CANCER GENOMICS & PROTEOMICS *3*: 311-316 (2006)

*Review*

# Basics of Diagnostic DNA Microarray Technology. Case Study: Hepatocellular Carcinoma

ANAS EL-ANEED and JOSEPH BANOUB

*Memorial University of Newfoundland, Biochemistry Department, St. John's, NL, A1C 5S7, Canada*

**Abstract.** *The DNA microarray technique is capable of identifying the expression of thousands of genes simultaneously. This technology has become instrumental in cancer research for diagnosis and biomarker discovery purposes. This mini-review will introduce DNA microarrays at the basic technical level and will focus on high risk Hepatocellular Carcinoma (HCC) patients, namely hepatitis B and C infected individuals.*

Of the 30,000-40,000 protein coding genes identified in the human genome project (1, 2), only a small number are "turned on" in a specific cell type. Changes at the genome level are the underlying cause of cancer development with some genes being "turned on" and others being "turned off"; in addition to multiple mutational alterations, such as those associated with the tumor suppressor gene *p53* (3). Monitoring these multiple genetic changes will assist in identifying many genes which are involved in the process that transforms normal into cancer cells and can identify potential diagnostic and prognostic cancer indicators.

In addition, genomic profiling is well-suited to the development of effective therapeutic strategies so that genes can be identified based on drug responsiveness and/or side effects (4). DNA microarrays can scan thousands of genes at the same time and identify specific "cancer" genes. It could be well argued that DNA microarrays are a complicated form of Northern Blotting. This is applicable at the tissue level rather than with serum; however, markers, once identified, can be specifically targeted within the blood as cancer lesions secrete many compounds into the circulatory system. This fact is particularly true in the case of liver diseases. Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related death worldwide (5) and late diagnosis is the main contributor to the high fatality rate associated with this disease. Currently, α-fetoprotein (AFP) is the only biomarker used to diagnose HCC. AFP is not always elevated in early stages of HCC (6) and can be highly elevated in non-cancerous cirrhotic liver diseases such as hepatitis B- and C-infected individuals (7). The following sections will discuss diagnostic DNA arrays with relevance to HCC research with respect to at-risk patients infected with hepatitis B (HBV) and/or C (HCV) viruses.

Traditional methods, such as Northern and Southern Blotting can monitor one gene at a time; DNA microarrays, however, can scan thousands of genes simultaneously and require a small quantity of sample. Similar to the Northern Blot technique, DNA microarrays rely on the specific base-pairing association of a DNA or RNA segment to its complementary version (*i.e.,* hybridization). The subject DNA or RNA is labeled (usually with a proper fluorescent dye) such that it can be detected once it is hybridized on the target surface which immobilizes known DNA segments (the target surface can be a glass or nylon membrane). These known DNA segments, on the support surface, are defined as probes. Thousands of different probes can be arrayed on the same surface and hence the name: DNA microarrays. These probes are usually either spotted cDNA (usually 300 bases in length) or *in situ* synthetic (or spotted) oligonucleotide segments (short 20-30 mer, or long 60-80 mer) (8, 9). Depending on the nature of the surface and the probes, each microarray can perform differently (10). In fact, different microarrays can complement each other and can be used in tandem to optimize the outcomes (11). It should be mentioned that some investigators use the term "probes" for the labeled nucleic acid materials within the sample rather than to the ones arrayed on the surface.

*Correspondence to:* Anas El-Aneed, Memorial University of Newfoundland, Biochemistry Department, St. John's, NL, A1C 5S7, Canada. Tel: (709) 737-4331, Fax: (709) 737-3395, e-mail: anas@mun.ca

*Key Words:* DNA microarray technology, hepatocellular carcinoma, hepatitis B, hepatitis C, review.

## Experimental and Analytical Procedures of Diagnostic DNA Microarrays

DNA microarrays were first introduced in 1995 (12) and since then thousands of studies utilizing this technique have been published. Cancer is by far the most investigated disease using DNA microarrays. A keyword search using "cancer" and "DNA microarray" on the PubMed database will result in over 1,100 hits in 2005 alone. By monitoring the expression of 6,000 genes among liver tissues of HCV-infected HCC patients, a significant genetic difference was observed among patients bearing the mutant *p53* gene (missense mutations in exons 5-9) compared to those bearing the wild type version of this gene (13). This kind of study is categorized under the supervised analysis of the DNA microarray data; that is to say, to identify groups of genes that are associated with a specific condition (in the above example, it is the mutation in the p53 gene). In supervised analysis of microarray data, the investigator is aware of the function of the evaluated genes and the analysis is commonly used to identify groups of genes that are related to pathological, biological or biophysical conditions. Based on gene expression profiles, prediction methods can be built and evaluated.

On the other hand, in unsupervised analysis, gene expression profiles are generated, without any pre-assumption or pre-knowledge of genes' functions. It relies only on gene expression profiles to classify (or group) cancer lesions. In a study that utilized unsupervised analysis, specific genes that are expressed in normal tissues including the liver were identified and their patterns within the cancerous lesions (based on published work) were evaluated (14). It was suggested that a set of liver-specific genes were well-suited for the classification of the degree of tumor differentiation (poor, moderate and well) (14).

The experimental steps leading to either supervised or unsupervised analysis are similar and are presented in Figure 1. In summary, total RNA is extracted from the cells (cancer *vs.* normal). cDNA based on the poly A nature of mRNA is synthesized by reverse transcription. Amplification of either cDNA (usually for cDNA microarray) or cRNA (for oligomicroarrays) is performed. During amplification, nucleic acids are labeled with two different fluorescent dyes, such as cyanine 3 (Cy3) for cancer cells and cyanine 5 (Cy5) for normal cells. Labeled cDNA or cRNA is applied simultaneously on the DNA microarray plate. Nucleic acids in the sample will anneal to their complementary DNA strands attached to the array slide. The color and intensity of the signal are correlated with the original concentration of mRNA in the cell. Genes which are over-expressed in the cancer lesion will be colored green (Cy3 is dominant), while the under-expressed ones are colored red (Cy5 is dominant). The yellow color will mean no change and no color means that the gene is absent in both cell groups. The final stage includes clustering (grouping) of similarly expressed genes so that genes which are "turned on/turned off" can be identified. A computer-based color scale can be established to determine the degree of repression or induction. Clustered genes can be visualized *via* color-coded matrices as shown in Figure 2.

Computational hierarchical clustering is widely used when dealing with microarray data (15) and it is common to perform two-way hierarchical clustering, for both the genes and the tested subjects (see Figure 2). In simple terms, the two genes which are very similar are paired together and the computer algorithm will merge those two genes such that they are considered one (pseudo gene) and then the genes which are next, in terms of similarity, are also joined together (15). This process is repeated until a giant cluster is formed. Figure 3 represents a theoretical cluster of 7 genes (designated alphabetically: A, B, C....etc). A correlation scale can be established with 1, for example, being highly correlated and –1 being not correlated (one increases while the other decreases). It is common to perform two-way clustering, as both the genes and the tested subjected are clustered (see Figure 2).

## Liver Cancer and Diagnostic DNA Microarrays

There have been many published work that utilize gene expression profiles and clustering to pin down specific genes, either as down-regulated or up-regulated, within the liver of HCC patients. Using a 12,800 gene microarray showed that 1.6% of those genes are up-regulated within HCC patients while 12.7% showed down-regulation (16). Growth rate, vascular invasion and *p53* expression levels can actually alter the gene expression pattern of HCC (17). A more recent investigation identified 703 genes, differentially expressed in HCC tissues compared to normal tissues (18). Among those 703 genes, there were 9 genes which encode for secretory proteins including the most common biomarker of HCC, AFP protein (18) Identification of secreted proteins has special significance since those candidates can be evaluated as biomarkers within the circulation. PGCP, a glutamate carboxypeptidase, and two secreted phospholipases A2 (PLA2G13 and PLA2G7) were considered as attractive secreted proteins, unique for HCV-infected HCC patients (19). Similarly, HSP70, a heat shock protein, was proposed as a molecular marker for early detection within HCC patients that bear either HCV or HBV (20).

Hepatocarcinogensis caused by HBV have different genetic pathways than that associated with HCV. It was shown, for example, in a comparative study, that HBV associated HCC resulted in the elevation of the gene expression of genes relating to metastasis, transcription and signal transduction (21). Similar set of genes were identified in a hepatocyte cell line (X18) which is conditioned to

overexpress the HBV-carcinogenic x protein (22). On the other hand, the genetic signature of HCV-linked HCC showed up-regulation of the genes responsible for detoxification and immune response (21). This discrepancy in the genetic profiles of HCV- and HBV-infected HCC patients can be attributed to the virological differences between these two infectious viruses. HBV is a DNA virus whose genetic material reaches the host cell nucleus. It usually functions without integrating into the host genome. In addition, HBV is known to encode for a protein known as the x protein (HBx protein) which is proposed to play a crucial role in HBV-mediated carcinogenesis. HCV, however, is an RNA virus that remains in the cytoplasm during its life cycle and is believed to contribute to HCC development mainly through chronic inflammation.

## Challenges of DNA Microarray Analysis

One of the challenges that researchers face when analyzing DNA data is the complex nature of the genes identified which requires deep knowledge, not restricted to a handful gene groups or specific pathways. Smith *et al.*, for example, proposed fifty potential marker genes, specific for HCV-infected liver cancer patients (19). Unfortunately, this study which used unsupervised analysis did not identify genes that could distinguish early HCC against advanced HCC. Genes were, however, further clustered based on their roles, such as cell-proliferation and cell cycle control, revealing more details of the molecular mechanism of hepatocarcinogenesis. Oxidative stress-related genes were linked to HCC development and as a novel finding, the study showed that a serine/threonine kinase (STK15) encoding gene was up-regulated in HCV-infected HCC pateints (19). STK15 is a centrosome-coupled kinase, associated with segregation abnormalities and aneuploidy in many cancerous lesions. Another interesting study indicated that the degree of similarities of gene expression patterns of HCV-HCC and HCV late cirrhosis were stronger than HCV-HCC and healthy individuals (23). This finding supports the theory that HCV contributes to HCC development mainly through its inflammatory-causing effect, namely liver cirrhosis.

Moreover, DNA arrays will probably identify genes which have unknown roles. Okabe *et al.* identified 165 genes as up-regulated within HCC patients and 170 as down-regulated; 69 genes from the former and 75 from the latter had unknown functions (24). DNA arrays can also be used to study the metastatic nature of a specific cancer type. Hepatitis-B positive HCC patients were categorized based on the metastatic nature of their cancer lesion. It was suggested that HCC with metastatic potential is evolutionary different than non-metastatic liver cancer (25). An 153 gene model was developed and successfully predicted 85% of the tested metastatic HCC patients (25). Such a model can serve as predictive tool for further therapies.

## General Concerns

DNA microarrays can monitor large number of genes and due to the complexity of such work, reproducibility seems to be the main concern within genomic research that focuses on diagnosis. Intra-individual changes can also affect the outcomes such as age, gender, smoking, nutrition states. Reproducibility concerns are common within different DNA platforms (10). The differences will probably increase when using home-made gene arrays, where quality control is a major concern. Applying rigid experimental design, however, showed similarities up to 90% within various DNA microarray platforms (26). More encouraging, microarrays are currently used in routine clinical prognostic screening in The Netherlands for breast cancer (27). Cross-reactivity can also be encountered within DNA arrays, and is termed "cross-talk". To increase validity of the DNA arrays experiments, analyses should be conducted by different investigators such that the samples are blinded during analysis. It is strongly recommended that after any DNA array conclusion, confirmation studies conducted with different samples in different laboratories should be performed. This is the only way to strongly validate any conclusion drawn from DNA array experiments.

## Conclusion

Despite the number of studies dealing with the molecular profiling of HCC with respect to HBV or HCV infections, scientific efforts should continue on the genetic front to enrich our understanding of the role of hepatitis in HCC development and to identify new biomarkers. To our knowledge, for example, there has been no large scale attempt to investigate co-infected HBV and HCV patients who develop primary liver cancer. Co-infection with those two pathogenic viruses can significantly increase the risk for HCC development due to the synergistic effects of the two infections (28, 29) and will most probably result in some changes at the molecular level. Such an investigation will increase the knowledge surrounding liver cancer and its relation to hepatitis infections.

In summary, high-throughput genomic techniques have revolutionized cancer research and moved it to a stage where many genes can be studied simultaneously. Valuable information regarding cancer development, therapy and diagnosis can now be obtained. For example, Welsh *et al.* (30) studied the genome profiles of cancer lesions with various anatomic sources including HCC and focused their study on genes which encode secreted proteins. The study suggested that MIC-1 gene, which plays a role in apoptosis and growth inhibition, is elevated within metastatic prostate, colorectal and breast cancer tissues (31). ELISA has also shown that this secreted protein is elevated in the serum suggesting its potential as a biomarker for these three

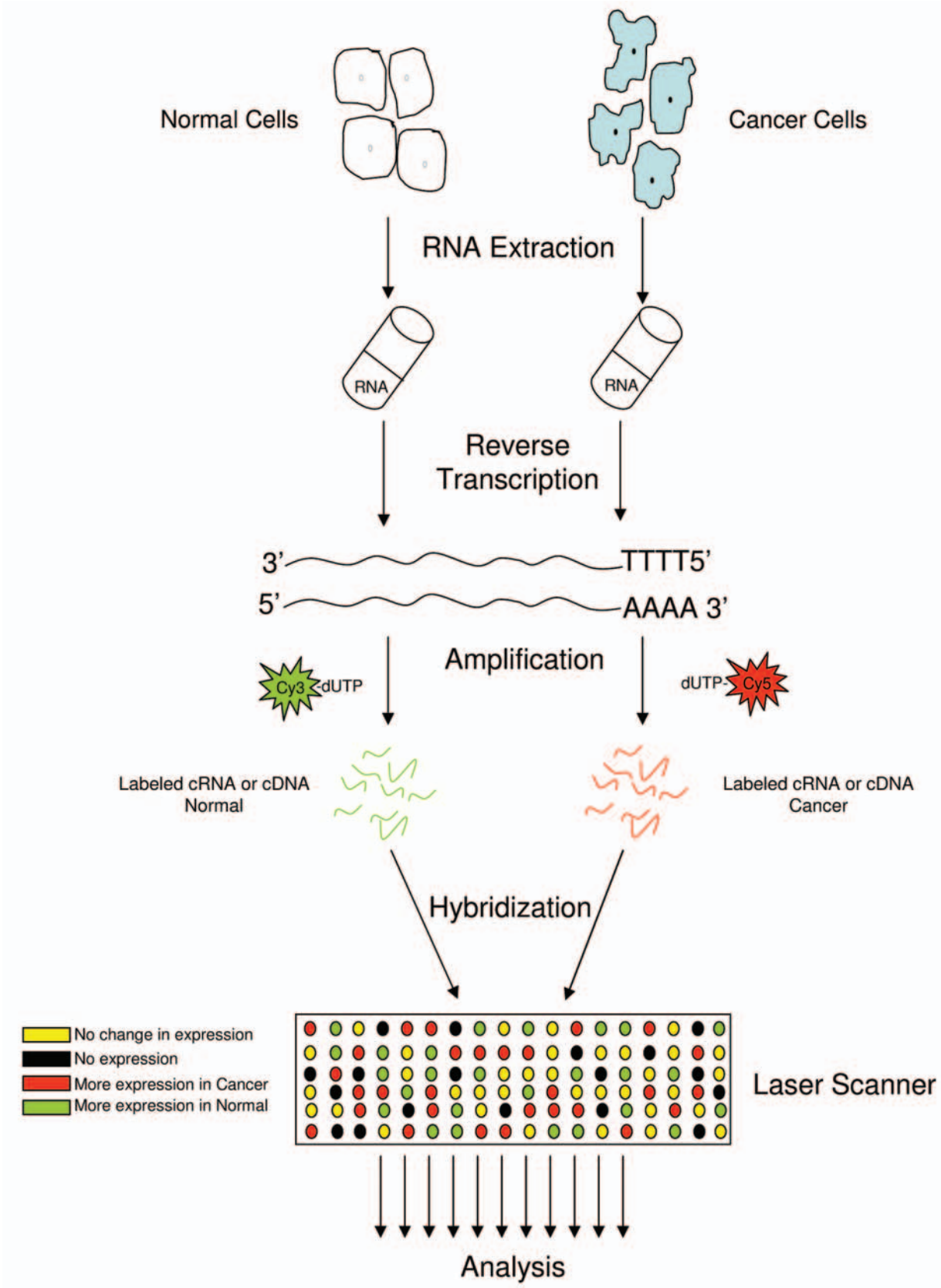


Figure 1. *Schematic representation of the experimental steps leading to expression DNA microarray analysis.*

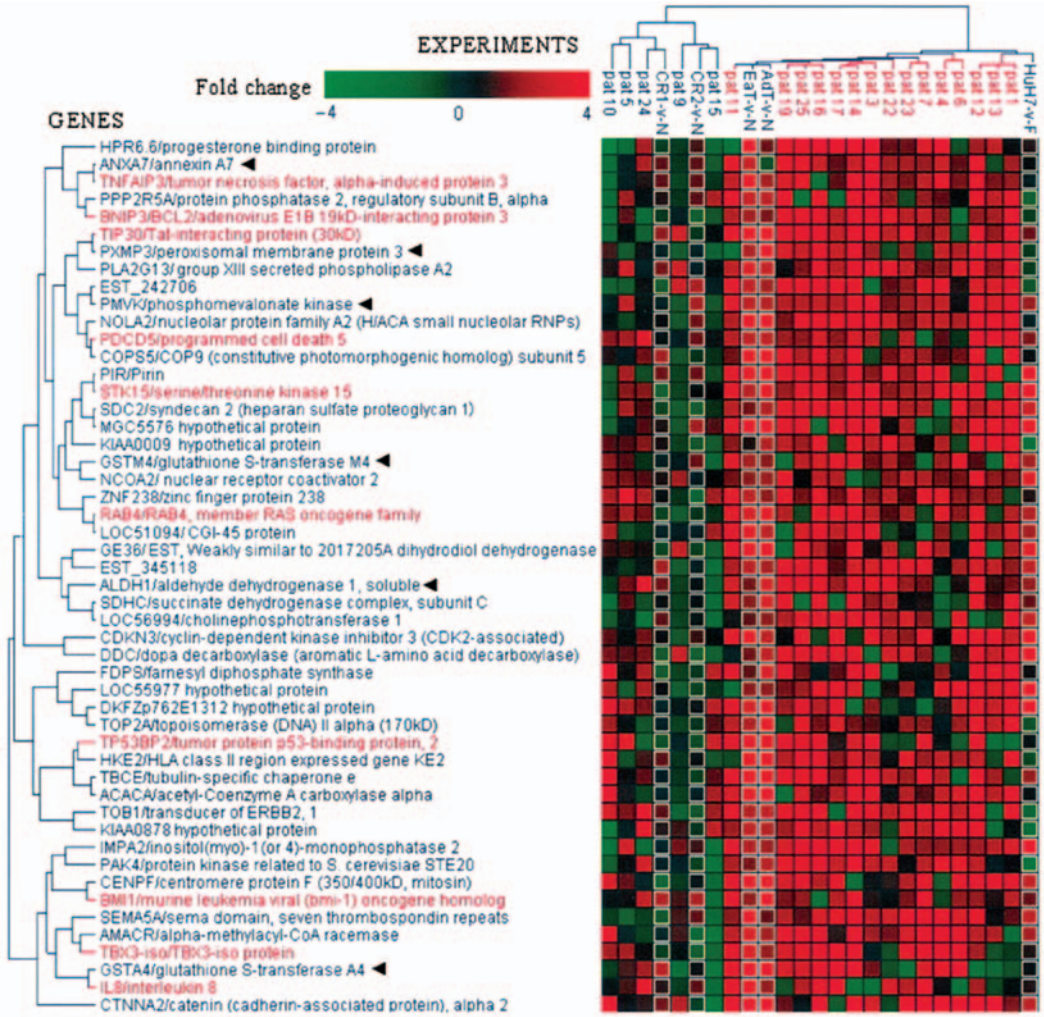


Figure 2. *Representation of computer-based color scale, reprinted with authorization from reference 19. The figure shows two-way hierarchical clustering results for 50 genes which were identified as potential HCC markers. Intensity of the green color indicates the degree or repression while the red color intensity is related to the degree of over-expression.*

different cancer lesions (30). For optimum outcomes, genomic approaches should work hand-in-hand with other methodologies such as proteomic approaches (32). The aim is to identify new molecular targets that may serve as biomarker indicators for early detection of cancer.

## Acknowledgements

The authors acknowledge the contribution of Mr. Howard Hooder with regard to proof reading and English corrections.

## References

1 Lander ES *et al*: Initial sequencing and analysis of the human genome. Nature *409*: 860-921, 2001.
2 Venter JC *et al*: The sequence of the human genome. Science *291*: 1304-1351, 2001.

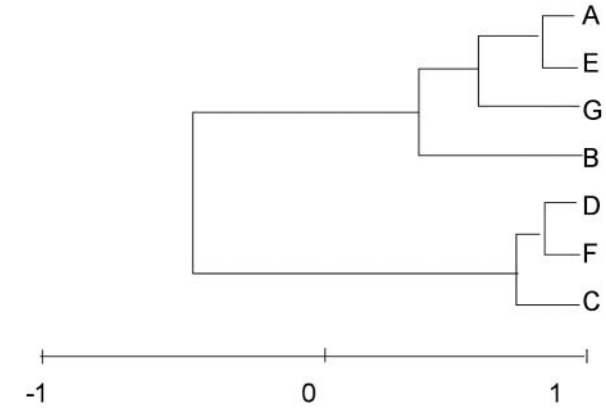


Figure 3. *Representation of hierarchical clustering of 7 genes designated A, B, C, D, E, F, and G. Two separate clusters are observed (C, F, and D) and (A, E, G, and B).*

3 Greenblatt MS, Bennett WP, Hollstein M and Harris CC: Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res *54*: 4855-4878, 1994.
4 Chin KV and Kong AN: Application of DNA microarrays in pharmacogenomics and toxicogenomics. Pharm Res *19*: 1773-1778, 2002.
5 Howe HL *et al*: Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst *93*: 824-842, 2001.
6 Song BC *et al*: Transforming growth factor-beta1 as a useful serologic marker of small hepatocellular carcinoma. Cancer *94*: 175-180, 2002.
7 Staley CA: *In*: Hepatology, A Textbook of Liver Disease, Vol. 2, Edn. 4th. Zakim D and Boyer TD (eds.). Elsevier Science, USA, Philadelphia, pp. 1371-1382, 2002.
8 Bucca G *et al*: Gene expression profiling of human cancers. Ann NY Acad Sci *1028*: 28-37, 2004.
9 Brentani RR *et al*: Gene expression arrays in cancer research: methods and applications. Crit Rev Oncol Hematol *54*: 95-105, 2005.
10 Marshall E: Getting the noise out of gene arrays. Science *306*: 630-631, 2004.
11 Jurata LW *et al*: Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures. J Neurosci Methods *138*: 173-188, 2004.
12 Schena M, Shalon D, Davis RW and Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*: 467-470, 1995.
13 Okada T *et al*: Gene expression profile linked to p53 status in hepatitis C virus-related hepatocellular carcinoma. FEBS Lett *555*: 583-590, 2003.
14 Ge X *et al*: Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics *86*: 127-141, 2005.
15 Eisen MB, Spellman PT, Brown PO and Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA *95*: 14863-14868, 1998.
16 Li Y *et al*: Discovery and analysis of hepatocellular carcinoma genes using cDNA microarrays. J Cancer Res Clin Oncol *128*: 369-379, 2002.
17 Chen X *et al*: Gene expression patterns in human liver cancers. Mol Biol Cell *13*: 1929-1939, 2002.
18 Patil MA *et al*: An integrated data analysis approach to characterize genes highly expressed in hepatocellular carcinoma. Oncogene *24*: 3737-3747, 2005.
19 Smith MW *et al*: Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. Cancer Res *63*: 859-864, 2003.
20 Chuma M *et al*: Expression profiling in multistage hepatocarcinogenesis: identification of HSP70 as a molecular marker of early hepatocellular carcinoma. Hepatology *37*: 198-207, 2003.
21 Iizuka N *et al*: Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. Cancer Res *62*: 3939-3944, 2002.
22 Ng RK *et al*: cDNA microarray analysis of early gene expression profiles associated with hepatitis B virus X protein-mediated hepatocarcinogenesis. Biochem Biophys Res Commun *322*: 827-835, 2004.
23 Mas VR *et al*: Hepatocellular carcinoma in HCV-infected patients awaiting liver transplantation: genes involved in tumor progression. Liver Transpl *10*: 607-620, 2004.
24 Okabe H *et al*: Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res *61*: 2129-2137, 2001.
25 Ye QH *et al*: Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med *9*: 416-423, 2003.
26 Horne J: On the bright side of microarrays. Nature Cell Biology *7*: 550, 2005.
27 Schubert CM: Microarray to be used as routine clinical screen. Nat Med *9*: 9, 2003.
28 Carter H, Robinson G, Hanlon C, Hailwood C and Massarotto A: Prevalence of hepatitis B and C infection in a methadone clinic population: implications for hepatitis B vaccination. N Z Med J *114*: 324-326, 2001.
29 Donato F, Boffetta P and Puoti M: A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. Int J Cancer *75*: 347-354, 1998.
30 Welsh JB *et al*: Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. Proc Natl Acad Sci USA *100*: 3410-3415, 2003.
31 Albertoni M *et al*: Anoxia induces macrophage inhibitory cytokine-1 (MIC-1) in glioblastoma cells independently of p53 and HIF-1. Oncogene *21*: 4212-4219, 2002.
32 El Aneed A and Banoub J: Proteomics and the diagnosis of hepatocellular carcinoma: focus on high risk hepatitis B and C patients. Anticancer Res *26*: 3293-3300, 2006.

*Received July 3, 2006*
*Accepted August 10, 2006*