Comparative analysis with three different microarray platforms of the oestrogen-responsive transcriptome from breast cancer cells

Margherita Mutarelli[(1),(2)], Walter Basile[(3)], Luigi Cicatiello[(3)], Claudio Scafoglio[(3)], Giovanni Colonna[(2)], Alessandro Weisz[(3)] Angelo Facchiano[(1),(2)]

[1] Istituto di Scienze dell'Alimentazione CNR, via Roma 52A/C, 83100 Avellino

[2] Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università degli Studi di Napoli, via de Crecchio, 80138 Napoli, Italy

[3] Dipartimento di Patologia Generale, Seconda Università degli Studi di Napoli, via de Crecchio, 80138 Napoli

Keywords. Microarray, Estrogen response, Breast cancer

The DNA microarray technique makes it possible to analyze the expression patterns of tens of thousands genes in a short time [1, 2]. The wide use of this technique and the rapidly improving different technologies available by several commercial and academic providers has led to the publication of thousands of results, extremely heterogeneous with respect to the type of technology used, to the kind of normalization and analysis subsequently applied to data an so on. This leads to a difficulty in collaborating and exchange data between groups with common research interest, whereas collaborations would be extremely useful due to the high cost of this techniques but also to the consideration that an experiment carefully designed could bring results relevant to different groups, each focusing on a different aspect of a main biological problem. So the awareness for the need of common standards or, at least, comparable technologies is emerging in the scientific community, as shown by the effort of the on-purpose Microarray Gene Expression Data (MGED) Society, which is trying to set up at least experimental methodology, ontology and data format standards [3].

In addition, it is important the ability of being able to compare newly produced data with preceding experiments, so to ensure of keeping high the value of results produced with equipment of the old generation[4]. Otherwise, a large amount of the work produced until the outcome of a new release of technology would be lost. This, considering that the huge amount of data produced is largely underexploited, would be a great loss for the scientific community. In fact, as analysis algorithms are improving, existing data can be re-analyzed to give more precise results, thus helping to adjust the planning of future experiments.

We thus started this work with the aim of evaluating the technical variability between three commonly used microarray platforms, such to adapt the first part of the analysis to the peculiarity of each technique, and the feasibility of a common subsequent analysis path, thus taking advantage of the different data-extraction abilities of the three. For this purpose, we used three different commercial chips to study the gene expression profiles of hormone-responsive breast cancer cells with and without stimulation with estradiol:

i) the Incyte 'UniGEM V 2.0' microarrays, containing over 14,000 PCR-amplified cDNAs, corresponding to 8286 unique genes, spotted at a high density pattern onto glass slides [2, 5];

ii) the Affymetrix technology, based on 25 nucleotide-long oligonucleotides directly synthesized on a GeneChip® array, representing more than 39,000 transcripts derived from approximately 33,000 unique human genes [6, 7];

iii) the Agilent 'Human 1A Oligo' Microarray consisting of 60-mer, *in situ* synthesized oligonucleotide probes for a total of about 18000 different genes [8, 9].

The RNA derived from human breast cancer cells (ZR-75.1) stimulated for 72 hrs with 17beta-estradiol after starvation in steroid-free medium for 4 days; the reference sample was derived from synchronized cells grown in steroid-free environment [10, 11]. The same samples were used to generate fluorescent targets to be hybridized on the different slides. Hybridization reactions were performed with four (for the Agilent slides) and two or three (for the other platforms) technical replicates, with a single (Incyte) or double (Agilent), balanced dye swap for competitive hybridizations.

A total combined number of 18,823 unique UniGene clusters were represented among the three platforms used. By focusing only on a subset of 5,733 genes that were present in all the chips, about 50% appeared to be significantly expressed and 25% genes resulted significantly regulated by 17beta-estradiol treatment in our experiment. A quite low overlapping was observed between the lists of regulated genes obtained by the three systems. We are working on understanding the conflicting results on some of the genes. The majority of genes were detected by only the Affymetrix platform, probably as a consequence of the higher sensitivity of this system, which allows the detection of some gene expression levels that are not identified with the other platforms. However, a number of genes was identified only by the cDNA and/or

oligonucleotide systems. Another possible experimental explanation is that the DNA sequences spotted on the arrays show different affinity for the target, so each slide has a particular pattern of probe-target annealing, although the same genes are represented on all the platforms. Finally, we are improving the data processing by statistical methods in order to allow the better understanding of the experimental results.

Ackowledgements

Research supported by: Italian Association for Cancer Research (AIRC, Investigator Grants 2003), Italian Ministry for Education, University and Research: FIRB Post-genomica (Grant RBNE0157EH_001/_003), European Commission (Contract BMH4-CT98-3433), Ministry of Health (Progetti Speciali 2000 and RF02/184), Second University of Naples (Ricerca di Ateneo 2002-2003). M. Mutarelli and C. Scafoglio are PhD students, supported by EU funds, of Dottorato di Ricerca in "Biologia Computazionale" and in "Oncologia medica e chirurgica ed Immunologia clinica", respectively, at the Second University of Naples.

References

1. AA. VV. The Chipping Forecast II *Nat. Genet. Suppl.* **32**:461-552 (2002).
2. DeRisi, J. *et al*. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet*. **14**:457-460 (1996).
3. Brazma, A. *et al*. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**: 365-371 (2001).
4. Nimgaonkar, A. *et al*. Reproducibility of gene expression across generations of Affymetrix microarrays. *BMC Bioinformatics* **4**:27-38 (2003).
5. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**:467-470 (1995).
6. Lipshutz, R.J. *et al.* Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**:442-444 (1995).
7. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**:1675-1680 (1996).
8. AA. VV. Development of an in situ synthesized oligonucleotide microarray for gene expression monitoring of the budding yeast Saccharomyces cerevisiae. *GSAC TIGR conference* (2001)
9. Janssen, D., The microarray debate continues. *Genomics & Proteomics* **3**(6):40-43 (2003).
10. Cicatiello, L., Natoli, G., Scafoglio, C., Altucci, L., Cancemi M., Facchiano, A., Calogero, R., Iazzetti, G., De Bortoli, M., Sfiligoi, C., Sismondi, P., Biglia, N., Bresciani, F. & Weisz, A. The gene expression program activated by estrogen in hormone-responsive human breast cancer cells. J. Mol. Endocrinol. (2004) *In press.*
11. Weisz, A., Basile, W., Scafoglio, C., Natoli, G., Altucci, L., Bresciani, F., Facchiano, A., Sismondi, P., Cicatiello, L. & De Bortoli, M. Molecular identification of ERalpha-positive breast cancer cells by the expression profile of an intrinsic set of hormone regulated genes. J. Cell Physiol. (2004). *In press.*
12. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet. Suppl.* **32**:496-501 (2002).
13. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.* **98**:5116-5121 (2001).
14. Lee, J.K. *et al*. Comparing cDNA and oligonucleotide array data: concordance of gene expression across platforms for the NCI-60 cancer cells. *Genome Biology* **4**:R82 (2003).