Getting the most out of comparative microarray data analysis: analysis of the estrogen-responsive transcriptome from breast cancer cells with four different microarray platforms

Mutarelli M (1,2), Scafoglio C (3,4), Cicatiello L (3,5), Colonna G (1),
Facchiano A (1,2), Weisz A (3,5)

(1) Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche,
Seconda Università degli Studi di Napoli
(2) Istituto di Scienze dell'Alimentazione CNR, Avellino
(3) Dipartimento di Patologia Generale, Seconda Università degli Studi di Napoli
(4) Department of Medicine, University of California S. Diego
(5) AIRC Naples Oncogenomics Center, Napoli (Italy)

Motivation

Presently, several commercial and academic providers offer printed DNA microarrays, also known as chips, prepared according to a variety of technologies. After a first generation of spotted cDNAs at a high density pattern onto a solid substrate such as a glass slide, the emerging standard is now the spotted or in-situ synthesis of short (25- to 30-mer) or longer oligonucleotides (50- to 70-mer) oligonucleotide probes directly onto a glass or silicon surface. A new and interesting innovation is the in-situ synthesis of the probes on beads, which are then randomly distributed on the chip surface. In order to evaluate the technical variability among different microarray platforms, we used four different commercial chips to study the gene expression profiles of hormone-responsive breast cancer cells following stimulation with estradiol. We decided to use only oligonucleotide-based arrays since previous comparative analysis have shown little reproducibility in cDNA microarray data and low overlapping in results among cDNA and oligo.

Methods

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 (G4110A) consisting of 60-mer, in situ synthesized oligonucleotide probes for a total of about 18000 different genes; iii) the Amersham CodeLink Human Whole Genome Bioarray, consisting of 30-mer, in situ synthesized oligonucleotide probes, for a total of about 52,000 different genes; iv) the Illumina Sentrix Human-6 (whole-genome) BeadArray, each containing 6 arrays which can be hybridized individually, consisting of 50-mer, synthesized in situ on beads randomly dispersed on the chip surface, for a total of about 46,000 different genes. The RNA derived from human breast cancer cells (ZR-75.1) stimulated for 72 hrs with 17β-estradiol (E2) after starvation in steroid-free medium for 4 days; the reference sample was derived from synchronized cells grown in steroid-free environment. A total intensity normalization was performed for the Agilent data, while the rma algorithm was used for the Affymetrix GeneChip expression data, and the same quantile normalization was performed for the Amersham Codelink data. As for the Illumina technology, since it was a new platform to our experience (and we had a lower number of technical replicates), we evaluated all four types of normalization provided by the Illumina BeadArray Image analysis software. 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) of about 0.01 for all platforms. Particular care was used to find corresponding probes among platforms, since in our experience and in literature it has been shown that simple matching of probes using manufacturer's annotation (such as Unigene cluster or GeneBank accession number) could result in a little overlap of resulting gene sets and inconsistencies probably due to different design of probe sequences and/or non-updated annotation. For this reason, we decided to use probe sequences to verify the actual association to a well-curated sequence database as the NCBI RefSeq.

Results

Results will be presented in a poster and discussed will all participants during the meeting. Research supported by: Italian Association for Cancer Research (AIRC, Investigator Grants 2003), Italian Ministry for Education, University and Research: FIRB Post-genomica (Grant RBNE0157EH), European Commission (Contract BMH4-CT98-3433), Ministry of Health (Progetti Speciali 2000 and RF02/184), Second University of Naples (Ricerca di Ateneo 2002-2003).

Contact email: mailto:margherita.mutarelli@unina2.it