

ESTissue: A novel method to identify gene expression profile EST based

Iacono M (1), Mignone F (1,2), Anselmo A (1), Pesole G (3)

(1) Dipartimento di Scienze Biomolecolari e Biotecnologie - Università degli Studi di Milano

(2) Dipartimento di Chimica Strutturale e Stereochimica Inorganica - Università degli Studi di Milano

(3) Dipartimento di Biochimica e Biologia Molecolare - Università di Bari

Motivation

Genes are differentially expressed in different tissues, in different developmental stages or in pathological conditions. Differences in expression levels are often caused by diverse regulatory control mechanisms operating at the transcriptional, post-transcriptional and post-translational levels. The possibility of identifying clusters of genes that share the same expression profile (in terms of tissue specificity or temporal patterns of expression) could be of great significance in the investigation of underlying regulatory mechanisms. The availability of data generated by highthroughput approaches such as Expressed Sequence Tags (ESTs) allows the *in silico* analysis of gene expression in different tissues and physiological conditions. Several methodologies have been developed to characterize gene expression profiles from EST data. However, all such approaches have their own particular limitations, including both methodological and data access issues.

Methods

Public databases currently accommodate around 7.6 million human ESTs, of these, approximately 6.3 million of these are clustered in the Unigene database. However, Unigene clusters are assembled without consideration of the origin (tissue, developmental stage etc) of individual sequences. The presence of an EST in a tissue-specific library implies that the given gene or isoform is expressed in that specific tissue (or condition), moreover, it is reasonable to assume that the level of expression of a gene should be correlated with the number of ESTs. We use the information available for each EST to compare the expression pattern of genes in different tissues and we are thus able to identify both tissue specific genes and housekeeping (ubiquitously expressed) genes.

Results

We have developed a system (available through a web interface) which allows the user to identify and download sets of coexpressed genes. Selection can be performed on tissue specificity, developmental stage or pathological state. An automated evaluation of expression profiles is also performed on user submitted genes.

Contact email: <mailto:graziano.pesole@unimi.it>