Analysis of the Mechanisms Mediating Tumor-Specific Changes in Gene Expression in Human Liver Tumors

Luis G. Acevedo, Mark Bieda, Roland Green, and Peggy J. Farnham

Abstract

There is widespread interest in efficient characterization of differences between tumor and normal samples. Here, we show an effective methodology for genome-scale characterization of tumors. Using matched normal and tumor samples from liver cancer patients, as well as non–cancer-related normal liver tissue, we first determined changes in gene expression as monitored on RNA expression arrays. We identified several hundred mRNAs that were consistently changed in the tumor samples. To characterize the mechanisms responsible for creation of the tumor-specific transcriptome, we performed chromatin immunoprecipitation on microarray experiments to assay binding of RNA polymerase II, H3me3K27, and H3me3K9 and DNA methylation in 25,000 promotor regions. These experiments identified changes in active and silenced regions of the genome in the tumor cells. Finally, we used a “virtual comparative genomic hybridization” method to identify copy number alterations in the tumor samples. Through comparison of RNA polymerase II binding, chromatin structure, DNA methylation, and copy number changes, we suggest that the major contributor to creation of the liver tumor transcriptome was changes in gene copy number. [Cancer Res 2008;68(8):2641–51]

Introduction

Liver cancer is the fifth most common cancer worldwide and the third most common cause of cancer mortality (1), and hepatocellular carcinoma (HCC) accounts for ~90% of primary liver cancers. Most cases of liver cancer occur in either sub-Saharan Africa or in Eastern Asia (China alone accounts for 50% of the world’s cases). The major risk factor differs in different areas. In most high-risk areas, the dominant risk factor is chronic hepatitis B virus (HBV) infection or consumption of aflatoxin B–contaminated food. Many of the high-rate Asian countries now vaccinate all newborns against HBV and thus HCC rates are decreasing in these areas. In contrast, rates of liver cancer in the low-risk areas, such as North and South America and Northern Europe, are increasing, perhaps due to an increased prevalence of hepatitis C virus infection in these areas. Accordingly, in the United States, HCC is the fastest growing cause of cancer-related death in men.

In attempts to characterize molecular changes that occur during the development of HCC, several previous studies have compared RNA expression levels in normal versus tumor liver using techniques such as directed analysis of the expression of candidate genes, serial analysis of gene expression assays, differential display, and hybridization of cDNA arrays (2–8). Gene expression alterations involving loss of checkpoint control have been commonly observed (see ref. 1 for a review). In particular, down-regulation of the p53 and RB tumor-suppressor pathways are affected at multiple levels. For example, the vast majority of human HCC overexpress gankyrin, which inhibits both RB and p53 checkpoint function (9, 10).

One approach toward the treatment of HCC is to attempt to reverse transcriptional deregulation in the tumor cells, thus restoring the normal liver transcriptome and the normal liver phenotype. For example, it has been shown that adenovirus-delivered gankyrin–specific inhibitory RNAs can reduce cell growth and enhance apoptosis in HCC cell culture lines (11). However, treatment of HCC in patients is more likely to involve therapies other than use of small interfering RNAs. One potential therapeutic option is to target the transcriptional machinery and/or chromatin structure that is responsible for deregulated expression of genes such as gankyrin. The success of this approach depends on a thorough understanding of the mechanisms by which genes are up- or down-regulated in the liver tumor tissue.

The expression of protein-coding genes can be regulated at different steps, including transcription initiation and elongation or mRNA processing, transport, stability, and translation. In a normal cell, much of the regulation is thought to occur at the level of transcription initiation. Changes in transcription initiation in a tumor cell could be achieved by modulating the activity of a promoter that is already active. For example, a highly expressed gene could be down-regulated by loss of an upstream transcriptional regulator (due to deletions or mutations in the tumor genome) or a modestly expressed gene could be up-regulated by the inappropriate tumor-specific expression of a transcriptional regulator that is normally not active in that differentiated cell type. Promoters regulated in such a manner generally show changes in the amount of general transcription factors, such as RNA polymerase II, bound to the core promoter region and/or to downstream regions of the gene. Another mechanism by which transcription initiation can be achieved is via changes in the chromatin structure of a gene region (see ref. 12 for a review). For example, in differentiated cells, many genes (in particular those which are expressed in pluripotent cells or in other highly differentiated cell types) are silenced by certain chromatin modifications. Specifically, DNA methylation and trimethylation of lysine 9 or lysine 27 of histone H3 (H3me3K9 or H3me3K27) are marks for silenced chromatin. Changes in chromatin structure have been previously associated with tumor development and thus gene expression of a tumor cell could be regulated by transitions of a particular gene region from an active to a silenced state, or vice versa. Finally, tumor cells can use another mechanism for altering
the levels of particular genes that is not generally used in normal cells. Many tumors show amplification or deletion of certain regions of the genome. Such genomic alterations (in particular copy number changes observed at a low frequency in a certain tumor type) can be a consequence of the genomic instability that is associated with tumor development. However, specific regions that display amplification or deletion at a high frequency in a tumor type may reflect a causal relationship between gene expression and tumor development.

Described below are our studies in which we have characterized the mechanisms by which the tumor-specific transcriptome is regulated in a human liver tumor. Interestingly, we find that the majority of genes that display altered RNA expression in HCC associated with tumor development. However, specific regions that display amplification or deletion at a high frequency in a certain tumor type) can be a consequence of the genomic instability that is

Materials and Methods

Tissue samples. Human normal adjacent tissues and tumor liver tissues were obtained from the National Disease Research Interchange National Development and Research Institutes (Philadelphia, PA) and from the University of Wisconsin-Madison Hospital (Madison, WI). Samples were surgically collected and preserved by flash freezing in liquid nitrogen followed by storage at −70 °C. Patient 1 (OD04962) was a 58-year-old Black male. Patient 2 (OD11735) was a 77-year-old male. Patient 3 (165) was an 81-year-old White female. Normal liver tissue (not cancer-related) was from patient 4 (00-559) who was a 50-year-old female. Purified normal hepatocytes (patient 5) were purchased from CellDirect as cryopreserved suspension hepatocytes that were purified by isodensity Percoll centrifugation. The HuH7 cell line was purchased from American Type Culture Collection; cells were grown at 37 °C in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin.

RNA arrays. RNA samples were prepared using Qiagen RNeasy Micro Kit (Qiagen) and then assayed using the Agilent Systems Bioanalyzer 2100 to ensure that high-quality RNA was used for the array experiments. The Illumina TotalPrep RNA amplification kit from Ambion (AM1791) was used to generate biotinylated, amplified RNA for hybridization with the Illumina Sentrix Expression BeadChips Human 6-v1 or HumanRef-8-v2. The human 6-v1 gene expression beadchips consisted of a six-array, two-stripe format comprising 48,000 probes per array. In this array, 24,000 probes were from reseq sequences and 24,000 from other Genbank sequences. Human Ref-8 v2 arrays consist of 24,000 probes from reseq sequences on an eight-stripe format array (see the Web site for more details concerning the Illumina arrays5). Arrays were processed as per manufacturer's instructions, scanned at medium photo-multiplier tube settings as recommended by the manufacturer, and analyzed using Bead Studio Software v. 2.3.41. Hybridization signals were analyzed using BeadStudio Gene Expression Module v.3 (Illumina). All transcripts were normalized and ranked according to the detection P value, which is calculated based on signals of negative controls. The detection P value is calculated as \( 1 - \frac{Y}{N} \), where \( Y \) is the rank of the gene signal relative to negative controls and \( N \) is the number of negative controls. P value can be interpreted as probability of seeing a certain signal level without specific probe-target hybridization. Expressed genes are indicated as those having a P value of 0. Data were normalized using the “average” method, which simply adjusts the intensities of two populations of gene expression values such that the means of the populations become equal. Differential expression was calculated using an algorithm provided by Bead Studio. Fold enrichment values were used to obtain the list of candidates with >1.5-fold change and a P value of 0. RNA expression analysis was performed by Hierarchical Clustering using Genesis 1.7.2 software,6 with the average linkage clustering as agglomeration rule (13). All of the genes from the Illumina platform were used for the clustering analysis.

Chromatin immunoprecipitation assays and amplicon preparation. Chromatin immunoprecipitation (ChiP) assays were performed as previously described (14) with the following modifications for liver tissues. Briefly, tissues were cut in small pieces with a razor blade, cross-linked in 1.5% formaldehyde for 15 min, processed in a Medimachine (BD Biosciences) using a 50 μm medicon to produce a liver cell suspension. Nuclear extracts were prepared and chromatin was sonicated using a Bioruptor Sonicator (Diagenode). Each chromatin immunoprecipitation assay was performed using between 7 and 15 μg of chromatin. A detailed protocol for ChiP assays in liver tissue is available online. In addition, a detailed protocol for ChIP miniaturation or “MicroChiP” is available (15). ChiP assays with the 5-methylcytidine antibody (Eurogentec) were performed using the ChIP-IT Express kit (Active Motif). For these assays, genomic DNA was extracted by shaking cells in digestion buffer [100 mmol/L NaCl, 10 mmol/L Tris-Cl (pH 8), 25 mmol/L EDTA (pH 8), 0.5% SDS] and 0.1 mg/mL Proteinase K for 12 to 18 h at 50°C and purified using a phenol-chloroform extraction method. Extracted DNA was sonicated to an average size of 800 bp, denatured at 95°C for 10 min, quickly chilled on ice and captured on magnetic beads following the protocol as described by the manufacturer.

DNA microarrays. Amplicons were applied to 5 kb promoter arrays (see Supplementary Table S1 and manufacturer's Web site7 for details). The labeling and hybridization of DNA samples for ChiP on microarray (ChiP-chip) analysis was performed by NimbleGen Systems, Inc. Briefly, each DNA sample (1 μg) was denatured in the presence of 5′-Cy3- or 5′-Cy5-labeled random nonamers (Trillink Biotechnologies) and incubated with 100 units (exo-) Klenow fragment (NEB) and deoxynucleotide triphosphate mix [6 mmol/L each in TE buffer: 10 mmol/L Tris-1 mmol/L EDTA (pH 7.4); Invitrogen] for 2 h at 37°C. Reactions were terminated by addition of 0.5 mol/L EDTA (pH 8.0), precipitated with isopropanol, and resuspended in water. Then, 13 μg of the Cy5-labeled ChiP sample and 13 μg of the Cy3-labeled total sample were mixed together, dried down, and resuspended in 40 μL of NimbleGen Hybridization Buffer (NimbleGen Systems) plus 1.5 μg of human COT1 DNA. After denaturation, hybridization was carried out in a MAUI Hybridization System (BioMicro Systems) for 18 h at 42°C. The arrays were washed using NimbleGen Wash Buffer System (NimbleGen Systems), dried by centrifugation, and scanned at 5-μm resolution using the GenePix 4000B scanner (Axon Instruments). Fluorescence intensity raw data were obtained from scanned images of the arrays using NIMBLESCAN 2.0 extraction software (NimbleGen Systems). For each spot on the array, log 2 ratios of the Cy5-labeled test sample versus the Cy3-labeled reference sample were calculated. Then, the biweight mean of this log 2 ratio was subtracted from each point; this procedure is approximately equivalent to mean-normalization of each channel.

Analysis of ChiP-chip data. The human 5-kb promoter array consists of a set of two arrays encompassing a total of ~25,000 human promoters. We refer to the two arrays in the set as “promoter 1” and “promoter 2.” Most promoters encompass ~5 kb of genomic sequence tiled with a 50-mer probe every ~100 nucleotides. To analyze these data, we used a variant of the Maxfour approach previously described (17) for our data. However, we used a 10-point window instead of the 4-point window used previously.

References

Further statistical exploration of this procedure will be presented elsewhere. As expected, we found that a comparison of repeated normal ChIP-chip experiments from the same samples showed an excellent correlation between experiments. However, there was a relatively small but clear deviation from the expected slope of 1 in these experiments, probably due to small variations in experimental procedures or arrays. Hence, we sought to normalize arrays to compensate for these effects in comparing tumor and normal samples. To do so, we relied on the expectation that a relatively small percentage of the ~25,000 promoters monitored in our experiments will differ in the tumor versus normal tissues. For one patient (patient 1), the slope of a fit of the tumor versus normal data for RNAPOLII was almost unity (1.03). In contrast, we found that for a second patient (patient 2), the slope of the tumor versus normal for RNAPOLII was 0.57, indicating a difference in the overall strength of ratio signal on the array. Hence, we corrected all values for these effects. To determine promoters that showed increased binding in tumor versus normal, we used the following variables: the difference (i.e., tumor-normal) value had to exceed 0.2 and the absolute value of the ratio of the difference to the normal had to be at least 0.5. To calculate the list of promoters that, conversely, showed decreased binding in tumor versus normal, we used the same variables, except that the difference was “normal-tumor” and the ratio was the difference/tumor value. Our goal was to identify promoters that showed changes in binding in the normal versus the tumor samples. Because tumor samples are often heterogeneous, we reasoned that many of the instances of differences in binding in the normal versus tumor samples might be modest. Therefore, we chose “very generous” variables to identify the initial set of promoters that potentially showed changes in binding in the tumor samples. We realized that many of these potential changes would be false positives due to the generous cutoffs used in the array analysis. In fact, we confirmed this by analyzing two technical replicates of the normal to tumor comparison from patient 1. For example, our variables identified 2,134 promoters that showed increased binding of RNAPOLII in the tumor in replicate A and 2,299 promoters that showed increased binding in replicate B. However, only 465 of these promoters showed the same direction change in both of the replicate experiments. This analysis tells us that the low overlap between patient 1 and patient 2 is to be expected due to the generous variables and the nature of microarray hybridization. To deal with this issue, but still allow us to identify modest changes in binding in the tumor samples, we used generous cutoff variables but then required that the promoters show the same directional change in two independent patients.

Virtual comparative genomic hybridization analysis. To analyze data, custom programs in perl, bash, and R were produced in a Linux environment. We used two normal cell arrays and two tumor cell arrays for human tissue data. For each patient, the genomic DNA channels of the arrays were quantile normalized and a single “merged” gDNA vector of values was produced using the median values. We calculated the average of all probes for a single promoter so that the promoter was represented by a single value. The same procedure was applied to the normal cell arrays. After calculation of the ratio of tumor/normal, the data was smoothed using a 21-point mean filter. Data were displayed using the snapCGH package in R. Confirmations of the amplified or deleted regions identified using the ChIP-chip assays were performed using PCR, which we have found is a reliable and cost-effective method for analysis of DNA. For these experiments, the number of PCR cycles was kept low to ensure that the signals were within the linear range of the assay, providing a semiquantitative analysis.

DAVID analysis. Functional annotations were performed using the program DAVID 2.1 (18). The same variables were used for all analyses presented in this study. These variables were Gene Ontology Molecular Function term, level 2; Interpro name in the Protein Domains section; and SP_PIR_Keys keywords in the Functional Categories section. Before performing the analyses, hypothetical genes and genes with no known function were removed from the list. After performing the analyses, all categories that represented <4% of the total number of genes were eliminated. In addition, redundant terms (e.g., transcriptional regulation and transcription factor activity) and noninformative terms (e.g., multigene family) were also eliminated.

Results

Characterization of the tumor-specific transcriptome. The first question we addressed was whether a section of liver tissue classified as normal, but taken from a patient having liver cancer, displayed gene expression patterns that were more closely related to the tumor sample from the same patient or to a normal sample from a person who did not have liver cancer. Using Illumina RNA expression arrays, we characterized RNA from two sets of matched normal and tumor liver samples, plus an additional sample of liver tissue from a patient who did not have liver cancer, a sample of purified normal hepatocytes, and Huh7 hepatoblastoma cells (a summary of all the arrays used in this study is provided as Supplementary Table S1; a table providing all RNA expression data is provided as Supplementary Table S2). The results were then analyzed in two ways. First, a cluster diagram was prepared that groups the samples into the most closely related subsets (Fig. 1A). We found that normal tissues all clustered together and the tumor tissues all clustered together. Thus, noncancerous tissue from a patient with HCC is more closely related to other normal tissue than to tumor from the same patient. Interestingly, purified normal hepatocytes and Huh-7 hepatoma cells showed RNA profiles quite different than the tissue samples. This is most likely due to the fact that only ~80% of the cells in the liver are hepatocytes; the liver also contains other cell types, such as Kupffer cells. Thus, in contrast to the purified hepatocytes and Huh7 cells, the gene expression profiles of the tissue reflect the transcriptome of both hepatocytes and Kupffer cells. As a second method of analysis, we identified RNAs that were deregulated in the two NAT (normal liver associated with a tumor) samples compared with liver from a patient who did not have liver cancer (normal) and compared with the two tumor samples from the matched patients (tumor). The numbers of mRNAs that showed increased or decreased levels in these pairwise comparisons are shown in Supplementary Table S3; the lists of genes resulting from each comparison are provided as Supplementary Tables S4–S21. We found that 2,324 and 1,587 mRNAs were up-regulated in the tumors of patient 1 and patient 2, respectively, with 86 genes in common in the two sets. We also found that 934 and 686 mRNAs were down-regulated in the tumors of patient 1 and patient 2, respectively, with 274 mRNAs in common in the two sets.

Using the program DAVID, we determined the functional classification of the set of up-regulated and down-regulated genes based on their Gene Ontology descriptions. In addition to classifying genes into different sets, the DAVID program provides a P value that indicates the probability that the set of genes was identified by chance based on the number of genes in the genome that fall into that particular category. The most highly represented categories and their P values for the dysregulated genes are shown in Fig. 1B and C. Interestingly, we found that ribonucleoproteins involved in protein synthesis were highly up-regulated in the liver tumors, whereas genes involved in liver functions (such as

---

**Footnotes:**

9 M. Birda et al., in preparation.

8 http://apps1.niaid.nih.gov/david/

10 http://david.abcc.ncifcrf.gov/
metallothionein) were down-regulated. Using more limited analyses, others have previously shown that ribosomal protein mRNAs are up-regulated (19) and that metallothionein is down-regulated (20) in HCC. Thus, we were confident that we had identified a high confidence set of mRNAs that are deregulated in HCC.

Our next goal was to determine the mechanisms by which these mRNAs are up- or down-regulated in an HCC. As discussed above, there are several different mechanisms by which the expression level of mRNAs can be controlled, including (a) changes in preinitiation complex formation (as monitored by levels of RNAPII bound to the promoter region); (b) changes in chromatin silencing (as monitored by levels of DNA methylation or modified histones, specifically H3me3K27 and H3me3K9); and (c) changes in gene copy number. We have investigated the role of these three mechanisms in development of a liver tumor using ChIP-chip technology; a summary of all the arrays used for the ChIP experiments is provided in Supplementary Table S1; the ChIP-chip data for RNAPII, H3me3K27, H3me3K9, and 5-meC is provided as Supplementary Table S22; and the lists of genes showing changes in binding of RNAPII, H3me3K27, H3me3K9, and 5-meC are provided as Supplementary Table S24 to S26.

**Reproducibility of ChIP-chip experiments.** Although ChIP assays are now a fairly standard experimental methodology, most studies have used cell lines grown in culture for the starting material. Because we are using small amounts of frozen tumor samples, we first modified the ChIP protocol for use with small numbers of cells (15) and then adapted the MicroChIP protocol for use with tumor samples (see Materials and Methods). After amplification (see Materials and Methods), the samples were hybridized to microarrays. For all the ChIP-chip experiments performed in this study, we used the NimbleGen two-array set that allowed analysis of ~5 kb regions (corresponding to ~4.2 kb upstream to 0.8 kb downstream of the transcription start sites) from ~25,000 promoters, with promoters being represented by ~50 probes spaced ~100 bp apart. After array normalization, the hybridization signals were divided by the total input signals to provide a fold-enrichment value for each 50-bp oligomer on the array. To identify binding sites on these promoter arrays, we used a modified Maxfour analysis program (see Materials and Methods). To show the reproducibility of our MicroChIP protocol for tissue samples, we performed two independent RNAPII and H3me3K27 ChIP-chip experiments using a matched set of normal and tumor liver samples. We found that these experiments led to highly reproducible results (Supplementary Fig. S2A and S2B). In addition, we found that, in accord with expectations (21), the silencing mark of H3me3K27 was almost entirely mutually exclusive with RNAPII in both normal and tumor samples (Supplementary Fig. S2C). Thus, we are confident that the
Identification of active promoters. As a first step toward characterizing the mechanisms responsible for changes in gene expression in the tumor samples, we compared levels of RNAPII bound to the promoter regions in matched normal and tumor liver samples from two different HCC patients. Using an antibody to RNAPII, ChIP samples were prepared from the four samples using between 7 and 15 μg of chromatin (4 × 10⁷–8 × 10⁷ cells). Amplicons were prepared from the ChIP samples and the quality of the amplicons was assessed by analysis of positive (the promoter for the largest subunit of RNAPII) and negative (ZNF44 and EVX1) control regions before array hybridization (see Supplementary Fig. S1). The amplicons were then labeled and hybridized to promoter arrays; promoters were ranked using the Maxfour analysis program (as described above); and the values for each promoter in the normal and tumor samples were compared. Comparison of repeated normal samples from the same patient revealed an excellent overall correlation. However, the slope of a linear fit could deviate significantly from 1, indicating that some arrays simply showed an overall stronger ratio signal than other arrays. We applied a simple procedure to estimate and compensate for intensity effects (see Materials and Methods for detailed description). To generate a generous list of promoters that were increased in tumor versus normal, we set a minimum difference between the tumor and normal samples and also a minimum enhancement value (see Materials and Methods). Examples of the hybridization data for binding of RNAPII are shown in Fig. 2A. The height of the columns indicates the fold enrichment of each oligomer from a particular promoter region in two sets of matched normal and tumor (two different patients) ChIP-chip experiments; examples shown represent cases in which binding is increased (or decreased) in both sets of matched samples. A summary of the changes of RNAPII binding in the normal versus tumor tissues is shown in Fig. 3A. Using our generous variables, we found that of the ∼25,000 promoters examined, ∼8% showed increased binding of RNAPII in the tumor and ∼5% showed decreased binding of RNAPII in the tumor. For patient 2, ∼9% showed increased binding of RNAPII and ∼14% showed decreased binding of RNAPII in the tumor sample. However, only 1% to 2% of the promoters showed consistent increases or decreases in binding in both tumor samples (compared with their matched normal samples).

Identification of silenced promoters. Changes in gene expression can also occur due to changes in chromatin structure. Regions of the genome bound by histone H3 trimethylated on lysine 9 or lysine 27 have been termed silenced chromatin, and regions of the genome bound by histone H3 trimethylated on lysine 9 or lysine 27 have been termed silenced chromatin, and regions of the genome whose compacted structure inhibits binding of site specific transcription factors and the general transcriptional machinery. To determine if changes in gene expression in the liver tumors was a consequence of changes in gene silencing, we performed ChIP assays using antibodies that specifically recognize H3me3K27 or H3me3K9. The ChIP samples were amplified and the quality of the amplicons was assessed by analysis of positive (EVX1 for H3me3K27 and ZNF44 for H3me3K9) and negative (the largest subunit of RNAPII) control regions (Supplementary Fig. S1). The amplicons were then labeled and hybridized to promoter arrays; promoters were ranked using the modified Maxfour analysis program; and the values for each promoter in the normal and tumor samples were compared as described above. Examples of the hybridization data for binding of H3me3K27 and H3me3K9 are shown in Fig. 2B and C and a summary of the changes is shown in Fig. 3A. We found that of the ∼25,000 promoters examined, ∼13% showed increased binding of H3me3K27 in the tumor and ∼13% showed decreased binding of H3me3K27 in the tumor. For patient 2, ∼6% showed increased binding of H3me3K27 and ∼6% showed decreased binding of H3me3K27 in the tumor sample. However, again, only ∼1% of the promoters showed consistent increases or decreases in H3me3K27 binding in both tumor samples (compared with their matched normal samples). We found even fewer changes in H3me3K9 binding. Of the ∼25,000 promoters examined, ∼8% showed increased binding of H3me3K9 in the tumor and ∼10% showed decreased binding of H3me3K9 in the tumor. For a second patient, ∼5% showed increased binding of H3me3K9 and ∼4% showed decreased binding of H3me3K9 in the tumor sample. However, less than ∼1% of the promoters showed consistent increases or decreases in H3me3K9 binding in both tumor samples (compared with their matched normal samples).

Although very few promoters showed consistent changes in binding of RNAPII, H3me3K27, or H3me3K9 in both of the patients, we could identify a small set of promoters that showed increased RNAPII and decreased H3me3K27 or H3me3K9 or decreased RNAPII and increased H3me3K27 or H3me3K9 in both of the patients. Interestingly, the sets of active genes that were silenced in the tumors (i.e., they showed a decrease in RNAPII binding and an increase in H3me3K27 and/or H3me3K9 binding) and the sets of silenced genes that were reactivated in the tumors (i.e., they showed an increase in RNAPII binding and a decrease in H3me3K27 and/or H3me3K9 binding) were highly enriched in specific gene ontology categories. Specifically, genes in the glycoprotein and transmembrane categories were silenced in the liver tumors, whereas genes in the categories of nuclear proteins, zinc fingers, transport, and coiled coil domains were reactivated in the tumors (Fig. 3B). A detailed ontology analysis of the genes regulated by consistent changes in binding of RNAPII, H3me3K27, and H3me3K9 in the tumor tissues is shown in Supplementary Fig. S3.

Analysis of copy number changes. Copy number variation is a hallmark of many cancers. Information on the prevalence and exact locations of these changes is of wide interest and may prove to be useful for differential diagnosis and, more fundamentally, may give insight into the differing patterns of various cancers. One widely used approach to assessing these changes is comparative genomic hybridization (CGH). Recently, high-density tiled arrays have been used with CGH to discover quite small changes in genomes (22). However, a high-resolution CGH analysis of the entire human genome would be an expensive undertaking, requiring 38 arrays (at a density of 380,000 probes per array) to 10 arrays (at a density of 2.1 million probes per array). In our ChIP-chip promoter array experiments, we apply both precipitated experimental sample and input genomic DNA to the arrays. Hence, we have genome-scale array data for genomic DNA from normal and tumor tissue. We reasoned that this genomic DNA data should reveal copy number changes in the tumor samples. In Supplementary Fig. S4, we show how use of the total genomic DNA channels from promoter ChIP-chip studies can be used to perform an in silico CGH experiment, an approach we term “virtual CGH” (vCGH). Briefly, we extract the hybridization signals from the input channel of the normal and the tumor ChIP-chip experiments. The data from the two normal (or two tumor) ChIP-chip experiments are quantile normalized and merged to produce a median value. We then divide the tumor signals by the normal signals (so that
amplified regions give values >0 and deleted regions give values <0), and then plot the resultant values using a windowing function; further details are provided in Materials and Methods.

To show the reproducibility of our vCGH analysis method, we performed two control experiments. First, we used our vCGH procedure on the two normal tissue samples from the cancer patients; this analysis led to essentially no differences (Fig. 4A). Second, we used the data from the two separate sets of ChIP-chip experiments performed on patient 1 (the ChIP experiments shown in Supplementary Fig. S2). Using the first set of input hybridization data from promoter array 1 (representing chromosomes 1–10), comparison of the normal versus tumor tissue identified a complex...
pattern of copy number differences. For example, we identified increased copies of promoters located at 1q, 6p, and 8q (Fig. 4B) and throughout the entire X chromosome (see below). We also observed loss of promoters at 6q and 8p (Fig. 4B). Importantly, a vCGH analysis using ChIP-chip data from a completely separate set of experiments for the same patient showed identical patterns (Fig. 4C). For a second patient (patient 2), the differences were weaker (perhaps due to contaminating normal tissue in the tumor sample), but the general pattern was observed (Fig. 4D). We confirmed the loss or gain of several of the identified regions using PCR analysis of DNA prepared from normal and tumor samples (Supplementary Fig. S5). Others have seen similar gains and losses of these chromosomal regions in HCC (23, 24). For example, we observe gains of 1q, 6p, and 8q; these same regions have been reported as copy number gains in 59%, 24%, and 38% of HCC. Similarly, we observe losses of 8p and 17p; these same regions have been previously reported to be lost in 41% and 49% of HCC.

In addition, our vCGH method can be applied to cell lines. In particular, vCGH can be used as a test of the “true identity” of a cell line. There has been an increasing realization that many cell lines are misidentified (25). A major concern is that many cell lines have been overtaken by HeLa cells, a hardy and fast-growing cell line that has been in culture for over 50 years. The contamination of other cell lines with HeLa cells does not necessarily occur in the investigators’ laboratory but rather the cell line may arrive from other sources already misidentified. Using input DNA tracks from ChIP-chip experiments, we compared HeLa to human ES cells. Our vCGH analysis of the HeLa cells showed a clear pattern of copy number changes (see Supplementary Fig. S6). We note that the well-characterized amplification of 5p in HeLa cells (26) was clearly seen in our vCGH analysis data. Having established the vCGH pattern of HeLa, we then used total DNA signals from ChIP-chip data from Ntera2 testicular embryonal carcinoma cells. The Ntera2 cells showed a distinct characteristic pattern that reproduced copy number changes that are often associated with testicular germ cell tumors (27), indicating that the tumor cell lines can be distinguished from HeLa cells using vCGH.

Mechanisms responsible for the tumor transcriptome. Having analyzed the sets of active and silenced promoters in normal and tumor cells and identified the promoters that show copy number changes in the tumor cells, we could return to the question posed earlier; that is, what mechanisms are responsible for mediating the changes in levels of RNAs in the HCC samples. Because the RNA and ChIP-chip data were performed using different platforms, it was first necessary to determine which of the deregulated RNAs had promoters represented on the NimbleGen arrays. We found that the promoters of 591 of the 865 mRNAs that were commonly up-regulated in both tumors and 195 of the 274 RNAs that were commonly down-regulated in both tumors were present on the NimbleGen promoter arrays. We found that the promoters of 591 of the 865 mRNAs that were commonly up-regulated in both tumors and 195 of the 274 RNAs that were commonly down-regulated in both tumors were present on the NimbleGen promoter arrays (Fig. 5A). Interestingly, increased binding of RNAPII occurred at only 5% and decreased levels of silenced chromatin (as monitored by decreases in either H3me3K27 or H3me3K9) occurred at only 2% in the genes up-regulated in the tumors. Similarly, decreased RNAPII binding occurred at only 3% and increased silencing occurred at only 3% of the genes down-regulated in tumors (Fig. 5B). In contrast, 53% of the promoters of the up-regulated mRNAs and 33% of the promoters for the down-regulated mRNAs displayed changes in

Figure 3. Analysis of active versus silenced promoters. A, the numbers of promoters that show changes in binding of RNAPII, H3me3K7, or H3me3K9 in the normal versus tumor samples. B, gene ontology categories that show coordinate changes in RNAPII and H3me3K27 or H3me3K9 binding.
copy number. Although copy number changes did seem to be responsible for a large percentage of the observed changes in gene expression, we noted that 40% and 61% of the up- and down-regulated genes, respectively, were not regulated by any of the tested mechanisms. Another mechanism that has been previously associated with changes in gene expression in human tumors is DNA methylation. To determine if DNA methylation was responsible for regulation of a subset of the tumor transcriptome, we performed ChIP experiments using an antibody that recognizes 5-methyl cytosine to precipitate methylated DNA from normal and tumor tissue of patient 1 (see Supplementary Table S22 for the ChIP-chip results monitoring DNA methylation in the normal and tumor samples). We found that 2% of the increased RNAs and 12% of the decreased RNAs were regulated by changes in DNA methylation (Fig. 5B; Supplementary Table S26). Thus, DNA methylation was responsible for the down-regulation of a significant number of genes in the tumor sample. However, changes in copy number remained the most predominant of the tested mechanisms for regulation of the tumor transcriptome (Fig. 5B). In fact, mapping of the location of the deregulated mRNAs from patient 1 onto the vCGH pattern of this same patient clearly shows the correlation of increased mRNA levels with increased copy number and of decreased mRNA levels with loss of genomic regions (Fig. 6A). There are several regions that show a high density of up-regulated RNAs in regions that do not show increases in gene copy number. These RNAs must have been increased by other mechanisms. For example, the first half of chromosome 14 shows a non-copy number–related increase in up-regulated RNAs. In this region, there are 22 promoters that show an increase in RNAPII binding in the tumor sample (see Supplementary Table S23 for a list of the promoters that show changes in binding of RNAPII in the tumor samples). We also mapped the location of the mRNAs commonly deregulated in both tumor RNA samples onto the vCGH pattern of patient 1 (Fig. 6B). Clearly, many of the commonly up- and down-regulated mRNAs map to regions having copy number changes.

Discussion
The goal of our study was to determine the extent to which several commonly used mechanisms of transcription regulation contribute to gene expression changes that occur during the development of a human liver tumor. Accordingly, we began our studies by identifying a set of transcripts that were dysregulated in both of the two independent liver tumors. We then examined changes in transcription initiation complex formation, gene

Figure 4. Analysis of copy number changes in HCC samples. vCGH analysis was performed comparing the normal DNA from both patients (A), the tumor versus normal DNA from two separate ChIP-chip experiments for patient 1 (B and C), and the tumor versus normal DNA for patient 2 (D). The Y axes indicate the log 2 ratio of the tumor to normal signals. Values >0 represent copy number gains in the tumor and values <0 represent copy number losses in the tumor. B, arrows, amplified (chromosome 1q and 6p), deleted (chromosome 8p), or control (chromosome 5q) regions analyzed in Supplementary Fig. S6.
silencing, and gene copy number for these deregulated genes. We found that, of these mechanisms, amplification and deletion of chromosomal regions was most often used to confer changes in mRNA expression in the liver tumor.

Although our study was not meant to be a comprehensive analysis of RNA changes in HCC, it is critical to note that we did identify sets of deregulated genes that (a) fit into several interesting gene ontology categories and (b) correspond to genes identified in previous studies of tumors. For example, we found that the most commonly down-regulated functional categories of genes were those from the metallothionein (e.g., MT1A, MT1E, MT1G, MT1M, MT1X, and MT2A) and cytochrome P450 (e.g., CYP1A2, CYP2C9, CYP21, CYP3A5, and CYP3A7) families. This suggests that a dedifferentiation of the liver has occurred in the tumor cells. Similarly, Xu et al. (3) performed an analysis of ESTs from hepatitis B–positive HCC samples and found that genes involved in liver function were down-regulated. More specifically, Tao et al. (20) performed a direct analysis of the metallothionein family in HCC and found that many metallothionein family members were down-regulated and suggested that down-regulated expression of this family may be a marker for hepatocarcinogenesis. We found that one of the most commonly up-regulated categories of genes included ribonucleoproteins involved in protein synthesis and transport. Similarly, Kondoh et al. (19) used differential display to identify five cDNAs that were up-regulated in HCC; all of these corresponded to ribosomal protein mRNAs. Interestingly, they also showed that the expression of these ribosomal protein mRNAs was high in three HCC cells lines irrespective of the growth state of the cell and suggested that activation of these genes is an important manifestation of the HCC phenotype. The other main category of genes that we discovered to be up-regulated in the HCC samples was nuclear proteins (see Supplementary Fig. S7 for categorized list of the nuclear proteins). Of the 194 nuclear proteins that were up-regulated in the HCC samples, 25 are involved in chromosomal biology and biogenesis, 12 are involved in nucleocytoplasmic transport, 24 are involved in DNA replication and repair, 32 are involved in RNA metabolism, and 67 are transcriptional regulators. Of the 67 transcriptional regulators, 38 (~50%) fit into the zinc ion binding/zinc finger category (Supplementary Fig. S7). Several of the up-regulated transcription factors have been previously associated with cancers. For example, FoxQ1 has been shown to be overexpressed in pancreatic cancer (28); Sox18 is up-regulated in gastric, pancreatic, breast, and embryonal tumor cell lines (29); and DP1 has been previously shown to be increased in expression in HCCs (30, 31).

As noted above, gankyrin mRNA is overexpressed in most HCC samples (9, 10); we also observed increased gankyrin mRNA in both of the HCC samples we analyzed. Detailed analysis of the gankyrin gene revealed that it was not increased in the tumor samples due to changes in RNAPII, H3me3K27, or H3me3K9 binding. Rather, the gankyrin gene is amplified. In fact, the amplification ratio of the gankyrin gene was in the top 3% of all amplified promoters. Overexpression of gankyrin can lead to degradation of p53 and Rb, contributing to genomic instability and oncogenic transformation. Rb functions to suppress the E2F family, a key regulator of cell proliferation. Interestingly, we also observed an increase in levels of DP1, the obligate heterodimeric partner of E2F1–E2F6, in the tumor samples. Expression analysis indicates that E2F3 and E2F4 are

Figure 5. Overview of the mechanisms regulating the HCC transcriptome. A, RNA and ChIP-chip platform comparisons. The RNAs that were commonly deregulated in the HCC samples were mapped to the NimbleGen arrays and then the number of each of these deregulated mRNAs that show changes in binding of RNAPII, H3me3K27, or H3me3K9 was determined. B, a pie chart demonstrating the percentages of deregulated mRNAs resulting from copy number changes (vCGH); changes in binding of RNAPII, H3me3K27, or H3me3K9; or changes in DNA methylation are shown for the mRNAs up- and down-regulated in liver tumors.
expressed at similar levels in the normal and tumor tissues, whereas the other E2Fs are not expressed in any of the liver samples. The increased expression of DP1 and decreased levels of Rb protein (due to gankyrin) may synergize to allow enhanced E2F3- or E2F4-mediated activation of key proliferation-responsive genes in the liver tumors.

In summary, we found that changes in preinitiation complex formation and changes in gene silencing do not play major roles in the development of the tested liver tumors. Rather, many gene expression changes in the tumors were a direct result of changes in gene copy number. We also note that 38% and 49% of the up- and down-regulated genes, respectively, were not regulated by any of the vCGH pattern.
the mechanisms. Recent estimates suggest that ~50% of human genes have alternative promoters (32). The set of alternative promoters for human genes is not well characterized and thus these promoters are not well represented on the arrays. Therefore, it is possible that some changes in RNA levels in the liver tumors were due to differential usage of novel alternative promoters. It is also possible that changes in transcription elongation and/or transcript stability contribute to the regulation of the set of genes in the “other” category.

Although our analysis was limited to a small number of liver tumors, our results suggest that clinical treatments with chromatin-modifying drugs, which have shown promise in the treatment of cancers of hematologic origin, may not be effective for HCC (see ref. 12 for a review of the clinical applications of epigenetic drugs). The mechanisms regulating the tumor transcriptome may differ for different tumor types and therefore RNA levels, chromatin structure, and gene copy number should all be analyzed to provide the required information to allow a rationale design of tumor type-specific chemotherapeutic regimens. Finally, we have outlined an efficient and cost-effective method to analyze copy number changes at all the known promoters regions in the human genome; this analysis can be performed at no extra cost using the data from ChIP-chip experiments.

Acknowledgments


Grant support: USPHS grants CA142250 and DK067889 (P.J. Farnham), and National Science Foundation grant 0225674 (M. Bieda).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank the members of the Farnham laboratory for helpful discussion. The DNA was labeled and the Illumina beadchips were processed and scanned at the University of California Davis Genome Center Expression Analysis Core Facility. The ChIP-chip and RNA microarray data have been deposited as GEO series GSE 10842 and GSE 10504.

References