A factor model to analyze heterogeneity in gene expression

Yuna Blum[1,2,3], Sandrine Lagarrigue[1,2] and David Causeur[3]

[1]Agrocampus Ouest, UMR598, Animal Genetics, 35000 Rennes, France
[2]INRA, UMR598, Animal Genetics, 35000 Rennes, France
[3]Agrocampus Ouest, Applied Mathematics Department, 35000 Rennes, France

Microarray technology allows the simultaneous analysis of thousands of genes within a single experiment. Classical approaches to analyze transcriptomic data ignore the gene dependence structure. This leads to correlation among test statistics which affects a strong control of the false discovery proportion.

We focus our study on a method called FAMT (Friguet *et al*, 2009) which captures the components of expression heterogeneity into factors. The relevance of factor modeling is first shown on illustrative gene expression data sets in simple situations of heterogeneity. We also use a real expression data set, primarily generated to map QTL for abdominal fatness in chickens (Le Mignon *et al*, 2009). FAMT provides functional information about a QTL region through a gene related to the fatness trait and controlled by this region (DHCR7) not observed by a classical approach. Then we interpret the independent factors extracted from this biological data set using known information about both experimental design and genes. We show that some factors may have different and complex origins, which can be related to particular metabolisms.

As we extract biological information from what was before simply considered as statistical noise, analyzing heterogeneity in gene expression yields a new point of view on transcriptomic data (Blum *et al*, 2009).

REFERENCES :

1. Friguet C, Kloareg M, Causeur D: A factor model approach to multiple testing under dependence, JASA, in press.
2. Le Mignon G, Desert C, Pitel F, Leroux S, Demeure O, Pitel F, Guernec G, Pitel F, Douaire M, P LR, Pitel F, Lagarrigue S: Using trancriptome profiling to refine QTL regions on chicken chromosome 5. BMC Genomics, in press.
3. Blum Y, Le Mignon G, Lagarrigue S, Causeur D: A factor model to analyze heterogeneity in gene expression. BMC Bioinformatics, submitted.