

# Molecular Profiling in the Diagnosis and Treatment of High Grade Sarcomas

*Cristina Antonescu, MD*

Department of Pathology

Memorial Sloan-Kettering Cancer Center

High density microarrays are used to measure in a comprehensive manner on a large scale (“profile”) gene expression patterns (when tumor RNA is tested) or gene copy number changes (when tumor genomic DNA is tested). Microarray-based studies are increasingly useful in addressing a wide variety of questions in sarcoma biology. During this talk, a brief description of the microarray methodology and data analysis will be provided, followed by few examples of the recent applications of gene expression in the diagnosis, classification and treatment of sarcomas.

## **I. DNA Microarray Technology**

**Expression profile** refers to the process of measuring the expression of thousands of individual genes simultaneously in a given tissue sample. The basic principle is hybridization-based.

**Microarrays**: a collection of spots on a solid surface (often a glass slide) arranged in neat rows and columns, so that the origin of each spot is known. Depending on the type, the spots can contain DNA sequences (as in gene microarrays), tissue fragments (tissue microarrays) protein (protein microarrays), etc.

### **Types of high-density microarrays:**

a) cDNA spotted arrays. Libraries of cDNA clones (20,000-40,000 most commonly expressed genes) are robotically arrayed on standard 1x3 glass microscope slides and usually produced by the investigator or by a core facility. This method requires competitive

hybridization to a mixture of reference and test mRNA, each labeled with a distinct fluorochrome (red-tumor and green-reference).

b) Affymetrix (Santa Clara, CA) oligonucleotide microarrays. Affymetrix uses similar equipment to that used for making silicon chips for computers, allowing mass production of chips at reasonable costs. Affymetrix uses masks to control synthesis of oligonucleotides on the surface of the chip. Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites. The process is repeated, activating different sets of sites and coupling different several hundred thousand squares (cells) each containing many copies of one oligonucleotide. So the result is several hundred thousand different oligos, each of them in millions of copies. It is not used for gene discovery because this approach requires knowledge of the sequence in order to choose probes for a given gene or EST. 25 base oligos are synthesized *in situ*, by light directed oligonucleotide synthesis. Each gene represented by at least 20 oligos distributed along the 5' to 3' length of the gene. Each is paired with a single base mismatch oligo. The arrays are constructed on rigid material (glass) they can be inverted and mounted in a temperature-controlled hybridization chamber. It is based on an important observation that single stranded DNA binds strongly to nitrocellulose membranes in a way that does not re-associate with each other but permits hybridization with cRNA. A fluorescently tagged nucleic acid sample injected into the chamber hybridizes to complementary oligonucleotides on the array. Laser excitation enters through the back of the glass support, fluorescence emission is collected by a lens and passes through a series of optical fibers to a sensitive detector. By scanning the laser beam (translating the array) a quantitative two-dimensional fluorescence image of hybridization intensity is quickly obtained.

**Probe redundancy:** the use of multiple oligonucleotides of different sequence designed to hybridize to different regions of the same RNA. The use of multiple independent detectors for the same molecule greatly improves signal-to-noise ratios, improves the accuracy of RNA quantitation, drastically reduces the rate of false positives and miscalls. An additional level of redundancy comes from the use of mismatch (MM) control probes that are identical with perfect match (PM) probes except for a single base difference in a central position. The MM probes act as specificity controls that allow the direct subtraction of both background and

cross-hybridization signals, and allow discrimination between “real” signals and those due to non-specific and semi-specific hybridization (hybridization of the intended RNA molecules produces more signal on PM probes than for MM probes, resulting in consistent patterns that are highly unlikely to occur by chance: the pattern recognition rules are codified in the analysis software).

c) array-based comparative genomic hybridization (array-CGH): uses microarrays with genomic DNA sequences and used to interrogate tumor DNA samples in order to provide information regarding gene copy number (amplification and deletions).

## **II. Data Analysis**

Data from a single 22,000-gene Affymetrix GeneChip occupies 100 MB of storage and require significant manipulation before interpretation. Normalization of all values to a mean of 500 is commonly used. Data cleanup is an important first step to reduce the noise, increase sensitivity, and manage artifacts of data set transformation. Filtering and statistical analyses constraints are applied to exclude those genes that did not vary significantly between comparison groups or that are not expressed at high enough levels. Data from a series of tumor samples with expression levels for thousands of genes can present a challenge for analysis. A method for data storage and retrieval in a database is essential.

**Hierarchical Clustering or Cluster Analysis**: uses standard statistical algorithms to arrange genes according to similarity in pattern of gene expression. The output is displayed graphically, conveying the clustering and the underlying expression data simultaneously. Relationship among genes is represented by a tree whose branch lengths reflects the degree of similarity between genes (dendrogram). This method recognizes groups of genes that are co-expressed, providing a new insight of their possible function. It also identifies samples whose gene expression patterns correlate. The basic idea is joining together points into clusters.

Appropriate selection of analysis tools will depend on the questions to be addressed. Certain key questions predominate cancer-related microarray research: “Can two types of cancer be discriminated? What genes discriminate them more clearly? Are there genes that

discriminate tumors from normals? Are there correlations between expression profiles and other molecular or pathological properties of the tumor? Are there correlations between expression profiles and clinical outcome and response to therapy? With what degree of certitude these results are not due to chance alone? Are there subsets within tumors of the same apparent class? Are these pathways relevant to the tumor phenotype or as potential targets for therapy?" These important questions and the need to develop the mathematical tools to address them have attracted the attention of computer scientists, engineer, and biostatisticians. Numerous computational approaches have been developed.

Microarray analysis whether supervised or unsupervised ultimately generates lists of genes that discriminates among samples. Making sense of these gene lists presents a significant challenge. Gene names can be misleading and the majority of genes are linked to little or no functional information. Most of the interpretations that arise from exploring microarray data should be considered hypotheses, rather than conclusions. Additional forms of experimentation is most likely necessary to establish a conclusive connection between a gene and the tumor in which is expressed.

**Validation the microarray data:** How reliable is the data? Tissue microarrays for in situ mRNA hybridization or immunohistochemistry provide the possibility of confirmatory studies on large number of samples.

### **III. Applications of the DNA microarray technology to cancer diagnosis and prognosis**

Cancer is a disease of disturbed genome function. Irrespective of whether this is the result of a point mutation, deletion, translocation, gene amplification, or methylation, the malignant phenotype is mediated by a characteristic pattern of gene expression. Identifying these genes whose expression differs between normal tissues and tumors and among tumor types is the focus of today's research. Investigators in the field initially performed studies in which tumors of different morphology or different primary sites were shown to have clearly distinguishable patterns of gene expression. This type of diagnostic classification by analysis of expression profile is called "class prediction". These proof-of-principle studies served to validate cDNA microarray technology and the researchers then shifted their focus to

identification of molecularly defined tumor entities that were inapparent by conventional pathologic analysis (“class discovery”). New unsuspected biological subsets were detected among cutaneous melanomas, breast carcinomas, and pediatric acute lymphoblastic leukemias. In other cases, there has been “class rediscovery” or “class confirmation”.

Variability in the results, more notable in prognostic studies than in class prediction, have already appeared in the literature and is most likely related to inter-laboratory and inter-platform reproducibility. For example, two large breast cancer expression profiling studies differed considerably in the results of unsupervised clustering: one identified 3 major subsets (ER-positive luminal cell type, basal cell type, and ERBB2-amplified type), while the other detected only 2 major subgroups according to ER status and lymphocytic infiltration. Moreover, in the latter study, well-established clinical markers, such as ERBB2 and ER were not found within a list of 70 genes linked to outcome.

#### **IV. Molecular Profiling in the Diagnosis and Treatment of High Grade Sarcomas**

Studies involving gene expression technology have been mainly applied as a source of diagnostic markers for sarcoma diagnosis and to their role in clarifying sarcoma classification. A general observation of microarray-based expression profiling studies of sarcomas is that translocation-associated sarcomas are robustly clustered by expression profiling using cDNA microarrays, whereas so-called complex karyotype sarcomas tend to be less tightly clustered. Thus, complex karyotype sarcomas that often show different gains and losses from case to case are also likely to show more variability in gene expression patterns, leading to less robust unsupervised clustering of the expression profiles. Indeed, this tumor group may be better studied by profiling of gene copy number changes using array-based CGH.

Our study on 51 soft tissue sarcoma specimens using hierarchical cluster analysis demonstrated distinct clusters for GIST, synovial sarcoma, clear cell sarcoma and round cell liposarcoma. Several fibrosarcoma tumors fall in close proximity to synovial sarcoma. Pleomorphic sarcomas exhibited overall poor correlation and consistency by boot strap

analysis. However, within this group certain prominent clusters were observed for both malignant fibrous histiocytoma (MFH) and leiomyosarcoma.

One example of “class confirmation” study in sarcomas includes the comparison of GIST versus leiomyosarcoma. Previous RNA expression profiling studies of different soft tissue sarcomas indicated that GIST expression profiles were distinct and quite homogeneous in part due to the unique derivation of GIST from ICC. GISTs are characterized by a distinctive transcriptional signature, as a result of overexpression of *KIT*, *PRKCθ*, *DOG1*, which can be applied in tumor diagnosis, even when compared with their closest pathologic mimic, leiomyosarcomas. Likewise, several studies have used microarrays to identify genes differentially expressed between *PAX-FKHR* fusion-positive alveolar and *PAX-FKHR* fusion-negative embryonal rhabdomyosarcoma.

Molecular profiling was also applied to clarify the conflicting data on the relationship of clear cell sarcoma (CCS) to cutaneous melanoma. While the two are clearly genetically distinct, as CCS lack the *BRAF* mutations commonly seen in melanomas, whereas melanomas do not contain the *EWS-ATF1* fusion, by expression analysis CCS share a melanocytic gene signature with melanomas.

In the study by Baird et al, the unsupervised hierarchical analysis on MFH cases alone identified 2 distinct groups: one carrying a muscle profile (myosin X, sarcoglycan β, tenascin C) and the second group revealed an immune regulatory gene profile (HEM1, MX1, DAP10). Interestingly, the storiform-pleomorphic and the myxoid subtypes equally populate both groups of tumors. However the distinction between MFH with myogenic differentiation versus inflammatory characteristics might have clinical relevance. Myogenic differentiation as reflected by IHC in MFH or pleomorphic sarcoma, NOS, has previously shown to have an adverse clinical outcome.

New IHC markers emerging from microarray studies include DOG1 in GIST, TLE1 in synovial sarcoma, AP2-β in alveolar rhabdomyosarcoma, and Apo D in dermatofibrosarcoma protuberans (DFSP).

Still in incipient phases, microarrays studies have been used to predict outcome or therapy response. Poor outcome Ewing sarcoma has been linked with alterations in cell cycle regulatory genes. In a recent study, West et al suggested that expression signatures of either fibromatosis or solitary fibrous tumor can be identified in the stromal component of most breast carcinoma and can be correlated with clinical outcome. The fibromatosis signature is associated with good outcome breast tumors, while solitary fibrous tumor profile is found in poor clinical outcome.

Three different studies have been assessed the correlation of transcriptional profile of pre-treatment osteosarcoma biopsy and response to chemotherapy. Disappointingly, limited overlap has been noted among the results of these studies, which can be explained as a consequence of different platforms or bioinformatics approaches used, differences in patient population or chemotherapy regimens applied, etc.

The ability of mining candidate genes for targeted therapies through microarray analysis has been validated with the GIST paradigm, where the overexpressed KIT or PDGFRA provided the biologic basis for imatinib mesylate therapy. Other examples include dermatofibrosarcoma protuberans, where through an autocrine/paracrine mechanism, the overexpressed PDGFB up-regulates its receptor, PDGFRB, therefore providing another susceptible target to selective tyrosine kinase inhibition. However, targeted therapy appears to be most successful in sarcomas with a distinct underlying molecular biology, and has been failed in sarcomas with non-recurrent complex karyotype. Since drugs target proteins rather than genes, protein level validation is critical and can be performed on the tissue microarray by using IHC and ISH techniques.

In summary, microarray studies applied so far in sarcomas support their classification into genetically simple and genetically complex categories and have provided useful diagnostic markers, as well as insight for novel and targeted therapeutic approaches. However, there is a considerable vacuum between the practical world of hospital-based

molecular diagnostic laboratories and some of the predictions prompted by high-throughput genomics work. Hopeful statements, such as one made in 1999 that “doctors will be offering gene expression profiles to some patients in the next 3 years”, failed to consider the many issues in moving complex assays from research laboratories to clinical laboratories. Because of the regulatory, billing, quality control, and test validation concerns, combined with limited resources, academic molecular diagnostic laboratories are extremely selective in their test menus. Furthermore, it is not clear yet that the present cost of microarray, large-scale expression profiling for “class prediction” is more cost-effective than established diagnostic approaches, i.e., histopathology supplemented in selected cases by IHC, cytogenetics, or molecular assays.

In short term however the impact of tumor expression profiling is in the identification of new diagnostic and prognostic markers, which can be studied individually by more conventional techniques. The availability of TMA accelerates the validation of these new IHC assays. The next step will require efforts in academic and commercial laboratories to generate new antibodies for the products of differentially expressed genes without currently available antibodies.

## REFERENCES

1. Ladanyi M, Gerald W. Expression profiling of human tumors. *Humana Press*. 2003
2. Ladanyi M, et al. Expression profiling of human tumors: the end of Surgical Pathology? *J Molec Diagn* 2001; 3:92-7.
3. Sorlie T et al. Gene expression patterns of breast carcinomas distinguish subclasses with clinical application. *Proc Natl Acad Sci U S A* 2001; 98:10869-74.
4. Perou CM et al. Molecular portraits of human breast tumors. *Nature* 2000; 406:747-52.
5. Khan J et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001; 7:673-9.
6. Nielsen TO, et al. Molecular characterization of soft tissue tumors: a gene expression study. *Lancet* 2002; 359:1301-7.
7. Linn SC et al. Gene expression patterns and gene copy number changes in dermatofibrosarcoma protuberans. *Am J Pathol* 2003; 163:2383-95.

8. Segal NH, et al. Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling. *J Clin Oncol* 2003; 21:1775-81.
9. Segal NH, et al. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *Am J Pathol* 2003; 163: 691-700.
10. Antonescu CR, et al. Gene expression in gastrointestinal stromal tumors is distinguished by *KIT* genotype and anatomic site. *Clin Cancer Res* 2004; 10:3282-90.
11. Bittner et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000; 406:536-40.
12. Allander SV, et al. Expression profiling of synovial sarcoma by cDNA microarrays: association of ERBB2, IGFBP2, and ELF3 with epithelial differentiation. *Am J Pathol* 2002; 161:1587-95.
13. Baird K, et al. Gene expression profiling of human sarcomas: insights into sarcoma biology. *Cancer Res* 2005; 65:9226-35.
14. Fritz B, et al. Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res* 2002; 62:2993-8.
15. Ohguri T, et al. Cytogenetic analysis of myxoid liposarcoma and myxofibrosarcoma by array-based comparative genomic hybridization. *J Clin Pathol* 2006; 59:978-83.
16. Heidenblad M, et al. Genomic profiling of bone and soft tissue tumors with supernumerary ring chromosomes using tiling resolution bacterial artificial chromosome microarrays. *Oncogene* 2006; 25:7106-16.
17. West RB, et al. Apo D in soft tissue tumors: a novel marker for dermatofibrosarcoma protuberans. *Am J Surg Pathol* 2004; 28: 1063-9.
18. West RB, et al. The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of *KIT* or *PDGFRA* mutation status. *Am J Pathol* 2004; 165:107-13.
19. West RB, et al. Determination of stromal signatures in breast carcinoma. *PLoS Biol* 2005; 3:e187.
20. West RB et al. A landscape effect in tenosynovial giant-cell tumor from activation of CSF1 expression by a translocation in a minority of tumor cells. *Proc Natl Acad Sci U S A* 2006;103:690-5.

21. Nielsen TO, et al. Tissue microarray validation of epidermal growth factor receptor and SALL2 in synovial sarcoma with comparison to tumors of similar histology. *Am J Pathol* 2003; 163:1449-56.
22. Nielsen TO. Microarray analysis of sarcomas. *Adv Anat Pathol* 2006; 13:166-73.
23. Ohali A, et al. *Oncogene* 2004; 23: 8997-9006.