Supporting Information for the article:

SOX17 Regulates Cholangiocyte Differentiation and Acts as a Tumor Suppressor in Cholangiocarcinoma

Short title: SOX17 promotes a normal biliary phenotype

Merino-Azpitarte M,[1] Lozano E,[1,2,3] Perugorria MJ,[1,3,4] Esparza-Baquer A,[1] Erice O,[1] Santos-Laso A,[1] O'Rourke CJ,[5] Andersen JB,[5] Jiménez-Agüero R,[1] Lacasta A,[1] D'Amato M,[1,4] Briz O,[2,3] Jalan-Sakrikar N,[6] Huebert RC,[6] Thelen KM,[7] Gradilone SA,[7] Aransay AM,[3,8] Lavín JL,[8] Fernández-Barrena MG,[9] Matheu A,[4,10] Marzioni M,[11] Gores GJ,[6] Bujanda L,[1,3] Marin JJG,[2,3] and Banales JM.[1,3,4]

[1]Department of Liver and Gastrointestinal Diseases, Biodonostia Research Institute – Donostia University Hospital –, University of the Basque Country (UPV/EHU), San Sebastián, Spain; [2]Experimental Hepatology and Drug Targeting (HEVEFARM), Biomedical Research Institute of Salamanca (IBSAL), University of Salamanca, Salamanca, Spain; [3]National Institute for the Study of Liver and Gastrointestinal Diseases (CIBERehd, "Instituto de Salud Carlos III"), Spain; [4]IKERBASQUE, Basque Foundation for Science; [5]Biotech Research and Innovation Centre, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [6]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA; [7]The Hormel Institute, University of Minnesota, Austin, MN, USA; [8]Genome Analysis Platform, CIC bioGUNE, Bizkaia Technology Park, Derio, Spain; [9]Hepatology Program, CIMA of the University of Navarra, Pamplona, Spain; [10]Neuro-Oncology Group, Biodonostia Institute, San Sebastian, Spain; [11]Department of Gastroenterology, "Università Politecnica delle Marche", Ancona, Italy.

Abbreviations: BSA, bovine serum albumin; CCA, cholangiocarcinoma; CDK4, cyclin-dependent kinase 4; CK7, cytokeratin 7; CK19, cytokeratin 19; DAPI, 4',6-diamidino-2-phenylindole; DMEM, dulbecco's modified Eagle's medium; DNMTs, DNA methyltransferases; DPBS, dulbecco's phosphate-buffered saline; ECL, enhanced chemoluminiscence; eGFP, enhanced green fluorescent protein; FBS, fetal bovine serum; FC, fold change; FW, forward; GAPDH, glyceraldehyde-3 phosphatase dehydrogenase; GO, gene ontology; HDAC6, histone deacetylase 6; HRP, horseradish peroxidase; IMF, immunofluorescence; KEGG, Kyoto encyclopedia of genes and genomes; MOI, multiplicity of infection; NaCl, sodium chloride; NaF, sodium fluoride; NHC, normal human cholangiocytes; O/N, overnight; ORF, open reading frame; P/S, penicillin/streptomycin; PEI, polyethylenimine; qPCR, quantitative polymerase chain reaction; RIPA, radio-immunoprecipitation assay; RV, reverse; S100A4, S100 calcium-binding protein A4; SDS, sodium dodecyl sulfate; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SFRP1, secreted frizzled-related protein 1; shRNA, small hairpin RNA; SOX17, SRY-related HMG-box 17; TBS, tris-buffered saline; Tris, tris(hydroxymethyl)aminomethane; WB, western blot; Wnt3a, wingless-type MMTV integration site family member 3A; ZO-1, zonula occluding 1.

Material and methods

Isolation and reprogramming of human myofibroblasts into cholangiocytes

Human myofibroblasts were isolated from biopsy specimens and then cultured and reprogrammed as previously described [1]. Briefly, isolated human myofibroblasts were transduced with transient Sendai viral vectors expressing octamer-binding transcription factor 4 (*OCT4*), SRY-related HMG-box 2 (*SOX2*), kruppel-like factor 4 (*KLF4*) and myc protoncogene protein (c*MYC*) in order to induce pluripotency. Induced pluripotent stem cells (iPSC) were seeded in culture plates pre-coated with Geltrex (Life Technologies) and cultured in mTeSR medium (STEMCELL Technologies) with 1% penicillin/streptomycin (P/S). iPSC were induced to definitive endoderm (DE) for 4 days using advanced “Roswell Park Memorial Institute” (RPMI) medium changed and supplemented daily with addition of 50 ng/mL Activin A and 50 ng/mL of recombinant human wingless-type MMTV integration site family member 3A (Wnt3a) in the presence of Matrigel (Corning). For hepatic specification (HS), DE cells were treated daily with 10 ng/mL fibroblast growth factor 2 (FGF2), 20 ng/mL bone morphogenic protein 4 (BMP4) and 50 ng/mL sonic hedgehog (SHH) for 4 days. HS cells were induced to hepatic progenitor (HP) cells for 4 days with daily exposure to 50 ng/mL SHH and 100 ng/mL recombinant Jagged-1. To generate iPSC-derived cholangiocytes (iDC), HP cells were treated for 4 days with H69 media [1] changed and supplemented daily with 10 ng/mL transforming growth factor β (TGFβ) and Geltrex was replaced by collagen (Corning) [1]. All growth factors were purchased from R&D Systems.

Isolation of normal human cholangiocytes

Normal human cholangiocytes (NHC) were isolated from bordering tissue samples obtained during surgery dissection of a local hepatic adenoma at the Mayo Clinic (Rochester, MN, USA); only tissue pieces informed as normal by an experienced pathologist were employed.

The procedure to isolate and culture NHC was carried out according to a novel protocol described by our group [2-4]. Briefly, liver tissue was cut in small pieces (approximately 1 mm^3) and placed in a 50 mL tube. Samples were digested for 30 min in a shaker bath at 37ºC with 25 mL of Dulbecco's modified Eagle's medium/Ham's F-12 nutrient mixture (DMEM/F-12) medium (Invitrogen) supplemented with 3% of fetal bovine serum (FBS), 1% of penicillin/streptomycin (P/S) (both from Invitrogen), 0.1% of BSA, 17 mg pronase, 12.5 mg type IV collagenase and 3 mg DNase (all three from Sigma). Digested tissue was sequentially filtered through 100 µm and 40 µm nylon meshes (Millipore, Bedford, MA). Trapped fragments between both meshes were collected, placed in a new 50 mL tube and incubated again for another 30 min with the aforementioned solution, but substituting pronase with 13 mg hyaluronidase (Sigma). Afterwards, a second series of sequential filtrations were performed and intrahepatic bile duct units ranging from 40 µm to 100 µm were resuspended in fully supplemented DMEM/F-12 medium and seeded on collagen-coated Cellstar flasks (Greiner Bio-One).

Quantitative polymerase chain reaction

Supplementary Table 2 shows the primers employed for quantitative polymerase chain reaction (qPCR). Expression of the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)

gene was used as housekeeping and the control group was considered 100% of expression.

Immunoblotting

Cells were seeded in a 6-well plate and left overnight (O/N) in quiescent medium (dulbecco's modified Eagle's medium (DMEM)/F12+Glutamax, Gibco). The following day, cells were scrapped with 80 µL of radio-immunoprecipitation assay (RIPA) lysis buffer [150 mM sodium chloride (NaCl), 50 mM tris(hydroxymethyl)aminomethane (Tris) pH 7.5, 0.1% sodium dodecyl sulfate (SDS), 1% Triton-100X, 0.5% sodium deoxycholate, protease inhibitor cocktail tablet (Roche) and phosphatase inhibitors (1 mM ortovanadate, 10 mM sodium fluoride (NaF), 100 mM β-glycerophosphate)] and incubated at -80ºC O/N for cell lysis and protein extraction. Afterwards, cells were centrifuged at 14,500 rpm for 10 min at 4°C. Supernatant was employed for protein measurement.

Changes in protein expression were detected by immunoblotting using 20 µg of protein from whole cell extract in 7.5 or 12.5% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and electro-transferred to a nitrocellulose membrane (BioRad). Once blocked with 0.5% skim milk powder/tris-buffered saline (TBS)-5% tween (TBS-Tween) (Milk) or 0.5% bovine serum albumin (BSA)/TBS-Tween (BSA), membranes were incubated O/N at 4°C with the appropriate primary antibody (Supplementary Table 3) at 1:1000 dilution in blocking solution. Horseradish peroxidase-conjugated secondary antibodies (Cell Signaling) at 1:5000 dilution in blocking solution (Milk or BSA) were incubated for 1h at room temperature and the *Novex® enhanced chemoluminiscence (ECL) horseradish peroxidase (HRP) Chemiluminiscent Substrate Reagent Kit* (Invitrogen) was used for further band visualization and quantitation with the *ChemiDoc™ MP System* (Bio-Rad). The β-actin protein expression was used as loading control.

Immunofluorescence in tissue samples

Human liver biopsies were fixed in 4% formaldehyde solution (Sigma) for 24 h and then embedded in paraffin (Merk Millipore). The resultant blocks were cut in 3-5 µm slices with a microtome and used for immunofluorescence techniques. Paraffin-embedded tissue samples were heated at 60ºC for 30 min and de-waxed in xylene. Afterwards, rehydration was carried out in decreasing grades of ethanol (100%, 96%, 70% and 50%). Antigenic unmasking was performed by boiling the tissue samples in Citrate buffer for 15 min. Samples were incubated O/N at 4ºC with the primary antibody in dulbecco's phosphate-buffered saline (DPBS) (1:100; Supplementary Table 3) or DPBS only as negative control. That step was followed by the incubation of the fluorescent secondary antibody (1:200) for 2 h and washed 3 times with DPBS. Finally, slides were mounted with a drop of *VECTASHIELD™* mounting medium with 4´,6-diamidino-2-phenylindole (DAPI, Vector laboratories). Pictures were taken with a *Nikon Digital Sight* camera under a fluorescence microscope (Eclipse 80i, Nikon) with the *NIS-elements AR 3.2 software* or with a *Zeiss LSM 510* confocal microscope [2].

Immunofluorescence in cell cultures

Cells were cultured on collagen-coated coverslips (Menzel-Gläser) in 24-well plates. Then, they were fixed with 1 mL of methanol for 10 min at -20ºC. Samples were washed 3 times with antigen retrieval solution (0.5% Triton-100X/PBS 1X) and incubated for 20 min with this solution at room temperature. Next, cells were incubated with blocking solution (5% FBS/1% BSA/PBS1X) for 30 min at room temperature and them with the primary antibody (1:100) (Supplementary Table 3) in 0.1% Triton-100X/1% BSA solution for 1h. Afterwards, cells were washed 3 times with 1% BSA/PBS 1X solution and incubated under darkness

with the corresponding fluorescent secondary antibody (1:1000) diluted in that same solution for 1.5 h. Finally, cells were washed 3 times with DPBS and coverslips were placed onto a microscope slide appropriate for immunofluorescence (Thermo Scientific) with a drop of *VECTASHIELD™* mounting medium with DAPI (Vector laboratories). Immunofluorescence images were obtained with a *Nikon Digital Sight* camera under a fluorescence microscope (Eclipse 80i, Nikon) with the *NIS-elements AR 3.2 software* or with a *Zeiss LSM 510* confocal microscope. Immunofluorescence of ciliary-associated proteins was performed in the same conditions but in cells under 7 days of confluence.

Cell infection with lentiviruses

NHC were infected by lentiviruses carrying small hairpin RNA (shRNA) against SRY-related HMG-box 17 (*SOX17)* or control (Santa Cruz) at a MOI (multiplicity of infection) of 3 and cholangiocarcinoma (CCA) human cells (EGI1) were infected with lentiviruses carrying *SOX17* or empty viruses (negative control) at a MOI 3.

Lentiviruses overexpressing SOX17 were produced as previously described [5]. Briefly, human *SOX17* open reading frame (ORF) was amplified from total RNA isolated from NHC cells by reverse transcription followed by high-fidelity PCR using *AccuPrime Pfx* DNA polymerase (Life Technologies), specific primers and an *Eppendorf Mastercycle ep gradient S Thermal Cycler* (Thermo Fisher). SOX17 cDNA was cloned into the *Pac*I site of the pWPI lentiviral vector under the regulation of the constitutive elongation factor 1α promoter. The identity of the cloning was confirmed by sequencing. The pWPI lentiviral vector also contains the enhanced green fluorescent protein (*eGFP*) gene. Recombinant lentiviruses were produced in HEK293T cells (ATCC, american type culture collection": CRL-11268) and using a standard polyethylenimine (PEI; Sigma) protocol. Thus, HEK293T cells were transfected with the pWPI-SOX17 vector or with the empty pWPI

vector (negative control) and the packaging plasmids psPAX2 and pMD2.G. To form complexes PEI:DNA 6 µg of pWPI-SOX17 (or empty pWPI), 6 µg of psPAX2 and 4.5 µg of pMD2.G were dissolved in 1.2 mL of saline solution and 60 µL of PEI. After 20 min of incubation at room temperature, the mixture was added in the HEK293T culture. The following 3 days the supernatant was collected, and finally was filtered (0.45 µm pore size) and ultracentrifuged (53,000 g, 120 min, 16ºC). The lentivirus concentration was determined by infecting HEK293T cells with serial dilutions of the viral solution in the presence of polybrene (Sigma) and the following analysis of eGFP-positive cells by a *FACSCalibur* flow cytometer (BD Biosciences).

Cell Death Analysis

Cell death was determined by using three different flow cytometry-based assays: Annexin-V (Invitrogen), Propidium Iodide (Invitrogen) and Caspase-3 activity (Phiphilux-G2D2, OncoImmunin® Inc), following the manufacturer's instructions. Briefly, 50,000 CCA (EGI1) and NHC cells were seeded in each well of a collagen-coated 24-well plate, and infected with Lent-SOX17 or Lent-control for 48 h before cell death analysis. Non-infected cells were also used as controls.

Cell proliferation

NHC and CCA (EGI1) cells (50,000) were seeded in each well of a collagen-coated 24-well plate, and then infected with Lent-shRNA-*SOX17* (or Lent-shRNA-control) and/or Lent-SOX17 (or Lent-control). Also, non-infected cells were used as controls. In CCA cells under Lent-SOX17 infection, the remaining living cells that survived apoptosis (48 h after infection) were reseeded in a 96-well plate (5,000 cells per well). Next, cell proliferation

was evaluated in the presence or absence of Wnt3a protein (R&D systems) for 48 h by using the *Cell Proliferation WST-1 Assay* (Roche) according to the manufacturer's instructions. EGI1 cells were cultured with DMEM/F-12 medium supplemented with 1% FBS and 1% P/S and incubated with 400 ng/mL of Wnt3a, and TFK1 cells were cultured in DMEM/F-12 medium supplemented with 1%P/S and incubated with 200 ng/mL of Wnt3a.

Anchorage-independent growth in soft agar

Non-infected or infected (with Lent-control or Lent-SOX17: MOI 3 for 96 h) CCA (EGI1) human cells were harvested and resuspended in 0.3% agarose (Sigma) and $8x10^4$ cells per well were seeded over a 0.5% agarose layer in 6-well plates and then supplemented with completed cholangiocyte medium (c.f. culture conditions in Supplementary data). After 3 weeks, colonies were fixed [25% methanol (Panreac)], stained [0.0001% crystal violet (Sigma)] and counted.

Cell senescence

NHC were used to analyze cell senescence over passages (5, 10 and 15) *in vitro*. In addition, low passages NHC (8) were infected for 48 h with Lent-shRNA-*SOX17* or Lent-shRNA-control. Non-infected cells were used as controls. In all cases 20,000 cells were seeded in each well of a collagen-coated 24-well plate. The senescence *β-galactosidase kit* (Cell signaling) protocol was followed according to the manufacturer's instructions.

Cell migration

CCA (EGI1) cells (50,000) were seeded in each well of a collagen-coated 24-well plate, and then infected with Lent-SOX17 or Lent-control. Non-infected cells were also used as controls. In CCA cells under Lent-SOX17 infection, the remaining living cells that survived apoptosis (48 h after infection) were reseeded (200,000 cells per well) in a collagen-coated 6-well plate with dulbecco's modified eagle's medium (DMEM)/F12+Glutamax and 1%P/S. Once cells reached 70-80% confluence, media was replaced by serum-free DMEM/F12+Glutamax/1%P/S for 24 h. Then, three longitudinal scratches were done in the surface of each well (under 100% cell confluence) with a 20 µL pipette tip and pictures of the same areas were taken to monitor the scratch area. After 24 h, cells were fixed and stained with 4% formaldehyde and 0.5% crystal-violet (Sigma) in PBS for 20 min. The wound-healing area was measured by using the *Image J* software and the ratio to the initial area was calculated.

Illumina mRNA expression array

Illumina Human HT12 v4 BeadChips (GPL10558) [6-9] were used to characterize gene expression in NHC and CCA (EGI1) cells. NHC uninfected or infected with Lent-shRNA-SOX17 or Lent-shRNA-control were collected 1 week after infection. On the other hand, CCA cells, both uninfected and infected with Lent-SOX17 or Lent-control were collected 6 h after infection. Total RNA was isolated using the *miRNeasy Micro kit* (Qiagen). The quality and purity of the samples was measured using a RNA *Nano Chip Bioanalyzer* (Agilent Technologies). 200 ng of each RNA sample were used for the array. The cRNA synthesis, amplification, labeling and hybridization of the samples were performed following the *Whole-Genome Gene Expression Direct Hybridization* protocol (Illumina Inc.).

The cRNAs of the samples were hybridized to the diverse gene-probes of the array and the gene expression levels of the samples were detected by a HiScan scanner (Illumina Inc.). Raw data were extracted with GenomeStudio analysis software (Illumina Inc.), in the form of GenomeStudio's Final Report (sample probe profile). First, raw expression data were background-corrected, log2-transformed and quantile-normalized using the *lumi* R package [7], available through the Bioconductor repository. Probes with a "detection p-value" lower than 0.01 in at least one sample were selected. For the detection of differentially expressed genes, a linear model was fitted to the probe data and empirical Bayes moderated t-statistics were calculated using the *limma* package [9] from Bioconductor. Only genes with differential fold-change (FC) >1.5 or <-1.5 and a p-value < 0.05 were considered as differentially expressed. Gene Ontology enrichment assessment was achieved according to gene ontology (GO) [6] and Kyoto encyclopedia of genes and genomes (KEGG) [8] database terms. A Hypergeometric test is applied to compute the relative abundance of detected terms, thus, calculating the functional enrichment. Additionally, complementary gene function analysis was performed with other database programs such as Panther (WebGestalt, String), Pubmed, Uniprot and Genecards.

Knock-down of DNMT1 mRNA expression with siRNAs and inhibition of DNMT activity pharmacologically

The DNA methyltransferase 1 *(DNMT1)* mRNA expression was knocked-down in CCA human cells (EGI1) by using specific small interference RNA (siRNA) oligos against *DNMT1* mRNA (siRNA-DNMT1; sc-35204, Santa Cruz). In parallel, negative control siRNAs (siRNA-control; sc-37007, Santa Cruz) and non-transfected cells were used as controls. Briefly, cells were seeded in 12-well collagen-coated plates (100,000 cells per well) and transfected with siRNA-*DNMT1* or siRNA-control using *Lipofectamine RNAiMAX*

(Invitrogen). Cells were collected 48h after siRNA transfection for the RNA expression analysis of *DNMT1* and *SOX17*. The protein expression of SOX17 was also measured in CCA cells (EGI1, TFK1 and Witt) under the presence or absence of 5-aza-2'-deoxycytidine (1 μM) (also known as decitabine) for 48h. On the other hand, the mRNA expression of *SOX17* in NHC was measured in the presence or absence of Wnt3a (100 ng/mL, R&D systems) in DMEM/F12+Glutamax/1%P/S culture media for 48h and with or without decitabine (pre-incubated for 18 h).

Analysis of SOX17 promoter methylation

The methylation status of *SOX17* promoter was studied using *Infinium HumanMethylation27k BeadChips*. Briefly, DNA was extracted from 48 CCA specimens obtained during surgical resection and 5 normal intrahepatic bile duct samples (Copenhagen cohort). Genomic DNA was isolated using *AllPrep DNA/RNA Universal kit* (Qiagen) and subjected to bisulfite modification with *EZ DNA Methylation-Gold kit* (Zymo Research). Genome-wide DNA methylation profiling was performed as described in the Illumina guidelines and bisulfate-converted-DNA-hybridized-chips scanned using *iScan system* (Illumina). Data were analyzed using RnBeads [10] (including Illumina normalization, removal of probes on the X chromosome and those containing 2 or more SNPs). DNA methylation was quantified using beta (β)-value metric (range: 0-1; 0 = 0% methylation, 1 = 100% methylation). Differentially methylated probes (DMPs) were called in cases where mean $\Delta\beta \geq 0.17$ or ≤ -0.17 and Fisher's P-value < 0.05.

To further investigate the extent of aberrant DNA methylation in *SOX17*, level 1 *Infinium HumanMethylation450 BeadChip* (Illumina Inc.) data were analyzed for 35 CCA patients and 6 surrounding normal specimens obtainable through the cancer genome atlas (TCGA). For 450k data, SNP-enriched probes (≥2 SNPs in 50 base pair probe sequence), X chromosome probes and poor quality probes (as determined by Greedycut algorithm) were excluded. Data were normalized using beta mixture interquartile (BMIQ) normalization [10] and processed methylation values exported as β-values. Differentially methylated probes (DMPs) and differentially methylated regions (DMRs) were computed using the *RnBeads-integrated limma* method (hierarchical linear model) followed by model fitting using an empirical Bayes approach.

Supplementary Table 1. Clinical information of the San Sebastian cohort of patients.

Patient ID	Disease	Gender	Age	Stage	Size (cm)	CA19-9	CEA
1	ND	M	60	N/A	N/A	N/A	N/A
2	ND	F	70	N/A	N/A	N/A	N/A
3	ND	M	65	N/A	N/A	N/A	N/A
4	ND	F	68	N/A	N/A	N/A	N/A
5	ND	F	76	N/A	N/A	N/A	N/A
6	ND	M	55	N/A	N/A	N/A	N/A
7	ND	M	62	N/A	N/A	N/A	N/A
8	ND	M	79	N/A	N/A	N/A	N/A
9	ND	M	65	N/A	N/A	N/A	N/A
10	ND	M	31	N/A	N/A	N/A	N/A
11	ND	M	50	N/A	N/A	N/A	N/A
12	ND	M	50	N/A	N/A	N/A	N/A
13	ND	F	77	N/A	N/A	N/A	N/A
14	ND	F	79	N/A	N/A	N/A	N/A
1	iCCA	M	85	II	3.8	-	-
2	dCCA	M	61	IIIB	1.5	73.7	1.8
3	iCCA	M	62	II	2,7	2963	2.3
4	iCCA	F	78	IVB	6	43.2	1.2
5	iCCA	M	70	II	2	19.7	1.9
6	iCCA	F	82	IIIA	5.5	3528	307.9
7	pCCA	F	47	II	3.2	394.1	1.6
8	iCCA	M	61	II	4	0.6	2.8
9	iCCA	F	64	IIIA	5	374	2.9
10	iCCA	M	80	IVB	1	0.7	4.6
11	dCCA	M	78	IIIA	2.3	16112	4.4
12	iCCA	M	68	IV	1.7	4.5	0.8
13	iCCA	M	60	II	5.2	19	2.9

Abbreviations: ID, identification; F, female; M, male; dCCA, distal CCA; iCCA, intrahepatic CCA; pCCA, perihilar CCA; N/A, not applicable; ND: no disease; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen.

Supplementary Table 2. Human primers used for qPCR.

Primers	Sequence 5'-3'
CDK4 FW	ATGGCTACCTCTCGATATGAGC
CDK4 RV	CATTGGGGACTCTCACACTCT
Cytokeratin 7 FW	ATCTTTGAGGCCCAGATTGC
Cytokeratin 7 RV	TTGATCTCATCATTCAGGGC
Cytokeratin 19 FW	CAACGAGAAGCTAACCATGC
Cytokeratin 19 RV	ATTGGCTTCGCATGTCACTC
Fibronectin FW	GGGCAACTCTGTCAACGAAG
Fibronectin RV	CACACCATTGTCATGGCACC
GAPDH FW	CCAAGGTCATCCATGACAAC
GAPDH RV	TGTCATACCAGGAAATGAGC
HDAC6 FW	GATGCTGACTACCTAGCTGC
HDAC6 RV	CGATGGACTTGGATGGTCTC
p16ink4a FW	GGGGGCACCAGAGGCAGT
p16ink4a RV	GGTTGTGGCGGGGGCAGTT
S100A4 FW	ACGTGTTGATCCTGACTGCT
S100A4 RV	CCTGTTGCTGTCCAAGTTGC
SFRP1 FW	CTACTGGCCCGAGATGCTTA
SFRP1 RV	GCTGGCACAGAGATGTTCAA
SOX17 FW	GTGGACCGCACGGAATTTG
SOX17 RV	GGAGATTCACACCGGAGTCA
ZO-1 FW	CGGTCCTCTGAGCCTGTAAG
ZO-1 RV	GGATCTACATGCGACGACAA
DNMT1 FW	GTACTGTAAGCACGGTCACC
DNMT1 RV	AGATCTTCTCCTGCATCAGC

DNMT3a FW	CCTGGAACTGCTACATGTGC
DNMT3a RV	TCCACCTGAATGCCCAAGTC
DNMT3b FW	AGGGAAGACTCGATCCTCGTC
DNMT3b RV	GTGTGTAGCTTAGCAGACTGG

(FW: Forward; RV: Reverse)

Supplementary Table 3. Antibodies used for western blot and/or immunofluorescence.

Antibody	Company	Reference	Use
Goat polyclonal anti-SOX17	R&D	AF1924	WB, IMF
Mouse monoclonal anti-acetylated α-tubulin	Sigma-Aldrich	T7451	IMF
Mouse monoclonal anti-CK7	Santa Cruz	sc-23876	WB
Rabbit polyclonal anti-γ-tubulin	Abcam	Ab11320	IMF
Mouse monoclonal anti-p-p53 (Ser15)	Cell signaling	#9286	WB
Mouse monoclonal anti-β-actin	Sigma	A5316	WB
Rabbit polyclonal anti-β-catenin	Cell signaling	#9581	WB, IMF
Rabbit polyclonal anti-Fibronectin	Abcam	ab2413	WB
Rabbit polyclonal anti-GAPDH	Abcam	ab22555	WB
Mouse monoclonal anti-KRT19 (CK19)	ARP	#03-61029	WB
Rabbit polyclonal anti-p21	Abcam	ab7960	WB
Rabbit polyclonal anti-p53	Novocastra (Leica)	NCL-p53-CM1	WB

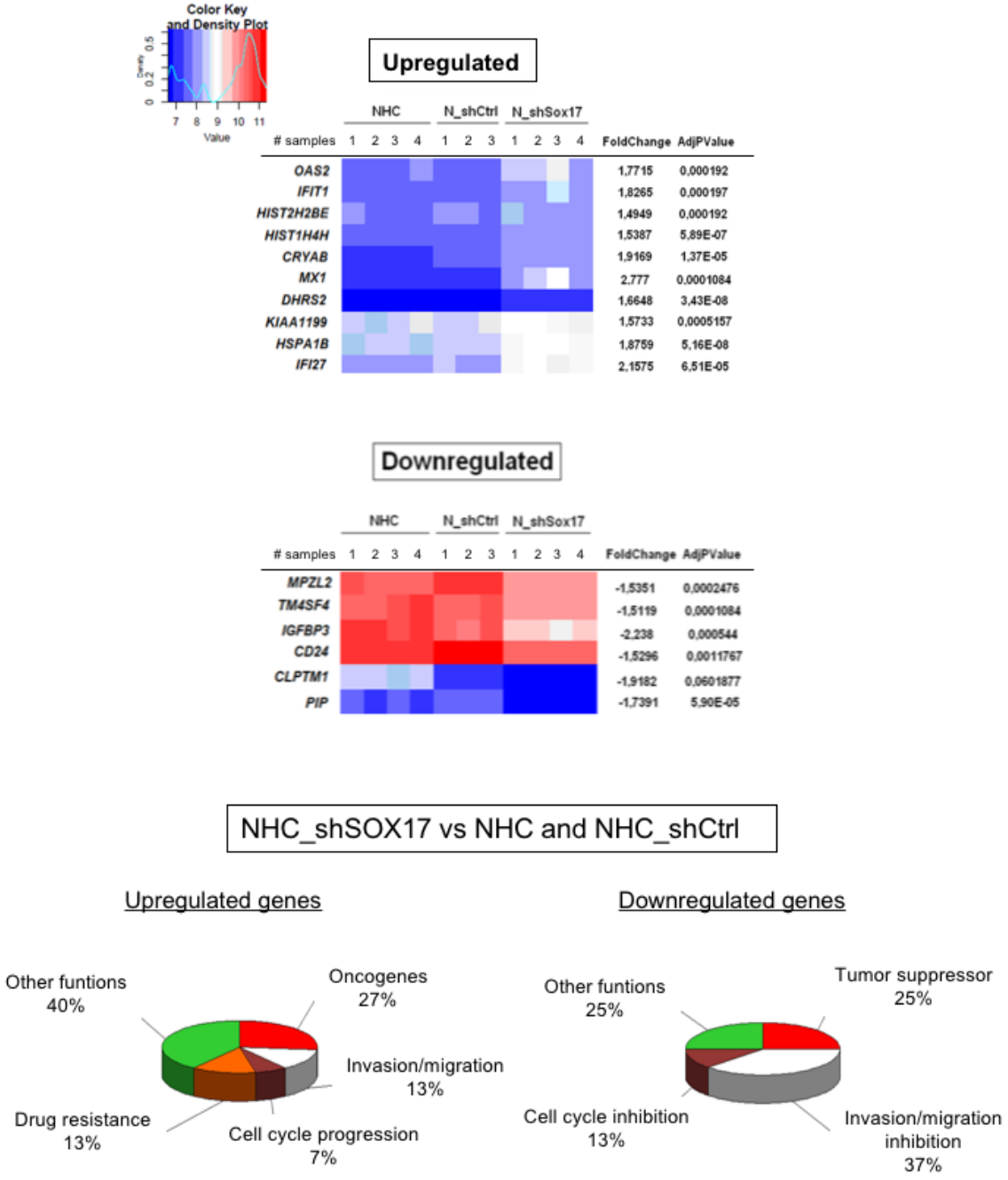
Rabbit polyclonal anti-p-β-catenin (Ser33/37/Thr41)	Cell signaling	#9561	WB
Rabbit polyclonal anti-SOX17	Abcam	ab89954	IMF
Anti-rabbit IgG, HRP-linked Antibody	Cell signaling	#7074	WB
Anti-mouse IgG, HRP-linked Antibody	Cell signaling	#7076	WB
Donkey anti-goat IgG-HRP	Santa Cruz	sc-2020	WB
Chicken anti-Goat IgG (H+L) Secondary Antibody, Alexa Fluor® 488 conjugate	ThermoFisher	A21467	IMF
Donkey anti-Goat IgG (H+L) Secondary Antibody, Alexa Fluor® 568 conjugate	ThermoFisher	A11057	IMF
Donkey anti-Mouse IgG Secondary Antibody, Alexa Fluor® 568 conjugate	ThermoFisher	A10037	IMF
Donkey anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor® 488 conjugate	ThermoFisher	A21206	IMF

Abbreviations: WB: western blot; IMF: immunofluorescence

Results

Effect of SOX17 downregulation in the cholangiocyte phenotype

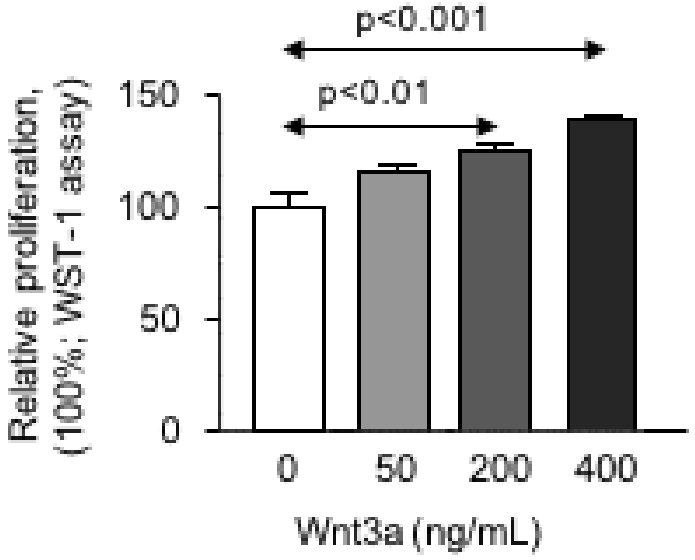
mRNA microarrays were carried out in order to test the effect of *SOX17* downregulation on the mRNA prolife of NHC. Heatmaps showed that most of the differentially-expressed genes were found overexpressed under SOX17 downregulation, which are mainly oncogenes, cell-cycle promoters, invasion/migration inducers and genes involved in drug resistance and cell survival (Supplementary Figure 1).



Supplementary Fig. 1. Heatmaps of the dysregulated genes and clustering by cellular functions. (n=3-4; Bayes moderated t-statistics)

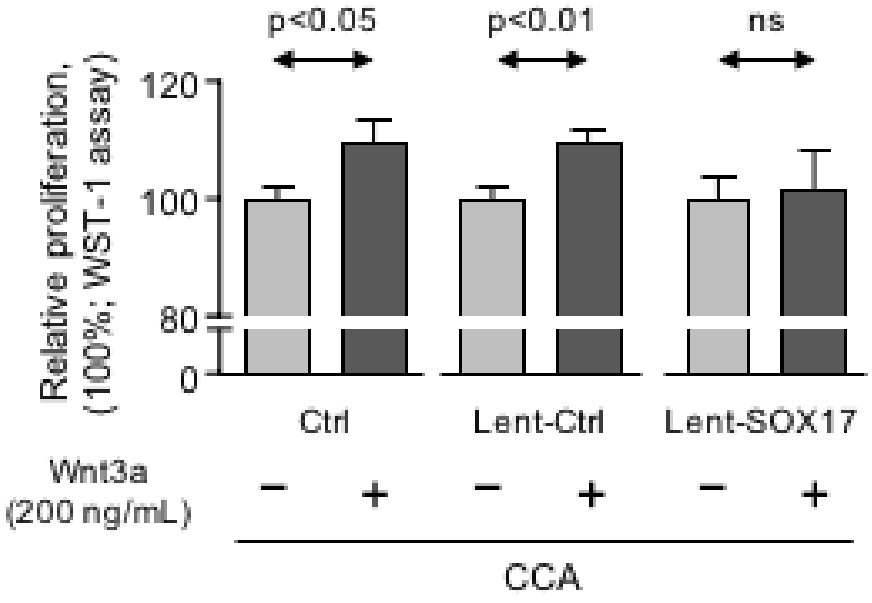
Role of Wnt3a in CCA cell proliferation

The presence of Wnt3a promotes the proliferation of CCA (EGI1) cells in a dose-dependent manner (Supplementary Figure 2).



Supplementary Fig. 2. Effect of Wnt3a on CCA cell proliferation (EGI1). (n=6 samples per condition; mean±SEM; p-value *vs* control, Student's t test).

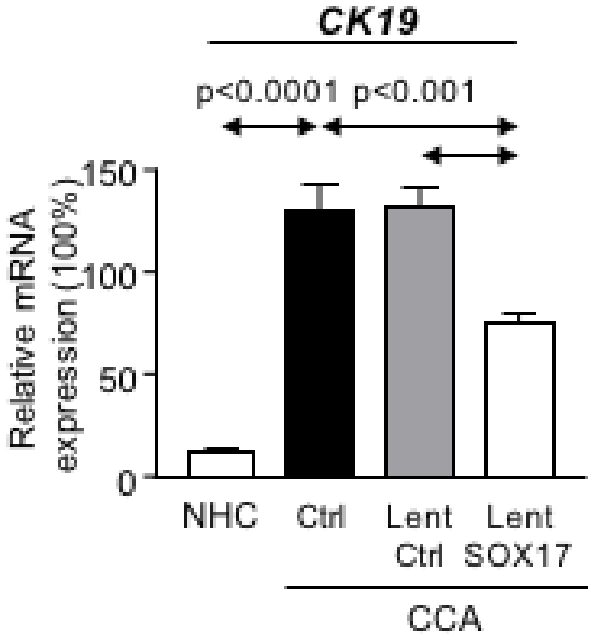
Similarly to the results obtained in EGI1 cells (Figure 6A), experimental overexpression of SOX17 in TFK1 (CCA) cells blocked the Wnt3a-dependent proliferation compared to controls (Supplementary Figure 3).



Supplementary Fig. 3. Role of SOX17 in the Wnt3a-dependent proliferation of TFK1 cells. (n=9-22 samples per condition; p-value *vs* controls, Student's t test).

Expression of CK19 in NHC and CCA cells in baseline conditions and under experimental modulation of SOX17 expression

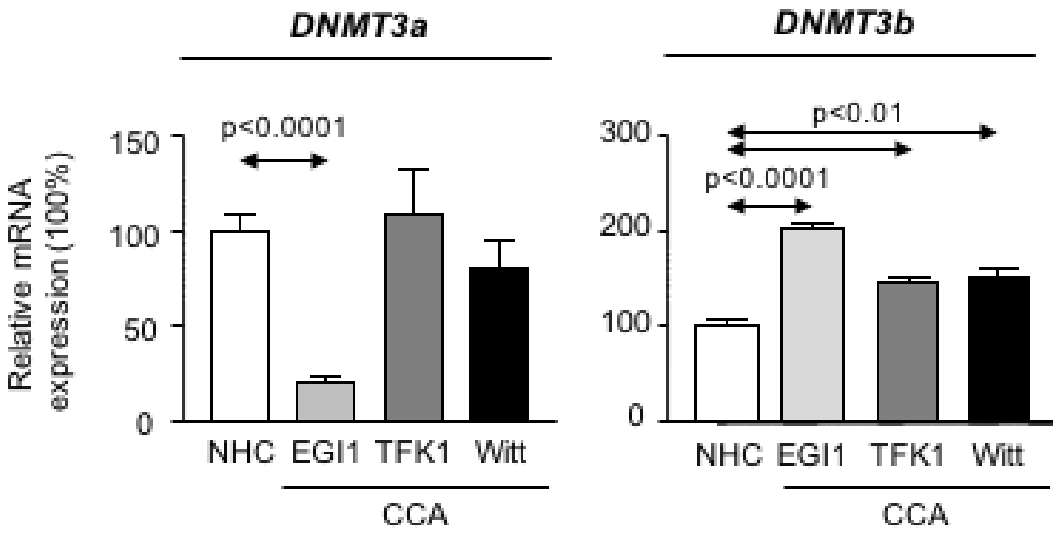
CK19 mRNA expression is upregulated in CCA (EGI1) cells compared to NHC under baseline conditions. Experimental overexpression of SOX17 in CCA cells inhibited the upregulated CK19 expression compared to control conditions (Supplementary Figure 4).



Supplementary Fig. 4. Expression of *CK19* in NHC and CCA cells under SOX17 modulation. (n=6-7 samples per condition; p-value *vs* controls, ANOVA+Tukey post hoc test).

Comparative expression of DNMTs 3a and 3b in normal human cholangiocytes and CCA human cells

DNMT3b is overexpressed in all three CCA cell lines (i.e. EGI1, TFK1 and Witt) compared to NHC, whereas *DNMT3a* showed no general expression changes between CCA and NHC (Supplementary Figure 5).



Supplementary Fig. 5. Expression of *DNMTs 3a* and *3b* in CCA *vs* NHC cells. (n=5-6 samples per condition; p-value *vs* control, Student's t test)

References

1. De Assuncao TM, Sun Y, Jalan-Sakrikar N, Drinane MC, Huang BQ, Li Y, et al. Development and characterization of human-induced pluripotent stem cell-derived cholangiocytes. *Lab Invest* 2015;95(6):684-696.
2. Banales JM, Saez E, Uriz M, Sarvide S, Urribarri AD, Splinter P, et al. Up-regulation of microRNA 506 leads to decreased Cl-/HCO3- anion exchanger 2 expression in biliary epithelium of patients with primary biliary cirrhosis. *Hepatology* 2012; 56(2):687-697.
3. Munoz-Garrido P, Marin JJ, Perugorria MJ, Urribarri AD, Erice O, Saez E, et al. Ursodeoxycholic acid inhibits hepatic cystogenesis in experimental models of polycystic liver disease. *J Hepatol* 2015;63(4):952-961.
4. Urribarri AD, Munoz-Garrido P, Perugorria MJ, Erice O, Merino-Azpitarte M, Arbelaiz A, et al. Inhibition of metalloprotease hyperactivity in cystic cholangiocytes halts the development of polycystic liver diseases. *Gut* 2014;63(10):1658-1667.
5. Lozano E, Monte MJ, Briz O, Hernandez-Hernandez A, Banales JM, Marin JJ, Macias RI. Enhanced antitumour drug delivery to cholangiocarcinoma through the apical sodium-dependent bile acid transporter (ASBT). *J Control Release* 2015;216:93-102.
6. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25-29.
7. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008, 24(13):1547-1548.
8. Kanehisa M. The KEGG database. *Novartis Found Symp* 2002, 247:91-101; discussion 101-103;119-128, 244-152.
9. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.
10. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* 2014;11(11):1138-1140.