Novel Statistical Methods for Gene-Environment Interactions in Complex Diseases

Rationale and Progress Report

Kung-Yee Liang Tom Louis Ingo Ruczinski

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

May 5, 2008

Specific Aims

1. Develop and evaluate new statistical methods to prioritize genes through proper ranking in genome-wide association (GWA) studies that address GxE interactions.

2. Develop and evaluate new statistical methods to localize causal genes as part of linkage and fine mapping studies while considering GxE interactions.

3. Develop and evaluate new statistical methods to identify higher order interactions between environmental variables and SNPs in candidate genes studies.

Specific Aims (cont.)

4. Adapt existing and develop new statistical methods to address imprecise and missing environmental and genetic measurements.

5. Develop and disseminate efficient algorithms for GxE analyses, and apply these methods in several ongoing genetic studies of complex diseases.

Localizing Disease Genes

With multiple markers in ASPs, one has

$$\begin{aligned} E(S(t)|x_1, x_2) &\equiv 1 + (1 - 2\theta_{t,\tau})^2 \times \{E(S(\tau)|x_1, x_2) - 1\} \\ &\equiv 1 + (1 - 2\theta_{t,\tau})^2 \times C(x_1, x_2) \end{aligned}$$

$S(t)$	IBD sharing at marker t.
τ	Location of disease gene.
x_1 and x_2	*Covariates* for sib one and two.
$C(x_1, x_2)$	Genetic effect at trait locus τ.

Localizing Disease Genes

Inference on $C(x_1, x_2)$ allows investigators to

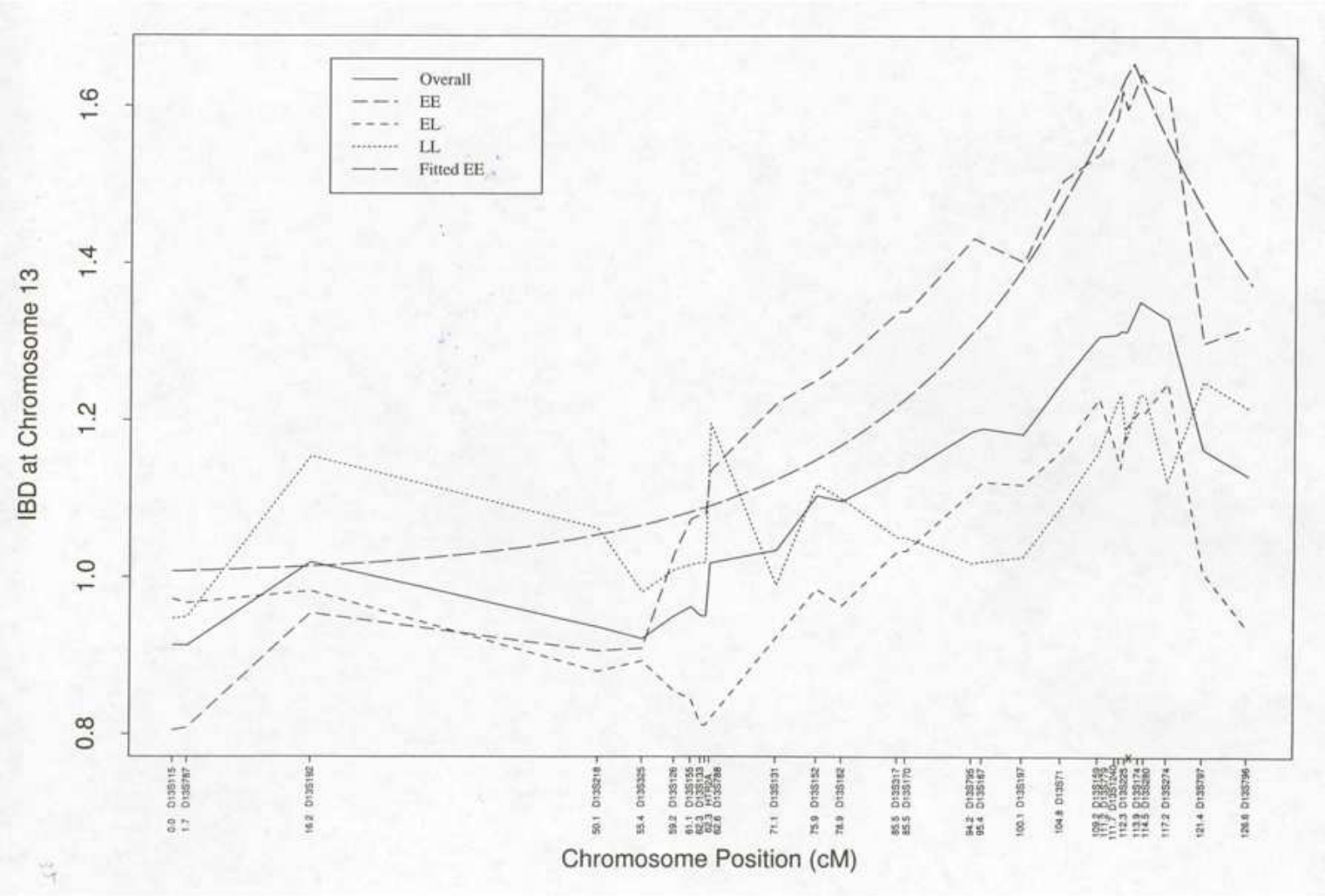
- Test hypothesis of GxE interaction,
- Enhance ability to detect genetic linkage,
- Provide more precise estimate of disease gene,
- Identify subgroup for intervention/prevention.

Similar ideas apply to fine mapping approaches for association studies with either population-based or family-based controls.

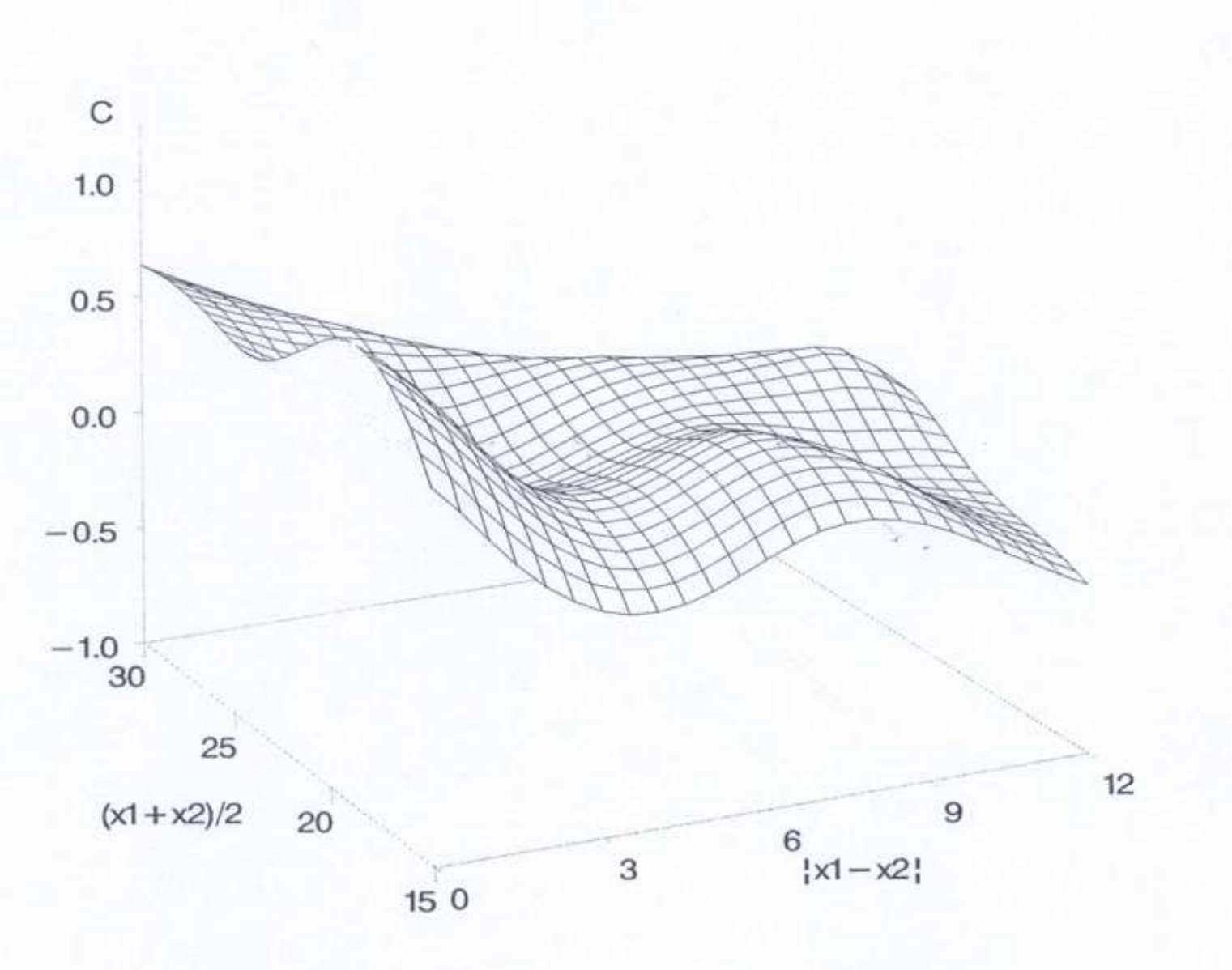
$C(x_1, x_2)$: the difference in probability in carrying the disease allele at trait locus between cases* and controls**

* Transmitted allele ** Non-transmitted allele

Localizing Disease Genes



Localizing Disease Genes



Extending Logic Regression Methods

We propose to enhance the currently existing logic regression methodology in several ways:

- Extend logic regression for trio data with affected probands. This includes the handling of missing data, simulation set-ups, and the actual analysis.
- Investigate alternatives to Monte carlo logic regression. In particular, derive more meaningful measures of variable importance.
- Investigate penalty terms and loss functions for the likelihoods used in logic regression.

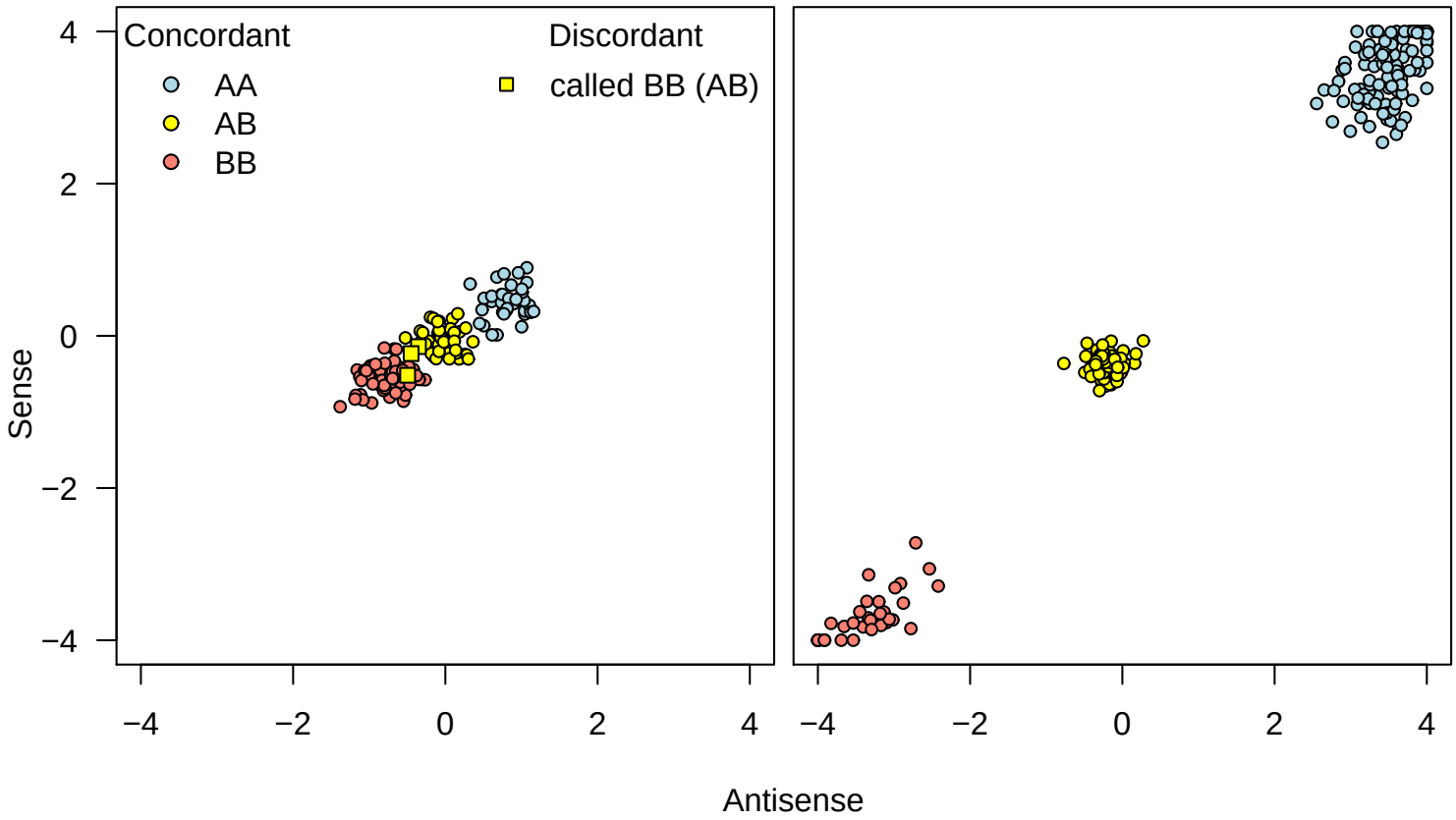
Addressing Imprecise and Missing Data

In genetic studies, genomic data can be missing, and genotyping errors can occur. The same holds true for environmental variables. In a stochastic sense, the structure of missing data and errors in variables are very similar. Missing data likelihood methods and multiple imputation are often practical and effective.

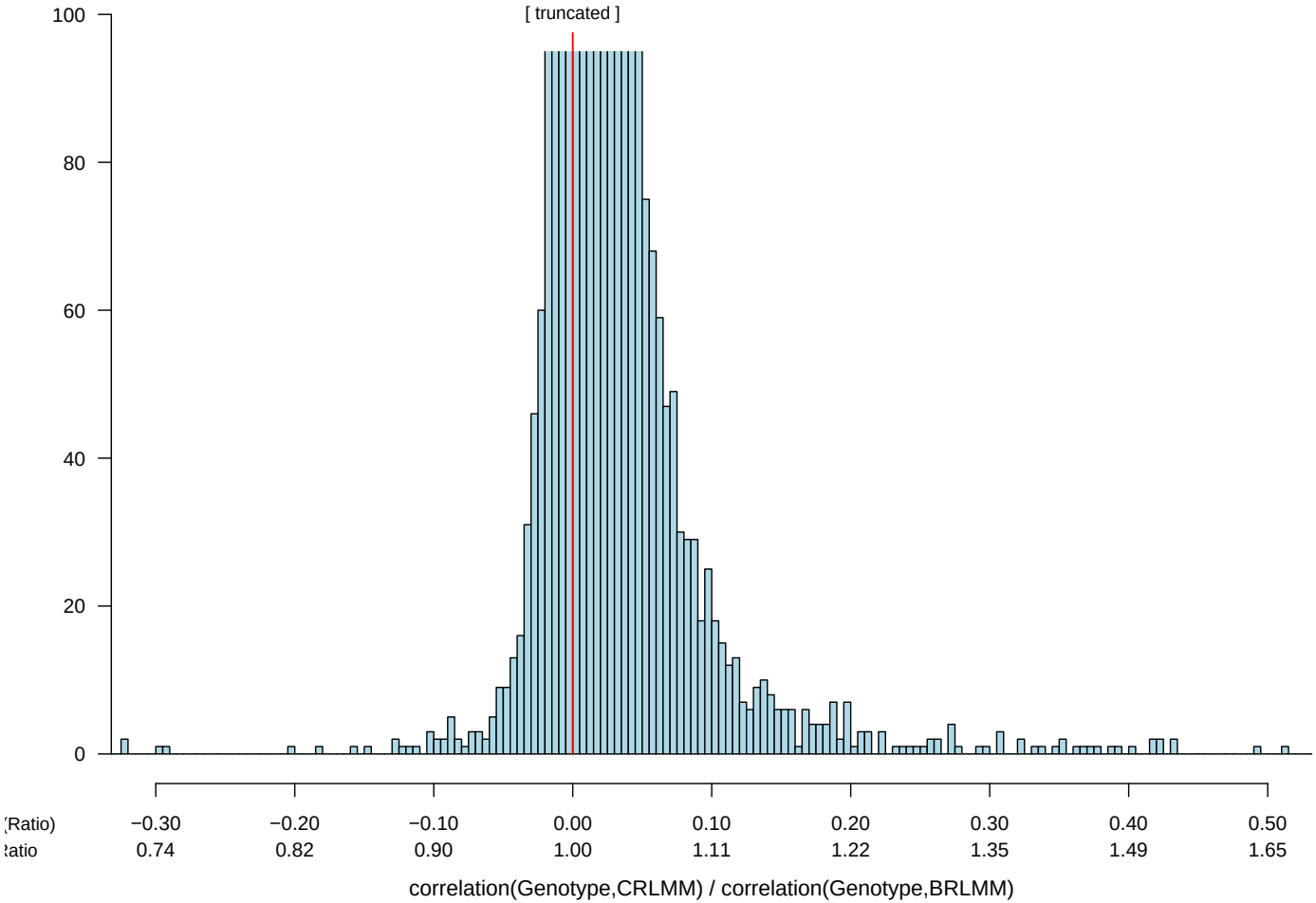
$\longrightarrow$ We develop methods that incorporate knowledge about genotype uncertainty, for example scoring functions for test of associations that take the genotype likelihoods derived from CRLMM into account.

Incorporating Genotype Uncertainty

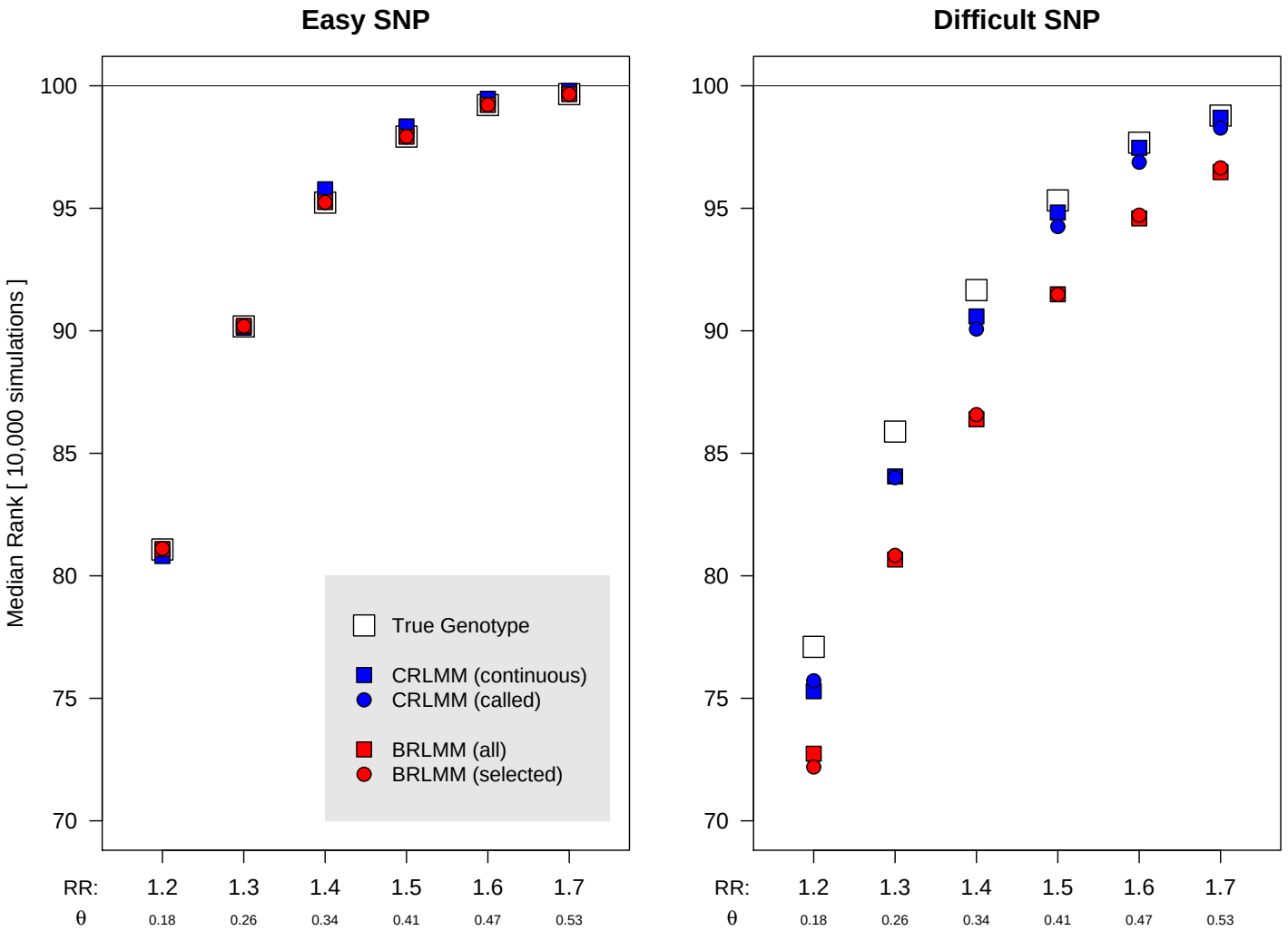
→ The confidence in genotype calls can differ substantially between SNPs!



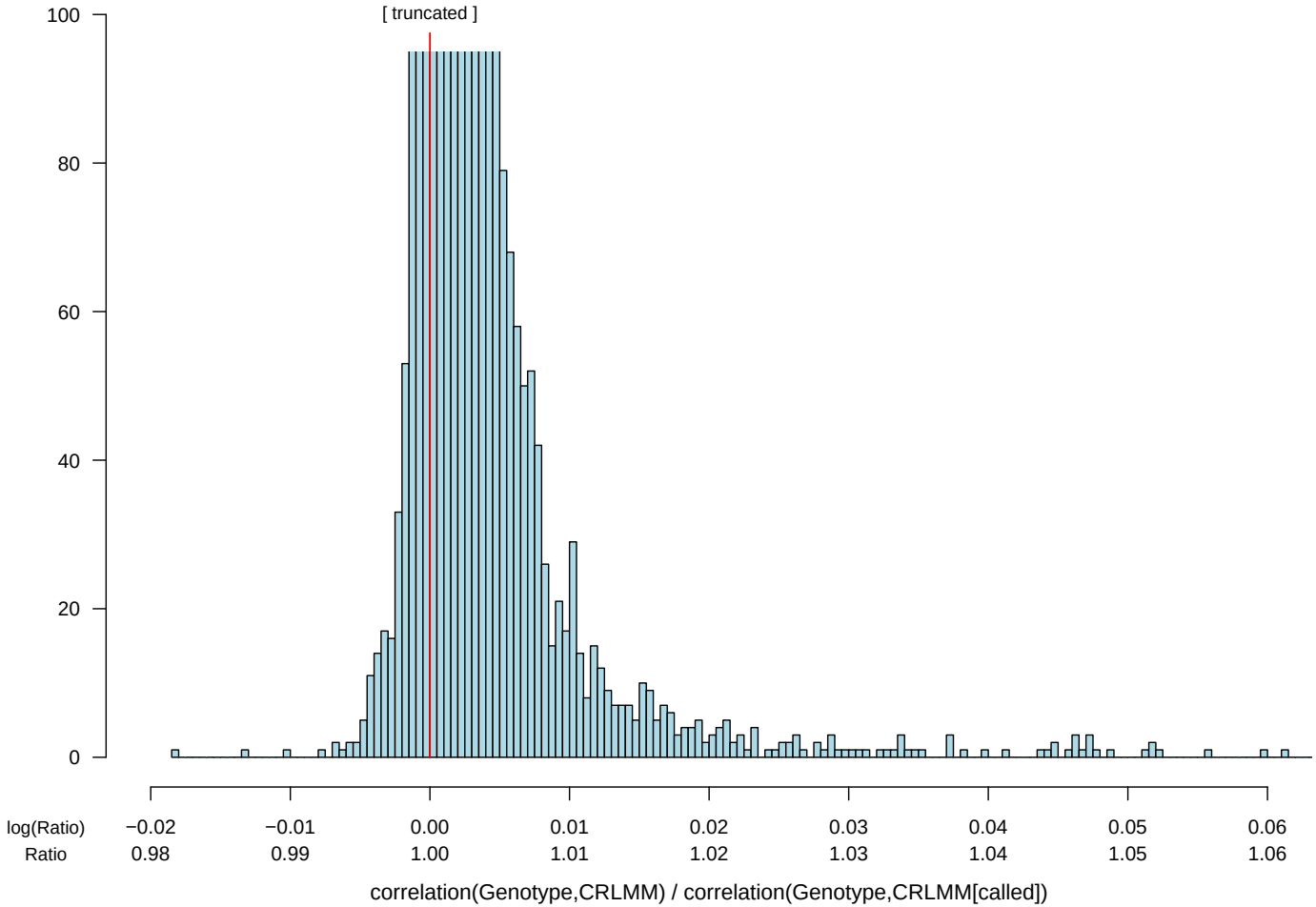
Incorporating Genotype Uncertainty



Incorporating Genotype Uncertainty



Incorporating Genotype Uncertainty



Prioritizing Genes in GWAs

- Based on ranking and selection approaches rather than on testing the null hypothesis of no GxE interaction.
- Requiring the specification and estimation of biologically relevant, SNP specific parameter (effect size) and specification of a joint prior distribution of these parameters.
- Ranking and selecting SNPs by combining estimates of effect sizes and their uncertainties.
- Incorporating knowledge obtained through linkage, fine mapping and candidate gene studies of GxE interaction.

Acknowledgments

- Ciprian Crainiceanu
- Qing Li
- Terri Beaty, Dani Fallin
- Rafa Irizarry, Benilton Carvalho
- Marvin Newhouse, Jiong Yang