# GENOME-WIDE ASSOCIATION MAPPING AND RARE ALLELES: FROM POPULATION GENOMICS TO PERSONALIZED MEDICINE

**FRANCISCO M. DE LA VEGA**,
Life Technologies Foster City, CA 94403, USA Francisco.delavega@lifetech.com

**CARLOS D. BUSTAMANTE**, and
Department of Genetics, Stanford University Stanford, CA, USA cdbustamante@stanford.edu

**SUZANNE M. LEAL**
Department of Molecular and Human Genetics, Baylor College of Medicine Houston, TX 77030, USA sleal@bcm.edu

Genome-wide associations studies (GWAS) have been very successful in identifying common genetic variation associated to numerous complex diseases [1]. However, most of the identified common genetic variants appear to confer modest risk and few causal alleles have been identified [2]. Furthermore, these associations account for a small portion of the total heritability of inherited disease variation [1]. This has led to the reexamination of the contribution of environment, gene-gene and gene-environment interactions, and rare genetic variants in complex diseases [1, 3, 4]. There is strong evidence that rare variants play an important role in complex disease etiology and may have larger genetic effects than common variants [2].

Currently, much of what we know regarding the contribution of rare genetic variants to disease risk is based on a limited number of phenotypes and candidate genes. However, rapid advancement of second generation sequencing technologies will invariably lead to widespread association studies comparing whole exome and eventually whole genome sequencing of cases and controls. A tremendous challenge for enabling these "next generation" medical genomic studies is developing statistical approaches for correlating rare genetic variants with disease outcome.

The analysis of rare variants is challenging since methods used for common variants are woefully underpowered. Therefore, methods that can deal with genetic heterogeneity at the trait-associated locus have been developed to analyze rare variants. These methods instead analyzing individual variants analyze variants within a region/gene as a group and usually rely on collapsing. They can be applied to both in cases vs. controls and quantitative trait studies are needed. The paper of Bansal et al. in this volume describes the application of a number of statistical methods for testing associations between rare variants in two genes to obesity. The authors considered the relative merits of the different methods as well as important implementation details, such as the leveraging of genomic annotations and determining p-values.

Knowledge of haplotypes can increase the power of GWAS studies and also highlight associations that are impossible to detect without haplotype phase (e.g. loss of heterozygosity). Even more complicated phase-dependent interactions of variants in linkage equilibrium have also been suggested as possible causes of missing heritability. In their work, Hallsorsson et al. formulate algorithmic strategies for haplotype phasing by multi-assembly of shared haplotypes from next-generation sequencing data. These methods would allow testing haplotypes harboring rare variants for association and potentially increase their explanatory power.

Since single SNP tests are often underpowered in rare variant association analysis, Zeggini and Asimit propose a locus-based method that has high power in the presence of rare variants and that incorporate base quality scores available for sequencing data. Their results suggest that this multi-marker approach may be best suited for smaller regions, or after some filtering to reduce the number of SNPs that are jointly tested to reduce loss of power due to multiple-testing adjustments.

Finally, the paper of Zhou et al., presents a penalized regression framework for association testing on sequence data, in the presence of both common and rare variants. This method also introduces the use of weights to incorporate available biological information on the variants. Although these tactics improve both false positive and false negative rates, they represent an incremental development and there is still significant room for improvement.

With the development of sequencing technologies and methods to detect complex trait rare variant associations many new and exciting discovery are imminent. The analysis of rare variants is still in its infancy and the next few years promises to produce many new methods to meet the special demands of analyzing this type of data.

## References

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Proc Natl Acad Sci U S A. 2009; 106(23):9362–9367. [PubMed: 19474294]

2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

3. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. PLoS Biol. 2010; 8(1):e1000294. [PubMed: 20126254]

4. McClellan J, King MC. Cell. 2010; 141(2):210–217. [PubMed: 20403315]