Comparative Functional Genomics for Identifying Models of Human Cancer

Ju-Seog Lee, Joe W. Grisham and Snorri S. Thorgeirsson*

Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4262, USA

* Correspondence should be addressed to S.S.T. (e-mail: snorri_thorgeirsson@nih.gov).
Abstract

Genetically modified mice with overexpressed and/or deleted genes have been extensively used to model human cancer. However it is uncertain to what extent the mouse models reproduce the corresponding cancers in humans. We have compared global gene expression patterns in human and mouse hepatocellular carcinomas (HCC) in an attempt to identify the mouse models that most extensively reproduce molecular pathways in the human tumors. The comparative analysis of gene expression patterns in murine and human HCC indicates that certain genetic mouse models closely reproduce the gene expression patterns of HCC in humans, while others do not. Identification of mouse models that reproduce the molecular features of specific human cancers (or subclasses of specific human cancers) promises to accelerate both the understanding of molecular pathogenesis of cancer and the discovery of therapeutic targets. We propose that this method, comparative functional genomics, can effectively be applied to the analysis of mouse models for other human cancers.
Introduction
Cancer develops via a complex multistage cellular and molecular process. Each stage includes genetic and/or epigenetic events that progressively transform normal cells into highly malignant derivatives. Several regulatory pathways are sequentially or simultaneously altered during tumor progression, resulting in growth autonomy, insensitivity to growth inhibition and apoptosis, augmented neoangiogenesis, and invasion and metastatic dissemination (1). Although many approaches, including genome-scale studies, provide insights into some of the stages in human tumorigenesis, the sequential analysis of tumor development in humans is difficult for obvious reasons. Experimental animal models, in particular mouse models, of carcinogenesis have permitted the examination of stages of neoplastic development in considerable detail. However, it is uncertain to what extent mouse models reproduce the molecular features that characterize human cancers. We have hypothesized that if regulatory elements of evolutionarily related species are conserved, gene expression signatures reflecting similar phenotypes in the species would also be conserved (2). To test this hypothesis, we investigated whether comparison of global expression patterns of orthologous genes in human and mouse HCCs, an approach we term “comparative functional genomics”, would identify similar and dissimilar tumor phenotypes (2). Here we discuss both the challenges and potential value of applying comparative functional genomics, to assess the accuracy of mouse models. Based on earlier results with HCC, we propose that comparative functional genomics can be used to assess the accuracy of mouse models for the cancers in other tissues.

Challenges in using mice to model cancer in humans
Technological advances during the last two decades have facilitated the generation of genetically engineered mice that mimic genetic alterations frequently observed in human cancers. By either selectively driving the overexpression or deletion (knocking out) of genes of interest in mice, the development of tumors in various tissues can be induced, and their similarity to human tumor tested (3:4). Indeed many mouse models of human cancers have been generated based on the notion that the same genetic alterations will lead to similar biological outcomes in both species. Although the information obtained from these genetically altered mouse models has provided answers to many critical questions related to the mechanisms of tumor progression (5:6), this approach has not always worked. This is well illustrated by the lack of retinoblastoma development when one or both alleles of the Rb gene are mutated or deleted in mice. Loss of Rb function alone is not sufficient to produce retinoblastoma in mice (7-10) which requires the additional loss of the Rb-related tumor suppressor, p107 (11). In contrast, inactivating mutations in both alleles of RB are both necessary and sufficient for the development of retinoblastoma in humans (12). Although the mouse model of retinoblastoma is not a precise molecular model of the human tumor, it has provided a deeper insight into the multiplicity of molecular mechanisms of tumor development. The retinoblastoma model clearly demonstrates, and there are other examples, the challenges that must be faced in generating and selecting accurate mouse models of specific human cancers due to subtle differences in molecular regulatory circuitry in the two species. Other major challenges to the wide use of mouse to models of specific human cancers include the fact that all types of human cancer are not yet reproducible in mice. It is,
however, reasonable to expect that this problem will be solved in the near future with
improved understanding of the biology of cancers and technological advances in the
capacity to control gene expression in cells of specific organs and tissues.

Although certain mouse tumor models reproduce similar cellular alterations and
histopathological patterns as the human counterparts, it is uncertain whether these models
recapitulate the molecular regulatory pathways that are altered in the human tumors.
Furthermore, it is evident that all human cancers are heterogeneous with respect to
natural history, molecular pathogenesis, and response to treatment. The issue of tumor
heterogeneity introduces a further complexity for the identification of the “best-fit”
mouse models for human cancers. There is, however, mounting evidence that distinct
genomic alterations in tumors, reflected in gene expression patterns, correlate with a
different clinical outcome in cancer patients. Studies of gene expression profiles of
diffuse large B cell lymphoma in humans provide one of the best examples of how
molecularly distinct subgroups of cancer display different clinical behaviors (13;14).
Somewhat similar observations have recently been made on mammary, prostate, and lung
cancers (15-19), and the number of clinically variant subgroups of different cancers that
are classified as a single type based on morphology alone will undoubtedly expand with
further analysis of global gene expression patterns. The identification of new subgroups
of human cancers introduces additional challenges to the application of mouse models to
analyze aspects of human cancers that cannot be studied readily in humans. It seems
likely that as more gene expression patterns of specific mouse tumors are analyzed, many
will also show heterogeneity of molecular development that is not reflected by the
morphology of the tumors. This dilemma may be addressed by directly comparing the
global gene expression patterns of mouse and human tumors through the “comparative
functional genomics” approach.

**Mouse models of hepatocarcinogenesis**

Hepatocellular carcinoma (HCC) is the fifth most common cancer in the world,
accounting for an estimated half million deaths annually (20). While HCC is a major
cancer in Southeast Asia and sub-Sahara Africa, the incidence of this tumor is much
lower in Western Europe and the United States. However, the incidence and mortality
rate of HCC have doubled in the United States over the past 25 years and this upward
trend is expected to continue over the next decades (21;22). Although the agents that
cause most of the HCC in humans are known (i.e. HBV and HCV infections, aflatoxin
exposure, alcoholism, nonalcoholic steatohepatitis, etc), and much is known about the
cellular and molecular alterations in human HCC (23), the molecular pathogenesis of
HCC is still not well understood (24-27).

HCC is also a common tumor in mice, and the morphologic expression of HCC in
mice resembles that in humans, although causative factors differ (28). Analysis of gene
expression patterns of chemically induced HCC in animal models has not yet elucidated
the molecular pathogenesis of this tumor, although these studies suggest that diverse
molecular pathways may lead to common patterns of morphologic features (29). In
addition to chemically induced liver tumors HCC development has also been produced in
genetically engineered mice generated to decipher molecular pathways that lead to this
tumor (Table 1).
Specific overexpression of genes such as *Myc*, *Tgfα*, *E2f1*, *Ccnd1*, and *Hras1 (G12V)* in the liver induces the development of HCC in mice (30-35). Induction of HCC development by expression of viral proteins in mouse liver is also documented (35-43). Gene knockout models have been generated in mice to study the participation of selected tumor suppressor genes in hepatocarcinogenesis (44-46). Chemical induction of HCC in mice with diethylnitrosamine (DENA) is a widely used model to investigate multiple aspects of hepatocarcinogenesis (47), and provides a useful base to compare HCC models induced by exogenous agents to genetically engineered models. Since HCC develops in each of these models, it is reasonable to presume that the all models can be useful for elucidating the molecular pathogenesis of HCC, including aspects of the regulation of HCC development in humans. However it is also evident that individual mouse models cannot be equally useful for identifying the molecular pathways involved in all variants of human HCC, due to species differences in liver physiology and the divergent impact of cancer-associated genes in the two species (48). Therefore the challenge is to identify the mouse models that most closely reproduce the different molecular subtypes of human HCC.

**Comparative genomics and comparative functional genomics**

Comparative studies of the genomic structure of evolutionarily related and unrelated species are useful for predicting gene structures and the number of active genes in the species (49). Such comparative studies have also identified cis-acting regulatory elements that regulate gene expression (50). Availability of genome sequences from many species has notably advanced comparative genomics in particular and the understanding of evolutionary biology in general. The neutral theory of molecular evolution, regardless of the controversy it has engendered among evolution theorists, provides a framework for the identification of functional DNA sequences in genomes of different species (51;52). The central hypothesis of the neutral theory of molecular evolution is that the vast majority of mutations in DNA sequences are neutral with respect to the fitness of an organism. While deleterious mutations are rapidly removed by selection, neutral mutations persist and follow a stochastic process of genetic drift through a population. Therefore, non-neutral DNA sequences (functional DNA sequences) must be conserved during evolution, whereas neutral mutations accumulate. This difference has allowed the identification of both protein coding sequences and functional noncoding sequences in a genome (49;50;53). We argue that if the regulatory elements of evolutionary related species are conserved it is reasonable to hypothesize that gene expression signatures that represent similar phenotypes in different species could also be conserved. This hypothesis forms the basis for our attempt to identify aberrant phenotypes reflecting molecular pathways that are conserved during cancer development in mice and humans.

We have recently investigated the possibility that the comparison of gene expression patterns of HCC from mice and humans would permit the direct identification of common aberrant molecular pathways involved in hepatocarcinogenesis. The involvement of common molecular pathways in subsets of HCC in both species could potentially allow the identification of mouse models that provide a “best-fit” for subclasses of human HCC that we found in a previous analysis of global gene expression (54). Gene expression patterns of mouse HCC were obtained from seven HCC mouse
models; two chemically induced (ciprofibrate and DENA), four transgenic (targeted overexpression of Myc, E2f1, Myc/E2f1, and Myc/Tgfa in the liver), and one knockout (Acox1−/−). Gene expression patterns of human HCC from 91 patients were analyzed in an independent study that uncovered two distinctive subgroups of HCC that were related to patient survival, and this association was validated by independent supervised methods (54). Orthologous human and mouse genes from both data sets were selected, and the gene expression data were integrated after standardizing the relative expression levels for both species (Figure 1). In hierarchical clustering analysis of integrated data, gene expression patterns of HCC from Myc, E2f1, and Myc/E2f1 mice had the highest similarity with those of the better survival group of human HCC, while the expression patterns of Myc/Tgfa and DENA-induced mouse HCC were most similar to those of the poorer survival group of human HCC (Figure 2). These results suggest that these two classes of mouse models might more closely recapitulate the molecular patterns of the two subclasses of human HCC. In contrast, gene expression patterns of HCC that develop in Acox1 knockout and ciprofibrate-treated mice were least similar to those observed in either subclass of human HCC. The development of HCC in these two models is driven by peroxisome proliferation in the liver (44;54-56). These results suggest that the process of hepatocarcinogenesis induced by peroxisome proliferation in mice progresses along molecular pathways that do not frequently occur in humans, supporting previous studies suggesting that humans are insensitive to hepatotoxic effects of peroxisome proliferation (57).

**Conclusion and Perspectives**

Although transgenic and knockout mouse models have improved our insight into the mechanistic aspects on human cancers, in most instances we still rely on the observed phenotypes with limited sets of genotypes to correlate the development of tumors in the two species. Aligning orthologous genomic sequence from different species has allowed the identification of evolutionary conserved sequences that might be functional regulatory elements (49;50;53). Therefore, it is reasonable to postulate that similar gene expression signatures exemplify phenotypes that are conserved in different species. Our work has demonstrated that gene expression signatures reflecting similar phenotypes are indeed conserved in mice and humans during the development of HCC, and that interspecies comparison of global gene expression patterns can identify mouse models that best mimic human HCC. Application of this approach, “comparative functional genomics” may similarly identify the most relevant mouse models for cancer of other tissues.

However, there are critical issues that must be considered before applying this approach. The fact that basal levels of gene expression differ between two species can complicate the analysis and interpretation of the integrated gene expression data. Use of a normal counterpart (or of the specific cell types involved) of the cancerous organ as a reference to measure gene expression ratios can lower the risk of false interpretation of analysis. Furthermore, standardization of data sets requires large numbers of human samples and multiple mouse models. Integration of data from a few human tumors and one or two mouse models will not provide sufficient discriminatory power to identify the most relevant mouse models.
The potential advantage to be realized from a comparative functional genomics analysis is the capability of selecting the animal models that best express the molecular features of human cancers. Discovery of subclasses of human tumors, such as HCC with homogenous patterns of gene expression that correlate with survival, suggests that appropriate animal models of these tumors, with similar patterns of gene expression, can be identified for each subclass. Indeed, we have identified mouse models of HCC that display patterns of gene expression similar to both the subclasses of human HCC (2). Identification of appropriate mouse models of human cancers by comparative functional genomics should accelerate the discovery of relevant therapeutic targets and the testing of candidate therapeutic agents.

Establishing a molecular relationship between mouse and human cancers may also provide the basis to investigate the early genomic and molecular events in neoplastic development. We have analyzed gene expression patterns only in the final stages of tumor (HCC) development and classified them in terms of survival and functional categories of genes. Although these aberrantly expressed genes are essential molecular elements in the malregulation of cellular physiology that produces the cancer phenotype, these results do not yet address the specific genetic aberrations that start the affected cell(s) on the molecular pathway to cancer, or the subsequent genomic and molecular changes that enable cells to progress toward the final malignant phenotype. Thus, future genomic studies with early, intermediate, and late precancerous cellular lesions will be necessary to identify the specific molecular pathways that are involved and the molecular components of the pathways that are aberrant. Using the conserved gene expression patterns in human and animal models as a “foundation”, future studies should identify the variety of aberrantly expressed genes, abnormal molecular species, and pathways that lead to this outcome. Because of the difficulties of acquiring preneoplastic lesions from human patients, these studies will need to be conducted simultaneously with tissues from animal models. Only when this is accomplished will we be able to understand the molecular pathogenesis of cancer and how closely mice (and other experimental animals) model cancer in humans.
Reference List


<table>
<thead>
<tr>
<th>Genes</th>
<th>Promoter</th>
<th>Strain</th>
<th>Latency period for HCC (month)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tgfa</strong></td>
<td>MT</td>
<td>CD1</td>
<td>13 to 15</td>
<td>(33)</td>
</tr>
<tr>
<td><strong>Myc</strong></td>
<td>Albumin</td>
<td>C57BL/6 x SJL</td>
<td>15 to 18 (HCA)</td>
<td>(35)</td>
</tr>
<tr>
<td><strong>Myc</strong></td>
<td>WHV</td>
<td>C57BL/6 x SJL</td>
<td>8 to 12</td>
<td>(32)</td>
</tr>
<tr>
<td><strong>Myc</strong></td>
<td>AAT</td>
<td>C57BL/6 x SJL</td>
<td>12</td>
<td>(58)</td>
</tr>
<tr>
<td><strong>Myc</strong></td>
<td>albumin</td>
<td>C57BL/6 x CBA/J</td>
<td>10 to 12</td>
<td>(34)</td>
</tr>
<tr>
<td><strong>Myc/cyto-MET</strong></td>
<td>WHV/AAT</td>
<td>C57BL/6 x SJL x DBA/2</td>
<td>3 to 5</td>
<td>(59)</td>
</tr>
<tr>
<td><strong>Myc/Tgfa</strong></td>
<td>albumin/MT</td>
<td>C57BL/6 x CBA/J x CD1</td>
<td>2 to 4</td>
<td>(34)</td>
</tr>
<tr>
<td><strong>Myc/HBX15</strong></td>
<td>HBX</td>
<td>C57BL/6 x SJL</td>
<td>3 to 6</td>
<td>(60)</td>
</tr>
<tr>
<td><strong>Myc/HBX</strong></td>
<td>WHV/AT3</td>
<td>C57BL/6 x DBA/2</td>
<td>5 to 7</td>
<td>(61)</td>
</tr>
<tr>
<td><strong>Egf</strong></td>
<td>albumin</td>
<td>CD2F1</td>
<td>6 to 7</td>
<td>(62)</td>
</tr>
<tr>
<td><strong>Myc/Egf</strong></td>
<td>ATT/alumin</td>
<td>C57BL/6 x SJL x CD2F1</td>
<td>3 to 4</td>
<td>(62)</td>
</tr>
<tr>
<td><strong>E2f1</strong></td>
<td>albumin</td>
<td>C57BL/6 x CBA/J</td>
<td>10 to 12</td>
<td>(30)</td>
</tr>
<tr>
<td><strong>My/2f1</strong></td>
<td>albumin</td>
<td>C57BL/6 x SJL</td>
<td>6</td>
<td>(63)</td>
</tr>
<tr>
<td><strong>H-Ras(G12V)</strong></td>
<td>albumin</td>
<td>C57BL/6 x SJL</td>
<td>5</td>
<td>(35)</td>
</tr>
<tr>
<td><strong>PML-RAR alpha</strong></td>
<td>MT</td>
<td>C57BL/6</td>
<td>4 to 6</td>
<td>(64)</td>
</tr>
<tr>
<td><strong>MET</strong></td>
<td>TRE &amp; hCMV</td>
<td>FVB/N</td>
<td>4 to 6</td>
<td>(65)</td>
</tr>
<tr>
<td><strong>SERPINA1</strong> (Z-allele)</td>
<td>native</td>
<td>C57BL/6 x DBA/2</td>
<td>12 to 16</td>
<td>(66)</td>
</tr>
<tr>
<td><strong>Plau</strong></td>
<td>albumin</td>
<td>C57BL/6 x SJL</td>
<td>10 to 16</td>
<td>(67)</td>
</tr>
<tr>
<td><strong>Hgf</strong></td>
<td>MT</td>
<td>FVB/N</td>
<td>17 to 19</td>
<td>(68)</td>
</tr>
<tr>
<td><strong>Ccnd1</strong></td>
<td>LFABP</td>
<td>C57BL/6</td>
<td>17</td>
<td>(31)</td>
</tr>
<tr>
<td><strong>Ccnd1/Tgfb1</strong></td>
<td>LFABP/alumin</td>
<td>C57BL/6 x CBA/J</td>
<td>12 to 15</td>
<td>(69)</td>
</tr>
<tr>
<td><strong>Tgfb1</strong></td>
<td>albumin</td>
<td>C57BL/6 x CBA/J</td>
<td>16 to 18</td>
<td>(70)</td>
</tr>
<tr>
<td><strong>HBsAg</strong></td>
<td>albumin</td>
<td>C57BL/6 x SJL</td>
<td>15 to 18</td>
<td>(42)</td>
</tr>
<tr>
<td><strong>HBsAg</strong></td>
<td>p21 (knock-in)</td>
<td>129S6/SvEvTac</td>
<td>15 to 18</td>
<td>(42)</td>
</tr>
<tr>
<td><strong>HBX</strong></td>
<td>p21 (knock-in)</td>
<td>129S6/SvEvTac</td>
<td>15 to 18</td>
<td>(42)</td>
</tr>
<tr>
<td><strong>HBX</strong></td>
<td>HBX native</td>
<td>CD1</td>
<td>11 to 15</td>
<td>(37)</td>
</tr>
<tr>
<td><strong>HBX</strong></td>
<td>HBX native</td>
<td>CD1</td>
<td>13 to 15</td>
<td>(38)</td>
</tr>
<tr>
<td><strong>HBX</strong></td>
<td>HBX native</td>
<td>C57BL/6 x DBA</td>
<td>11 to 18</td>
<td>(43)</td>
</tr>
<tr>
<td><strong>SV40 T-ag</strong></td>
<td>AAT</td>
<td>C57BL/6</td>
<td>2 to 3</td>
<td>(41)</td>
</tr>
<tr>
<td><strong>SV40 T-ag</strong></td>
<td>AAT</td>
<td>C57BL/6 x SJL</td>
<td>3 to 7</td>
<td>(58)</td>
</tr>
<tr>
<td><strong>SV40 T-ag</strong></td>
<td>albumin</td>
<td>C57BL/6 x SJL</td>
<td>3 to 5</td>
<td>(35)</td>
</tr>
<tr>
<td><strong>SV40 T-ag</strong></td>
<td>MUP</td>
<td>C57BL/10Ros-p(d) x C3Hf/HeRos</td>
<td>6</td>
<td>(40)</td>
</tr>
<tr>
<td><strong>HCV core</strong></td>
<td>HBX</td>
<td>C57BL/6</td>
<td>16 to 19</td>
<td>(39)</td>
</tr>
</tbody>
</table>

**Knockout models**

| Acox1/-/- | 129/Ola, C57BL/6 | 10 to 15 | (44) |
| Lkb1+/+ | 129/SvJ | 10 to 12 | (46) |
| Mdr2(ABcb2)/-/- | 129/OlaHsd | 18 | (45) |

WHV: Woodchuck Hepatitis Virus; AAT: α-1-Antitrypsin; MUP: Major Urinary Protein; LFABP: Liver Fatty Acid Binding Protein; HCA: Hepatocellular Adenoma; MT: Metallothionein; AT3: Antithrombin 3
Fig. 1 Comparative functional genomics. Gene expression data of mouse (seven models) and human HCC tissues (two subclasses) were collected independently using spotted dual-channel microarray platforms. In both series of experiments, normal livers were used as a reference to measure gene expression ratios. Before the integration of the two independent data sets, orthologous genes present in both microarrays were selected for further analysis. After an initial data reduction to focus on genes whose expression varied non-trivially across the samples, gene expression ratios were then standardized to a mean of zero and a standard deviation of one separately in each data set. By applying unsupervised (hierarchical clustering) and supervised analysis (prediction models) of gene expression patterns, mouse models that best or least mimic human conditions can be identified. The “best fit” mouse models can then be used to test hypotheses on tumor progression that are generated by analysis of cross-species gene expression patterns or from other experimental data. These models will be also extremely valuable for testing both potential therapeutic targets identified in human study and preclinical trials of drugs.

Fig. 2. Cluster analysis of integrated human and mouse HCC. (A) Unsupervised hierarchical cluster analysis using distance matrices based on Pearson correlation coefficients of integrated 68 mouse and 91 human HCC tumors. Orthologous genes with an expression ratio that has at least 2-fold difference relative to reference in at least 10% of tissues in one of both data sets were selected for hierarchical analysis (1,650 genes). Red and blue color bars represent human and mouse HCC tissues respectively. Identities of each HCC tissues were shown in right end of each row. (B) Correlation heat map depicting overall similarity and dissimilarity of gene expression pattern among human and mouse HCC tissues. The red and green color in cells reflect the correlation between gene expression patterns of two HCC tissues compared as indicated in the scale bar. Figure is adopted and modified from Lee et al. (2).
Selection of orthologous genes

Standardization and integration of data sets

Supervised and unsupervised analysis of data

Identify best-fit mouse models for each subclass of human HCC
Identify conserved gene expression patterns in mouse and human HCC
Identify potential therapeutic targets in human HCC

Build hypothesis

Test hypothesis

Validation and preclinical trial of targets