

# Researching the roles of rare genetic variants in disease

November 15 2013, by Seth Palmer

---

As scientists work toward further personalizing medical treatment through genomics, heritability—the proportion of observed variation in a particular trait that can be attributed to inherited genetic factors—is key to understanding more precisely how a person's DNA contributes to risk factors for such hereditary diseases as Alzheimer's, Parkinson's syndrome and various cancers.

The process of determining heritability, however, is tedious and often fruitless, as [genetic variation](#) can be extremely difficult to assess, according to says Marylyn Ritchie, associate professor of biochemistry and molecular biology at Penn State and director of the Center for Systems Genomics, part of the Huck Institutes of the Life Sciences. Studies often require thousands of participants in both "case" and "control" groups, and in the case of [rare genetic disorders](#), tens or even hundreds of thousands of participants might be required in order to generate enough data to link a given mutation or set of mutations to a particular condition.

"Working with DNA sequence data, you'll get the variants in the genome that are common and shared among people, and then you'll also get rare variation—base changes that are unique to individuals or at least less-common in a population," Ritchie explains. "We typically do studies with thousands of people, but to study rare variation, you either need to get tens or hundreds of thousands of people—which is not cost-effective—or you need to do some other type of analysis to try to work with those rare variants. So we're trying to develop new algorithms and

tools to analyze those data."

Rather than analyzing each DNA base independently, a common approach to studying rare genetic variation is to use a software program to "bin" together all the variants within a gene and count how many of the subjects with a disease have any variation in that gene. Those data are then compared with data from a control group in order to find out which variants may be significant in the context of the disease.

"That looks like a promising approach," Ritchie says, "but the limitation is that the researcher has to annotate and subsequently bin the data in a very manual way, and it's a very arduous process—it takes a lot of effort, and you can only annotate and bin the variants based on what knowledge you already have or what you can gather from other data sources to figure out how they go together."

So Ritchie and her colleagues developed a computer program called BioBin to automate the annotation process with genomic data compiled from a number of public databases.

"What we've done," says Ritchie, "is written an algorithm and a software package to go with it that will—in an automated way—process all the sequence data that you have, annotate either what gene or region of the genome that sequence belongs to, whether it's in a coding or regulatory region, part of a pathway, in an evolutionarily conserved region or one that's undergoing natural selection, or if it's between genes, and then bin all of the variants together based on these different functional definitions. And you can export those data to do association testing—comparing cases and controls to see whether their genetic pathways are different, if people with a disease have more variation in certain pathways or regulatory regions or evolutionarily conserved regions than unaffected individuals."

Since developing BioBin, Ritchie and her lab have used it to analyze several genomic datasets from dbGaP—the database of Genotypes and Phenotypes, hosted by the National Center for Biotechnology Information—in addition to performing a proof-of-concept analysis with the newly released 1000 Genomes Phase I data.

"We used the 1000 Genomes data to compare genetic variation between 14 ancestry groups from different continents," Ritchie says, "where there should be a lot of variation because of the differences in ancestry; and we showed that with our tool, you can pick up the genes and pathways that are different between the populations.

"We've also used BioBin to study variation in individuals with Kabuki syndrome—which is a rare disease—and we've applied it to a cystic fibrosis (CF) dataset that we're still working with, trying to figure out if there is underlying genetic variation that makes certain people with CF more susceptible to a severe lung infection called *Pseudomonas aeruginosa*, which occurs in a lot of CF children; the infection doesn't occur in all individuals with CF, so it does seem like there is some either genetic or environmental susceptibility that's also there, and we're looking to see if there's genetic susceptibility based on rare variants."

With an eye toward the translational research that will bring the benefits of her work to the public, Ritchie sees applications being developed using BioBin and [genomic data](#) to personalize medical treatment—particularly chemotherapeutic drugs for cancer patients.

"A lot of my collaborators are working with cancer drugs," says Ritchie, "and we're doing some analyses of those data now to figure out if we can understand what genetic variation is explaining responses to different chemotherapeutics. We're also doing some work related to cardiovascular traits. We've done the traditional analyses to figure out the common variation that explains response, but now we're doing

analyses of gene-gene and gene-environment interactions—taking a systems biology approach to looking at not only DNA variation, but also gene-expression variation, and using BioBin to look for interactions between common variants and rare variants.

"We're working right now to make the software available to the public—it's on our website, and has been downloaded by a few groups, but we're writing the paper that goes with it to explain how to use it and everything it can be used for. Our future work, then, will be adding additional data sources, other ways to bin the data, and expanding our analysis to other disease-related sequencing datasets."

**More information:** *BMC Medical Genomics* 2013, 6(Suppl 2):S6 [DOI: 10.1186/1755-8794-6-S2-S6](https://doi.org/10.1186/1755-8794-6-S2-S6)

Provided by Pennsylvania State University

Citation: Researching the roles of rare genetic variants in disease (2013, November 15) retrieved 4 May 2024 from <https://medicalxpress.com/news/2013-11-roles-rare-genetic-variants-disease.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>
--