Time-course whole-genome microarray analysis of estrogen effects on hormone-responsive breast cancer cells

ID - 238

Mutarelli Margherita[1,2], Cicatiello Luigi[1,3], Ravo Maria[1,4], Grober Oli' Maria Victoria[1,5], Facchiano Angelo[2], Angelini Claudia[6], Weisz Alessandro[1,3]

[1]Department of General Pathology - Second University of Naples, Naples

[2]Institute of Food Sciences, National Research Council (ISA-CNR), Avellino

[3]AIRC Naples Oncogenomics Center

[4]Research Doctorate 'Ph.D. in Toxicology v Oncology and Molecular Pathology', University of Cagliari, Cagliari

[5]Research Doctorate 'Pathology of Cellular Signal Transduction', Second University of Naples

[6]Istituto per le Applicazioni del Calcolo, National Research Council (IAC-CNR), Naples

Motivation

Estrogens (E2) are key regulators in many biological processes, along with a highly recognized role in breast cancer. Estrogen actions are particularly suited to be studied with microarray-based technologies since the Estrogen Receptor (ER) is a transcription factor and its action affects so many different processes within the cell that only a wide-scale technique can avoid loosing key aspects of its multi-purpose action. We studied the effect of a stimulation with a mitogenic dose of E2 on two different breast cancer cell lines in order to focus our attention primarily on a E2 core response and not on cell type specific responses. The microarray platform used has been recently developed and includes some interesting technical innovations: one for all, it includes an average of 30-40 replicates for each probe sequence, which are randomly distributed on the chip surface. This allows the signal to be more robust, respect to local fluctuations in hybridization efficiency or spatial effects, and provides a standard deviation relative to each probe-specific hybridization signal.

Methods

RNA was extracted from human breast cancer cell-lines (ZR-75.1 and MCF-7) stimulated with E2 following starvation in steroid-free medium for 4 days; RNA samples were extracted before or after 1, 2, 4, 6, 8, 12, 16, 20, 24, 28 and 32 hours hormonal stimulation and the sample before stimulation was taken as reference. RNA extraction and hybridization reactions were performed with Illumina Human WG-6 BeadChip protocols following manufacturers recommendations. In order to investigate the influence of pre-processing procedures, five types of normalization were applied on data, three of them present in the Illumina BeadStudio analysis software and two of them performed using R/Bioconductor statistical environment [1]. The probes with more than 15% missing values were filtered out from the analysis. Automatic selection of genes whose time-dependent profile was significantly different from zero was performed using BATS (Bayesian Analysis Time Series) software [2], a newly-developed tool with a truly functional approach which treats the gene expression temporal profile as a global signal instead of a collection of independent time-points (as commonly done in most of existing methods). After having identified estrogen responsive genes for each cell line, another Bayesian functional based software, Splinecluster [3], was used to obtain cluster analysis of the genes common to the two cell lines. Finally, an analysis of the genes with a common expression profile, taking into account a slight difference in cell-cycle timing, was conducted.

Results

We evaluated the performance of different normalization algorithms in terms of overlap between detected genes. We applied all normalizations available on the proprietary software plus a standard method developed for single-channel microarrays (quantile) and a newly-developed method (lumi), which is the first one to use the probe standard deviation, a measure until now unique to this platform. The only non-linear algorithm present in BeadStudio had to be discarded since the data produced was not normalized between the arrays, thus requiring a further normalization step that we decided not to apply to avoid excessively manipulation on data.

Since BATS software allows the user to treat the noise present in data, which has to be removed to detect and estimate gene signal, in different ways (thoroughly described in its documentation) we have used eight combination of error models and parameters choice, to eventually select a list of gene expression profiles significantly affected by the E2 treatment, regardless of fluctuation in data due to the normalization algorithm applied or the way noise is estimated. Selection was found robust with respect to changes in parameters and type of normalization.

We obtained a list of 630 genes common to the two cell lines, which underwent cluster analysis separately with the corresponding expression signals in the two datasets. We show the amount of overlap between the profile clusters in the two cell lines, both as gene composition and expression profile over time, as well as

an analysis of the presence of known cell cycle markers.

References 1. Gentleman R et al Bioconductor: Open software development for computational biology and bioinformatics Genome Biology, 5:R80 (2004) 2. Angelini C, De Canditiis D, Mutarelli M and Petsky M Bayesian Approach to estimation and testing in time course microarray experiments Tech Rep IAC-CNR n.317/06 (2006) 3. Heard NA, Holmes CC and Stephens DA A quantitative study of gene regulation involved in the Immune response of Anopheline Mosquitoes: An application of Bayesian hierarchical clustering of curves J Amer Statist Assoc 101, 18-29 (2006)

Acknowledgements Research supported by: MIUR (PRIN 2005063915_003 and 2006069030_003), Reg. Campania (LR5-2003), UE (LSHM-CT2005-018652), SUN, CNR-Bioinformatics Project, Oncoproteomics Project Conv.n.527B/2A/10 and AIRC.

Email: margherita.mutarelli@isa.cnr.it