

### **POSTER PRESENTATION**

**Open Access** 

# A new set-valued system identification approach to identifying rare genetic variants for ordered categorical phenotype

Wenjian Bi<sup>1†</sup>, Guolian Kang<sup>2\*†</sup>, Yuehua Cui<sup>3</sup>, Yun Li<sup>4</sup>, Christine M Hartford<sup>5</sup>, Wing Leung<sup>5,6</sup>, Ji-Feng Zhang<sup>1</sup>

From UT-KBRIN Bioinformatics Summit 2014 Cadiz, KY, USA. 11-13 April 2014

### **Background**

For phenotype-genotype association studies that involve a phenotype with ordered multiple response categories, we usually either regroup multiple categories of the phenotype into two categories of "cases" and "controls" and then apply the standard logistic regression (LG) model [1,2], or apply a non-parametric method of Spearman rank correlation [3] or parametric method of ordered logistic (orderLG) regression model [4] which accounts for the ordinal nature of the phenotype. However, these approaches may lose statistical power if the phenotype is obtained by categorizing an observed or complicated unmeasured or immeasurable continuous phenotype or if the underlying genetic variants are rare.

#### Materials and methods

Therefore, we propose a set-valued (SV) system method, which assumes that the underlying continuous phenotype follows a normal distribution, to identify genetic variants associated with an ordinal categorical phenotype. We couple this model with a set-valued system identification method to identify all underlying key system parameters.

#### Results

Simulation studies show that SV well controlled the Type I error rate. In the comparison among LG, SV and orderLG methods, LG had significantly lower power than both SV and orderLG due to the disregard of the ordinal nature of the phenotype, and SV had similar or higher power than orderLG. Additionally, the SV association

parameter estimate was 2.7-28.7 fold less variable than the orderLG association parameter estimate. Less variability in the association parameter estimate translates to greater power and robustness across the spectrum of minor allele frequencies. These advantages are most pronounced for rare variants or even common variants when sample size is small. For instance, in a simulation with data generated from an additive orderedLG model with an odds ratio of 7.4 for a phenotype with three categories, a single nucleotide polymorphism with minor allele frequency of 0.75% and sample size of 999 (333 per category), the power of SV, orderLG and LG models were 70%, 40% and <1%, respectively, at a significance level of 10<sup>-6</sup>. When applied to a real data set, the set of variants identified by LG and orderLG was a subset of those identified by SV. Thus, SV can be a competitive alternative to LG or orderLG in genetic association studies such as candidate gene, genome-wide association studies or next generation sequencing studies, for ordered categorical phenotype.

#### Authors' details

<sup>1</sup>Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PRC. <sup>2</sup>Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. <sup>3</sup>Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA. <sup>4</sup>Department of Biostatistics, University of North Carolina, 3101 McGavran-Greenberg Hall, Chapel Hill, NC 27599, USA. <sup>5</sup>Department of Bone Marrow Transplantation and Cellular Therapy, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. <sup>6</sup>Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN, 38163, USA.

Published: 29 September 2014

#### References

 Treviño LR, Shimasaki N, Yang W, Panetta JC, Cheng C, Pei D, Chan D, Sparreboom A, Giacomini KM, Pui CH, Evans WE, Relling MV: Germline genetic variation in an organic anion transporter polypeptide associated

Full list of author information is available at the end of the article



<sup>\*</sup> Correspondence: Guolian.Kang@stjude.org

<sup>†</sup> Contributed equally

<sup>&</sup>lt;sup>2</sup>Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105 USA

- with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol* 2009, **27(35)**:5972-5978.
- Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, Kubo M, Jenkins GD, Batzler A, Shepherd L, Pater J, Wang L, Ellis MJ, Stearns V, Rohrer DC, Goetz MP, Pritchard KI, Flockhart DA, Nakamura Y, Weinshilboum RM: Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. J Clin Oncol 2010, 28(31):4674-4682.
- Png E, Thalamuthu A, Ong RT, Snippe H, Boland GJ, Seielstad M: A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region. Hum Mol Genet 2011, 20(19):3893-3898.
- Yang JJ, Cheng C, Yang W, Pei D, Cao X, Fan Y, Pounds S, Treviño LR, French D, Campana D, Downing JR, Evans WE, Pui C, Devidas M, Bowman WP, Camitta BM, Willman C, Davies SM, Borowitz MJ, Carroll WL, Hunger SP, Relling MV: Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. JAMA 2009, 301(4):393-403.

#### doi:10.1186/1471-2105-15-S10-P29

Cite this article as: Bi et al.: A new set-valued system identification approach to identifying rare genetic variants for ordered categorical phenotype. BMC Bioinformatics 2014 15(Suppl 10):P29.

## Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit

