# The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach

Stefania Bortoluzzi1, Fabio d'Alessi1, Chiara Romualdi2 and Gian Antonio Danieli1.

Department of Biology1 and Department of Statistics2, University of Padua (danieli@bio.unipd.it)

In human genomics, the high throughput analysis of general databases promises to provide relevant information and to produce novel biolgical knowledge. This strategy might be applied as well to the reconstruction of transcriptional profiles of different tissues or of the same tissue in different developmental stages (Bortoluzzi and Danieli, 1998).
We developed a novel software tool, in order to mine the UniGene database, to retrieve extended datasets and to merge them. By applying this tool, information on 4,080 UniGene clusters were retrieved, belonging to three adult human skeletal muscle cDNA libraries, selected for being unnormalised or unsubtracted. The software processed the resulting records, which were sorted out according to specific criteria. In particular, for the present work, the field 'ESTs SEQUENCES' was considered. For each entry, the number of ESTs obtained from skeletal muscle cDNA libraries was annotated. If additional ESTs, obtained from different tissues, were reported in the record, their presence was also automatically annotated by the program, to produce information on the expression of the corresponding gene in different human tissues.
The level of expression of each gene was estimated as a percentage of the total transcriptional activity, by computing the number of the skeletal muscle ESTs corresponding to a given entry, over the total number of skeletal muscle ESTs reported for all the entries included in the catalogue. The basic assumption is that the number of detected ESTs per gene is a function of the transcript frequency in the population of mRNAs. The expressed genes were classified according to their different levels of expression. Only 10% of genes expressed in muscle resulted transcribed at high level. They contribute to more than 50% of the total transcriptional activity of the tissue.
The large majority of genes expressed in the adult skeletal muscle appear to be active also in several different tissues. Most skeletal muscle genes were found in at least one additional tissue and about 89% of them in more than 4 additional tissues. Forty-seven entries (1.2% of the total) were found only in cDNA libraries obtained from human skeletal muscle.
The validation of this "in silico" approach was attempted by a comparison with SAGE data on genes expressed in skeletal muscle, reported in the Rochester SAGE catalogue (Welle et al, 1999), release July 1999. We considered the expression data concerning the 295 tags corresponding to fully annotated genes highly expressed in skeletal muscle. The reciprocal correspondence between 120 genes belonging to both catalogues was established, by using the GenBank ID and the levels of expression of each gene were compared. The results obtained by the two methods showed a statistically significant concordance (Bortoluzzi et al., 2000).
Transcriptional profiles of the adult human skeletal muscle and of the adult human retina, produced in our laboratory, are published at the dedicated web site GETProfiles (http://telethon.bio.unipd.it/GETProfiles/index.html).

REFERENCES

Bortoluzzi, S. and G.A. Danieli. 1999. Towards an in silico analysis of transcription patterns. Trends Genet 15: 118-119.
Welle, S., K. Bhatt, and C.A. Thornton. 1999. Inventory of high-abundance mRNAs in skeletal muscle of normal men. Genome Res 9: 506-513.
Bortoluzzi S., F. díAlessi, C. Romualdi and G.A. Danieli 2000. The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach. (Genome Res 10, in press)