Gene Expression database as a tool to identify potentially interesting Oncological Targets

ID - 111

Bosotti Roberta[1], Ukmar Giorgio[1], Calogero Raffaele[2], Isacchi Antonella[1]

[1]Genomics Unit, Biotechnology Department, Nerviano Medical Sciences S.r.l.,Nerviano (MI)

[2]Genomics and Bioinformatics Unit, Dipartimento di Scienze Cliniche e Biologiche c/o Az. Ospedaliera S. Luigi, Orbassano (TO)

Motivation

Data on gene differential expression between normal and tumor samples can be used to get insight about putative physiological gene function, identify targets overexpressed in tumor vs normal cells or make hypothesis about possible side effects of target inhibition.Unfortunately systematic information at protein level is usually not available. Protein Atlas (www.proteinatlas.org) is a very relevant source of information but so far it covers only ~1300 proteins.Gene expression levels, analysed by Microarray (MA) technology, can partially compensate for the gap, giving information at mRNA level but for the whole genome.

Thanks to the effort of public consortiums, thousands of MA experiments have been collected in dedicated databases (ArrayExpress, GEO,), although they are not organized in a way easy accessible for data mining and comparison.An attempt to create a user-friendly interface for expression data in normal tissues as well as cell lines has been done by Novartis (Novartis Gene Expression Atlas), with some limitations. Gene Expression Atlas is a database collecting MA expression data performed on cell and tissue samples coming from various commercial sources (Clinomics Biosciences, AllCells, CloneticsBioWhittaker,) provided with an interface to search for gene expression on a gene by gene basis. It includes several cell lines and is easy to use but tissue source is not consistent and is poorly characterized, donor information is not available and only few tumor samples are included (mainly leukemia). Moreover samples are pooled and performed in duplicate.

Methods

Searching GEO database we noticed two interesting data collections: GSE3950 (~300 chips from 65 diverse normal tissue types) and GSE2109 (~1500 chips from ~1500 different patients with several tumor types). In both studies a particular attention was dedicated to standardize the procedure for data collection and storage. On all the samples microarrays were performed using Affimetrix technology (hgu133plus2 chip).

The objective of the present work was to create a database to store gene expression data together with available clinical information form the two dataset and provide an interface to query the database and comparing different conditions.For this purpose a database was created in Oracle and interfaced with Spotfire.

Results

The database allows the comparison of gene expression level across tissues, displaying histograms of normal and tumor samples. Expression data are directly linked to clinical information derived from patients (histology, grade, sex, relaps, treatment type,). Filtering and sorting options allow to focus on a specific tissue of interest and to make hypotesis about gene expression profile and histology (ie.

adenoma vs. carcinoma). Heatmaps and gene profiles are also possible to display.The database can be a starting point to identify potential new targets based on differential expression profile tumor vs normal

Email: roberta.bosotti@nervianoms.com