A new method to integrate gene expression, chromosome location and pathway enrichment in microarray gene expression experiments

Martini P[1,2], Cagnin S[1,2], Romualdi C[2], Lanfranchi G[1,2]

[1] CRIBI Biotechnology Centre, University of Padova, Padova, Italy
[2] Department of Biology, University of Padova, Padova, Italy

Motivation

In the last decade several new technologies allowed researchers to produce a large amount of data in a relative small time laps. For example, microarrays are able to quantify the expression of ten of thausands of genes in a single experiment, usually stored on dedicated databases. Thus, nowadays, we have to solve a challanging task: providing a full and comprehensive explanation of the biological information underneath lists of differentially expressed genes sharing expression profiles. Bioinformatics instruments able to analyse, categorize and explain this large amount of data will be of great help in this context.

Methods

Here we develop a new approach to the global expression data mining (STEPa) and to the global pathway analysis (SPA). These approaches are based on two assumptions:
1) genes are distributed in a non-random way across genome (cluster profile) and
2) non-random distribution affects the expression level of group of related genes.
STEPa (Simple Test of Expression Pattern) is a new algorithm performing chromosome mapping of gene expression values, taking into account the regional contest of gene. STEPa works using a combination of two sliding windows across each chromosome to detect regions with significant expression variations. SPA (Simple Pathway Analysis) is a new procedure that allows the compution of global expression values for specific groups of genes (more than 7000 groups built on: Gene Ontology (GO) annotation (6415), KEGG (172), ByoCyc (161), BioCarta (313), SuperArray (51) and WikiPathway (111)). Pathway score is based on expression values adjusted by chromosome expression profiles built with STEPa (Figure 1).

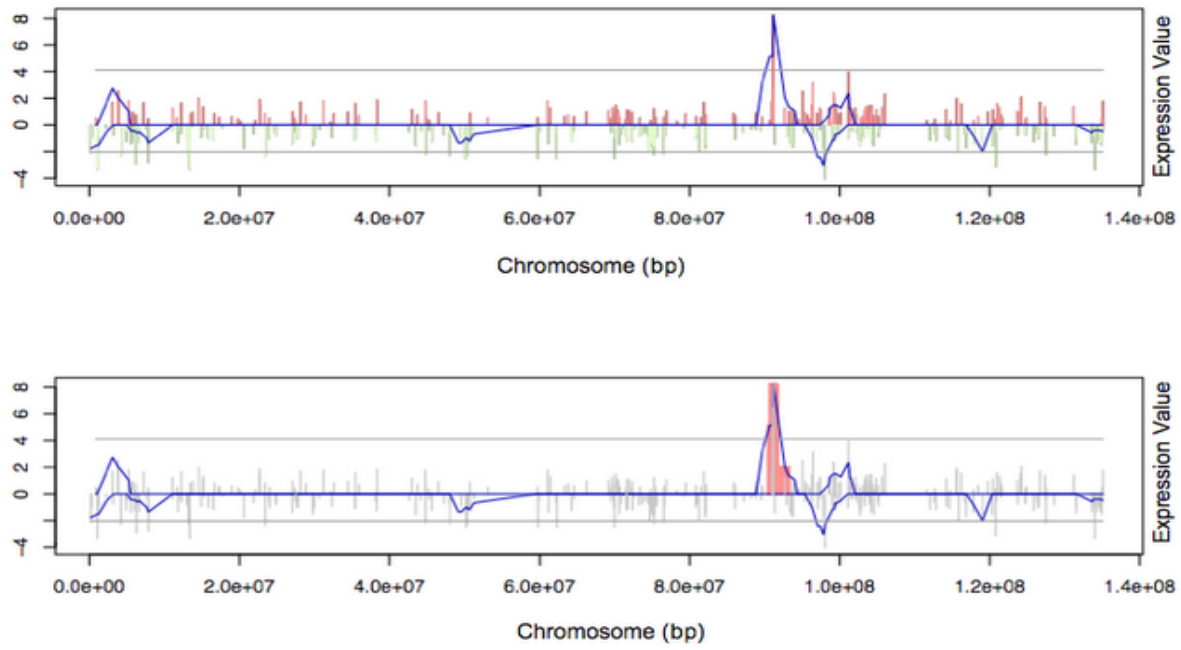

Figure1 Example of graphical result of STEPa analysis on chromosome 10. In the upper part of pictured are represented expression value (up-regulated genes in red, down-regulated genes in green) ordered by

chromosomal position and STEPa profile of chromosome 10 (blue line). Below is represented the profile of chromosome 10 highlighting significant varied region.

Results

Custom data set of inflammatory myopathies has been used as case study: we analyzed 8 Dermatomyositis (DM) and 32 Polymyositis (PM) expression data derived from home made human microarrays platform (GEO ID: GPL6647) based on two-color hybridization and, 12 healthy controls (Ctrl) , 9 Dermatomyositis (DM) and 21 Juvenile Dermatomyositis (JDM) derived from public Affymetrix datasets available at GEO database. Although other chromosome mapping procedures have been recently proposed, STEPa is the only one which allows the user to perform meta-analysis across data set because platform-independent. Using STEPa we built a profile for each chromosome, discriminating between different pathological states and highlighting interesting chromosome regions. Exploring these regions we identify a strong signature of interferon stimulated genes in DM, results that are consistent with published studies carried out by Greenberg et al 2005. Finally, STEPa is the first approach of chromosome mapping applied in pathology with no translocation. With SPA we reach substantially the same results, highlighting significant pathways for each pathological state, confirming
i) data produced by Greenberg et al 2005,
ii) interferon signature for DM and
iii) introducing new interesting pathways like proteasome degradation.
These two tools allow the integration between GO and functional pathways with altered chromosomal regions in order to construct an interaction network involved in pathology studied.

Contact : paolo.martini@unipd.it