Materials and Methods

All patients participating in this study gave informed consent prior to surgery. All tissues were surgically resected, snap frozen in liquid nitrogen within a half hour after the resection, and stored at –80° C. In most of the cases, both tumor and adjacent non-tumor tissues were collected. Total RNA was extracted with RNeasy kit (Qiagen) and mRNA was isolated from total RNA using FastTrack (Invitrogen) or Poly(A)Pure (Ambion) mRNA purification kit. mRNA from cell lines was purified directly with FastTrack (Invitrogen) kit. We combined mRNA from the following cells, in equal quantities, to make the reference pool: HepG2, SNU398, SNU1, Jurkat, RPMI and CCD-1070SK.

Microarray procedure

23075 cDNA clones, representing about 17,400 genes, were mechanically printed onto treated glass microscope slides, as previously described(6) (http://cmgm.stanford.edu/pbrown/array.html). Approximately 18700 of the clones were obtained from Research Genetics and 4300 clones were obtained directly from CGAP (http://www.ncbi.nlm.nih.gov/ncicgap/) The hybridizations were performed as previously described(5). A detailed protocol is available at: http://cmgm.stanford.edu/pbrown/protocols/5_hyb_human.html. In brief, 2 micrograms of sample mRNA and 2 micrograms of reference mRNA were labeled with Cy5-dUTP and Cy3-dUTP (Amersham) respectively, using Reverse Transcriptase (GibcoBRL) for 2 hours at 42° C. The two labeled cDNA probes were separated from unincorporated nucleotides by filtration, mixed and hybridized to microarray at 65° C overnight. After hybridization, each microarray was washed with 2xSSC, 0.03%SDS for 5 minutes at 65° C, then with 1xSSC for 5 minutes and 0.1xSSC for 5 minutes, both at room temperature. The array was then scanned using GenePix 4000A microarray scanner (Axon

Instruments). Array CGH was performed as described(32) The detailed protocol is available at: http://cmgm.stanford.edu/pbrown/protocols/4_genomic.html

Data Analysis

Primary data collection and analysis were carried out using GenePix Pro 3.0 (Axon Instruments). Areas of the array with obvious blemishes were manually flagged and excluded from subsequent analysis. The raw data were deposited into Stanford Microarray Database(33) at: http://genome-www4.stanford.edu/MicroArray/SMD/index.html. All non-flagged array elements for which the fluorescent intensity in each channel was greater than 1.5 times the local background were considered well measured. Genes for which fewer than 75% of measurements across all the samples in this study met this standard were excluded from further analysis. We selected for further analysis genes whose expression level differed by at least three fold, in at least four samples, from their mean expression level across all samples. We applied a hierarchical clustering algorithm both to the genes and arrays using the Pearson correlation coefficient as the measure of similarity, and average linkage clustering, as described(10). The results were visualized and analyzed with TreeView (M. Eisen; http://rana.lbl.gov). We used two-sample t-statistics to identify genes that were differentially expressed in two sets of samples. The statistical significance of the differential expression of any gene was assessed by computing a p-value for each gene, representing the chance of observing a test statistic at least as large (in absolute value) as the value actually obtained. No specific parametric form was assumed for the distribution of the test statistics. To determine the p-value, we used a permutation procedure in which the class labels of the samples were permuted 500,000 times, and for each permutation, two-sample t-statistics were computed for each gene. The permutation p-value for a particular gene is the proportion of the permutations (out

of 500,000) in which the permuted test statistic exceeds the observed test statistic in absolute values. Any gene for which this p-value was less than 0.001 was considered to be differentially expressed. The corresponding “per-family Type 1 error rate, PFER”, that is, the expected number of false positives for such a multiple test procedure is PFER = number of genes x 0.001. Alternatively, the nominal “false discovery rate, FDR”, or expected proportion of false positive among the genes declared differentially expressed, is FDR = PFER/number of genes declared expressed.

Immunohistochemistry

Immunohistochemistry was performed as described(7). Antibody MY10, for CD34, was used at 1:10 dilution (Becton & Dickinson). Antibody DO-7, for p53 ,was used at 1:100 dilution (Dako).

Southern Analysis

Genomic DNA was extracted using the DNeasy extraction kit (Qiagen), digested overnight with HindIII or EcoRI, and resolved by electrophoresis through a 1% agarose gel. After depurination, denaturation, and neutralization, the gel was transferred overnight with 10x SSC to a nylon Hybond N+ membrane (AP Biotech). HBV-specific sequences were detected by hybridization with a PCR-amplified copy of the complete HBV genome, from blood of a patient with HBV infection, labeled with fluorescein by random primed labeling (AP Biotech). Detection was performed with anti-fluorescein antibody conjugated with horseradish peroxidase, followed with ECL development (AP Biotech), and exposure to X-ray film (Kodak).