More information: Casale FP, et al. (2015). Efficient set tests for the genetic analysis of correlated traits. *Nature Methods*. Published online 15 June; [DOI: 10.1038/nmeth.3439](https://doi.org/10.1038/nmeth.3439)

Provided by European Molecular Biology Laboratory

Citation: Complex, large-scale genome analysis made easier (2015, June 16) retrieved 30 April 2024 from <https://medicalxpress.com/news/2015-06-complex-large-scale-genome-analysis-easier.html>

<p>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.</p>
--