

TOOLBOX

Algorithm fine-tunes analysis of rare variants

BY JESSICA WRIGHT

22 AUGUST 2012

Signal spots: Microarrays can detect the presence of genetic variants among large groups of individuals.

A new algorithm increases the accuracy of techniques that detect rare genetic variants among populations, according to a study published 27 July in *Bioinformatics*¹.

Researchers often seek to link rare single nucleotide polymorphisms (SNPs), which alter a single DNA base, with the likelihood of having a disease. For example, SNPs that lower the activity of a molecular pathway and are more common in individuals with autism than in controls suggest a link between the pathway and the disorder.

To identify SNPs in thousands of samples, researchers use a technique called microarray analysis. This method uses radioactive probes that bind to different base pairs, generating a signal. For example, a B variant at a site that is typically A can be either AB or BB — depending on which variant is inherited from parents — each leading to a different signal.

A standard microarray analysis, such as one called GenCall, defines specific cutoffs for each category. A signal that does not fit any of the categories is designated a 'no-call,' and can occur with a poor reading or if DNA at that site is missing because of a mutation or a deletion.

In the new study, researchers analyzed DNA from 9,380 samples in an ongoing schizophrenia

study. After using GenCall, they calculated new thresholds for the categories based on the deviations from the mean signal for each category. This second step, which they dub zCall, can reclassify signals on the cusp between two categories.

The researchers verified zCall's ability to correctly identify SNPs using two datasets in which the protein-coding parts of the genome have been sequenced: 947 samples from a Swedish dataset and 369 samples from the [ARRA Autism Sequencing Collaborative](#).

Adding zCall improved the accuracy of microarray analysis from about 92.5 percent to 99 percent for the Swedish dataset and 93 percent to 99 percent for samples from the ARRA collaborative. ZCall is available for use [online](#).

References:

1: Goldstein J.I. *et al.* *Bioinformatics* Epub ahead of print [PubMed](#)