

Poster presentation

Open Access

Evaluation of pooled allelotyping versus individual genotyping for genome-wide association analysis of complex disease

Siddharth Pratap*¹, Scott M Williams² and Shawn E Levy³

Address: ¹Dept. of Microbial Pathogenesis and Immune Response, Meharry Medical College, Nashville, TN 37208, USA, ²Dept. of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN 37232, USA and ³Dept. of Biomedical Informatics, Vanderbilt University, Nashville, TN 37232, USA

Email: Siddharth Pratap* - spratap@mmc.edu

* Corresponding author

from UT-ORNL-KBRIN Bioinformatics Summit 2008
Cadiz, KY, USA. 28–30 March 2008

Published: 8 July 2008

BMC Bioinformatics 2008, 9(Suppl 7):P11 doi:10.1186/1471-2105-9-S7-P11

This abstract is available from: <http://www.biomedcentral.com/1471-2105/9/S7/P11>

© 2008 Pratap et al; licensee BioMed Central Ltd.

Background

Recent advances in genotyping techniques and genomic knowledge via the Hap Map and Human Genome projects allow for true Genome-Wide Association (GWA) analysis for common complex diseases such as heart disease, diabetes, and Alzheimer's. A major obstacle in GWA analysis is the prohibitively high cost of genotyping the possibly thousands of individuals necessary to achieve statistical significance of results. One potential solution is to pool the DNA of case and control populations and to determine the genotype allele frequency differences in these populations by pooled allelotyping. While pooling can dramatically save time and money, it also adds sources of error. Our work has created a system process that allows for direct evaluation and comparison of pooled allelotyping to individual genotyping for GWA association analysis of complex disease.

Materials and methods

Complex disease penetrance functions were calculated for a 3 locus bi-allelic model with additive or multiplicative allelic spectrums using GenomeSIM software [1]. Penetrance probabilities were calculated for genotypes having from 0 to 6 disease-associated alleles. All probability functions used a base penetrance probability of 10% disease risk to account for environmental influences on disease risk. A total of 25,000 individual genotype files were created, each comprised of 10,000 SNPs with 3 disease-associated loci imbedded within. Custom MATLAB scripts

were used to make *in silico* "pseudo-pools" for pooled allelotyping from the individual genotype files. HAPLOVIEW software was used to conduct individual genotyping association analysis [2]. A modified version of the Pooled DNA Analyzer (PDA) program was used for pooled association analysis [3].

Conclusion

Power analysis was conducted for individual genotyping and pooled allelotyping with allele frequency estimation error from levels from 1% to 5% (see Figure 1). Our results show that pooling errors have a very large effect on the overall statistical significance of a pooled GWA study. Even a pooling error of 1% shifted the minimum resolvable relative risk (RR) with 80% power from (1.33–1.5) in individual genotyping to (1.5–1.67) in pooling. Pooling with 2% error had a minimum resolvable RR of (1.67–1.83). Pooling with 3% error resolved at RR (2.0–2.33). Further, pooling with 4% and 5% error of was not able to achieve 80% power at any of the levels of relative risk tested. Thus, pooled GWA studies may be limited to resolving complex disease associated variants with medium to high relative risks ratios.

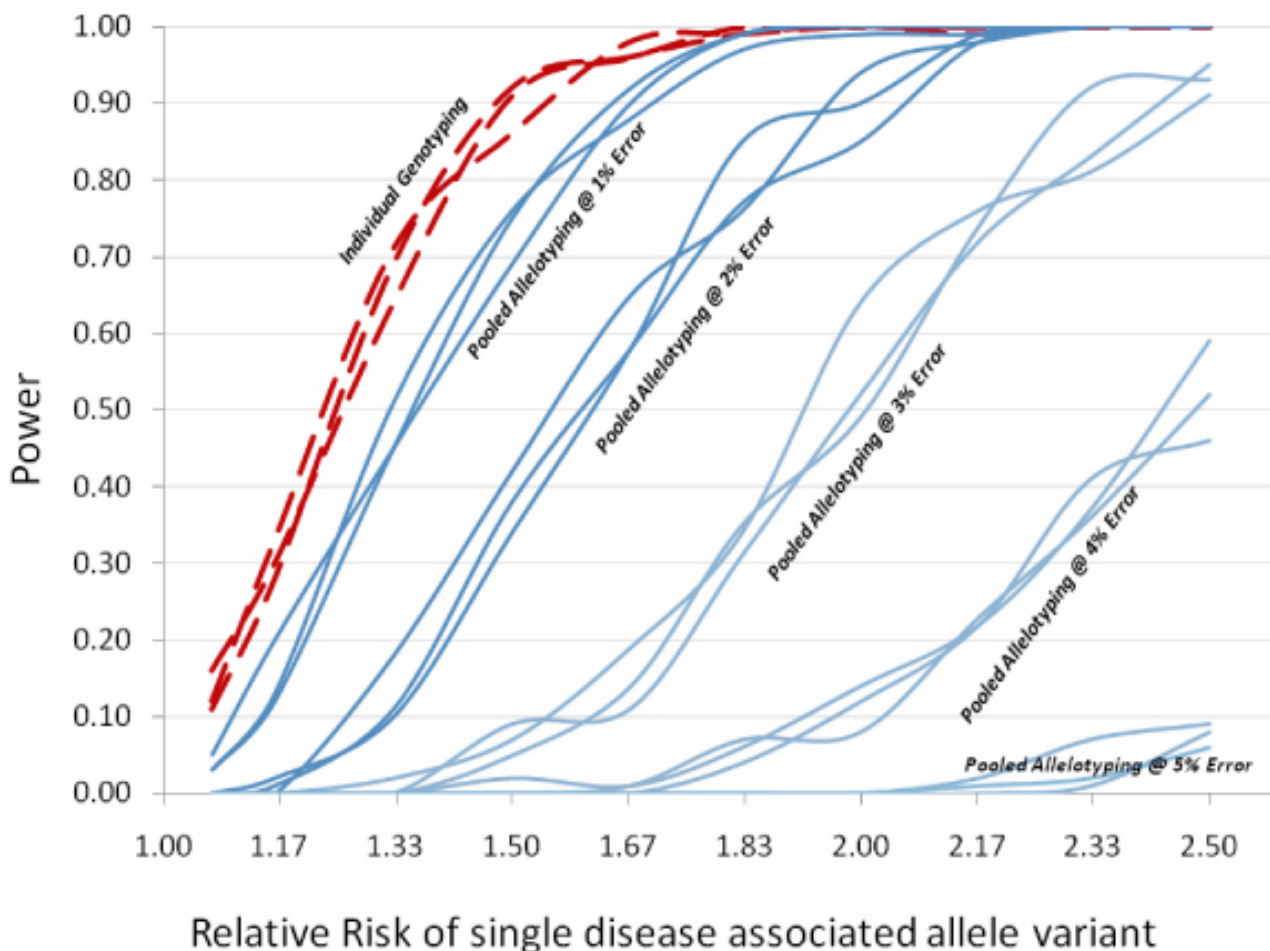


Figure 1
Association analysis power curve for individual genotyping versus pooled allelotyping. Individual genotyping: red dashed lines. Pooled allelotyping with 1% to 5% allele frequency estimation error: blue solid lines. Power is defined as the percent of simulations where disease associated loci had association test p-values more significant than ($p\text{-value} < 0.05$). Results are the average of 100 simulations at each of the relative risk levels listed on the X-axis.

Acknowledgements

S.P. supported by the National Library of Medicine, National Institutes of Health Grant (T15 007450-03) at the Vanderbilt University Department of Biomedical Informatics.

References

1. Dudek S, Motsinger AA, Velez DR, Williams SM, Ritchie MD: **Data simulation software for whole-genome association and other studies in human genetics.** *Pacific Symposium on Biocomputing* 2006, **11**:499-510.
2. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**(2):263-265.
3. Yang HC, Pan CC, Lin CY, Fann CS: **PDA: pooled DNA analyzer.** *BMC Bioinformatics* 2006, **7**(1):233.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."
 Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

