Identification of new putative mitotic genes through coexpression analysis

Damasco C (1) Grassi E (1) Ala U (1) Gatti M (2) Di Cunto F (1)

(1) Dipartimento di Genetica, Biologia e Biochimica dell'Università di Torino,
Via Santena 5 bis - I-10126 Torino, Italy
(2) Dipartimento di Genetica e Biologia Molecolare, Istituto Pasteur-Fondazione Cenci Bolognetti,
Istituto di Biologia ePatologia Molecolari del CNR, Università di Roma "La Sapienza",
P.le A. Moro 5, 00185 Roma, Italy

Motivation

Microarrays are one of the most powerful techniques for the analysis and the measurement of gene expression. In the last few years, a huge amount of expression data, obtained with this technology from several model organisms has become available. Therefore, the development of processing tools aimed to extract useful functional information from published primary data has become a very important computational biology subject. We have previously described a data-mining approach called CLOE (Pellegrino et al., BMC Bioinformatics 2004) based on meta-analysis of microarray datasets from pairs of species, which evaluates genes coexpression and its phylogenetic conservation among species. With this approach, it's possible to make high confidence predictions about proteins' function and interaction. Since the approach can be applied to any couple of species, we have investigated the possibility of massively using it to identify new putative mitotic genes. In particular, we concentrated on Drosophila Melanogaster, which allows high-throughput experimental validation by RNAi. Furthermore we are exploring the possibility of using coexpression analysis for identifying the best positional candidate gene for mitotic mutations mapped to wide genomic loci by classical genetic approaches.

Methods

DNA microarray have been obtained from published studies performed with both Affymetrix (De Gregorio et al., PNAS 2001) and cDNA platforms. In particular, the data on which the best positional candidate predictor is based and the mammalian data used for combined analysis have been collected from the Stanford Microarray Database (http://smd.stanford.edu/cgibin/search/QuerySetup.pl). All the probes were completely re-mapped to the most significant biological databases, and in particular, Unigene, Entrez Gene and Ensembl. The orthology relationships were assigned on the base of HomoloGene tables. The computational analysis starts with the selection of database's probes referred to genes that compose a locus. For every probe we generate a list of all the other dataset's probes, ordered by a coexpression index. To this regard, we use the Pearson's correlation coefficient. For the functional statistical analysis, we introduce annotations of the Gene Ontology (GO) project. The basic mechanism of the predictor is to assign to every gene a functional score, based on the ranks of the genes annotated to the relevant GO function in the corresponding coexpression list. To gain a further level of confidence, the analysis is then performed on the orthologous genes of other species, such as human and mouse.

Results

The systematic use of our approach has led to the identification of a huge number of candidate Drosophila mitotic genes. For many of them, the prediction has been successfully validated by RNAi, leading to the discovery of new mitotic functions. Besides to this results, we will present the preliminary description and validation of our best candidate prediction tool.

Contact email: ferdinando.dicunto@unito.it