TRANSPATH® Report 2, 0001 (2004)

Organ/tissue expression patterns based on UniGene EST frequency calculations and Cytomer®

Alexander Kel, Ellen Fricke, TRANSPATH_Team

BIOBASE GmbH
Halchtersche Strasse 33
D-38304 Wolfenbuettel, Germany

We made an estimation of the gene expression levels in different tissues based on the UniGene database (release of March, 2003). The following steps were performed:

1. For all UniGene clusters (***G*** - is the total number of clusters for human genes), we collected all of the information about libraries that was used for sequencing the ESTs included in the clusters.

2. From the list of libraries we excluded all cancer-related libraries. Only the libraries of normal tissues and organs were considered.

3. All tissue and organ names used in the selected list of libraries were linked to corresponding terms in the Cytomer® database [1]. All libraries linked to the same term were grouped into "organ library groups" (e.g. liver library group). ***P***- is the total number of formed organ library groups.

4. Compute n_{ij} which is the number of ESTs of a gene ***i*** (Unigene cluster) linked to a organ library group ***j*** .

5. Compute $a_{ij} = \frac{n_{ij}}{\sum_{i=1,G} n_{ij}}$ - an "abundancy score" of each gene in each group.

6. Compute $\bar{a}_{i\bullet} = \frac{\sum_{j=1,P} a_{ij}}{P}$ - an average "abundancy score" of gene ***i***. And

$sigma(\bar{a}_{i\bullet}) = \sqrt{\frac{\sum_{j=1,P}(a_{ij} - \bar{a}_{i\bullet})^2}{P(P-1)}}$ - standard error.

7. Compute $s_{ij} = \frac{a_{ij}}{\sum_{j=1,P} a_{ij}}$ - a "specificity score" of each gene in each group.

8. Compute $\bar{s}_{\bullet j} = \frac{\sum_{i=1,G} s_{ij}}{G}$ - an average "specificity score" of group ***j***. And

$sigma(\bar{s}_{\bullet j}) = \sqrt{\frac{\sum_{i=1,G}(s_{ij} - \bar{s}_{\bullet j})^2}{G(G-1)}}$ - standard error.

9. Compute $da_{ij} = \frac{a_{ij} - \bar{a}_{i\bullet}}{sigma(\bar{a}_{i\bullet})}$ - an "abundancy parameter" which shows a difference of the abundancy score from an average abundancy score. In the case of $\boldsymbol{da}_{ij} \geq 2.0$, we can consider this difference as statistically significant.

10. Compute $ds_{ij} = \frac{s_{ij} - \bar{s}_{\bullet j}}{sigma(\bar{s}_{\bullet j})}$ - a "specificity parameter" which shows a difference of the specificity score from an average specificity score. In the case of $\boldsymbol{ds}_{ij} \geq 2.0$, we can consider this difference as statistically significant.

11. We used the following default cut-offs to establish links between genes ***i*** and tissue/organs ***j*** and to form groups of genes expressed in a given tissue/organ: $\boldsymbol{da}_{ij} \geq 2.0$ AND $\boldsymbol{ds}_{ij} \geq 2.0$ AND $\boldsymbol{n}_{ij} \geq 2$.

[1] PMID: 15089753
Wingender, E.
TRANSFAC®, TRANSPATH® and CYTOMER® as starting points for an ontology of regulatory networks
In Silico Biol. 4, 55-61 (2004).