A multifactorial integrative analysis scheme for combined mRNA and microRNA expression datasets

Alain Sewer[1], Vincenzo Belcastro[1], Arnd Hengstermann[2], Carole Mathis[1], and Julia Hoeng[1]
[1]Philip Morris Products SA, Neuchâtel, Switzerland
[2]Philip Morris Research Laboratories GmbH, Cologne, Germany

Introduction and Objectives

Extracting the relevant signal from high-throughput 'omics' datasets is a challenging task, which is essentially due to the high noise level arising from both biological and technical origins. Strategies to combine data from different sources have proved to be powerful, improving the quality of the results and allowing more successful interpretation. Here, we consider the case of combined microRNA and mRNA expression data from a multi-factorial *in vitro* experiment.

Our objective was to develop methodologies to extract mechanistic information from microRNA expression data: first to characterize the microRNA response to the experimental treatment and then to use complementary data sources such as microRNA target predictions and mRNA expression data derived from the same experiment to integrate the relevant microRNAs into regulatory interaction networks associated with the biological processes of interest.

Materials and Methods

Experiment. Organotypic human-derived bronchial epithelial airway cell cultures (MatTek) were directly exposed to cigarette smoke (CS) in a time- and dose-dependent manner using several exposure and post-exposure times (see table below giving the number of collected samples: **green**=CS-treatment and **red**=normal air control).

Experimental Factors		Post-exposure time *[hours]*				
		0.5	*2*	*4*	*24*	*48*
Exposure time *[minutes]*	*7*	3|3	3|3	3|3	3|3	3|3
	14	3|3	3|3	3|3	3|3	3|3
	21	3|3	3|3	3|3	3|3	3|3
	28	3|3	3|3	3|3	3|3	3|3

Expression data. Exiqon miRCURY LNA™ technology was used to profile the microRNAs and the Affymetrix HG-U133 Plus 2.0 platform was used to generate the mRNA expression data.
MicroRNA target predictions. TargetScan results were used to provide a list of candidate microRNA-mRNA regulatory interactions [1], which will be further filtered based on the results of the combined expression analysis.
Computational methods. Differential expressions were determined using linear models based on the experimental design [2]. Expression values are written as

$$E_{gc}^{(m)} = \beta_0^{(m)} + \beta_T^{(m)} \cdot T_c + \beta_D^{(m)} \cdot D_c + \beta_{TD}^{(m)} \cdot T_c \cdot D_c + \varepsilon$$

where m = {mRNA or microRNA at fixed post-exposure time-points}, the βs are the corresponding model coefficients, T = {1,0} - mean({1,0}) = {1/2,-1/2} describes the centered treatment values, D = {7,14,21,28} - mean({7,14,21,28}) = {-10.5,-6.5,6.5, 10.5} are the centered exposure times. The treatment-induced differential expression are:

$$\Delta E_{gc}^{(m)} = E_{gc}^{(m)} - E_{gc}^{(m)}\Big|_{\text{control}} = E_{gc}^{(m)} - (\beta_0^{(m)} - \beta_T^{(m)} \cdot |T_c| + \beta_D^{(m)} \cdot D_c - \beta_{TD}^{(m)} \cdot |T_c| \cdot D_c) + \varepsilon' = \beta_T^{(m)} + \beta_{TD}^{(m)} \cdot D_c + \varepsilon''$$

Combined analysis was based on the Pearson correlations R_{gm} between a pair of identically-sized vectors $\mathbf{x}_g$ and $\mathbf{y}_m$:

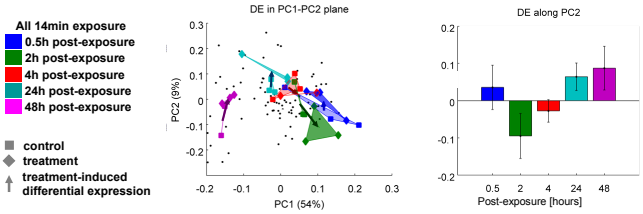
$$R = \frac{\sum_c (x_c - \bar{x}) \cdot (y_c - \bar{y})}{\sqrt{\sum_c (x_c - \bar{x})^2} \cdot \sqrt{\sum_c (y_c - \bar{y})^2}}$$

In the present case, $\mathbf{x}_g$ was constructed from the mRNA expression matrix $\Delta E_{gc}^{\text{mRNA}}$ and $\mathbf{y}_m$ from the microRNA expression matrices $\Delta E_{mc}^{\text{microRNA}}$, using the respective linear models defined above.

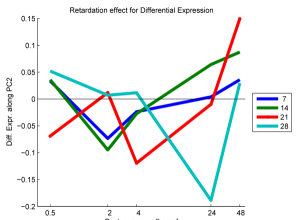
Results

Principal Component Analysis on the microRNA expression matrix

PCA on E_{mc}^{microRNA} indicated that treatment-induced differential expression (DE) is captured by the second principal component (PC2) and was therefore not the main source of variance in the data, which turned out to be the post-exposure time contained in PC1. PCA additionally showed that the CS-treatment accounted for a comparatively small fraction of the total variance (9%, while PC1 accounted for 54%).



PCA revealed a delay in response effect in the treatment-induced response along PC2: the longer the exposure time, the later the response along PC2. The same behavior was observed in gene expression.

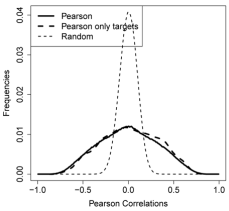


The calculations of the treatment-induced differential expressions based on the linear models revealed a subset of around 30 microRNAs that significantly responded to the CS exposure. Several distinct response patterns were observed.

Results *(continued)*

Combined analysis

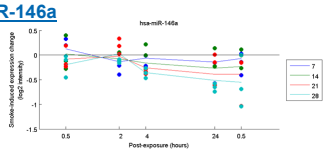
The implemented approach consisted of the following steps:
(1) match the samples between the two datasets,
(2) integrate the microRNA target mRNAs (TargetScan),
(3) extract the treatment-induced response using the linear models,
(4) separate according to post-exposure time,
(5) select the relevant target mRNAs.



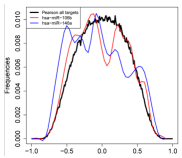
The figure shows that the overall distributions of the Pearson correlations between mRNA and microRNA expression values display an improved signal after completion of steps (1) and (2).

Application of the combined analysis: miR-146a

miR-146a is a microRNA involved in the cellular inflammatory response [3]. The application of the linear model to E_{mc}^{microRNA} revealed a statistically significant down-regulation from 4 hours onward.

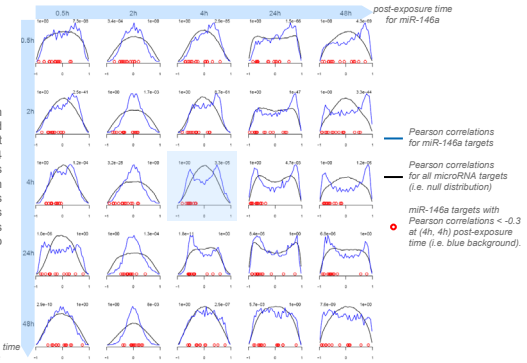


The results of steps (1) to (3) of the integrative analyses showed an enrichment of target genes for both negative and positive correlations [4]. Displaying enrichment with histograms allows a comparison with the results of a similar study [5].



Steps (4) and (5) revealed that the observed signal describes two distinct processes:
- Early (0.5, 2 and 4 hours) co-transcription of miR-146a with a first set of potential mRNA targets. At 2 hours, miR-146a expression positively correlated with genes involved in *inflammatory* response (also described in [6]).
- Intermediate (4 and 24 hours) down-regulation of a second set of potential mRNA targets. At 4 and 24 hours miR-146a expression negatively correlated with mRNAs of genes involved in *oxidative stress* and *cellular respiration*, and *catabolic process*, respectively.
- Late (48 hours) recovery of the mRNA expressions.

Positive correlation between miR-146a and a mRNA, for a fixed post exposure time (e.g., 4 hours), means coherence in expressions changes across exposure times (e.g. 28 minutes), and is not directly related to microRNA fold changes.



These relationships suggest the following kinetics (individual genes are currently investigated):

Summary and Conclusion

The selected computational tools were decisive in filtering the results and thereby ensuring their reliability. In particular, computing the microRNA differential expressions required a carefully chosen linear model to produce a satisfactory outcome. In spite of the incomplete knowledge of the microRNA functions, the integration of several data sources enabled the identification of distinct biological processes, as demonstrated for the case of miR-146a.

References

[1] Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets, Lewis BP *et al.*, Cell. 2005.
[2] Linear models and empirical bayes methods for assessing differential expression in microarray experiments, Smyth GK, Stat Appl Genet Mol Biol. 2004.
[3] NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses, KD Taganov *et al.*, PNAS 2006.
[4] MicroRNA-mediated Feedback and Feedforward Loops are Recurrent Network Motifs in Mammals, Tsang J *et al.*, Mol. Cell. 2007.
[5] MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium, Schembri F *et al.*, PNAS 2009.
[6] Divergent intracellular pathways regulate interleukin-1beta-induced miR-146a and miR-146b expression and chemokine release in human alveolar epithelial cells, Perry MM *et al.*, FEBS Lett 2009

PMI RESEARCH & DEVELOPMENT

The 8th Annual RECOMB Satellite on Regulatory Genomics 2011
Barcelona, Spain
18-19 October

Philip Morris International Research & Development, Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland
T: +41 58 242 21 11, **F**: +41 58 242 28 11, **W**: www.philipmorrisinternational.com