# Report1

This paper focuses on the signaling and bioenergetics pathways involved in the exit of T cells from the quiescent state. After obtaining the protein expression data, the main lines of analysis are divided into the following three steps.

The first step used one-way ANOVA to identify proteins that changed significantly between any two time points. One-way ANOVA focuses on the differences in the means of two or more aggregates due to the influence of a single factor. In that paper, the factor here refers to time. Expression differences can generally be ranked by comparing p-values. It is worth noting that here the two totals already obey a normal distribution by default. Strictly speaking, the researcher needs to make a test whether the protein expression distribution obeys normal distribution before doing a single anova. Considering that 8431 protein expression data are to be tested here, non-parametric tests such as rank sum test can also be used here. To a certain extent, the computational effort is reduced.

The second part focuses on unsupervised clustering of previously significantly different expression data using weighted gene correlation network analysis (WGCNA). The main idea of WGCNA is to aggregate genes into modules one by one and then calculate a value (eigengene) to represent these modules, which can then be linked to the features of the sample (by calculating the correlation between eigengene and features), so that we can filter the modules of interest and study the genes in them. The process is divided into the following four steps.

(1) Calculation of dissimilarity between every two genes by topological overlap

First, the network structure of the gene is assumed to be a scale free network which means that there is a power law distribution between the connectivity of the nodes and their frequencies. Next is the calculation of the adjacency matrix given by equation

.

(2) Clustering tree

Using the previously obtained adjacency matrix, the gene tree can be obtained by employing hierarchical clustering. Of course, other approaches can also be taken.

1. Tree shearing

By shearing the tree it is possible to divide the genes with consistent expression into the same modules.

1. Merge similar modules

The eigengene of each module is calculated and the similar modules are merged by clustering based on these eigengene. Module eigengene refers to the first principal component of gene expression in a gene module, and this value is used to represent the gene expression profile of the module.

The final step of the analysis is to put the previously calculated whole proteome clusters (WPC) into Protein-Protein Interaction Networks (PPI network) for research analysis. Several pathways of T-cell activation were identified.

In general, the steps and ideas of the study are very clear, but there are some minor problems. One of them is the previously mentioned problem of verifying the Gaussian distribution, and another exists in the second step of the computation regarding the scale-free network. The paper does not mention how the researchers determined that the network is a scale-free network, and the choice of beta value. This all needs further verification and clarification.

Translated with www.DeepL.com/Translator (free version)