

## **Meta-analysis of breast cancer microarray data: reliable identification of up- and down-regulated genes.**

Kondrakhin Yu.V.<sup>1,2,#</sup>, Poroikov V.V.<sup>3</sup>, Sharipov R.N.<sup>4,1,2</sup>, Kel A.E.<sup>5</sup> and Kolpakov F.A.<sup>1,2,\*</sup>

<sup>1</sup>Institute of Systems Biology, Novosibirsk, Russia;

<sup>2</sup>Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia;

<sup>3</sup>Institute of Biomedical Chemistry, RAMS, Moscow, Russia;

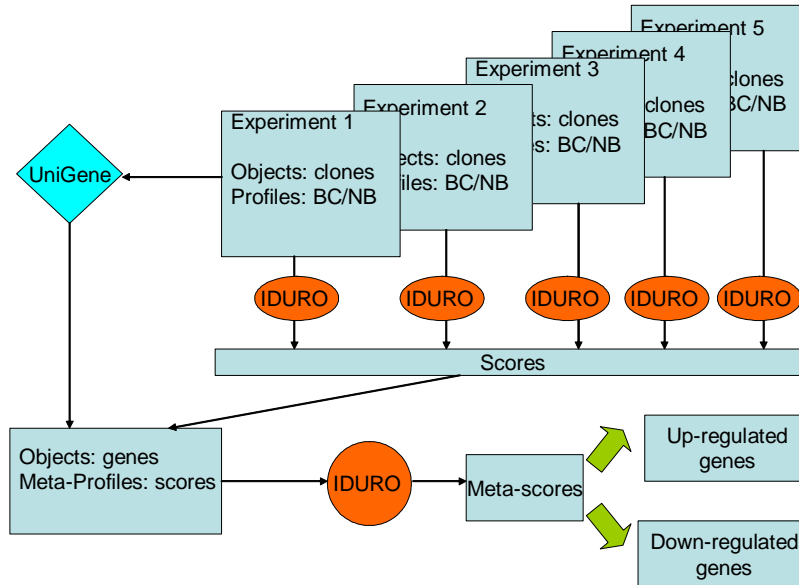
<sup>4</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia; and

<sup>5</sup>BIOBASE GmbH, Wolfenbuettel, Germany

\*[fedor@biouml.org](mailto:fedor@biouml.org)

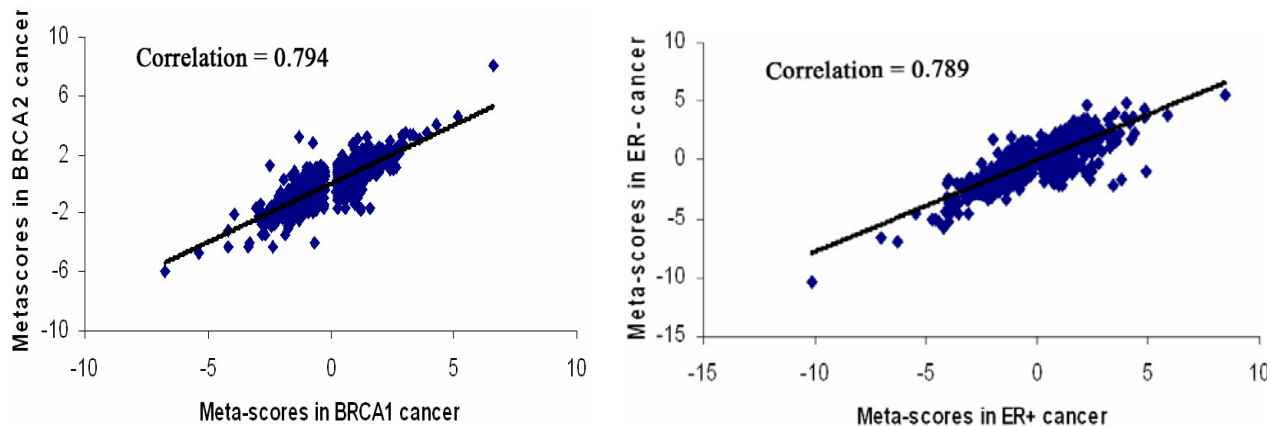
The general goal of meta-analysis is to obtain reliable results by processing integrated data sets derived from independent experimental studies. Different meta-analysis schemes have been proposed to address issues raised by cancer profiling studies. Thus, a meta-signature associated with breast cancer prognosis was derived in [1]. The significantly co-occurred genes were identified in [2] on the basis of meta-analysis of gene expression signatures published in the literature. The main purpose of our meta-analysis is to identify the up- and down-regulated genes in breast cancer with respect to normal breast tissues. To our knowledge, meta-analysis of microarray data on breast cancer has not been performed with this purpose so far.

To identify the up- and down-regulated genes we analyzed five independent data sets published (see Figure 1) [3-7]. For this purpose, we have developed the novel method IDURO (Identification of Down- and Up-Regulated Objects) based on optimization of parameters of hyper-geometrical and binominal distributions. The meta-analysis is two-step. At step 1 IDURO applied to the clone profiles consisted of the BC/NB measurements (Breast Cancer / Normal Breast) and at step 2 it applied to the meta-profiles consisted of scores obtained previously. Thus, at step 1 IDURO assigned a score to each analyzed clone and at step 2 IDURO assigned a meta-score to each analyzed gene. Scores and meta-scores were defined up to the sign as the base-ten log-transformations of the optimized p-values. If the meta-score was positive or negative, then gene was regarded as up- or down-regulated, respectively. Decision about statistical significance for the differentially expressed genes was made by striking a reasonable balance between the alternative characteristics, the False Discovery Rate and the Miss Rate. The FDR controls the relative number of genes that are not dysregulated, among those identified as significantly dysregulated, on the one hand; the miss rate controls the relative number of genes that are truly dysregulated, among those identified as non-significantly dysregulated, on the other hand.



**Figure 1.** The general scheme of meta-analysis.

The developed method demonstrated that the differences in the breast cancer subtypes are due to the differential expression of small number of genes, whereas the behavior of most genes is the same for the breast cancer subtypes. Figure 2(left) demonstrates the observed gene expression similarity between BRCA1 and BRCA2 subtypes, and figure 2(right) – between estrogen receptor positive (ER(+)) and estrogen receptor negative (ER(-)) breast cancer subtypes.



**Figure 2.** Relationship between: left) BRCA1 and BRCA2 subtypes of breast cancer and right) ER(+) and ER(-) breast cancers.

We also performed comparative analysis of regulatory regions ([-1000bp, +1000bp] – relative transcription start site) of the 400 most differentially expressed genes. The potential

transcription binding sites were recognized on the base of matrix approach by using matrices from TRANSFAC. Further discriminant analysis allowed to reveal a number of transcription factors, for which the concentration of binding sites differed significantly in:

- up-regulated genes: NF-Y (TRANSFAQ matrices M00287, M00209 and M00775), YY1 (M01035, M00793), E2F (M00516, M00736, M00737), c-Ets-1 (M00339), PITX2 (M00482), Crx (M00623), SREBP-1 (M00220), and STAT3 (M00497);

- down-regulated genes: SRF (M01007, M00922 and M00152), HEB (M00698), TTF1 (M00432), HSF (M00163), FOXO3 (M00477), FOXO4 (M00476), ATF3 (M00513).

1. Shen R, Ghosh D and Chinnaiyan AM. Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* **5**:94 (2004).
2. Finocchiaro DF, Mancuso FM, Cittaro D and Muller H. Graph-based identification of cancer signaling pathways from published gene expression signatures using PubLiME. *Nucleic Acids Res* **35**:2343-2355 (2007).
3. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, *et al.* Gene expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**, 539-48 (2001).
4. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, *et al.* Gene expression profiles of human breast cancer progression. *Proc. Natl Acad. Sci. USA* **100**, 5974-9 (2003).
5. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869-74 (2001).
6. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**, 9212-17 (1999).
7. Zhao H, Langerod A, Ji Y, Nowels KW, Nesland JM, Tibshirani R, Bukholm IK, Karesen R, Botstein D, Borresen-Dale AL, Jeffrey SS. *et al.* Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell* **15**, 2523-36 (2004).

*This research was supported by European Committee grant №037590 “Net2Drug”.*