Constructing Association Network in Two Stage Analysis: An Application of Backward Genotype-Trait Association (BGTA) Algorithm to NARAC Data

Yuejing Ding1*, Tian Zheng1, Iuliana Ionita-Laza2, Lei Cong1 and Shaw-Hwa Lo1

1Department of Statistics, Columbia University, New York, NY 10027; 2Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115.

Objective
- Identify RA susceptibility genes;
- Explore gene-gene interactions;
- Construct association network among different loci;
- Data: NARAC SNP scans data for Rheumatoid Arthritis (RA) analyzed in Amos et al. (2006) (GAW 15 problem 2);
- Family data were divided into “case” and “control” groups (349 “controls” and 474 “cases”).

2. Two-stage SNP Selection

1. BGTA Algorithm
- Markers: M1, M2, ..., M5407
- Random select a subset
- Compute GTD, GTA scores
- Greedy screening
- Repeat B times
- Final returned marker sets
- Exceptional high GTD score or exceptional high return frequency

2. Important Susceptibility Loci with Extremely High Joint GTD Score

Conclusions
- We found 39 loci which are strongly associated with RA disease, about 2/3 of which were previously found in the literature, including the well known HLA region, PADI4, CTLA4, SLC22A4, RUNX1. 1/3 of loci we found are new discovery.
- Besides susceptibility loci, 5 association sub-network, one central and four satellite, are constructed and graphed.
- It is seen that some of the loci work as a “hub”, such as 1p36.22, 2q33.3, 5p15.33, 8q24.23, 14q11.2, 14q12, and 18p11.21, and they all interacting with some other regions.

Acknowledgment
The research was supported by NIH grant R01 GM070789.


* Correspondences: yding@stat.columbia.edu