The majority of DNA-microarray based gene expression profiling studies on human hepatocellular carcinoma (HCC) has focused on identifying genes associated with clinicopathological features of HCC patients. Although notable success has been achieved, this approach still faces significant challenges due to the heterogeneous nature of HCC (and other cancers) as well as the many confounding factors embedded in gene expression profile data. However, these limitations are being overcome by improved bioinformatics and sophisticated analyses. Also, application of cross comparison of multiple gene expression data sets from human tumors and animal models are facilitating the identification of critical regulatory modules in the expression profiles. The success of this new experimental approach, comparative functional genomics, suggests that integration of independent data sets will enhance our ability to identify key regulatory elements in tumor development. Furthermore, integrating gene expression profiles with data from DNA sequence information in promoters, array-based CGH, and expression of non-coding genes (i.e., microRNAs) will further increase the reliability and significance of the biological and clinical inferences drawn from the data. The pace of current progress in the cancer profiling field, combined with the advances in high-throughput technologies in genomics and proteomics, as well as in bioinformatics, promises to yield unprecedented biological insights from the integrative (or systems) analysis of the combined cancer genomics database. The predicted beneficial impact of this “new integrative biology” on diagnosis, treatment and prevention of liver cancer and indeed cancer in general is enormous. (HEPATOLOGY 2006;43:S145-S150.)

Hepatocellular carcinoma (HCC) is the fifth most common cancer world-wide and one of the most deadly cancers with approximately 600,000 yearly deaths. However, during the past 25 years striking advances have been made in our understanding of hepatocellular carcinoma (HCC). The causes of more than 85% of HCC cases are known (hepatitis viruses B and C, aflatoxin B1, ethanol), but new risk factors continue to emerge (type II diabetes and obesity). Therapies directed toward the eradication and/or prevention of viral infection have already reduced the incidence of HCC developing in some populations, although the overall incidence of HCC continues to rise, especially in Western Europe and the United States.

Much is also known about the morphologies of cells and tissues that precede and accompany the development of HCC in humans, allowing earlier diagnosis in some instances. This insight has come largely from the detailed analysis of HCC development in experimental animals and comparison of the results with HCC in humans. A variety of genomic and molecular alterations have been detected in fully developed HCC and references cited therein) and to a lesser extent in morphologically defined preneoplastic precursor lesions. However, this database has been obtained largely in a fragmented and uncoordinated manner, typically in studies in which aberrations in single chromosome loci/genes (the latter frequently the most popular “oncogene” or “tumor suppressor” gene of the day) have been analyzed in a group of HCC from humans and experimental animals. Although in aggregate these data are useful, they have not led to a coherent understanding of the mechanisms of HCC development, or to the identification of critical genomic and/or molecular aberrations that improve the precision of diagnosis or serve as sites for therapeutic interventions. This situation appears to reflect the limited number of genes that have been examined simultaneously, a strategy that does not
allow the assessment of complex multigenic molecular pathways.

Recent studies in both experimental animals and humans have begun to generate more comprehensive data that may ultimately define molecular regulatory pathways involved in HCC development. These new studies include analysis of complex patterns of genomic aberrations, including locus deletions and changes in gene expression, by microarray-based technologies that allow the study of multiple genes simultaneously.\(^9\) Data from studies employing gene expression profiling have already enabled the identification of complex patterns of gene expression associated with differential longevity among human patients with HCC.\(^12\) Of perhaps more importance to the eventual identification of molecular pathways that drive HCC development are comparative studies that seek to identify conserved gene expression patterns common to HCC of both humans and of experimental animals. These studies compare gene expression profiles in human HCC with patterns from HCC that have been produced in experimental animals by transgenic modification of specific target genes, a technique that can disclose otherwise occult patterns of gene expression in the human tumors. The remainder of this review focuses on the discussion of integrative and comparative functional genomics in the study of HCC development.

### Integrative Functional Genomics of HCC

The notion that a disease condition may represent disruption of the normal network structures of an organ system through either genetic perturbations and/or deleterious environmental agents is emerging as the central tenet of the systematic approaches currently adopted to understand complex diseases.\(^13\) Cancer is a complex disease that emerges from multiple spontaneous and/or inherited mutations that induce dramatic changes in expression patterns of genes and proteins that function in networks controlling critical cellular events.\(^5\) Characterization of the complex molecular networks that drive HCC development has been facilitated by the emergence of novel high-throughput technologies, including DNA microarrays that can simultaneously detect the expression levels of thousands of genes. This new technology has been used successfully to predict clinical outcome and survival as well as to classify different types of cancer.\(^14\)

Here we will briefly illustrate how gene expression profiling of HCC can provide a better understanding of the molecular pathogenesis of the disease and then focus on the application of integrative functional genomic approach utilizing comparative genomics to enrich the human gene expression database on HCC. Finally we will discuss possible future clinical applications of the gene expression technology.

### Gene Expression Profiling and Pathogenesis of HCC

Transcriptional analysis of HCC, in common with other tumors, has identified dysregulation of numerous molecular pathways (for review see 14-16). For example, these pathways are associated with cell proliferation and cell cycle regulation, apoptosis, angiogenesis, protein degradation as well as cell signaling, transcriptional regulation and immune response (particularly in HCV and HBV associated HCC). However, the data obtained from gene expression profiling suggest that the dysregulation of potential oncogenic pathways is quite heterogeneous in human HCC.\(^17\) The heterogeneous nature of HCC (as well as other human cancers) combined with the analytical aspects of the DNA microarray technology (i.e., massive data-output harboring several sources of variability from the biological samples, hybridization protocols, scanning and image analysis) has made it difficult (and to some extent still is) to accurately and reproducibly classify HCC. The standard strategy for estimating the accuracy of a classification method is to apply a training-validation approach in which the training set is utilized to identify the molecular signature and the validation set is used to estimate the degree of misclassifications. Michiels and colleagues\(^18\) reanalyzed data from seven large published studies that attempted to predict prognosis of cancer patients (including HCC patients) on the basis of DNA microarray analysis. The authors found that the list of genes identified as predictors of prognosis was highly unstable, and the selected molecular signatures strongly depended on the selection of patients in the training sets. Also in all but one study, the proportion that was misclassified decreased as the number of patients in the training set increased. Moreover the results revealed that five of the seven studies did not classify patients better than chance. The authors concluded that, because of inadequate validation of the results, these studies were overoptimistic and they recommended the use of validation by repeated random sampling. They also emphasized that studies with large sample size are needed before expression profiling can be utilized in the clinic. It is clear from this analysis that inadequate validation of molecular signatures used for prognostic classification of HCC (and other cancers) may be one of the major source of failures in applying “molecular signatures” in the clinic.

In a recent global gene expression analysis of human HCC, we identified two distinctive subclasses that are highly associated with the survival of the patients.\(^12\) This approach also identified a limited number of genes that
accurately predicted the length of survival and provided new molecular insights into the pathogenesis of HCC. For example, information obtained from knowledge-based annotation of the 406 survival genes uncovered some of the molecular pathways responsible for the biological differences observed in the two subclasses of HCC. As expected, measurement of cell proliferation and apoptotic rates in both subclasses strongly support the principle that imbalance between cell proliferation and cell death is a central feature of tumors. Indeed, the proliferation and apoptosis measurements provided the best quantitative separation of the two survival subclasses. However, additional issues were also underscored. For example, the low survival subclass displayed higher expression of genes involved in ubiquitin and histone modification, suggesting an etiological involvement of these processes in accelerating the progression of HCC. This is of interest since it is well established that the ubiquitin system is often deregulated in cancers. Also, the degree of ubiquitination has also been proposed as a possible predictive marker for recurrence of human HCC. The deregulated components in ubiquitin-mediated protein degradation may therefore provide attractive therapeutic targets for novel HCC treatment modalities.

While dysregulation in proliferation and survival pathways are common to all cancers, there are other pathways that may be specific for certain types of cancers. Defining and characterizing the general and specific biological pathways that are aberrant in neoplastic diseases has a clear and important implication for clinical oncology. The DNA microarray platform technology is uniquely suited to address this issue and to provide a “global map” of shared and unique molecular “modules” characteristic of human and experimentally induced cancers (for review see 15). These molecular modules consist of functionally related genes that are members of the same biological pathway, have a shared structural motif, are expressed in a specific tissue, or are induced by a specific stimulus. Consequently, the modular approach attempts to go beyond the clustering and identification of gene signatures to identify the biological processes that are disrupted during the disease process. This process is further helped by the modular attribute that permits the analysis and discovery of a coherent set of changes in expression too small to detect when analyzing expression profiles of individual genes in isolation. The general application of the module-based transcriptional analysis of cancer has provided important insights into the complexity of the pathways that create and maintain tumors and is beginning to uncover some of the mechanisms that are responsible for cancer progression. The power of the modular approach is well illustrated by Segal et al. These authors applied a module-level analysis of a “cancer compendium” from multiple studies to obtain a global view of shared and unique molecular modules in human cancer. They identified gene sets with similar behavior across arrays, combined them into modules and used these modules to characterize a variety of clinical conditions (e.g., tumor stage and type) by the combination of activated and deactivated modules. These data were combined into a “cancer module map” which showed that activation and repression of some modules (e.g., cell cycle) was shared across multiple cancer types and could be related to general tumorigenic processes. Other modules (e.g., growth inhibitory modules) were more specific for the tissue of origin or progression of a particular tumor. Most importantly, the modular map characterizes each condition by a particular combination of modular activity, and thereby provides insight into the mechanistic aspects of specific tumors. The evident value of applying the higher-level gene module analysis for the molecular characterization of human cancer, when combined with the fact that a modular approach can be applied uniformly to multiple datasets from different tumor types, makes this analytic methodology particularly attractive.

Comparative Functional Genomics of HCC

Modeling of human cancer in genetically engineered mice has been intensively pursued over the last two decades, even though tumorigenesis in mice does not fully parallel that in humans. Indeed, several notable differences have been documented between human and mouse tumorigenesis. These observations, together with the fact that human tumors are generally quite heterogeneous, have raised questions about how accurate surrogate mouse models are for humans, particularly for analyzing the molecular pathways of carcinogenesis. Recent studies, including our own, have provided important new information on the usefulness of modeling human cancer in genetically engineered mice.

Our approach is based on the hypothesis that, since regulatory genomic elements of evolutionarily related species are conserved, gene expression signatures reflecting similar phenotypes in the species would also be conserved. To test this hypothesis, we investigated whether comparison of global expression patterns of orthologous genes in human and mouse HCC would identify similar and dissimilar tumor phenotypes, and thus allow the identification of the best-fit mouse models for human HCC. The results from global gene expression analysis of human and mouse HCC have allowed us to identify mouse models that both closely mimic human HCC as well as those that show very little similarity to the human disease (Fig.2., reference 9). Another example il-
Fig. 1. Gene expression profiling of human HCC and application of comparative functional genomics to identify best-fit mouse models to study human cancer. (A) Hierarchical clustering of 91 HCC tumors. Genes with an expression ratio that had at least a twofold difference relative to reference in at least 9 tissues were selected for hierarchical analysis (4,187 gene features). The data are presented in matrix format in which rows represent the individual gene and columns represent each tissue. Each cell in the matrix represents the expression level of a gene feature in an individual tissue. The red and green color in cells reflect high and low expression levels, respectively, as indicated in the scale bar (log2 transformed scale). (B) Significant association of gene expression patterns with patients survival. Kaplan-Meier plot of overall survival of HCC patients grouped on the basis of gene expression profiling shown in panel A. (C) Comparative functional genomics of HCC. Cluster analysis of integrated human and mouse HCC. Unsupervised hierarchical cluster analysis of integrated 68 mouse and 91 human HCC tumors. The data are presented in matrix format in which columns represent individual gene and rows represent each tissue. Red and blue bars represent human and mouse HCC tissues, respectively. The identity of each HCC tissue is shown at the end of each row (for details see reference 9). (D) Phenotypic similarities between HCCs generated in the transgenic mouse models and subclass A and B of human HCC. These models should also be particularly valuable for testing both potential therapeutic targets identified in human studies and preclinical trials of drugs.

Illustrating the usefulness of genetically engineered mouse models comes from comparing gene expression profiles in a mouse model of prostate cancer overexpressing the myc gene with those obtained from the human disease. To define the functional role of the c-myc oncogene in human prostate cancer, these investigators generated transgenic mice expressing human c-myc in the mouse prostate. All mice developed invasive adenocarcinoma of the prostate. Gene expression profiling identified a myc prostate cancer expression signature, which included the putative human tumor suppressor NXK3.1. The myc-specific gene expression signature in the mouse model permitted the definition of a subset of myc-like human cancer that is probably driven by myc amplification or by other mechanisms embedded in the myc activation pathways. This approach further illustrates how genomic technologies can be applied to mouse cancer models to extract important information from human tumor databases for both clinical oncology and the understanding of molecular pathogenesis of cancer.

A recent work by Sweet-Cordero et al. further emphasizes the usefulness of genetically modified mouse models to probe human cancers. These investigators found that, although a gene-expression signature of KRAS2 activation could not be identified when analyzing human tumors with known KRAS2 mutation status alone, integrating data from Kras2-mediated lung cancer mouse model and human data uncovered a gene-expression signature of KRAS2 mutation in human lung cancer. Using these data the investigators were able to identify both a pattern of gene expression indicative of KRAS2 mutation and potential effectors of oncogenic KRAS2 activity in human cancer.

These novel approaches establish a molecular relationship between the mouse models and the human cancers (for review see Lee and Thorgeirsson, Oncogene, in press). The clear gain to be realized from adopting the genomics technologies is to connect molecular pathogenic features of human cancer to mouse models, and indeed other experimental animal models, with a greater level of confidence.

**Toward Integrative Functional Genomics**

Success of gene expression profiling of HCC is highlighted by discovering consistent gene expression patterns associated with histological or clinical phenotypes and discovering subtypes of cancers previously unrecognized with conventional methods. This approach promises to provide diagnostic and prognostic markers that can be clinically used in the near future. The research focus is now shifting toward identifying genetic determinants that are components of the specific regulatory pathways altered in cancers with the aim of discovering novel therapeutic targets. However, identification and characterization of altered regulatory pathways in tumors is challenging, because many of the regulatory switches are controlled post-transcriptionally (i.e., localization, modi-
fication, interaction with other key regulatory proteins). Nevertheless, oncogenic pathways evolved from normal cells by resurrecting pre-existing pathways in different ways or by recombining components of the pathways in a novel fashion that supports tumorigenesis. Therefore, using gene expression profiling studies to map and refine biological pathway maps in the developing and normal adult human liver under a variety of physiological conditions (i.e., generating biological pathway modules) might provide insight into the relevance of these pathways to the development of HCC. In addition, gene expression data from animal models (particularly genetically-defined mouse models), as well as gene expression signatures characteristic for all the cell types in the liver, need to be collected and integrated into the gene expression patterns from human HCC. This approach, integrative functional genomics, will further help not only to stratify patients into more clinically homogeneous groups, but also uncover the origins of tumor cells and distinct pathways involved in the molecular pathogenesis of HCC. In future, heterogeneous patients can be stratified for clinical trials based on molecular features. Target specific treatments can then be applied to homogeneous groups of patients most likely respond to the treatment.

The DNA microarray technology has provided an extraordinary opportunity to perform integrative analyses of the cancer transcriptome. The results of array-based gene expression profiles have impacted both clinical decision-making in oncology and advanced our understanding of cancer biology, as well as facilitated the development of more effective therapies (for reviews see 16, 28-33). The pace of current progress in the cancer profiling field, combined with the advances in other high-throughput technologies in genomics and proteomics, and in bioinformatics promises to yield unprecedented biological insights from the integrative (or systems) analysis of the combined cancer genomics database. As we transit from the genomic to the post-genomic period, a number of the predictions made during the genomics era will no doubt be realized (Fig. 3). In the immediate future the focus will undoubtedly be on using the current genomic technologies to improve the diagnosis and treatment of
cancer, and one can certainly expect significant progress in the treatment of liver cancer. When HEPATOLOGY celebrates the 50th anniversary in the fullness of the post-genomics era, one may safely predict that the practice of clinical hepatology and medicine in general will have shifted into a predictive and preventive phase (Fig. 3). At that time liver cancer may be both effectively prevented and treated.

References