Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Yong KJ, Gao C, Lim JSJ, et al. Oncofetal gene *SALL4* in aggressive hepatocellular carcinoma. N Engl J Med 2013;368:2266-76. DOI: 10.1056/NEJMoa1300297

Supplementary Appendix

Oncofetal Gene *SALL4* in Aggressive Hepatocellular Carcinoma

Kol Jia Yong, B.Sc., Chong Gao, M.D., Ph.D., Joline SJ Lim, M.B.,B.S., Benedict Yan, M.B.,B.S., Henry Yang, Ph.D., Todor Dimitrov, Ph.D., Akira Kawasaki, M.D., Ph.D., Chee Wee Ong, M. Sc., Kwong-Fai Wong, Ph.D., Sanghoon Lee, Ph.D., Sharada Ravikumar, M.D., Ph.D., Supriya Srivastava, M.D., Xi Tian, B.S., Ronnie T. Poon, M.B.,B.S., Ph.D., Sheung Tat Fan, M.D., D.Sc., John M. Luk, D.Med.Sc., Yock Young Dan, M.B.,B.S., Ph.D., Manuel Salto-Tellez, M.D., Li Chai, M.D., Daniel G. Tenen, M.D.

TABLE OF CONTENTS

SUPPLEMENTARY METHODS

DEFINITION OF CLINICAL OUTCOME

Preoperative diagnosis of HCC was based on imaging criteria on CT or MRI scan, and if necessary, by percutaneous biopsy. Diagnosis of HCC was confirmed by histology. HCC stage was analyzed according to the conventional TNM and BCLC staging criteria.[1] Child-Pugh score were computed to document the levels of derangement of liver disease of the patients at diagnosis.[2] Similar HCC treatment algorithms based on international and Asian treatment guidelines were used in both cohorts.[1,3,4] Generally, patients were subjected to curative surgery if they were at BCLC stage 0, A or B, resection was however also offered to selected good risk patients with BCLC stage C tumors, such as those with portal vein branch invasion. In our patient cohorts, 70-80% of HCC patients were positive for Hepatitis B infection, and oral nucleoside and nucleotide analogs were initiated if HBV DNA was elevated. Curative surgery was defined as complete resection of the tumor with clear microscopic margin, and no residual tumors detected by CT scan or angiography at 1 month after surgery. All post-operative patients were subjected to regular surveillance as outpatients with standard protocol including 2-6 monthly CT scan imaging of the liver and measurement of serum alpha-fetoprotein level to monitor for tumor recurrence. Any suspected intrahepatic recurrence was confirmed by hepatic angiography, post-lipiodol CT scan, and if necessary, percutaneous fine-needle aspiration cytology. Recurrent tumors were treated with re-resection, ablation, transarterial chemoembolization or systemic therapy as appropriate. In our cohorts, 80% of patients were followed for up to 2 years. Only patients who died as a result of HCC or its related mortality were included in our analysis.

IMMUNOHISTOCHEMISTRY

Paraffin tissue sections of 4 μm were deparaffinized with Histoclear and hydrated in graded ethanols. Antigen retrieval was performed by boiling at 120°C in high pH target retrieval solution for 5 minutes in a pressure cooker for SALL4 IHC, or heating in citrate buffer at 95°C for 30 minutes for Ki-67 IHC. Non-specific signal was blocked by peroxidase block for 10 minutes at room temperature, followed by protein block for 30 minutes at room temperature. Primary antibodies were incubated at room temperature for one hour in a humidified chamber, followed by HRP-conjugated secondary antibody incubation for 30 minutes at room temperature. Antibody binding was revealed by DAB and reaction was stopped by immersion of tissue sections in distilled water once brown color appeared. Tissue sections were counterstained by hematoxylin, dehydrated in graded ethanols and mounted. The following antibodies were used: SALL4 (Santa Cruz, CA, USA #sc-101147) and Ki-67 (Novus Biologicals, Littleton, CO, USA #NB110-89717). All reagents for immunohistochemistry were from Dako (Dako, Glostrup, Denmark A/S). Appropriate positive and negative controls were included for each run of IHC. Only nuclear staining was considered positive for SALL4. For IHC on tissue microarrays, SALL4 expression was scored according to the percentage of tumor cells stained positive for SALL4, with 0 denotes less than 5% of tumor cells stained positive, 1 denotes 5 – 30% of tumor cells stained positive, 2 denotes 31 – 50% of tumor cells stained positive, 3 denotes 51 – 80% of tumor cells stained positive, 4 denotes >80% of tumor cells stained positive. SALL4 expression in tissue microarrays was scored by a pathologist and two researchers independently.

CELL CULTURE

HCC cell lines were maintained in either Dulbecco’s Modified Eagle Medium (DMEM) or RPMI medium supplemented with 10% fetal bovine serum (FBS) at 37°C in a humidified atmosphere of

5% CO_2. Immortalized hepatocyte cell lines, THLE-2 and THLE-3, were maintained in BEGM medium supplemented with growth factors (Lonza, Basel, Switzerland) in pre-coated tissue culture flasks at 37°C in a humidified atmosphere of 5% CO_2 as recommended by ATCC.

MICROARRAY DATASETS

For profiling HCC samples versus normal controls, datasets from the GEO database with the accession numbers of GSE6222, GSE6864, & GSE29721 were used. For comparing HCC samples of high and low SALL4 expression with primary hepatocytes (Hep) and human fetal liver (HFL) samples, appropriate samples from the following GEO datasets (GSE6222, GSE6764, GSE9843, GSE15238, GSE18269, GSE23343, GSE29721, & GSE33606) were utilized. For comparison of SNU-398 samples with primary hepatocytes and human fetal liver samples, Hep and HFL samples were taken from GEO datasets, GSE23034 & GSE23413, respectively. Microarray data for the Hong Kong cohort of primary HCCs was deposited in GEO database with the accession number GSE25097, and the software/ methods used for molecular profiling and computation were previously described.[5] SNU-398 samples with SALL4 knocked down were submitted to GEO database with the following accession number: GSE35965.

MICROARRAY DATA NORMALIZATION

For Affymetrix data, all CEL files were analyzed together using the Robust Multichip Average method to obtain the gene expression intensities.[6] For Illumina Beadchip data, raw data with background subtraction were used for all samples. Normalization was then performed across all samples based on the Cross Correlation method[7] using R script, and normalized data were further log2-tranformed.

CLUSTERING, HEATMAPS AND GENE SET ENRICHMENT ANALYSIS (GSEA)

Hierarchical clustering with average linkage was used in all clustering. For clustering HCC SALL4 high and low samples with human fetal liver and hepatocyte samples, the mean of the four group means (HCC SALL4 high, HCC SALL4 low, human fetal liver, and hepatocyte) was subtracted from the log2-transformed normalized data prior to clustering. For clustering SNU-398 SALL4 knocked down data, human fetal liver and hepatocyte genes with differential expression between the cell line and primary cells were excluded by using only genes with no significant changes between any one of SUN-398 SALL4 knocked down (KD) and wildtype control (WT) and any one of primary cells to remove cell line and primary cell expression differences. The cutoff for no significant fold change used is 1.5. The mean of the four group means (SNU-398 KD, SNU-398 WT, hepatocyte and human fetal liver) was then also subtracted from the log2-transformed normalized data prior to clustering. Genes with no significant changes between the four groups of samples were not represented in both heatmaps to show clear patterns. Gene set enrichment analysis (GSEA) was performed by using normalized data using GSEA v2.0 tool (http://www.broad.mit.edu/gsea/). For GSEA, we first carried out hierarchical clustering to separate primary HCCs into high and low SALL4 groups and obtained four subclusters. Comparing high and low SALL4 HCCs from two of the subclusters (a total of 12 high SALL4 primary HCCs and 43 low SALL4 HCCs) yielded the data reported in Figure 3. P values were obtained by applying Kolmogorov-Smirnov test across different gene sets.

QUANTITATIVE REAL-TIME RT-PCR

Total RNA was extracted using Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA) and treated with RNase-free DNase (Qiagen) following manufacturer's instructions. Reverse transcription was carried out using High Capacity Reverse Transcription Kit with RNase inhibitor (Applied

Biosystems, Foster City, CA). Quantitative PCR was performed using GoTaq qPCR Master Mix (Promega, Madison, WI). Amplification was done with a Corbett Rotor Gene 6000 (Qiagen) using the following parameters: 95°C (10 min), 40 cycles of 95°C (15s), 60°C (60s), and 72°C (20s). All measurements were performed in duplicate. The following primers were used:

SALL4 F: 5'-GCGAGCTTTTACCACCAAAG-3'

SALL4 R: 5'-CACAACAGGGTCCACATTCA-3'

*SALL4A*F: 5'-TCCTGGAAACCACATCCTTC-3'

SALL4A R: 5'-ATGTGCCAGGAACTTCAACC-3'

SALL4B F: 5'-GGTGGATGTCAAACCCAAAG-3'

SALL4B R: 5'-ATGTGCCAGGAACTTCAACC-3'

ACTB F: 5'-CAGAGCCTCGCCTTTGCCGATC-3'

ACTB R: 5'-CATCCATGGTGAGCTGGCGGCG-3'

VIRAL TRANSDUCTION OF SHRNAS

Lentiviruses expressing scrambled shRNAs or SALL4-specific shRNAs were packaged by transfection of 293T cells with lentiviral vector pLL3.7 or pLKO.1. 24 hours and 48 hours post-transfection, viruses were harvested and filtered through 0.45 μm filters. Virus titers were determined by using infected 3T3 cells by the conventional ways. MOI 5 to 10 was used depending on individual cell lines. Transduction of HCC cells were carried out using spinoculation protocol. Briefly, virus and 8 μg/mL polybrene (Santa Cruz) were added to the trypsinized cells and let settle for an hour at 37°C in a humidified atmosphere of 5% CO_2, centrifuged at 2,200 RPM at 37°C for 90 minutes and incubated overnight at 37°C in a humidified atmosphere of 5% CO_2. Transduction efficiency was determined by GFP expression by FACS analysis. The following shRNAs were used:

Scr shRNA 1:

GGGTACGGTCAGGCAGCTTCTTTCAAGAGAAGAAGCTGCCTGACCGTACCCTTTTTTC;

Scr shRNA 2:

CAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT;

shSALL4 1:

GGCCTTGAAACAAGCCAAGCTATTCAAGAGATAGCTTGGCTTGTTTCAAGGCCTTTTTTC;

shSALL4 2:

TGCTATTTAGCCAAAGGCAAATTCAAGAGATTTGCCTTTGGCTAAATAGCTTTTTTC

WESTERN BLOT

Total cell lysates were harvested in NP-40 lysis buffer (150mM NaCl, 1% Nonidet P-40, 50mM Tris, pH8.0, protease inhibitor cocktail) and protein concentrations were determined by BCA protein assay (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Equal amounts of proteins from each lysate were resolved by SDS-PAGE and then transferred to nitrocellulose or PVDF membranes. Membranes were blocked with buffer containing 5% skim milk and 0.1% Tween 20 in PBS overnight at 4°C or 1 hour at room temperature with gentle shaking. Primary antibodies were incubated for one hour (α-tubulin and β-actin) or two hours (SALL4) at room temperature or overnight at 4°C (PTEN, AKT, pAKT, caspase-3, and cleaved caspase-3) with gentle shaking, followed by secondary antibody incubation at room temperature for one hour with gentle shaking. The following antibodies were used: SALL4 (Santa Cruz #sc-101147), β-actin (Santa Cruz #sc-47778), α-tubulin (Sigma #T6074), PTEN (Cell Signaling #9559), total AKT (Cell Signaling #2966), pAKT (Cell Signaling #9271), caspase-3 (Cell Signaling #9668), and cleaved caspase-3 (Cell Signaling #9664).

CELL VIABILITY ASSAY

5000 – 7000 cells were seeded in each well of a microtiter plate in 100 µL of medium. Cells from each treatment were seeded in duplicate. Controls using the same medium without cells were set up in parallel. At various time points, 317 µg/mL of 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (CellTiter96 AQ_{ueous} One Solution Cell Proliferation Assay, Promega) was added to each well. After two to four hours incubation, depending on cell line, of the tetrazolium salt, absorbance at 490nm was read by a microplate reader.

CASPASE 3/7 ASSAY

Capase 3/7 luminescent-based assay was carried out following manufacturer's instructions (Promega). Equal number of infected HCC cells was seeded in microtiter plates compatible for luminescence assay and substrate for Caspase 3/7 was added four days post-transduction to detect caspase activity.

MICROARRAY

Total RNA was extracted using Qiagen RNeasy Mini Kit (Qiagen) following manufacturer's instructions. RNA quantity was assessed using Nanodrop (Thermo Fisher Scientific, Wilmington, DE) and RNA integrity was analyzed using Nano chip for Eukaryotes on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). A RIN of 8.0 and above was considered to indicate a satisfactory sample quality. Gene expression array analysis was performed using Affymetrix GeneChip Human Gene 1.0ST Array system (Affymetrix, Santa Clara, CA) according to manufacturer's instructions. Briefly, for each sample, 200 ng of total RNA in a volume of 5 µl was amplified using the Applause WT Amp ST, RNA Amplification System (NuGen, San Carlos, CA)

and labeled with Encore Biotin Module, Post–Amplification System (NuGen). A total of 2.5 µg of labeled cDNA, along with GeneChip Hybridization Control reagents (Affymetrix), was injected into an Affymetrix GeneChip Human Gene 1.0 ST Array. The chips were incubated for 18 hours at 45°C and rotated at 60 RPM to allow hybridization. The chips were then washed and stained using GeneChip Hybridization Wash and Stain Kit (Affymetrix) using the Affymetrix GeneChip Fluidics Station 450. Stained arrays were scanned on Affymetrix GeneChip Scanner 3000 7G.

IN VITRO PEPTIDE TREATMENT ASSAY

SALL4 (nonmutant) and control (mutant and scrambled) peptides were synthesized (Biosynthesis Inc., Lewisville, Texas) using standard solid phase peptide synthesis chemistry and purified by the manufacturer to 95% purity. HCC cells were grown in 6-well plates to 50-70% confluence 24 hours prior to peptide treatment. For each treatment, 100 µL 1×PBS + 1µL diluted peptide was set up. Chariot reagent (Active Motif Inc., Carlsbad, CA), used as a peptide carrier, was diluted 1:10 in distilled water. Diluted peptide and diluted Chariot reagent were mixed gently and incubated for 30 minutes at room temperature. Medium was aspirated from the well and cells were overlaid with the mixture. 400 µL serum-free medium was then added and cells were incubated for one hour at 37°C in a humidified atmosphere of 5% CO_2. 500 µL complete medium was added one hour later. For determination of cell viability, the following doses of peptides were used: 0, 5 and 20 µM of nonmutant, mutant, or scrambled peptides, or 0, 50 and 100 nM of trichostatin A (TSA). Cell viability was examined 72 hours after peptide treatment. For western blot analysis, 20 µM of nonmutant or scrambled peptide, or 100nM of TSA was used to treat the cells. Cell lysates were harvested 72 hours following peptide treatment. 400 nM of SF1670 PTEN inhibitor was used throughout all experiments.

IN VIVO TUMORIGENICITY ASSAY

4- to 8-week-old NOD/SCID mice were used. Animal work was done at CHB with approval from IACUC. For loss of function studies, 10 million of SNU-398 or 6 million of HuH-7 cells infected with viruses expressing scrambled shRNA or SALL4-specific shRNA (total 300 µL of cell suspension and matrigel in 1:1 ratio) were injected subcutaneously into the right flank of NOD/SCID mice. For TAT-peptide treatment studies, 3 million SNU-398 cells were resuspended in 150 µL PBS, mixed with matrigel in a 1:1 ratio, and then transplanted subcutaneously into the left flank of each mouse. 56 mg/ kg body weight of TAT-mutant or 60 mg/ kg body weight of TAT-nonmutant peptides were administered by intraperitoneal injection for five consecutive days starting from the day of subcutaneous HCC cells transplantation. For all studies, after injection, mice were examined and tumor volumes were measured at various time points. Tumor volume was calculated by using the formula, tumor volume = $\pi/6$ X larger diameter X $(\text{smaller diameter})^2$. Tumor samples were processed for routine histology examination. Mice were sacrificed when tumors were too large to be compatible with life and survival analysis was done.

AUTHOR CONTRIBUTIONS

DGT, LC, MST and KJY designed the study. KJY, CG, JSJL, TD, SR, JML and DYY performed the experiments and acquired data. KJY, JSJL, RTP, STF, JML, YYD, MST, DGT and LC analyzed and interpreted acquired data. CWO, KFW and SS carried out statistical analysis. BY, SS and MST analyzed and interpreted histopatological data. CG, AK and XT provided technical supports for all the *in vivo* experiments. SL provided molecular biology technical supports. HY acquired and analyzed global gene expression and CNV data. KJY wrote the first draft of the manuscript. KJY, BY, JSJL, RTP, STF, JML, DYY, MST, DGT and LC participated in scientific discussion and drafting of the manuscript. MST, LC and DGT decided to publish the paper.

SUPPLEMENTARY FIGURES

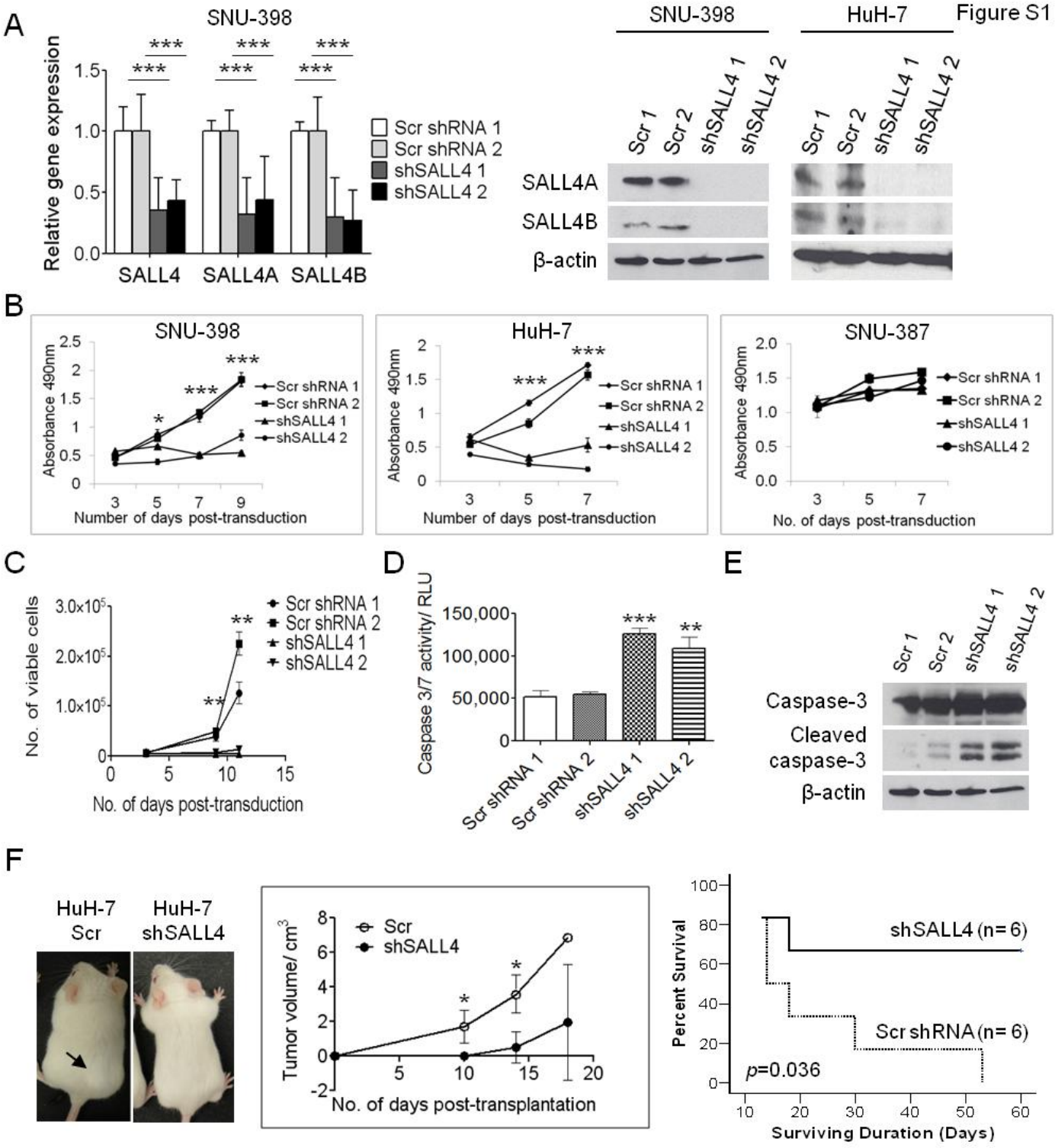

Figure S1. Loss of SALL4 leads to decreased HCC cell viability and tumorigenicity. (A) Left panel: qPCR analysis of relative *SALL4, SALL4A* and *SALL4B* expression in SNU-398 cells four days post-transduction. Error bars indicate standard error of three replicates; expression was normalized to *ACTB* and plotted relative to scrambled controls. *** P < 0.001. Right panels: Western blot analysis of SALL4 expression in SNU-398 and HuH-7 cells four days post-transduction. (B) MTS analysis of HCC cell viability upon *SALL4* gene knockdown by scrambled control shRNAs (Scr shRNA) or SALL4-specific shRNAs (shSALL4) in SNU-398 (left), HuH-7 (middle) and SNU-387 (right) cells. Error bars indicate standard error of three replicates. * P < 0.05; *** P < 0.0001. (C) Growth curve of SNU-398. ** P < 0.01. (D) Caspase 3/7 assay shows an increase in apoptosis in SALL4-knockdown SNU-398 cells at day 4 post-transduction. ** P < 0.01; *** P < 0.0001. (E) Western blot shows the expression of total caspase 3 and cleaved caspase 3 in SNU-398 cells four days post-transduction. (F) Effects of *SALL4* gene knockdown on tumorigenicity of HCC cells. Left panel: representative images show immunocompromised mice transplanted with scrambled shRNA 1- or shSALL4 1-treated HuH-7 cells. Arrow shows subcutaneous tumor on the right flank of the mouse. Middle panel: tumor volumes of the subcutaneous tumors. Statistical significance of the final time point could not be determined because there was only one mouse left in the scrambled control group. * P < 0.05. Right panel: Kaplan-Meier plot shows poorer survival advantage for mice harboring SNU-398 or HuH-7 cells infected with scrambled control shRNA 1 as compared to mice harboring cells infected with shSALL4 1. N = 12; P = 0.04.

Figure S2

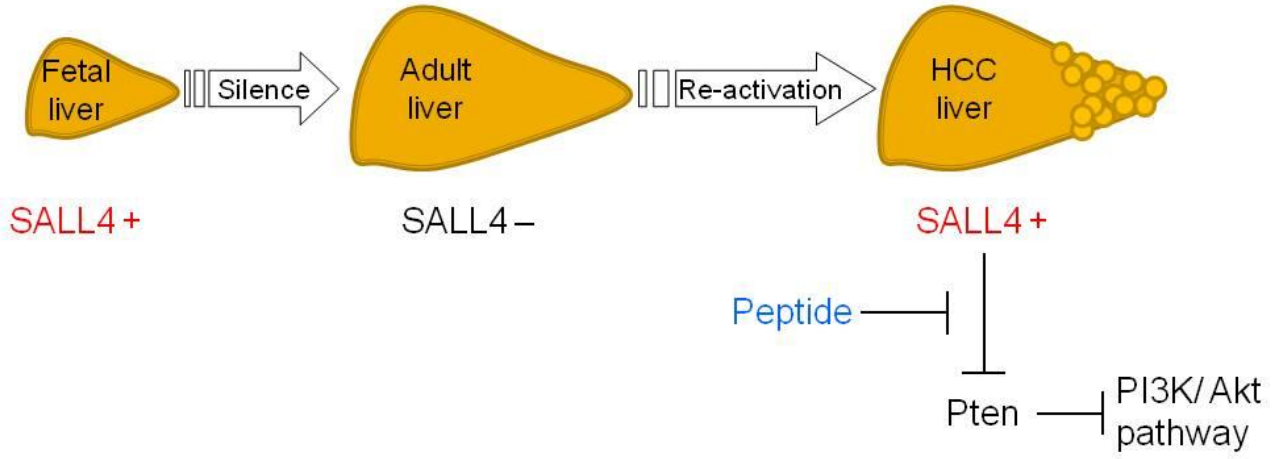

Figure S2. SALL4 is a novel oncofetal protein in HCC. In healthy humans, SALL4 is expressed in fetal liver but silenced in mature adult liver. In a subgroup of HCC livers, SALL4 is re-activated and plays a functional role in hepatocarcinogenesis by silencing the tumor suppressor PTEN through the recruitment of the NuRD complex. A therapeutic peptide can be used to block the interaction between SALL4 and the NuRD complex, thereby activating *PTEN* transcription. Upregulation of PTEN expression leads to downregulation of pAKT level and silencing of the PI3K/AKT survival signaling, resulting in decreased HCC cell viability and tumorigenicity. We propose SALL4 to be a novel oncofetal protein that can be specifically targeted for treatment of a subgroup of aggressive HCCs with SALL4 expression.

SUPPLEMENTARY TABLES

Table S1. Demographic and Clinicopathological Characteristics of the Patients (Singapore Cohort) and Association between SALL4 Expression with Clinicopathological Parameters.[a,b]

Clinicopathological Parameters	Frequency (%)	HCC MMPM		
		SALL4^{neg}	SALL4^{pos}	P value
Age, years				
< mean (56.2)	38 (48.1)	16	22	1.00
≥ mean	41 (51.9)	18	23	
Gender				
Female	15 (19)	6	9	1.00
Male	64 (81)	28	36	
Ethnicity				
Chinese	62 (78.5)	28	34	0.58
Non-Chinese	17 (21.5)	6	11	
Tumor grade				
I & II	69 (87.3)	48	21	1.00
III & IV	10 (12.7)	7	3	
TNM stage				
I & II	64 (81)	28	36	0.75
III	15 (19)	6	9	
Child-Pugh class				
A and B	68 (86.1)	30	38	0.75
C	11 (13.9)	4	7	

Clinicopathological Parameters	**Frequency (%)**	**HCC MMPM**		
		SALL4^{neg}	**SALL4^{pos}**	**P value**
BCLC stage				
A	57 (72.2)	22	35	0.22
B	22 (27.8)	12	10	
HBsAg (n=76)				
Negative	24 (31.6)	10	14	0.81
Positive	52 (68.4)	24	28	
HCV (n=78)				
Negative	74 (94.9)	34	40	0.13
Positive	4 (5.1)	0	4	
Serum AFP, ng/dl (n=70)				
< 10.0	21 (30)	11	10	0.31
≥ 10.0	49 (70)	19	30	
Tumor size, cm (n=78)				
< mean (6.2)	49 (62.8)	18	31	0.24
≥ mean	29 (37.2)	15	14	
Multinodularity				
No	56 (70.9)	22	34	0.33
Yes	23 (29.1)	12	11	
Lymphovascular invasion (n=53)				
No	47 (88.7)	19	28	1.00
Yes	6 (11.3)	2	4	

Clinicopathological Parameters	**Frequency (%)**	**HCC MMPM**		
		SALL4^{neg}	**SALL4^{pos}**	**P value**
Satellitosis				
No	67 (84.8)	27	40	0.34
Yes	12 (15.2)	7	5	
Portal hypertension				
No	66 (83.5)	28	38	1.00
Yes	13 (16.5)	6	7	
Pre-Op Treatment				
No	55 (69.6)	25	30	0.63
Yes	24 (30.4)	9	15	
Recurrence (n=37)				
Early (<2 years)	9 (24.3)	5	4	0.46
Late (≥2 years)	28 (75.7)	11	17	

HCC, hepatocellular carcinoma; MMPM, morpho-molecular prognostic model; SD, standard deviation; AFP, alpha-fetoprotein; TNM, tumor-node-metastasis

SALL4^{neg}: IHC score 0; SALL4^{pos}: IHC score +1 to +4

[a]N=79 (unless otherwise noted in brackets).

[b]Fisher's exact test

Table S2. SALL4 expression status of primary HCC tissues from the Singapore cohort

IHC score[a]	Frequency, *n*	Percentage (%)
0	76	44.4
1	75	43.9
2	12	7.02
3	4	2.34
4	4	2.34
Total	171	100

[a]SALL4 expression was scored according to published scoring criteria used for germ cell tumors[8-10], with some modifications. We scored SALL4 expression according to the percentage of tumor cells stained positive for SALL4, 0: no tumor cells stained positive, 1: 1 – 30% of tumor cells stained positive, 2: 31 – 50% of tumor cells stained positive, 3: 51 – 80% of tumor cells stained positive, 4: >80% of tumor cells stained positive.

Table S3. Demographic and Clinicopathological Characteristics of the Patients (HK Cohort) and Association between SALL4 Expression with Clinicopathological Parameters.[a,b]

Clinicopathological Parameters	Frequency (%)	SALL4 mRNA expression		
		< median	≥ median	P values
Sex				
Male	179 (78.5)	95	84	0.08
Female	49 (21.5)	19	30	
Age				
< 55	105 (46.1)	44	61	**0.02**
≥ 55	123 (53.9)	70	53	
Alcohol (n=217)				
No	133 (61.3)	65	68	0.52
Moderate/Heavy	84 (38.7)	45	39	
Tumor size, cm (n=217)				
< 5	77 (33.8)	37	40	0.67
≥ 5	151 (66.2)	77	74	
Alpha fetoprotein, ng/mL				
< 100	116 (50.9)	78	38	**>0.001**
≥ 100	112 (49.1)	36	76	
HBsAg				
Negative	31 (13.6)	22	9	**0.01**
Positive	197 (86.4)	92	105	
Histological differentiation (n=191)				
Well	37 (19.4)	24	13	**0.04**
Moderate/Poor	154 (80.6)	77	77	
BCLC stage (n=94)				
0 & A	67 (71.3)	39	28	0.38
B & C	27 (28.7)	13	14	
TNM stage				
Early (I, II)	103 (45.2)	56	47	0.27
Late (III, IV)	125 (54.8)	58	67	

Clinicopathological Parameters	**Frequency (%)**	**SALL4 mRNA expression**		
		< median	**≥ median**	**P values**
Early Recurrence (n=139)				
Absent	40 (28.8)	23	17	0.05
Present	99 (71.2)	39	60	
Late Recurrence (n=89)				
Absent	68 (76.4)	39	29	0.71
Present	21 (23.6)	13	8	
Venous infiltration (n=227)				
Absent	114 (50.2)	64	50	0.10
Present	113 (49.8)	49	64	

[a]N=228 (unless otherwise noted in brackets).

[b]Chi-square test

Table S4. Univariate overall survival, early recurrence (< 2 years) and late recurrence (≥ 2 years) analysis for clinicopathological features and SALL4 expression (SG cohort).[a]

	Overall Survival			Early Recurrence			Late Recurrence		
	HR	**95% CI**	**P value**	**HR**	**95% CI**	**P value**	**HR**	**95% CI**	**P value**
Clinicopathological parameters									
Age (<56.2 *vs* ≥56.2)	0.6	0.27-1.36	0.22	-	-	-	0.42	0.11-1.59	0.20
Gender (female *vs* male)	0.65	0.3-1.42	0.28	-	-	-	0.19	0.05-0.67	**0.01**
Ethnicity (chinese *vs* non-Chinese)	1.45	0.49-4.32	0.22	1.30	0.12-14.51	0.83	3.39	0.39-29.33	0.27
Tumour Grade (I&II *vs* III&V)	0.87	0.26-2.88	0.82	-	-	-	1.16	0.34-3.92	0.81
TNM (I&II *vs* III&IV)	1.51	0.56-0.41	0.42	-	-	-	1.22	0.32-4.63	0.77
Child-Pugh (A&B *vs* C)	2.22	0.82-6.0	0.12	-	-	-	1.47	0.31-6.85	0.63
BCLC stage (A *vs* B)	2.08	0.9-4.83	0.09	-	-	-	1.32	0.38-4.53	0.66
HBsAG (negative *vs* positive)	1.27	0.49-3.28	0.62	0.45	0.03-7.37	0.57	0.36	0.09-1.45	0.15
HCV (negative *vs* positive)	0.79	0.11-5.87	0.81	-	-	-	1.48	0.19-11.72	0.71
Serum AFP, ng/dl (<20.0 *vs* ≥20.0)	2.22	0.89-5.45	0.08	-	-	-	5.76	0.71-46.47	1.00
Tumor size, cm (<6.2 *vs* ≥6.2)	0.78	0.33-1.83	0.57	-	-	-	0.8	0.23-2.78	0.73
Multinodularity (no *vs* yes)	1.67	0.73-3.84	0.22	-	-	-	1.32	0.39-4.53	0.66
Lymphovascular invasion (no *vs* yes)	1.32	0.17-10.46	0.79	-	-	-	-	-	-
Satellitosis (no *vs* yes)	0.97	0.33-2.86	0.96	0.75	0.07-7.94	0.81	0.79	0.17-3.67	0.77
Portal hypertension (no *vs* yes)	0.99	0.34-2.91	0.98	-	-	-	0.76	0.09-5.98	0.79
Pre-op treatment (no *vs* yes)	1.74	0.77-3.91	0.18	-	-	-	1.73	0.51-5.9	0.38
Protein expression									
SALL4 (positive *vs* negative)	2.92	1.15-7.39	**0.02**	0.39	0.04-3.78	0.42	5.35	1.14-25.06	**0.03**

[a]Univariate analysis, Cox proportional hazards regression model.

Bold values denoted statistical significance at α=0.05; Dashes indicate no modeling as coefficients did not converge.

SALL4 negative: IHC score 0; SALL4 positive: IHC score +1 to +4.

HR, hazard risk ratio; CI, confidence interval; TNM, tumor-node-metastasis.

Table S5. Univariate overall survival, early recurrence (< 2 years) and late recurrence (≥ 2 years) analysis for clinicopathological features and SALL4 expression (HK cohort).[a]

	Overall survival	**Early recurrence**	**Late recurrence**
	P value	P value	P value
Clinicopathological parameters			
AFP, ng/mL (< 100 *vs* ≥ 100)	**0.01**	**<0.01**	0.06
No. of tumor nodule (< 3 *vs* ≥ 3)	**0.01**	**< 0.001**	0.83
Venous infiltration (Y *vs* N)	**< 0.001**	**< 0.001**	0.26
BCLC stage (0 & A *vs* B &C)	0.09	0.41	0.65
Sex	0.92	0.47	0.57
Age	0.53	0.28	0.92
Alcohol	0.23	0.13	0.31
Tumor size	**<0.01**	0.10	0.57
HBsAg	0.55	0.31	0.14
Histological differentiation	0.08	0.22	0.32
TNM stage	**< 0.001**	**<0.01**	0.06
Gene expression			
SALL4 (< median *vs* ≥ median)	**<0.01**	**0.02**	0.56

[a]Univariate analysis, Kaplan-Meier analysis.

Table S6. Multivariate overall survival, early recurrence (< 2 years) and late recurrence (≥ 2 years) analysis for the clinicopathological features and SALL4 expression (SG cohort).[a]

	Overall Survival			Early Recurrence			Late Recurrence		
	HR	**95% CI**	**P value**	**HR**	**95% CI**	**P value**	**HR**	**95% CI**	**P value**
Clinicopathological parameters									
Serum AFP, ng/dl (<20.0 vs ≥20.0)	2.08	0.83-5.22	0.12	-	-	-	3.2	0.25-40.48	0.37
BCLC stage (A vs B)	1.84	0.75-4.52	0.18	-	-	-	0.74	0.16-3.27	0.69
Protein expression									
SALL4 (positive vs negative)	2.87	1.09-7.52	**0.03**	0.62	0.05-6.99	0.69	2.7	0.41-18.72	0.29

[a]Multivariate analysis, Cox proportional hazard regression model. Variables were adopted for their prognostic significance (p<0.1) by univariate analysis. Dashes indicate no modelling as coefficients did not converge.

SALL4 negative: IHC score 0; SALL4 positive: IHC score +1 to +4.

HR, hazard ratio; CI, confidence interval

Table S7. Multivariate overall survival, early recurrence (< 2 years) and late recurrence (≥ 2 years) analysis for the clinicopathological features and SALL4 expression (HK cohort).[a]

	Overall survival			Early recurrence			Late recurrence		
	HR	**95% CI**	**P value**	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**
Clinicopathological									
Serum AFP, ng/mL (<100 *vs* ≥ 100)	-	-	-	-	-	-	0.31	0.11 – 0.86	**0.03**
No. of tumor nodules (< 3 *vs* ≥ 3)	-	-	-	2.33	1.45 – 3.76	**< 0.001**	-	-	-
Tumor size, cm (<5 *vs* ≥5)	1.71	1.05 – 2.76	**0.03**	-	-	-	-	-	-
TNM stage (I&II *vs* III&IV)	-	-	-	-	-	-	0.34	0.14 – 0.83	**0.02**
Venous infiltration (Y *vs* N)	0.28	0.17 – 0.44	**< 0.001**	0.42	0.27 – 0.66	**< 0.001**	-	-	-
Gene expression									
SALL4 (< median *vs* ≥ median)	1.52	1.00 – 2.32	**0.05**	1.67	1.11 – 2.51	**0.01**	-	-	-

[a]Multivariate analysis, Cox proportional hazard regression model. Dashes indicate no modeling as coefficients did not converge.

SUPPLEMENTARY REFERENCES

1. de Lope CR, Tremosini S, Forner A, Reig M, Bruix J. Management of HCC. J Hepatol 2012;56 Suppl 1:S75-87.

2. El-Serag HB. Hepatocellular carcinoma. N Engl J Med 2011;365:1118-27.

3. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020-2.

4. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012;56:908-43.

5. Sung WK, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765-9.

6. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:e15.

7. Chua SW, Vijayakumar P, Nissom PM, Yam CY, Wong VV, Yang H. A novel normalization method for effective removal of systematic variation in microarray data. Nucleic Acids Res 2006;34:e38.

8. Wang F, Liu A, Peng Y, et al. Diagnostic utility of SALL4 in extragonadal yolk sac tumors: an immunohistochemical study of 59 cases with comparison to placental-like alkaline phosphatase, alpha-fetoprotein, and glypican-3. Am J Surg Pathol 2009;33:1529-39.

9. Cao D, Guo S, Allan RW, Molberg KH, Peng Y. SALL4 is a novel sensitive and specific marker of ovarian primitive germ cell tumors and is particularly useful in distinguishing yolk sac tumor from clear cell carcinoma. Am J Surg Pathol 2009;33:894-904.

10. Mei K, Liu A, Allan RW, et al. Diagnostic utility of SALL4 in primary germ cell tumors of the central nervous system: a study of 77 cases. Mod Pathol 2009;22:1628-36.