SUPPLEMENTAL MATERIAL.

Transcriptomic and Genomic analysis of Human Hepatocellular Carcinomas and Hepatoblastomas.

Jian-Hua Luo, Baoguo Ren, Sergei Keryanov, George C. Tseng#, Uma N.M. Rao, Satdarshan P. Monga, Steven Strom, Anthony J. Demetris, Michael Nalesnik, Yan P. Yu, Sarangarajan Ranganathan and George K. Michalopoulos##

Department of Pathology, School of Medicine, and Biostatistics (#), Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261

Materials and Methods

Sample preparation: Fresh liver tissues, recovered immediately from the operating room after removal, were dissected and trimmed to obtain pure tumor (completely free of normal hepatic tissue) or non-tumor liver tissues (free of tumor cells). Microdissection was coupled with sequential frozen and permanent section analyses to confirm the purity and homogeneity of the samples: gross and microscopic analyses were performed by pathologists. For tumor tissues, only samples with purity >90% were selected. For donor liver tissues, obtained at the time of organ-donation from brain-dead cadavers or live donors, samples free of any pathological alterations were selected, including cirrhosis or neoplastic and preneoplastic changes. For liver tissues adjacent to cancer, samples free of cancer cells, or any obvious dysplastic alterations, were selected. All tissues were processed and frozen within 30 minutes after removal. These tissues were then homogenized and subjected to analysis. All patients with hepatocellular carcinoma have at least a 5 years follow-up with regular evaluations for relapse or the presence of metastasis. Protocols for tissue banking, tissue anonymization and tissue processing were approved by institutional review board.

Affymetrix chip analysis: The overall analysis was conducted using Affymetrix chips u95av2, u95b and u95c.

cRNA preparation: Total RNA was extracted and purified with Qiagen RNeasy kit (Qiagen, San Diego, CA). Five micrograms of total RNA were used in the first strand cDNA synthesis with T7-d(T)$_{24}$ primer (GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(dT)$_{24}$) by Superscript^{TM} II (GIBCO-BRL, Rockville, MD). The second strand cDNA synthesis was carried out at 16°C by adding E. coli DNA ligase, E. coli DNA polymerase I and RnaseH in the reaction. This was followed by the addition of T4 DNA polymerase to blunt the ends of newly synthesized cDNA. The cDNA was purified through phenol/chloroform and ethanol precipitation. The purified cDNA were then incubated at 37°C for 4 hours in

an *in vitro* transcription reaction to produce cRNA labeled with biotin using MEGAscript^{TM} system (Ambion, Inc, Austin, TX).

Affymetrix chip hybridization: Fifteen to 20 µg of cRNA were fragmented by incubating in a buffer containing 200 mM Tris-acetate, pH 8.1, 500 mM KOAc, 150 mM MgOAc at 95°C for 35 minutes. The fragmented cRNA were then hybridized with a pre-equilibrated Affymetrix chip at 45°C for 14-16 hours. After the hybridization cocktails were removed, the chips were then washed in a fluidic station with low-stringency buffer (6x SSPE, 0.01% Tween 20, 0.005% antifoam) for 10 cycles (2 mixes/cycle) and stringent buffer (100 mM MES, 0.1 M NaCl and 0.01% Tween 20) for 4 cycles (15 mixes/cycle), and stained with SAPE (Strepto-avidin Phycoerythrin). This was followed by incubation with biotinylated mouse anti-avidin antibody, and restained with SAPE. The chips were scanned in a HP ChipScanner (Affymetrix Inc, Santa Clara, CA) to detect hybridization signals. For quality assurance, all samples were run on Affymetrix test-3 chips to evaluate the integrity of RNA, samples with RNA 3'/5' ratios less than 2.5 were accepted for further analysis.

Data Analysis:

Preprocessing and filtering:

Preprocessing including normalization and assessing expression intensity values were preformed in dChip (http://www.dchip.org) software and expression intensity measurements were obtained for 37891 genes. It was noticed that intensity measurements under 100 in Affymetrix array are usually not reliable in previous reports and genes with expression intensity measurement of less than 100 were excluded in the calculations (1, 2) . Such "filtering" resulted in 8398 remaining genes. We further filtered out genes with coefficient of variation (i.e. standard deviation divided by mean) less than 0.4. This resulted in 1496 genes used for further analysis. The expression values are then taken into logarithmic transformation (base 2). For sample hierarchical clustering, we performed a second filtering procedure to delete irrelevant genes. This follows previously published methodology (3), in which the tightest and most stable clusters are identified in a sequential manner through an analysis of the tendency of genes to be grouped together

under repeated re-sampling. Following this approach, a number of 419 genes in 15 clusters with tightly correlated patterns were identified. The remaining 1077 genes had a scattered distribution and low correlation with any other groups of genes. We thus filtered out the 1077 scattered genes and perform sample hierarchical clustering with the 419 tightly clustered genes.

Detecting differentially expressed genes:

The software SAM (Significance Analysis for Microarrays) (4)) was used for detecting the number of differentially expressed genes required for definitive separation of two categories of tissues. The method applied a more robust version of t-statistics and utilized permutation analysis to assess the statistical significance while maintaining the inter-gene correlation structure. The multiple comparison problem (namely thousands of genes are simultaneously compared at the same time) was considered and integrated in the analysis (FDR controlled at 5%). The software is available from http://www-stat.stanford.edu/~tibs/SAM .

Classification analysis:

A nearest shrunken centroid methods (4) developed for microarray sample classification is used for developing and evaluating classification models that distinguishes normal vs. tumor groups, normal vs. adjacent groups, tumor vs. adjacent groups, and simultaneously classifying three groups. The method is modified from traditional nearest centroid classification method and in contrast to other traditional methods the gene selection procedure is embedded in the training of the classification model. By introducing the shrinkage threshold, random genes that contribute no discriminant power but induce random noises in the classification models are eliminated. Varying the shrinkage threshold will result in different number of genes involved in the classification model and different cross validation error rate. Ten-fold cross validation is applied to avoid over-fitting in error rate evaluation. More details of the method is described in the supplement material. We applied the software PAM (Prediction Analysis for Microarrays) publicly available from http://www-stat.stanford.edu/~tibs/PAM/ to implement the method in our paper.

Cluster analysis:

Cluster analysis can be used to identify groups of genes (co-regulated genes) or groups of samples (subtypes of tumor) with similar expression pattern. Tight clustering (3) for clustering genes was applied as an unsupervised filtering procedure, as discussed above. The clustered genes are then used for hierarchical clustering of samples. Average linkage and Euclidean distance are used for hierarchical clustering throughout this paper.

Statistical analysis of differentially expressed individual genes between tumors, adjacent tissue and donor livers.

Data for intensity expression of genes (based on analysis by algorithms provided by Affymetrix) in all tissue cases were assembled on a single Excel spreadsheet. Each column was a separate tissue sample and each row represented one gene. Comparison between different groups was done by statistical analysis of the difference of the means, using non-paired t-test. In further comparisons between different groups, data were sorted to eliminate genes in which the difference of the means between the compared groups was not statistically significant. To determine statistical significance, Benjamini-Hochberg procedure (5) is applied to account for multiple comparison (FDR=5%) and to prevent spurious appearance of statistical significance often occurring when a p-value of 0.05 is used across the board. Genes with p-value higher than that defined by the Benjamini-Hochberg procedure were not included in the calculations or discussed separately. The genes with statistically significant differences in expression were sorted by descending rank of the ratio of the mean between the two categories. This revealed the highest over-expressed and under-expressed genes for each comparison. Data are shown in Supplemental tables 1 to 8 and contain the ratio of the means, p-value of the difference of the means, and the gene identifiers (GenBank number, chromosome locus). EST sequences were not included in the tables.

Oligonucleotide genome array analysis.

Target DNA preparation. Genomic DNA of the cell lines were extracted with QIAamp DNA mini kit (QIAGEN, CA), digested with *KpnI* and *NdeI* (New England Biolab, Boston). Two micrograms of DNA were amplified and labeled through PCR, using 5'-end Cy5 labeled random 12 mer for 50 cycle of the following sequence: 95°C for 30 seconds, 28°C 5 minute, 45 °C 5 minutes, 72 °C 3 minutes. To remove the residual primers and dNTP, the amplified DNA was purified through microcon YM-10 (Millipore, MA). The procedure followed the manufacturer's recommendation. To evaluate the completeness of deamination of cytosine, a PCR using primers corresponding to non-CpG island region of human β-actin was performed. The signal intensities of the wild type and the deaminated version of the same region were evaluated. It appeared that only at the pH 5.0 to 5.4, high ratios of deamination/wild type signals were achieved. The experiments of this set of data were performed at the condition of pH 5.0, a condition we considered optimal.

Oligonucleotide genome array analysis.

Target DNA preparation. Genomic DNA of the cell lines were extracted with QIAamp DNA mini kit (QIAGEN, CA), digested with *KpnI* and *NdeI* (New England Biolab, Boston). Two micrograms of DNA were amplified and labeled through PCR, using 5'-end Cy5 labeled random 12 mer for 50 cycle of the following sequence: 95°C for 30 seconds, 28°C 5 minute, 45 °C 5 minutes, 72 °C 3 minutes. To remove the residual primers and dNTP, the amplified DNA was purified through microcon YM-10 (Millipore, MA). The procedure followed the manufacturer's recommendation.

Oligonucleotide array preparation. The GOLD SEAL micro slides were pretreated in 10%NaOH and 60% ethanol for 1 hour. The slides were rinsed with deionized water thoroughly to remove the residual NaOH. The cleaned slides were then transferred to poly-l-lysine solution (25 ml poly-l-lysine (SIGMA), 25 ml PBS, 200 ml deionized water) for coating. After rinsing once with water, poly-l-lysine coated slides were printed with oligo probes in an Affymetrix GMS-417 arrayer. Four thousand 21-23 base oligonucleotides (obtained from MWG biotech, Inc) corresponding to the last exons of 4000 genes and EST were titrated in 3xSSC buffer in a final concentration of 50 pmol/μl. These probes were printed on poly-l-lysine coated slides in triplicates with

linear pattern by GMS-417 arrayer. Oligonucleotides corresponding to λ phage DNA were used as negative controls, printed at the end of the slides. Cy5, Cy3 or Hex labeled oligo were also printed at the end of the array in order to track the quality of printing and cross linking in the preparation. DNA cross-linking on the slides was carried out by baking the slides at 50°C overnight and an additional 60 mj of UV irradiation in StrataLinker 1800 (STRATAGENE). Unbound probes were washed by l-methyl-2-pyrrolindinone solution (60 mM sodium borate, 160 mM succinic anhydride, SIGMA) and water.

Target DNA Hybridization and data analysis. Fifteen microliters of Cy5 labeled genomic DNA in 1x hybridization buffer were hybridized with the array in a slide hybridization chamber (Fisher biotech) at 45°C overnight in a dark and humid incubator. Post-hybridization wash was carried out by a series of washes, 0.01% SDS in 2XSSC, 1xSSC, and 0.1x SSC. Slides were then scanned with a GMS-428 array scanner (AFFYMETRIX, CA). The scanning resolution was set at 10 μm. The auto gain setting determined the signal strength by selecting an appropriate gain setting based on the intensity range of the image. Jaguar 2.0 software (AFFYMETRIX) was used to compute the background signal and background subtraction. The algorithm threshold is determined as 75 percentile of background + 1.5 X IQR background. Each duplicated set of data was obtained in the identical condition. To compare the data from array to array, the normalization was performed by factoring the pixel average intensity of each array to an arbitrary level of 10,000 units. The hybridization signals were analyzed with pair-wise Student T-test.

Immunohistochemistry staining and tissue micro-array analysis. Formalin-fixed and paraffin-embedded human liver tissues including HCC, normal liver samples and hepatic tissue samples from livers containing HCC were used. The tissue micro-array contains 484 samples of hepatic tissue, including 319 "benign" (normal liver or tissue adjacent to HCC) liver tissues and 165 HCC. For immunostaining, 4-micron thick sections of tissue array were cut and mounted on glass slides. The sections were heated at 60°C for 12 hours and de-paraffinized in xylene and ethanol. Antigen retrieval was performed using

25 mM sodium citrate buffer (pH9.0) at 90°C for 15 minutes, followed by treatment of 3% H$_2$O$_2$ to block endogenous peroxidase. The slides were incubated at room temperature for 2 hours with anti-GPC3 (BioMosaic, Inc) and anti-TIEG (Abnova, Inc)antibodies at 1:200 and 1:400 dilution, respectively. The sections were then incubated with horseradish peroxidase conjugated anti-mouse IgG for 30 minutes at room temperature. This was followed by incubating the section with 3,3' diaminobenzidine solution (DAKO, Carpentaria, CA) to develop staining color. Hematoxylin was used for counterstaining. The specificity of immunostaining were verified by incubating the similar slides without primary antibody.

1.   Yu YP, Landsittel D, Jing L, Nelson J, Ren B, Liu L, McDonald C, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol 2004;22:2790-2799.

2.   Luo JH, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, et al. Gene expression analysis of prostate cancers. Mol Carcinog 2002;33:25-35.

3.   Tseng GC, Wong WH. Tight clustering: a resampling-based approach for identifying stable and tight patterns in data. Biometrics 2005;61:10-16.

4.   Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99:6567-6572.

5.   Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811-818.

SUPPLEMENTAL TABLES.

SUPPLEMENTAL TABLE 1

GENES INCREASED IN HCC VS. NL

Gene Name	Chromosome locus	GenBank #	HCC/NL	HCC/NL p-value <0.0254
glypican 3	xq26.1	U50410	19.15	0.0000036175
interferon, alpha-inducible protein 27	14q32	AI991845	6.40	0.0000915099
Paternally expressed 10	7q21	AB028974	5.25	0.0055376577
collagen, type I, alpha 2	7q22.1	AI983633	4.81	0.0088750212
regulator of G-protein signalling 5	1q23.1	AA928698	4.15	0.0000000391
galectin-3 internal gene /// lectin, galactoside-binding, soluble, 3 (galectin 3)	14q21-q22	AB006780	4.01	0.0000000247
SPARC-like 1 (mast9, hevin)	4q22.1	X86693	3.69	0.0000082339
EGF-like repeats and discoidin I-like domains 3	5q14	AL047586	3.51	0.0009467139
metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	11q13.1	AW007987	3.46	0.0000002334
glutamine synthase	1q31	X59834	3.32	0.0001848689
lumican	12q21.3-q22	U21128	3.30	0.0001919501
metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	11q13.1	N78139	3.29	0.0000001593
osteopontin	4q21-q25	AF052124	3.27	0.0009783037
regulator of G-protein signalling 1	1q31	S59049	2.98	0.0039291163
ubiquitin D	6p21.3	AL031983	2.98	0.0001326005
collagen, type IV, alpha 1	13q34	M11315	2.97	0.0000053599
chemokine (C-X-C motif) ligand 10	4q21	X02530	2.96	0.0229160849
multiple C2-domains with two transmembrane regions 1	5q15	T03872	2.88	0.0017126823
chemokine (C-C motif) ligand 20	2q33-q37	U64197	2.74	0.0213926489
glutathione S-transferase A2	6p12.1	M16594	2.69	0.0002695450
interleukin 8	4q13-q21	M28130	2.57	0.0005650233
Kruppel-like factor 5 (intestinal)	13q22.1	AI815057	2.57	0.0000014563
Chondroitin sulfate proteoglycan 2 (versican)	5q14.3	X15998	2.56	0.0014551462
A kinase (PRKA) anchor protein (gravin) 12	6q24-q25	U81607	2.56	0.0057094836

chemokine (C-X-C motif) ligand 11	4q21.2	AF030514	2.55	0.0055464132
mitogen-activated protein kinase 8	10q11.22	L26318	2.54	0.0220696723
cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	D00408	2.53	0.0082907407
chemokine (C-C motif) ligand 8	17q11.2	Y16645	2.50	0.0221740334
collagen, type III, alpha 1	2q31	X06700	2.48	0.0007899668
chromosome 17 open reading frame 27	17q25.3	AI655728	2.48	0.0000089774
cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	D00408	2.48	0.0042930664
aldo-keto reductase family 1, member B10 (aldose reductase)	7q33	U37100	2.47	0.0005915918
CAP, adenylate cyclase-associated protein, 2 (yeast)	6p22.3	U02390	2.47	0.0000004375
WW domain binding protein 5	xq22.2	AI934584	2.45	0.0003147586
geminin, DNA replication inhibitor	6p22.2	AI803434	2.42	0.0029438978
transmembrane 4 superfamily member 3	12q14.1-q21.1	M35252	2.42	0.0002025513
hypothetical protein DKFZp434B044	16q24.1	AI535730	2.40	0.0000449826
caveolin 1, caveolae protein, 22kDa	7q31.1	AF070648	2.37	0.0018321930
nerve growth factor receptor (TNFRSF16) associated protein 1	xq22.2	AI879010	2.37	0.0000842820
chromosome 9 open reading frame 5	9q31	AI684486	2.36	0.0009557755
immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	AI564488	2.36	0.0215349518
major histocompatibility complex, class I, F	6p21.3	X17093	2.34	0.0001856766
ribosomal protein S6 kinase, 90kDa, polypeptide 3	xp22.2-p22.1	AI368015	2.32	0.0000003430
alpha-fetoprotein	4q11-q13	J00077	2.31	0.0195499222
organic solute transporter alpha	3q29	AA442232	2.31	0.0000599900
major histocompatibility complex, class II, DP beta 1	6p21.3	M83664	2.30	0.0000430268
lysozyme	12q15 /// 19q13.4	AF009005	2.29	0.0005193007
UDP glycosyltransferase 2 family, polypeptide B15	4q13	T50788	2.28	0.0006262232
chromosome 17 open reading frame 27	17q25.3	AW014019	2.28	0.0000302026
family with sequence similarity 49, member B	8q24.21	H24185	2.26	0.0000741508
immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	AI610470	2.26	0.0000000169

interferon, alpha-inducible protein (clone IFI-6-16)	1p35	U22970	2.26	0.0000043638
chemokine (C-X-C motif) receptor 4	2q21	L06797	2.25	0.0000237304
UDP-glucose ceramide glucosyltransferase	9q31	W72338	2.24	0.0000247090
interleukin 6	7p21	X04430	2.24	0.0179114114
vesicle amine transport protein 1 homolog (T californica)	17q21	AW016496	2.22	0.0018721846
chromosome 10 open reading frame 10	10q11.21	AB022718	2.20	0.0001772102
glutathione S-transferase	6p12.1	AI199811	2.19	0.0020480711
tumor protein p53 binding protein, 2	1q42.1	U58334	2.18	0.0000000302
actin, alpha 2, smooth muscle, aorta	10q23.3	X13839	2.18	0.0002984270
hypothetical protein DKFZp762C1112	8q21.3	AI422867	2.17	0.0011136122
transmembrane 4 superfamily member 1	3q21-q25	AI445461	2.17	0.0000313997
glutathione S-transferase A2 /// glutathione S-transferase A1	6p12.1	M16594	2.17	0.0076510444
hypothetical protein MGC11324	4q21.23	AI052530	2.16	0.0000482200
GTP binding protein overexpressed in skeletal muscle	8q13-q21	U10550	2.15	0.0000002304
chemokine (C-C motif) ligand 2	17q11.2-q21.1	M26683	2.15	0.0173077999
cell division cycle 2, G1 to S and G2 to M	10q21.1	Y00272	2.15	0.0008701684
thrombospondin 2	6q27	L12350	2.14	0.0002967788
armadillo repeat containing 1	8q13.1	AL110454	2.13	0.0001765297
cadherin 11, type 2, OB-cadherin (osteoblast)	16q22.1	D21255	2.13	0.0004255250
cyclic AMP phosphoprotein, 19 kD	15q21.2	AI525176	2.13	0.0000000108
regulator of G-protein signalling 2, 24kDa	1q31	L13463	2.12	0.0021340953
regulator of G-protein signalling 5	1q23.1	AB008109	2.11	0.0000002747
immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	M80899	2.11	0.0000000502
meningioma expressed antigen 5 (hyaluronidase)	10q24.1-q24.3	AI344277	2.10	0.0000027740
collagen, type V, alpha 2	2q14-q32	Y14690	2.10	0.0000048801
dUTP pyrophosphatase	15q15-q21.1	U31930	2.09	0.0000719365
CAP, adenylate cyclase-associated protein, 2 (yeast)	6p22.3	N90755	2.08	0.0000115830
solute carrier family 1 (glial high affinity glutamate transporter), member 2	11p13-p12	AW005111	2.08	0.0034482303
vimentin	10p13	M14144	2.08	0.0000134092

high-mobility group box 1	13q12	AI690476	2.07	0.0000014822
osteonectin	5q31.3-q32	J03040	2.07	0.0194326532
dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)	2q24.3	X60708	2.07	0.0000808056
ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	8q24.1	L35594	2.07	0.0001084712
tight junction protein 2 (zona occludens 2)	9q13-q21	AA161241	2.06	0.0136573064
Interferon regulatory factor 2 binding protein 2	1q42.3	AI819847	2.06	0.0000027116
eukaryotic translation elongation factor 1 alpha 2	20q13.3	L10340	2.06	0.0075766269
eukaryotic translation initiation factor 3, subunit 6 48kDa	8q22-q23	AW015717	2.05	0.0000528209
zinc finger motif enhancer binding protein 2	1p22.2	AA159179	2.04	0.0000432515
Microtubule-associated protein 2	2q34-q35	AA167714	2.04	0.0000232673
translocase of outer mitochondrial membrane 20 homolog (yeast)	1q42	D13641	2.04	0.0000002807
UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)	7q34-q36	AI985274	2.04	0.0000090862
arrestin domain containing 3	5q14.3	AA587245	2.03	0.0000057152
brother of CDO	3q13.2	AA628405	2.03	0.0000042158
WD repeat domain 26	1q42.12	AL042428	2.03	0.0000000269
Zinc finger protein 281	1q32.1	AA121673	2.03	0.0000008892
NAD(P)H dehydrogenase, quinone 1	16q22.1	M81600	2.02	0.0024662151
KIAA0888 protein	5q13.3	AB020695	2.02	0.0005463522
Homo sapiens, clone IMAGE:4096273, mRNA	---	AI887641	2.02	0.0000124322
interferon-induced protein with tetratricopeptide repeats 1	10q25-q26	M24594	2.02	0.0000239752
SRY (sex determining region Y)-box 4	6p22.3	AI535884	2.01	0.0000007228
tumor necrosis factor receptor superfamily, member 21	6p21.1-12.2	AI807277	2.00	0.0003400044
CK2 interacting protein 1; HQ0024c protein	1q21.2	AA524401	2.00	0.0000013204

OTHER GENES OF BIOLOGIC SIGNIFICANCE (0.045>P-VALUE>0.0254)				**P-VALUE<0.045**
pancreatitis-associated protein	2p12	L15533	5.91	0.0340178611
cholesterol 7-alpha hydroxylase (CYP7	8q11-q12	M93133	3.34	0.0315047386

lactate dehydrogenase B	12p12.2-p12.1	T03298	2.94	0.0377702426
hypothetical protein DKFZp434L142	4q32.1	AI913548	2.49	0.0441194645
carboxypeptidase E	4q32.3	X51405	2.49	0.0419248491
angiopoietin-like 1	1q25.2	AA621478	2.12	0.0436001729

SUPPLEMENTAL TABLE 2

GENES UNDEREXPRESSED IN HCC VS. NL

Gene Name	Chromosome locus	GenBank #	HCC/NL	HCC/NL p-value <0.0254
serum amyloid A1	11p15.1	AA829286	0.12	0.000000000000000
hypothetical protein dJ462O23.2	1p36.12-p35.1	U90918	0.12	0.000000000000053
myosin XVB, pseudogene	17q25.1	AI671062	0.14	0.000000001933547
chemokine (C-C motif) ligand 7	17q11.2-q12	X72308	0.20	0.000000000005214
ficolin (collagen/fibrinogen domain containing) 3 (Hakata antigen)	1p36.11	D88587	0.20	0.000000000427027
hypothetical protein LOC286440	xq22.3	AL034399	0.20	0.000000000095405
hepcidin antimicrobial peptide	19q13.1	AI937227	0.22	0.000000002731361
hypothetical protein MGC3130	17q21.31	AI003902	0.22	0.000000000000006
motile sperm domain containing 2	xp22.2	AI744083	0.24	0.000000000199388
C-reactive protein, pentraxin-related	1q21-q23	X56692	0.24	0.000000000000021
hypothetical protein FLJ14490	1p34.2	AI968937	0.25	0.000000110444537
hypothetical protein FLJ35036	3q23	AI928164	0.29	0.000000000000349
metallothionein 1G	16q13	J03910	0.30	0.000173664228547
complement component 9	5p14-p12	K02766	0.30	0.000000001958046
serum amyloid A4, constitutive	11p15.1-p14	M81349	0.31	0.000000000016444
leucine-rich alpha-2-glycoprotein 1	19p13.3	T61106	0.31	0.000000000041315
survival of motor neuron protein interacting protein 1	14q13	AF027150	0.32	0.000000000000009
phospholipase A2, group IIA (platelets, synovial fluid)	1p35	M22430	0.32	0.000045262446502
orosomucoid 1	9q31-q32	X02544	0.32	0.000000000000000
ankyrin repeat and BTB (POZ) domain containing 1	3q21	AI823581	0.32	0.000000000000153
small proline-rich protein 2B	1q21-q22	M21302	0.32	0.000000000000007
hypothetical protein LOC284412	19q13.13	AA454184	0.33	0.000000000000006

alpha-1 antiproteinase, antitrypsin, member 3	14q32.1	X68733	0.33	0.000000000000003
activating transcription factor 5	19q13.3	AB021663	0.33	0.000000000235743
GalNAc-T2	1q41-q42	NM_004481	0.33	0.000000658717489
complement component 1, r subcomponent-like	12p13.31	N91560	0.33	0.000000001166802
angiopoietin-like 2	9q34	AW006802	0.34	0.000000000000033
lipopolysaccharide binding protein	20q11.23-q12	AF013512	0.34	0.000000000278815
peroxiredoxin 4	xp22.11	U25182	0.34	0.000000000012940
Acetyl-Coenzyme A acetyltransferase 1 (acetoacetyl Coenzyme A thiolase)	11q22.3-q23.1	D10511	0.34	0.000000001361116
proline-rich protein BstNI subfamily 1	12p13.2	S80905	0.35	0.000000000001179
cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	9p21	L36844	0.36	0.000000003535510
apolipoprotein F	12q13.3	L27050	0.36	0.000028589546313
calneuron 1	7q11	R40920	0.36	0.000000119852557
fibrinogen, gamma polypeptide	4q28	AI568934	0.36	0.000000000001409
Phosphodiesterase 8B	5q13.3	N49949	0.36	0.000000057117872
fibrinogen-like 1	8p22-p21.3	D14446	0.38	0.000000000001485
Transmembrane 6 superfamily member 1	15q24-q26	AA411025	0.38	0.000000000006004
chromodomain helicase DNA binding protein 1-like	1q12	AI422099	0.39	0.000000000780052
tubby like protein 1	6p21.3	U82468	0.39	0.000000000000733
hypothetical protein FLJ14490	1p34.2	AL042399	0.39	0.000000003979694
serine protease inhibitor, Kazal type 1	5q32	AI961220	0.39	0.000984458555811
hypothetical protein FLJ21439	15q14	AA780049	0.39	0.000000000141688
solute carrier family 13 (sodium-dependent citrate transporter), member 5	17p13.1	U92029	0.39	0.000009558174031
LIM domain binding 3	10q22.3-q23.2	AJ133766	0.39	0.000000000000000
metallothionein 1B (functional)	16q13	M13485	0.40	0.000000002529257
Hypothetical protein MGC35361	7q22.1	W25633	0.41	0.000000000021584
metallothionein 1A (functional)	16q13	K01383	0.41	0.000000003858776
hypothetical protein FLJ21945	2p23.3	AI637685	0.41	0.000000000000114

coagulation factor IX (plasma thromboplastic component, Christmas disease, hemophilia B)	xq27.1-q27.2	K02402	0.41	0.000000023122134
metallothionein 1H	16q13	R93527	0.41	0.000000000010371
solute carrier family 39 (zinc transporter), member 14	8p21.3	D31887	0.42	0.000000000021736
serine hydroxymethyltransferase 1 (soluble)	17p11.2	L11931	0.42	0.000000000403428
metallothionein 1E (functional)	16q13	M10942	0.42	0.000000026305271
hemopexin	11p15.5-p15.4	M36803	0.42	0.000000000000228
hypothetical protein FLJ30990	2q31.2	AI753091	0.42	0.000000000000000
fibrinogen, A alpha polypeptide	4q28	M64982	0.42	0.000000000000001
protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65)	11q23.2	AW009884	0.42	0.000059593721753
autocrine motility factor receptor	16q21	M63175	0.42	0.000000041007995
metallothionein 1X	16q13	AA224832	0.43	0.000004119579961
Carboxypeptidase N, polypeptide 2, 83kD	8p23-p22	J05158	0.43	0.000000209552876
RAS, dexamethasone-induced 1	17p11.2	AI968384	0.43	0.000129436187925
Male sterility domain containing 1	12p11.22	H16791	0.43	0.000005557969861
haptoglobin	16q22.1	X00442	0.43	0.000000000000000
protocadherin 8	13q14.3-q21.1	AF061573	0.43	0.000118170471666
chromosome 18 open reading frame 10	18q12.2	AA143321	0.43	0.000000000672585
Cell division cycle and apoptosis regulator 1	10q21.3	W73136	0.43	0.000000000000822
ferrochelatase (protoporphyria)	18q21.3	D00726	0.43	0.000000000110240
STAT3	17q21.31	AA057753	0.44	0.000000047075189
Hermansky-Pudlak syndrome 3	3q24	AA219354	0.44	0.000000013449438
transcriptional regulating factor 1	6p21.1-p12.1	AW006022	0.44	0.000000000000149
metallothionein 1F (functional)	16q13	H68340	0.44	0.000000000073106
protein kinase C, epsilon	2p21	AA039557	0.44	0.000419388468108
UDP glycosyltransferase 2 family, polypeptide B15	4q13	U08854	0.44	0.000000000280543
kelch-like 5 (Drosophila)	4p14	AA806103	0.45	0.000122943416737
brain protein 44-like	6q27	AI749098	0.45	0.000000004326065

Hermansky-Pudlak syndrome 5	11p14	AB023234	0.45	0.000009255623340
IQ motif containing F3	3p21.2	AA398133	0.45	0.000000063715203
complement factor H-related 4	1q32	X98337	0.46	0.002269653901947
corticotropin releasing hormone binding protein	5q11.2-q13.3	X58022	0.46	0.000001135915625
ceruloplasmin (ferroxidase)	3q23-q25	M13699	0.46	0.000000000276892
alanine-glyoxylate aminotransferase 2	5p13	N65959	0.46	0.000021687906929
phosphoinositide-3-kinase adaptor protein 1	10q24.1	AI762879	0.46	0.000076156952664
regulator of G-protein signalling like 1	1q25	AA757244	0.47	0.000000364908218
Potassium voltage-gated channel, KQT-like subfamily, member 1	11p15	AA029517	0.47	0.000000000037555
hypothetical protein FLJ12484	15q26.1	AI039246	0.47	0.000000000140227
hypothetical protein FLJ32452	12q13.3	T92427	0.47	0.000000000000080
lactate dehydrogenase A	11p15.4	X02152	0.47	0.000000019952819
chromosome 10 open reading frame 104	10q22.1	AA034078	0.48	0.000000002091168
complement factor H-related 3	1q32	X68679	0.48	0.002358214732103
IQ motif and Sec7 domain 1	3p25.2	AB018306	0.48	0.000000000023987
metallothionein 3 (growth inhibitory factor (neurotrophic))	16q13	M93311	0.48	0.000000000432174
Syntaxin 17	9q31.1	AA600810	0.48	0.000000724345388
phosphodiesterase 3B, cGMP-inhibited	11p15.1	AI651372	0.48	0.000000000166674
hydroxyacid oxidase (glycolate oxidase) 1	20p12	AB024079	0.48	0.000000000018111
thioredoxin domain containing 7 (protein disulfide isomerase)	2p25.1	D49489	0.49	0.000000000002609
aquaporin 3	9p13	N74607	0.49	0.000000001494900
tumor rejection antigen (gp96) 1	12q24.2-q24.3	X15187	0.49	0.000000779827620
insulin-like growth factor 1 (IGF1)	12q22-q23	X57025	0.49	0.000015035416112
complement component 4 binding protein, beta	1q32	L11245	0.49	0.000000000519597
nicotinamide N-methyltransferase	11q23.1	U08021	0.49	0.000000004839254
heat shock 70kDa protein 6 (HSP70B')	1q23	X51757	0.49	0.000000000002860
solute carrier family 17 (sodium phosphate), member 2	6p21.3	U90544	0.49	0.000000542372879

hypothetical protein FLJ11712	13q14.3	AA181196	0.49	0.000000001581265
cytochrome P450, family 19, subfamily A, polypeptide 1	15q21.1	X13589	0.49	0.000000000108822
Prickle-like 2 (Drosophila)	3p14.1	AA412673	0.49	0.000000000006880
testis-specific kinase substrate	19q13.3	AL041339	0.49	0.000000000000000
surfeit 4	9q34.2	R26138	0.49	0.000002001823259
proline synthetase co-transcribed homolog (bacterial)	8p11.2	AB018566	0.49	0.000000000003264
phosphoserine aminotransferase 1	9q21.2	AI971498	0.49	0.000000085739025
ocular development-associated gene	7q21-q22	AI337901	0.50	0.000000000000004
TRIAD3 protein	7p22.1	W27535	0.50	0.000000000049396
alpha-2-glycoprotein 1, zinc	7q22.1	X59766	0.50	0.000000089786283
tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	8p23.1	AF082558	0.50	0.000000003727625
putative lymphocyte G0/G1 switch gene	1q32.2-q41	M69199	0.50	0.000352700769927
fibrinogen, gamma polypeptide	4q28	AI989422	0.50	0.000000000000000
stress-associated endoplasmic reticulum protein 1	3q25.1	AA887132	0.50	0.000006354912141
erythropoietin receptor	19p13.3-p13.2	X97671	0.51	0.000000000061953
hypothetical protein LOC253842	9q34.11	AA176289	0.51	0.000000010132654
gamma-glutamyltransferase-like activity 1	22q11.23	M64099	0.51	0.000000000000000
paired immunoglobin-like type 2 receptor beta	7q22.1	AI954159	0.51	0.000000000008891
chromosome 20 open reading frame 3	20p11.22-p11.21	AI032386	0.51	0.000000026777403
type I transmembrane C-type lectin receptor DCL-1	2q24.2	D14664	0.51	0.000001933894422
isocitrate dehydrogenase 2 (NADP+), mitochondrial	15q26.1	X69433	0.51	0.000000003515240
solute carrier family 27 (fatty acid transporter), member 2	15q21.2	D88308	0.51	0.000003386440407
cell cycle progression 1	15q21.1	AI770067	0.51	0.000000071468934
solute carrier family 1, member 1	9p24	AI928365	0.51	0.000086704663365
superoxide dismutase 2, mitochondrial	6q25.3	X07834	0.51	0.000034071165644
glutaredoxin (thioltransferase)	5q14	X76648	0.51	0.000000246860457

SUPPLEMENTAL TABLE 3   GENES OVER-EXPRESSED IN HCC VS. AT

Gene Name	Chromosome locus	Unigene #	HCC/AT	HCC/AT P-value<0.00473
glypican 3	xq26.1	U50410	5.88	0.0000362681
collagen, type I, alpha 2	7q22.1	AI610692	3.39	0.0003167752
serine protease inhibitor, Kazal type 1	5q32	AI961220	2.32	0.0024584918
heat shock 27kDa protein 1	7q11.23	AI635057	2.20	0.0003581711
spondin 2, extracellular matrix protein	4p16.3	AW007426	2.11	0.0000379429
chromobox homolog 4 (Pc class homolog, Drosophila)	17q25.3	AI570531	2.03	0.0026225816
OTHER GENES OF BIOLOGIC SIGNIFICANCE (0.05>P-VALUE>0.00473)				**HCC/AT P-value<0.04**
Paternally expressed 10	7q21	AB028974	4.33	0.0082611345
EDIL-3 (EGF-like repeats and discoidin I-like domains 3)	5q14	AL047586	2.35	0.0070002213
multiple C2-domains with two transmembrane regions 1	5q15	T03872	2.27	0.0066686433
osteopontin	4q21-q25	AF052124	2.14	0.0141018017
osteonectin	5q31.3-q32	J03040	2.04	0.0216456200
alpha-fetoprotein	4q11-q13	J00077	2.00	0.0385721372

SUPPLEMENTAL TABLE 4 — GENES DECREASED IN HCC VS. ADJACENT TISSUE

Gene name	Chromosome locus	GenBank #	HCC/AT	HCC/AT P-value<0.00473
hypothetical protein MGC27165	14q32.33	S71043	0.33	0.0008780901
metallothionein 1G	16q13	J03910	0.37	0.0004223274
serum amyloid A1	11p15.1	AA829286	0.38	0.0017176571
immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	AJ001564	0.39	0.0043959232
ficolin (collagen/fibrinogen domain containing) 3 (Hakata antigen)	1p36.11	Hs.333383	0.40	0.0000272951
metallothionein 1X	16q13	AA224832	0.40	0.0000118953
immunoglobulin lambda constant 2 (Kern-Oz- marker)	22q11.2	J03011	0.40	0.0027603778
immunoglobulin kappa constant	2p12	M63438	0.41	0.0006603977
Immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	4q21	AI660656	0.41	0.0010205442
immunoglobulin lambda constant 2 (Kern-Oz- marker)	22q11.2	X79783	0.42	0.0034711767
hypothetical protein dJ462O23.2	1p36.12-p35.1	U90918	0.43	0.0042217231
butyrylcholinesterase	3q26.1-q26.2	M16541	0.44	0.0013549129
chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	4q21	U81234	0.45	0.0018071612
aldo-keto reductase family 1, member D1 (delta 4-3-ketosteroid-5-beta-reductase)	7q32-q33	Z28339	0.46	0.0008431942
cytochrome P450, family 2, subfamily A, polypeptide 6	19q13.2	X13930	0.48	0.0004069825
hemoglobin, beta	11p15.5	M25079	0.48	0.0006770709
butyrobetaine (gamma), 2-oxoglutarate dioxygenase (gamma-butyrobetaine hydroxylase) 1	11p14.2	AF082868	0.48	0.0008914462
interleukin 1 receptor accessory protein	3q28	AI189680	0.48	0.0005000738
solute carrier family 22 (organic cation transporter), member 1	6q26	X98332	0.48	0.0000043178
metallothionein 1B	16q13	M13485	0.49	0.0000019231
metallothionein 1E	16q13	M10942	0.49	0.0000070106

metallothionein 1A	16q13	K01383	0.49	0.0000044712
apolipoprotein F	12q13.3	L27050	0.49	0.0000293196
serum amyloid A4, constitutive	11p15.1-p14	M81349	0.50	0.0000009402
fructose-1,6-bisphosphatase 1	9q22.3	U21931	0.51	0.0000076049
tryptophan 2,3-dioxygenase	4q31-q32	U32989	0.52	0.0000242939
metallothionein 1H	16q13	R93527	0.52	0.0000016996
carbonic anhydrase II	8q22	J03037	0.54	0.0000042713
afamin	4q11-q13	L32140	0.54	0.0000649205
chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	10q11.1	U19495	0.54	0.0000706483
complement component 9	5p14-p12	K02766	0.55	0.0010604775
GADD45G	9q22.1-q22.2	AI952982	0.55	0.0005290806
pre-B-cell colony enhancing factor 1	7q22.3	U02020	0.55	0.0022326511
putative lymphocyte G0/G1 switch gene	1q32.2-q41	M69199	0.55	0.0017120131
alpha-2-glycoprotein 1, zinc	7q22.1	X59766	0.55	0.0000002134
orosomucoid 1	9q31-q32	X02544	0.55	0.0000000547
tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	8q24	AB008822	0.56	0.0034819907
hypothetical protein MGC3130	17q21.31	AI003902	0.56	0.0017847883
nicotinamide N-methyltransferase	11q23.1	U08021	0.56	0.0000254343
growth hormone receptor	5p13-p12	X06562	0.56	0.0010974508
metallothionein 1F (functional)	16q13	M10943	0.56	0.0000071957
mannose-binding lectin (protein C) 2, soluble (opsonic defect)	10q11.2-q21	X15422	0.56	0.0001612084
cytochrome P450, family 1, subfamily A, polypeptide 2	15q22-qter	M31667	0.56	0.0016336504
insulin-like growth factor binding protein 3	7p13-p12	M35878	0.57	0.0000138101

OTHER GENES OF BIOLOGIC SIGNIFICANCE (0.03>P-VALUE>0.00473)   **HCC/AT P-value<0.03**

immunoglobulin heavy locus /// immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	AI828404	0.39	0.0362948126
interleukin 8	4q13-q21	M28130	0.40	0.0077674876
immunoglobulin heavy constant gamma 1 (G1m marker)	14q32.33	AI564488	0.43	0.0173043713

immunoglobulin lambda constant 2 (Kern-Oz- marker)	22q11.2	AI872510	0.43	0.0276011471
myosin XVB, pseudogene	17q25.1	AI671062	0.45	0.0166134427
similar to bK246H3.1 (immunoglobulin lambda-like polypeptide 1, pre-B-cell specific)	22q11.23	AI932613	0.50	0.0061802509
chemokine (C-C motif) ligand 7	17q11.2-q12	X72308	0.51	0.0148936989
hepcidin antimicrobial peptide	19q13.1	AI937227	0.53	0.0156796619
PAI1	7q21.3-q22	M14083	0.54	0.0177070194
complement factor H-related 4	1q32	X98337	0.56	0.0157961419

SUPPLEMENTAL TABLE 5   GENES INCREASED IN HPBL VS. HCC

Gene Name	Chromosome locus	GenBank #	HPBL/HCC	HPBL/HCC P value <0.01069
mitogen-inducible gene 6	1p36.12-36.33	AI571452	3.88	0.0004427442
immediate early response 3 interacting protein 1	18q12	AI755071	3.40	0.0006995293
FLJ21963 protein	12q21.31	AI123815	3.18	0.0073770191
Pyrophosphatase (inorganic) 1 (PP)	---	AA592983	2.94	0.0001680994
PTPN13: Protein tyrosine phosphatase, non-receptor type 13	---	AI820966	2.83	0.0054976986
MALAT1: Metastasis associated lung adenocarcinoma transcript 1	---	AI935353	2.81	0.0001036320
hypothetical protein FLJ30596	5p13.2	AI962601	2.64	0.0000249897
transforming growth factor, beta 1	19q13.1	X02812	2.58	0.0000990358
zinc finger protein 204	6p21.3	AF033199	2.58	0.0082274144
CLTC: Clathrin, heavy polypeptide (Hc)	---	W90250	2.55	0.0086559249
ELL2: Elongation factor, RNA polymerase II,	---	H60379	2.47	0.0000203235
adipose differentiation-related protein	9p22.1	X97324	2.45	0.0087383662
ASXL2: Additional sex combs like 2 (Drosophila)	---	AA774820	2.44	0.0000037098
insulin-like growth factor binding protein 1	7p13-p12	M74587	2.44	0.0000694010
ATP-binding cassette, sub-family A (ABC1), member 1	9q31.1	AL048433	2.42	0.0057252055
Similar to golgin-67 isoform	---	AA281618	2.42	0.0047901900
tetratricopeptide repeat domain 16	9q34.11	R26574	2.37	0.0000000175
LOC399491 protein	16p13.1	AI984573	2.34	0.0044804702
cytochrome P450, family 3, subfamily A, polypeptide 4	7q21.1	D00003	2.25	0.0104921808
cytochrome P450, family 2, subfamily E, polypeptide 1	10q24.3-qter	J02625	2.25	0.0012746009
integral membrane protein 2B	13q14.3	AI499172	2.23	0.0022903305
elongation factor, RNA polymerase II, 2	5q15	C75510	2.17	0.0000007238
NFE2L2: Nuclear factor (erythroid-derived 2)-like 2	---	AW023229	2.15	0.0046690831
leukocyte cell-derived chemotaxin 2	5q31.1-q32	D63521	2.12	0.0048468010
KLF6: Kruppel-like factor 6	---	AI887641	2.11	0.0052141206
G antigen, family D, 3	xp11.22-p11.21	R33105	2.11	0.0039857902
BAX inhibitor 1	---	C06024	2.11	0.0012581437

ATPase, Ca++ transporting, plasma membrane 1	12q21.3	AI337321	2.10	0.0043151043
phospholipid scramblase 4	3q24	AW021496	2.09	0.0073092203
DNA polymerase-transactivated protein 6	2q33.1	AL040007	2.06	0.0039865566
ATPase family, AAA domain containing 2	8q24.13	AI139629	2.06	0.0000481250
ZBTB1: Zinc finger and BTB domain containing 1	---	AA001367	2.06	0.0001583248
formin binding protein 1	9q34	AA828277	2.05	0.0084224591
inter-alpha (globulin) inhibitor H2	10p15	X07173	2.04	0.0002306217
hypothetical protein LOC339047	16p12.3	AI954084	2.02	0.0058510966
SNRPN: Small nuclear ribonucleoprotein polypeptide N	---	AI149693	2.01	0.0006638551
chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	10q11.1	U19495	2.01	0.0070566924
FN1: Fibronectin 1	---	AI333596	2.00	0.0000058106
kelch-like 9 (Drosophila)	9p22	R61304	2.00	0.0008170343

OTHER GENES OF BIOLOGIC SIGNIFICANCE (0.045>P-VALUE>0.01069)

				P-VALUE<0.045
delta-like 1 (DLK1)	14q32.13	U15979	7.64	0.0115700000
insulin-like growth factor 2 (somatomedin A)	11p15.5	J03242	6.71	0.0435177239
Bicaudal D homolog 1 (Drosophila) (BICD1)	---	AA130857	3.87	0.0197199949
PRIM1: Primase, polypeptide 1	---	AA831707	3.13	0.0286859706
secreted phosphoprotein 2, 24kDa	2q37-qter	U20530	3.09	0.0120080512
ficolin (collagen/fibrinogen domain containing) 3 (Hakata antigen)	1p36.11	D88587	3.01	0.0374420967
cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	D00408	2.98	0.0419050791
Hermansky-Pudlak syndrome 3	3q24	AA219354	2.75	0.0277720427
v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	22q13.1	AL021977	2.52	0.0235371422
Similar to Kruppel-like factor 7 (ubiquitous); ubiquitous Kruppel-like transcription factor	3q26.2	T80868	2.48	0.0184916248
lipase, hepatic	15q21-q23	X07228	2.37	0.0208695829
cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	D00408	2.36	0.0392834721
Brain expressed X-linked-like 1	Xq22.1-q22.3	AL035494	2.30	0.0233413734
ras homolog gene family, member B	2p24	M12174	2.26	0.0421167559
FGL2: Fibrinogen-like 2	---	AI432401	2.19	0.0315095133

complement component 1, q subcomponent, gamma polypeptide	1p36.11	AI193100	2.18	0.0267366061
SPTBN1: Spectrin, beta, non-erythrocytic 1	---	T94862	2.13	0.0150509162
hypothetical protein DKFZp434P162	15q13.1	AI620911	2.09	0.0154767399
sex comb on midleg-like 2 (Drosophila) /// sex comb on midleg-like 1 (Drosophila)	xp22 /// xp22.2-p22.1	AW022735	2.08	0.0422693708
OTU domain, ubiquitin aldehyde binding 2	14q32.13	H10030	2.06	0.0279499182
hypothetical protein LOC149832	20p13	AW016356	2.05	0.0411846001
EIF2C2: Eukaryotic translation initiation factor 2C, 2	---	AI942410	2.02	0.0190942628
inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	2p25	AA557557	2.02	0.0127808932
afamin (New albumin gene 3' adjacent to the alpha 1-fetoprotein locus)	4q11-q13	L32140	2.02	0.0359937233

SUPPLEMENTAL TABLE 6

GENES DECREASED IN HPBL VS. HCC

Gene Name	Chromosome locus	GenBank #	HPBL/HCC	HPBL/HCC P value <0.01069
interferon, alpha-inducible protein 27	14q32	AI991845	0.23	0.0004606865
aldo-keto reductase family 1, member B10 (aldose reductase)	7q33	U37100	0.23	0.0000183474
chitinase 3-like 1 (cartilage glycoprotein-39)	1q32.1	Y08374	0.23	0.0089454398
lectin, galactoside-binding, soluble, 4 (galectin 4)	19q13.2	AB006781	0.24	0.0000023073
osteopontin	4q21-q25	J04765	0.25	0.0026621901
Hypothetical LOC389048	2q21.1	AA172056	0.33	0.0000198720
ADRA1A: Adrenergic, alpha-1A-, receptor	---	W73851	0.35	0.0104266175
Mature T-cell proliferation 1	Xq28	AI769948	0.35	0.0000000040
TRIM36: Tripartite motif-containing 36	---	AA398338	0.36	0.0000000098
cut-like 2 (Drosophila)	12q24.11-q24.12	AB006631	0.37	0.0000635269
Zinc finger protein 322B	9q22.33	AI971103	0.37	0.0000000089
eukaryotic translation elongation factor 1 alpha 2	20q13.3	L10340	0.37	0.0024320201
cytochrome P450, family 1, subfamily B, polypeptide 1	2p21	U03688	0.38	0.0004808591
hemochromatosis type 2 (juvenile)	1q21.1	AA948096	0.39	0.0015679671
RALBP1 associated Eps domain containing 1	---	H02904	0.41	0.0000000107
Zinc finger protein 291	15q24	N47807	0.41	0.0000000270
Transcribed locus, weakly similar to XP_375099.1 hypothetical protein LOC283585 [Homo sapiens]	---	AA040039	0.42	0.0016219506
ubiquilin (ubiquitin-like) 2	xp11.23-p11.1	AI190896	0.42	0.0000206437
Kinase interacting with leukemia-associated gene (stathmin)	1q23.3	AI765553	0.42	0.0005322896
EGLN1: Egl nine homolog 1 (C. elegans)	---	AI208803	0.42	0.0000000308
DnaJ (Hsp40) homolog, subfamily B, member 11	3q27.3	AI819362	0.42	0.0000000096
Immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	4q21	AI660656	0.43	0.0096696567
myosin, light polypeptide kinase	3q21	AA526844	0.43	0.0014652567
KLHL6: Kelch-like 6 (Drosophila)	---	AI024669	0.43	0.0000000380
UDP glycosyltransferase 2 family, polypeptide B17	4q13	U59209	0.44	0.0040834299

cytochrome P450, family 1, subfamily B, polypeptide 1	2p21	U03688	0.44	0.0013519903
transmembrane protein 14A	6p12.3	W84548	0.44	0.0004592319
glutathione peroxidase 2 (gastrointestinal)	14q24.1	X53463	0.44	0.0003670465
HEAT repeat containing 1	---	AI880842	0.44	0.0000003720
DEP domain containing 6	8q24.12	AI799804	0.44	0.0000000002
hypothetical protein DKFZp762C1112	8q21.3	AI422867	0.44	0.0024055656
PDE6A: Phosphodiesterase 6A, cGMP-specific, rod, alpha	---	C20707	0.45	0.0000003736
synaptotagmin VII	11q12-q13.1	AI554660	0.45	0.0005471215
MBNL2: Muscleblind-like 2 (Drosophila)	---	AA885123	0.45	0.0000022916
CDNA clone IMAGE:6167132, partial cds	---	N52198	0.45	0.0000000760
SOX17: SRY (sex determining region Y)-box 17	---	AI821669	0.45	0.0000063847
chromosome 17 open reading frame 27	17q25.3	AI655728	0.45	0.0002074674
PRO0149 protein	16p13.2	AW044607	0.46	0.0000000154
armadillo repeat containing 1	8q13.1	AL110454	0.46	0.0009731269
Microtubule-associated protein 2	2q34-q35	AA167714	0.46	0.0000731971
senescence downregulated leo1-like	15q21.2	AA074729	0.46	0.0000000235
transmembrane protein 16A	11q13.3	AI703454	0.46	0.0000059670
Transcribed locus, weakly similar to NP_079012.2 gasdermin domain containing 1 [Homo sapiens]	---	AA142976	0.47	0.0052556217
similar to CDC10	---	AI298966	0.47	0.0000024168
CDNA: FLJ23131 fis, clone LNG08502	---	W44633	0.47	0.0000127147
hypothetical protein FLJ23834	7q22.3	AA974855	0.47	0.0000003093
hypothetical protein LOC286359	9q22.33	AI027273	0.47	0.0000041186
LYRIC/3D3	8q22.1	AA398463	0.47	0.0000000002
geminin, DNA replication inhibitor	6p22.2	AI968057	0.47	0.0002215481
GPP34-related protein	1q21.2	AI804436	0.47	0.0000155005
Calpain 12	19q13.2	AI168605	0.48	0.0000000238
RAB GTPase activating protein 1-like	---	H48933	0.48	0.0000000192
mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme B	5q35	AI741591	0.48	0.0000220904
FLJ16030 protein	1p34.2	AI653297	0.48	0.0000000604
CDNA FLJ33492 fis, clone BRAMY2003681	---	AA398544	0.49	0.0000007227
membrane-associated protein 17	1p33	AI749525	0.49	0.0001184959

CDNA: FLJ22256 fis, clone HRC02860	---	R89028	0.49	0.0000001406
transmembrane 4 superfamily member 3	12q14.1-q21.1	M35252	0.49	0.0071245163
stromal cell-derived factor 2-like 1	22q11.21	AI927836	0.49	0.0000000197
NAD(P)H dehydrogenase, quinone 1	16q22.1	M81600	0.49	0.0046321066
PEX14: Peroxisomal biogenesis factor 14	---	AI291049	0.49	0.0000000301
hypothetical protein LOC284244	18q12.1	AF070541	0.49	0.0000011813
acid acyltransferase-epsilon	8p23.1	AA056538	0.49	0.0042168044
Immunoglobulin lambda constant 2 (Kern-Oz- marker)	22q11.2	AF043586	0.50	0.0000594687
UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 (GalNAc-T3)	2q24-q31	X92689	0.50	0.0000012623
interferon, alpha-inducible protein 27	14q32	X67325	0.50	0.0019156894
KIAA1598	10q25.3	AI744645	0.50	0.0000969389
Ropporin, rhophilin associated protein 1	3q21.1	AL079759	0.50	0.0000005776
ERO1-like (S. cerevisiae)	14q22.1	AI742422	0.50	0.0000308840
Zinc finger protein 322B	9q22.33	AA523657	0.50	0.0000006386
TRNT1: TRNA nucleotidyl transferase, CCA-adding, 1	---	R91427	0.50	0.0000000000

SUPPLEMENTAL TABLE 7

GENES INCREASED IN CLUSTER A VS. B

Gene Name	Chromosome locus	Unigene #	RATIO A/B	P VALUE A/B<0.00117
GABA transporter 1 (GAT1)	3p25-p24	AI003579	3.16	0.0000069406
apobec-1 complementation factor	10q11.23	T78476	2.72	0.0002318619
kininogen 1	3q27	M11437	2.57	0.0000140889
transcription elongation factor A (SII), 3	1p36.12	AA873235	2.51	0.0000199372
checkpoint suppressor 1	14q24.3-q32.11	U68723	2.49	0.0004609957
chromosome 10 open reading frame 58	10q22.3	AA127810	2.48	0.0011252955
Erb-B3	12q13	AI565773	2.35	0.0005635270
glucose-6-phosphatase	17q21	U01120	2.30	0.0000371943
Mitogen-activated protein kinase kinase kinase 2	2q14.3	AI125541	2.20	0.0006127848
acyl-Coenzyme A oxidase 1, palmitoyl	17q24-17q25	S69189	2.18	0.0005146575
ATP-binding cassette, sub-family C (CFTR/MRP), member 6	16p13.1	U66689	2.17	0.0000903006
bile acid Coenzyme A: amino acid N-acyltransferase (glycine N-choloyltransferase)	9q22.3	L34081	2.15	0.0002419904
cytochrome P450, family 3, subfamily A, polypeptide 5	7q21.1	J04813	2.14	0.0003990136
chromosome 10 open reading frame 58	10q22.3	AI589570	2.10	0.0000857447
ankyrin repeat domain 15	9p24.3	D79994	2.09	0.0009150277
pleckstrin and Sec7 domain containing 3	8pter-p23.3	AW015495	2.08	0.0005384379
aldehyde dehydrogenase 3 family, member A2	17p11.2	U46689	2.08	0.0006250162

OTHER GENES OF BIOLOGIC SIGNIFICANCE (0.04>P-VALUE>0.00117)				P-VALUE<0.02
cytochrome P450, family 7, subfamily A, polypeptide 1	8q11-q12	M93133	6.82	0.0217036793
glycerol-3-phosphate acyltransferase, mitochondrial	10q25.2	AI934361	3.94	0.0231947333
cytochrome P450, family 3, subfamily A, polypeptide 4	7q21.1	J04449	3.69	0.0040865500
Paternally expressed 10	7q21	AB028974	3.69	0.0352837896
glial high affinity glutamate transporter, member 2	11p13-p12	AW005111	2.98	0.0036559913
insulin receptor substrate 1 (IRS1)	2q36	S85963	2.82	0.0015252748

UDP glycosyltransferase 2 family, polypeptide B17	4q13	U59209	2.73	0.0093112841
LIM domain 7	13q22.2	AA100793	2.55	0.0044695161
glutamate-cysteine ligase, catalytic subunit	6p12	N47894	2.50	0.0012046468
Alcohol dehydrogenase IB (class I), beta polypeptide	4q21-q23	AA708155	2.47	0.0082192957
cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	D00408	2.47	0.0014214275
alcohol dehydrogenase 6 (class V)	4q23	X63359	2.39	0.0077438123
transmembrane protein 14A	6p12.3	W84548	2.37	0.0035002905
nuclear factor I/A	1p31.3-p31.2	AI750575	2.36	0.0018488175
phosphoribosyl pyrophosphate synthetase-associated protein 1	17q24-q25	D61391	2.35	0.0035949947
peroxisomal D3,D2-enoyl-CoA isomerase	6p24.3	AI564314	2.35	0.0012222229
PDZ domain containing 1	1q21	AF012281	2.33	0.0035335918
armadillo repeat containing 1	8q13.1	AL110454	2.33	0.0033871380
Zinc finger protein 281	1q32.1	AI796083	2.33	0.0043431974
UDP glycosyltransferase 2 family, polypeptide B17	4q13	U59209	2.33	0.0276288815
glutathione S-transferase A2	6p12.1	M16594	2.33	0.0080702781
ring finger protein 128	xq22.3	AI569980	2.25	0.0248967399
dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)	2q24.3	X60708	2.23	0.0060494370
Transcribed locus, weakly similar to XP_375099.1 hypothetical protein LOC283585 [Homo sapiens]	---	AA040039	2.22	0.0337114195
hemochromatosis type 2 (juvenile)	1q21.1	AA948096	2.21	0.0047648034
dicarbonyl/L-xylulose reductase	17q25.3	AI382182	2.21	0.0120478600
Immunoglobulin superfamily, member 4	11q23.2	AL080181	2.20	0.0022241441
NAD(P)H dehydrogenase, quinone 1	16q22.1	M81600	2.16	0.0328150487
emopamil binding protein-like	13q12-q13	AI858023	2.16	0.0039792506
3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	5p14-p13	AL048840	2.15	0.0056963187
metastasis suppressor 1	8p22	AB007889	2.15	0.0024933541
aldo-keto reductase family 1, member C4 (chlordecone reductase; 3-alpha hydroxysteroid dehydrogenase, type I; dihydrodiol dehydrogenase 4)	10p15-p14	M33375	2.13	0.0048617526
CDNA: FLJ22642 fis, clone HSI06970	---	AI700633	2.12	0.0068885540
microsomal triglyceride transfer protein (large polypeptide, 88kDa)	4q24	X91148	2.11	0.0019859427
aldehyde dehydrogenase 5 family, member A1 (succinate-semialdehyde dehydrogenase)	6p22.2-p22.3	AL031230	2.11	0.0179278309

nuclear ubiquitous casein kinase and cyclin-dependent kinase substrate	1q32.1	T62753	2.10	0.0042172952
sorbitol dehydrogenase	15q15.3	L29254	2.10	0.0014126633
3-hydroxysteroid epimerase	12q13	U89281	2.10	0.0052990662
glutamate-cysteine ligase, catalytic subunit	6p12	AI971137	2.10	0.0070197442
alcohol dehydrogenase 1A (class I), alpha polypeptide	4q21-q23	M12963	2.09	0.0376757055
chromosome 6 open reading frame 97	6q25.1	AI734267	2.09	0.0209332286
phosphoserine phosphatase	7p15.2-p15.1	Y10275	2.06	0.0244312186
leukocyte cell-derived chemotaxin 2	5q31.1-q32	D63521	2.05	0.0017713747
myotubularin related protein 4	17q22-q23	AB014547	2.05	0.0058549400
membrane-associated RING-CH protein VI	5p15.2	AB011169	2.02	0.0022853495
G antigen 8 /// G antigen 2 /// G antigen 1	xp11.4-p11.2 /// xp11.23	U19142	2.02	0.0083730353
WW domain containing E3 ubiquitin protein ligase 1	8q21	U96113	2.02	0.0061768211
Solute carrier family 25, member 29	14q32.2	AA610175	2.01	0.0365618757
interferon-induced protein with tetratricopeptide repeats 1	10q25-q26	M24594	2.01	0.0039989813
period homolog 3 (Drosophila)	1p36.23	AA161496	2.00	0.0033203108
chromosome 9 open reading frame 5	9q31	AI684486	2.00	0.0368756359
nuclear receptor subfamily 4, group A, member 3	9q22	U12767	2.00	0.0156033922

SUPPLEMENTAL TABLE 8

GENES DECREASED IN HCC
CLUSTER A VS. CLUSTER B

Gene Name	Chromosome locus	GenBank #	RATIO A/B	P VALUE<0.00117
C-reactive protein, pentraxin-related	1q21-q23	X56692	0.25	0.0010991291
TIMP-1	xp11.3-p11.23	D11139	0.28	0.0000000326
RAS, dexamethasone-induced 1	17p11.2	AI968384	0.30	0.0006040517
Fc fragment of IgG, low affinity IIa, receptor for (CD32)	1q23	M31932	0.32	0.0003886579
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	10p14-p15	AI765775	0.34	0.0005417389
hypothetical protein DKFZp434B044	16q24.1	AI535730	0.34	0.0002362676
DNA-damage-inducible transcript 4	10pter-q26.12	AW015428	0.34	0.0006275021
proteoglycan 1, secretory granule	10q22.1	X17042	0.38	0.0002084695
ADAM metallopeptidase with thrombospondin type 1 motif, 1	21q21.2	AI810627	0.39	0.0001785133
collagen, type IV, alpha 1	13q34	M11315	0.44	0.0002031832
interferon, gamma-inducible protein 16	1q22	M63838	0.44	0.0000312833
insulin-like growth factor binding protein 7	4q12	L19182	0.44	0.0005439604
endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	9q22.1-q22.2	AA149845	0.45	0.0003460972
UDP-glucose ceramide glucosyltransferase	9q31	W72338	0.46	0.0001634680
adrenomedullin	11p15.4	D14874	0.46	0.0007827728
transmembrane 4 superfamily member 1	3q21-q25	AI445461	0.46	0.0000478726
HIF1 alpha	14q21-q24	U22431	0.47	0.0001203476
GTP binding protein overexpressed in skeletal muscle	8q13-q21	U10550	0.47	0.0007195899
transmembrane 4 superfamily member 1	3q21-q25	M90657	0.50	0.0000445029
vimentin	10p13	M14144	0.50	0.0005944523
collagen, type VI, alpha 2	21q22.3	X15882	0.50	0.0000951248

OTHER GENES OF BIOLOGIC SIGNIFICANCE
 (0.04>P-VALUE>0.00117)   **P-VALUE<0.004**

Gene Name	Chromosome locus	GenBank #	RATIO A/B	P VALUE
chemokine (C-C motif) ligand 7	17q11.2-q12	X72308	0.18	0.0021677482

regulator of G-protein signalling 2, 24kDa	1q31	L13463	0.25	0.0029116812
suppressor of cytokine signaling 3	17q25.3	AI680350	0.25	0.0031724985
Hypothetical LOC387763	11p11.2	AW001427	0.32	0.0016098195
methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase	2p13.1	X16396	0.35	0.0020875143
SOCS 3 (suppressor of cytokine signaling 3)	17q25.3	AI244908	0.36	0.0039003029
transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	20q12	M55153	0.36	0.0026570994
jun B proto-oncogene	19p13.2	M29039	0.37	0.0020085329
thymosin, beta 10	2p11.2	AA417033	0.39	0.0025738484
complement component 7	5p13	J03507	0.39	0.0037438594
insulin-like growth factor binding protein 5	2q33-q36	L27560	0.39	0.0014035038
cold shock domain protein A	12p13.1	M24069	0.40	0.0013905954
immediate early response 3	6p21.3	S81914	0.40	0.0023629963
cytochrome P450, family 1, subfamily B, polypeptide 1	2p21	U03688	0.43	0.0019217561
DNA-damage-inducible transcript 4	10pter-q26.12	AA618602	0.44	0.0028931311
Growth arrest and DNA-damage-inducible, beta	19p13.3	AF078077	0.45	0.0018280021
phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila)	1p31	L20971	0.47	0.0020348715
nicotinamide N-methyltransferase	11q23.1	U08021	0.47	0.0014650268
v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	22q13.1	AL021977	0.47	0.0033264086
metallothionein 1H	16q13	R93527	0.48	0.0028008904
prostaglandin D2 synthase 21kDa (brain)	9q34.2-q34.3	M98539	0.49	0.0015819456
chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	4q21	X54489	0.50	0.0012834706