

Analysis of the estrogen-responsive transcriptome from breast cancer cells: a comparative analysis with three different microarray platforms

Luigi Cicatiello¹, Claudio Scafoglio¹, Angelo Facchiano^{2,3}, Margherita Mutarelli^{2,3}, and Alessandro Weisz¹

¹ Dipartimento di Patologia Generale, Seconda Università degli Studi di Napoli, via de Crecchio, 80138 Napoli, ² Istituto di Scienze dell'Alimentazione CNR, via Roma 52A/C, 83100 Avellino and

³ Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università degli Studi di Napoli, via de Crecchio, 80138 Napoli, Italy

The DNA microarray technique is a powerful method that provides the opportunity to analyze the expression patterns of tens of thousands genes in a short time [1, 2]. Presently, several commercial and academic providers offer printed DNA microarrays, also known as chips, prepared according to a variety of technologies. In some cases, amplified cDNAs are spotted at a high density pattern onto a solid substrate such as a glass slide [3, 16], while, on the other hand, a commercial supplier provides oligonucleotide arrays constructed by synthesizing approximately 20- to 25-mer oligonucleotide probes directly onto a glass or silicon surface using photo-lithographic technology [13, 14]. A third generation technology is however emerging, as an attempt of overcoming the drawbacks of the other two platforms, which makes use of longer oligonucleotides (50- to 70-mer) onto the array [4], thus representing a compromise between the other two technologies.

Due to the powerful nature of microarray analysis, the number of relevant publications is increasing exponentially and it would be interesting to use this huge amount of data available to directly compare newly produced data sets. Also, cross-platform utilization of data from different technologies has the potential to reduce the need to duplicate experiments and also to make feasible and fruitful collaborations among research groups which may use different techniques, but all this requires corresponding measurements to be comparable [5, 6].

In order to evaluate the technical variability among different microarray platforms, we used three different commercial chips to study the gene expression profiles of hormone-responsive breast cancer cells following stimulation with estradiol [7-12]. The same samples were used to generate fluorescent targets to be hybridized on the different slides. The following microarray platforms were used:

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

- ii) the Agilent ‘Human 1A Oligo’ Microarray consisting of 60-mer, *in situ* synthesized oligonucleotide probes for a total of about 18000 different genes;
- iii) the Incyte ‘UniGEM V 2.0’ microarrays, containing over 14,000 PCR-amplified cDNAs, corresponding to 8286 unique genes.

The RNA derived from human breast cancer cells (ZR-75.1) stimulated for 72 hrs with 17 β -estradiol (E2) after starvation in steroid-free medium for 4 days; the reference sample was derived from synchronized cells grown in steroid-free environment. Hybridization reactions were performed with three (for the Incyte chips) or four (for the other platforms) technical replicates, with a single (Incyte) or double (Agilent), balanced dye swap for competitive hybridizations. A total intensity normalization [15] was performed for the Incyte and the Agilent data, while the Affymetrix algorithm was used for the GeneChip expression data. Selection of significantly regulated genes was performed through the ‘Significance Analysis of Microarrays’ (SAM) software, setting the Delta value to gain a false discovery rate (FDR) ranging from 0.02 to 5% [17].

A total combined number of 18,823 unique UniGene clusters were represented among the three platforms used. In order to compare the sensibility and the reliability of the three microarrays, we focused on a subset of 5,733 genes that were present in all the chips. A total number of 2,494 common genes appeared to be significantly expressed. Among these, 1,058 genes resulted significantly regulated by E2 treatment in our experiment. The following table summarizes the results:

	Selected genes	Induced genes	Inhibited genes	Inverted regulation
Agilent + Incyte + Affymetrix	47	30	17	0
Agilent + Incyte only	3	2	0	1
Agilent + Affymetrix only	50	29	18	3
Incyte + Affymetrix only	206	146	59	1
Agilent only	34	22	12	-
Incyte only	58	40	18	-
Affymetrix only	660	398	262	-
Total	1058	667	386	5

Table I – *Details of the results obtained in this study.*

Surprisingly, a very low overlapping was observed between the gene lists generated with the three systems: only 47 genes (4.4% of total) resulted regulated in all the three data sets, with full agreement among the platforms concerning induced (30) and inhibited (17) genes. The regulation of

259 genes (24.5%) could be detected with two of the array systems used (206 by Affymetrix and Incyte, 50 by Affymetrix and Agilent, 3 by Incyte and Agilent), with few cases of genes showing opposite behaviour (up- vs down-regulation). The majority of genes (752, 71.1% of total) were detected, however, by only one of the methods used, and in particular 660 genes were identified only by the Affymetrix platform.

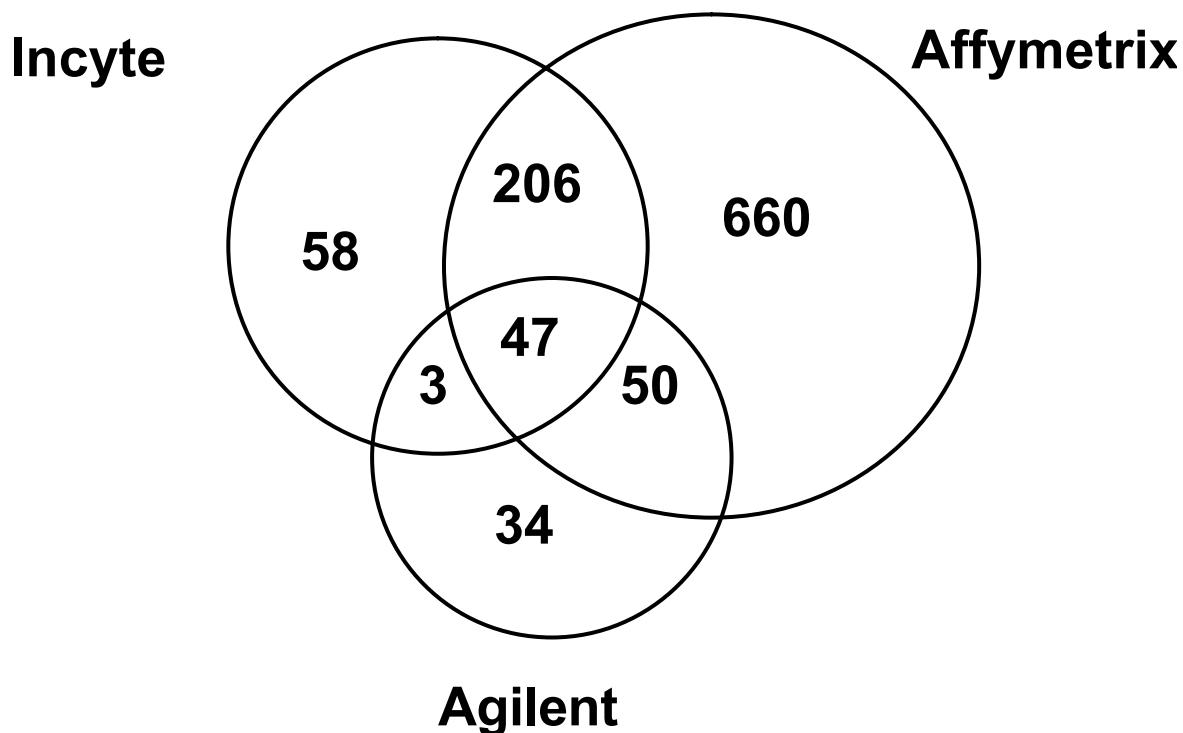


Figure 1 - Venn diagram of the comparison of the three platforms used for this study.

One reason for the observed discrepancy could be the higher sensitivity of the Affymetrix system, which allows the detection of some gene expression levels that are not identified with the other platforms. This cannot account, however, for the genes identified only by the cDNA and/or oligonucleotide systems. Another possible 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.

Whatever the explanation, our results point out an important issue in microarray data analysis; the data deriving from the three platforms were processed with the same statistical procedures and the overlapping genes were only 4.4% of the common genes studied. Should we consider really estrogen-responsive only those 47 genes whose differential expression was detected by all the three

systems? Or should we trust each single platform, whether confirmed or not by the others, thus extending our list to 1,058 genes? Should we search an experimental validation with other techniques (i.e. RT-PCR) before believing in the microarray results, thus enormously reducing the field of our analysis?

We believe that the answer to these questions will be brought about only by a strong attempt to standardization of the technique through all its steps, from probe selection to data mining, among many different laboratories across all the world [18].

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), European Commission (Contract BMH4-CT98-3433), Ministry of Health (Progetti Speciali 2000 and RF02/184), Second University of Naples (Ricerca di Ateneo 2002-2003).

C. Scafoglio is a PhD student of Dottorato di Ricerca in Oncologia medica e chirurgica ed Immunologia clinica (XVIII° ciclo), Seconda Università degli Studi di Napoli.

M. Mutarelli is a PhD student of Dottorato di Biologia Computazionale (XVIII° ciclo), Seconda Università degli Studi di Napoli.

Bibliography

1. AA. VV. The Chipping Forecast II *Nat. Genet. Suppl.* **32**:461-552 (2002).
2. Ambrosino, C., Cicatiello, L., Scafoglio, C., Altucci, L. & Weisz A. A Combinatorial Approach to Gene Expression Analysis: DNA Microarrays. In: Combinatorial Chemistry and Technology, Principles, Methods and Applications, 2nd ed. (Eds. Miertus, S., Fassina, G.), Marcel Dekker, New York, 2003. *In press*.
3. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**:457-460 (1996).
4. Janssen, D., The microarray debate continues. *Genomics & Proteomics* **3**(6):40-43 (2003).
5. Kothapalli, R., Yoder, S.J., Mane S. & Loughran jr., T.P. Microarray results: how accurate are they? *BMC Bioinformatics* **3**:22-31 (2002).
6. Kuo, W.P., Jenssen, T., Butte, A.J., Ohno-Machado L. & Kohane I.S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18**:405-412 (2002).
7. Cicatiello, L., Natoli, G., Scafoglio, C., Altucci, L., Cancemi M., Facchiano, A., Calogero, R., Iazzetti, G., De Bortoli, M., Sfiligoi, C., Sisoni, P., Biglia, N., Bresciani, F. & Weisz, A. The gene expression program activated by estrogen in hormone-responsive human breast cancer cells. *J. Mol. Endocrinol.* *In press*
8. Weisz, A., Cicatiello, L., and Bresciani, F. Microarray-based gene expression monitoring of hormone-responsive human breast cancer cells during estrogen-induced G1 progression. EMBO Workshop 'Structure and Function of Nuclear Receptors', Villefranche-sur-Mer, France, Abstracts Book, pg. 60. 1999
9. Sorbello, V., Fuso, L., Sfiligoi, C., Scafoglio, C., Ponzone, R., Biglia, N., Weisz, A., Sisoni, P. & De Bortoli, M. Quantitative *real-time* RT-PCR analysis of eight novel oestrogen-regulated genes in breast cancer. *Int. J. Biol. Marker* **18**: 123-129, 2003
10. Weisz, A., Basile, W., Scafoglio, C., Natoli, G., Altucci, L., Bresciani, F., Facchiano, A., Sisoni, 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.* 2003. *In press*
11. Cicatiello, L., Facchiano, A., Caligero, R., De Bortoli, M., Bresciani, F., & Weisz, A. Gene expression monitoring in hormone-responsive human breast cancer cells during estrogen-induced cell cycle progression. Oncogenomics: Dissecting Cancer through Genome Research. AACR/Nature Genetics Joint Conference, Jan. 25-27, 2001, Tucson, AZ (U.S.A.). Abs 60, pg. 95. 2001. http://www.nature.com/cgi-taf/DynaPage.taf?file=/ng/journal/v27/n4s/full/ng0401supp_95a.html&_UserReference=C0A804EC4650B9B7E02E44E479153B0D3D72
12. Weisz, A., Cicatiello, L., Bresciani, F., Facchiano, A., Caligero, R., Iazzetti, G., Sfiligoi, C., De Bortoli, M.,

- & Sismondi, P. Gene expression profiling of human breast cancer ZR-75.1 cell response to estrogen. EMBO Workshop 'Nuclear Receptor Structure and Function', Erice (Italy), Abstracts Book, P69. 2001
13. Lipshutz, R.J. *et al.* Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**:442-444 (1995).
 14. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**:1675-1680 (1996).
 15. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet. Suppl.* **32**:496-501 (2002).
 16. 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).
 17. 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).
 18. Facchiano, A. & Weisz, A. Internet tools for the analysis of gene expression by data bases integration. Proceedings of the NETTAB 2001 Workshop: 'CORBA and XML: Toward a bioinformatics integrated network environment' (Eds. Iannotta, B., F. Malusa, D. Marra, P. Romano), pp. 99-101, 2001.