Supplementary information

Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model

Emilie A. Bard-Chapeau[1], Anh-Tuan Nguyen[1], Alistair G. Rust[2], Ahmed Sayadi[1], Philip Lee[3], Belinda Q Chua[1], Lee-Sun New[4], Johann de Jong[5], Jerrold M. Ward[1], Christopher KY. Chin[1], Valerie Chew[6], Han Chong Toh[7], Jean-Pierre Abastado[6], Touati Benoukraf[8], Richie Soong[8], Frederic A. Bard[1], Adam J. Dupuy[9], Randy L. Johnson[10], George K. Radda[3], Eric CY. Chan[4], Lodewyk FA.Wessels[5], David J. Adams[2], Nancy A. Jenkins[1,11*] and Neal G. Copeland[1,11*]
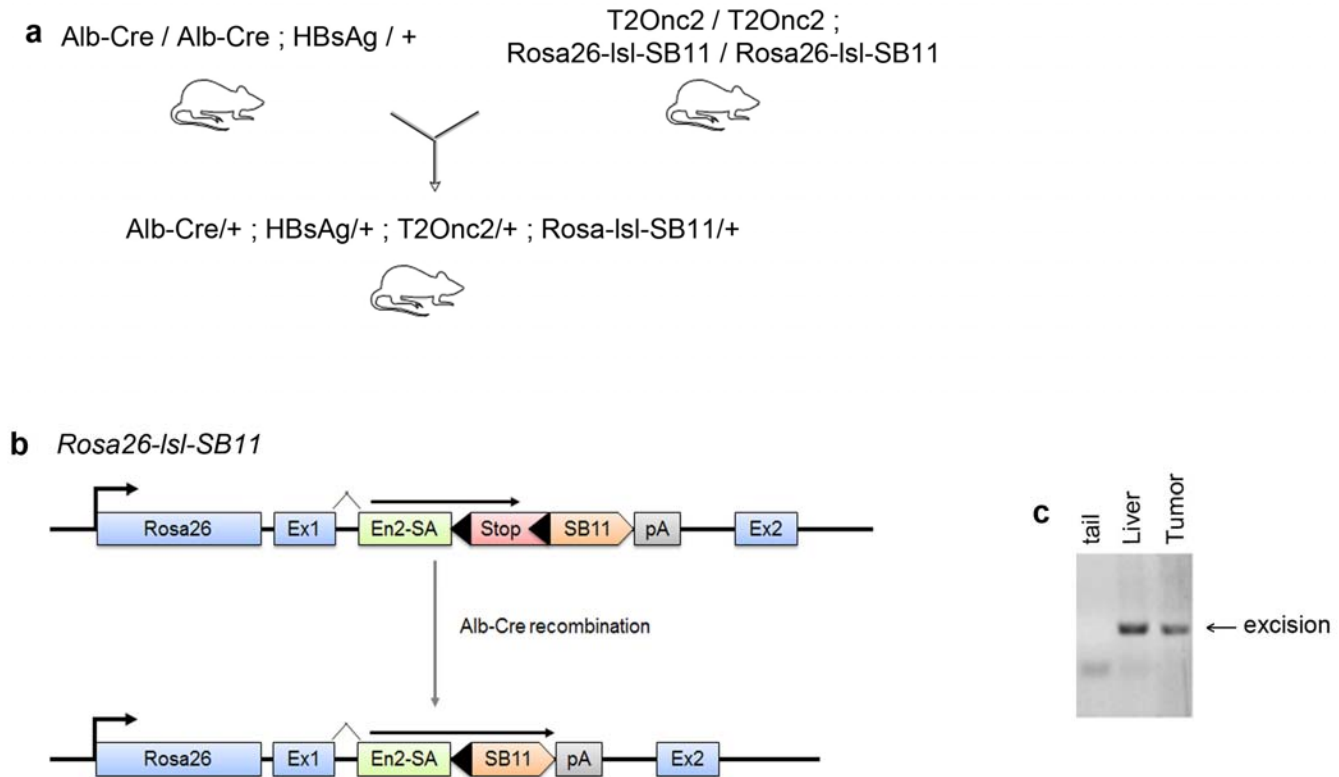
Supplementary Figures and Legends 1-19

Supplementary Tables 1-7

Supplementary Dataset 1

Supplementary References

Supplementary Figures and Legends

Supplementary Figure 1 | Generation of experimental animals used in the HCC transposon screen
a, Scheme of the genetic cross used to generate the Liver-SB/HBsAg animals. The T2/Onc2 (6113) SB transposon concatamer inserted in mouse chromosome 1 carries ~358 transposon copies[1]. High transposon copy number was important for obtaining insertions that mimic the genetic heterogeneity of HCC. The donor transposon concatamer is mobilized by SB transposase expression. The conditional transposase inserted in the Rosa26 locus, Rosa26-lsl-SB11[2], is expressed after Cre recombination. A Cre recombinase under control of albumin (Alb) promoter[3] was used to activate transposase selectively in hepatocytes. The HBsAg transgene was used to promote hepatic chronic inflammation and sensitize for HCC tumorigenesis. **b**, Schematic for the activation of Rosa26-lsl-SB11 by Alb-Cre recombination, specific to hepatocytes. **c**, Efficient mobilization of the transposon in non-tumorous liver tissue of a Liver-SB/HBsAg mouse (72.3 week-old) and in tumor tissue, but not in tail. PCR-based excision analysis, as previously described [4].
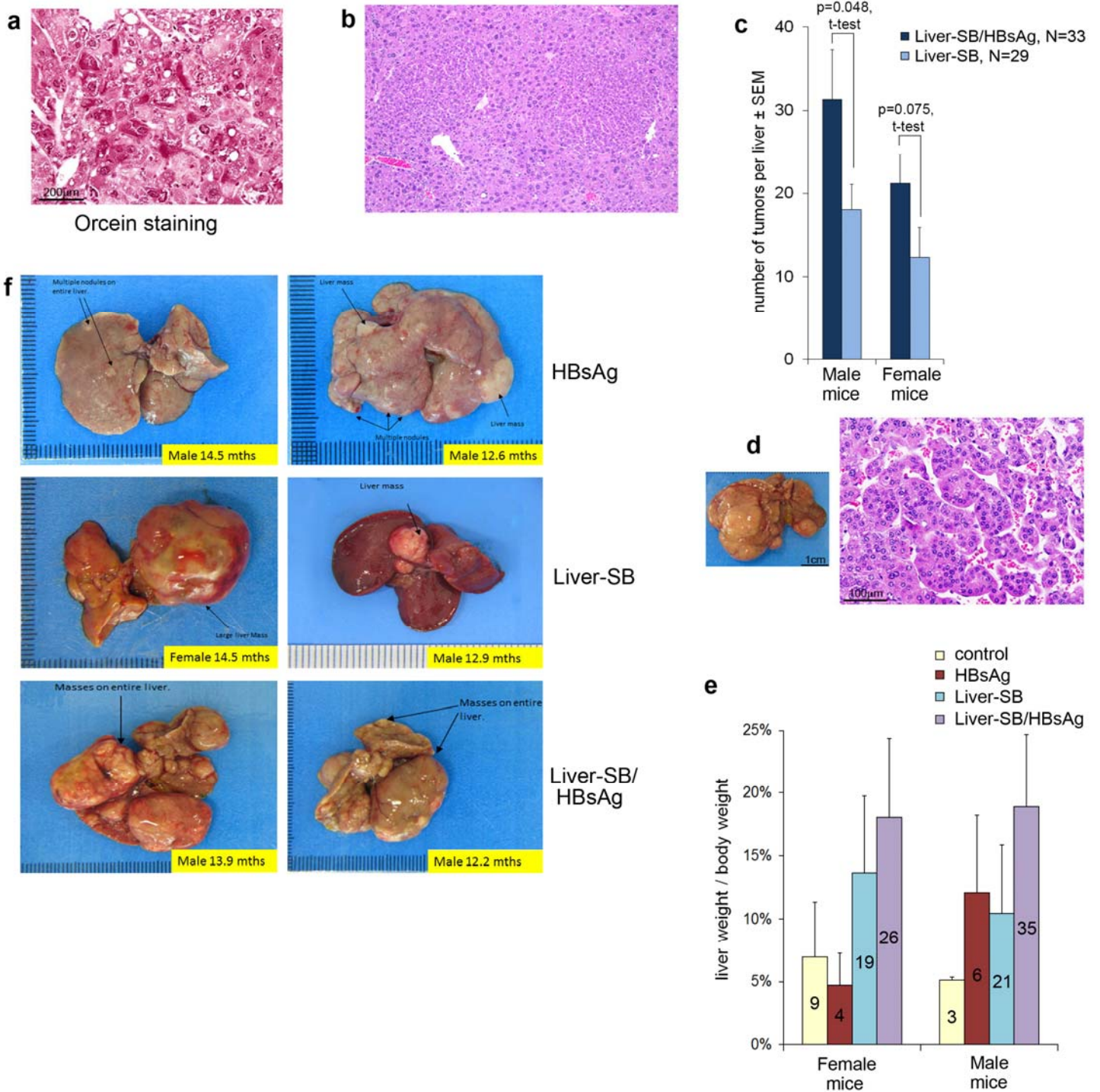
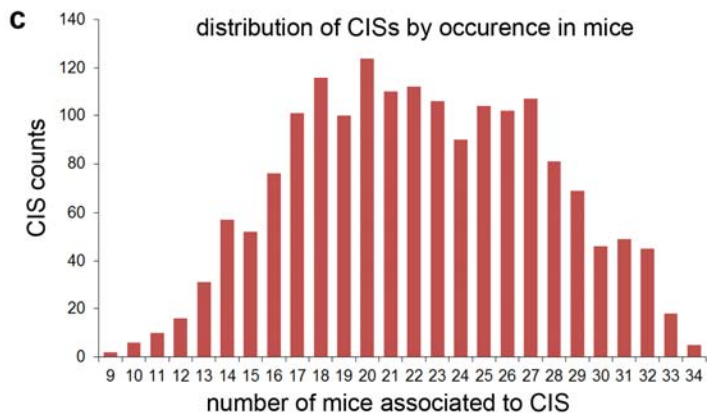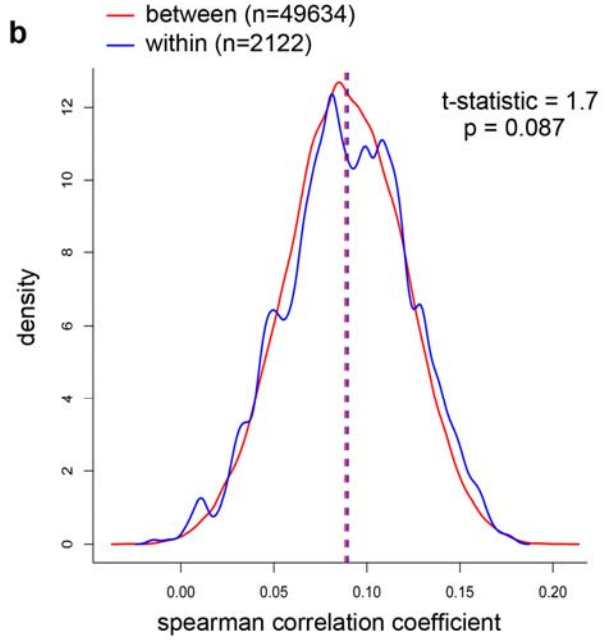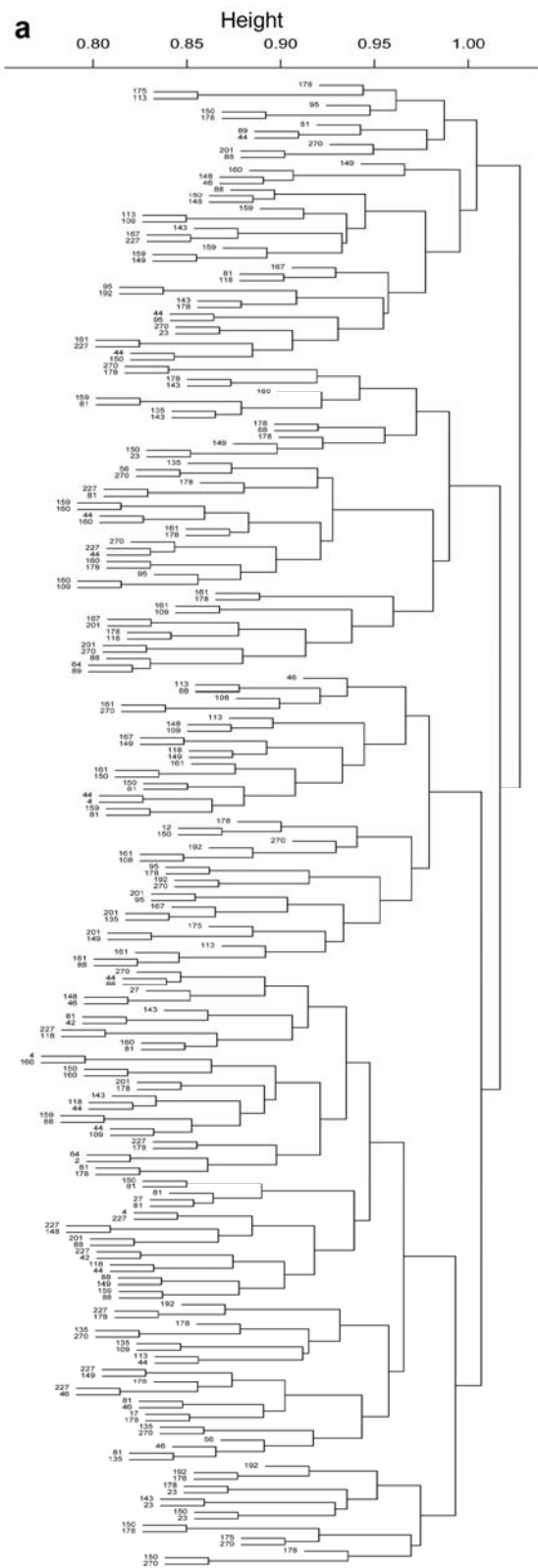**Supplementary Figure 2 | Liver pathology in HBsAg/Liver-SB mice**

a, Orcein staining of the liver of a 20.4 week-old Liver-SB/HBsAg male mouse. Orcein stains the HBsAg antigen in dark red and shows the characteristic ground glass appearance of hepatocytes seen in the human disease [5]. This ground glass appearance results from the accumulation of the HBsAg in the endoplasmic reticulum. **b**, Preneoplastic hepatocellular foci of hyperplastic hepatocytes, with inflammatory mononuclear cells in a 19.7 week-old Liver-SB/HBsAg male mouse. **c**, The total number of grossly visible tumors was counted in euthanized moribund Liver-SB/HBsAg and Liver-SB livers. The count includes intrahepatic tumors; the lobes were cut and tumor nodules seen inside the lobes also counted. Two-tailed unpaired t test was used to compare Liver-SB and Liver-SB/HBsAg data. Error bars represent the standard deviation. **d**, Hepatocellular carcinoma in a male Liver-SB/HBsAg mouse at 63.6 weeks of age. Macroscopic picture (left panel) showing multiple large liver tumors, and H&E staining (right panel) showing trabecular HCC. **e**, Percentage of

liver weight compared to body weight for sick animals. The littermate control mice were not sick but were necropsied at comparable ages (60 to 120 weeks of age). The number of mice assessed is shown in the bars. For both the female and male mice, a one-way ANOVA showed that the means differed significantly between the four groups (females: $F=12$, $p=5.5.10^{-6}$; males: $F=15$, $p=1.6.10^{-7}$). More specifically, a Tukey's HSD test showed significant differences ($p<0.05$) for female Control vs. Liver-SB, female Control vs. Liver-SB/HBsAg, male Control vs. Liver-SB/HBsAg, female HBsAg vs. Liver-SB, female HBsAg vs. LiverSB/HBsAg , male HBsAg vs. LiverSB/HBsAg, and male LiverSB vs. LiverSB/HBsAg. Most liver weight increases were due to liver tumor burden. Error bars represent the standard deviation. **f**, Gross pictures of the livers of animals 12.2 to 14.5 months of age. Livers with the most tumors were in the HBsAg/Liver-SB group. Liver-SB mice had the fewest liver masses. The scales represent mm.

| Mouse ID	Number of tumors per mouse
mouse2	1
mouse12	1
mouse17	1
mouse27	2
mouse42	2
mouse56	2
mouse64	2
mouse89	2
mouse108	2
mouse4	3
mouse175	3
mouse23	5
mouse148	5
mouse167	5
mouse46	6
mouse95	6
mouse109	6
mouse113	6
mouse118	6
mouse192	6
mouse135	7
mouse143	7
mouse149	8
mouse159	8
mouse201	8
mouse160	10
mouse161	10
mouse44	11
mouse88	11
mouse150	11
mouse227	12
mouse81	14
mouse270	14
mouse178	25

Supplementary Figure 3 | Distribution of the 228 tumors used for cancer gene identification
A total of 250 tumors were deep sequenced to characterize their transposon insertions. When the tumors were harvested from the same animal they were from different lobes, or when from the same lobe they were not continuous. However, we cannot exclude that some tumors from the same animal display genetic similarities. We pruned genetically similar tumors based on their transposon insertion sites (see Supplementary Methods) and removed 22 tumors that showed partial genetic similarity with others. For further analyses, we utilized insertional data from the remaining 228 large tumors. They were 22 to 178 mm in size and obtained from 34 moribund mice at 10.6 to 19.3 months of age (110 tumors from female and 118 tumors from male mice).
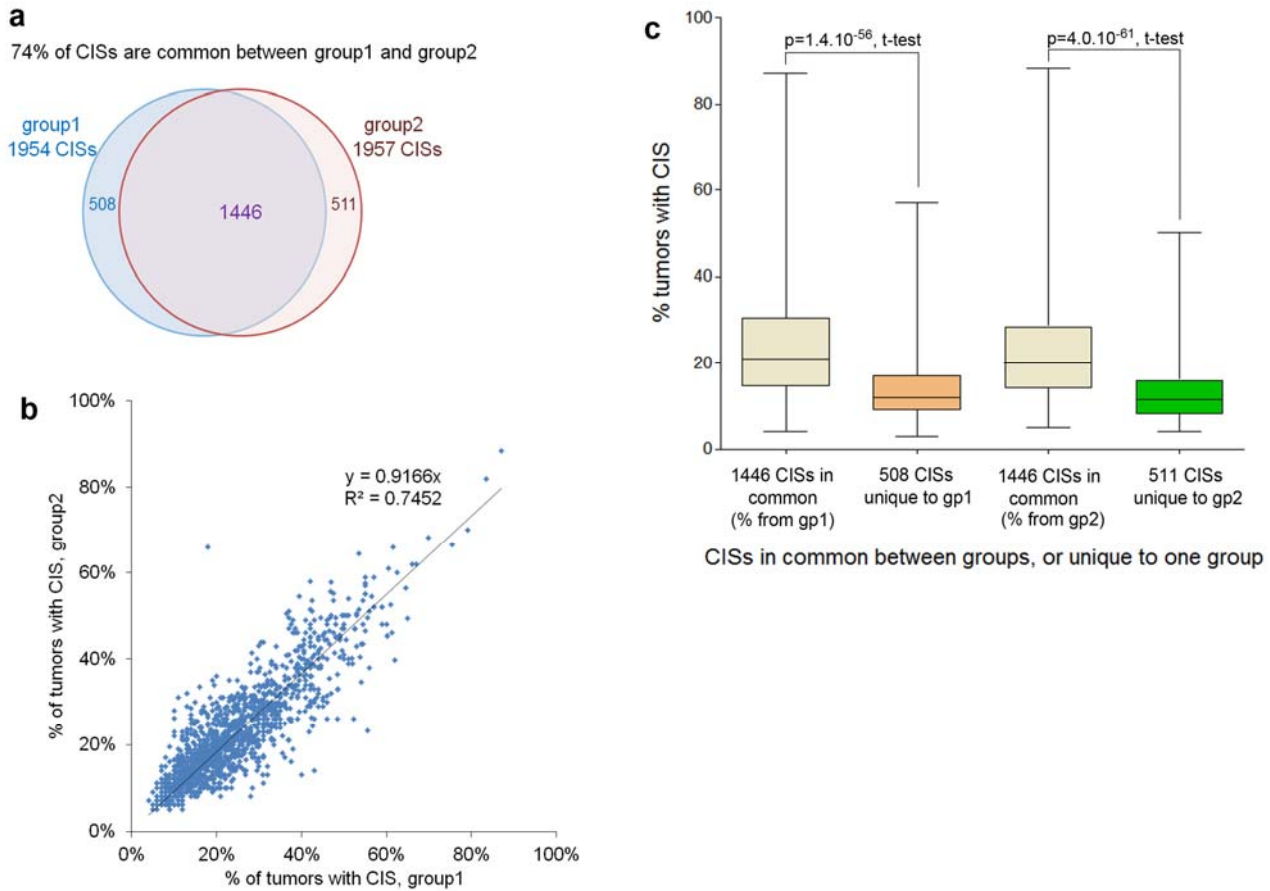
a

b

between (n=49634)
within (n=2122)

t-statistic = 1.7
p = 0.087

density

spearman correlation coefficient

c

distribution of CISs by occurence in mice

CIS counts

number of mice associated to CIS

6

Supplementary Figure 4 | Tumors are genetically dissimilar
a, To determine whether tumors isolated from the same mouse were truly independent, we clustered the tumors based on their GKC-derived CIS patterns using complete linkage hierarchical clustering with Spearman correlation. The mouse ID numbers are indicated. **b**, Distributions of the Spearman correlation coefficients for tumors within (blue curve) or between (red curve) mice. The median Spearman correlation coefficients for tumors from the same or different mice were not significantly different from each other, indicating that tumors from the same mouse were just as dissimilar, on average, as tumors originating from different mice. Moreover, the correlation coefficient was less than 0.1, indicating no positive correlation in both cases. **c**, Distribution of CISs based on their mutation rate in tumors from different mice. The CIS counts represent the number of CISs found in a particular number of animals. The number of mice associated to CIS represents the total number of animals in which a particular CIS is identified. All CISs are mutated in tumors from at least nine different mice.
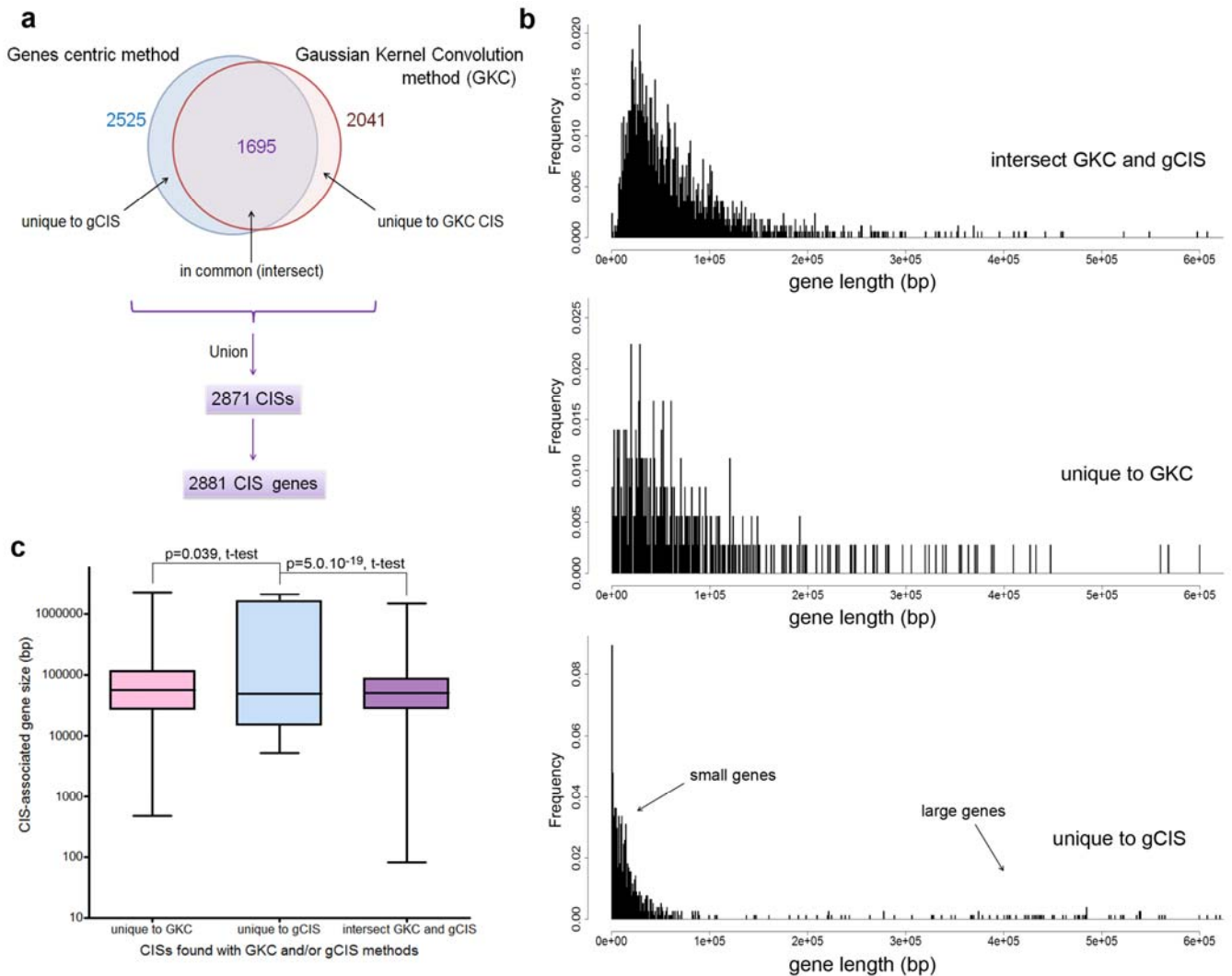
Supplementary Figure 5 | CISs identified by the GKC and gCIS methods

a, The gene centric method (gCIS)[6] identified 2525 CISs and GKC[7] identified 2041 CISs. The overlap of the genomic regions of GKC CISs with gCISs was higher than expected by chance (83.0% overlap, $p<2.10^{-16}$, Chi-square with Yates correction). The union of these two methods gave 2871 CIS loci that represented 2881 candidate cancer genes. This was because two genes can be annotated to one GKC CIS when the loci overlap or when the CIS is located between two genes. **b,** CIS genes were ordered by size. Gene lengths were obtained from the Mouse Genome Informatics (MGI) database, miRNAs were excluded. **c**, Box-plot representing the variation of gene size associated with CISs of various categories. (**b,c**) Very large and very small genes are poorly identified by the GKC method due to the use of fixed kernel widths (15K to 240K).

a

74% of CISs are common between group1 and group2

group1
1954 CISs

group2
1957 CISs

508   1446   511

b

$y = 0.9166x$
$R^2 = 0.7452$

% of tumors with CIS, group2

% of tumors with CIS, group1

c

% tumors with CIS

$p=1.4.10^{-56}$, t-test

$p=4.0.10^{-61}$, t-test

1446 CISs in common (% from gp1) | 508 CISs unique to gp1 | 1446 CISs in common (% from gp2) | 511 CISs unique to gp2

CISs in common between groups, or unique to one group

Supplementary Figure 6 | CIS calling for two subgroups of 100 tumors leads to similar results
Two subgroups of 100 tumors each from 17 independent mice were used for CIS calling with both GKC and gCIS methods. **a**, There was a 74% overlap in the CISs identified in tumors from group 1 and group 2, which was much higher than expected by chance ($p<2.10^{-16}$ Chi-Square with Yates correction).   **b**, There is a strong linear correlation in the frequency a CIS was mutated in the two subgroups of tumors. **c**, CISs that are unique to one subgroup, or are shared between two subgroups, are plotted against the percent of tumors with an insertion in the CIS. CISs found in both groups are of higher frequency, while the CIS that differs between group 1 and 2 are around the significance cut-off. The CISs close to the cut-off are the ones that differ most.

9
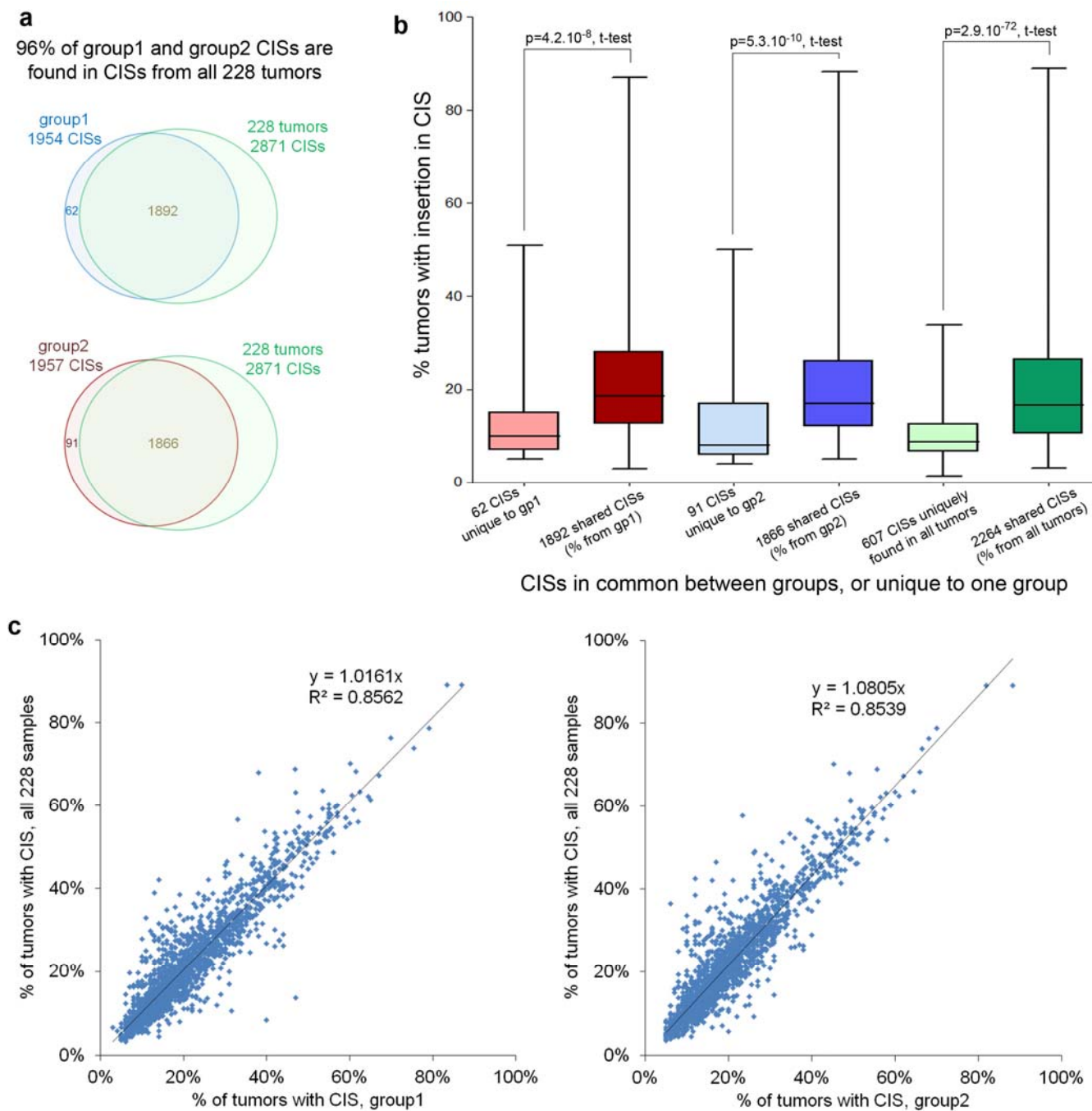
Supplementary Figure 7 | SB/HBV screen is reaching saturation
a, Overlap of the CISs from group 1 and group 2 tumors with the CISs obtained from all 228 tumors (see Supplementary Figure 6). The overlap was 96% on average and much higher than expected by chance (p values less than 2.10^{-16} Chi-Square with Yates correction), implying a good depth of coverage of our screen. **b**, CISs that are unique or common between group 1 and 2 and the CISs from 228 tumors are represented according to their presence in a certain proportion of tumors. The percent of CIS occurrence in tumors are significantly different (unpaired t-test) between CISs in common or found uniquely. **c**, Linear correlation between the frequency a CIS is mutated in all 228 tumors and the frequency it is mutated in group 1 or group 2 tumors. No significant variation was detected (unpaired t-test).

|	Organism	Number of Genes in dataset	Number of genes of other dataset, excluding genes of our donor chromosome	Number of Genes of our dataset, excluding genes of donor chromosome	Number of overlapping genes	% coverage of other dataset	Enrichment P-value (Chi-square, Yates correction)
Our Liver-SB/HBV CIS genes	Mouse	2881	n/a	2817 (chr15)	n/a	n/a	n/a
Liver SB screen in p53 Liver-KO background; (ref 17) transposon on chr15	Mouse	18	18	2817	16	88.8%	$< 2.10^{-16}$
Liver SB screen in Sav1 Liver-KO background; transposon on chr15	Mouse	116	113	2817	90	79.6%	$< 2.10^{-16}$
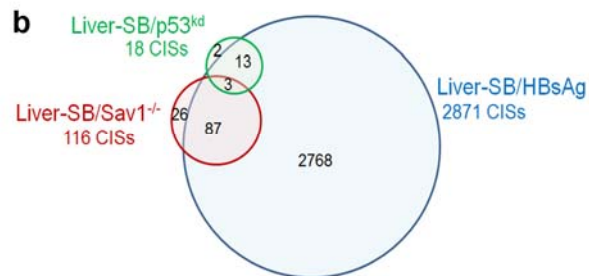
Supplementary Figure 8 | Overlap between the CIS genes and CISs identified in three different Liver-SB screens

a, We compared our list of 2881 CIS genes with the CIS genes identified in two other smaller-scale Liver-SB screens. We found 16 of the 18 CISs previously identified tumors induced in a p53-deficient background[8] among our list of CIS genes. In addition, in another Liver-SB screen done in a Sav1-deficient background, described in Supplementary methods and Supplementary Table 1e, 90 of the 116 CIS genes identified in the screen were also among our list of CIS genes. **b**, Diagram showing the overlap between the CISs found in the three Liver-SB screens.
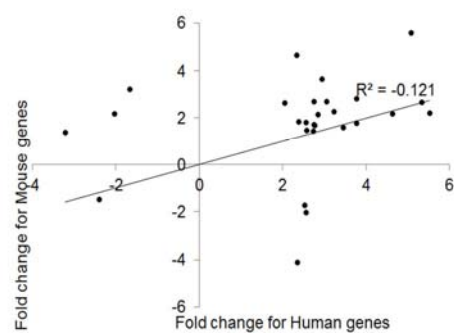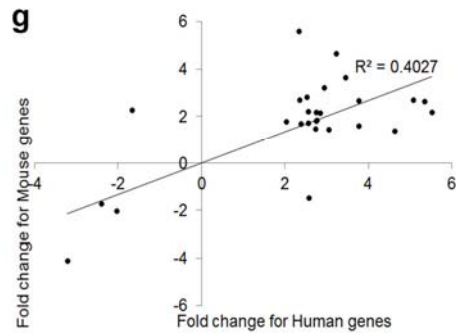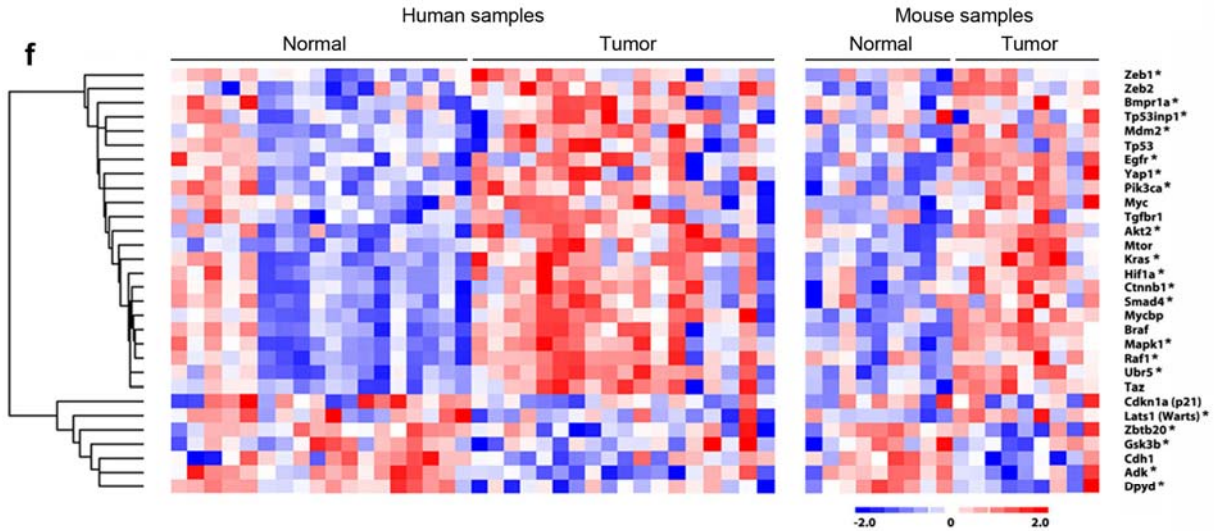
Supplementary Figure 9 | CIS genes associated with highest sequence read counts
a, We obtained CISs from SB insertions with sequencing read counts ≥ 4. These CISs were plotted according to their median values of sequence reads obtained for insertions in the CIS and their frequency of occurrence in tumors. The CISs highlighted in red are the outliers, and are considered to be strong driver genes in SB-induced HCC. **b**, Mapping of SB insertions in the imprinted locus on Chr12 reveals a transposon orientation bias. Most insertions are in the same strand as Rtl1 gene, suggesting an activation of downstream genes and/or microRNAs. The scale represents the mm9 mus musculus genome.

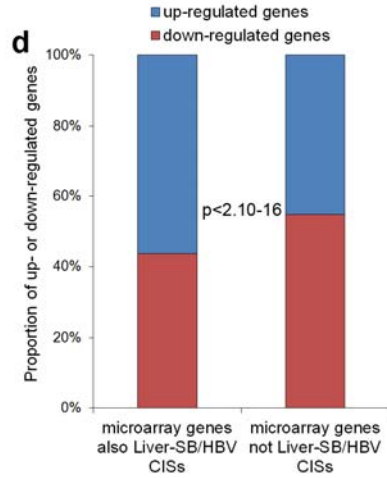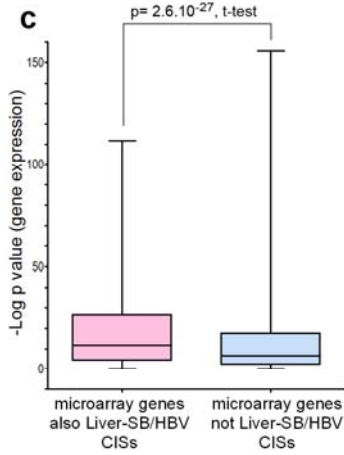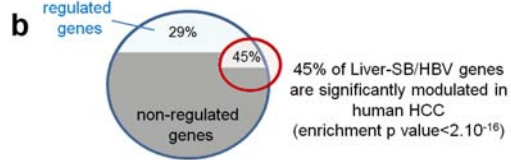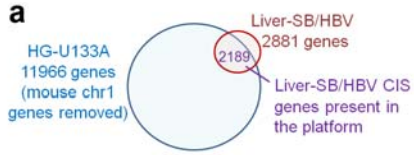|	Organism	Number of CISs in dataset	Number of overlapping genes	% coverage of other dataset	Enrichment p value (Chi-square, Yates correction)
Our Liver-SB/HBV list of CIS genes	Mouse	2881	n/a	n/a	n/a
CRC Sleeping-Beauty screen (ref 21) (transposon on chr1)	Mouse	1033	626	60.6	$< 2.10^{-16}$
PDAC SB screens (refs 18,23) (transposon on chr1)	Mouse	1649	1006	61	$< 2.10^{-16}$
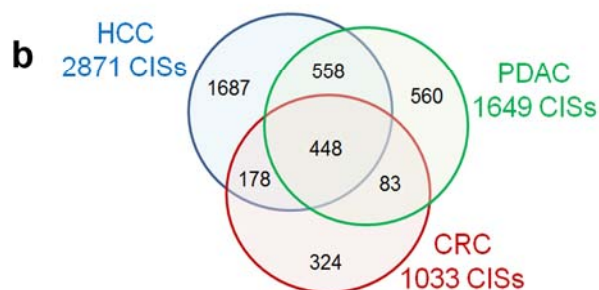
Supplementary Figure 10 | HCC CIS genes are mutated in other types of cancer
a, We compared our list of HCC CIS genes with the list of CIS genes identified in CRC[7] and PDAC[9,10]. The overlaps were then calculated as well as p values for enrichment of the overlap (2x2 contingency table, Chi-square with Yates' correction). **b**, About 61% of the genes identified in CRC[7] and PDAC[9,10] were also identified in HCC. These latter genes likely represent common cancer genes selected in many types of cancer.
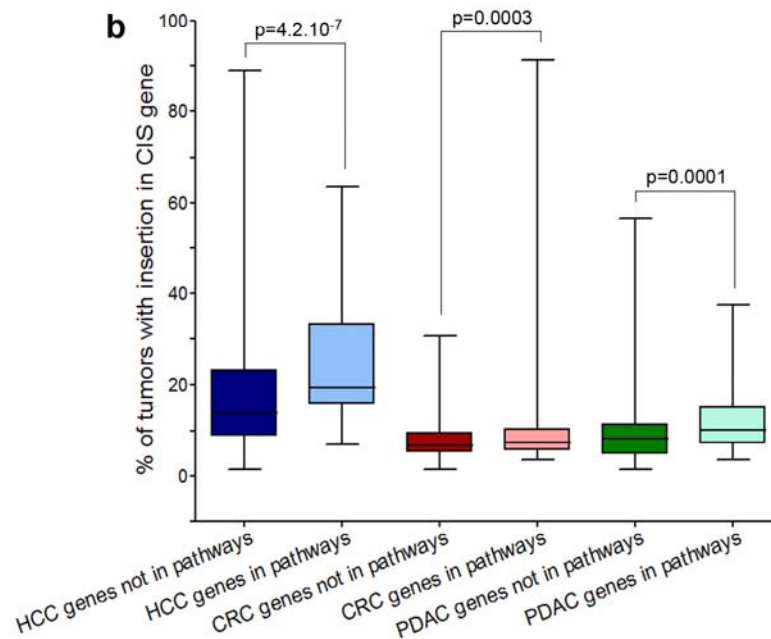
a HG-U133A 11966 genes (mouse chr1 genes removed) | Liver-SB/HBV 2881 genes | 2189 | Liver-SB/HBV CIS genes present in the platform

b regulated genes | 29% | 45% | non-regulated genes | 45% of Liver-SB/HBV genes are significantly modulated in human HCC (enrichment p value<2.10⁻¹⁶)

c p= 2.6.10⁻²⁷, t-test | -Log p value (gene expression) | microarray genes also Liver-SB/HBV CISs | microarray genes not Liver-SB/HBV CISs

d up-regulated genes | down-regulated genes | Proportion of up- or down-regulated genes | p<2.10-16 | microarray genes also Liver-SB/HBV CISs | microarray genes not Liver-SB/HBV CISs

e % of non regulated genes in HCC | % of significantly regulated genes in HCC | % of genes represented in the microarray | All microarray genes | Genes also found by Liver-SB/HBV | Genes also found by CRC-SB | Genes also found by PDAC-SB

f Human samples — Normal, Tumor | Mouse samples — Normal, Tumor

Zeb1*, Zeb2, Bmpr1a*, Tp53inp1*, Mdm2*, Tp53, Egfr*, Yap1*, Pik3ca*, Myc, Tgfbr1, Akt2*, Mtor, Kras*, Hif1a*, Ctnnb1*, Smad4*, Mycbp, Braf, Mapk1*, Raf1*, Ubr5*, Taz, Cdkn1a (p21), Lats1 (Warts)*, Zbtb20*, Gsk3b*, Cdh1, Adk*, Dpyd*

-2.0   0   2.0

g R² = 0.4027 | Fold change for Mouse genes | Fold change for Human genes

R² = -0.121 | Fold change for Mouse genes | Fold change for Human genes

14

Supplementary Figure 11 | CIS genes found in the Liver-SB/HBV screen are also misexpressed in human HCC

Microarray data[11] from 223 HCC patients containing data from tumor tissue and adjacent normal tissue (GSE14520). **a**, Diagram representing the human genes present on the Affymetrix microarray and corresponding orthologs of mouse HCC CIS genes. All genes mapping to mouse chromosome 1 were excluded. **b**, The proportion of differentially expressed genes was higher for the HCC CIS genes than for all other genes present in the microarray. Enrichment p value was $<2.10^{-16}$ (2x2 contingency table, Chi-square with Yates' correction). **c**, Boxplot showing the Log transformed p values for differential expression of genes between HCC tissue and adjacent non-tumor tissues. The p values are significantly lower for HCC CIS genes. **d**, Proportion of all CIS genes and non-CIS genes found to be up- or down-regulated in the microarray datasets. **e**, HCC CIS genes showed more misregulation on the arrays compared with CRC and PDAC CIS genes ($p<2.10^{-16}$, Chi-Square with Yates correction for all). **f**, High-throughput qRT-PCR validation and clustering analysis represented as a heatmap shows similar expression pattern differences between tumor and adjacent non-tumor samples for 18 HBV-positive patient samples and 9 HBsAg/Liver-SB samples. We assessed 30 genes (including 21 Liver-SB/HBV CIS genes) representing several key oncogenic pathways (Myc, Pi3k/Akt/mTor, Ras, p53, Tgfβ, Wnt/β-catenin) in addition to three of our most highly mutated genes (Adk, Dpyd, Zbtb20). **g**, Correlation of the RT-qPCR fold change in expression between mouse and human samples (from Supplementary Table 4). The left panel shows a positive correlation, whereas the right panel shows no correlation when using randomized values.

|	p value for enrichment of PDAC CIS genes in pathways	p value for enrichment of CRC CIS genes in pathways	p value for enrichment of HCC/HBV CIS genes in pathways	p value for enrichment of HCC/Savi1LiverKO CIS genes in pathways
Akt pathway	0.0015	0.0064	5.4E-04	0.2513
Wnt pathway	1.6E-09	< 2.2e-16	5.5E-06	0.0019
Ras/Erk pathway	6.3E-09	2.0E-07	3.7E-12	1.1E-07
p53 regulation	2.0E-15	3.2E-10	2.8E-09	1.3E-05
Tgfb/Bmp pathway	6.2E-05	7.8E-05	1.6E-05	0.3652
Hippo pathway	5.8E-05	2.5E-04	1.0E-06	6.0E-05
IL6 pathway	0.1894	0.7756	0.1203	0.0239
Hh pathway	0.8175	0.8638	0.6598	0.1286

Supplementary Figure 12 | HCC, CRC and PDAC CIS genes are enriched for genes in major cancer signaling pathways

a, Enrichment of CIS genes in major cancer signaling pathways . Pathway genes and CIS genes are listed in Supplementary Table 5. Pathway analysis was performed for HCC, CRC[7] and PDAC[9,10] CIS genes, and the smaller-scale HCC transposon screen performed in a Sav1-deficient background (described in Supplementary Methods and Supplementary Table 1e). The p values for enrichment (based on number of genes targeted per pathway) were calculated using 2x2 contingency tables, Chi-Square with Yates correction, two-tailed p values. **b**, These box-plots represent the frequency of CISs that either belong to these pathways or not. CIS genes from these pathways are significantly enriched in CISs of higher frequency (two-tailed p value, unpaired t-test).
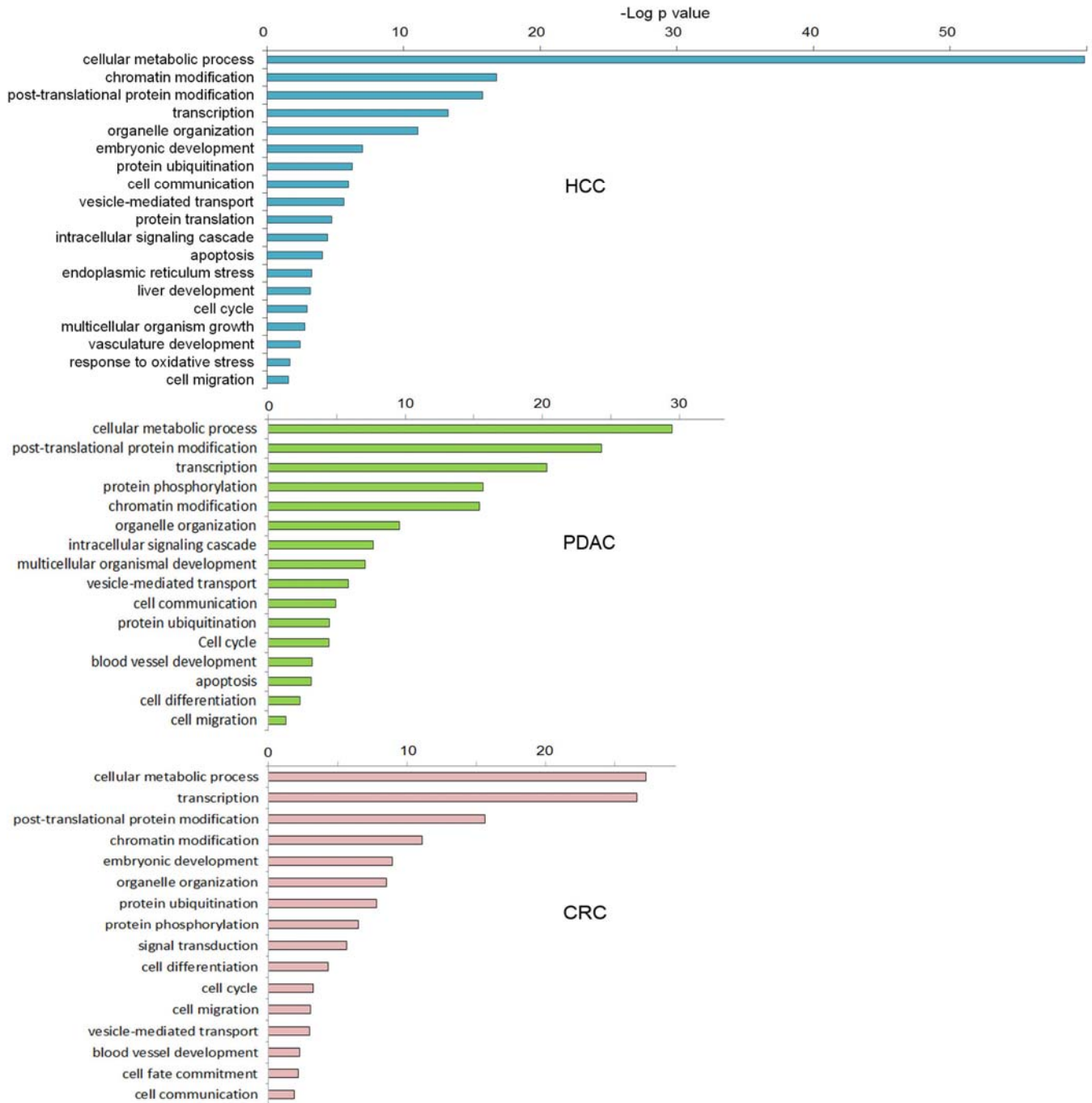
Supplementary Figure 13 | Gene ontology analysis for HCC, PDAC and CRC CIS genes
The lists of CIS genes were uploaded into DAVID bioinformatics[12] and the enrichment in biological processes was obtained. The bars represent the –log10 of the p values for enrichment. Only significant results are shown (p<0.05). Although the PDAC and CRC mutagenesis screens are not approaching saturation, most ontologies show relatively similar p values. HCC display a much higher enrichment in genes related to metabolic processes.
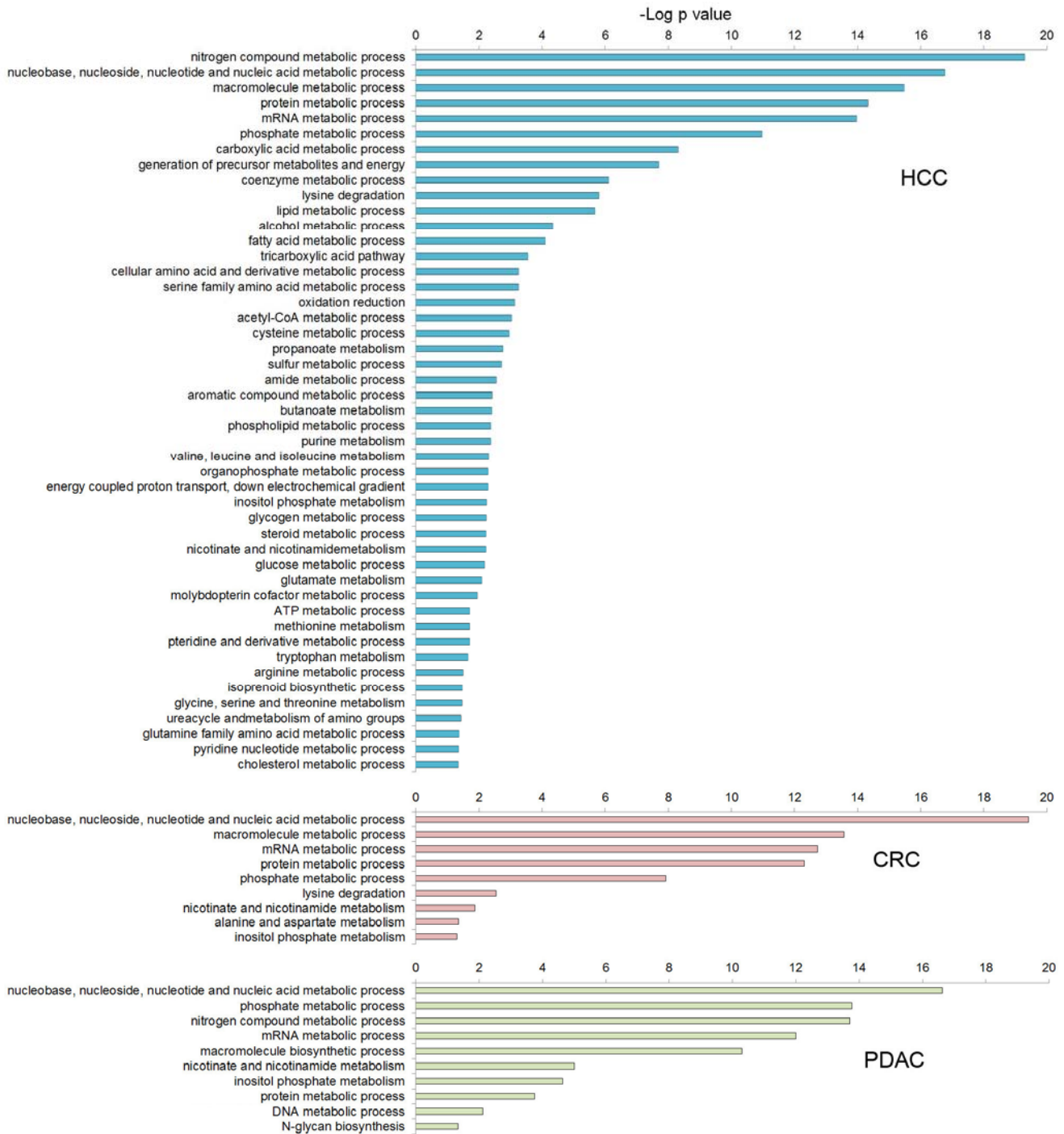
Supplementary Figure 14 | Gene ontology analysis focused on metabolic processes with CIS genes obtained from CRC[7], PDAC[9,10] and Liver-SB/HBV studies

The lists of genes were uploaded into Ingenuity Pathway Analysis and DAVID bioinformatics[12] and enrichments for metabolic biological processes were obtained (children terms of "cellular metabolic process" in DAVID). The bars represent the –Log10 of the p values for enrichment. Only significant results are shown (p<0.05).
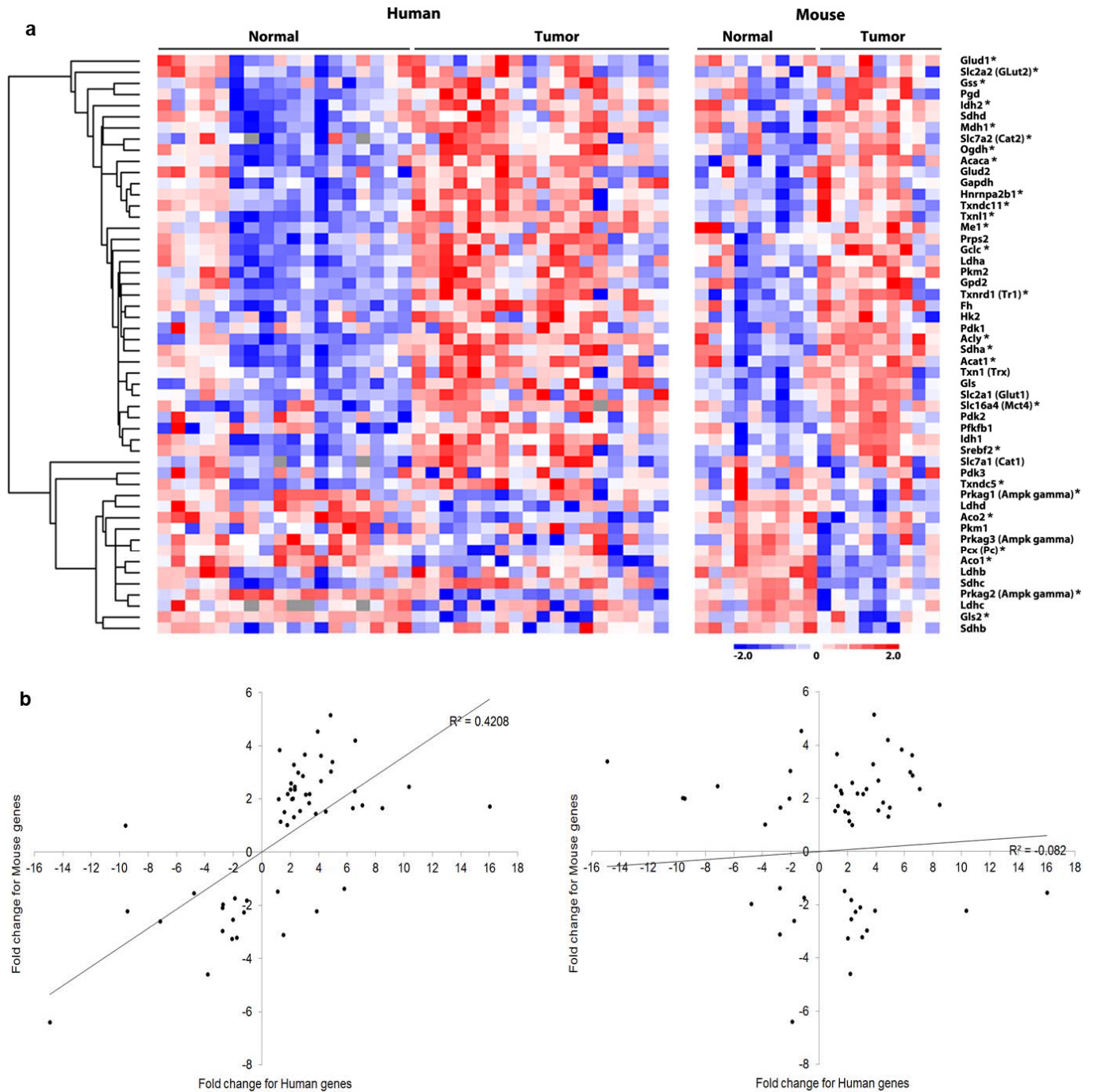
Supplementary Figure 15 | Molecular pathogenesis linked to metabolism in liver tumors
a, Heatmap representing the hierarchical clustering of gene expression levels of 52 metabolic genes from key metabolic processes like glycolysis, glutaminolysis, redox regulation, pentose phosphate pathways, and the TCA cycle. Tumor samples from 9 HBsAg/Liver-SB mice and 18 HBV-positive human HCC patients are shown. Tumor tissues and adjacent non-tumor tissues were assessed. CIS genes listed in Supplementary Table 1c are labeled with *. **b**, Correlation of average fold change expression values between the mouse and human samples (from Supplementary Table 4). Each dot represents one gene. The left panel shows a positive correlation between the two species, whereas the right panel shows no correlation when using randomized values.

19
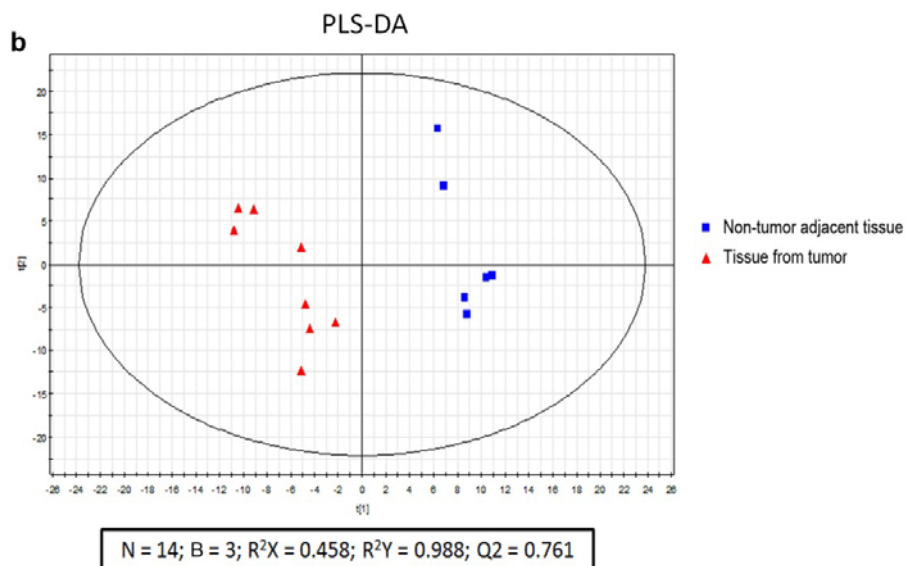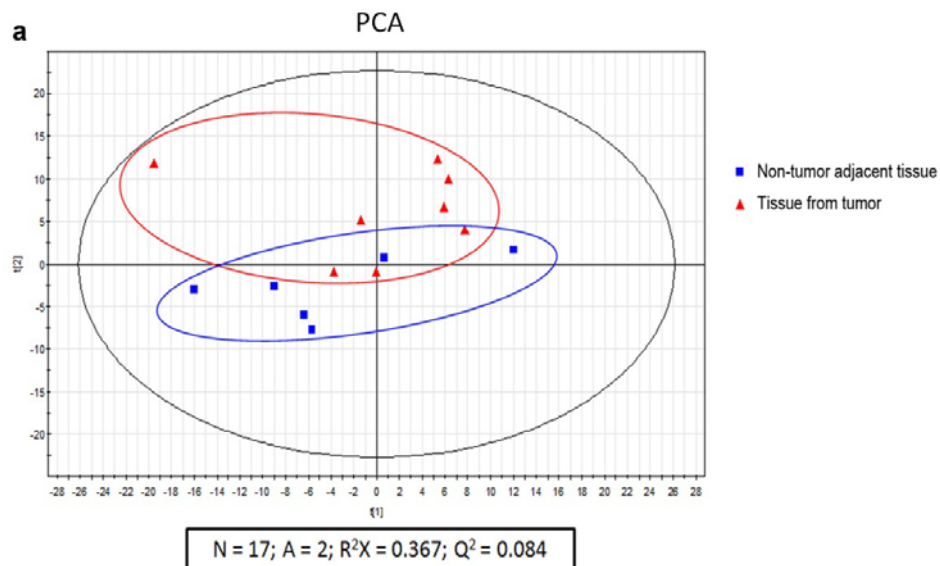
Supplementary Figure 16 | Quality control of metabolomics data

Gas chromatography mass spectrometry was performed on tumors larger than 5mm and adjacent normal tissue for each animal. Principal component (a) and Partial Least Square-Discriminant (b) analyses of the data grouped most samples consistently to their provenance from tumor or control tissue. (N = number of observations; A = number of components; R2X and Q2 = PCA validation parameters, B= latent variables). **a**, The analytical performance of GC/TOFMS was found to be satisfactory as the non-supervised PCA scores plot showed that all the QC samples were clustered in greater proximity compared to the cancer and control groups. **b**, The supervised PLS-DA model and its associated validation parameters further demonstrated that the prediction model was robust [5 Latent variables, $R^2Y = 0.989$, Q^2(cum) = 0. 778]. Validation of the PLS-DA model using permutation analysis confirmed that the model was not due to chance correlation (data not shown). The marker metabolites were identified on the basis of the pre-defined criteria and the PLS-DA model constructed between the normal matched (NM) normal and tumor (TU) tissues. The marker metabolites and their associated parameters (variable importance value = VIP, *p* values and fold change) are summarized in Supplementary Table 7.
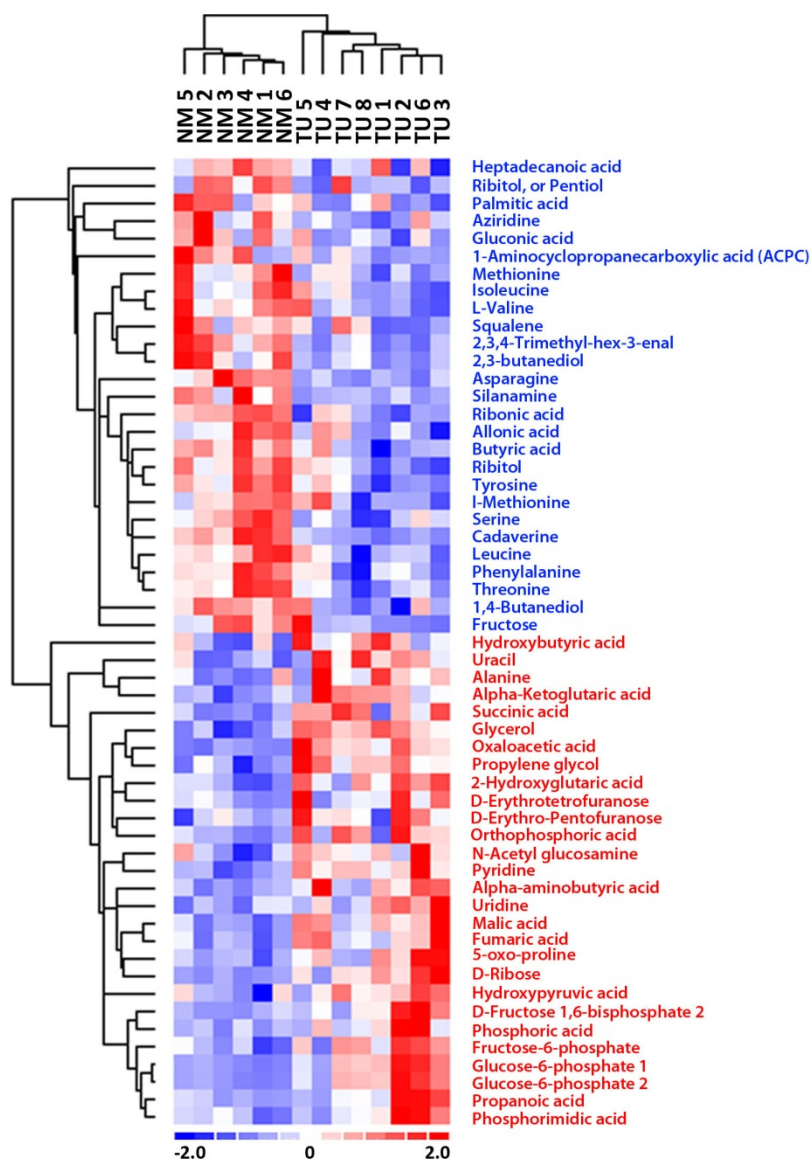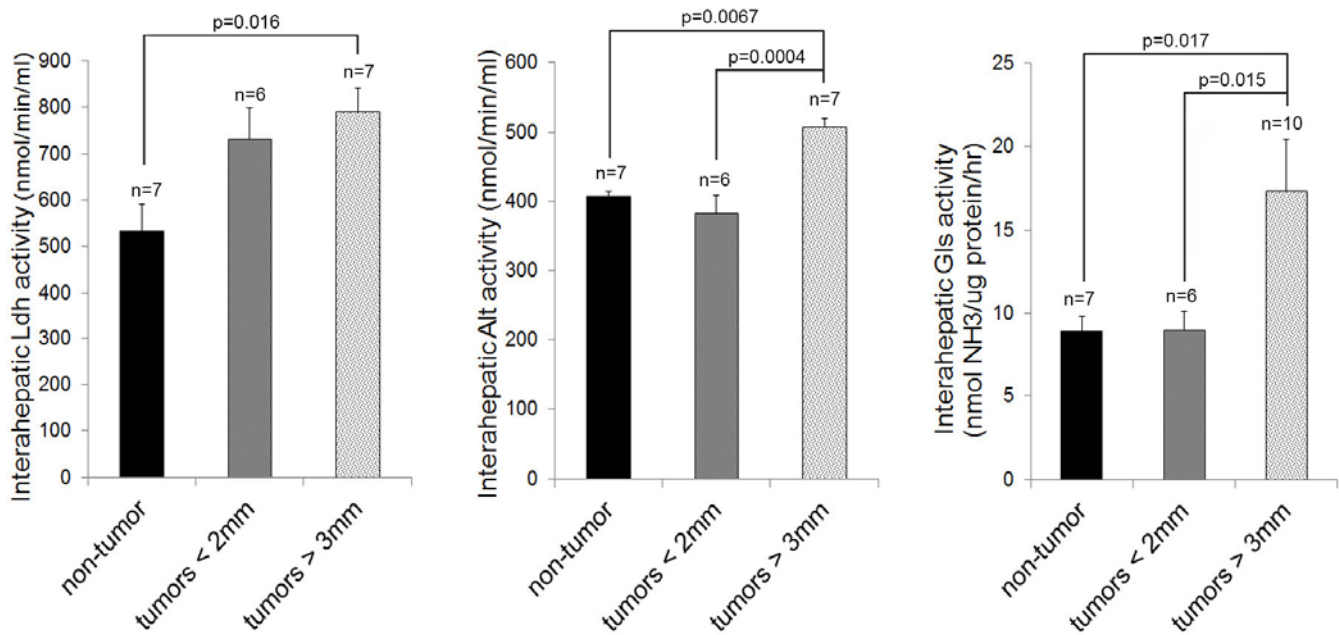
Supplementary Figure 17 | Untargeted metabolic profiling for liver tumors and non-tumor adjacent tissue
Heatmap representing hierarchical clustering analysis of the 56 metabolites found to be significantly changed in tumors (TU) and adjacent non-tumor (NM) liver tissue of HBsAg/Liver-SB mice ($p<0.01$, see Supplementary Table 7).
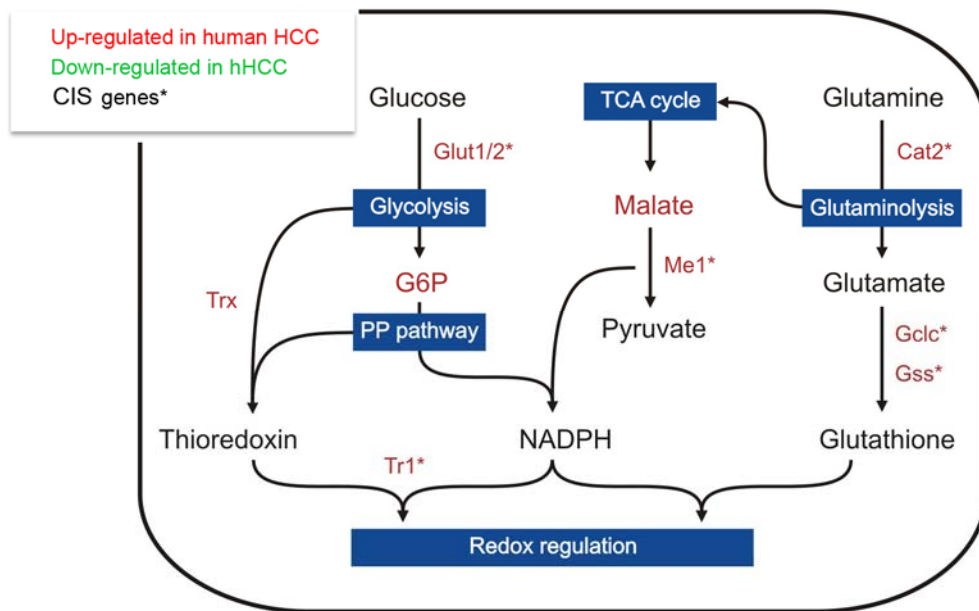
Supplementary Figure 18 | Mapping of various datasets to redox pathways disrupted in HCC
To understand the metabolic pathways disrupted at the genetic, mRNA, and metabolite levels, we used data from the Liver-SB/HBV screen (Supplementary Table 1c), human expression data[11], and metabolomics results (Supplementary Table 7). Metabolites are written with larger font size, and genes with smaller font size. Red or green fonts mean the gene expression or metabolite amount is increased or decreased, respectively, in HCC tissue versus adjacent non tumor tissue. CIS genes listed in Supplementary Table 1c are labeled with *. Rapid tumor cell growth and proliferation induce a high level of reactive oxidative stress (ROS) which hampers tumor cell survival[13]. Liver HCC tumor counteracted the effects of ROS by employing antioxidant defenses including NADPH, glutathione, and thioredoxin. The conversion of glutamine-derived malate to pyruvate by Me1 (malic enzyme 1) gene and the activation of PP pathway may produce NADPH. The synthesis of glutathione may happen through glutaminolysis (Gclc, Gss) and thioredoxin production may be induced by its gene expression and its enhanced metabolic processing.

Supplementary Figure 19 | Biochemical metabolic assays confirm the predictions from the integrated datasets

Enzymatic activities in normal tissue and tumors. Increased reduction of pyruvate to lactate by lactate dehydrogenase (Ldh) in tumors illustrates the 'Warburg effect', in which a higher NADH/NAD$^+$ turnover is required for aerobic glycolysis. In large tumors, elevated alanine transaminase (Alt) activity enables rapid conversion of alanine into pyruvate, in order to support increased pyruvate anaplerosis. Large tumors also include glutamine as an anaplerotic carbon source to supplement the increased metabolic fuel requirement. Error bars represent the standard error of the mean (SEM). Kruskal-Wallis tests were performed for each of the cases Ldh, Alt, and Gls. They showed that in each of these cases, the group means were different (FDR-corrected p-value = 0.027 for Ldh, 0.005 for Alt, 0.0178 for Gls). More specifically, Nemenyi tests showed significant differences between specific groups as indicated on the graphs (p represents Nemenyi's adjusted p value). When no indication is mentioned, there was no significant difference. Spearman correlation showed that the all enzymes activities significantly increased with tumor size (rho=0.51, FDR-corrected p=0.0051 for Ldh, rho=0.65, FDR-corrected p=0.0051 for Alt, and rho=0.58, FDR-corrected p=0.0051 for Gls). The numbers of samples assessed in each group are indicated above the bars.

Supplementary Tables

Supplementary Table 1 | List of CISs obtained from liver tissues.
a, CISs obtained from 228 Liver-SB/HBV hepatic tumors with the gene centric method. **b**, CISs obtained from 228 Liver-SB/HBV hepatic tumors with the GKC method. **c**, Union of (a) and (b). The genes also found in other transposon-based mutagenesis screens[7,9,10] are indicated. **d**, The CISs obtained from SB insertions with sequencing read counts $\geq$ 4. The CISs highlighted in blue indicate the outliers considered to be strong driver genes induced by SB in HCC. **e**, CISs obtained from liver tumors in a Sav1 liver-knockout background (screen described in Supplementary Materials). Sfi1 and En2 genes were removed from all lists as they are known artifacts due to multiple copies of this gene in the mouse genome, or presence in the transposon sequence, respectively.

Supplementary Table 2 | Cross-species comparison between HCC CIS genes and human expression data.
The microarray data from 223 human HCC patient samples (GSE14520), tumor tissue and adjacent non-tumor tissue[11] were used. The expression results are indicated for all the CIS genes of the Liver-SB screen. The genes highlighted in grey are below the threshold of significance (significance: absolute value of fold change >1.5 and adjusted p value <10^{-5}).

Supplementary Table 3 | List of HCC candidate cancer genes that are also found mutated in human cancer, or mis-regulated in human HCC
CISs highlighted in red are found either mutated or mis-expressed in human cancer (see Table 1 and Figure 3).

Supplementary Table 4 | Cross-species comparison of several genes of interest by high-content RT-qPCR (Biomark™). The fold changes were calculated with 10 out of 18 HBV-positive HCC samples and 6 out of 9 HBsAg-positive Liver-SB mouse tumor samples that belonged to the most representative HCC gene signature (Supplementary Figures 11f and 15a). Gene expression levels in tumors and non-tumor adjacent tissues were calculated, as well as two-tailed p values (unpaired t-test). Primer sequences used for the Biomark™ qPCR validation are indicated.

Supplementary Table 5 | CIS genes mapping to cancer signaling pathways
The tables list genes of several canonical cancer signaling pathways. The CIS genes obtained in HCC-SB, PDAC-SB[9,10] and CRC-SB[7] screens are mentioned with their percentage of occurrence.

Supplementary Table 6 | Lists of HCC CIS genes classified by biological or metabolic processes.
a, Classification by biological processes using DAVID bioinformatics[12]. **b**, Classification by metabolic processes using both DAVID bioinformatics and Ingenuity Pathway Analysis.

Supplementary Table 7 | List of marker metabolites found misregulated in liver tumors.
Variable importance value (VIP), p values from t-test and Welch test, and fold change for tumor tissue versus adjacent normal tissue are indicated.

Supplementary Dataset 1 | Non-redundant transposon insertions.
This file contains BED files listing non-redundant insertions from the 228 large liver tumors.

Supplementary References

1   Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221-226 (2005).

2   Dupuy, A. J. *et al.* A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. *Cancer Res* **69**, 8150-8156 (2009).

3   Postic, C. & Magnuson, M. A. DNA excision in liver by an albumin-Cre transgene occurs progressively with age. *Genesis* **26**, 149-150 (2000).

4   Dupuy, A. J. *et al.* Mammalian germ-line transgenesis by transposition. *Proc Natl Acad Sci U S A* **99**, 4495-4499 (2002).

5   Portmann, B., Galbraith, R. M., Eddleston, A. L., Zuckerman, A. J. & Williams, R. Detection of HBSAG in fixed liver tissue - use of a modified immunofluorescent technique and comparison with histochemical methods. *Gut* **17**, 1-9 (1976).

6   Brett, B. T. *et al.* Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS One* **6**, e24668 (2011).

7   March, H. N. *et al.* Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat Genet* **43**, 1202-1209 (2011).

8   Keng, V. W. *et al.* A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol* **27**, 264-274 (2009).

9   Mann, K. M. *et al.* Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A* (2012).

10   Perez-Mancera, P. A. *et al.* The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature* **486**, 266-270 (2012).

11   Roessler, S. *et al.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res* **70**, 10202-10212 (2010).

12   Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

13   Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nat Rev Cancer* **11**, 85-95 (2011).