Analysis of co-regulated mRNA / miRNA pairs by *CoExpress* software tool

CENTRE DE RECHERCHE PUBLIC SANTÉ
MICROARRAY CENTER

UNIVERSITÉ DU LUXEMBOURG

Petr V. Nazarov[1], Viktar Khutko, Stephanie Schmitz[2], Arnaud Muller[1], Stephanie Kreis[2], Laurent Vallar[1]

petr.nazarov@crp-sante.lu

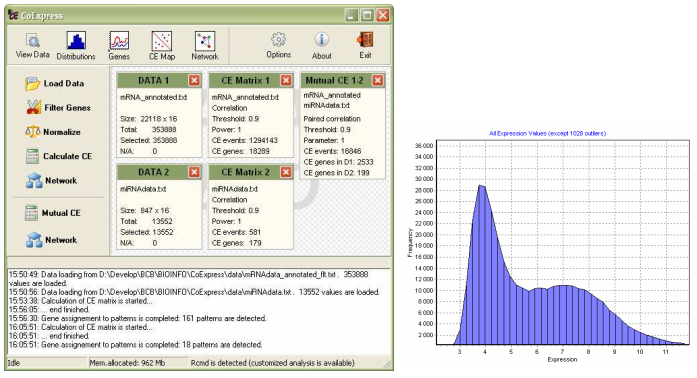
Introduction

Small noncoding miRNAs influence fundamental biological processes by ultimately altering the expression levels of proteins either through degradation of mRNA or through interference with mRNA translation. miRNAs tend to have long half lives, stable and characteristic expression profiles and therefore represent promising candidates as disease markers and therapeutic targets. Insights into miRNA functions and miRNA target genes can be obtained from simultaneous analyses of full genome transcription profiles and miRNA levels derived from the same sample types. Therefore, there is a need for effective and user-friendly tools for fast analysis of co-regulation between miRNAs and mRNAs.

Tool

Here we propose a free, stand-alone **software tool CoExpress** for the interactive co-expression (CE) analysis of mRNA and miRNA microarray data. The software is a user-friendly and allows on-the-fly study of CE, including:

- expression data **normalization** and R-based preprocessing (optionally);
- building and visualization of **CE matrix** using correlation or mutual information metrics for 2 data sets;
- clustering, visualization and filtering of **CE profiles**;
- building matrix of **mutual CE** for 2 datasets.



a) b)

Fig. 1. CoExpress interface with 2 datasets loaded (a), distribution of values in dataset (b)

Technical Notes

The software tool exists in two versions:

- **Windows-based version** for an interactive data analysis and visualization
- **Linux command line version** for multithreading analysis of big datasets

The properties and comparative description of both variants are given below:

Table 1. Comparative description of two versions of CoExpress

Parameters	Windows	Linux
Maximum genes	~ 20 000	> 100 000
Maximum arrays	< 1 000	> 1 000
Multi-CPU support	-	+
Graphical User Inerface	+	-
Compiller	bcc32	gcc
Time for CE calculation on a big dataset*		
1 CPU	3h 45m	55 m
8 CPU	n/a	7m
Time for CE calculation on a small dataset**		
1 CPU	1m 26s	1m 13s

(*) 2428 Affymetrix arrays with 19894 genes were used
(**) 17 Agilent two-color arrays with 15375 genes were used

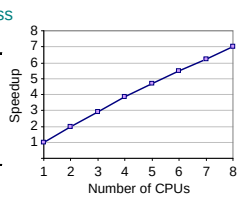
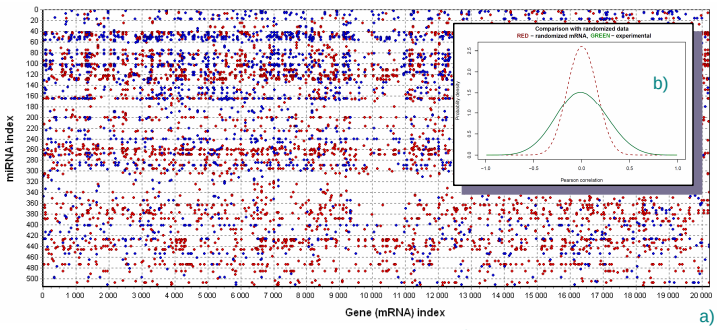


Fig. 2. Speedup with the increase of number of CPUs (on a big data set*)

For effective computing well-validated **boost** and **pthread** C++ libraries were used.

Results and Validation

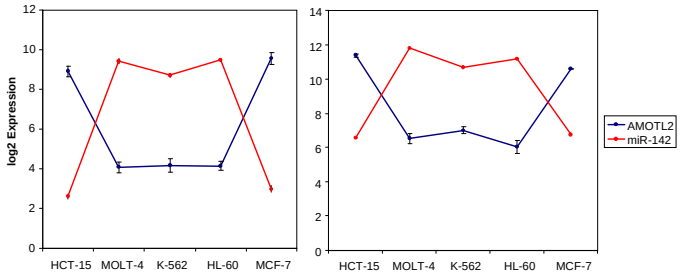
The performance of the software was tested using public mRNA and miRNA expression data from 14 various cell lines (A498, ACHN, CAKI1, CCRFCEM, HCT15, HL60, K562, MALME3M, MCF7, MOLT4, NCIH226, NCIH522, RPMI8226, SKOV3). Data from 42 Affymetrix® HGU133plus2 arrays and 14 miRNA custom microarray experiments were downloaded from public repositories, normalized and analyzed. We have detected 7423 co-expression events between 2533 mRNAs and 199 miRNAs with r^2>0.6 (see Fig 3). Remarkably, many negative co-expression events were detected in this two-dataset CE analysis, whereas in one-data set the vast majority of CE was positive.



a)

Fig. 3. Matrix for mRNA / miRNA co-expression for $r^2 > 0.6$ (a). Red dots show positive events, blue indicate negative. The results of bootstrapping validation are given in insertion (b), suggesting low p-values for the selected threshold.

22 of the most prominent mRNA-miRNA co-expression events were validated by qPCR, which showed good concordance with the results of co-expression analysis. A representative pattern from this analysis is presented in Fig. 4.



a) b)

Fig. 4. Example of expression pattern from qPCR (a) and corresponding patterns from public microarray analysis (b).

Concluding Remarks

Prediction based on experimental data is much more promising that prediction by computational programs. We present here a new program CoExpress, which can be applied to own or publicly available microarray data. In these data the method predicts possible gene targets for all miRNAs. We confirmed a high prediction power of this method by qPCR validation of predicted targets.

The **new version of CoExpress** (in development) is a cross-platform and multi-tread tool. We have used Qt SDK for GUI and cross-platform compilation. **CoExpress** will be further developed towards introducing advanced network reconstruction methods and integration with public databases and prediction analysis tools.

The current version of CoExpress is available for downloading from **www.bioinformatics.lu**

This work has been supported by National Research Fund of Luxembourg (FNR), University of Luxembourg and Public Research Centre Sante.

1. **Microarray Center, CRP-Santé, 84, Val Fleuri, L-1130, Luxembourg**
2. **Life Sciences Research Unit, University of Luxembourg, 162A, av. de la Faïencerie L-1511, Luxembourg**

Fonds National de la Recherche Luxembourg