# Supplementary Material and Methods

**Processing of CGH data.** Genomic coordinates of RefSeq and UniGene in hg17 (http://genome.ucsc.edu) were annotated to both CGH and gene expression data. The expression levels of multiple tagged probes on the same gene locus were averaged. For the comparison of CGH and gene expression data, the genes with more than 4 missing values in the corresponding 15 HCC expression profile and the genes in sex chromosomes were excluded, and finally 9,981 genes were used.

**Quantitation of copy numbers of genomic DNA.** DNA copy numbers of genomic DNA samples were estimated by real-time qPCR PCR conditions were 95 °C for 10 min and 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. Each sample was assayed in triplicate and a control normal liver DNA was included in every assay. Each sample DNA copy number of interested genes was measured as $2^{-Ct}$ and normalized with that of the *GAPDH*, and the relative values compared to the normal liver sample were calculated. The sequences of the various primers are available upon request.

**Determination of the window size for the TCM.** TCM with different window sizes are constructed by selecting different number of neighboring genes. The threshold for each TCM is calculated based on 100 random simulation tests by randomly ordering the chromosomal locations of the 139 HCC dataset. The numbers of the genes above the given thresholds (*i.e.*, number of significant genes) for each TCM with different window size are plotted (data not shown). The window size of neighboring genes for TCM is determined to 24 which is the largest number of neighboring genes before the plateau.

**Validation of the prognostic classification of 50 or 30 corCNA genes in independent data sets.** To validate the robustness of the prognostic groups identified by 50 corCNA genes and 30 corCNA genes, class prediction was performed with two independent HCC datasets of SNU (n = 65) and GSE6764 (n = 65). All the SNU data set samples were comprised with HBV-related HCC, while GSE6764 dataset

samples were comprised with HCV-related HCC. Each data set was normalized by RMA method, and the GSE6764 dataset was further normalized by subtracting average expression levels of normal livers (n = 10). For the validation of GSE6764 dataset, preneoplastic samples with cirrhosis (n =17) and dysplasia (n =13) were excluded, and 35 HCC samples with early (n = 18), advanced (n = 7), and very advanced (n = 10) neoplastic stages were used. For the multiple tagged genes, the probe with the largest magnitude (i.e., sum of the squares of expression values in each sample) was selected as a representative probe. Class prediction of independent data sets was performed using 10-fold cross validation methods and six different class prediction algorithms, Compound Covariate Predictor (CCP), Linear Discriminant Analysis (LDA), Nearest Centroid (NC), k-Nearest Neighbor (1-NN, 3-NN), and Support Vector Machine (SVM) implemented in BRB-Array Tools (http://linus.nci.nih.gov/BRB-ArrayTools). Before applying class prediction algorithms, each dataset was standardized independently by transforming the expression of each gene to mean of 0 and standard deviation of 1. The predictors with the lowest cross-validation mis-classification rate in the train dataset were selected, and then applied to test datasets using six classification prediction algorithms. Since the survival information for SNU and GSE6764 datasets were not available, the class prediction rate in SNU dataset were estimated by log-rank tests on the patients' recurrence-free survival, and the accuracy for classifying the "very advanced" HCC samples (aa, n = 10) was used for GSE6764 dataset.

**Western blot analysis.** Protein extracts from siRNA treated cells were resolved by SDS-PAGE and Western blot analyses performed with antibodies for p70 S6 kinase (Cell Signaling), phospho-p70 S6 kinase (Thr421/Ser424, Cell Signaling), and β-actin (Chemicon). Immunoreactive proteins were detected by ECL system (GE Healthcare).