Visualization of Proteomics Data Integrated with KEGG Metabolic Data Using R and Bioconductor

Ermir Qeli[1,*] , **Christian Panse**[2,*] , **Christian Ahrens**[1,*]

1. Center for Model Organism Proteomes (C-MOP)
2. Functional Genomics Center Zurich (FGCZ)

* Contact emails: ermir.qeli@molbio.uzh.ch,cp@fgcz.ethz.ch, christian.ahrens@molbio.uzh.ch

Keywords: Proteomics, KEGG, Metabolic Pathways, R, KEGGSOAP, Heatmaps.

Metabolic pathways represent series of chemical reactions occurring within a cell, where each reaction is catalyzed by enzymes. These enzymes are nothing else than proteins. Here we go after the question of how to integrate and visualize quantitative or qualitative proteomics data in the context of metabolic pathways

For illustrating the idea, we focus on the organism *A. thaliana* and we extracted all its possible pathways stored in KEGG (Kanehisa et al, 2000) using KEGGSOAP together with the annotated enzymes present in these pathways.

Furthermore, we extracted from PRIDE tissue specific proteomics data of *A. thaliana* (Bärenfaller et al, 2008) and compared them with the recently acquired pollen proteome (Grobei et al, 2009). From these data, protein spectral counts were calculated, which provide a rough estimation of protein abundance in the respective tissue. For each metabolic pathway, a $m \times n$ matrix with m genes(enzymes) and n tissues with the log values of the respective spectral counts for each gene model for the specific tissue was generated. This matrix is then visualized as a heatmap, where one can compare the enzymatic activity between the different tissues and see which enzymes are active in a specific tissue and inactive elsewhere.

Figure 1 shows only one such heatmap, namely the one related to biosynthesis of phenylalanine, tyrosine and tryptophan. According to literature, the activity in this pathway for pollen should be quite low. The color intensity for the highlighted column for pollen in Figure 1 indicates that this is really the case. Similar pictures are generated for all pathways of *A. thaliana* annotated in KEGG (more than 100). As the scripts written are generic in nature, they can be applied also for other organisms annotated in KEGG where proteomics data is available and are available upon request from the authors.

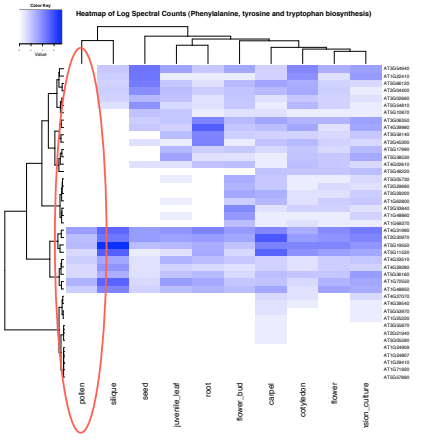

Figure 1: Heatmap showing the enzymatic activity for different *A. Thaliana* tissues in phenylalanine, tyrosine and tryptophan biosynthesis pathway.

References

Kanehisa M, Goto S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/.

Bärenfaller et al (2008). Genome-Scale Proteomics Reveals Arabidopsis thaliana Gene Models and Proteome Dynamics. *Science*, 320(5878) , 938-41 .

Grobei et al (2009). Deterministic protein inference for shotgun proteomics data of the *Arabidopsis thaliana* mature pollen provides new insights into pollen development and function. *in revision*.

Zhang J, Gentleman R. (2009). Client-side SOAP access KEGG, http://www.bioconductor.org/packages/2.3/bioc/html/KEGGSOAP.html/.