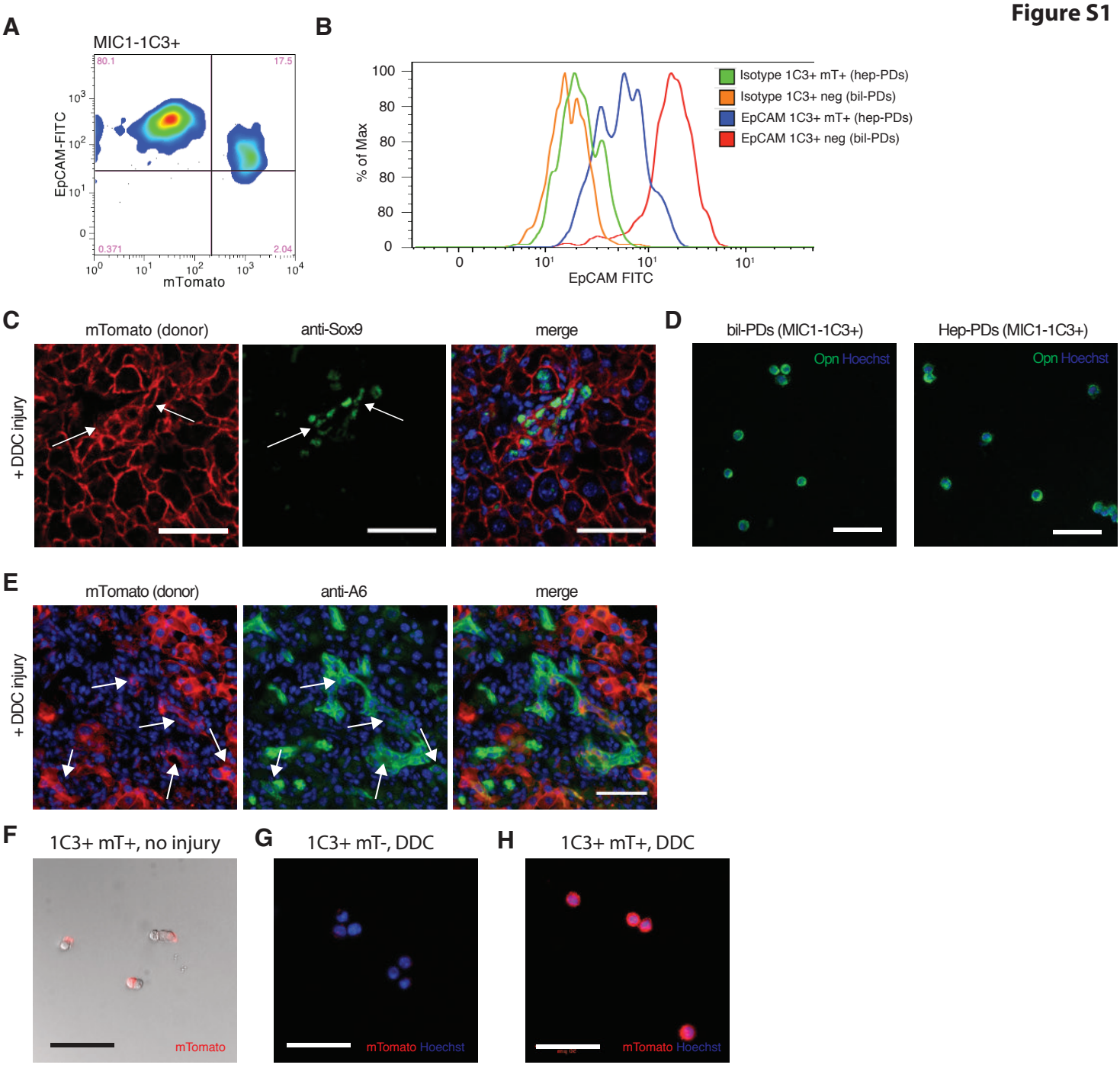

Figure S1, related to Fig 1: Hepatocyte-derived proliferative ducts express multiple ductal markers

A) EpCAM staining intensity was consistently half-log lower in hepPDs (1C3+ mTomato+) compared with bilPDs (1C3+ mTomato-) by FACS. B) Histogram of EpCAM staining intensity for two ductal populations and isotype controls, respectively (representative plot from 3 biological replicates). C) hepPDs (mTomato+) are Sox9+ and arrange in ducts after 6 weeks DDC treatment D) MIC1-1C3+ FACS isolated ducts derived from biliary epithelium or hepatocytes show equivalent Opn staining intensity. F) Rare MIC1-1C3+ mTomato-high+ trigger-pulse-width low FACS events from an uninjured chimera appeared mTomato+ due to attached membrane fragments and doublet cells and not engraftment of donor progenitors. G) mTomato-negative MIC1-1C3+ or H) mTomato—MIC1-1C3+ from a DDC treated chimeric are shown by comparison (fixed, counterstained).

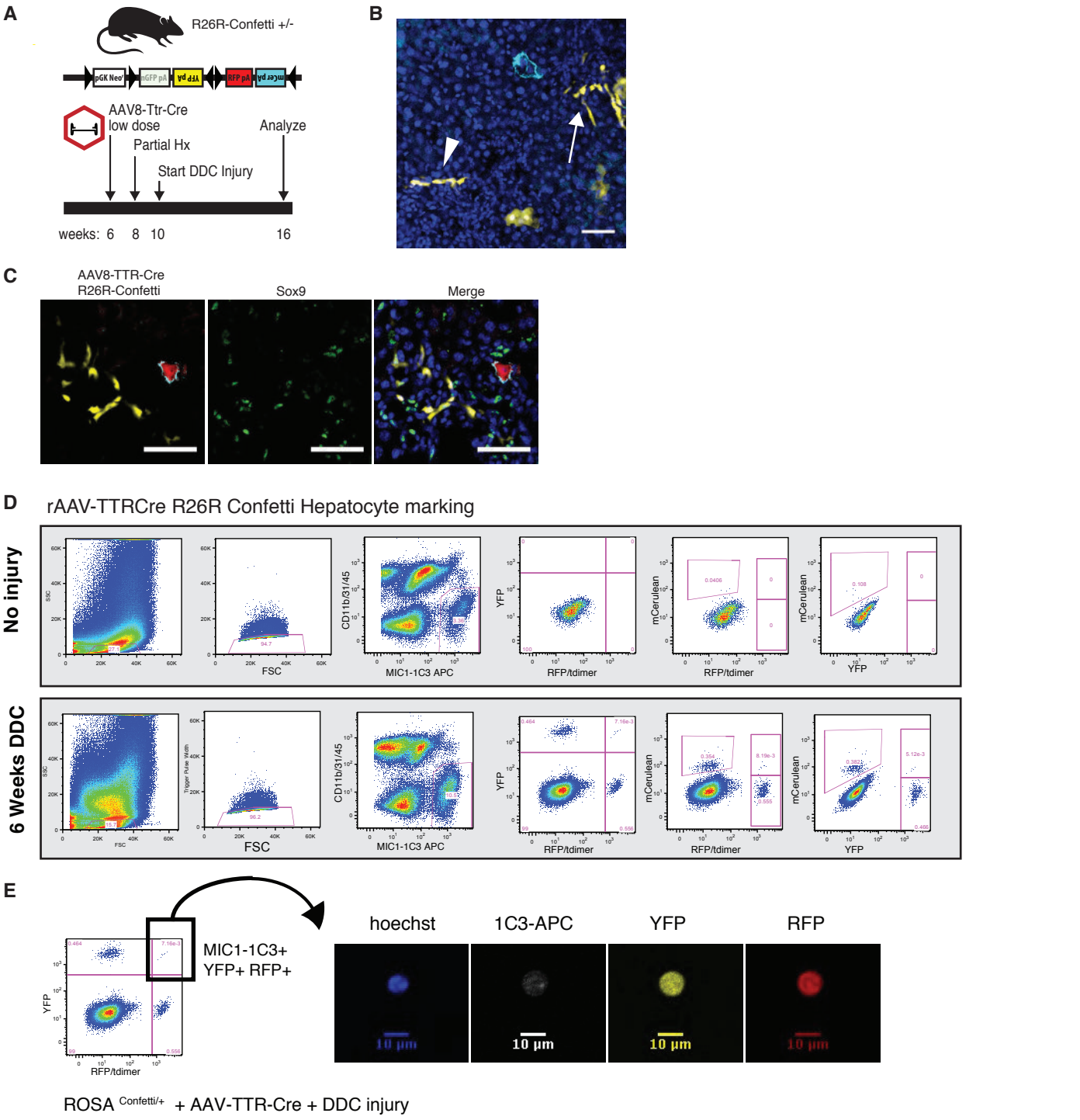

Figure S2, related to Fig. 2: Hepatocytes labeled with rAAV8 clonally expand and undergo ductal metaplasia
A) Animals were given low dose rAAV8-Ttr-Cre (2e10 vg), administered partial hepatectomy to deplete rAAV particles, and 6 weeks 0.1% DDC to initate an oval cell response. B) Confocal analysis of thick sections indicated single hepatocytes proliferated into clusters of 20-40 cells. Clusters of 24 cells (arrowhead) and >31 cells (arrow) are shown. C) Hepatocyte-derived ductal cells co-expressed Sox9 and Opn(not shown), bars = 50 μm. D) FACS based analysis indicated that rAAV-Ttr-Cre did not mark MIC1-1C3 cells without injury (top) but 2% of hepatocytes were marked in this animal (not shown). Following 6 weeks DDC injury (bottom) Cre-marked MIC1-1C3 cells were detected. E) Approximately 1% of hepatocyte-derived progenitors in R26R-Confetti heterozygous mice expressed 2 colors of the Confetti reporter. The R26R-Confetti was present at 1 copy per diploid genome, suggesting that double-positive cells hepatocyte-derived progenitor cells were polyploidy. Microscopy confirmed cells were not doublets.

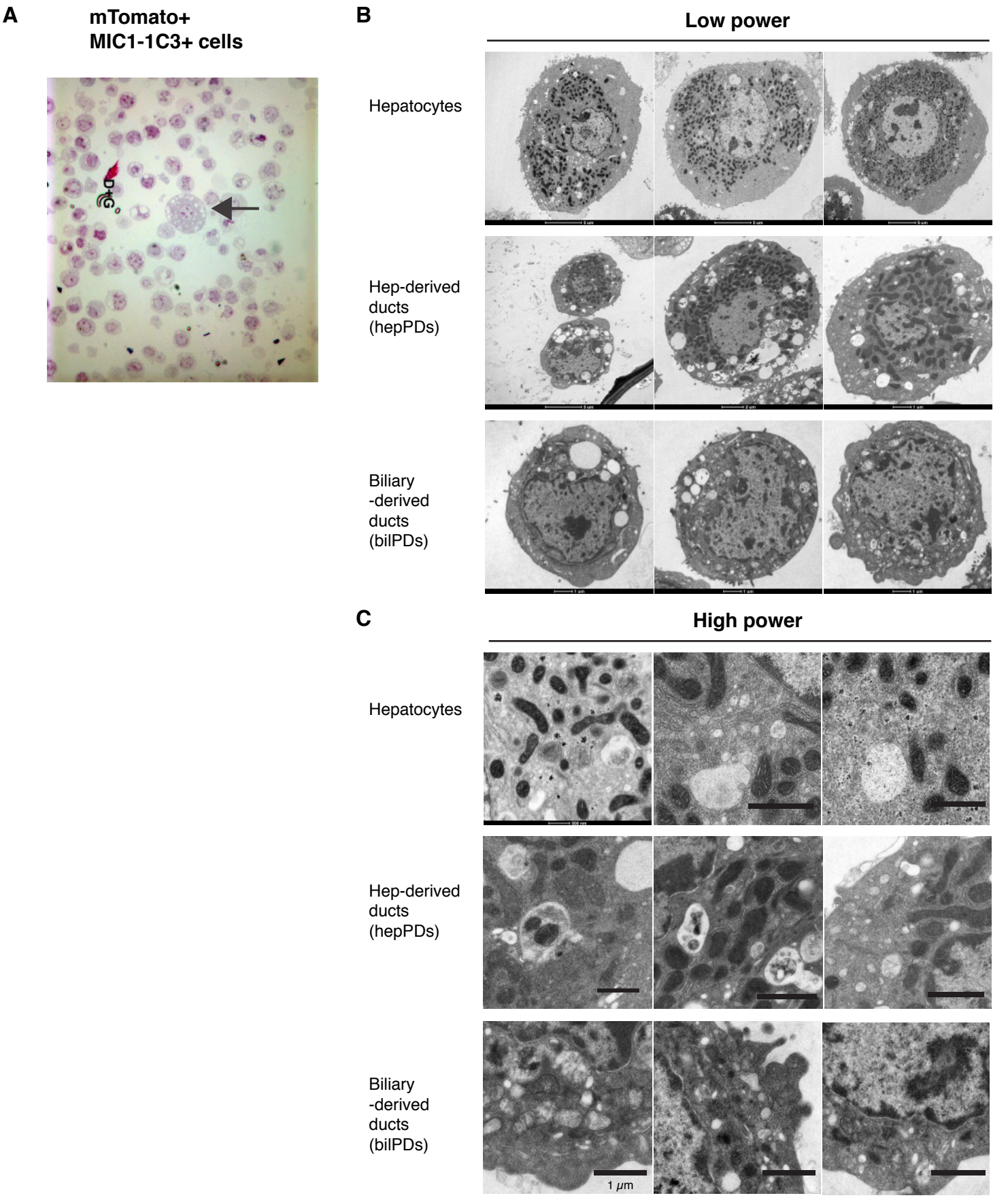

Figure S3, related to Fig. 2: Additional microscopy images of sorted liver cells
A) Plastic section of hep-PDs with a single contaminating hepatocytes present (arrow) shown for size comparison.
B) Low power TEM images of FACS sorted hepatocytes, hep-PDs, and bil-PDs and C) high power (bar length 1μm, or as indicated).

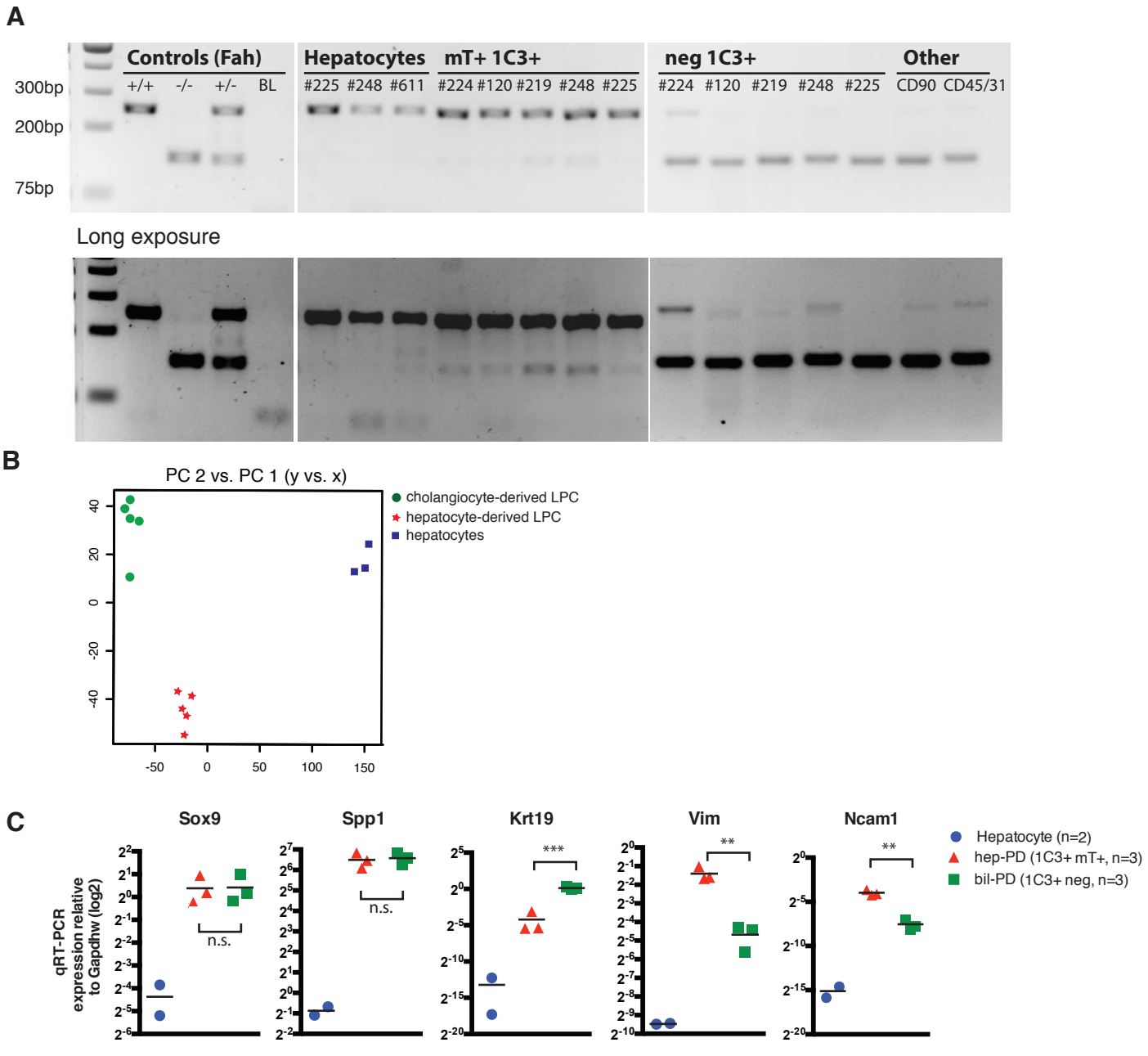

Figure S4, related to Fig. 3: MIC1-1C3+ cells separated by mTomato phenotype are genetically and phenotypically distinct

A) 3-primer DNA genotyping assay for Fah on FACS isolated MIC1-1C3+ cells confirms that mTomato was a reliable marker of cell origin. All samples were run on the same multi-row gel. Long exposure shows trace-level cross contamination in FACS samples but not PCR controls. B) Principle-component analysis of RNA-seq data separates three FACS sorted populations by gene expression. C) qRT-PCR expression levels of duct-associate transcripts normalized to Gapdh in FACS sorted 1C3+ cells or hepatocytes. mRNA levels were not significantly different in the two ductal populations for Sox9 ($p=0.94$) and Spp1 ($p=0.76$). Krt19, Vim, and Ncam1 were differentially expressed in the two types of proliferative ducts (unpaired t-test, $p < 0.01$ **, $p<0.001$***).

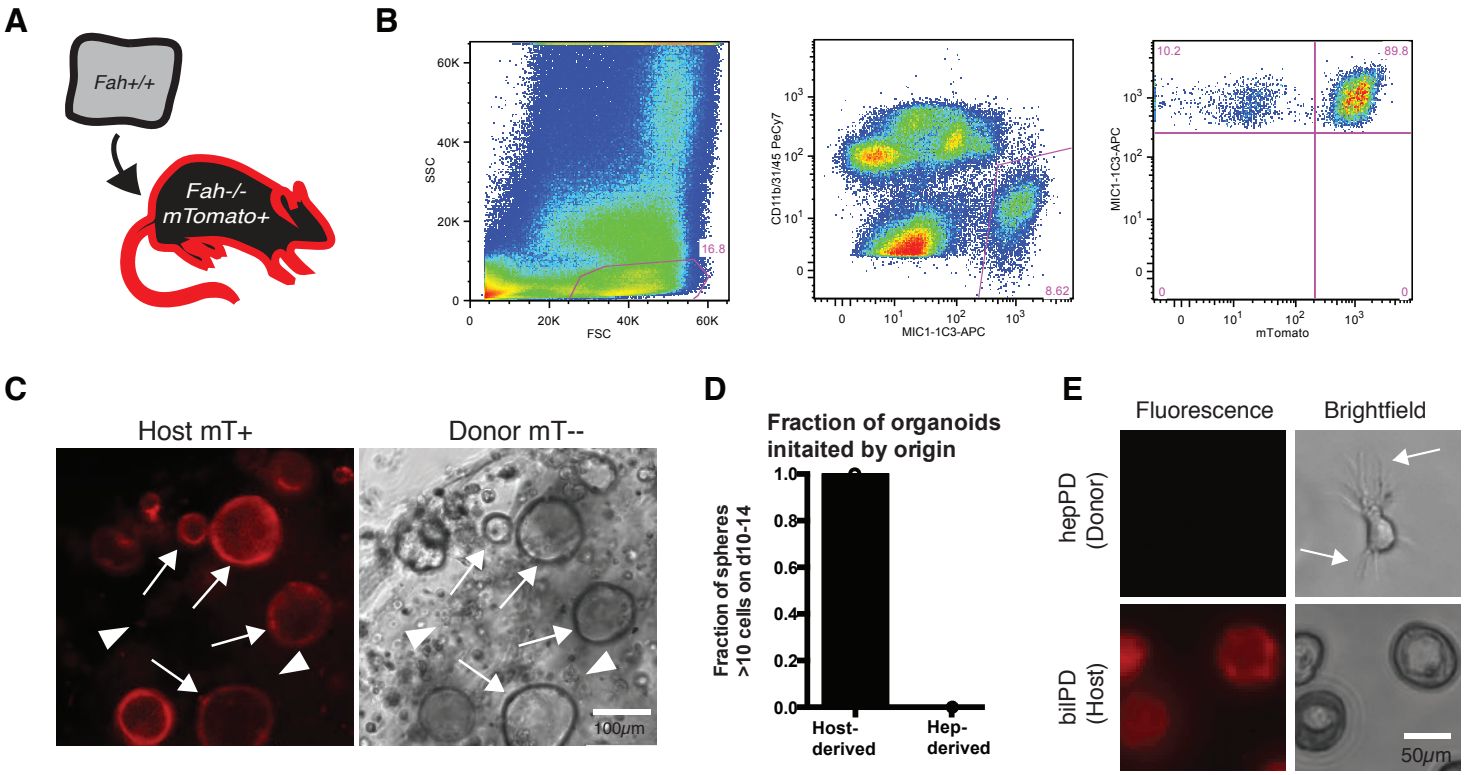

Figure S5, related to Fig. 4. Color swap indicates *in vitro* differences are not an artifact of the mTomato fluorescent protein

A) Wild-type hepatocytes were transplanted into ROSA-mTomato Fah-/- mice, allowed to repopulate for 10 weeks, and given DDC for 6 weeks. B) CD31- CD45- CD11b- MIC1-1C3+ cells were analyzed for mTomato status following DDC injury. MIC1-1C3+ mTomato- are hepPDs and 1C3+ mTomato+ are bilPDs. C) Unsorted non-parenchymal cells were seeded into matrigel for organoid formation assay. mTomato+ cells formed organoids (arrowhead) but mTomato-negative cells (arrow) did not. D) All organoids were mTomato+, indicating a bile duct/host source. E) FACS sorted hepPDs (mTomato--) did not form organoids but extended filipoida, while host-derived bilPDs (mTomato+) initiated self-renewing organoid.

Figure S6

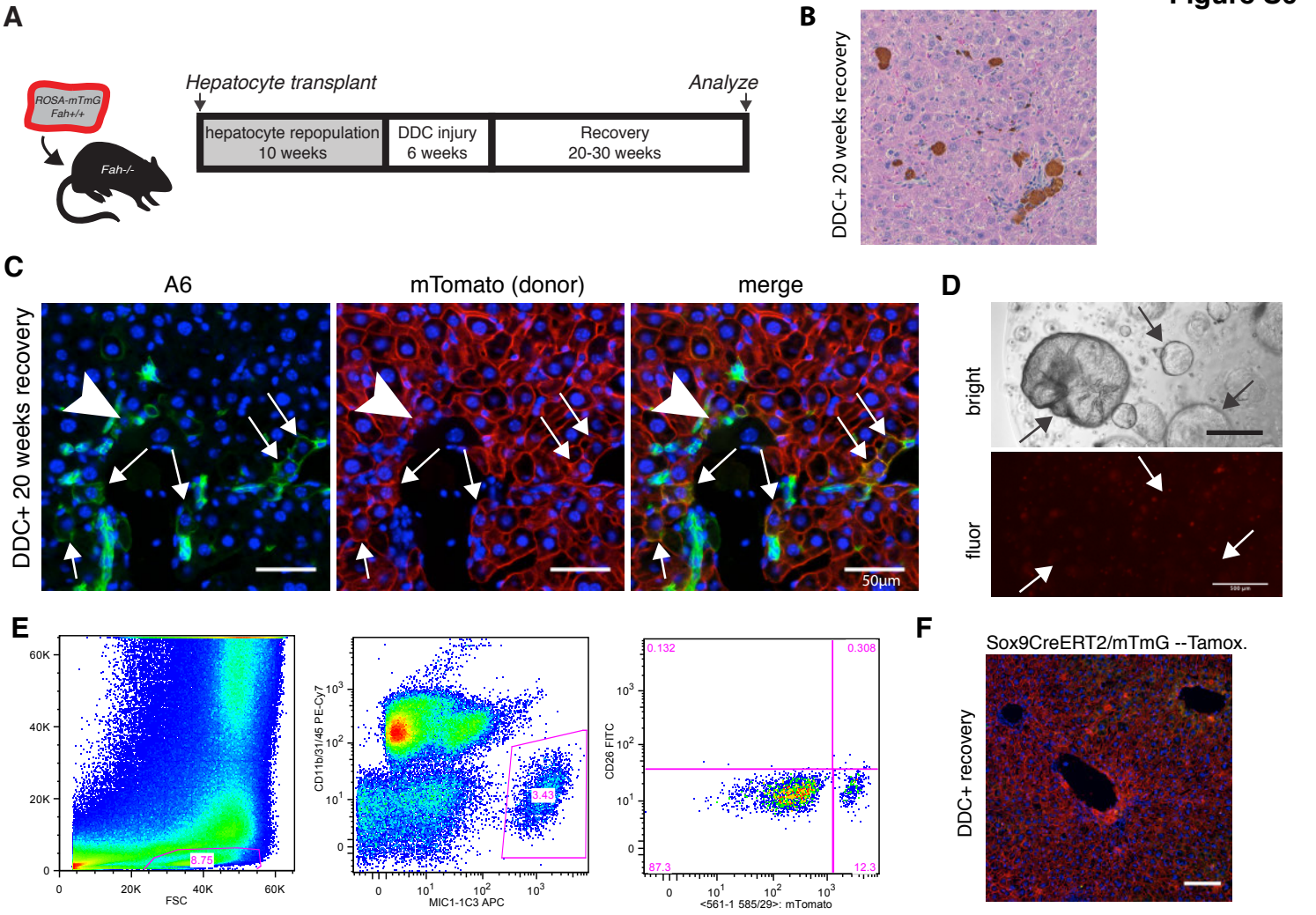

Figure S6, related to Fig. 5: Fate of hepPDs in long-term recovery from DDC injury and additional characterization of reverted hepatocytes

A) mTomato+ hepatocyte chimeric mice were allowed to recovery from DDC injury for 20-30 weeks (n=3). B) After 20 weeks on normal diet, livers showed residual cholestasis, inflammation, and porphyrin accumulation by H&E (bar = 50μm). C) A6-low hepatocytes (arrow) adjacent to biliary structures colocalized with ROSA-mTomato+ indicating they were hepatocyte-derived. Residual A6-high mTomato+ hepPDs were still present (arrowhead). D) Greater than 10% of MIC1-1C3+ cells were hepatocyte-derived by FACS-based analysis (n=3). E) All organoids were still host-derived (mTomato-negative) indicating that hepatocytes did not transdifferentiate into organoid forming progenitors after 20 weeks of recovery (bar = 500μm, representative image, n=2 animals). F) Sox9-CreERT2 mTmG hepatocyte chimeric mice injured with DDC did not show direct Sox9-CreERT2 marking in the absence of tamoxifen. G) Sox9-CreERT2 marked hepatocytes (hepPDderived; arrows) were positive for hepatocyte markers Hnf4a and H) Mup following a 4 week recovery period. I) Serial-transplanted Sox9-CreERT2 marked hepatocytes (arrows) were MUP+.

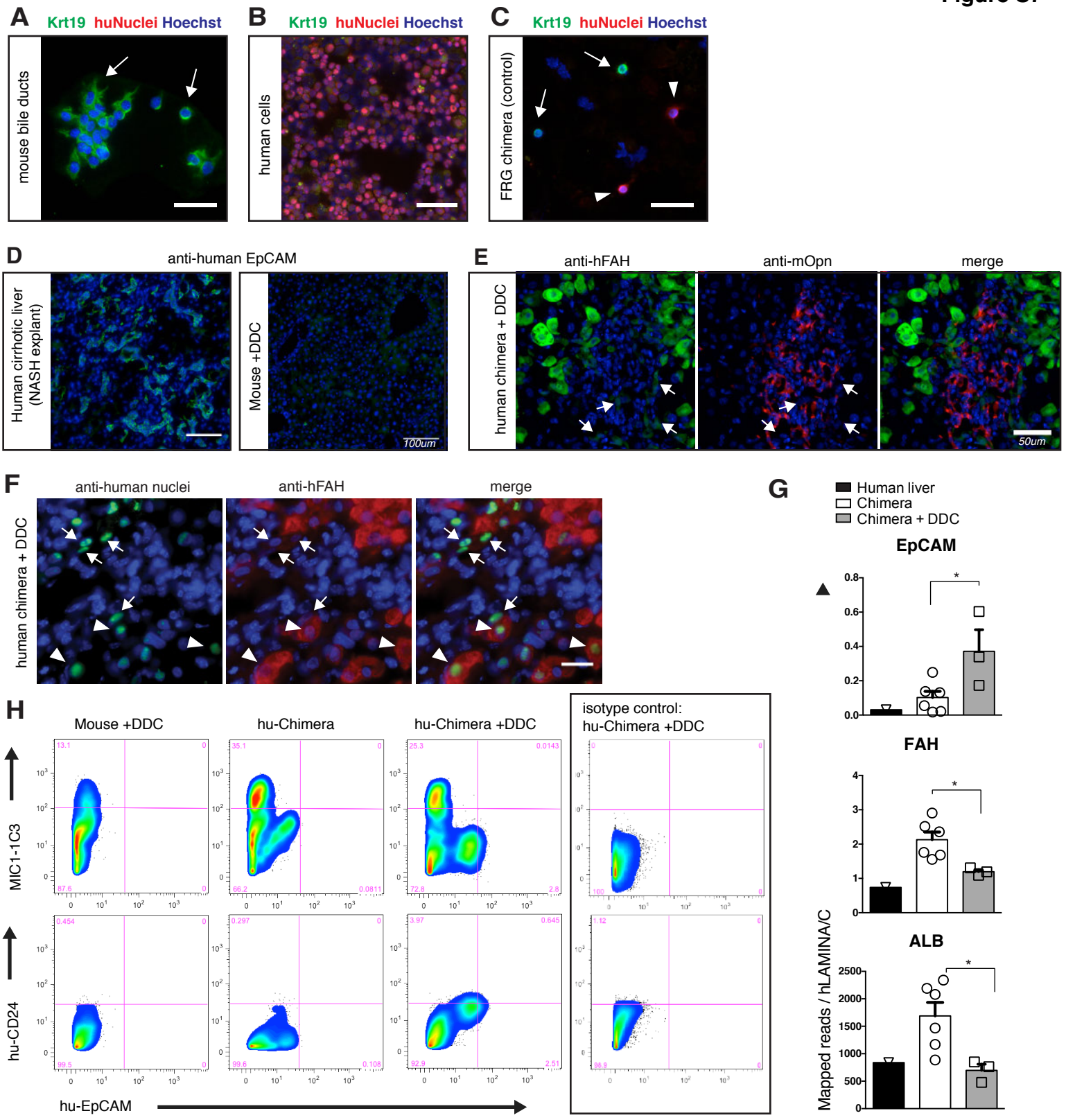

Figure S7, related to Fig. 7: Identification of human hepatocyte-derived cells in chimeric livers with additional markers
A) Krt19 (green) and human nuclear antigen staining in cultured mouse bile duct cells (cytospin) B) human blood and endothelium and C) dissociated cells from a humanized FRG mouse liver. Krt19+ cell (arrows) and human nuclear antigen+ cells (arrowhead) did not co-localize in a repopulated chimeric mice, suggesting that human Krt19+ cells were not present before injury (n=2 chimeric mice). D) Anti-human EpCAM identified ductal reactions in cirrhotic human liver (non-alcoholic steatohepatitis explant) but not DDC-injured mouse liver tissue. E) Human hepatocyte-derived ducts assume a ductal morphology (arrow), express low levels of FAH, and are in close contact with mouse-Opn+ cells. F) Human hepPDs (arrows, human nuclear antigen+), have ovoid nuclei and scant cytoplasm, and stain faintly for hepatocyte gene FAH. Hepatocytes (arrowhead) stain strongly for FAH, have round nuclei and abundant cytoplasm. G) Additional RNA-seq data from chimeric livers show DDC injury is associated with significantly greater human EpCAM and significantly less FAH and ALB mRNA compared with uninjured chimeras chimeric livers (t-test, $p < 0.05$ *). H) FACS analysis for mouse MIC1-1C3, human CD24, and human EpCAM in DDC treated mice, uninjured humanized FRG mice and DDC injured FRG mice (isotype control provided for comparision, far right).

Gene symbol	Hepatocytes(n=3) (RPKM)	s.d.	hepPDs (n=5) (RPKM)	s.d.	bilPDs (n=5) (RPKM)	s.d.	Fold diff: hepPD / Hep	hepPD / bilPD
Hepatocyte								
Alb	37559.47	9091.36	4653.57	1406.89	2872.78	489.03	**0.12**	**1.62**
Hgd	314.05	26.73	11.76	4.89	10.71	5.80	**0.04**	**1.10**
Cyp7a1	40.12	19.99	0.84	0.65	1.34	1.20	**0.02**	**0.63**
Ttr	3108.40	351.13	705.55	149.94	463.34	38.86	**0.23**	**1.52**
F9	135.63	11.16	43.14	4.85	6.54	3.31	**0.32**	**2.52**
Fah	292.62	57.61	58.87	7.36	14.014*	4.37	**0.20**	**N/A**
Tdo2	510.59	132.18	47.25	25.76	29.44	17.16	**0.09**	**1.61**
Hnf4a	243.50	5.19	130.29	30.12	58.34	10.81	**0.54**	**2.23**
Progenitor/cholangiocyte								
Sox9	8.84	4.48	92.45	15.41	102.56	29.11	**10.46**	**0.90**
Spp1 (Opn)	796.50	198.50	17559.80	1095.90	18618.33	587.62	**22.05**	**0.94**
Hnf1b	12.77	1.17	158.86	12.92	176.43	21.44	**12.44**	**0.90**
Krt19	0.36	0.29	42.54	23.45	657.89	65.25	**119.27**	**0.06**
Lgr5	1.34	0.52	0.24	0.10	0.07	0.05	**0.18**	**3.70**
Cd44	11.24	2.97	178.73	37.13	150.82	35.34	**15.90**	**1.19**
Cd24	2.02	0.84	481.16	61.65	676.05	85.61	**238.59**	**0.71**
Nes	0.04	0.03	8.54	2.06	2.94	0.47	**213.60**	**2.91**
Kit	0.14	0.06	7.22	2.06	2.94	0.51	**52.83**	**2.46**
Epcam	0.83	0.14	85.30	29.48	545.47	51.95	**103.19**	**0.16**
Itga3 (MIC1-1C3)	6.79	0.71	157.60	15.29	154.36	6.57	**23.20**	**1.02**
Cftr	0.02	0.03	11.50	3.61	49.78	5.23	**492.86**	**0.23**
Tacstd2 (Trop2)	0.02	0.02	1.16	0.57	38.75	11.30	**57.90**	**0.03**
Tnfrsf12a (Fn14)	116.01	13.28	314.77	63.33	278.23	45.34	**2.71**	**1.13**
Prom1 (Cd133)	10.96	15.11	26.59	9.19	139.14	17.75	**2.43**	**0.19**
Jag1	1.51	0.65	75.73	9.86	77.66	23.22	**50.15**	**0.98**
Notch2	14.43	1.46	56.36	8.42	51.75	4.03	**3.90**	**1.09**
EMT-related								
Zeb1	5.04	0.26	12.55	0.85	3.44	0.90	**2.49**	**3.65**
Vim	9.12	4.74	211.34	33.53	45.79	25.42	**23.16**	**4.62**
Cdh2	29.47	2.37	60.19	4.68	13.19	5.24	**2.04**	**4.56**
Ncam1	0.01	0.01	7.03	2.82	1.36	0.70	**1054.80**	**5.16**
Smo	1.38	0.46	36.47	2.16	35.23	2.65	**26.42**	**1.04**
Snai1	0.09	0.10	0.09	0.04	0.88	0.52	**1.02**	**0.10**
Snai2	3.86	1.37	1.52	0.33	0.50	0.23	**0.39**	**3.02**
Twist1	0.01	0.01	0.19	0.07	0.17	0.12	**27.90**	**1.12**
Cdh1	100.13	9.08	377.64	36.59	553.60	39.05	**3.77**	**0.68**
Other cell types								
Ptprc (Cd45)	0.22	0.18	1.46	0.98	2.89	1.97	**6.62**	**0.50**
Pecam1 (Cd31)	0.44	0.32	1.64	1.06	3.96	2.56	**3.76**	**0.42**
Gfap	0.00	0.01	0.08	0.04	0.30	0.07	**22.80**	**0.25**
Thy1	0.02	0.00	0.40	0.49	0.77	0.67	**20.10**	**0.52**
Col1a1	0.20	0.17	3.68	1.47	35.64	39.97	**18.41**	**0.10**
Desmin	0.05	0.02	0.30	0.13	1.55	1.11	**5.92**	**0.19**
Yap1	30.93	0.06	52.73	2.76	42.73	3.30	**1.70**	**1.23**
Ctgf	0.85	0.15	205.68	52.08	494.72	83.77	**241.98**	**0.42**
Cyr61	71.07	11.90	1037.55	326.93	1056.27	401.06	**14.60**	**0.98**
AmotL2	6.40	2.39	57.40	8.31	92.56	17.35	**8.96**	**0.62**
Rbfox3 (NeuN)	0.00	0.01	0.02	0.01	0.09	0.08	**4.80**	**2.52**
Afp	2.96	0.67	3.48	0.68	2.62	1.85	**1.18**	**2.52**

*Denotes gene deletion at exon5

Table S1: Gene expression of selected genes between cell types in DDC injury

RNAseq expression data (RPKM) for FACS sorted hepatocytes, hepPDs, and bilPDs for selected genes of interest. Mean and standard deviations values are shown below (relates to Figure 3).

Table S2: Genesets enriched in hepPDs versus Hepatocytes

Geneset DETAILS	SIZE	NES	NOM p-val	FDR q-val
KEGG_NOTCH_SIGNALING_PATHWAY	42	1.6	0.001	0.073
KEGG_BASAL_CELL_CARCINOMA	54	1.61	0.001	0.114
KEGG_MELANOGENESIS	93	1.48	0.001	0.144
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	83	1.42	0.004	0.151
KEGG_ARRHYTHMOGENIC_RV_CARDIOMYOPATHY	74	1.49	0	0.151
KEGG_FOCAL_ADHESION	184	1.38	0.001	0.161
KEGG_PATHWAYS_IN_CANCER	304	1.39	0	0.162
KEGG_EPITHELIAL_CELL_SIGNALING_IN_H_PYLORI	63	1.42	0.005	0.162
KEGG_ECM_RECEPTOR_INTERACTION	80	1.4	0.009	0.169
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	64	1.39	0.013	0.171
KEGG_DILATED_CARDIOMYOPATHY	89	1.42	0.008	0.172
KEGG_AXON_GUIDANCE	125	1.5	0	0.173
KEGG_SPLICEOSOME	94	1.39	0.002	0.174
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	116	1.42	0	0.178
KEGG_HEDGEHOG_SIGNALING_PATHWAY	54	1.51	0.004	0.187
KEGG_WNT_SIGNALING_PATHWAY	140	1.43	0.001	0.188
KEGG_COLORECTAL_CANCER	60	1.36	0.027	0.191
KEGG_TIGHT_JUNCTION	119	1.45	0.002	0.191
KEGG_MAPK_SIGNALING_PATHWAY	249	1.43	0	0.193
KEGG_LEISHMANIA_INFECTION	47	1.35	0.037	0.203
KEGG_RENAL_CELL_CARCINOMA	67	1.34	0.026	0.207
KEGG_OLFACTORY_TRANSDUCTION	28	1.34	0.069	0.211
KEGG_ENDOCYTOSIS	150	1.33	0.005	0.219
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	104	1.32	0.01	0.22
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	192	1.33	0.001	0.223
KEGG_ERBB_SIGNALING_PATHWAY	83	1.3	0.034	0.224
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	84	1.3	0.038	0.226
KEGG_CELL_CYCLE	113	1.31	0.007	0.227
KEGG_CHRONIC_MYELOID_LEUKEMIA	70	1.3	0.043	0.233
KEGG_ENDOMETRIAL_CANCER	49	1.29	0.077	0.233
KEGG_TGF_BETA_SIGNALING_PATHWAY	84	1.29	0.037	0.234
KEGG_VEGF_SIGNALING_PATHWAY	69	1.31	0.043	0.235
KEGG_ADHERENS_JUNCTION	72	1.27	0.034	0.236
KEGG_SMALL_CELL_LUNG_CANCER	80	1.29	0.048	0.236
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	15	1.31	0.111	0.236
KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	42	1.28	0.09	0.238
KEGG_DNA_REPLICATION	33	1.28	0.107	0.242
KEGG_GNRH_SIGNALING_PATHWAY	94	1.28	0.042	0.245

Table S2: Genesets enriched in hepPDs versus Hepatocytes. Differentially expressed genesets identified with GSEA geneset analysis (KEGG) (relates to Figure 3).

Table S3: Genesets enriched in hepPDs versus bilPDs

Geneset DETAILS	SIZE	NES	NOM p-val	FDR q-val
KEGG_TYROSINE_METABOLISM	36	2.35	0	0
KEGG_FATTY_ACID_METABOLISM	33	2.25	0	0.001
KEGG_BETA_ALANINE_METABOLISM	22	2.18	0	0.001
KEGG_TRYPTOPHAN_METABOLISM	37	2.15	0	0
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	48	2.13	0	0
KEGG_BUTANOATE_METABOLISM	30	2.11	0	0
KEGG_DRUG_METABOLISM_CYTOCHROME_P450	42	2.05	0	0.001
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	41	2.03	0	0.001
KEGG_PHENYLALANINE_METABOLISM	18	2	0	0.001
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	30	1.99	0	0.001
KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	18	1.93	0	0.002
KEGG_ASCORBATE_AND_ALDARATE_METABOLISM	13	1.92	0.003	0.002
KEGG_HISTIDINE_METABOLISM	27	1.92	0.004	0.002
KEGG_PEROXISOME	71	1.91	0	0.002
KEGG_PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	14	1.87	0	0.003
KEGG_DRUG_METABOLISM_OTHER_ENZYMES	27	1.85	0	0.005
KEGG_LIMONENE_AND_PINENE_DEGRADATION	10	1.81	0.003	0.007
KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	1.73	0.004	0.013
KEGG_FOLATE_BIOSYNTHESIS	10	1.71	0.011	0.014
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	32	1.7	0	0.016
KEGG_PROPANOATE_METABOLISM	30	1.67	0.004	0.021
KEGG_CITRATE_CYCLE_TCA_CYCLE	28	1.63	0	0.028
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	15	1.6	0.023	0.036
KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	31	1.58	0.008	0.04
KEGG_ARGININE_AND_PROLINE_METABOLISM	50	1.55	0.005	0.048
KEGG_RETINOL_METABOLISM	31	1.54	0.021	0.049
KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG	19	1.54	0.038	0.048
KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	29	1.54	0.025	0.048
KEGG_NICOTINATE_AND_NICOTINAMIDE_METABOLISM	21	1.53	0.031	0.049
KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS	15	1.52	0.038	0.051
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	39	1.48	0.028	0.067
KEGG_GLUTATHIONE_METABOLISM	45	1.47	0.052	0.069
KEGG_LYSINE_DEGRADATION	43	1.44	0.038	0.08
KEGG_ABC_TRANSPORTERS	40	1.43	0.049	0.083
KEGG_RIBOFLAVIN_METABOLISM	16	1.39	0.103	0.104
KEGG_PROTEASOME	43	1.38	0.045	0.105
KEGG_SELENOAMINO_ACID_METABOLISM	21	1.35	0.108	0.121
KEGG_PPAR_SIGNALING_PATHWAY	61	1.33	0.048	0.135
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	54	1.26	0.087	0.194
KEGG_LYSOSOME	114	1.23	0.054	0.219
KEGG_PENTOSE_PHOSPHATE_PATHWAY	24	1.21	0.154	0.237
KEGG_STEROID_HORMONE_BIOSYNTHESIS	31	1.21	0.151	0.237
KEGG_REGULATION_OF_AUTOPHAGY	28	1.19	0.162	0.246

Table S3: Genesets enriched in hepPDs versus bilPDs. Differentially expressed genesets identified with GSEA geneset analysis (KEGG) (relates to Figure 3).

Gene	30141_ Chimera	30154_ Chimera	1794_Chimera_DDC	30012_ Chimera	225_Mouse_Heps	225_Mouse_MIC11c3	Normal Human Liver	F19b1_ Chimera	F19b2_ Chimera	F19b3_ Chimera	Human Cholang-iocytes	1431_Chimera_DDC	3567_Chimera_DDC	Refseq ID	% identity (species-species)	% gaps
ACTB	847	900	7471	9280	27	151	6803	2667	1502	3817	74351	981	1692	NM_001101.3	90	1
ALB	484250	655243	691371	1325310	0	0	640934	760431	259616	433514	283474	162014	349546	NM_000477.5	77	1
CD24A	46	56	2280	200	0	0	244	47	11	16	17697	435	231	NM_001792.3	78	8
CD44	0	1	430	35	0	0	91	0	0	3	192	51	16	NM_013230.2	79	3
CDH2	145	257	887	1066	0	1	350	475	82	213	2970	151	216	NM_000610.3	88	1
FAH	642	792	1721	1994	0	0	564	767	275	822	242	231	537	NM_001454.3	85	0
FOXJ1	0	0	3	0	0	0	0	0	0	0	141	0	0	NM_007197.3	87	1
FZD10	0	0	0	0	0	0	0	0	0	0	0	0	0	NM_000137.2	86	1
GAPDH	1794	2397	9850	7841	1	1	3581	5210	3109	7165	9557	1451	2255	NM_000181.3	89	0
GUSB	176	248	1620	828	0	0	534	272	97	314	1158	115	411	NM_001289745.1	79	1
HNF1B	15	26	247	77	4	67	88	31	10	45	3355	38	44	NM_000458.2	86	2
JAG1	5	12	162	12	1	17	76	28	17	45	5597	31	26	NM_000214.2	89	1
KRT19	0	1	0	0	0	0	129	0	1	0	5702	4	0	NM_002276.4	84	1
KRT7	7	16	2826	88	0	0	99	1	4	3	16575	314	253	NM_005556.3	85	1
LMNA	221	315	1444	1133	0	0	764	325	177	489	18240	212	411	NM_170707.3	84	1
SMO	233	273	282	913	0	0	164	158	52	151	323	40	94	NM_000582.2	88	1
SOX9	36	39	538	146	0	0	71	49	23	59	8267	65	110	NM_005631.4	89	2
SPP1	26	57	4413	66	0	0	116	0	0	2	6501	322	616	NM_000346.3	80	1
TTR	9348	13098	12529	32752	0	0	21720	7799	2576	6200	898	1509	5176	NM_000371.3	81	0
VIM	1	4	262	10	0	0	752	0	4	0	11407	36	6	NM_003380.3	90	1
ZEB1	70	51	129	182	19	28	90	108	46	37	18	52	35	NM_001128128.2	86	1
EpCAM	4	7	489	62	0	0	24	45	44	64	4317	128	71	NM_002354.2	80	1
mActb	222153	208173	193101	178538	107822	26466	0	182431	180821	211433	0	351304	178593	NM_007393.3		
mAlb	2079696	2490783	3465719	2514398	1969623	280099	0	2376240	1759517	1758750	0	4236799	2065659	NM_009654.3		
mCd24A	1725	1775	1229	1530	12501	3225	0	490	784	966	0	1842	1197	NM_007664.4		
mCd44	554	960	977	711	600	10937	0	869	700	884	0	975	696	NM_009846.2		
mCdh2	657	1080	805	840	945	13385	0	623	681	840	0	1131	712	NM_009851.2		
mFah	3828	37385	31513	9864	32438	637838	0	2063	2403	2266	0	98928	22385	NM_008240.3		
mFoxj1	62	95	529	276	3	1115	0	133	304	375	0	1454	767	NM_175284.3		
mFzd10	41	166	592	245	2	1102	11	80	67	119	307	794	470	NM_010176.4		
mGapdh	370	1229	503	329	172	10816	0	769	362	412	2	684	485	NM_010368.1		
mGusb	706	1233	1960	911	108	34171	0	1297	4621	4319	0	4774	1426	NM_001289726.1		
mHnf1B	182	482	682	297	172	5800	0	479	422	756	0	753	338	NM_001291268.1		
mJag1	32	334	310	90	18	4966	0	16	19	30	0	384	262	NM_013822.5		
mKrt19	1751	2025	3471	1679	2361	3460	0	2437	2291	3071	0	3311	1697	NM_008471.2		
mKrt7	2685	3004	4899	2843	5911	12193	0	7306	6218	8930	0	6034	3195	NM_033073.3		
mLmna	61817	59751	51286	47340	80796	44334	1	49978	76233	88338	11	69571	51227	NM_001002011.3		
mSmo	3101	3440	3211	2350	5879	14168	0	4049	4325	5069	1	4480	2625	NM_001204201.1		
mSox9	795	924	570	602	721	1901	1	1157	1202	1299	0	992	560	NM_176996.4		
mSpp1	1	0	4	0	17	329	0	6	2	0	0	3	3	NM_011448.4		
mTtr	1118	1448	5322	2720	469	11764	0	991	792	1032	5	7153	2479	NM_013697.5		
mVim	260	458	2790	883	2175	30612	0	758	399	400	0	3707	1388	NM_011701.4		
mZeb1	15081	12558	34180	16161	28419	64649	41	71139	82192	107686	527	47776	24449	NM_011546.3		
mEpCam	85	322	691	317	53	4660	0	274	242	236	0	1988	822	NM_008532.2		
TOTAL	2,894,477	3,501,089	4,542,118	4,164,540	2,251,204	1,213,595	677,224	3,481,679	2,391,957	2,649,870	467,518	5,010,901	2,722,288			
Human	497,862	673,786	738,465	1,381,933	52	265	677,170	778,368	267,602	452,895	466,665	168,052	361,675			
Mouse	2,396,615	2,827,303	3,803,653	2,782,607	2,251,152	1,213,330	54	2,703,311	2,124,355	2,196,975	853	4,842,849	2,360,613			
% human tags	17.20%	19.25%	16.26%	33.18%	0.00%	0.02%	99.99%	22.36%	11.19%	17.09%	99.82%	3.35%	13.29%			
Erroneous alignment					0.0023%	0.0218%					0.1825%					

Table S4: Total transcript-mapped reads in 22 selected genes from DDC treated humanized mice and controls.
Unique RNAseq transcripts from chimeric liver samples and controls (relates to Figure 7).

Supplemental Experimental Procedures

Detailed mouse strain information

The following mice used in this study are available from Jackson Labs.

Short name	Scientific name	Jax Stock #	Reference
ROSA-mTmG	B6.129(Cg)-*Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo*/J	007676	(Muzumdar et al., 2007)
R26R-Confetti	B6.Cg-Gt(ROSA)26Sortm1(CAG-Brainbow2.1)Cle/J	017492	(Snippert et al., 2010)
Sox9-CreERT2	Tg(Sox9-cre/ERT2)1Msan/J	018829	(Kopp et al., 2011)
Fah$^{-/-}$	Fah tm1Mgo	N/A	(Grompe et al., 1993)

Immunohistochemistry

Antibody or antigen	IgG Type	Dilution factor	Use	Product	Source
MIC1-1C3	Rat mAB	100	FACS	NBP1-18961	Novus Biologics
OC2-2F8	Rat mAB	20	FACS	Gift	Craig Dorrell, OHSU
Cd31	Rat mAB	100	FACS	561410	BD Biosciences
Cd45	Rat mAB	100	FACS, IF	552848	BD Biosciences
Cd11b	Rat mAB	100	FACS, IF	552850	BD Biosciences
Fah	Rabbit pAB	500	IF	Custom	Grompe Lab
Ostepontin	Goat pAB	100	IF	AF808	R&D systems
A6	Rat mAB	50	IF	Gift	Valentina Factor, NCI
Krt19	Rabbit pAB	500	IF	Gift	Xin Wang, U. Minn.
Sox9	Rabbit pAB	500	IF	AB5535	Millipore, Inc
Hnf4a	Rabbit pAB	100	IF	sc-8987	Santa Cruz Bio
CD44	Mouse mAB	100	FACS	560533	BD Biosciences
EPCAM	Mouse mAB	50	FACS, IF	F0860	Dako

PFA-fixed tissues were cut into 8μm sections on a freezing microtome (Cryostat, Leica), washed 2 x 15' in PBS, permeablized in 0.2-0.5% Triton-X 100 and blocked in 5% normal donkey serum for 2 hours to overnight. Primary antibodies were applied for 1 hour at room temperature or overnight at 4 degrees in a humidity chamber. Secondary antibodies (Donkey) and conjugated to AlexaFluor 488 or AlexaFluor 647 were used (Jackson Immunoresearch). Tissues were counterstained with Hoechst 33342 and mounted in Prolong Gold Antifade mounting media (Life Technologies). For EdU analysis, a 100mg/kg pulse was given by intraperitoneal injection 6 hours prior to sacrifice. The Click-iT labeling protocol was used according to the manufacturer instructions (Life Technologies) followed by co-immunohistochemistry. mTomato and mGFP fluorophores were directly observed without secondary antibody detection. Up to four-color images were captured on an Axio Imager M2 with color MRm and monochrome MRc 5 cameras (Carl Zeiss) running Xen Blue 2011. Large area images

were acquired with the Tiles module and stitched. This system is part of the Advanced Light Microscopy Core at the Jungers Center, OHSU Portland, Oregon.

Rnaseq data analysis

The first 4 bases from each read were trimmed and the subsequent 44 bases were aligned to the *Mus musculus* genome (UCSC mm9) using Bowtie version 0.12.7. Tag alignment was implemented using custom R scripts (http://www.R-project.org/) according to UCSC Refseq gene models. Only reads that uniquely aligned to the genome were counted. Gene expression levels were measured by RPKM (reads per kilobase of exon per million reads). For comparative gene expression analysis, we used the edgeR package with general linear model (Bioconductor 3.2.4)(Robinson et al., 2010). Significant genes were selected based on a false-discovery corrected p-value (q-value) cutoff of 0.01, a fold change >2, and RPKM > 1 in either group. Gene ontology analysis was performed using GSEA software (Broad Institute) using the canonical pathway KEGG modules. Unsupervised clustering was performed using Ward's method with log2 RPKM values for each gene.

Human-mouse chimeric liver samples were aligned to custom transcriptome index comprising 21 mouse and matched human Refseq RNA sequences (see Supplemental Table 4) with Bowtie version 0.12.7 set to allow no mismatches. Only tags that uniquely aligned to a single transcript were counted. Tags counts were compiled into a matrix using custom R-scripts, which are available upon request. Tag alignment was outputted to wiggle track format to examine the distributions of tags across each gene model as quality control. Expression levels were normalized to various housekeeping genes including GUSB, GAPDH, LMNA, and BACT. RNAseq fastq files and RPKM matrix were deposited into the NCBI database under accession numbers GSE55552 and GSE58679.

RT-PCR

RT-PCR was performed as previously described(Tarlow et al., 2014). Mouse primers were used at 60 °C annealing,
Alb1 F- GCG CAGA TGA CAG GGC GGA A
Alb1 R – GTG CCG TAG CAT GCG GGA GG
Gapdh F – AAG GTC GGT GTG AAC GGA TTT GG
Gapdh R - CGTTGAATTTGCCGTGAGTGGAG.

Human specific primers were used at 66 °C annealing
LMNA F – CCT CCT CGC CCT CCA AGA GC
LMNA R – AGA TGC GGG CAA GGA TGC AG
KRT19 F – CCT CCC GCG ACT ACA GCC ACT A
KRT19 R – CCA CTT GGC CCC TCA GCG TA.

Semiquantiative *Fah* genotyping PCR was performed at 62 °C with 3 primers
Wt – GGA TTG GGA AGA CAA TAG CAG GC

Mut – TGA GAG GAG GGT ACT GGC AGC TAC
Com. – TTG CCT CTG AAC ATA ATG CCA AC

Statistics
Calculations were performed with the Prism 6.0 statistical software package (Graphpad, Inc). Parametric pairwise or paired t-tests were performed where appropriate for image analysis data. Significance levels were defined as $p< 0.05$, $p<0.01$*, or $p<0.001$***.

Supplemental References

Grompe, M., Al-Dhalimy, M., Finegold, M., Ou, C.N., Burlingame, T., Kennaway, N.G., and Soriano, P. (1993). Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. Genes Dev. *7*, 2298–2307.

Kopp, J.L., Dubois, C.L., Schaffer, A.E., Hao, E., Shih, H.P., Seymour, P.A., Ma, J., and Sander, M. (2011). Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. Development *138*, 653–665.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593–605.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. Cell *143*, 134–144.

Tarlow, B.D., Finegold, M.J., and Grompe, M. (2014). Clonal tracing of Sox9+ liver progenitors in oval cell injury. Hepatology *60,* 278-289.