DNA microarrays in cancer diagnosis and prognosis

Microarranjos de DNA no diagnóstico e prognóstico do câncer

Oswaldo Keith Okamoto*

ABSTRACT
This review discusses recent advances in our understanding of the human genome and the application of derived technologies in the medical area. It focuses on the use of DNA microarray for diagnosis, prognosis, and therapeutic purposes in oncology, and the resultant impact it might have on routine clinical practice.

Keywords: Genomics; Oligonucleotide array sequence analysis; Neoplasms; Gene expression; Tumor markers, biological

INTRODUCTION
With the completion of the human genome sequencing(1), a powerful research resource has been made available. Large genomic data sets are being constantly generated and deposited into public databases, pushing biological and biomedical research to a whole new level. Researchers now face the important task of decoding this immense amount of data into useful information. The identification of key genes and regulatory pathways, for instance, may help uncover the biochemical and physiological processes related to certain pathological conditions.

Studying the biological role of genes and how they are orchestrated in the genome is the aim of Functional Genomics. In the medical area, this type of knowledge is being exploited as a basis for the understanding of human diseases. Related advances in technology such as microarrays allow genome-wide analyses of gene expression in biological samples of interest. Besides its impact in the basic research field, high-throughput microarray technology has boosted the biotechnological and pharmaceutical industries, with the promise of improving diagnostics and treatment options.

The microarray technology
DNA microarrays are a powerful tool to investigate biological processes through the large-scale analysis of genes from a particular cell, tissue, or organism. In this system, thousands of probes (usually cDNA or oligonucleotides), each representing a unique gene, are orderly spotted/synthesized onto glass slides or similar solid supports. This platform is also known as DNA chips. The principle of the technique relies on the specific hybridization between the probes on the chips and the sample targets (fluorescent-labeled DNA, cDNA or cRNA molecules).

Generally, DNA microarrays are used to monitor changes in gene expression levels. In this type of application, a single microarray assay can simultaneously analyze the expression of thousands of genes through the quantification of their respective transcripts (mRNA levels). Gene expression is estimated by comparing the relative amount of mRNA in two distinct cell populations. The target mRNA from control and test samples are labeled with fluorescent dyes and after hybridization, the amount of mRNA bound to the chip is estimated by the intensity of the corresponding light emission. The mRNA levels in control and test samples
are then compared and the differential expression data given as a ratio or fold change.

Biologically relevant information can be extracted from the microarray data by further computational analysis. Genes similarly expressed may be clustered, generating expression profiles that allow a global view of the cellular transcriptional program in a given condition. Furthermore, genes consistently up- or down regulated provide a clue to the pathways involved in a cell’s response to its microenvironment (figure 1).

DNA microarrays are also valuable tools to study genomic gains and losses as well as mutations in genes. For instance, changes in the number of copies of a particular gene involved in a disease can be determined by a Comparative Genomic Hybridization approach. In this case, sample targets hybridizing to the microarray probes are fragments of genomic DNA. Differences in gene copies between control and test samples can be evaluated as described for the gene expression analysis.

The study of Single Nucleotide Polymorphisms (SNP) also uses genomic DNA as sample targets. Mutation microarray analysis, however, examine only one sample at a time. Similarly to the expression profiles, SNP patterns can be generated and associated to a particular state or disease.

Molecular signatures as a tool for diagnosis and prognosis of cancers

During pathological conditions, the cellular activity is dynamically regulated through specific changes in gene expression. Hence, microarray-generated expression data can be used as a correlate of a particular cell phenotype. Such defined and specific profiles of gene expression are referred to as molecular signatures, which could be useful to identify certain clinical conditions, class, or phase of diseases.

Indeed, this novel concept of molecular signatures is the basis for the latest generation of diagnosis tests that are being developed in this post-genomic era. It could facilitate premature and precise identification of complex multifactorial diseases such as cancer, providing the basis for a more efficient therapy. In addition, molecular signatures may provide valuable prognostic information since tumors that are morphologically similar may have different molecular characteristics associated with the clinical outcome of the disease.

The medical implications of the microarray technology are illustrated in many recent publications in the literature(2-4). A great portion of these expression profiling studies are being conducted in patients with breast cancer. Using DNA microarrays, Martin and colleagues(5) were able to identify 170 genes differentially expressed in invasive breast tumors cells. Up to 10-fold increases in the expression of a group of 12 particular genes could be detected in tumor cells disseminated in the blood of 48 patients. Similar results were obtained by other techniques such as real time-PCR, attesting to the validity of the microarray data. Using these 12 genes as molecular markers, the researchers were able to correctly identify 77% of patients with untreated invasive breast cancer, demonstrating the applicability of blood-based expression tests for diagnosis of solid tumors.

Figure 1. Global gene expression analysis using DNA chips. RNA is extracted from biological samples, labeled with fluorescent dyes and hybridized onto microarrays. Normalized fluorescence intensities are captured with proper scanners and transformed into gene expression data. Data mining software generate expression profiles that can be compared across different biological samples (key to colored squares: red = over expression; green = under expression; black = no differences in expression). Genes with similar expression patterns can be clustered and used as “molecular signatures” to discriminate groups of samples. In the example provided, genes comprising signatures A or B are over-expressed or under-expressed in tumor samples relative to normal tissue, respectively.
An independent study identified 70 out of 25,000 genes on microarray chips that correlate with diagnosis and prognosis of breast cancer\(^6\). Molecular signatures associated with either “poor” or “good” prognosis were generated based on the expression profiles of these 70 marker genes in 78 patients with sporadic breast tumors, divided into two groups: those who developed metastasis within five years of initial diagnosis and those who did not have metastasis within the same period. Prediction of disease outcome using the 70-gene prognosis classifier was validated by analysis of a second group of 19 patients, which resulted in 89\% of correct predictions. Therefore, the authors confirmed the prognostic power of their classifier which indicates that women under 55 years of age, diagnosed with breast cancer, and presenting a molecular signature associated with poor prognosis have up to 28-fold odds ratio (95\% confidence interval) to develop metastasis within five years. The same study also identified a molecular signature consisting in 100 genes differentially expressed in patients carrying BRCA1 mutations, an ancillary parameter for a precise diagnosis of hereditary breast cancer.

Recently, a more general study conducted at Harvard/MIT compared gene expression profiles of primary and metastatic adenocarcinomas of multiple types (lung, breast, prostate, colorectal, uterus, ovary), from different patients\(^7\). From this analysis, a molecular signature comprised by eight up-regulated and nine down-regulated genes was correlated to metastasis and poor clinical outcome (\(p<0.03\)). Most genes up-regulated function in protein translation, chromosome segregation, and cytoskeleton. The transcription factor RUNX1 is one of the genes repressed in metastatic tumors, consistent with its attributed tumor suppressor activity\(^8\).

Studies like these point out the utility of microarray-generated molecular signatures to help infer the clinical outcome of cancers at the time of tumor diagnosis. The same approach is being used to improve diagnosis of subtypes of tumors with different levels of malignancy. For instance, molecular signatures have been reported to distinguish low and high grade astrocytomas, the main type of tumor in the central nervous system\(^9\)-\(^10\). After proper validation in larger groups of patients by independent studies, the expression of a defined subset of genes could be monitored to help patient stratification and disease management.

**The contribution of proteomics**

Gene expression is translated into proteins of diverse roles. A cell’s function is determined by an intricate net of protein interactions that is dynamically regulated in response to internal and external signals. Proteomics is the large scale analysis of proteins which allows the analysis of post-transcriptional modifications not identifiable by DNA microarrays. The complementary information extracted from proteomics is therefore highly valuable for the understanding of molecular pathways involved in the development of complex diseases.

Using Two-Dimensional Gel Electrophoresis, a simple biochemical technique, researchers from the University of Palermo were able to identify 44 proteins that are elevated in cultured breast cancer cells\(^11\). Amino acid sequencing revealed that most of these proteins function in glycolysis. The increased energy supply is believed to favor the tumor’s ability to recruit new blood vessels and thus grow.

Global protein levels have also been analyzed in the serum of women with ovarian cancer and compared with those of healthy women. In this study, mass spectroscopy, a more sophisticated analytic tool, was used to compare profiles of about 15,200 proteins in the serum of 100 patients. It was reported a protein pattern associated with ovarian cancer that presented a positive predictive value of 94\% when used to screen an independent set of 116 serum samples of diseased and unaffected women.

Protein microarrays are being developed and applied for diagnosis purposes in the same way as DNA chips. The developing field of proteomics may have a profound impact on therapeutics by facilitating the discovery of new drug targets. One example in breast cancer is the HER-2/neu oncogene, used as prognostic factor and predictor of therapy response. Immunotherapy with humanized anti-HER-2 monoclonal antibody (Herceptin) is already being used to treat patients with encouraging results\(^12\)-\(^13\).

**Concluding remarks**

The advent of genomic technologies to study complex diseases brings a new paradigm to translational medicine. While traditional methods for the study of underlying mechanisms of diseases focus on restricted factors, genomic approaches facilitate discovering of numerous genetic markers and associated pathways in a short time frame. As the technology becomes more mature, the natural tendency is the incorporation of the generated knowledge into clinical decision-making. Likewise, the related area of pharmacogenomics is evidently changing the way pharmaceutical companies are designing the new generation of drugs. Medical genomics is an evolving field and, although the
implementation of genomic-derived technologies in medicine will still take time to occur, it is advancing fast in oncology especially in breast cancer. The promise of customized therapeutic strategies and development of affordable smart drugs lays an exciting future ahead. However, this scenario relies on constant investment in scientific research and education. An intimate connection between Medical Centers, Academic Institutions, and bio-pharmaceutical companies in the form of partnerships or collaborations will be critical for technological development, training, and assimilation of the knowledge into clinical practice.

REFERENCES


Figure 1. Axial FLAIR image: abnormal hypersignal in the polar and periventricular regions of both temporal lobes and focal ischemia in the pons.

Figure 2. Axial FLAIR image: diffuse hypersignal of periventricular and subcortical white matter in both cerebral hemispheres associated with lacunar infarct areas.