nature research

Corresponding author(s): Holger Willenbring

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

## Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample sizes of sufficient statistical power were chosen based on previously published studies using similar analyses (PMIDs 27210761, 24631193, 21282722, 22105172).

2. Data exclusions

   Describe any data exclusions.

   For ink injection, livers that were misinjected or improperly dehydrated were excluded. For RNA-seq, samples with RIN below 7.7 or RNA yield below 15 ng were excluded. For sparse-labeling clonal analysis, clones that extended to and potentially beyond the x, y or z boundaries were excluded. For Sirius-red-staining quantification, tissue sections that were damaged or folded were excluded.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   For each series of experiments, attempts at replication were successful. A statement describing experimental replication for all figures is included in the Quantification and statistics section of the Methods.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Male and female mice of the indicated age and genotype were chosen randomly for inclusion in experiments, with the exception of RNA-seq where only female mice were used.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Researchers were not blinded when analyzing results. The hepatic phenotype of Alb-Cre +/-;Rbpjf/f;Hnf6f/f mice is distinguishable by gross examination.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ Test values indicating whether an effect is present<br>*Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.* |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation) |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Imaging: Thin sections and cultured cells were imaged using Openlab software (PerkinElmer). Fiji (ref. 50) and/or Photoshop (Adobe) software was used to process (brightness, contrast and gamma) and merge channels. Images of thick sections for 3D analysis of connectivity and clonal analysis were processed and analyzed using Imaris (Bitplane) or Volocity (PerkinElmer) software. For R26R-Confetti+/- mouse and pSMAD3 analysis, images were acquired using NIS elements software (Nikon Center of Excellence Confocal Imaging Core at CCHMC) and processed and analyzed using Imaris software. Gen5 Microplate Reader and Imager Software was used to capture images for Sirius-red-staining analysis on a Cytation 5 cell imaging multi-mode reader (BioTek).

Cholangiocyte and hepatocyte isolation: Cells were analyzed and sorted on a FACSAria III using FACSDiva software (BD Biosciences). FlowJo (FlowJo, LLC) was used to analyze data and generate charts.

RNA-seq: Raw reads were aligned to the mm10 mouse genome with annotations provided by UCSC using CobWEB, a proprietary Burrows-Wheeler Transform method.

Western blot: Quantification was performed using ImageJ.

Quantification and statistics: For clonal analysis in R26R-Confetti+/- mice, 30-40 µm z-stack images of wsCK-positive DBA-negative pBDs and wsCK-positive DBA-positive hBDs were visualized in 3D with Imaris software. The module Surfaces within Imaris was used to render a 3D surface created on an intensity value on a per channel basis. For Sirius-red-staining quantification, Fiji software was used to set a threshold for Sirius-red-positive area within each lobe and the percent of the total area that was Sirius red positive was measured and is reported. Charts were generated in Prism 6 or 7 (GraphPad).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

All materials are commercially available, with the exception of unique materials, which are available from the authors on reasonable request.

## 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Primary and secondary antibodies are detailed below and in the Supplementary Tables 2 and 3, respectively. All antibodies (except OC2-2F8) are commercially available and validation is provided in the data sheets of the manufacturer. Validation of OC2-2F8 is provided in PMID 27257763.

We used the following primary antibodies for staining of cells and tissue samples. They are listed as antigen first, followed by dilution, host, supplier, catalog number and clone/lot number as applicable.

1) ACTIN, 1:500, Mouse, Seven Hill Bioreagents, LMAB-C4, Clone C4
2) acTUB, 1:100, Mouse, Abcam, ab24610, Clone 6-11B-1
3) CD11b-PE-Cy7, 1:100, Rat, Tonbo Bio, 60-0112, C0112012616602, C0112012616603
4) CD31-PE-Cy7, 1:100, Rat, BioLegend, 102417, B212261, B212262
5) CD45-PE-Cy7, 1:100, Rat, Tonbo Bio, 60-0451, C0451080615602, C0451080615603
6) CK7, 1:25, Rabbit, Ventana, 790-4462, Clone SP52
7) CK19, 1:100, Rabbit, Abbomax, 602-670, 1101-766
8) CK19, 1:50, Rat, DSHB, Troma-III-s, 8G4, 48
9) DBA-biotin, 1:100-300, Vector Labs, B-1035, ZA0417
10) DBA-FITC, 1:80, Vector Labs, FL-1031-2, ZA0722
11) DESMIN, 1:100, Rabbit, Thermo Fisher Scientific, RB9014P0, 9014P1604F, 9014P1604G
12) EPCAM, 1:100-500, Rat, BD Biosciences, 552370, 7089730
13) EPCAM-BV421, 1:100, Rat, BioLegend, 118225, B204922
14) F4/80, 1:100, Rat, BioRad, MCA497GA, Clone A3-1, Batch 1014, 1610
15) GFP, 1:200-500, Chicken, Abcam, ab13970, GR236651-10, GR236651-21
16) GFP, 1:500, Rabbit, Abcam, ab6556, GR277888-1
17) HNF1, 1:50, Rabbit, Santa Cruz, SC-8986, G3112
18) KI67, 1:200, Rabbit, Thermo Fisher Scientific, RM-9106-S0, Clone SP6, Lot 9106S1607G
19) LYVE1, 1:200, Rabbit, ReliaTech, 103-PA50AG, 1208R1, 1210R03
20) MUP, 1:250, Goat, Cedarlane, GAM/MUP, 5023
21) OC2-2F8, 1:20, Rat, Craig Dorrell, ref. 55
22) OPN, 1:250, Goat, R&D Systems, AF808, BDO0613101, BDO0617041
23) pSMAD3, 1:400-500 (IF), 1:1,000 (WB), Rabbit, Abcam, ab52903, GR128879-61
24) pSMAD3, 1:400, Rabbit, Cell Signaling, 9520, Clone C25A9
25) SOX9, 1:500, Rabbit, MilliporeSigma, AB5535, 2847051, 2975230
26) SSTR2, 1:500, Rabbit, Bioss, bs-1138R, 9A25V1
27) wsCK, 1:200, Rabbit, Dako, Z0622, 10088577

We used the following secondary antibodies for staining of cells and tissue samples. They are listed as reactivity first, followed by species, fluorochrome, dilution, supplier, catalog number and lot number as applicable.

1) Chicken, Donkey, Alexa Fluor 488, 1:500, Jackson Immuno, 703-545-155, 116967, 113085
2) Goat, Donkey, Alexa Fluor 488, 1:500, Thermo Fisher Scientific, A11055, 1869589
3) Goat, Donkey, Alexa Fluor 594, 1:500, Jackson Immuno, 705-585-147, 106994
4) Goat, Donkey, Alexa Fluor 594, 1:500, Thermo Fisher Scientific, A11058, 1445994
5) Goat, Donkey, Alexa Fluor 647, 1:200, Thermo Fisher Scientific, A21447, 1841382, 1661244
6) Goat, Donkey, Biotin, 1:1,000, Jackson Immuno, 705-065-003, 77098
7) Goat, Donkey, Cy3, 1:500, Jackson Immuno, 705-165-147, 114067
8) Mouse, Horse, HRP, 1:1,000, Cell Signaling, 7076, 27
9) Rabbit, Goat, Alexa Fluor 488, 1:300, Jackson Immuno, 111-545-144, 105878
10) Rabbit, Donkey, Alexa Fluor 555, 1:200-500, Thermo Fisher Scientific, A31572, 1917920
11) Rabbit, Donkey, Alexa Fluor 647, 1:200, Jackson Immuno, 711-605-152, 118217
12) Rabbit, Donkey, Biotin, 1:1,000, Jackson Immuno, 711-065-152, 128874
13) Rabbit, Donkey, Cy3, 1:500, Jackson Immuno, 711-165-152, 117211, 105494
14) Rabbit, Goat, HRP, 1:1,000, Cell Signaling, 7074, 26
15) Rabbit, Goat, HRP, 1:500, Jackson Immuno, 111-035-144, 91036
16) Rat, Donkey, Alexa Fluor 488, 1:500, Thermo Fisher Scientific, A21208, 1789917
17) Rat, Donkey, Alexa Fluor 647, 1:200, Jackson Immuno, 712-605-153, 123845, 110001, 127806
18) Rat, Donkey, Cy2, 1:300, Jackson Immuno, 712-225-150, 102982
19) Rat, Goat, PE, 1:200, BD Biosciences, 550767, 5107876
20) Streptavidin, Alexa Fluor 594, 1:300, Thermo Fisher Scientific, S-11227, 1704463
21) Streptavidin, Cy2, 1:250, Jackson Immuno, 016-220-084, 84799
22) Streptavidin, Cy3, 1:300, Jackson Immuno, 016-160-084, 125000
23) Streptavidin, Dylight 649, 1:300, Vector Labs, SA-5649, Y0718, Z0320
24) TSA Biotin Kit Streptavidin, HRP, 1:100, Perkin Elmer, NEL700A001KT, 2075089
25) Tyramide Plus, Cy3, 1:300, Perkin Elmer, NEL744001KT, 2310803
26) Vectastain Elite ABC Reagent, Vector Labs, PK-7100, ZD0404

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Alb-Cre+/-;Rbpjf/f;Hnf6f/f mice (mixed background) were previously reported (ref. 2,26). R26R-RFP+/+ (ref. 39) (C57BL/6) and R26NZG+/+ (ref. 40) (FVB) mice were used. Flp-reporter mice were generated by crossing R26NZG+/+ mice with EIIa-Cre+/+ (ref. 41) mice (C57BL/6) to remove the Cre-reporter element and then crossing out the EIIa-Cre. These R26ZG+/+ mice were crossed with Alb-Cre+/-;Rbpjf/f;Hnf6f/f mice to generate Alb-Cre+/-;Rbpjf/f;Hnf6f/f;R26ZG+/+ mice. R26R-Confetti+/+ (ref. 42) mice (C57BL/6) were crossed with Alb-Cre +/-;Rbpjf/f;Hnf6f/f mice to generate Alb-Cre+/-;Rbpjf/f;Hnf6f/f;R26R-Confetti+/- mice. Tgfbr2f/f (ref. 43) mice (C57BL/6) were crossed with Alb-Cre+/-;Rbpjf/f;Hnf6f/f mice to generate Alb-Cre+/-;Rbpjf/f;Hnf6f/f;Tgfbr2f/f mice. Because Tgfbr2 (68.39 cM) and Hnf6 (41.93 cM) are both on chromosome 9, recombinants were generated at 0.1356 (35/258 mice) observed frequency (0.26 expected frequency). Different founder recombinants were intercrossed to generate Alb-Cre+/-;Rbpjf/f;Hnf6f/f;Tgfbr2f/f mice. R26R-ZsGreen+/+ (ref. 44) mice (C57BL/6) were crossed with Alb-Cre+/- mice to generate Alb-Cre+/-;R26R-ZsGreen+/+ mice. Alb-Cre+/-;Rbpjf/f;Hnf6f/f mice were crossed with Rag2-/- (ref. 45);Il2rg-/- (ref. 46) mice (mixed background) to generate Alb-Cre+/-;Rbpjf/f;Hnf6f/f;Rag2-/-;Il2rg-/- mice. Male and female mice of the indicated age and genotype were chosen randomly for inclusion in experiments. The mouse used as a positive control for biliary gene expression was a C57BL/6 wildtype mouse fed choline-deficient diet (MP Biomedicals) and given 0.15% (w/v) ethionine (Sigma-Aldrich) in the drinking water (CDE diet) for 3 weeks. All mice were kept under barrier conditions. All procedures were approved by the Institutional Animal Care and Use Committee at UCSF or CCHMC.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The following human liver tissue was used: 1) Regenerative nodule and nonregenerated tissue from a 3-year-old male with ALGS (explanted liver). 2) Regenerative nodule and nonregenerated tissue from a 15-year-old male with ALGS (liver biopsies). 3) Histologically normal tissue from a 35-year-old male with normal JAG1 and NOTCH2 sequences (liver resection for metastasis of rectal adenocarcinoma). Samples were obtained with patient consent and approval from the Commission Cantonale d'Ethique de la Recherché CCER (1) or the UCSF Institutional Review Board (2 and 3). Full clinical and pathological characterization of patients is provided in Extended Data Table 1.

nature research

Corresponding author(s): Holger Willenbring

☐ Initial submission ☒ Revised version ☐ Final submission

# Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

## Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

## Methodological details

5. Describe the sample preparation.

Detailed information is provided in the Methods section under "Cholangiocyte isolation" and "Hepatocyte isolation."

For cholangiocytes, nonparenchymal liver cells were isolated from >P115 Alb-Cre+/-;Rbpjf/f;Hnf6f/f and Rbpjf/f;Hnf6f/f mice as previously described (ref. 51). Cells were resuspended at 1.0E+7 cells/mL in DMEM/2% FBS and blocked with Mouse Fc Block (BD Biosciences) for 30 minutes. Cells were incubated with fluorochrome-conjugated antibodies (Supplementary Table 2) and DBA-FITC (Vector Laboratories) for 30 minutes, washed with cold DPBS 3 times and resuspended in DMEM/2% FBS. Sytox Red (Thermo Fisher Scientific) was added to label dead cells prior to sorting. Unstained and single-stained cells were used for compensation. Specificity of DBA binding was verified with a GalNAc (Sigma)-blocked control as previously described (ref. 52). CD11b–CD31–CD45–EPCAM+DBA– cells were collected as peripheral cholangiocytes and CD11b–CD31–CD45–EPCAM+DBA+ were collected as hilar cholangiocytes. Cells were either sorted into DMEM/2% FBS, pelleted and snap frozen, or sorted directly into extraction buffer for RNA purification.

Hepatocytes were isolated from >P115 Alb-Cre+/-;Rbpjf/f;Hnf6f/f mice by 2-step collagenase perfusion followed by purification through a Percoll gradient. Cells were resuspended at 1.0E+6 cells/100 µL in Hanks Buffer with 10% FBS and incubated with OC2-2F8 antibody for 1 hour on ice. Cells were washed with cold DPBS 2 times and resuspended in Hanks/10% FBS. Fluorochrome-conjugated secondary antibody (Supplementary Table 3) was added and cells were incubated for 30 minutes on ice followed by 2 washes with cold DPBS. Cells were blocked with 5% normal rat serum (Jackson Immuno) in Hanks Buffer for 10 minutes on ice. Cells were then incubated with fluorochrome-conjugated antibodies (Supplementary Table 2) for 30 minutes on ice. Cells were then washed in cold DPBS 3 times and resuspended in Williams E medium/2% FBS. Sytox Red (Thermo Fisher Scientific) was added to label dead cells prior to sorting. Unstained and single-stained cells were used for compensation. CD11b–CD31–CD45–EPCAM–OC2-2F8+ cells were collected as hepatocytes. Cells were either sorted into DMEM/2% FBS, pelleted and snap frozen, or sorted directly into extraction buffer for RNA purification.

| | |
|---|---|
| 6. Identify the instrument used for data collection. | FACSAria III (BD Biosciences) |
| 7. Describe the software used to collect and analyze the flow cytometry data. | FACSDiva (BD Biosciences) was used to collect data and FlowJo (FlowJo, LLC) was used to analyze data and generate charts/figures. |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | CD11b–CD31–CD45–EPCAM+DBA– cells represent 0.1% of the parent population and CD11b–CD31–CD45–EPCAM+DBA+ cells represent 0.1% of the parent population. CD11b–CD31–CD45–EPCAM– OC2-2F8+ cells represent 34.3% of the parent population. |
| 9. Describe the gating strategy used. | For cholangiocyte isolation, based on the pattern of FSC-A/SSC-A, nonparenchymal liver cells were selected for analysis. Singlets were gated according to the pattern of FSC-W vs. FSC-H followed by SSC-W vs. SSC-H. Positive vs. negative populations were determined by antibody staining. CD11b–CD31–CD45–EPCAM+DBA– cells were collected as peripheral cholangiocytes and CD11b–CD31–CD45–EPCAM+DBA+ as hilar cholangiocytes. For hepatocyte isolation, cells for analysis were first gated based on the pattern of FSC-A/SSC-A. Singlets were gated according to the pattern of SSC-W vs. SSC-H followed by FSC-W vs. FSC-H. CD11b–CD31–CD45–EPCAM–OC2-2F8+ cells were collected as hepatocytes. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒