CD133^{+} Intercellsome Mediates Direct Cell-Cell Communication to Offset Intracellular Signal Deficit

Kota Kaneko, Yan Liang, Qing Liu, Wendy S. Chen and Gen-Sheng Feng *

Department of Pathology, Division of Biological Sciences, and Moores Cancer Center, University of California at San Diego, La Jolla, CA 92093

* **Corresponding author:** E-mail: gfeng@health.ucsd.edu

Abstract

CD133 (prominin 1) is widely viewed as a cancer stem cell marker, in association with drug resistance and cancer recurrence. Herein we report that in response to defective RTK-Shp2-Ras-Erk signaling, heterogeneous hepatocytes with upregulated CD133 clustered to proliferate during liver regeneration. Similarly, pharmaceutical inhibition of proliferative signaling induced CD133 expression transiently in various cancer cell types of different tissue origins, indicating an inherent mechanism of drug resistance. Although previously known as a cell surface protein, super-resolution and electron microscopy localized CD133 mainly on intracellular vesicles trafficking between cells directly, which we name "intercellsome". Overexpression of CD133 altered trafficking and morphology of intercellsomes, and isolated $CD133^+$ intercellsomes were enriched with immediate early gene (IEG) transcripts. Single-cell RNA sequencing revealed lower intracellular diversity (entropy) of IEG transcripts in proliferative signal-deficient cells, which was remedied by intercellular mRNA exchanges between $CD133^+$ cells. CD133-deficient cells were more sensitive to proliferative signal inhibition in livers and intestinal organoids. These data identify a long-sought CD133 function and a new mechanism of intercellular communication for cell proliferation under stress in healthy and cancer cells.

Main

CD133, also called Prominin-1 or AC133, was originally cloned as a pentaspan plasma membrane glycoprotein, predominantly expressed in various epithelial cells and hematopoietic stem cells [1-3], and evolutionarily conserved in invertebrates and vertebrates [4-6]. A body of literature has documented CD133 as a marker for stem/progenitor cells in numerous tissues, including skin, bone marrow, brain, kidney, liver, prostate and pancreas [7-9], although its function remains elusive and conflicting data exist regarding its expression profiles [7,10-12]. Generation of knockin mice with a reporter gene *LacZ* under control of the endogenous CD133 promoter revealed that CD133 expression was not restricted to stem/progenitor cells but was also detected in differentiated epithelial cells [9,13]. Homozygous CD133 knockout mice displayed healthy embryonic and postnatal development, with no severe defect observed in adult stem cells [9,13], except photoreceptor degeneration and loss of vision [14,15]. Consistently, genetic mutations were identified in *PROM1* gene in familial macular degeneration [16,17], which were phenocopied in transgenic mouse models [18]. CD133 is enriched in apical plasma membrane protrusions in neuro- and intestinal epithelial cells [19,20]. In particular, CD133 was shown to modulate the topology/morphology of cellular membrane and dynamics of microvilli [21,22]. However, the exact function and mechanism of CD133 remain to be deciphered in stem and non-stem cells.

Remarkably, CD133 has been widely regarded as a biomarker for cancer stem cells (CSC) in solid tumors in brain, colon, liver, lung, pancreas and prostate, etc [10,11,23]. However, along with the controversy on the existence of CSCs [24], the significance of CD133 expression in CSC or certain tumor cell subpopulations is unclear. CD133 expression was detected in cancer cells that did not display CSC properties, and CD133-negative tumor cells were able to form clonospheres in culture and to generate tumors in immune-deficient mice [9,23]. Regardless, $CD133^+$ cancer cells were often shown more resistant to various radio/chemo-therapeutics [23,25], which apparently contribute to tumor recurrence. Indeed, numerous studies suggest a prognostic value of CD133 expression in overall survival and progression-free survival in solid tumors [23,26]. Now, one urgent issue is to elucidate CD133 function in cancer progression and drug resistance, in association with the controversial CSC population.

Liver regeneration has served as an ideal model to dissect mechanisms that drive cell proliferation in vivo, with the unique regenerative capacity conserved from animals to humans [27-29]. One frequently used experimental platform for liver regeneration is partial hepatectomy (PHx) of two-thirds liver in rodents. Although re-entering cell cycle of quiescent

hepatocytes in adult liver is believed to be a primary mechanism of liver regeneration, possible involvement of stem/progenitor cells has also been explored [30,31]. However, conflicting data exist, regarding the sources and contributions of putative stem/progenitor cells in regeneration and hepato-carcinogenesis following hepatic injuries [30-33], although the identity and mechanism are yet to be deciphered.

Shp2/Ptpn11, an SH2-containing tyrosine phosphatase, is a key molecule that promotes signaling from receptor tyrosine kinases (RTKs) to the Ras-Erk pathway [32-36]. Interestingly, Shp2 has become an extremely hot pharmaceutical target in oncological treatment [32,37,38]. Ablating Shp2 in hepatocytes led to impaired hepatocyte proliferation following PHx, and Shp2 was required for signal relay from MET and EGFR, etc [39]. However, liver regeneration was impaired but not blocked by removal of Shp2 or other pro-proliferative molecules in hepatocytes [27-29], and hepatocellular cancer was even aggravated in Shp2-deficient liver [40]. Thus, one critical question is why and how a small number of hepatocytes overcome the diminished mitogenic signaling, to re-gain replicative capacity for regeneration and tumorigenesis.

In this study, we found that actively dividing hepatocytes in Shp2-deficient livers formed distinctive colonies and highly expressed CD133, but were negative for other stem cell markers. Following PHx, the CD133$^+$ cell colonies appeared transiently and disappeared quickly after completion of liver regeneration. A clonal tracing experiment revealed that the CD133$^+$ cell population was not generated by clonal expansion of a pre-existing unique cell type but originated from heterogeneous mature hepatocytes. This phenotype was also observed in MET-deficient livers following liver injuries, and even in liver tumors following proliferative signal suppression. CD133 expression was upregulated in a variety of cancer cell lines from different organs under proliferative signal inhibition, indicating a commonly shared mechanism. Remarkably, we identified and isolated CD133$^+$ intracellular vesicles, which were enriched with transcripts of immediate early responsive genes (IEGs) and trafficking directly between tightly contacting cells in colonies. Thus, we give a name “intercellsome” here, for vesicle that mediates direct cell-cell communication. Single cell RNA-sequencing (scRNA-seq) demonstrated that cells defective for proliferative signaling exhibited lower intracellular transcriptional diversity (entropy), which was remedied in CD133$^+$ cell clusters. These data reveal a long-sought CD133 function in normal and cancer cells and elucidate a mechanism by which cells strive to proliferate under proliferative signal deficit.

Results

CD133$^+$ proliferating cell clusters emerge transiently in regenerating liver

We chose mouse liver regeneration to dissect mechanisms of cell proliferation in mammals under physiological context. In previous experiments, we showed that removal of Shp2 suppressed Ras-Erk signaling and hepatocyte division following PHx in mice [41]. However, it was unclear why certain Shp2-negative hepatocytes managed to divide when proliferative signaling was impaired in general. To address this issue, we examined carefully the regenerating process in hepatocyte-specific Shp2 knockout (SKO, *Shp2$^{flox/flox}$:Alb-Cre*) liver. At 2 days after PHx, WT and SKO livers exhibited similar overall and histological phenotypes (Extended Data Fig. 1a,b). Immunostaining for Ki67 showed significantly lower numbers of proliferating hepatocytes in SKO than WT livers, with no difference for non-parenchymal cells (NPCs) where Shp2 was not deleted (Fig. 1a, Extended Data Fig. 1c,d). Intriguingly, we identified certain areas that were enriched with proliferating hepatocytes in the SKO liver (Fig. 1a; arrow). This peculiar proliferation pattern was especially cryptic, because these areas did not display severe inflammation or different phenotypes of hepatocytes except active proliferation (Fig. 1a; right panel). Therefore, we searched extensively for biomarker molecules, which could distinguish the actively proliferating cells in the particular area, including GFAP, EpCAM, CK19, CD133, Sox9, CD44 and α-fetoprotein (AFP) (Extended Data Fig. 1e-h). Of note, CD133 was the only molecule that was highly expressed in the area enriched with proliferating hepatocytes (Fig. 1b). CD133 was expressed in the clustered hepatocytes in a continuous manner, and the percentages of Ki67$^+$ hepatocytes were significantly higher in CD133$^+$ than CD133$^-$ areas in regenerating livers (Fig. 1c,d). These patchy areas appeared 2 days after PHx, and underwent expansion until 3 weeks, whereas WT livers showed few CD133$^+$ hepatocyte clusters (Fig. 1e). Liver/body weight ratios were comparable between SKO and WT mice by 3 weeks (Extended Data Fig. 1i), suggesting that active division of the CD133$^+$ cell population compensated the weakened proliferative capacity of CD133$^-$ hepatocytes in SKO liver.

The CD133$^+$ hepatocytes also expressed high levels of E-cadherin and displayed tight cell-cell contact in the colonies (Fig. 2a, Extended Data Fig. 1j,k). CD133 was localized to the apical side of hepatocytes (Fig. 2a), and was rarely detected in other types of cells located on their basal sides. At 3 weeks after PHx, the colonies appeared more distinctive even in H&E staining, which contained smaller hepatocytes with compact structure, albeit no difference observed for NPC types and the vasculature (Fig. 2b). Immunostaining for

sinusoidal endothelial cells and HGF did not show any unique features in the colonies otherwise (Extended Data Fig. 1l,m). The distinct colonies merged into surrounding tissues 5 weeks after PHx, as evidenced by loss of CD133 expression and disappearance of clear colony edges (Fig. 2c). Therefore, the CD133$^+$ colonies of proliferating hepatocytes were composed transiently during liver regeneration.

Due to the compacting colony architecture 3 weeks after PHx, we observed clear difference in light scattering/refraction in unstained thick liver sections (Fig. 2d), which allowed manual dissection of the colonies out of liver sections for analysis. Immunoblotting confirmed enrichment of CD133$^+$ and E-cadherin$^+$ cells but did not detect Shp2 protein in the isolated colonies, indicating their origin from Shp2-deleted hepatocytes (Fig. 2d,e). CD133 was also expressed in oval cells, which can differentiate into both hepatocytes and bile duct epithelial cells, similar to the potential of hepatoblasts (Fig. 2f). However, we did not observe expression of other oval cell markers, including EpCAM, AFP and Sox9, in the colony lysates (Fig. 2e), consistent with the immunostaining data (Extended Data Fig. 1e-h). Indeed, hepatocytes in the colonies did not exhibit oval cell-like morphology throughout the regenerative process (Fig. 1a,e, 2a-c). CD133 expression in colonized hepatocytes was even higher than the embryonic liver at E17.5, which highly expressed AFP, a hepatoblast marker (Fig. 2e). The molecular size of CD133 expressed in the colonies was slightly different from that in bile duct (Fig. 2e), likely reflecting variations in glycosylation [42,43]. Further, the colony-forming cells expressed a mature hepatocyte marker HNF4α, at similar levels to non-colony areas (Fig. 2e). These results indicate that the clustered CD133$^+$ cells in SKO liver were not hepatoblasts or oval cells but were mature hepatocytes that acquired unique proliferation potential following PHx.

The colonies originate from heterogeneous mature hepatocytes

To determine the origin of these colonized hepatocytes by clonal tracing, we transfected vectors carrying a GFP and a *sleeping beauty* (SB) transposon [44] into liver by hydrodynamic tail vein injection (HTVi) [45,46], which labeled a few hepatocytes with GFP in a mosaic pattern (Fig. 3a,b) [47,48]. We then performed PHx and analyzed regenerating livers 3 weeks later (Fig. 3a,b). The resulting colonies in SKO livers contained both GFP$^+$ and GFP$^-$ cells (Fig. 3b), indicating multiple origins of hepatocytes. When a large GFP$^+$ cell clone was located at a colony edge, we did not see recruitment of GFP$^-$ cells, even though the

surrounding hepatocytes were mostly GFP-negative (Fig. 3c). Thus, the colonies were expanded by proliferation of cells from inside rather than by continuous recruitment of surrounding hepatocytes (Extended Data Fig. 2a). By measuring the clone sizes of GFP^+ hepatocytes inside and outside, we found that hepatocytes within the colonies divided much more actively than those located outside (Fig. 3d,e), suggesting that clustering enabled hepatocytes to divide actively (Fig. 3f). The starting colony size was estimated to be approximately 10 hepatocytes, based on the GFP^+ clone/colony size ratios (Fig. 3g, Extended Data Fig. 2b). The initial colony size should be large enough to include a variety of hepatocytes, signifying that the colony-forming potential does not require a special property of hepatocytes. The existence of mononuclear and binuclear cells [49] in the colonies 2 days after PHx further supports their heterogeneity in origin (Extended Data Fig. 2c).

We reintroduced *WT* Shp2 together with GFP into SKO liver in a mosaic manner by HTVi, and analyzed the liver tissues 2 days after PHx (Extended Data Fig. 3a). Re-introducing Shp2 indeed restored the *WT* hepatocyte proliferation pattern in a cell-autonomous manner (Extended Data Fig. 3b,c). Although the colonies were mainly composed of Shp2-negative hepatocytes (Extended Data Fig. 3d), we observed $Shp2^+$ cells inside the colonies, which were also $CD133^+$ (Extended Data Fig. 3e). Thus, Shp2 deficiency was not a requirement to join the colony, although $Shp2^+$ cells disturbed colony morphology (Extended Data Fig. 3e). Shp2-negative hepatocytes sent out paracrine signals to trigger clustering with surrounding hepatocytes, including $Shp2^+$ cells in the neighborhood.

A similar pattern of colony formation was observed in cultures of primary hepatocytes isolated from SKO liver. Shp2-negative hepatocytes formed unique colonies of $Ki67^+$ cells in vitro, resembling the observed clusters in liver sections (Fig. 3h, Extended Data Fig. 2d), while proliferating hepatocytes were distributed rather randomly in the WT culture (Extended Data Fig. 2d). E-cadherin expression in cell-cell contacts distinguished these colonies (Fig. 3h), featuring their similar properties to those in vivo. This phenotype was independent of colony sizes and therefore was not due to high cell density in the area (Fig. 3h). The colony formation in primary hepatocyte culture also confirms their origin from mature hepatocytes.

CD133 upregulation is a common mechanism in response to proliferative signal deficit

CCl_4 is metabolized by hepatocytes around the central veins into a hyper-oxidative chemical that causes necrosis and apoptosis in the area [50,51]. We examined CD133 induction in the liver following CCl_4-induced injury, which involves cell death and inflammation,

largely different microenvironment from the PHx model. At 2 days after CCl_4 injection, we detected distinctive $CD133^+$ hepatocyte clusters in SKO livers with higher proliferation rate than the surrounding areas (Fig. 4a). Although CCl_4 also induced ductular reaction by $CD133^+$ bile duct epithelial cells [31], these areas were not associated with the $CD133^+$ hepatocyte clusters (Fig. 4a,b; arrowheads). Consistent with the PHx model, while bile duct epithelial cells were $EpCAM^+$, $CD133^+$ hepatocytes did not express EpCAM (Fig. 4a). The $CD133^+$ hepatocyte colonies appeared around the injured area, consistent with a notion of midlobular hepatocytes located next to a pericentral injury site in this model [52].

We examined the effect of hepatocyte-specific deletion of HGF receptor MET in mice. Similar $CD133^+$ proliferating hepatocyte clusters were induced by PHx or CCl_4 injection in MET-deficient livers (Fig. 4c). These results suggest that CD133 induction is not restricted to a specific gene deletion or an injury model, but rather is a response to impaired proliferative signaling. Indeed, treatment of PLC liver cancer cells in vitro with Shp2 inhibitor (SHP099) or Mek inhibitor (Trametinib), or a Shp2-targeting CRISPR vector, also upregulated CD133 expression (Fig. 4d,e), indicating that CD133 induction is an inherent cellular response to impaired signaling through the Ras-Erk pathway. Upregulated CD133 expression was independent of other stem cell markers and thus did not reflect forced selection of a stem cell population (Fig. 4d). Moreover, CD133 induction was rapid and dramatic in a manner that cannot be explained by selective expansion of pre-existing $CD133^+$ cells.

To explore if CD133 induction is a widely conserved mechanism, we examined a variety of cell lines derived from various organs and animals. Treatment with Shp2 or Mek inhibitors upregulated CD133 expression in mammary epithelial cells MCF10A (Fig. 4f,g), glioblastoma cells GSC3028 and GSC23, insulinoma cells MIN6, colon adenocarcinoma cells MC38, and cervical cancer cells Hela as well as kidney fibroblasts Cos7 (Fig. 4h). Together, these results indicate that CD133 upregulation under proliferative signal deficit is a common responding mechanism in various cell types.

Upregulated CD133 is mainly located on filament fibers connecting neighboring cells

Since the $CD133^+$ cell clustering was apparently triggered by local intercellular signal, we examined expression of molecules known to mediate intercellular communication at short distance [53-58]. We detected significantly higher expression of Wnt7a, Wnt10a and Shh in SKO liver (Extended Data Fig. 3f). Consistently, porcupine (PORCN), a protein that

resides in the ER and modifies Wnt for secretion [56,57,59,60], was highly expressed in clustered hepatocytes (Extended Data Fig. 3g), suggesting a role of local Wnt signaling in colony assembly. PORCN expression was also strongly associated with the colonies in cultured primary SKO hepatocytes in vitro (Extended Data Fig. 3h), and PORCN protein amounts were higher in SKO than WT liver lysates 2 days after PHx (Extended Data Fig. 3i). Furthermore, the mRNA level of Wnt10a was drastically elevated in response to PHx in SKO liver (Extended Data Fig. 3j). To probe the cellular resource, we isolated hepatocytes and NPCs from SKO livers 2 days after PHx. Wnt10a was mainly produced by hepatocytes, whereas the major source of Shh might not be hepatocytes (Extended Data Fig. 3k). These results suggest that the Wnt10a paracrine signal from SKO hepatocytes might be a local inducer of the colonies. Treatment of human HCC cells with Trametinib upregulated Wnt10a expression, and additional inhibition of PORCN suppressed CD133 induction (Extended Data Fig. 3l), suggesting a role of Wnt10a in CD133 induction under proliferative signal deficiency.

Interestingly, we observed CD133 on filament-like structures in colonized E-cadherin^{+} cells in primary SKO hepatocyte culture (Fig. 5a). These CD133^{+} filament-like structures bridged neighboring cells in the colonies (Fig. 5a,b), suggesting a direct cell-cell communication event. This direct communication model is consistent with a distinctive feature of clustered cells with extremely tight contacts in vivo and in vitro (Fig. 1b,d, 2a, 3f,h). As the filament-like staining pattern of CD133 observed here was not reported previously, we examined PLC cells with an antibody against human CD133. Indeed, we observed similar CD133^{+} filament structures inside and also connecting the cells (Fig. 5c). To rule out a possible artifact in immunostaining for CD133, we expressed a Myc-tagged CD133 fusion protein in PLC cells and stained with antibodies against CD133 and the Myc-tag. The two signals overlapped on the filament-like structure, confirming the subcellular localization of CD133 (Extended Data Fig. 4a,b). As CD133 protein possesses five transmembrane domains and a membrane targeting signal peptide, we postulated that the observed filament-like CD133 staining represents CD133^{+} vesicles on filaments. Indeed, the CD133 signal co-localized with tubulin filaments (Extended Data Fig. 4c). CD133 overexpression led to a thicker pattern of CD133^{+} filaments, especially at drastically high expression levels (Extended Data Fig. 4a-c). Similar CD133^{+} structure was induced by Shp2 and Met inhibitors in MCF10A, HeLa and MC38 cells (Fig. 5d, Extended Data Fig. 4d). Thus, despite a previous view on CD133 as a cell surface molecule, this study demonstrated that CD133 is

mainly located on a trafficking machinery in cells.

CD133$^+$ intercellsome is a new vesicle type induced under proliferative signal deficit

The filamentous staining profile of CD133 was different from the conventional droplet-like patterns often observed with endosomes or exosomes in a multi-vesicular body. To unequivocally determine subcellular localization of CD133 and the CD133$^+$ filaments, we performed stochastic optical reconstruction microscopy (STORM) imaging. Indeed, the CD133$^+$ filaments observed at low resolution were CD133$^+$ vesicles tightly associated with and distributed on tubulin filaments at super-resolution (Fig. 5e). CD133 overexpression modified the morphology and topology of the vesicles (Fig. 5f), suggesting a role of CD133 in the vesicle assembly and dynamics. Quantitative co-localization analysis confirmed physical location of CD133$^+$ vesicles on tubulin filaments, while the value decreased after CD133 overexpression (Fig. 5e,f, Extended Data Fig. 4e). The lowered co-localization value following CD133 overexpression likely reflected altered morphology of the vesicles, stretching away from the tubulin filaments, which also explained the bulky appearance in low resolution images (Extended Data Fig. 4a-c). While this function is consistent with a reported role of CD133 on lipid membrane morphology/topology [21,22], we observed CD133 on cell surface membrane only when exogenous CD133 was over-expressed (Extended Data Fig. 4b,f). In agreement with STORM imaging, immunofluorescence and immunogold electron microscopy on a cryo-ultramicrotome section of the colonies in regenerating SKO liver further confirmed localization of CD133 on intracellular vesicles enriched between the apical lumens, rather than distributed on the apical membrane surface (Fig. 5g, Extended Data Fig. 4g).

We isolated CD133$^+$ vesicles from PLC cells using magnetic beads bound with anti-CD133 antibody, with the vesicle yields dramatically increased after pre-treatment of the cells with Trametinib (Fig. 5h). Notably, CD133$^+$ vesicles did not contain exosome markers CD9, CD63 or CD81, supporting their identity as a new type of vesicle, different from the classical exosome (Fig. 5h). Co-staining of endosome markers, RAB5A and RAB7, showed no overlap with CD133, with no CD133 signal detected in the multivesicular body where exosomes accumulated (Extended Data Fig. 5a). Evidently, CD133 was located on a filament network that connects neighboring cells, instead of endosomes or exosomes. Therefore, we give a name intercellsome to this new type of vesicle identified here.

Intercellsome contains unique cargos and travels between neighbor cells

We isolated CD133$^+$ intercellsomes from SKO liver lysates following PHx, and found that the vesicles were not enriched with signaling molecules, such as Stat3 and Erk1/2, compared to the CD133$^-$ fractions (Extended Data Fig. 5b), suggesting that proliferative signaling proteins were not the most critical cargos. We then shifted attention to the RNA components in the vesicles. CD133$^+$ intercellsomes were strikingly different from CD133$^-$ fractions in the overall RNA profiling and contained minimal amount of rRNAs or micro-RNAs, with the majority being mRNAs (Fig. 6a). The expression of immediate early-responsive genes (IEGs), including *c-Jun, JunB, and c-Myc,* were upregulated rapidly following PHx [61,62]. qRT-PCR analysis showed that these mitogenic mRNAs were enriched in CD133$^+$ intercellsomes, with no enrichment detected for housekeeping gene transcripts, such as *Gapdh* and *Ppia*, (Fig. 6b, Extended Data Fig. 5c), showing selective inclusion of mRNAs. Similar results were obtained with CD133$^+$ intercellsomes isolated from PLC cells treated with Trametinib (Extended Data Fig. 5d).

We performed RNA-seq analysis of CD133$^+$ intercellsomes isolated from Trametinib-treated PLC cells, extracellular vesicles (EVs) secreted by the same cells, and the whole cells. Small RNAs including miRNAs, snoRNAs and snRNAs were less abundant in CD133$^+$ intercellsomes than the whole cells (Fig. 6c). In contrast, miRNAs were enriched in the EVs compared to the whole cells (Fig. 6c). Although the majority of RNA species detected in EVs were still mRNAs, this data could be influenced by the library preparation with total RNAs, which missed lots of small RNAs. Nevertheless, the enrichment of miRNAs in the EVs was in sharp contrast to significantly enriched mRNAs in CD133$^+$ intercellsomes (Fig. 6c), clearly distinguishing the two types of vesicle. Compared to the whole cells, 4309 mRNAs were upregulated, with 3974 mRNAs downregulated, in CD133$^+$ intercellsomes (Fig. 6c). The enrichment of IEGs in CD133$^+$ intercellsomes was also confirmed in the RNA-seq data, while most of these transcripts were at low levels in the EVs (Fig. 6d, Extended Data Fig. 5e). Consistently, single molecule in situ hybridization showed localization of *c-Myc* mRNA on CD133$^+$ filaments, verifying the IEG existence in intercellsomes (Fig. 6e,f).

To demonstrate direct shuttling of intercellsomes between cells, we established a GFP-labelled cell line that stably expressed a CD133-Myc-tag fusion protein. After mixing these cells with mCherry-expressing cells without the fusion protein, we traced the Myc-tag signal (Fig. 6g). Indeed, the CD133-Myc-tag protein was transported from GFP$^+$ cells to the neighboring mCherry$^+$ cells through the bridge (Fig. 6h). Importantly, neither GFP nor

mCherry signal travelled between the cells, indicating selective CD133$^+$ vesicle traffic. Taken together, these results suggest that intercellsome is a unique vesicle type, moving between neighbor cells and containing particular cargos that are different from those in EVs.

mRNA sharing converts intercellular heterogeneity into intracellular diversity of IEGs

HuR is an RNA-binding protein that binds to AU-rich elements in IEG transcripts for their stabilization, and shuttles between nucleus and cytoplasm. This protein was enriched in CD133$^+$ intercellsomes, as revealed by immunoblotting (Fig. 5h, 7a) and was localized to the apical side of CD133$^+$ hepatocytes (Extended Data Fig. 5f). Besides the nuclei, immunostaining detected HuR in some peri-nuclear areas (Fig. 7b; arrows) and co-localization with CD133 in PLC cells (Fig. 7b). The CD133$^+$ filaments bridged the neighboring cells with HuR visible on the bridging filaments (Fig. 7b; arrowheads), suggesting migration of HuR-bound mRNAs directly from one cell to another. Treatment of PLC or MCF10a cells with Shp2 or Mek inhibitors markedly increased the density of HuR on CD133$^+$ filaments (Fig. 7c, Extended Data Fig. 5g), with even brighter HuR signals than the nuclei. Given that CD133 was found to modify the vesicle patterns (Fig. 5e,f, Extended Data Fig. 4a-c), we tested if CD133 overexpression also affected HuR trafficking. Indeed, the pattern of HuR changed together with alterations in the CD133$^+$ machinery (Fig. 7d). Furthermore, lipofection of PLC cells with isolated CD133$^+$ intercellsomes enhanced cyclin D1 expression, supporting a notion that mitogenic mRNAs were carried in the vesicles (Fig. 7e,f, Extended Data Fig. 5d,h). Collectively, these results suggest that CD133$^+$ intercellsomes are upregulated in cells to remedy mitogenic signal deficit.

We wondered how the exchange of mitogenic mRNAs within cell colonies could boost proliferative signals. If the cells with impaired signaling had equally reduced amounts and types of mitogenic mRNAs, exchange would not be beneficial. We hypothesized that the expression of different IEGs was variably downregulated, which triggered intercellular exchange of mRNAs through CD133$^+$ intercellsomes (Fig. 7g). This process may convert intercellular heterogeneity to higher intracellular IEG diversity (calculated as entropy), without increasing total IEG expression levels. To test this theory, we performed scRNA-seq of WT and SKO hepatocytes at 4 hrs and 2 days following PHx, using our newly established protocol for liver research [63,64]_ENREF_62_ENREF_63. Hepatocyte identity was confirmed by unsupervised clustering and analyses of multiple markers, with contaminated NPCs excluded from further analyses (Extended Data Fig. 6). We focused attentions on IEGs,

which were highly induced following PHx as documented in the public database (Methods), and were also detected in isolated CD133$^+$ vesicles. First, we checked if Shp2 deletion suppressed expression of different IEGs to variable extents. Indeed, the heatmap demonstrated stochastic expression profiles of different IEG species in individual cells (Extended Data Fig. 7a). CD133$^+$ cells were easily identified in SKO liver due to its high expression (Extended Data Fig. 7b). Of note, CD133$^+$ cells were not enriched with particular hepatocyte subtype markers or stem cell markers, compared to CD133$^-$ cells in SKO liver (Extended Data Fig. 7c). Principal component analysis (PCA) of the IEG expression profiles showed overlap of all samples with each other, instead of partitioning into distinct groups, suggesting no biased expression of specific IEGs in the CD133$^+$ cell population in SKO liver (Extended Data Fig. 7d). Nonetheless, clustering of CD133$^+$ cells in the center of PCA plot showed that despite the stochastic expression pattern, CD133$^+$ cells were featured by relatively similar and stable levels of all IEG species. By measuring the overall expression amounts of all IEGs in the selected list, we found that the total IEG levels were higher in WT than SKO cells at 4 hr after PHx (Fig. 7h). We next analyzed IEG diversity within individual cells and variations in the expression pattern among cells. For fair comparison, the analysis was focused on Cyclin D1$^+$ cells in the periportal area, to avoid influences of cell cycle and zonal location in the liver (Extended Data Fig. 7e, Methods). Indeed, the intracellular IEG diversity was high and intercellular variations were low among CD133$^+$ hepatocytes, relative to other groups (Fig. 7i,j). This data distinguished CD133$^+$ cells of SKO liver from WT hepatocytes and also revealed that their increased entropy of IEGs was not due to restoration of a WT-like expression pattern (Fig. 7g). A plot with SKO hepatocytes also showed clearly higher IEG diversity in CD133$^+$ than CD133$^-$ cells, with similar total IEG expression levels (Fig. 7k). Consistent with the significance of the selected IEGs for cell proliferation, we observed a positive correlation between Cyclin D1, IEG diversity and the total IEG levels (Fig. 7k). When simulating the mRNA exchange, the plot pattern of CD133$^-$ cells in SKO liver moved toward a striking resemblance to the CD133$^+$ population, increasing IEG diversity within each cell (Fig. 7l,m). Mathematical modeling demonstrated that exchanging 1/12 of IEGs in a cell with 4 other cells twice was sufficient to reach the distinctive IEG profile of CD133$^+$ cells (Fig. 7l,m, Extended Data Fig. 8a-c, Methods). Together, these data suggest that the enhanced proliferative phenotype of CD133$^+$ cells is driven by reciprocal mRNA sharing within clusters, despite signaling deficiency in the whole population.

CD133 is required to sustain stem and non-stem cell proliferation under signal deficit

Given the compensatory CD133 upregulation as a resistance mechanism, we reasoned that a defective phenotype associated with CD133 loss could be amplified in a compound mutant also deficient for a pro-proliferative molecule. To test this theory, we crossed the *Prom1* KO mouse with a conditional Shp2$^{flox/flox}$ mouse line, and deleted Shp2 in hepatocytes by AAV-Cre virus infection. At 2 days after PHx, hepatocyte proliferation in E-Cad^{+} colonies was significantly suppressed in the Prom1/Shp2 double KO hepatocytes, compared to SKO control (Fig. 8a,b), although Prom1 KO alone did not inhibit Shp2^{+} hepatocyte proliferation (Fig. 8b). We also isolated hepatocytes from SKO and Shp2/Prom1 DKO (Prom1 KO; Shp2$^{flox/flox}$; Alb-Cre) livers, and analyzed the cell colonies formed in culture in vitro. Cell proliferation in E-Cad^{+} colonies was markedly lower in DKO than SKO hepatocytes (Fig. 8c,d). However, the basal low levels of proliferation were similar between SKO and DKO non-colonized cells, consistent with a role of intercellsomes in direct cell-cell communication within colonies. We also observed higher proliferation rates of DKO hepatocytes within E-Cad^{+} colonies than non-colonized cells in culture, suggesting that CD133 is unlikely the sole mediator of the intercellular communication event. Next, we labeled SKO hepatocytes by injecting AAV-GFP into SKO mice, and mixed the GFP-labelled SKO cells with DKO hepatocytes at 1:10 ratio in culture (Fig. 8e). If CD133 had a cell-autonomous function, SKO cells in a DKO cell pool were expected to proliferate similarly as in a pure SKO cell pool. For example, if CD133 serves as a receptor on cell surface, SKO cells in the mixture should still receive a ligand signal from DKO cells. By contrast, the SKO cells scattered within DKO colonies only proliferated as much as DKO cells (Fig. 8f,g), arguing against a receptor-like or cell-autonomous role of CD133. CD133 removal in DKO cells did not abolish direct cell-cell contact, with tight cell adhesions maintained, as shown by E-cad staining (Fig. 8f).

We performed scRNA-seq with hepatocytes isolated from DKO liver 2 days after PHx (Extended Data Fig. 9). In the absence of CD133, we chose Porcn as an alternative marker for colonized cells, as this protein was enriched in intercellsome-inducing SKO hepatocytes (Extended Data Fig. 3g,h). After applying MAGIC (Markov affinity-based graph imputation of cells) [65], we found that Porcn^{+} SKO hepatocytes indeed expressed higher CD133 transcripts (Fig. 8h). Porcn^{+} DKO hepatocytes also expressed higher *Prom1-CreERT2 mutant* transcripts (Fig. 8h), suggesting that the "Porcn^{+} population" could be used as a marker for colonies induced in both SKO and DKO cells. Next, we checked Cyclin D1 in the Porcn^{+}

population and observed lower levels in DKO than SKO cells, indicating impaired proliferation (Fig. 8i). According to our model (Fig. 7g), if the IEG mRNA exchange was impaired, the IEG variations would be higher among Porcn$^+$ DKO cells relative to Porcn$^+$ SKO hepatocytes. Indeed, the IEG variation among cells did not decrease in Porcn$^+$ population in DKO cell pool as it did in the SKO cell pool (Fig. 8j), suggesting impaired intercellular communication and mRNA exchange after CD133 removal.

Given the abundant CD133 expression in intestinal crypt, we isolated the crypt cells from WT and Prom1 KO mice for organoid culture in vitro. Consistent with previous data [7], CD133 was expressed in intestinal crypt cells, likely including both the Lgr5$^+$ stem cells and the more progenitor-like transit amplifying cells (Extended Data Fig. 10a). In both the tissues and the organoids, CD133 was enriched to the apical side, similar to its expression profile in the liver. Consistent to previous data, no significant difference in cell proliferation was observed between WT and Prom1 KO organoids (Extended Data Fig. 10b,c). However, treatment with Trametinib exhibited more prominent inhibition on cell proliferation in the crypt buds of Prom1 KO organoids than the WT control (Extended Data Fig. 10b,c), indicating higher sensitivity of CD133-deficient intestinal progenitor cells to impaired Ras-Erk signaling. Together, these data illustrate a functional role of CD133 in different cell types for strive to proliferate under signal deficit.

Discussion

Originally aimed at deciphering hepatocyte proliferation under proliferative signal deficit, this study has unveiled a new intercellular communication mechanism shared in various cell types to resist signal inhibition. Interestingly, this was mediated by CD133, a popular and yet elusive CSC marker. By taking multidisciplinary approaches, especially super-resolution microscopy and scRNA-seq, we identified intercellsome, a new type of CD133$^+$ vesicle for direct cell-cell exchange of mitogenic mRNAs. Given that cells strive to survive through autophagy under nutritional deficiency, we believe this is an inherent mechanism by which cells endeavor to divide under mitogenic signal deficiency.

CD133 was known as a stem cell marker in normal and tumor tissues, although its expression is not restricted to the stem cell populations in either case. We found that CD133 expression was transiently induced in Shp2- and MET-deficient hepatocytes in regenerating livers damaged by PHx or CCl_4. The CD133 upregulation was not accompanied by increase in other stem cell markers or associated with cell lineage transition. Further, the reversable

CD133 induction was detected in various cell types following suppression of RTK-Ras-Erk signaling by Shp2 and Mek inhibitors. These data on inducible and dynamic CD133 expression suggest that its expression pattern is not strictly associated with cell types but rather is modulated by intra- and inter-cellular signals, which may clarify many conflicting data in the literature. Given that proliferative signaling can be frequently disturbed in health and disease, upregulated CD133 expression and function is likely a conserved mechanism for cell proliferation under stress. Notably, we observed basal levels of CD133 expression and CD133$^+$ intercellsomes in various cell types under physiological conditions, which can be markedly stimulated by various stresses.

This study identified CD133$^+$ filaments that connect neighboring cells (Fig. 5). Super-resolution microscopy and immuno-EM further demonstrated that CD133 was primarily located on intercellsomes, rather than on plasma membrane. CD133 evidently plays a critical role in assembly and function of intercellsomes, which mediate intercellular sharing of materials between tightly contacted cells with minimal diffusion. It remains to be elucidated if the CD133$^+$ vesicles migrate directly between neighboring cells or through a process of exo- and endo-cytosis. However, this newly disclosed role of CD133 explains why CD133/Prom 1 deletion did not affect general development in mice, as CD133$^+$ vesicles are mainly generated in stress responses. However, CD133 is unlikely the sole molecule responsible for formation of intercellsomes; other molecules may also participate in this important cellular event. It is also plausible that cells use a CD133-independent mechanism to cope with stress in the absence of CD133. Further, this study did not exclude a possibility that CD133 has other functions in cell activities, independent of intercellsomes. Although we found that its expression on intercellsomes was widely conserved, CD133 might play a different role on cell surface membrane in a context that was not studied here. CD133$^+$ vesicles were reportedly detected in internal and external body fluids [66], although their cell origin and functional significance are unclear. It is conceivable that these reported CD133$^+$ vesicles could be intercellsomes released from damaged cells rather than excreted by intact cells.

Unlike CD63/CD9/CD81-positive exosomes or EVs [67], CD133$^+$ intercellsomes are enriched with mRNAs rather than miRNAs, especially the IEG transcripts. scRNA-seq analysis showed that the exchange of mRNAs between CD133$^+$ cells converted intercellular stochastic heterogeneity into higher intracellular diversity (entropy). Simulating the process with math modeling suggests that the reciprocal mRNA exchange among defective cells is

sufficient to restore a proliferative capacity of CD133^{+} hepatocytes despite the deficient proliferative signaling in the entire population. The diversity of mitogenic mRNAs is more important than upregulation of individual genes during cell proliferation. Therefore, while CD133-mediated mRNA sharing to increase the mRNA diversity is also used by non-stem cells for compensation of signal deficiency, stem cells may more often require this mechanism to maintain their functions, especially under stress. In agreement with our data, recent scRNA-seq analyses suggested intracellular transcriptional diversity (entropy) as a common feature of stem cells [68-70]. The intracellular transcriptional diversity can be calculated as Shannon's entropy or a count of gene transcript species within individual cells. The high diversity of mRNAs within individual cells explains why stem cells can shift their identity via differentiation or maturation, during which the transcriptional diversity is narrowed down for more specific lineages. This property may explain why stem cells are more resilient in general, as higher transcriptional diversity makes a more robust system that can buffer various disturbances. Therefore, the intracellular transcriptional diversity is not merely a signature of stem cells but is functionally required for such robustness. Despite common features of various stem cells, there is no universal molecule that can identify or define all types of stem cells. Massive genomic data analyses have revealed different sets of biomarkers expressed in various stem cells. However, the intracellular diversity is likely a common property of all stem cells as well as actively dividing non-stem cells, which is maintained at least in part by CD133^{+} intercellsomes identified in this study.

Despite the controversy on the concept of CSCs and also on CD133 as its biomarker, many reports have documented a significant correlation between CD133 expression, drug resistance and patients' prognosis. The CD133^{+} cancer cells are reportedly more resistant to radio/chemotherapies and thus likely responsible for tumor recurrence [25,71-73]. We observed markedly increased expression of CD133 in several tumor cell lines following treatment with growth-inhibitory compounds, which are already used in clinical treatment or trials (Figure 4). These data suggest a plasticity of CD133 expression in acute response of tumor cells to anti-tumor drugs, and a group of tumor cells may manage to proliferate through CD133-mediated intercellular communication upon inhibition of proliferative signals. This swift response to drug treatment does not require new mutations and is unrelated to the CSC identity. If considering "stemness" as a trait rather than an identity, it is likely that any cancer cells acquiring high transcriptional diversity may show some signs of stemness, such as drug resistance. The CD133^{+} intercellsomes are assembled

and disassembled transiently, independent of the degree of differentiation. Thus, this study reveals a long-sought CD133 function in cancer and may clarify the controversial issues regarding CD133 and the so-called cancer stem cells. Simultaneously suppressing intracellular proliferative signaling and disrupting the compensatory intercellsome function may be a new strategy to overcome drug resistance and tumor relapse.

Acknowledgments

We thank Dr. E. Wang, and our lab members for reagents and helpful discussion. Single cell RNA-seq was conducted at the IGM Genomics Center, UCSD. Some cell images were taken at the UCSD School of Medicine Microscopy Core (supported by NINDS P30NS047101), super-resolution microscopy was performed at Moores Cancer Center Microscopy Core (supported by P30CA23100-28), and cryo-ultramicrotome and electron microscopy were performed at Electron Microscopy Core Facility, UCSD. This work was supported by NIH grants (R01DK128320, R01CA236074 and R01CA239629) to G.S.F. W.S.C. was supported by an NIH training grant (T32CA009523).

Author contributions

K.K. and G.S.F. conceived the project. G.S.F. supervised the project, secured funding and provided the reagents. Under the supervision, K.K. designed and performed the experiments and developed the model; Y.L. did the scRNA-seq and data analysis, and was the developer of all the unique codes in the study; Q.L. contributed to cell culture experiments and W.S.C. provided liver tissue sections. K.K. and G.S.F. discussed the data analysis and interpretation, built the concepts and wrote the manuscript.

Data availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The codes generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Gen-Sheng Feng (gfeng@ucsd.edu).

References

1 Yin, A. H. *et al.* AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* **90**, 5002-5012 (1997).

2 Weigmann, A., Corbeil, D., Hellwig, A. & Huttner, W. B. Prominin, a novel microvilli-specific polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells. *Proc Natl Acad Sci U S A* **94**, 12425-12430, doi:10.1073/pnas.94.23.12425 (1997).

3 Miraglia, S. *et al.* A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. *Blood* **90**, 5013-5021 (1997).

4 Weigmann, A., Corbeil, D., Hellwig, A. & Huttner, W. B. Prominin, a novel microvilli-specific polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells. *Proceedings of the National Academy of Sciences* **94**, 12425-12430, doi:10.1073/pnas.94.23.12425 (1997).

5 Corbeil, D., Karbanová, J., Fargeas, C. A. & Jászai, J. in *Prominin-1 (CD133): New Insights on Stem & Cancer Stem Cell Biology* (ed Denis Corbeil) 3-24 (Springer New York, 2013).

6 Russo, M. T., Racioppi, C., Zanetti, L. & Ristoratore, F. Expression of a single prominin homolog in the embryo of the model chordate Ciona intestinalis. *Gene Expression Patterns* **15**, 38-45, doi:https://doi.org/10.1016/j.gep.2014.04.001 (2014).

7 Barker, N., Bartfeld, S. & Clevers, H. Tissue-Resident Adult Stem Cell Populations of Rapidly Self-Renewing Organs. *Cell Stem Cell* **7**, 656-670, doi:10.1016/j.stem.2010.11.016 (2010).

8 Kuang, S., Gillespie, M. A. & Rudnicki, M. A. Niche Regulation of Muscle Satellite Cell Self-Renewal and Differentiation. *Cell Stem Cell* **2**, 22-31, doi:10.1016/j.stem.2007.12.012 (2008).

9 Shmelkov, S. V. *et al.* CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J Clin Invest* **118**, 2111-2120, doi:10.1172/JCI34401 (2008).

10 Clevers, H. The cancer stem cell: premises, promises and challenges. *Nature Medicine* **17**, 313, doi:10.1038/nm.2304 (2011).

11 Medema, J. P. Cancer stem cells: The challenges ahead. *Nature Cell Biology* **15**, 338, doi:10.1038/ncb2717 (2013).

12 Zhu, L. *et al.* Multi-organ Mapping of Cancer Risk. *Cell* **166**, 1132-1146.e1137, doi:https://doi.org/10.1016/j.cell.2016.07.045 (2016).

13 Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457**, 603-607, doi:10.1038/nature07589 (2009).

14 Zacchigna, S. *et al.* Loss of the cholesterol-binding protein prominin-1/CD133 causes disk dysmorphogenesis and photoreceptor degeneration. *J Neurosci* **29**, 2297-2308, doi:10.1523/JNEUROSCI.2034-08.2009 (2009).

15 Dellett, M. *et al.* Genetic background and light-dependent progression of photoreceptor cell degeneration in Prominin-1 knockout mice. *Invest Ophthalmol Vis Sci* **56**, 164-176, doi:10.1167/iovs.14-15479 (2014).

16 Maw, M. A. *et al.* A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. *Hum Mol Genet* **9**, 27-34, doi:10.1093/hmg/9.1.27 (2000).

17 Zhang, Q. *et al.* Severe retinitis pigmentosa mapped to 4p15 and associated with a novel mutation in the PROM1 gene. *Hum Genet* **122**, 293-299, doi:10.1007/s00439-

007-0395-2 (2007).
18 Yang, Z. *et al.* Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. *J Clin Invest* **118**, 2908-2916, doi:10.1172/JCI35891 (2008).
19 Corbeil, D. *et al.* The human AC133 hematopoietic stem cell antigen is also expressed in epithelial cells and targeted to plasma membrane protrusions. *J Biol Chem* **275**, 5512-5520, doi:10.1074/jbc.275.8.5512 (2000).
20 Marzesco, A. M. *et al.* Release of extracellular membrane particles carrying the stem cell marker prominin-1 (CD133) from neural progenitors and other epithelial cells. *J Cell Sci* **118**, 2849-2858, doi:10.1242/jcs.02439 (2005).
21 Röper, K., Corbeil, D. & Huttner, W. B. Retention of prominin in microvilli reveals distinct cholesterol-based lipid micro-domains in the apical plasma membrane. *Nature Cell Biology* **2**, 582, doi:10.1038/35023524 https://www.nature.com/articles/ncb0900_582#supplementary-information (2000).
22 Thamm, K. *et al.* Prominin-1 (CD133) modulates the architecture and dynamics of microvilli. *Traffic* **20**, 39-60, doi:https://doi.org/10.1111/tra.12618 (2019).
23 Glumac, P. M. & LeBeau, A. M. The role of CD133 in cancer: a concise review. *Clin Transl Med* **7**, 18, doi:10.1186/s40169-018-0198-1 (2018).
24 Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337, doi:10.1038/nature12624 (2013).
25 Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760, doi:10.1038/nature05236 (2006).
26 Grosse-Gehling, P. *et al.* CD133 as a biomarker for putative cancer stem cells in solid tumours: limitations, problems and challenges. *J Pathol* **229**, 355-378, doi:10.1002/path.4086 (2013).
27 Taub, R. Liver regeneration: from myth to mechanism. *Nat Rev Mol Cell Biol* **5**, 836-847 (2004).
28 Fausto, N., Campbell, J. S. & Riehle, K. J. Liver regeneration. *Hepatology* **43**, S45-S53, doi:doi:10.1002/hep.20969 (2006).
29 Michalopoulos, G. K. & DeFrances, M. C. Liver Regeneration. *Science* **276**, 60-66, doi:10.1126/science.276.5309.60 (1997).
30 Michalopoulos, G. K. Hepatostat: Liver regeneration and normal liver tissue maintenance. *Hepatology* **65**, 1384-1392, doi:10.1002/hep.28988 (2017).
31 Miyajima, A., Tanaka, M. & Itoh, T. Stem/Progenitor Cells in Liver Development, Homeostasis, Regeneration, and Reprogramming. *Cell Stem Cell* **14**, 561-574, doi:https://doi.org/10.1016/j.stem.2014.04.010 (2014).
32 Chen, Y.-N. P. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148, doi:10.1038/nature18621 https://www.nature.com/articles/nature18621#supplementary-information (2016).
33 Tajan, M., de Rocca Serra, A., Valet, P., Edouard, T. & Yart, A. SHP2 sails from physiology to pathology. *European Journal of Medical Genetics* **58**, 509-525, doi:https://doi.org/10.1016/j.ejmg.2015.08.005 (2015).
34 Grossmann, K. S., Rosario, M., Birchmeier, C. & Birchmeier, W. The tyrosine phosphatase Shp2 in development and cancer. *Advances in cancer research* **106**, 53-89, doi:10.1016/s0065-230x(10)06002-1 (2010).
35 Matozaki, T., Murata, Y., Saito, Y., Okazawa, H. & Ohnishi, H. Protein tyrosine phosphatase SHP-2: a proto-oncogene product that promotes Ras activation. *Cancer science* **100**, 1786-1793, doi:10.1111/j.1349-7006.2009.01257.x (2009).
36 Chan, R. J. & Feng, G. S. PTPN11 is the first identified proto-oncogene that encodes

a tyrosine phosphatase. *Blood* **109**, 862-867 (2007).
37 Vainonen, J. P., Momeny, M. & Westermarck, J. Druggable cancer phosphatases. *Science Translational Medicine* **13**, eabe2967, doi:doi:10.1126/scitranslmed.abe2967 (2021).
38 Kerr, D. L., Haderk, F. & Bivona, T. G. Allosteric SHP2 inhibitors in cancer: Targeting the intersection of RAS, resistance, and the immune microenvironment. *Current Opinion in Chemical Biology* **62**, 1-12, doi:https://doi.org/10.1016/j.cbpa.2020.11.007 (2021).
39 Bard-Chapeau, E. A. *et al.* Concerted Functions of Gab1 and Shp2 in Liver Regeneration and Hepatoprotection. *Molecular and Cellular Biology* **26**, 4664-4674, doi:10.1128/mcb.02253-05 (2006).
40 Bard-Chapeau, Emilie A. *et al.* Ptpn11/Shp2 Acts as a Tumor Suppressor in Hepatocellular Carcinogenesis. *Cancer Cell* **19**, 629-639, doi:https://doi.org/10.1016/j.ccr.2011.03.023 (2011).
41 Bard-Chapeau, E. A. *et al.* Concerted functions of Gab1 and Shp2 in liver regeneration and hepatoprotection. *Mol Cell Biol* **26**, 4664-4674, doi:26/12/4664 [pii] 10.1128/MCB.02253-05 (2006).
42 Barkeer, S., Chugh, S., Batra, S. K. & Ponnusamy, M. P. Glycosylation of Cancer Stem Cells: Function in Stemness, Tumorigenesis, and Metastasis. *Neoplasia (New York, N.Y.)* **20**, 813-825, doi:10.1016/j.neo.2018.06.001 (2018).
43 Kemper, K. *et al.* The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. *Cancer research* **70**, 719-729, doi:10.1158/0008-5472.can-09-1820 (2010).
44 Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvák, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* **91**, 501-510, doi:https://doi.org/10.1016/S0092-8674(00)80436-5 (1997).
45 Zhang, G., Budker, V. & Wolff, J. A. High Levels of Foreign Gene Expression in Hepatocytes after Tail Vein Injections of Naked Plasmid DNA. *Human Gene Therapy* **10**, 1735-1737, doi:10.1089/10430349950017734 (1999).
46 Liu, F., Song, Y. K. & Liu, D. Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Therapy* **6**, 1258, doi:10.1038/sj.gt.3300947 (1999).
47 Mikkelsen, J. G. *et al.* Helper-Independent sleeping beauty Transposon–Transposase vectors for efficient nonviral gene delivery and persistent gene expression in vivo. *Molecular Therapy* **8**, 654-665, doi:https://doi.org/10.1016/S1525-0016(03)00216-8 (2003).
48 Yant, S. R. *et al.* Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature Genetics* **25**, 35, doi:10.1038/75568 (2000).
49 Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707, doi:10.1038/nature09414 https://www.nature.com/articles/nature09414#supplementary-information (2010).
50 Weber, L. W., Boll, M. & Stampfl, A. Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model. *Critical reviews in toxicology* **33**, 105-136, doi:10.1080/713611034 (2003).
51 Forbes, S. J. & Newsome, P. N. Liver regeneration — mechanisms and models to clinical application. *Nature Reviews Gastroenterology &Amp; Hepatology* **13**, 473, doi:10.1038/nrgastro.2016.97 (2016).
52 Wei, Y. *et al.* Liver homeostasis is maintained by midlobular zone 2 hepatocytes.

Science **371**, doi:10.1126/science.abb1625 (2021).
53 Tsiairis, Charisios D. & Aulehla, A. Self-Organization of Embryonic Genetic Oscillators into Spatiotemporal Wave Patterns. *Cell* **164**, 656-667, doi:10.1016/j.cell.2016.01.028 (2016).
54 Rogers, K. W. & Schier, A. F. Morphogen Gradients: From Generation to Interpretation. *Annual Review of Cell and Developmental Biology* **27**, 377-407, doi:10.1146/annurev-cellbio-092910-154148 (2011).
55 Lander, A. D. Morpheus Unbound: Reimagining the Morphogen Gradient. *Cell* **128**, 245-256, doi:10.1016/j.cell.2007.01.004 (2007).
56 Mikels, A. J. & Nusse, R. Wnts as ligands: processing, secretion and reception. *Oncogene* **25**, 7461, doi:10.1038/sj.onc.1210053 (2006).
57 Nusse, R. Wnts and Hedgehogs: lipid-modified proteins and similarities in signaling mechanisms at the cell surface. *Development* **130**, 5297-5305, doi:10.1242/dev.00821 (2003).
58 Teleman, A. A., Strigini, M. & Cohen, S. M. Shaping Morphogen Gradients. *Cell* **105**, 559-562, doi:https://doi.org/10.1016/S0092-8674(01)00377-4 (2001).
59 Tammela, T. *et al.* A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. *Nature* **545**, 355, doi:10.1038/nature22334 https://www.nature.com/articles/nature22334#supplementary-information (2017).
60 Clevers, H., Loh, K. M. & Nusse, R. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, doi:10.1126/science.1248012 (2014).
61 Fausto, N. Liver regeneration. *Journal of Hepatology* **32**, 19-31, doi:https://doi.org/10.1016/S0168-8278(00)80412-2 (2000).
62 Haber, B. A., Mohn, K. L., Diamond, R. H. & Taub, R. Induction patterns of 70 genes during nine days after hepatectomy define the temporal course of liver regeneration. *The Journal of clinical investigation* **91**, 1319-1326, doi:10.1172/JCI116332 (1993).
63 Liang, Y. *et al.* Temporal analyses of postnatal liver development and maturation by single-cell transcriptomics. *Developmental Cell* **57**, 398-414.e395, doi:https://doi.org/10.1016/j.devcel.2022.01.004 (2022).
64 Chen, W. S. *et al.* Single-cell transcriptomics reveals opposing roles of Shp2 in Myc-driven liver tumor cells and microenvironment. *Cell Rep* **37**, 109974, doi:10.1016/j.celrep.2021.109974 (2021).
65 van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e727, doi:10.1016/j.cell.2018.05.061 (2018).
66 Marzesco, A. M. Prominin-1-containing membrane vesicles: origins, formation, and utility. *Adv Exp Med Biol* **777**, 41-54, doi:10.1007/978-1-4614-5894-4_3 (2013).
67 Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology* **9**, 654, doi:10.1038/ncb1596 https://www.nature.com/articles/ncb1596#supplementary-information (2007).
68 Grün, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266-277, doi:https://doi.org/10.1016/j.stem.2016.05.010 (2016).
69 Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nature Communications* **8**, 15599, doi:10.1038/ncomms15599 https://www.nature.com/articles/ncomms15599#supplementary-information (2017).
70 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental

potential. *Science* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

71 Ma, S., Lee, T. K., Zheng, B. J., Chan, K. W. & Guan, X. Y. CD133+ HCC cancer stem cells confer chemoresistance by preferential expression of the Akt/PKB survival pathway. *Oncogene* **27**, 1749-1758, doi:10.1038/sj.onc.1210811 (2008).

72 Sarvi, S. *et al.* CD133+ cancer stem-like cells in small cell lung cancer are highly tumorigenic and chemoresistant but sensitive to a novel neuropeptide antagonist. *Cancer Res* **74**, 1554-1565, doi:10.1158/0008-5472.CAN-13-1541 (2014).

73 Bertolini, G. *et al.* Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. *Proc Natl Acad Sci U S A* **106**, 16281-16286, doi:10.1073/pnas.0905653106 (2009).

74 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

75 Su, A. I., Guidotti, L. G., Pezacki, J. P., Chisari, F. V. & Schultz, P. G. Gene expression during the priming phase of liver regeneration after partial hepatectomy in mice. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11181-11186, doi:10.1073/pnas.122359899 (2002).

76 Locker, J. *et al.* A common set of immediate-early response genes in liver regeneration and hyperplasia. *Hepatology* **38**, 314-325, doi:10.1053/jhep.2003.50299 (2003).

77 Miao, Z., Deng, K., Wang, X. & Zhang, X. DEsingle for detecting three types of differential expression in single-cell RNA-seq data. *Bioinformatics* **34**, 3223-3224, doi:10.1093/bioinformatics/bty332 (2018).

78 Wang, T., Li, B., Nelson, C. E. & Nabavi, S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinformatics* **20**, 40, doi:10.1186/s12859-019-2599-6 (2019).

Methods

Animals. Hepatocyte-specific Shp2 knockout (*Alb-Cre; Shp2*$^{flox/flox}$) or SKO mice were generated as described previously [40]. c-Met$^{flox/flox}$ mice (FVB;129P2-*Met*^{tm1Sst}/J) were purchased from the Jackson Laboratory and bred with Alb-Cre mice to produce hepatocyte specific c-Met KO mice. Prom1 KO mice (B6N;129S-Prom1$^{tm1(cre/ERT2)Gilb}$/J) were purchased from the Jackson Laboratory and bread with Shp2$^{flox/flox}$ and SKO mice. Shp2$^{flox/flox}$ littermates were used as WT control mice. All experimental procedures were approved by the UCSD Institutional Animal Care and Use Committee (IACUC). PHx was performed on mice at age of 8-10 weeks with the use of Buprenorphine hydrochloride (Par Pharmaceutical) as a pain relief for the surgery. CCl_4 (20% in 100 μl corn oil/20g of body weight; Sigma) was intraperitoneally injected into 8-10 week-old mice. AAV-Cre (AAV8.TBG.PI.Cre.rBG) was purchased from the Penn Vector Core and injected into the tail vein ($2x10^{11}$/28 gbw). For chemical induction of hepatocellular carcinoma (HCC), DEN (25 mg/kg; N0258-1G; Sigma) was injected intraperitoneally into mice at postnatal day 15. BrdU (1 mg/mice; Sigma) was intraperitoneally injected twice a day for 3 days.

Cell culture and experiments. PLC/PRF/5 cells were cultured in DMEM medium supplemented with 10% fetal bovine serum. The MCF10A cells were cultured in DMEM-F12 medium (Gibco) with 5% horse serum (Gibco) and 1% penicillin-streptomycin supplemented with 20 ng/mL human EGF(Gemini), 0.5 μg/mL hydrocortisone, 100 ng/mL cholera toxin (Sigma-Aldrich), and 10 μg/mL human insulin (Gemini). The cultures were maintained at 37 ºC in a humidified environment containing 5% CO2. For inhibitors treatment, PLC/PRF/5 cells were treated with HGF and SHP099 (10 μM; CHEMIETEK; CT-SHP099) or trametinib (10 nM; APExBIO; A3018) for 2-6 days after cell plating and starvation overnight at the second day. MCF10A cells were treated with SHP099 (20 μM) or trametinib (10 nM) for 2-6 days after cell plating. Medium was changed every two days. Cells were transfected with CD133-Myc-Tag expression vector (HG15024-CM; SinoBiological) using Lipofectamine 3000 transfection reagent (Invitrogen). For lentivector transfection, the fusion protein and fluorescent proteins were subcloned into pLJM1 vector. HEK293T cells were transfected with the lentivectors together with psPAX2 and pCMV-VSV-G, using Lipofectamine 3000 (Thermo Fisher Scientific). Supernatant was collected, centrifuged at 3000g for 5min, filtered through 0.45μm filter and used to transfect the PLC cells. The lentiCRISPR v2 was used for the CRISPR experiment.

Immunostaining. Freshly frozen sections or fixed frozen sections were used for immunofluorescent staining. Freshly frozen sections and cultured cells were fixed by cold 4% PFA in PBS, cold acetone or cold methanol for 30 min to overnight, depending on the antibodies used. Antigen retrieval was performed for some stainings for 10 min to 1 hour, depending on the antibody used. Antibodies for Ki67 (14-5698-80; eBioscience), HNF4α (sc-8987; Santa Cruz Biotechnology), mouse CD133 (14-1331-82; eBioscience), human CD133 (86781; CST), E-cadherin (sc-7870; Santa Cruz Biotechnology), GFP (04404-84; Nacalai Tesque), Shp2 (sc-280; Santa Cruz Biotechnology), Porcupine (ab105543; abcam), β-catenin (sc-7199; Santa Cruz Biotechnology), CHMP2B (ab157208; abcam), HuR (ab200342; abcam, and 66549-1-Ig; proteintech), VE-Cad (AF1002; R&D SYSTEMS), BrdU (555627; BD Biosciences), PECAM (13-0311-81; eBioscience), HGF (ab83760; abcam), GFAP (Z0334; Dako), EpCAM (552370; BD Biosciences), Sox9 (AB5535: Millipore), CD44v6 (GTX75661; GeneTex) and AFP (AF5369; R&D SYSTEMS) were used as primary antibodies. Fluorescent images were taken with a monochrome camera, and pseudo colors were applied for each channel. STORM images were taken with Nikon A1R TIRF STORM Microscope at the Moores Cancer Center. Cryo-ultramicrotome sectioning, fluorescence staining, immunogold staining and electron-microscopy (EM) were performed by the Electron Microscopy Core Facility at UCSD.

In situ hybridization. Freshly frozen sections of the regenerating SKO liver and PLC cell line treated with MEKi were subjected to HCR RNA-FISH, using probes and reagents purchased from Molecular Instruments after the CD133 staining. Freshly frozen sections were fixed with PFA for 30min, washed with PBS and further fixed/permeabilized with EtOH overnight. PLC was fixed with MeOH overnight. The fixed sections and cells were then immunostained for CD133 using the antibodies described above. After the immunostaining, HCR RNA-FISH was performed according to the protocol provided by Molecular Instruments. ProtectRNA RNase Inhibitor (R7397; SIGMA) was added to all the reagents used. The images were acquired using Leica SP8 confocal microscopy.

Primary hepatocyte isolation and culture. Each liver was perfused with 10 ml of HBSS (without Ca^{2+} or Mg^{+}) containing 0.5 mM EDTA, 4 ml of HBSS (without Ca^{2+} or Mg^{+}) and 10 ml of collagenase solution [2 mg/ml collagenase H (11074059001; Roche), 0.1 mg/ml DNaseI (10104159001; Roche), 5 mM $CaCl_2$ and 0.9 mM MgCl in HBSS] sequentially. All the solutions were pre-warmed to 37 °C before use. Livers were dissected and further incubated

in 7 ml of collagenase solution at 37 °C for 5 min and dissociated. Dissociated cells were passed through cell strainer and centrifuged at 50 g for 5 min to precipitate hepatocytes. For qRT-PCR, the pellet was immediately dissolved in Trizol for the hepatocyte fraction. The supernatant containing NPCs was washed with PBS at 50 g for 5 min three times to eliminate hepatocytes, and precipitated at 400 g for 5 min and dissolved in Trizol and used as NPC fraction. For hepatocyte culture, the hepatocyte pellet after the first centrifugation was resuspended and washed with PBS at 50 g for 5 min three times to eliminate NPCs, resuspended in the culture media [Williams' Media E supplemented with Primary Hepatocyte Maintenance Supplements (CM4000; Gibco) as instructed] and seeded on collagen-coated plates.

Intestinal organoid culture. After dissecting the intestinal tube from the mice, the tubes were washed with PBS, by flushing with syringe. The villi were gently scraped off and the tissues were cut to smaller pieces. The tissue fragments were rinsed with PBS three times, and then incubated in PBS with 1mM EDTA for 30 minutes. After the incubation, the fragments were put in fresh PBS and rigorously shaken, and the cells/debris that came off were discarded. The tissue fragments were then incubated in PBS with 5mM EDTA for one hour. After the incubation, the fragments in the solution were rigorously shaken and pipetted, and the cell clumps that came off were used as the isolated crypts. The crypts were passed through 100μm cell strainer and spun down and washed with PBS at 300 x g for 10 minutes three times. The isolated crypt fragments were embedded in Matrigel and cultured in DMEM/F12 containing B27 (17504044; Gibco), N2 supplement (17502048; Gibco), 1.25 mM n-Acetylcysteine, 50 ng/ml EGF, 100 ng/ml Noggin and 500 ng/ml R-spondin 1. Organoids were passaged in the same media, and treated with 50 nM trametinib.

Immunoblotting. Proteins were extracted in RIPA buffer, and immunoblotting was performed using standard protocols. In addition to antibodies used for immunostaining, antibodies for EpCAM (bs-1513R; Bioss), HNF4α (GTX89532; GeneTex), GAPDH (5174; CST), β-actin (A5316; Sigma), Stat3 (9132; CST), Erk1/2 (4695; CST) and Cyclin D1 (sc-20044; Santa Cruz Biotechnology) were used as primary antibodies.

Hydrodynamic tail vein injection. Plasmids pT3-elongation factor 1α (EF1α)-Shp2 (5 μg/ml), pT3-EF1α-GFP (5 μg/ml), p-cytomegalovirus (pCMV)-sleeping beauty (SB) transposase (1.5 μg/ml), or were diluted in PBS. Total volume of 2 ml/20 g.b.w. was rapidly injected into tail vein. The GFP-expressing vector has transposon sequences, and when injected

together with the SB transposase-expressing vector, the GFP expressing sequence was integrated into the genome. Plasmids were prepared using the GenElute HP Endotoxin-Free Plasmid Maxiprep Kit (NA0410; Sigma). Silica fines in the final product were carefully eliminated by additional centrifugation at 12,000 g for 30 min.

Isolation of CD133$^+$ vesicles. Liver perfusion was similar to hepatocyte isolation as described above, except that collagenase Type IV (17104019; Gibco) was used instead of collagenase H to minimize unspecific digestion of proteins other than collagen. The livers were dissected, immersed in the collagenase solution and dissociated and mildly lysed. Cells and debris were removed by centrifugation at 4,000 g for 10 min three times, followed by filtration with 0.45 μm PES filter. We obtained around 2 ml of solution per mouse at this point. EDTA was added to suppress the collagenase activity and sodium azide was added to prevent possible degradation by bacterial growth. For PLC cells, cells were lysed in PBS by Dounce homogenizer, and cells and debris was removed similarly. FITC conjugated anti-CD133 antibody (1:500 dilution; 11-1331-82; Invitrogen) was added, and the solution was incubated at 4 °C overnight. Objects labeled with anti-CD133 antibody were then targeted with magnetic particles, using FITC selection cocktail (18558; STEMCELL TECHNOLOGY), and isolated using MACS MS column (130-042-201; Miltenyi Biotec). CD133$^+$ fraction was directly eluted with either lysis buffer for immunoblotting or lysis buffer for RNA isolation using RNeasy Plus Micro Kit (74034; QIAGEN). Isolation protocol for total RNAs including small RNAs was used according to the manufacturer's instruction with the RNeasy Plus Micro Kit. For the agarose gel electrophoresis, total RNAs were normalized to same concentrations, mixed with RNA loading dye (final concentration of 47.5% formamide, 0.01% SDS, 0.01% bromophenol blue, 0.005% Xylene Cyanol FF, 0.01% SYBR safe, 0.5 mM EDTA), denatured at 65 °C for 15 min, cooled on ice, and loaded to the gel. For treatment of cells with the isolated vesicles, Anti-PE MultiSort Kit (130-090-757; Miltenyi) was used instead of FITC selection cocktail, to enzymatically detach the vesicles from the magnetic beads. As the vesicles were not likely secreted and endocytosed naturally, we forced the cells to take the vesicles by using Lipofectamine 3000.

Isolation of extracellular vesicles. Extracellular vesicles were isolated from the culture supernatant using Total Exosome Isolation Reagent (Invitrogen). Debris or cell contaminants were removed by centrifugation, larger vesicles such as apoptotic bodies were subsequently removed by filtration, and extracellular vesicles were precipitated according to

the manufacturer's instruction. Total RNAs were extracted in the same way as the extraction from CD133+ vesicles described above.

qRT-PCR. RNAs extracted by either Trizol or RNeasy Plus Micro Kit were reverse transcribed to cDNAs using High-Capacity cDNA Reverse Transcription Kit (4368814; Applied Biosystems), and quantitative PCR was performed using DyNAmo Flash SYBR Green qPCR Kit (F415S; Thermo Scientific). Primer sequences are listed in supplementary Table S1.

Single cell RNA library construction and sequencing. Single hepatocytes were isolated from WT at 4 hrs (4h) and 2 days (2d), SKO at 4h and 2d, and DKO at 2d after PHx, by a two-step in-situ perfusion with collagenase H as described above. To remove dead cells and debris, the pelleted hepatocytes were resuspended in 45% Percoll (sigma-aldrich) and centrifuged at 50g for 10 min. Cells were further washed with PBS and counted with hemocytometer. The isolated hepatocytes were loaded onto a 10x Chromium Controller and then partitioned into nanoliter-scale Gel Beads-In-Emulsion (GEMs). The volume of single cell suspension was calculated in order to generate 5000 GEMs per sample. Libraries were constructed with Chromium Single Cell 3' Reagent Kits (v3 chemistry, 10x Genomics). Sequencing was performed using HiSeq 4000 (Illumina) at IGM Genomics Center, University of California, San Diego, with the following read length: Read 1, 28 bp, including 16 bp cell barcode and 12 bp unique molecular identifier (UMI); Read 2, 98 bp transcript insert; i7 sample index, 8 bp. For the single cell RNA-seq data, GEO accession number: GSE169320

Single cell RNA-seq data preprocessing. The paired reads from WT and SKO samples were aligned to mouse reference genome GRCm38 using CellRanger package (v3.0.2). And the reads from DKO were aligned to modified GRCm38, with Prom1-CreERT2 sequence manually added. An expression matrix, with each row representing a gene and each column representing a cell, was generated for each sample, and then filtered for downstream analysis. In brief, for each sample, genes expressed in less than 3 cells were removed; cells failed to meet following criteria were removed: 1) the number of genes detected in each cell should be over 200 and less than 5000; 2) the UMIs of mitochondrial genes should be less than 20% of total UMI; 3) the UMIs of a cell should be less than 20,000 to avoid non-singlet (i.e. transcriptome representing more than one cell).

Detection of cell subpopulations. To identify subpopulations within hepatocytes

and select clusters for further analyses, we examined each sample using R package Seurat (v2.4) [74]. First, the raw expression matrix was normalized by the total expression, multiplied by scale factor 10,000, and log-transformed. Next, we regressed on the number of UMIs per cell as well as mitochondrial gene percentage. After normalization and scaling, z-scored residuals were stored for dimensionality reduction and clustering (Extended Data Fig. 6, Fig. 9a). The hepatocytes showed high expression of *Hnf4a* and *Asgr1*. And other clusters with expression of *Ptprc*, *Lyve1*, *Pecam1*, *Lrat*, *Hba-a1*, and *Krt19* (contaminant immune cells, endothelial cells, hepatic stellate cells, and erythrocytes) were removed. Next, for hepatocytes in each sample, PCA was performed first, and 75 PCs were used for tSNE analysis with other parameters, including perplexity = 30 and maximum iteration = 2000. To find hepatocyte subpopulations, the same PCs were imported into FindClusters from Seurat. All samples were also colored by *Alb* (marker for hepatocytes), *Cyp2f2* (marker for periportal hepatocytes) and *Cyp2e1* (marker for pericentral hepatocytes) expression. $CD133^+$ cells of SKO liver (2d after PHx) were not limited to, but abundantly identified in one of the two clusters of periportal hepatocytes (80%), which was consistent with our observations on tissues of the PHx model. After comparing differentially expressed genes in two periportal clusters, we found the cluster enriched with $CD133^+$ cells in SKO 2d sample, had higher *Alb* and *Hamp* expression, presumably reflecting a population that is located in a particular zone within the periportal area. The same population (Alb^{high}, $Hamp^{high}$, $Cyp2f2^+$) was also observed in other samples. For fair comparison, further analysis was limited to this periportal population to avoid variances due to spatial locations within the liver tissue.

Calculation of total IEGs' expression levels. To obtain the IEG list for analysis, we checked data in public database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) and publications [75,76], for genes expressed during liver regeneration or after liver injury. Among the genes upregulated, we eliminated the genes that were irrelevant to cell growth or those acting as negative feedback regulator. As a result, we obtained 12 IEGs for further analyses, including *Fos, Jun, Egr1, Ier2, Atf3, Junb, Myc, Crem, Ets2, Ier3, Lepr, Egfr*. To integrate the total expression level of the 12 genes, we defined a score based on zero-inflated Negative Binomial (ZINB) model. ZINB regression is for modeling count variable with excessive zeros. It was first introduced to model single-cell RNA-seq data by Miao et al. [77], and it had shown high accuracy in differential expression analysis compared to other models [78]. The score for total IEGs expression was computed as follows. 1) Within a group of cells,

we first fitted counts of each gene by a ZINB model:

$$f_{ZINB}(n;\theta,r,p) \equiv P(N=n) = \theta \cdot I(n=0) + (1-\theta) \cdot f_{NB}(n;r,p)$$

$$= \theta \cdot I(n=0) + (1-\theta) \cdot \binom{n+r-1}{n} p^n (1-p)^r$$

where θ is a proportion of 'real' zeros in this group of cells ('real' zeros represent that the gene was not expressed in some cells, instead of the dropout caused by mRNA capture procedure); r and p were parameters from negative binomial distribution, representing size and probability, respectively; n is the median normalized count [77] of genes. ZINB model connects 'real' zeros and a negative binomial model with θ and indicator function I. The observed excessive zeros from single-cell RNA-seq dataset might come from either 'real' zeros or zeros in negative binomial model. 2) Therefore, the expected count or mean for each gene was calculated as follow:

$$\mu_{ZINB} = P(\text{gene not expressed}) \cdot 0 + P(\text{gene expressed}) \cdot \mu_{NB}$$

$$= (1-\theta) \cdot \frac{pr}{1-p}$$

For each group of cell c and each gene g, we calculated the expected count as $\mu_{c,g}$. With the same method, the expected count of gene g among all cells used for this analysis was calculated as $\mu_{a,g}$. 3) For cell cluster c, the score of total IEGs expression level was defined as:

$$S_c = 2^{\frac{1}{N_g}\sum_g \log_2\left(\frac{\mu_{c,g}}{\mu_{a,g}}+0.001\right)}, \text{ where } N_g = 12$$

4) For a single cell in a group, a, for example, we calculated its binary distance to all other cells and selected 20 closest cells as a subgroup, called c_a. The score for c_a was calculated following step 1) to 3) as described above and plotted as a dot in Figure 6h.

IEG diversity within each cell and variation among cells. We first used Shannon's entropy of all cells in a group to indicate the IEG diversity. The entropy E_j of cell j was computed as:

$$E_j = -\sum_{i=1}^{N} p_{i,j} \log_2 P_{i,j}$$
$$p_{i,j} \equiv r_{i,j} / \sum_i r_{i,j}$$
$$r_{i,j} \equiv count_{i,j} / \mu_i$$

where $count_{i,j}$ represents the median normalized count of gene i in cell j, and μ_i equals the expected count among all cells used for analysis. The higher entropy defined above practically reflects 1) more species of IEGs are expressed by a cell, as well as 2) less overall deviation from mean expression values of IEGs within a cell. To describe the variation among cell subpopulation, we used cell pair-wise Euclidean distance with normalized and scaled data by Seurat. However, in order to do fair comparison between samples when using scaled data, we first integrated four datasets by Canonical Correlation Analysis (CCA) with Seurat, before calculating the distance. Finally, the significance of diversity and variation between cell groups were decided by Wilcoxon rank sum test and adjusted by FDR.

RNA exchange simulation. We plotted IEG diversity (entropy) against total IEG expression levels in SKO-2d sample. The cells were first divided into CD133$^-$ and CD133$^+$ groups. Within each group, all cells were ordered by their entropies from low to high. Starting from the first cell, we applied a moving average to entropies, with a sliding window of 20 cells and a step size of 5. Within each window, we also computed the total IEG expression value as described above. We then plotted IEG entropy against total expression level for CD133$^-$ and CD133$^+$ groups. Every sliding window with 20 single cells was treated as a meta-cell. Next, we examined how the cells would behave in the plot, when simulating the RNA exchange among CD133- cells. The basic idea was to randomly divide all CD133$^-$ cells into small groups with size N and make cells within the same group exchange IEGs with each other. We tested four possible RNA exchange models (Extended Data Fig. 6e):

- Model 1: Within a small group of size N, for each IEG, the cell with the highest expression gives X of its RNA to each of the other (N-1) cells; the corresponding RNA level is reduced by (N-1)*X in the donor cell.

$$\begin{cases} N = 3,5,7 \\ X = \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{3}{16}, \frac{1}{4} \end{cases}$$

- Model 2: Within a small group of size N, for each IEG, RNA is transferred as described in Model 1, but RNA level is maintained in the donor cell because of its continuous transcription.

$$\begin{cases} N = 3,5,7 \\ X = \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{3}{16}, \frac{1}{4} \end{cases}$$

- Model 3: Within a small group of size N, for each IEG, there are two rounds of RNA exchange. Considering that the CD133 expression was already detected (in RNA level) at 4 hours after PHx (Extended Data Fig. 6b), this is likely the reasonable model, as they should have multiple chances of mRNA exchange while they change their profile along the time until 2 days after PHx. The first round is the same as Model 1: recipients get X_1 of the RNA in a donor cell and the donor cell loses (N-1)*X_1. After the first round of exchange is finished, the cell with highest expression now is designated as a new donor. During second round, new donor loses (N-1)*X_2 of its RNA and each of the other (N-1) cells get X_2.

$$\begin{cases} N = 3,5,7 \\ X_1 = \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{3}{16}, \frac{1}{4} \\ X_2 = \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{3}{16}, \frac{1}{4} \end{cases}$$

- Model 4 (exchange all): Within a small group of size N (N = 3, 5, 7), for each IEG and each cell, the RNA levels become the same, which equal to the mean values before exchange.

After the RNA exchange simulation was applied, the IEG diversity (entropy) and total IEG expression were re-calculated in order to track their changes. For Figure 6m, the IEG diversity was calculated as described above, and the analysis included both Cyclin D1-negative and -positive cells.

MAGIC data imputation. To set the threshold of *Porcn* expression and select Porcn+ cells in SKO and SKO 2d after PHx, the data imputation method called Markov Affinity-based Graph Imputation of Cells (MAGIC) was performed. MAGIC was developed

to fill in missing transcripts, via sharing information across similar cells. We made density plot with imputated *Porcn* expression values and determined a threshold that could separate two peaks. Cells were grouped into Porcn- and Porcn+ subpopulations based on this threshold. The MAGIC imputated expression levels of other genes were visualized and compared using violin plots as indicated in this manuscript.

Bulk RNAseq sample preparation and data processing. Total RNAs extracted from vesicles as described above were subjected to ribosome-depleted total RNA library prep (Illumina) at UCSD Genomics Core. Sequencing was performed at UCSD Genomics Core on the NovaSeq 6000 platform. Raw reads data quality control was performed using fastqc, sequenced reads were mapped to the hg19 reference genome using STAR, and the number of mapped reads to each gene was counted by htseq-count. The R package DESeq2 was then used to normalize raw reads and perform differential expression analysis between CD133+ intercellsomes vs whole cell samples, as well as EVs vs whole cell samples. The significantly expressed genes were determined with adjusted p value < 0.05.

Heatmap. For single cell RNA-seq, we randomly selected 54 cells from the same periportal cluster as described above, to make the heatmap. The raw counts were log-transformed with pseudo-count = 1. For bulk RNA-seq, we made two heatmaps to compare gene expression between CD133+ intercellsomes and EVs. In Fig. 6d, the variance stabilizing transformed values of interested genes from CD133+ intercellsomes and EVs were extracted and scaled. In Extended Data Fig. 5e, the transformed values from CD133+ intercellsomes and EVs were further standardized to whole cells.

Statistical analysis. All statistical analyses of data are described in the corresponding figure legends. All statistical analyses were performed with GraphPad Prism 7, except R was used for the single cell RNA-seq analyses.

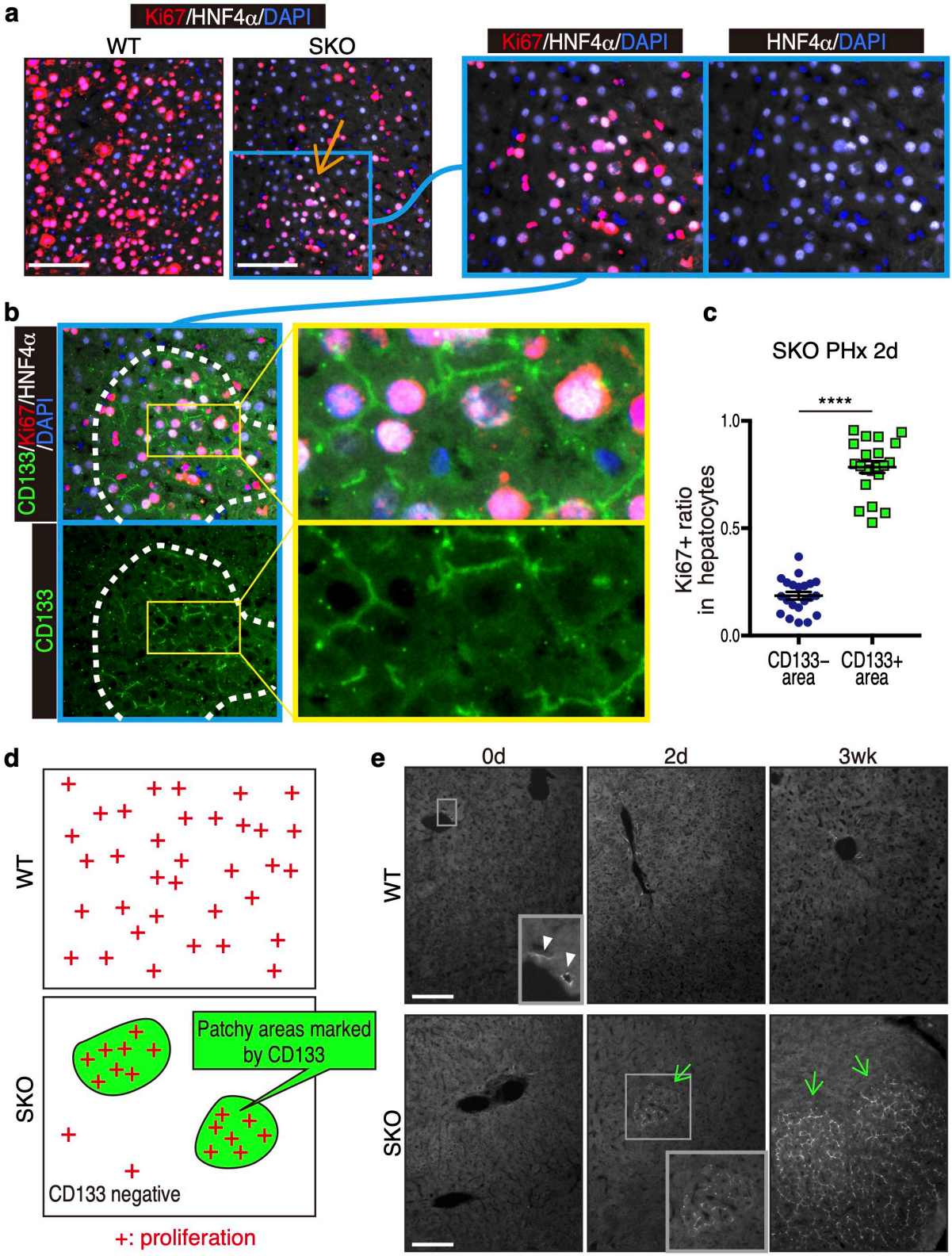


Fig. 1| Identification of a patchy hepatocyte proliferation pattern in Shp2-deficient (SKO) liver after PHx

a, **b**, Immunofluorescence on liver tissue sections 2 days after PHx. HNF4α is a hepatocyte

marker and Ki67 is a proliferation marker. Arrow in **a** points to an area enriched with proliferating hepatocytes. Dashed line in **b** shows an area with continuous CD133 expression. **c**, Quantification of proliferating rate in hepatocytes in CD133-positive and -negative areas in SKO livers 2 days after PHx. Each dot indicates one area. Data were collected from 3 mice. Means ± SEM are shown. ****$P < 0.0001$ (two-tailed unpaired t test).
d, While WT hepatocytes proliferated at high frequency everywhere, proliferating hepatocytes in SKO liver were mostly located in patchy areas marked by CD133 expression. **e**, Immunofluorescence of CD133 on liver tissues at day 0 or 2, and 3 weeks (0d, 2d, 3 wk) after PHx. $CD133^+$ hepatocyte clusters were only found in SKO livers after PHx (light green arrows). In WT livers, CD133 expression was only seen in bile duct epithelial cells (arrowheads).

Scale bars, 100 μm (**a**, **e**).

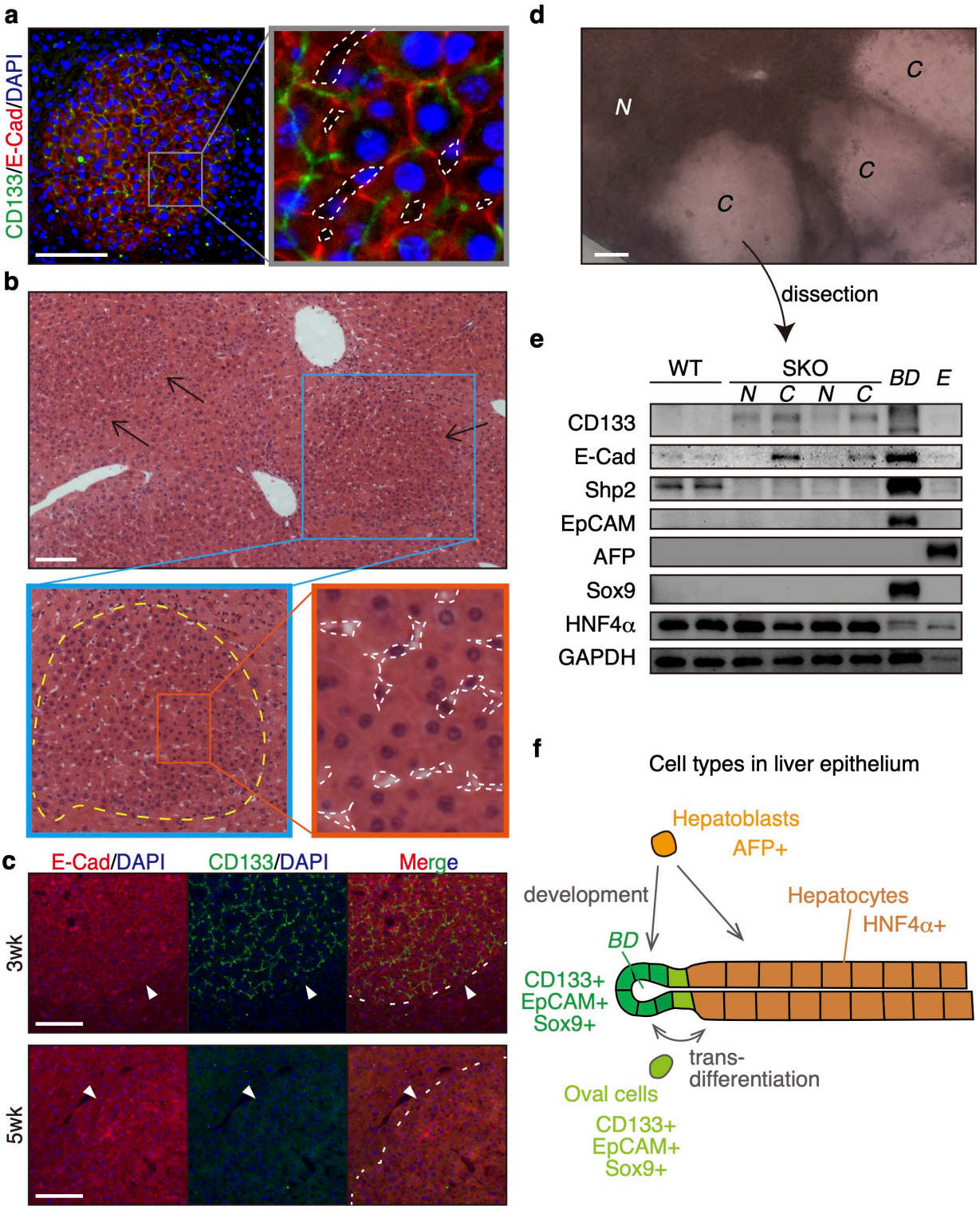


Fig. 2| CD133$^+$ colonies represent a unique regeneration process in Shp2-deficient liver

a, Representative image of CD133$^+$ colonies shown by immunofluorescence on SKO liver tissues 3 weeks after PHx. Dashed lines, vasculatures.

b, H&E staining of SKO liver sections 3 weeks after PHx. Yellow dashed line: boundary

between the colony and surrounding tissue; white dashed lines: vasculatures.

c, Immunofluorescence on SKO liver sections at indicated time points after PHx. Arrowheads and dashed lines indicate the boundaries between the colonies and surrounding tissues, which was clear at 3 weeks but disappeared at 5 weeks.

d, Unstained thick tissue section of SKO liver 3 weeks after PHx.

e, Colonies (*C*) and non-colony control areas (*N*) were dissected out from thick tissue sections of SKO livers 3 weeks after PHx, and analyzed by immunoblotting. Random areas from WT livers 3 weeks after PHx, bile duct (*BD*) and E17.5 liver (*E*) were used for comparison.

f, Illustration of epithelial lineages in the liver.

Scale bars, 100 μm (**a**-**d**).

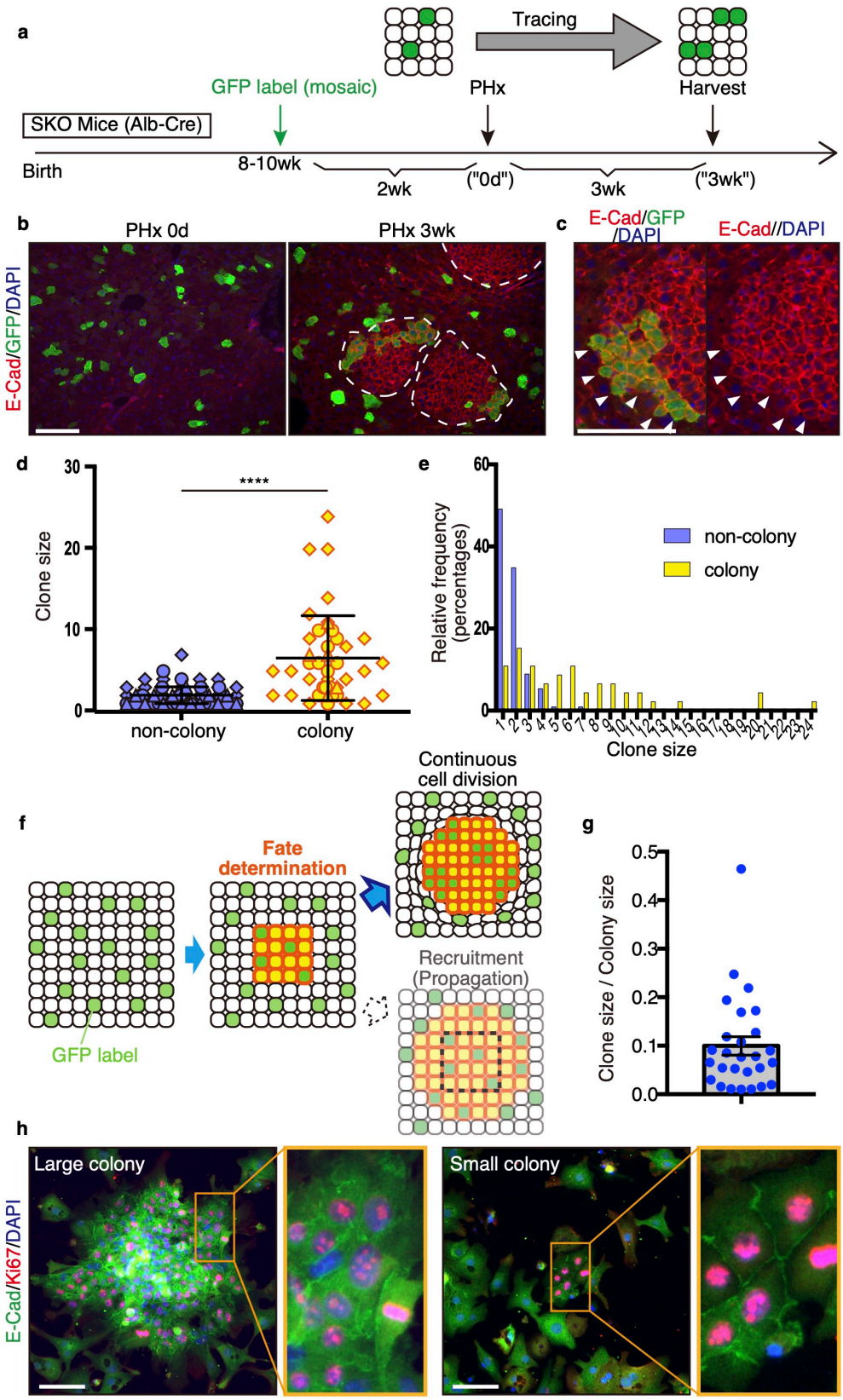


Fig. 3| Mitogenic signal deficiency induced tight grouping of cells that leads to continuous growth

a, Schematic illustration of clonal tracing performed in SKO livers during regeneration.

b, Liver tissues were examined before (PHx 0d) or 3 weeks after the surgery (PHx 3wk). E-Cad shows the colony structures (arrows).

c, GFP-labeled clone in a colony 3 weeks after PHx. Note that no unlabeled cells were detected on this edge (arrowheads).

d, Cell numbers of GFP-labeled hepatocyte clones in colony and non-colony areas, counted on sections. Data were collected 3 weeks after PHx. Means ± SD are shown, n = 3, ****$P < 0.0001$ (Mann-Whitney test).

e, Distribution analysis of clone sizes from **c**.

f, Clonal dynamics of colony-forming hepatocytes.

g, Clone/colony size ratios to estimate the original colony size (see also Extended Data Fig. 2b). Mean ± SEM is shown.

h, Immunofluorescence on the colonies (E-Cad^{+} clusters) in Shp2-deficient hepatocyte culture in vitro. Note the other non-colony cells with low proliferation rate, indicating the patchy proliferation in the colonies. See also Extended Data Fig. 2d.

Scale bars, 100 μm (**b**, **c**, **h**).

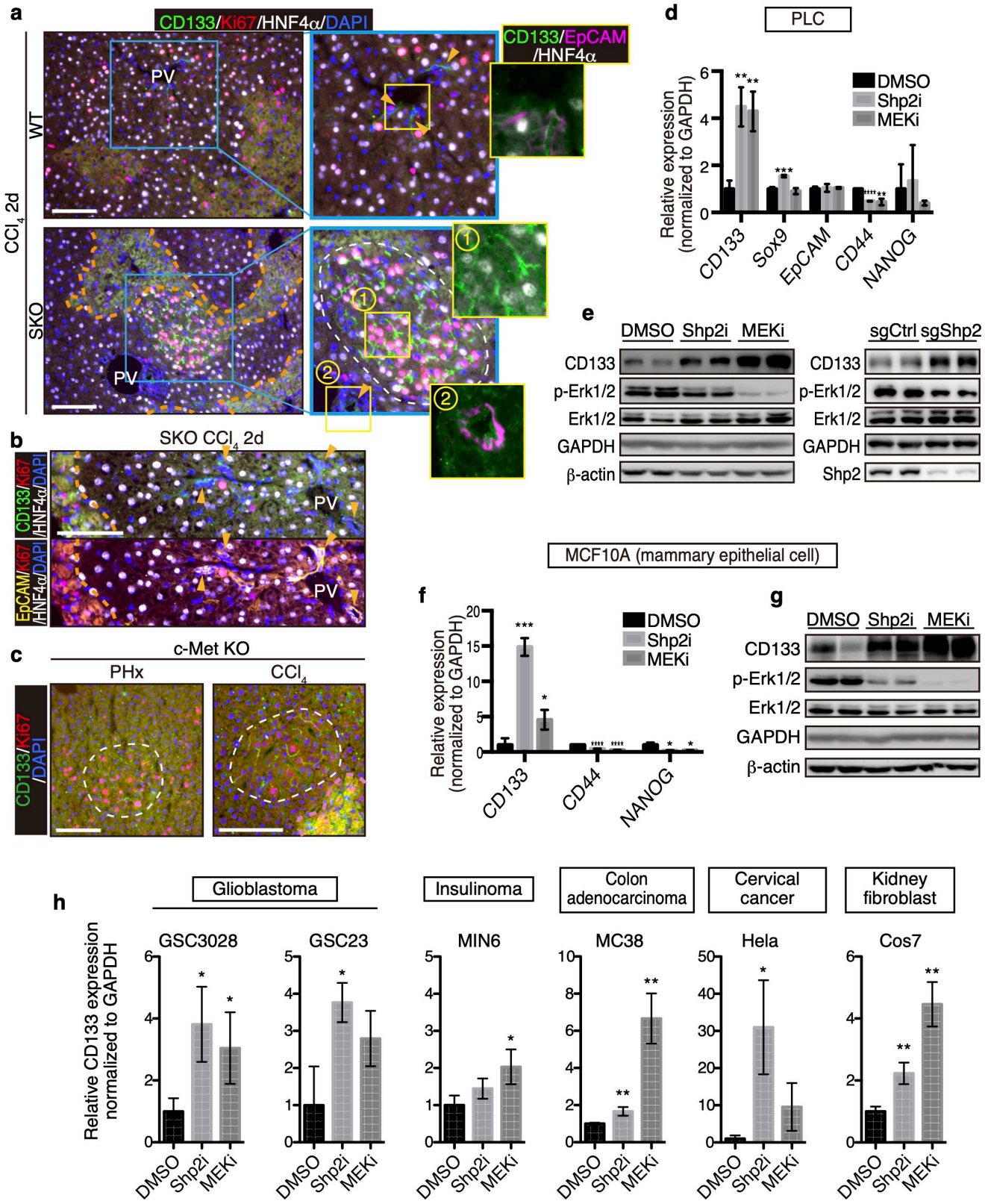


Fig. 4| The CD133 induction during signal deficiency is a widely conserved mechanism

a, **b**, Immunofluorescence of liver sections 2 days after CCl_4 injection. Proliferating hepatocytes were scattered in WT livers, whereas they were highly concentrated in $CD133^+$

colonies (white dashed line) in SKO livers as shown in **a**. $CD133^+/EpCAM^+/HNF4\alpha^-$ bile duct epithelial cells (arrowheads) were not associated with the colonies. Pink dashed lines, injured areas. PV, portal vein.

c, Immunofluorescence of $Met^{hep-/-}$ liver sections 2 days after PHx or CCl4 injection. White dashed lines, $CD133^+$ colonies. Pink dashed line, injured area.

d, qRT-PCR analysis of PLC cell lysates treated with Shp2 or MEK inhibitors (Shp2i and MEKi). **P < 0.01, ***P < 0.001, ****P < 0.0001 (two-tailed unpaired t test, each compared with DMSO treatment). Means ± SD from 3 replicates are shown.

e, Immunoblotting of PLC cell lysates treated with inhibitors or transfected with Shp2 targeting CRISPR vector. Guide RNA targeting the AAVS1 safe harbor site was used as a control (sgCtrl).

f, qRT-PCR analysis of MCF10A cell lysates treated with the inhibitors. *P < 0.05, ***P < 0.001, ****P < 0.0001 (two-tailed unpaired t test, each compared with DMSO treatment). Means ± SD from 3 replicates are shown.

g, Immunoblotting of MCF10A cell lysates treated with inhibitors.

h, qRT-PCR analysis of lysates from various cell lines treated with Shp2 or MEK inhibitors. *P < 0.05, ***P < 0.001, ****P < 0.0001 (two-tailed unpaired t test, each compared with DMSO treatment). Means ± SD from 3 replicates are shown.

Scale bars, 100 μm (**a**-**c**).

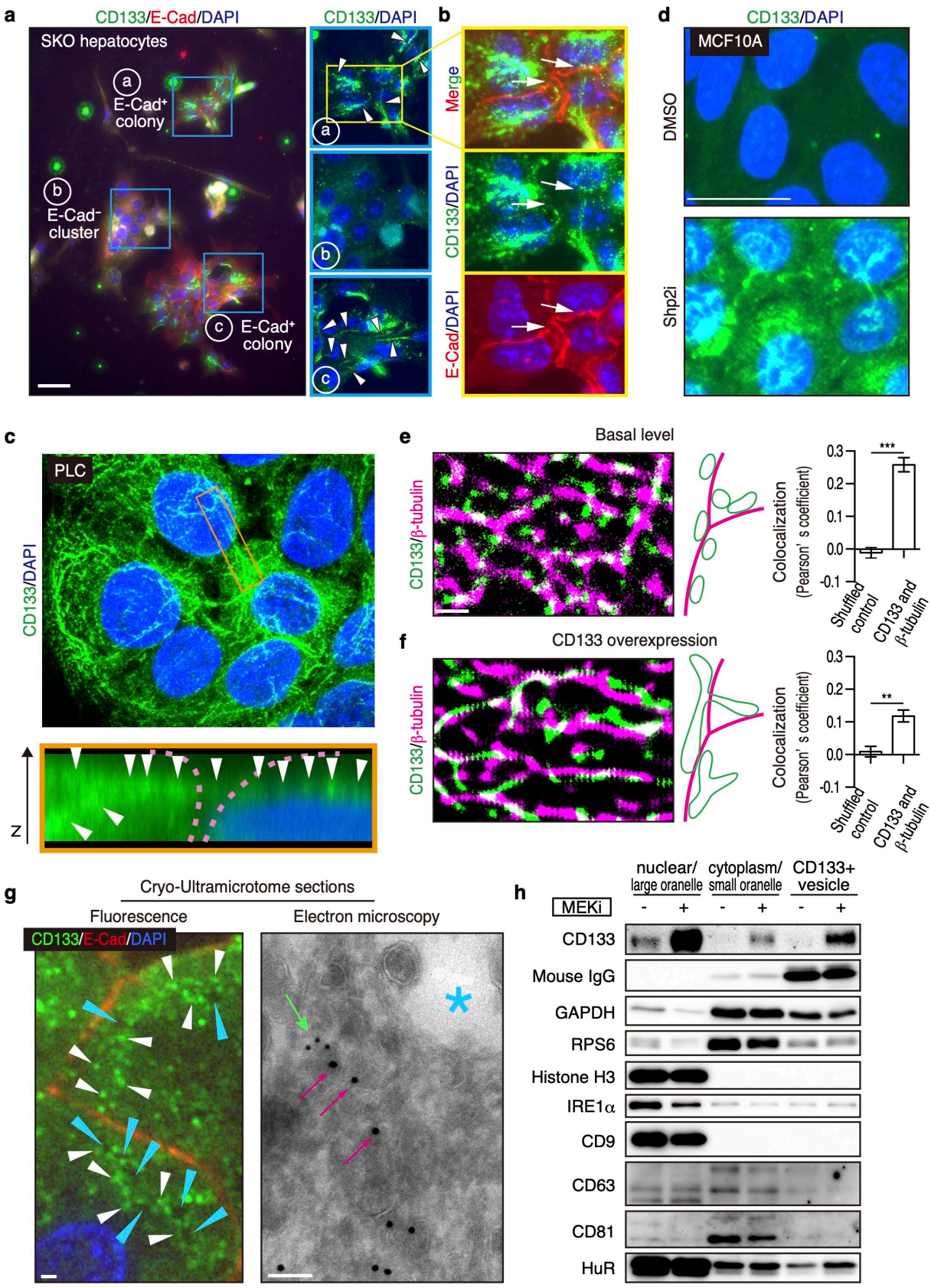


Fig. 5| Mitogenic signal-deficient cells induce CD133^{+} intercellsome

a, **b**, Immunofluorescence on in vitro colonies of primary hepatocytes from SKO liver.

CD133 was localized at filament-like structures in E-Cad$^+$ colonies as shown by arrowheads in **a**, which were connected between different hepatocytes as shown by arrows in **b**.

c, 3D-reconstituted confocal image of immunofluorescence on PLC cells. Lower panel shows the Z-plane section of the orange box area. Arrowheads indicate the CD133 signal on continuous filament like structures bridged between neighboring cells. Pink dashed lines indicate the cell surface.

d, Immunofluorescence of MCF10A cells treated with Shp2 inhibitor.

e, **f**, Super-resolution STORM images of immunofluorescence on PLC cells without (**e**) or with (**f**) CD133 overexpression. Colocalization of CD133 and β-tubulin was analyzed as Pearson's coefficient. Mismatched green and magenta channels from shuffled ROIs were measured as controls (Extended Data Fig. 4d). Means ± SEM from 6 images are shown. **P < 0.01, ***P < 0.001 (two-tailed unpaired t test).

g, Immunofluorescence and Immuno-Gold EM images of cryo-ultramicrotome sections of SKO liver tissue after PHx. Cyan arrowheads and asterisk indicate apical lumens. White arrowheads indicate the CD133 signals aligned between apical lumens of neighboring cells. Light green arrow, CD133 staining (12 nm colloidal gold); Magenta arrows, α-tubulin staining (18 nm colloidal gold).

h, Immunoblotting of CD133$^+$ vesicles isolated from MEK inhibitor (MEKi) -treated PLC cells. Markers for different fractions were analyzed. CD133 antibody used for the vesicle isolation was IgG produced in mouse, which was detected by anti-mouse IgG antibody, showing efficient capture by the beads. Despite the efficient capture, the DMSO-treated PLC cells did not have much CD133+ vesicle to be bound with the antibody.

Scale bars, 100 μm (**a**), 25 μm (**d**), 1 μm (**e**, **f**, Fluorescence in **g**), 100 nm (EM in **g**).

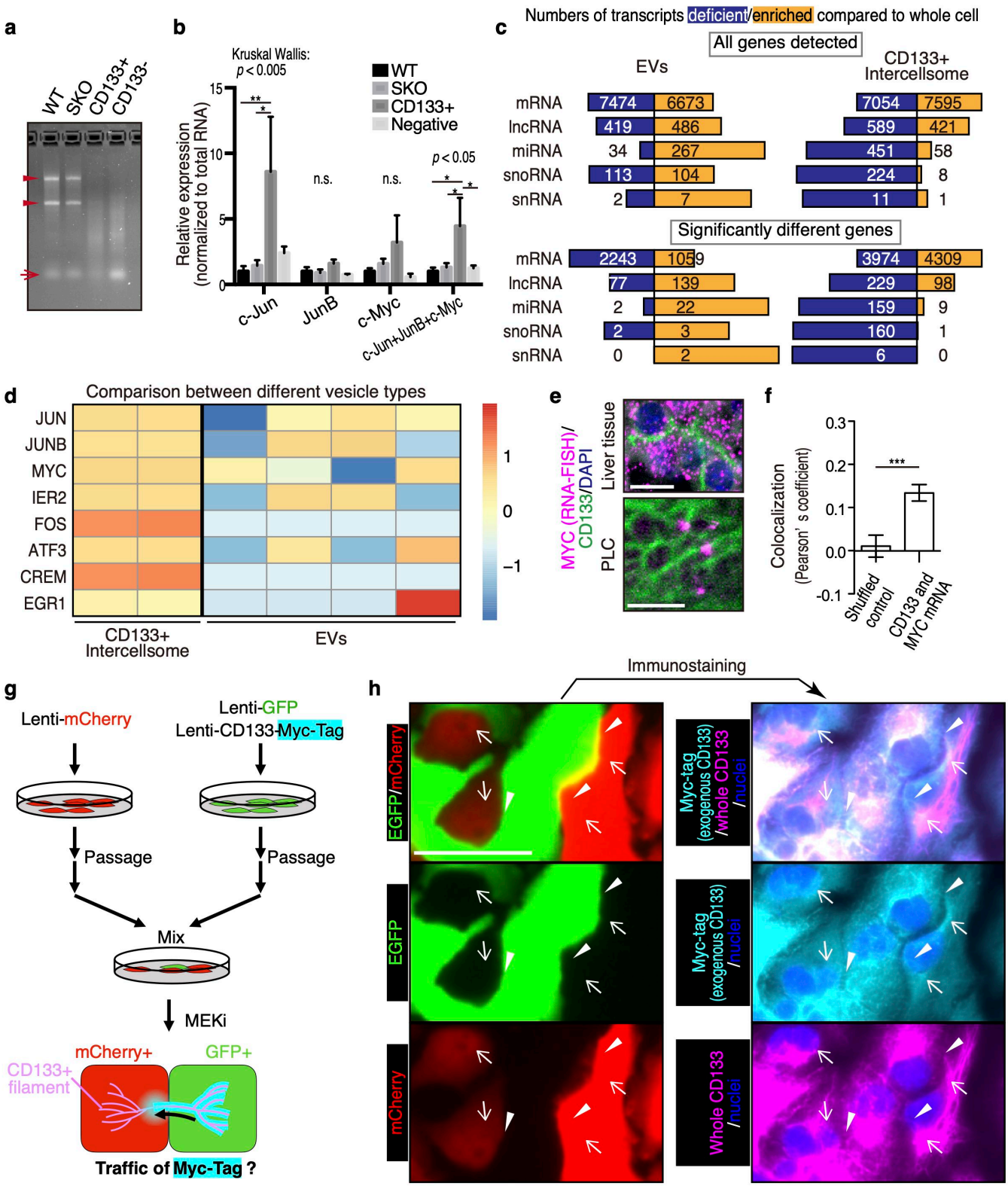


Fig. 6| CD133^{+} intercellsome contains mitogenic mRNAs and traffics between neighbor cells

a, Agarose gel electrophoresis of total RNAs extracted from WT and SKO livers and CD133-

positive vesicle and negative fractions from SKO liver after PHx. Arrowheads show rRNAs and arrow shows microRNAs.

b, qRT-PCR analysis of RNAs extracted form WT (3 mice) and SKO tissues (3 mice) and from $CD133^{+}$ vesicles and $CD133^{-}$ fractions (4 mice for both). Means ± SEM are shown. *P < 0.05, **P < 0.01 (uncorrected Dunn's multiple comparison test, performed after Kruscal-Wallis test). n.s., not significant.

c, RNA-seq analysis of the different cell types and the whole cells. The bars indicate proportions between numbers of deficient and enriched gene transcripts in each RNA types.

d, Comparison of IEG contents between the different vesicle types with RNA-seq.

e, RNA-FISH for MYC mRNA and immunostaining for CD133.

f, Quantitative colocalization analysis of CD133 and MYC mRNA in PLC cells shown in **e**. Means ± SEM from 14 images are shown.

g, Experimental design to detect the traffic of CD133+ intercellsome between neighbor cells.

h, Immunostaining of Myc-tag and CD133 on GFP+ and mCherry+ PLC cells mixed as shown in **g**. Note that Myc-tag only indicates exogenous CD133, while CD133 indicates both endogenous and exogenous CD133. Myc-tag was primarily detected in the GFP+ cells, but also detected on the bridges (arrowheads) and in the mCherry+ cells (arrows). GFP was not detected at the same locations (arrowheads and arrows), indicating the specific traffic of CD133-Myc-tag.

Scale bars, 10 μm (e) and 50 μm (**h**)

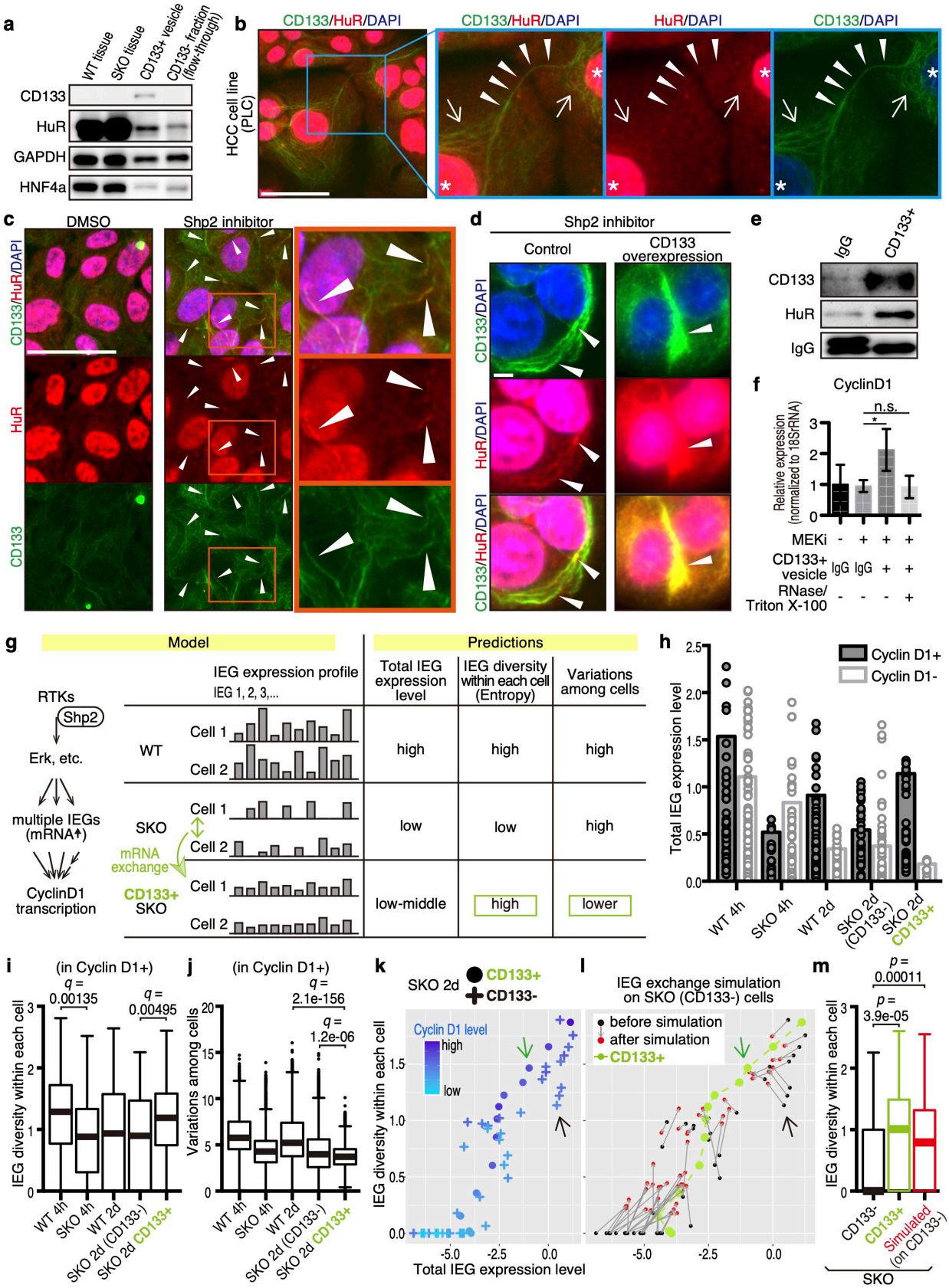


Fig. 7| CD133-mediated mRNA sharing converts intercellular heterogeneity into intracellular diversity of IEGs

a, Immunoblotting of HuR in the CD133$^+$ vesicles from SKO liver after PHx. GAPDH and HNF4α were used as controls for cytoplasmic and nuclear fractions, respectively.
b, Immunofluorescence on PLC cells. Arrows, peri-nuclear areas enriched with HuR in the cytoplasm; Arrowheads show colocalization of HuR on the CD133$^+$ filament bridging two cells.
c, Immunofluorescence on PLC cells treated with the Shp2 inhibitor (SHP099). Arrowheads show strong localization of HuR on the CD133$^+$ filaments.
d, Immunofluorescence on PLC cells transfected with CD133 expression vector treated with a Shp2 inhibitor (SHP099). Arrowheads show strong localization of HuR on the CD133$^+$ filaments.
e, Immunoblotting of CD133$^+$ vesicles isolated from MEK inhibitor-treated PLC cells.
f, qRT-PCR analysis of PLC cell lysates after treatment with CD133$^+$ vesicles isolated from MEK inhibitor-treated PLC cells. RNase and Triton X-100 were used to digest the RNA content of the vesicles. *$P < 0.05$ (two-tailed unpaired t test). n.s., not statistically significant. Means ± SD from 3 replicates are shown.
g, A model and predictions for single cell RNA-seq data analysis.
h, Total IEG expression levels. Cyclin D1-positive and -negative cells were analyzed separately to evaluate the influence of cell cycle on the IEG analysis. See Methods section for what the bars and dots represent.
i, **j**, Box plots (Tukey's) of IEG diversity within each cell (calculated as entropy) and IEG variations among cells. Analyses were focused on cyclin D1$^+$ cells in all groups for fair comparison.
k, **l**, Plot of intracellular IEG diversity against total IEG expression levels in SKO hepatocytes 2 days after PHx. Blue color gradient indicates cyclin D1 expression levels. For the simulation in **l**, the parameters used were: Group of 5 cells, X=1/12, model 3 (see also Extended Data Fig. 6e and Methods). Green and Black arrows show typical profiles of CD133-positive and -negative cells, respectively.
m, Box plot (Tukey's) of intracellular IEG diversity after simulation. The analysis was not limited to cyclin D1-positive or -negative cells. Note the simulation of the IEG exchange attracted the cells from cyclin D1-low profile to cyclin D1-high profile (**k**-**m**). Statistics was performed by Wilcoxon rank sum test adjusted by FDR in **i**, **j** and **m**.

Scale bars, 50 μm (**b**, **c**), 5 μm (**d**).

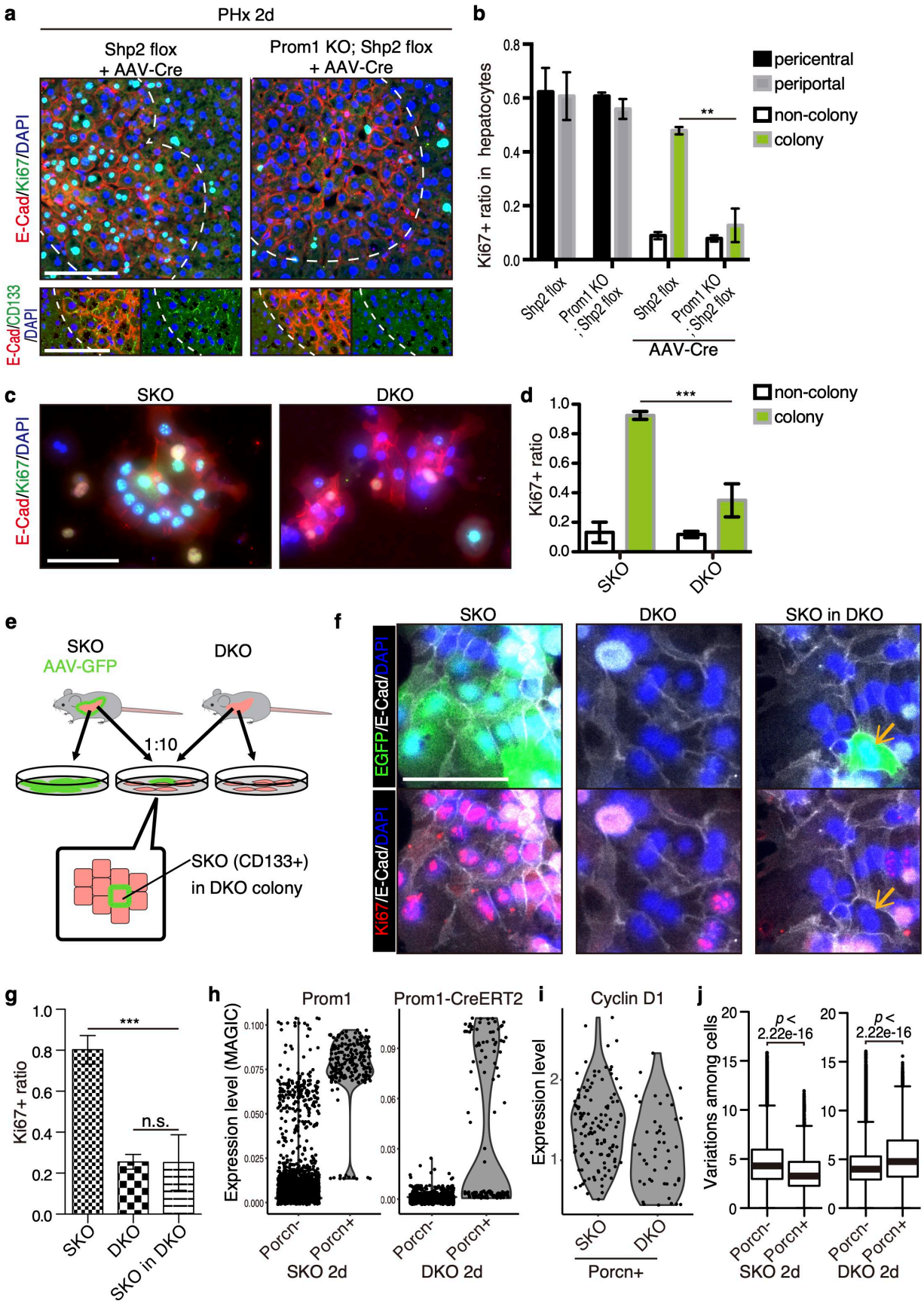


Fig. 8| CD133 is required for cell proliferation under proliferative signal deficit

a, **b**, Immunofluorescence (**a**) on liver sections of WT and Prom1 KO mice 2 days after PHx with or without Shp2 deletion using AAV-Cre and quantification of $Ki67^{+}$ ratio in hepatocytes in the indicated genotypes (**b**). Separate analyses of pericentral and periportal hepatocytes showed insignificance of zonal difference. **P < 0.01, (two-tailed unpaired t test). Means ± SEM are shown. n = 3, 3, 4 and 5 mice, respectively.

c, **d**, Immunofluorescence (**c**) on primary hepatocytes isolated from SKO and Shp2/Prom1 double KO (DKO) mouse livers and quantification of Ki67+ ratio (**d**). Images of representative colonies are shown. ***P < 0.001, (two-tailed unpaired t test). Means ± SD from 4 wells are shown.

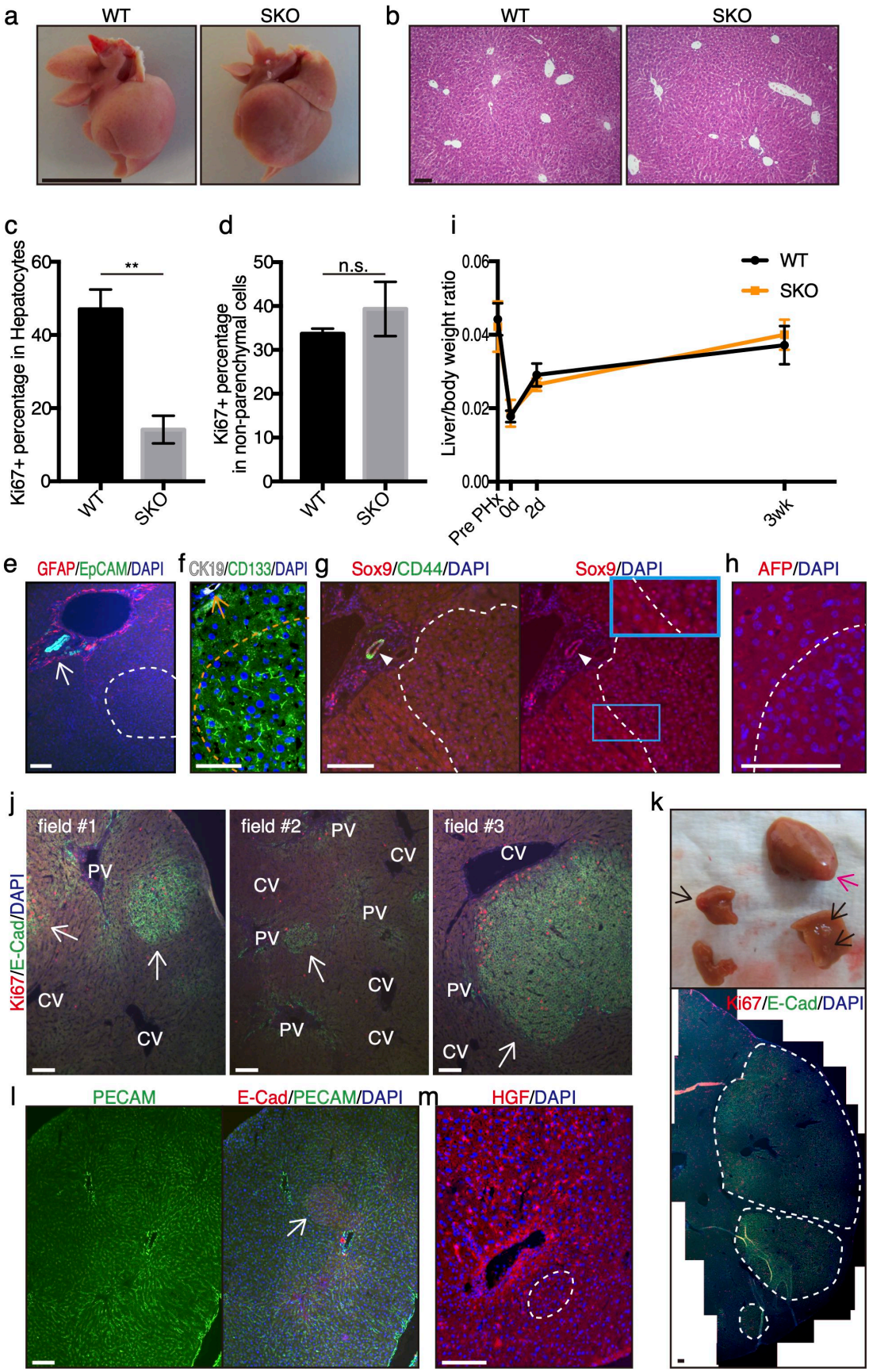
e, Experimental design with primary hepatocytes isolated from GFP labeled SKO liver and unlabeled DKO liver. E-Cad+ colonies were analyzed.

f, Immunofluorescence images of E-Cad+ colonies in SKO, DKO and mixed culture as shown in **e**. The arrows show a GFP+ SKO cell forming part of the colony with the surrounding DKO cells.

g, Quantification of the Ki67 ratio in E-Cad+ colony-forming cells shown in **f**. ***P < 0.001 (two-tailed unpaired t test). n.s., not statistically significant. Means ± SD from 3 wells are shown.

h-**j**, scRNA-seq analyses of DKO liver compared to SKO liver 2 days after PHx. Porcn was used as a marker to represent the Prom1 promoter active hepatocyte populations in both SKO and DKO (**h**). A proliferation marker CyclinD1 was reduced in Porcn+ DKO hepatocytes compared to Porcn+ SKO hepatocytes (**i**). The IEG variations among cells were analyzed (**j**). See also a model in Fig. 7g, regarding the predicted relationship between the intercellsome exchange and IEG variations among cells.

Scale bars, 100 μm (**a**, **c**, **f**).



Extended Data Fig. 1| Liver regeneration of WT and SKO livers after PHx

a, General morphology of livers 2 days after PHx.

b, H&E staining of liver sections 2 days after PHx.

c, **d**, Ki67^{+} percentages in HNF4α^{+} hepatocytes (**c**) and HNF4α^{-} NPCs (**d**) 2 days after PHx. Means ± SEM are shown, n=3 per group, **P < 0.01 (two-tailed unpaired t test). n.s., not significant.

e, EpCAM was highly expressed in biliary epithelial cells (arrow), but not in the colony (dashed line). Stellate cells (GFAP) were not altered around the colony.

f, CK19 was highly expressed in biliary epithelial cells (arrow), but not in the colony (dashed line). CD133 was highly expressed in the colony.

g, Sox9 and CD44 were highly expressed in biliary epithelial cells (arrowhead), but was not upregulated in the colony (dashed line) compared to surrounding hepatocytes.

h, AFP showed no difference between the colony (dashed line) and the surrounding tissue.

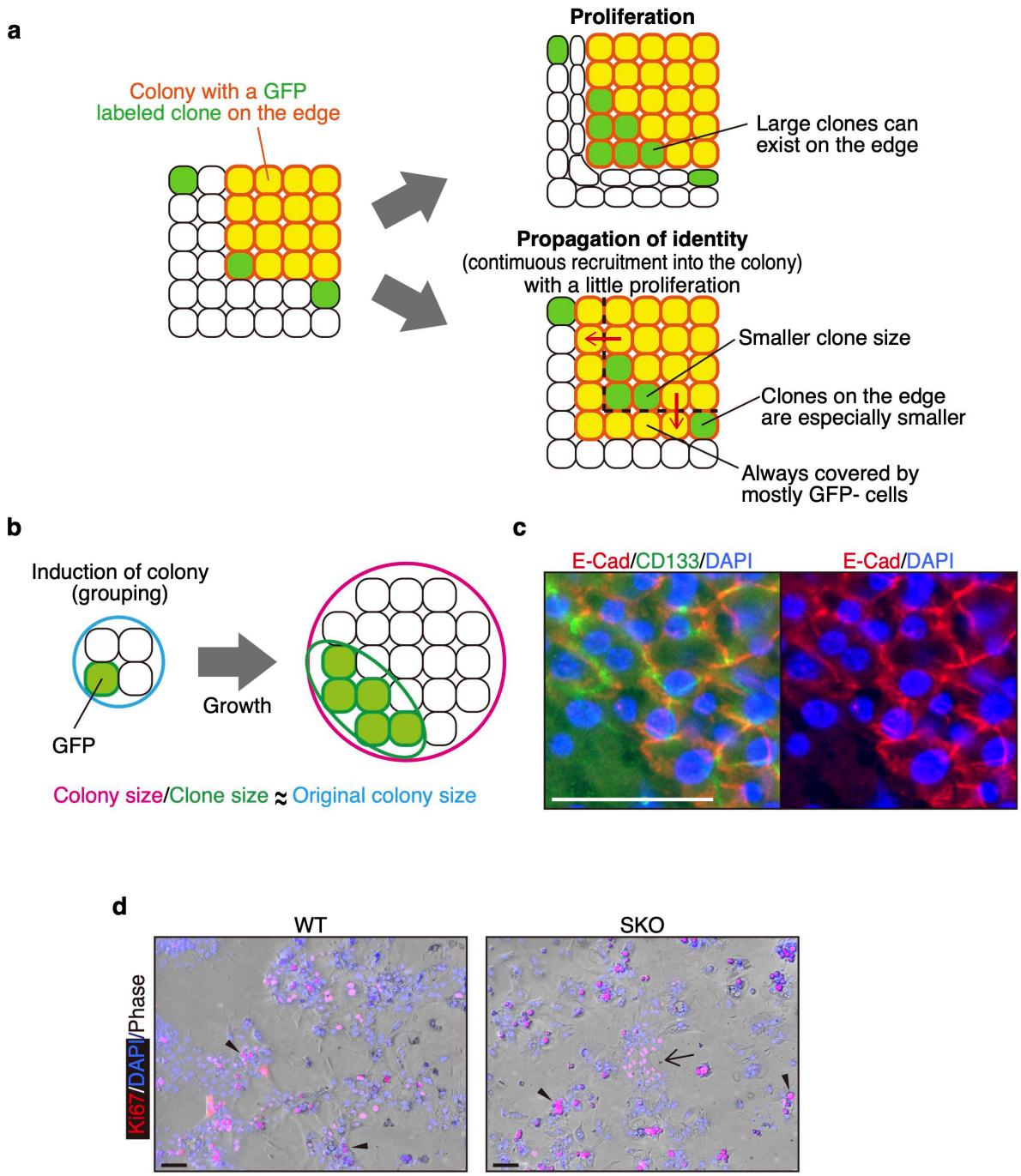
i, Liver/body weight ratios after PHx. Means ± SEM from 3 or more mice analyzed for each time point and group are shown.

j, **k**, Immunofluorescent staining of SKO liver sections 3 weeks after PHx. Colonies were at various sizes and locations as shown by arrows in **j**. Macroscopic colonies were found as shown by arrows and dashed lines in **k**. Immunofluorescent image in **k** corresponds to the magenta arrow in the liver image. PV, portal vein; CV, central vein.

l, Vasculature shown by PECAM did not distinguish the colony (arrow).

m, HGF expressed by NPCs was not concentrated in the colony (dashed line).

Scale bars, 1 cm (**a**) 100 μm (**b**, **e-g**, **i-l**).



Extended Data Fig. 2| Clonal analysis of hepatocytes

a, Predicted outcomes of different dynamics.

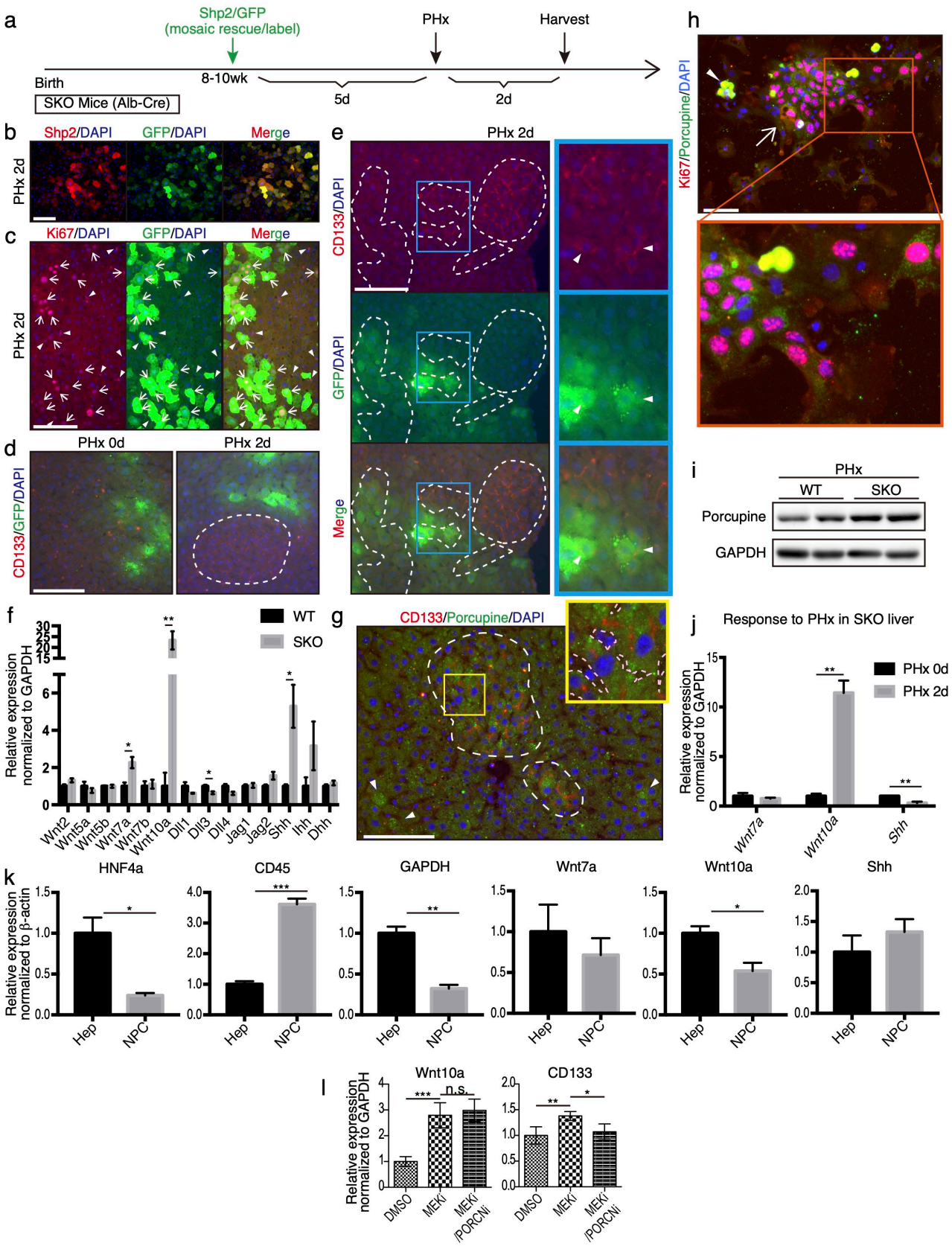
b, Colony size/clone size ratios can be used to roughly estimate the original colony sizes.

c, Immunofluorescence of SKO liver sections 2 days after PHx, showing mixture of binuclear and mononuclear hepatocytes in the colony.

d, Immunofluorescence of primary hepatocyte cultures from WT and SKO mice. The arrow indicates an in vitro colony in Shp2-deficient hepatocyte culture. Arrowheads, dying cells

with high autofluorescence.

Scale bars, 50 μm (**c**) and 100 μm (**d**).



Extended Data Fig. 3| Molecular analysis of intercellular signals during colony induction

a, Schematic illustration of mosaic rescue experiment of Shp2 in Shp2-deficient liver.

b, Immunofluorescence showing Shp2-rescued hepatocytes also labeled with GFP.

c, GFP^+ hepatocytes were proliferating ($Ki67^+$; arrows), while other proliferating cells were not hepatocytes (arrowheads). Areas with colonies were not included here.

d, Colonies were frequently constituted by GFP-negative hepatocytes (dashed line). Orange dots: non-specific stains of debris caused by HTVi.

e, Some GFP^+ hepatocytes showed expression of CD133 (arrowheads). Dashed lines show CD133-positive areas.

f, qRT-PCR analysis of WT and SKO liver lysates 2 days after PHx. n=3, means ± SEM are shown, *P < 0.05, **P < 0.01 (two-tailed unpaired t test).

g, Immunofluorescence on SKO liver section 2 days after PHx. White dashed lines mark $Porcupine^+$ cells enriched in $CD133^+$ colonies, rarely detected in non-colony area (arrowheads). Pink dashed lines: vasculature.

h, Immunofluorescence on the colonies (arrows) in Shp2-deficient hepatocyte culture in vitro. Arrowhead, dying cells with high autofluorescence.

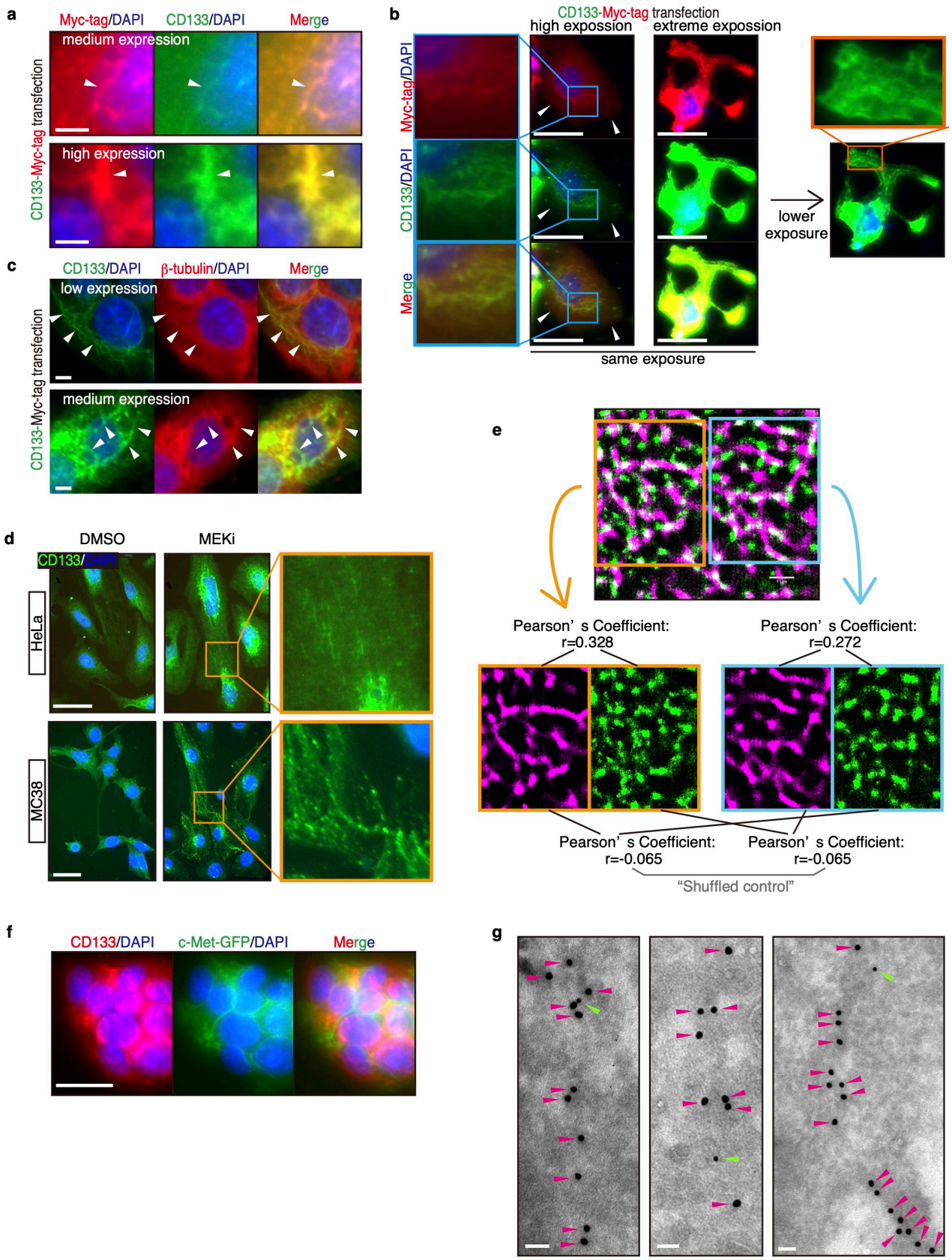
i, Immunoblot analysis of WT and SKO liver lysates 2 days after PHx.

j, qRT-PCR analysis of SKO liver tissue lysates without PHx (0d, 3 mice) and with PHx (2d, 3 mice). Means ± SEM are shown. **P < 0.01 (two-tailed unpaired t test).

k, qRT-PCR analysis of hepatocyte and NPC fractions from SKO livers (3 mice) 2 days after PHx. HNF4α and CD45 were used as positive controls for the fractionation. β-actin was used for normalization, because GAPDH was highly expressed by hepatocytes. Means ± SEM are shown. *P < 0.05, **P < 0.01, ***P < 0.001 (two-tailed unpaired t test).

l, qPCR analyses of PLC cells treated with Mek inhibitor and Porcn inhibitor. Means ± SD from 4 wells are shown. *P < 0.05, **P < 0.01, (two-tailed unpaired t test). n.s., not significant.

Scale bars, 100 μm (**b**-**e**, **g**, **h**).



Extended Data Fig. 4| Analyses of CD133 localization

a-c, Immunofluorescence on PLC cells transfected with CD133-Myc-tag fusion protein

expression construct. Cells with different expression levels are shown as representatives. In **a** and **c**, exposure times for the CD133 signals were adjusted separately to clearly demonstrate the patterns, rather than intensities. Note that cells with higher CD133 expression displayed bulky pattern of the filaments, with maintained colocalization with tubulin filaments. Arrowheads, $CD133^{+}$ filaments. As shown in **b**, the overexpressed CD133 primarily localize to the filaments, without detectable membrane localization (arrowheads). With extreme overexpression, CD133 can also localize to the cellular membrane surface, altering the morphology of the surface (right panels).

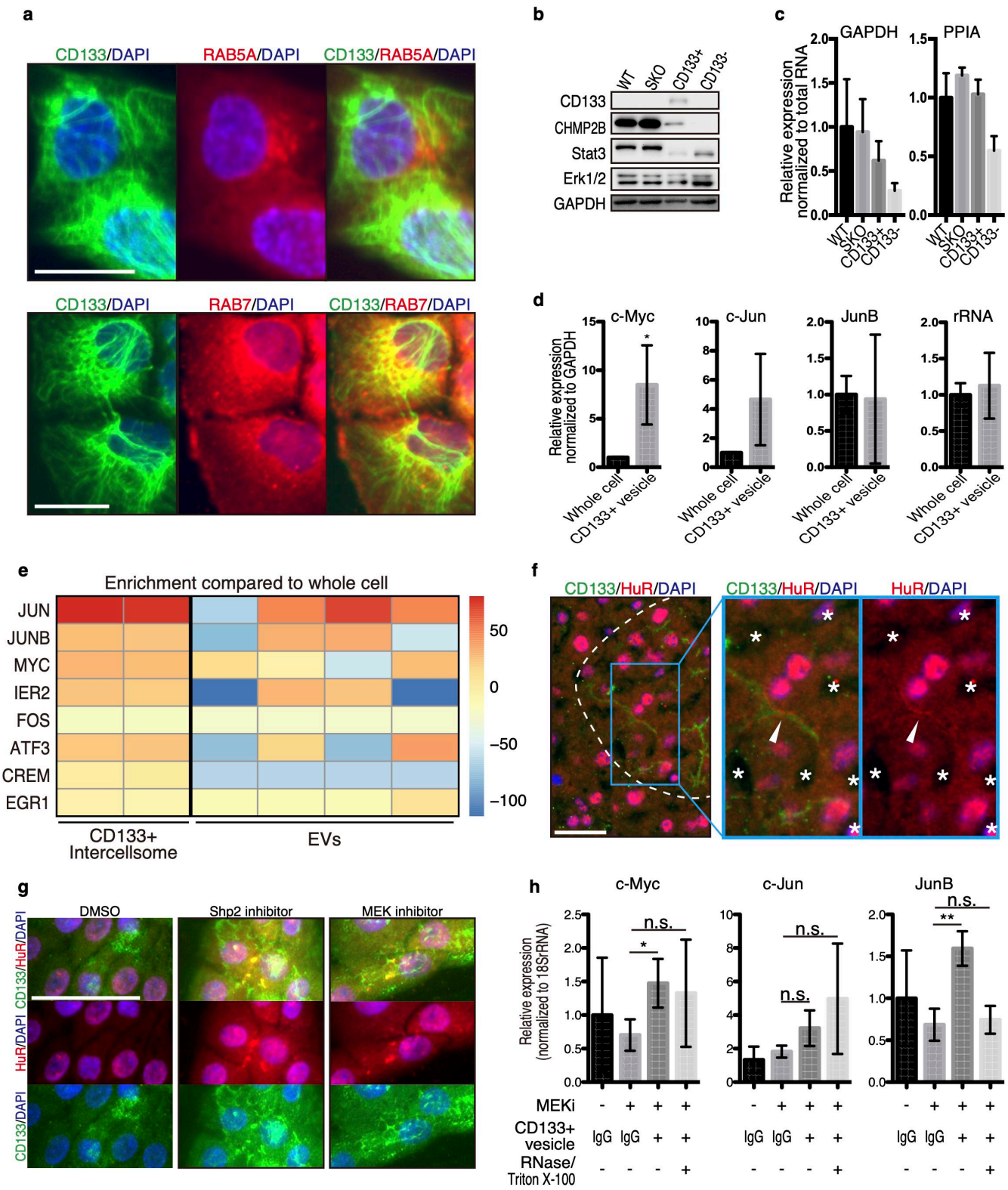
d, Immunofluorescence on HeLa and MC38 cells.

e, Colocalization of CD133 and β-tubulin was analyzed as Pearson's coefficient. Mismatched green and magenta channels from shuffled ROIs were measured as controls.

f, Immunofluorescence on PLC cells transfected with cMet-GFP. cMet, an HGF receptor, shows the cell surface. CD133 did not colocalize with cMet-GFP on the cell surface, but instead localized to the filaments.

g, Immuno-Gold EM images of cryo-ultramicrotome sections of SKO liver tissue after PHx. Light green arrowheads, CD133 staining (12nm colloidal gold); Magenta arrowheads, α-tubulin staining (18nm colloidal gold).

Scale bars, 5 μm (**a**, **c**), 50μm (**b**, **d**), 25μm (**f**) and 50nm (**g**).



Extended Data Fig. 5| Analyses of CD133^{+} vesicles

a, Immunofluorescence for endosomal markers on PLC cells.

b, Immunoblotting of CD133+ vesicles isolated from SKO liver after PHx. CHMP2B is a

protein that is generally involved in vesicle formation.

c, qRT-PCR analysis of housekeeping genes in RNAs extracted from WT (3 mice) and SKO tissues (3 mice) and from $CD133^+$ vesicles and CD133-negative fractions (4 mice for both). Means ± SEM are shown.

d, qRT-PCR analysis of RNAs extracted from $CD133^+$ vesicles from PLC cells treated with MEK inhibitor.

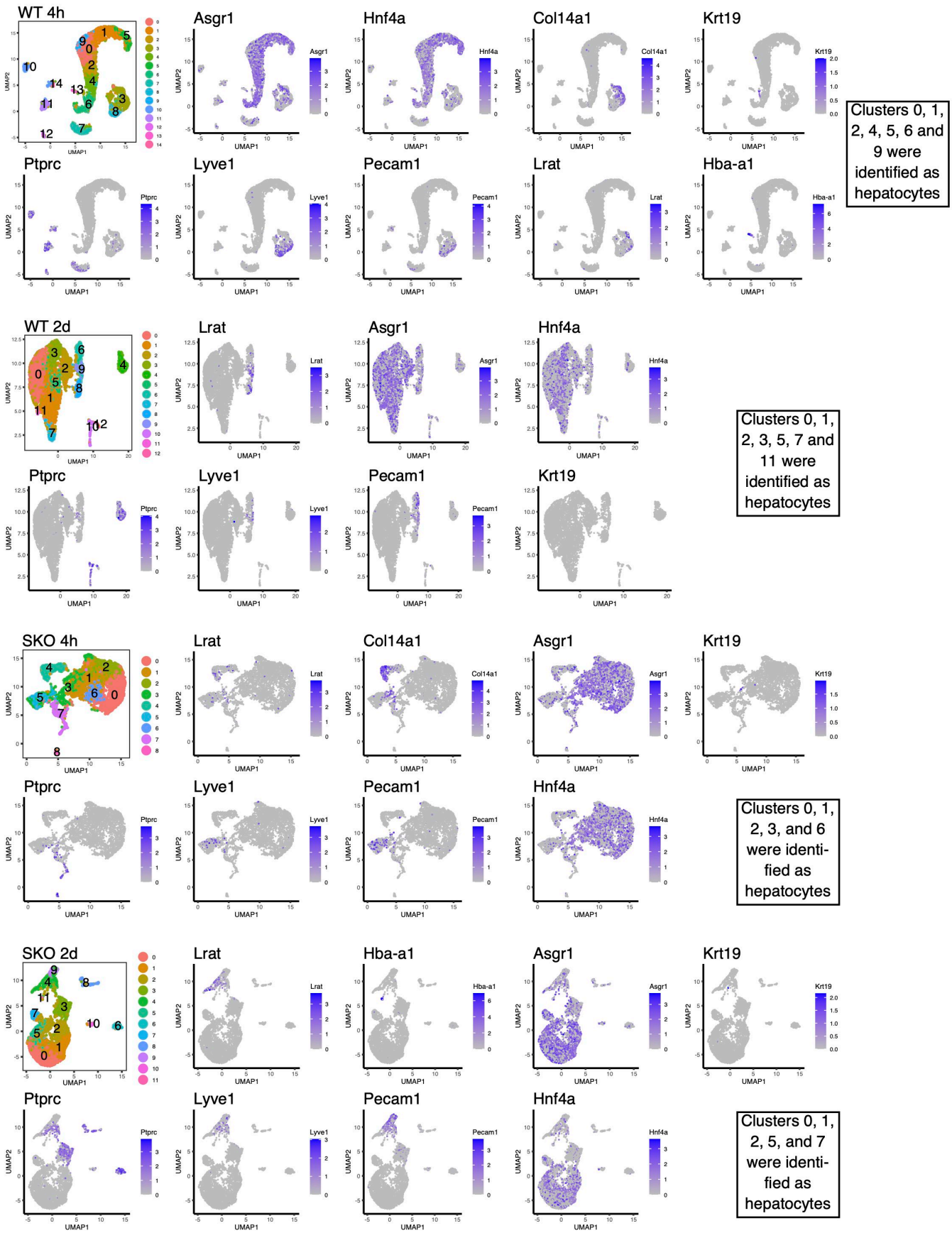
e, RNA-seq analysis showing enrichment scores of IEG contents in the two vesicle types compared to the whole cells.

f, Immunofluorescence on SKO liver 2 days after PHx. Asterisks indicates vasculature.

g, Immunofluorescence on MCF10A cells treated with inhibitors.

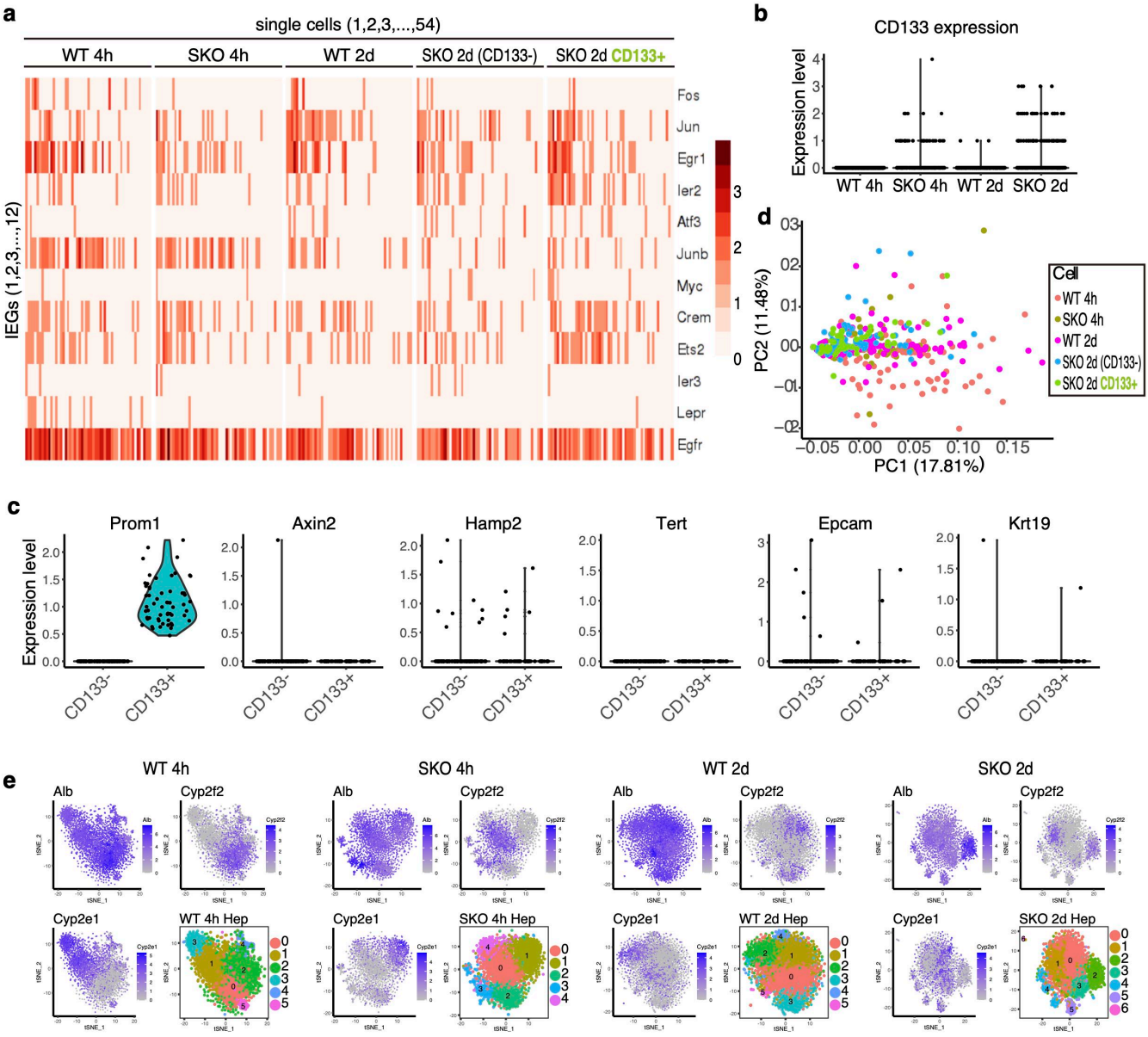
h, qRT-PCR analysis of PLC cell lysates after treatment with $CD133^+$ vesicles isolated from MEK inhibitor-treated PLC cells. RNase and Triton X-100 were used to digest the RNA content of the vesicles. *$P < 0.05$, **$P < 0.01$ (two-tailed unpaired t test). n.s., not statistically significant. Means ± SD from 3 replicates are shown.

Scale bars, 25μm (**a**) and 50μm (**f**, **g**).



Extended Data Fig. 6| Identification of hepatocytes in the scRNA-seq data set

UMAP analysis of the isolated cells from indicated livers. Hepatocyte markers and NPC markers were analyzed after the unsupervised clustering.



Extended Data Fig. 7| Single cell RNA-seq analysis of WT and SKO livers after PHx

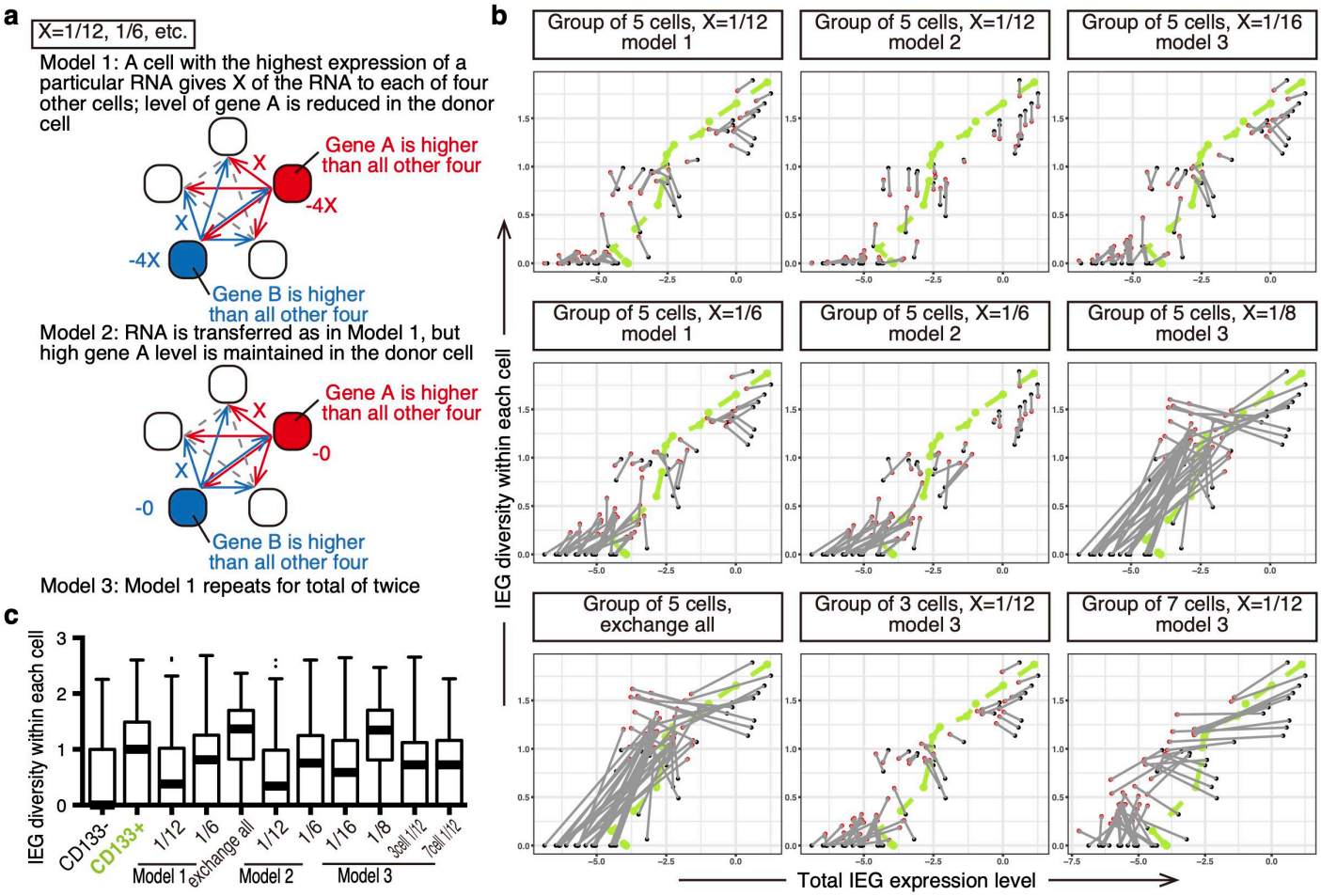
a, Heatmap analysis of single hepatocytes.

b, Detection of CD133 expression in the single cell RNA-seq data.

c, Expression levels of indicated hepatocyte subtype markers and stem cell-like markers. CD133- and CD133+ hepatocytes in the SKO liver 2 days after PHx were compared.

d, Principal component analysis with the IEGs.

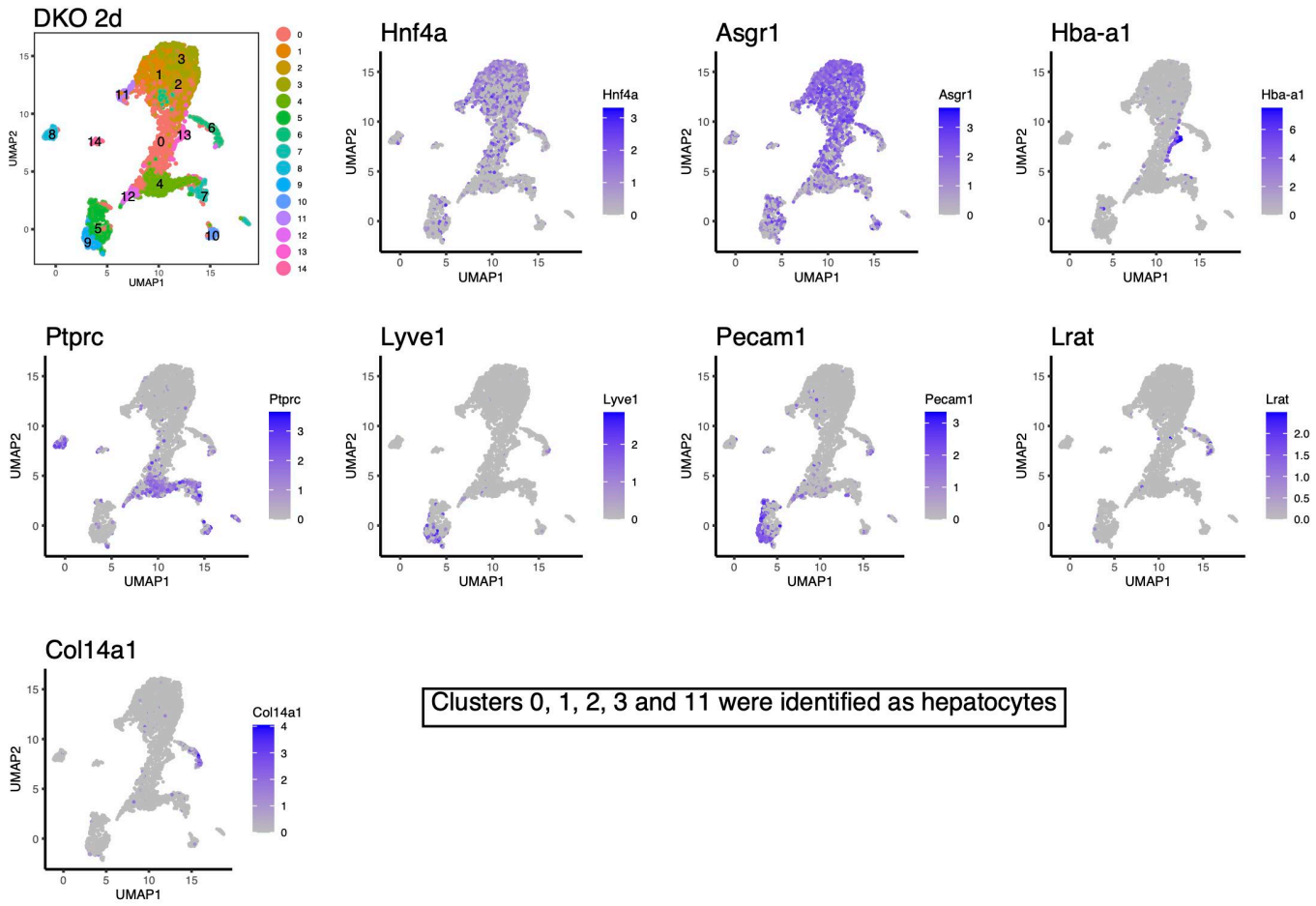
e, tSNE analysis. Clusters likely reflect spatial locations within the tissue.



Extended Data Fig. 8| Simulation analysis of the IEG exchange with scRNA-seq of the regenerating SKO liver

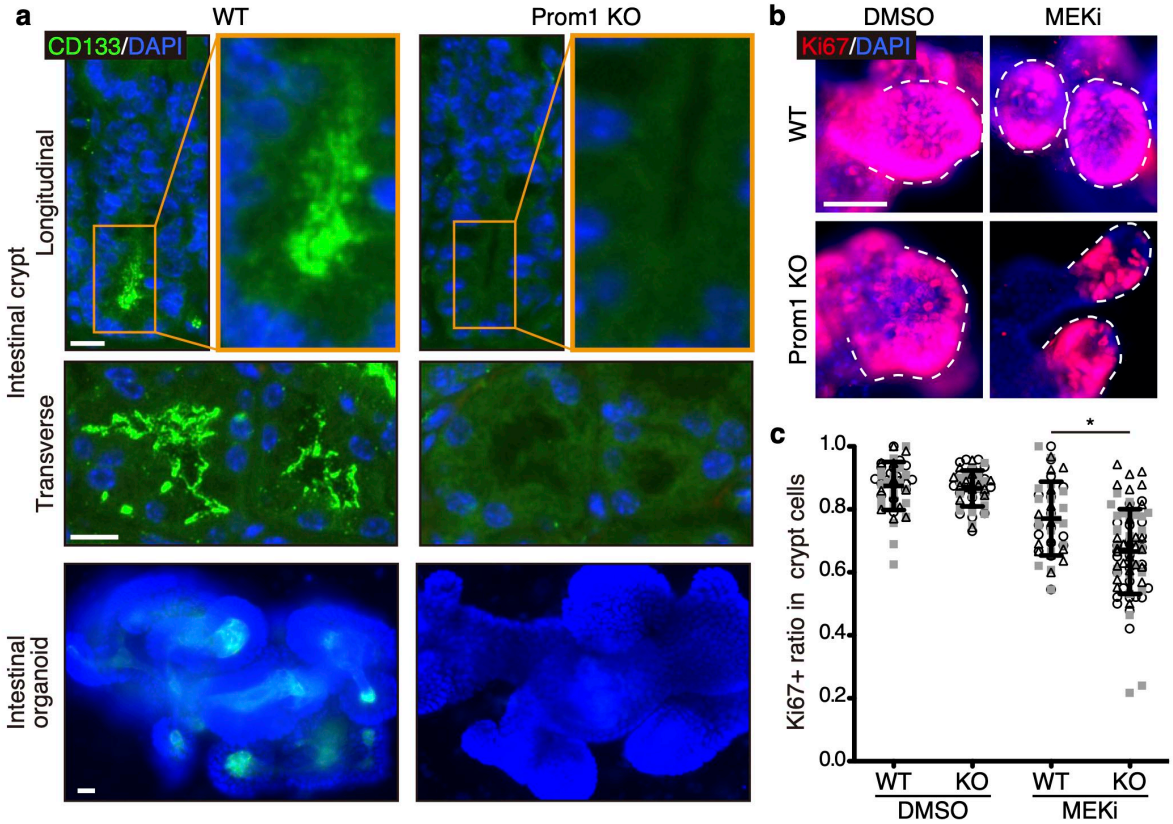
a, Possible models tested for the simulation. Examples with 5 communicating cells are shown. X is a variable that reflects the amount of RNA exchanged between each pair of cells.

b, **c**, Simulation with different RNA exchange model, different variable X and cell numbers.



Extended Data Fig. 9| ScRNA-seq analysis of DKO liver after PHx.

UMAP analysis of the isolated cells to identify hepatocyte populations. Hepatocyte markers and NPC markers were analyzed after the unsupervised clustering.



Extended Data Fig. 10| CD133 is required for proliferation of crypt cells in the intestinal organoids during signal deficit

a, Immunofluorescence on WT and Prom1 KO mouse intestinal tissue sections and intestinal organoids.

b, **c**, Immunofluorescence (**b**) on WT and Prom1 KO mouse intestinal organoids treated with MEK inhibitor (MEKi) and quantification of Ki67+ ratio in the crypt cells (**c**). Dashed lines indicate the crypt buds. *$P < 0.05$, (two-tailed unpaired t test). Means ± SD from 3 wells are shown. Each dot represents each crypt buds, and symbols indicate each well.

Scale bars, 20 μm (**a**) and 50 μm (**b**).