FAST BREAKING PAPERS - 2009

February 2009

PDF

Rafael Irizarry talks with *ScienceWatch.com* and answers a few questions about this month's Fast Breaking Paper in the field of Mathematics.

Article Title: Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data
Authors: Carvalho, B;Bengtsson, H;Speed, TP;Irizarry, RA
Journal: BIOSTATISTICS
Volume: 8
Issue: 2
Page: 485-499
Year: APR 2007
* Johns Hopkins Univ, Dept Biostat, Baltimore, MD 21205 USA.
* Johns Hopkins Univ, Dept Biostat, Baltimore, MD 21205 USA.
* Univ Calif Berkeley, Dept Stat, Berkeley, CA 94720 USA.
* Walter & Eliza Hall Inst Med Res, Div Genet & Bioinformat, Melbourne, Vic, Australia.

SW: Why do you think your paper is highly cited? Does it describe a new discovery, methodology, or synthesis of knowledge?

Genome-wide association studies (GWAS) are used to discover genes underlying heritable disorders. The number of GWAS has skyrocketed in the past two years. Microarrays are the genotype calling technology of choice in GWAS as they permit exploration of more than a million single nucleotide polymorphisms (SNPs) simultaneously.

The starting point for the statistical analyses is to convert raw microarray intensities into genotype calls. We have much experience analyzing raw data and, in this paper, we describe our solution to genotype calling. We made sure to make the method robust to batch effects. It turns out the batch effect is quite problematic in GWAS as large datasets are processed on different days, utilizing different PCR reactions. GWAS data analysts are realizing our new methodology is a better solution than default procedures.

"...it is rarely the case that GWAS randomize or control for plate when storing samples."

SW: Would you summarize the significance of your paper in layman's terms?

A logistics problem with large GWAS is that processing occurs in batches. Because DNA samples are stored in 96-well plates and robots make it convenient to run all samples in a plate at once, plates are usually confounded with hybridization times. To make matters worse, it is rarely the case that GWAS randomize or control for plate when storing samples.

Therefore, it is common that plate and outcome of interest are confounded. Thus, if genotype algorithms do not appropriately adjust for batches, it will be difficult, if not impossible, to distinguish real from artificial associations. Our algorithm appears to be more robust to batch effects than other methods.

Changing algorithms can greatly reduce the chance of false positives and therefore increase our chances of making significant findings.

SW: How did you become involved in this research, and were there any problems along the way?

I have collaborated with researchers involved in GWAS. I believed that, given my expertise in the analysis of raw microarray data, I could be of some help.

SW: Where do you see your research leading in the future?

I will continue to develop methods for raw data from high-throughput technologies with the hopes of improving signal-to-noise ratios.

Rafael A. Irizarry
Professor
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD, USA

.•> ***Additional Information:*** Rafael Irizarry has been named a Current Classics scientist (Math.) for Apr. 2008. Also view a commentary from a past New Hot Paper feature, and a Podcast (MP3¦WMA) added Sep. 16, 2008.

Keywords: genome-wide association studies, underlying heritable disorders, microarrays, single nucleotide polymorphisms, raw microarray intensities, genotype calls, raw microarray data, high-throughput technologies, signal-to-noise ratios.

PDF

back to top

2009 : February 2009 - Fast Breaking Papers : Rafael Irizarry

Scientific Home | About Scientific | Site Search | Site Map