Identification of asthma-related *trans*-acting epistatic eQTLs using Model-Based Multifactor Dimensionality Reduction (MB-MDR)

Bessonov K[1][2], Croteau-Chonka D[3], Qi W[3], Carey VJ[4], Raby BA[3], Van Steen K [1][2]
[1] Systems and Modeling Unit, Montefiore Institute, University of Liege, Liege, Belgium; **[2]** Bioinformatics and Modeling, GIGA-R, University of Liege, Liege, Belgium
[3] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA;
[4] Center for Biostatistics in AIDS Research; Harvard School of Public Health, Boston ,MA,USA

Introduction

Epistasis is likely to underlie most complex traits, including gene expression, yet it is very difficult to detect using standard approaches. SNPs located inside a gene coding region or in its vicinity (i.e. ≤2 Mb from each 5' and 3' side) can influence the corresponding gene expression levels. These expression quantitative trait loci (eQTLs) are referred to as *cis*SNPs. In contrast, eQTLs that are outside the aforementioned gene range can also influence the gene's expression, in which case, they are called *trans*SNPS to that gene. In this study we considered significant *cis*SNPs previously identified via generalized least squares (GLS) regression modeling. We then identified those genes transcripts whose expression is regulated by *cis/trans*SNP interaction. In this work we aimed at identifying transcripts whose expression is regulated by a *cis/trans*SNP interactions using Model-Based Multifactor Dimensionality Reduction (MB-MDR) [2]. This model-free approach to detect ***trans*-epistasis** involves reducing a high-dimensional *G*x*G* space to *G*x*G* factor levels that either exhibit high evidence, low evidence or no evidence at all for their association to gene expression levels of interest.

Our protocol was applied on real-life data from the childhood asthma management program (CAMP) [1]. It involved coupling a traditional *a priori* eQTL search to an *a posteriori* *trans*-epistasis analysis to identify genetic modifiers to statistically significant *cis*SNPs. Such an approach allows to reveal previously unreported inter-dependencies that may be important in understanding of biological mechanisms underlying human complex diseases such as asthma.

The proposed protocol identified a large number *trans*-epistasis gene-gene effects of eQTLs.

Methods

Our CAMP based dataset [1] consisted of 153 caucassian non-smoker subjects with ages ranging between 16-25. Each selected subject was represented by 19,451 gene expression values and 507,945 SNPs. Preliminary eQTL analysis was done within GLS regression framework. The FDR was controlled at 0.05, generating a final list of 1844 significant *cisSNPs*.

Data preparation for MB-MDR interaction analysis

Genotype and expression sample IDs were matched. The gene coding regions were identified using GenomicFeatures library in R. Two Mb ($2x10^6$ bp) windows in either direction were added delimiting the effective range for *cis*SNPs on gene expression. The SNPs outside this range where defined as *trans*SNPs. For MB-MDR runs the gene expression values for each eQTL were rank transformed to normality using GenABEL library functions in R and analyzed consecutively using default options in quantitative trait MB-MDR analysis. MB-MDR was modified to work under multi-trait context (trait = gene expression value). The assessment of significant *cis/trans*SNP interactions was based on marginal permutation-derived null distributions (999 data replicates). All *cis/trans*SNP pairs with a p-value p≤0.001 (not corrected for multiple-testing) where retained for further analysis

Determination of trans-epistasis threshold

The "*trans*-epistasis effect" on a particular *cis*SNP was defined as the ratio of significant modifier effects to the *cis*SNP under consideration (*cis/trans*SNP pairs with MB-MDR marginal p≤0.001) out of a total number of 507,945 possible pairings with the *cis*SNP. We randomly sampled on average 200 MB-MDR marginal p-values, for each of 1844 eQTLs, and repeated this process 100 times. For each replicate we then computed the 1844 *trans*-epistasis effects. The distribution of *trans*-epistasis effects appeared to be consistent across replicates. The null distribution of the 1st replicate is shown in **Fig . 2B**. From these distributions we derived a critical value for the *trans*-espistasis effect. It was fixed at $0.00124 \pm 2.4x10^{-8}$ using α significance level of 0.05 and one-sided alternative hypothesis. A total of 1,763 out of 1,844 *cis*SNPs remained after this *trans*-epistasis effect filtering.

KEGG pathway enrichment and annotation

The KEGG pathway database, containing gene mappings to 186 known biological pathways, was used to annotate the genes corresponding to the 1,763 eQTLs obtained from the previous step. Notably, some genes did not map to any pathway or were mapped to multiple pathways (**Fig.3**).

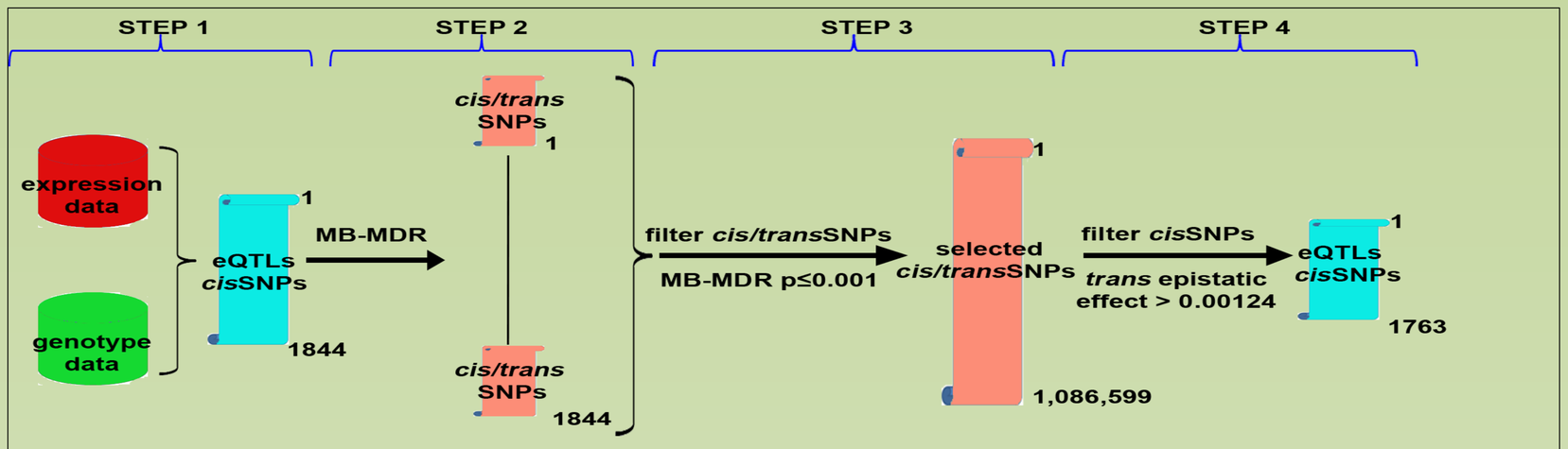

Fig 1: Outline of *trans*-epistatis eQTL analysis method. **Step 1:** Identify significant eQTLs via GLS; **Step 2**: Identify *trans*SNPs interacting with significant *cis*SNPs from step 1. **Step 3:** Filter *cis/trans* SNP pairs with MB-MDR marginal p≤ 0.001; **Step 4:** filter eQTLs with *trans*-epistasis effect > 0.00124 (based on 100 random samples) and annotate eQTLs corresponding genes to KEGG pathways

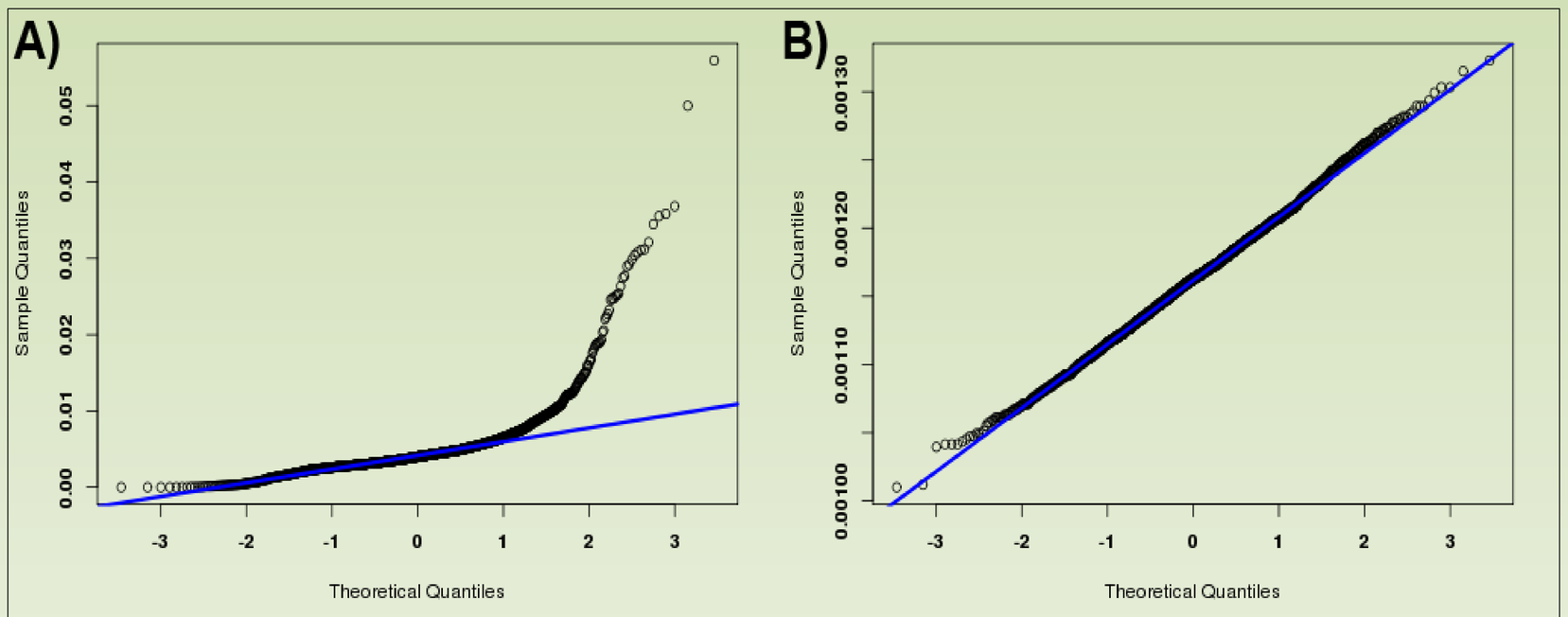

Fig 2: Q-Q plots assessing 1844 *trans*-epistasis effects based on **A)** CAMP and **B)** randomly permuted data against normal distribution. The **blue line** passes through 1st and 3rd quartiles (0.25,0.75). They y-axis refers to *trans*-epistasis effect with theoretical 0 to 1 range

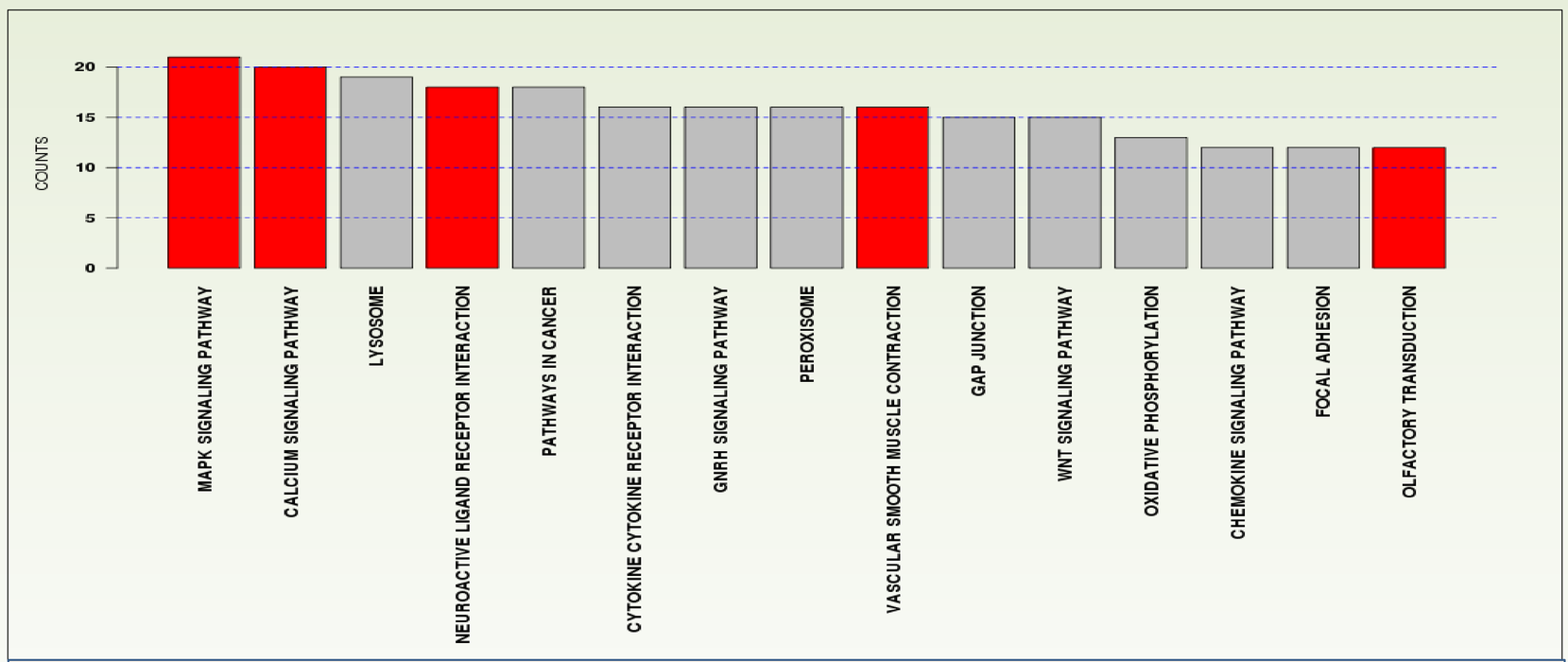

Fig 3: KEGG pathways enrichment based on the list of genes linked to eQTLs with significant *trans*-epistasis effect (1763 genes). The top 15 KEGG pathways with the highest number of occurrences within this list are shown. Asthma-related pathways are indicated in red.

Results and Conclusions

- MB-MDR can be successfully used on the genome-wide scale in eQTL studies for *trans*-epistasis assessement
- *Trans*-epistasis effect expressed as a ratio is able to characterize genome-wide impacts of *cis*SNPs
- The maximum *trans*-epistasis effect was 0.056 (compared to a mean of 0.00116 in "null data")
- More than 95% (1763/1844) of the significant *cis*SNPs in our data were significantly modified by *trans*SNPs
- The distribution of *trans*-epistasis values significantly deviated from normality (**Fig.2-A**)
- Low association was found between **Fig.1-Step 1** (GLS based) and **Fig.1-Step 4** (based on *trans*-epistasis effects) rankings of significant *cis*SNPs (Pearson r^2= 0.00046)
- Annotation to KEGG pathways of genes associated with doubly selected eQTLs (GLS and MB-MDR) suggest that asthma CAMP cohort patients may have aberrations associated to vascular smooth muscle contraction, intra-celular signaling and protein degradation pathways (**Fig.3**)

Contact: Kyrylo Bessonov - **kbessonov@ulg.ac.be**

References

[1] Childhood Asthma Management Program Research Group. **The Childhood Asthma Management Program (CAMP): design, rationale, and methods**. Control Clin Trials. 1999 Feb;20(1): 91-120.
[2] Cattaert T, Calle ML, Dudek SM, Mahachie John JM, Van Lishout F, Urrea V, Ritchie MD, Van Steen K: **"Model-based multifactor dimensionality reduction for detecting epistasis in case-control data in the presence of noise."** Ann Hum Genet 2011, 75:78-89.
[3] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004). **"The KEGG resource for deciphering the genome".** Nucleic Acids Res 32, 2004 (Database issue): D277–80