Epigenetic remodelling licences adult cholangiocytes for organoid formation and liver regeneration

Luigi Aloia*[1,2,3], Mikel Alexander McKie*[1,2,3], Grégoire Vernaz*[1,4,5], Lucía Cordero-Espinoza[1,2,3], Niya Aleksieva[6], Jelle van den Ameele[1,2], Francesco Antonica[2], Berta Font-Cunill[1,2,3], Alexander Raven[6], Riccardo Aiese Cigliano[7], German Belenguer[1,8], Richard L. Mort[9], Andrea H. Brand[1,2], Magdalena Zernicka-Goetz[2], Stuart J. Forbes[6], Eric A. Miska[1,4,5] and Meritxell Huch[1,2,3,8 $]

[1]The Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, CB2 1QN, UK.

[2]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, UK.

[3]Wellcome Trust - Medical Research Council Stem Cell Institute, University of Cambridge, Gleeson Building, Tennis Court Road, Cambridge CB2 1QR, UK.

[4]Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

[5]Wellcome Sanger Institute, Hinxton, CB10 1SA, UK

[6]MRC Centre for Regenerative Medicine, University of Edinburgh, 5 Little France Drive, Edinburgh, EH16 4UU, UK

[7]Sequentia Biotech SL, Calle Comte D'Urgell 240, Barcelona, Spain

[8] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany

[9]Division of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Bailrigg, Lancaster, UK

* these authors contributed equally

[$] corresponding author:

Meritxell Huch

m.huch@gurdon.cam.ac.uk;

and

huch@mpi-cbg.de

Abstract

Upon severe or chronic liver injury, adult ductal cells (cholangiocytes) contribute to regeneration by restoring both hepatocytes and cholangiocytes. Recently, we showed that ductal cells clonally expand as self-renewing liver organoids that retain their differentiation capacity into both hepatocytes and ductal cells. However, the molecular mechanisms by which adult ductal-committed cells acquire cellular plasticity, initiate organoids and regenerate the damaged tissue remain largely unknown.

Here, we describe that, during organoid initiation and *in vivo* following tissue damage, ductal cells undergo a transient, genome-wide, remodelling of their transcriptome and epigenome. TET1-mediated hydroxymethylation licences differentiated ductal cells to initiate organoids and activate the regenerative programme through the transcriptional regulation of stem-cell genes and regenerative pathways including the YAP/Hippo.

Our results argue in favour of the remodelling of genomic methylome/hydroxymethylome landscapes as a general mechanism by which differentiated cells exit a committed state in response to tissue damage.

The adult liver exhibits low physiological turnover, however it has an efficient regenerative ability following damage. Upon tissue injury, if hepatocyte proliferation is compromised, resident, lineage-restricted ductal cells (cholangiocytes) acquire cellular plasticity to regenerate both, cholangiocytes and hepatocytes[1-9]. Similarly, *in vitro*, ductal cells grown as clonal organoids become bi-potential, express stem/progenitor markers, including *Lgr5*[4,10,11], *Foxl1*[7] and *Trop2*[12], and differentiate into both ductal and hepatocyte-like cells *in vitro* and mature hepatocytes *in vivo,* upon transplantation[4,13,14]. However, the molecular mechanisms by which adult committed cells exit their lineage-restricted state, initiate proliferating organoids and respond to damage by generating both ductal cells and hepatocytes remain largely unknown.

During development, epigenetic mechanisms are imposed to ensure that differentiated cells remain lineage-restricted[15]. In mammals, 5-methylcytosine (5mC) is the most common DNA modification and is associated to gene repression at promoter and enhancer level[16-20]. DNA demethylation might occur passively, due to loss of DNA methylation maintenance during replication or via the conversion of 5mC to 5hmC by the Ten-eleven translocation (TET) family of methylcytosine dioxygenase enzymes[21,22], which results in dilution of 5hmC through DNA replication[23]. Moreover, cytosine demethylation can be achieved by a replication-independent mechanism mediated by TETs, whereby 5mC is converted to 5hmC, which can be further oxidized and replaced with an unmodified cytosine[24,25].

Erasure of 5mC and TET1 activity are essential for resetting the genome for pluripotency, germ-cell specification, imprinting and somatic cell reprogramming[26-30]. During development and postnatal life, *Tet1* is essential to maintain the intestinal stem cell pool[31], while *Tet2* and *Tet3* are required to induce postnatal demethylation in hepatocytes[32]. However, whether epigenetic mechanisms and/or DNA-methylation/hydroxymethylation play a role in the acquisition of cellular plasticity in adult differentiated cells during the regenerative response has not been investigated yet.

Here, we report that in the liver, during the response to tissue damage, adult resident ductal cells undergo a genome-wide remodelling of their transcriptional and methylome/hydroxymethylome landscapes in the absence of ectopic genetic manipulation. We identify TET1-mediated hydroxymethylation and its downstream

84 regulation of ErbB/MAPK and YAP/Hippo signalling pathways as one of the
85 epigenetic mechanisms required for lineage-restricted ductal cells to acquire cellular
86 plasticity, establish liver organoids and elicit a full regenerative response.
87

Results

Adult non-proliferative ductal cells undergo genome-wide changes in their transcriptional landscape during organoid initiation and as a response to tissue damage

We recently reported a liver organoid culture system that allows the clonal and long-term expansion of mouse[4] and human[13] liver ductal cells as self-renewing bi-potent organoids capable of differentiating into ductal and hepatocyte-like cells *in vitro* and *in vivo*[4,13,14,33,34]. Using the pan-ductal marker EpCAM after excluding hematopoietic and endothelial cells (see methods) we isolated pure populations of ductal cells capable of generating organoid cultures from undamaged liver with ~15% efficiency (Extended Data Figure 1a). To confirm that organoid formation is not due to a subpopulation of proliferating ductal cells, we isolated $EpCAM^+$ cells from *R26Fucci2a* mice[35] and tracked their cell cycle dynamics. As reported[36], we found that virtually all $EpCAM^+$ ductal cells are arrested in G1/G0 ($mCherry^+/mVenus^-/EpCAM^+$) (Figure 1a-b and Extended Data Figure 1b), indicating that the organoid initiating cells are non-proliferative (Figure 1c). To investigate the molecular basis that endows adult committed ductal cells to initiate bi-potent organoids, we first estimated the time required for the cells to enter the S/G2/M phase. We found that first entry into S-phase takes ~40h from isolation, while subsequent G1 phases shortened to ~15h (Figure 1d-e, Extended Data Figure 1c and Movie 1).

Next, we performed genome-wide gene expression analysis (RNA-sequencing) in cells isolated directly from the undamaged tissue (0h), cells collected prior to entry in S-phase (12h and 24h) and after proliferation initiation (48h and organoid stage, 6 days). We found that adult differentiated ductal cells undergo profound transcriptional changes during the initiation and formation of organoid cultures. We identified >3,000 genes differentially expressed (DE) during the first 24h, prior to S-phase, while 900 genes changed after proliferation started (48h *vs* organoids) indicating that most of the organoid transcriptional signature is established within 48h in culture (Figure 2a-b, Extended Data Figure 2a and Supplementary Dataset 1).

We classified the differentially expressed genes into 10 clusters. Genes in cluster 3 and 7 (increased expression from 48h-onwards), were mainly enriched in cell-cycle, while genes in cluster 5, whose expression precedes the onset of proliferation (starts at 12h and peaks at 24h), were significantly enriched for chromatin regulators (Figure

2b-c). Of note, 55% (383 out of 698) of the genes from an epigenetic modifiers' list[37] were differentially expressed, including Polycomb, SWI/SNF members and TETs, while some ductal markers were transiently down-regulated (Figure 2d-e and Extended Data Figure 2b). These results suggested that epigenetic mechanisms might be prominently involved in the initiation of liver organoids from non-proliferative, lineage-restricted ductal cells.

Organoids mimic many aspects of the tissue-of-origin in a dish[38], however, they have not been used to study the molecular mechanisms of tissue regeneration. Therefore, we opted to benchmark our organoid cultures to the *in vivo* response to tissue damage by studying the transcriptional changes that occur *in vivo* after injury and compare these to our organoid findings. For that, we induced liver damage to adult mice by administering a 0.1% 3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) supplemented diet (Figure 2f). Proliferation initiation began at day 3 (d3) and peaked at day 5 (d5) of damage (Figure 2g). Interestingly, also *in vivo*, ductal cells undergo significant genome-wide changes of their transcriptional landscape, with >1,500 genes differentially expressed between the undamaged and any of the two damage time points (Supplementary Dataset 1 and Extended Data Figure 2c-e). Notably, most of the transcriptional changes occur at d3, before the significant increase of proliferation, resembling our *in vitro* observations.

Interestingly, 71.4% of the DE genes *in vivo* were also found as DE genes *in vitro* (1,108 out of 1,552 genes) and presented similar expression patterns. Specifically, epigenetic regulators such as *Tet1, Hdac7, Uhrf1 or Dnmt1,* hepatoblast markers (*Foxa3*, *Sox4)* or ductal markers presented similar patterns (Figure 2h-i and Extended Data Figure 3a).

Altogether, these results reveal that both, *in vivo* and *in vitro*, ductal cells undergo a global rewiring of their transcriptional landscape as a response to tissue damage, and validate organoids as a model to study some molecular mechanistic aspects of tissue regeneration.

TET1 catalytic activity is required for organoid initiation and expansion

To identify potential epigenetic regulators required for the activation of ductal cells during organoid initiation, we selected some of the DE epigenetic modifiers during the first 24h and assessed the effect of their loss-of-function (siRNA knock-down) on organoid initiation. We found that depletion of *Tet1* significantly impaired organoid

formation, while *Tet2* knock-down exhibited a reduction, but was not statistically significant (Figure 3a and Extended Data Figure 3b).

Thus, we further investigated the role of TET1 in organoid initiation and expansion. For that, we generated 2 independent TET1 mutant alleles: (1) a conditional allele ($Tet1^{flx/flx}$) enabling the spatiotemporal control of TET1 deletion and (2) a hypomorphic allele ($Tet1^{hypo}$) which displays ~35% of *Tet1* mRNA and protein levels ($Tet1^{hypo/hypo}$) compared to WT littermates (Extended Data Figure 3c-e and Supplementary Table 1).

We found that ablation of TET1 in FACS-sorted ductal cells derived from $Rosa^{CreERT2}/Tet1^{flx/flx}$ abrogated organoid formation (Figure 3b), in agreement with the siRNA results (see Figure 3a and Extended Data Figure 3b). In addition, TET1 depletion in established organoids impaired their expansion (Extended Data Figure 3f). Organoids generated from the *Tet1* hypomorphic mutant mice ($Tet1^{hypo/hypo}$) exhibited reduced 5hmC levels and expansion defects, despite that they could be established (Figure 3c-e and Extended Data Figure 3g-k). Organoids derived from heterozygous or *WT* littermates displayed no growth defects (Extended Data Figure 3h-k). Importantly, ectopic expression of full-length TET1 cDNA (hypo-OE organoids), but not a catalytically inactive mutant ($TET1^{H1671Y/D1673A}$)[29,39] (hypo-$OE^{cat.mut.}$ organoids), rescued all these phenotypes (Figure 3c-e and Extended Data Figure 3g/k). Altogether, these results demonstrated that the catalytic activity of TET1 is required to initiate and propagate liver organoids from lineage-restricted, non-proliferative, ductal cells.

Genome-wide changes in DNA methylation/hydroxymethylation occur during the activation of ductal cells following damage

Given the crucial role of TET1-mediated hydroxymethylation in organoid initiation, we speculated that epigenetic regulation of DNA methylation and hydroxymethylation levels could be involved in the ductal regenerative response to damage *in vivo*. For that, we quantified the levels of DNA methylation at single base resolution by Whole Genome Bisulfite Sequencing (WGBS) in genomic DNA extracted from $EpCAM^{+}$ ductal cells sorted from undamaged and d3 and d5 DDC-damaged livers (Figure 4a, Extended Data Figure 4a-c, Supplementary Dataset 2). WGBS revealed a global increase in cytosine modification (5mC and/or 5hmC) at d3 after damage, while d5 and undamaged controls showed similar global levels (Figure

4b) although modifications occurred in the same CpG only in ~50% of the cases across the time points analysed (Extended Data Figure 4d). Next, we identified the differential levels of cytosine modification in defined regions in a CpG context (DMRs) (Extended Data Figure 4e-f). At d3, the majority of DMRs represented a gain of modified cytosine (mCpG) compared to undamaged (68%) whereas at d5 and between both damage time points, these were mainly associated with a loss in mCpG (56%, and 75%, respectively) (Figure 4c and Extended Data Figure 4g). We then analysed the levels of mCpG at the TSS (+/- 500bp) of genes transcriptionally up-regulated after damage. From all up-regulated genes, 32.6% (337 out of 1032) showed decreased methylation/hydroxymethylation levels at d3 (Figure 4d-e and Extended Data Figure 4h), suggestive of a potential role of demethylation in their transcriptional activation.

Of note, we also found that a significant proportion of all up-regulated genes (34%, 349 genes out of 1032) presented increased levels of mCpG (Figure 4f and Extended Data Figure 4h). Since WGBS cannot discriminate between 5mC and 5hmC, we hypothesized that this could be explained by an increased 5hmC. Hence, we performed Reduced Representation of Hydroxymethylation Profiles (RRHP), to identify 5hmC at single base resolution in the same DNA samples used for the WGBS (see Figure 4a and Supplementary Dataset 2). Consistent to 5hmC immunofluorescence stainings on ductal cells upon *in vivo* damage in *WT* mice or upon β1 integrin deletion (a damage model of duct-mediated hepatocyte regeneration[9]) and during organoid initiation (Extended Data Figure 5a-c), RRHP showed increased 5hmC sites upon damage (Figure 4g and Extended Data Figure 5d).

To identify 5hmC regulated targets, we analysed 3,581 genes showing differential hydroxymethylation levels i.e., presenting ≥4 unique 5hmC sites at their TSS, either in undamaged or after damage. Of note, >95% of these genes (3,450 genes) had acquired *de novo* 5hmC sites at d3, prior to proliferation, while most of these *de novo* marks were lost at d5, suggestive of a significant transient reshaping of the hydroxymethylome as a response to damage and prior to cell proliferation (Figure 4h-j and Extended Data Figure 5e). Notably, 5hmC levels did not increase in CpG islands (CGI) outside TSS (Extended Data Figure 5f).

The differentially hydroxymethylated genes could be classified in six clusters (1-6), with clusters 2-4 presenting increased 5hmC at day 3 and reduced levels at day 5 and cluster 6 (140 genes) showing overall increased 5hmC levels at day 5 (Figure 4j and

Extended Data Figure 5g). When overlapping genes with increased 5hmC with genes differentially expressed *in vivo* we found 154 genes transcriptionally up-regulated (Figure 4k and Supplementary Dataset 1). Interestingly, some of these also presented increased cytosine modifications in the WGBS at d3, prior to proliferation, hence explaining, at least in part, the observed dichotomy between the increased levels of modified cytosine in the WGBS and the increase in transcription. Among these, we found genes involved in liver regeneration signalling pathways (e.g. *Erbb2*)[40] and liver development (*Foxa3, Sox4*)[41] (Figure 4l). In addition, 84 genes showing differential 5hmC levels were also down-regulated *in vivo*, including negative regulators of the BMP pathway (*Bambi)* and genes important for hepatocyte differentiation (*Cebpa* and *Atf3)* (Extended Data Figure 5h and Supplementary Dataset 1).

Altogether, our genome-wide analyses suggest that transient increase in hydroxymethylation levels might facilitate the acquisition of cellular plasticity in ductal cells and subsequent initiation of the response to damage.

TET1 induces ductal cell plasticity through the regulation of the YAP/Hippo and ErbB/MAPK signalling pathways

Our findings indicate that hydroxymethylation levels rise upon damage in genes/pathways relevant for liver regeneration, at the time where *Tet1* expression is increased, and before the onset of proliferation. Therefore, we next sought to elucidate TET1-regulated genes involved in the acquisition of cellular plasticity during liver regeneration. Hence, we investigated TET1 genomic occupancy by performing Targeted DamID-seq (DNA Adenine Methyltransferase IDentification sequencing)[42,43] (Extended Data Figure 6a). We found 5,102 TET1 specific peaks, 56% of which were in actively transcribed regions (Extended Data Figure 6b-c and Supplementary Dataset 3). We next identified TET1 targets by overlapping the peaks to a +/-2Kb region around the TSS. We found 2,358 TET1 target genes in liver organoids, 88% of which shared an H3K4me3 peak, indicating that TET1 binding at TSS occurs mostly in transcriptionally active genes (Figure 5a). These were involved in cell-cycle, transcription and chromatin organisation, among others (Extended Data Figure 6d).

Notably, we identified TET1 binding on stem-cell genes such as *Lgr5*[10]*, Axin2*[44,45]

and *Lrig1*[46], the known TET1-target *Cdk1*[47], epigenetic regulators (*Cbx3, Ezh2, Dnmt1, Hdac1*) and liver development transcription factors (*Onecut1* and *Onecut2*) (Figure 5b and Supplementary dataset 3). TET1 and 5hmC levels were increased before transcription of the stem-cell genes *Lgr5*, *Trop2* and *Sulf2,* while both, *Lgr5* mRNA and 5hmC were reduced in organoids with low levels of TET1 ($TET1^{hypo/hypo}$) and could be rescued by ectopic expression of TET1 (Figure 5c and Extended Data Figure 6e-g). TET1-dependent 5hmC might co-operate with the existing transcriptional regulatory machinery, as the recruitment of TET1 to *Lgr5,* a TCF4 target[48], paralleled the binding of TCF4/*Tcf7l2* to the locus (Figure 5c). As expected, no TET1 binding or changes in 5mC/5hmC were detected in genes not expressed, including the hepatoblast marker *Afp* and hepatocyte marker *Alb* (Figure 5b and Extended Data Figure 6g). Of note, some TET1 targets were also up-regulated *in vivo* (see Figure 4, Extended Data Figure 4h and Supplementary Dataset 4). The overlap between TET1 targets and DE genes *in vivo* and *in vitro* (see Figure 2h) suggests that TET1 mainly functions as a transcriptional activator in liver ductal cells (Figure 5d and Supplementary Dataset 1).

To further elucidate the mechanism by which TET1-mediated hydroxymethylation regulate organoid formation and liver regeneration we performed KEGG pathway enrichment analysis on TET1 targets that were also differentially hydroxymethylated upon damage *in vivo*. This revealed a significant enrichment on several components/targets of signalling pathways including mTOR, ErbB, MAPK and YAP/Hippo, among others (Figure 6a, Supplementary Dataset 2).

Interestingly, mTOR, ErbB, MAPK and YAP/Hippo have been extensively described to be essential for liver regeneration *in vivo*[40,49-53]. Additionally, YAP/Hippo and mTOR have been recently identified as required for intestinal[54] and liver[50] organoid expansion. Therefore, we hypothesized that the direct regulation of these pro-regenerative pathways could explain the mechanism by which TET1 facilitates the acquisition of cellular plasticity in liver ductal cells upon tissue injury or during organoid initiation. We first validated TET1 occupancy by ChIP-qPCR on selected TET1 targets [ErbB and MAPK (*Egfr, Foxo3, Socsc2, Jun*) and YAP/Hippo (*Wwtr1/Taz, Tead1, Gadd45b and Ctgf*)] (Figure 6b). Next, we assessed whether their expression was TET1 dependent, by evaluating their mRNA levels following TET1 depletion in $Rosa^{CreERT2}/Tet1^{flx/flx}$ organoids. We found a consistent down-regulation of YAP/Hippo pathway components such as *Wwtr1/Taz* and *Tead1* and target genes

such as *Gadd45b* and *Ctgf* upon TET1 knock-down (Figure 6c). The expression levels of these, except for *Gadd45b*, were rescued in TET1 hypo-OE organoids (Figure 6d). For several of the components and targets of the ErbB/MAPK pathways (*Egfr*, *Foxo3, Jun*) we detected both, up- or down-regulation following TET1 knock-down (Figure 6c).

Thus, we evaluated whether TET1-dependent regulation of these pathways is involved in the acquisition of cellular plasticity during organoid formation. We confirmed TET1 binding to some of these targets at 18hrs after seeding (Figure 6e). To elucidate whether ErbB, MAPK and YAP/Hippo signalling act down-stream of TET1, we then supplemented the cultures with small molecule inhibitors of the aforementioned pathways (Gefitinib (EGFRi), PD032509 (MEKi) and Verteporfin (YAPi)) for the first 18h in culture (0-18hrs), i.e., before TET1 binding, and at 18hrs-48hrs, i.e., after TET1 binding, and evaluated organoid formation efficiency 6 days later. Treatment at 18-48hrs, once TET1 is bound to its targets, induced a significant decrease of organoid formation, thus suggesting that the regulation of ErbB, MAPK and YAP/Hippo signalling could represent one of the mechanisms by which TET1 positively regulates organoid formation from mature liver ductal cells (Figure 6f). Conversely, treatment before TET1 binding (0-18h) or inhibition of FGFR1/3 did not cause any significant effect on organoid formation (Figure 6f and Extended Data Figure 7a). mTOR inhibition instead, resulted in ablation of organoid formation regardless of the time of supplementation, suggesting that either this pathway is essential during the first 18h for ductal cell survival *in vitro* or is not regulated by TET1 (Extended Data Figure 7a). Thus, our results suggest that TET1 promotes the acquisition of cellular plasticity in ductal cells, at least in part, via the regulation of YAP/Hippo and ErbB, MAPK signalling pathways.

TET1 is required for ductal-mediated hepatocyte and cholangiocyte regeneration

To elucidate whether TET1 is relevant for liver regeneration, we induced liver damage to the *Tet1* hypomorphic and ductal specific *Tet1* mutant mice. As damage paradigms, we opted for three different models: (1) acute damage with 5 days DDC treatment; (2) chronic damage caused by repetitive doses of DDC and (3) a damage model where hepatocyte proliferation is impaired by over-expression of p21 and

ductal cells have been demonstrated to regenerate both themselves and hepatocytes [2,8,9] (Supplementary Table 1).

To address the role of TET1 during acute liver damage we used the TET1 hypomorphic allele *(Tet1$^{hypo/hypo}$),* since the conditional *Rosa^{CreERT2} /Tet1$^{flx/flx}$* exhibited partial lethality upon Cre induction, in agreement with the published TET1 full KO[31] (Supplementary Table 1). *Tet1$^{hypo/hypo}$* mice presented no obvious phenotype under homeostasis (Extended Data Figure 8a-d). However, upon damage, it exhibited significantly lower number of proliferating liver ductal cells (Ki67^{+}/OPN^{+} cells) and absolute number of liver ductal cells when compared to *WT* control littermates (Figures 7a-b and Extended Data Figure 8e-h). Notably, this reduced proliferation of the ductal compartment was not explained by differences in the extent of liver damage between genotypes (Extended Data Figure 8b and d).

Interestingly, upon chronic liver damage, *Tet1$^{hypo/hypo}$* mice presented extended fibrosis (Figure 7c-d). Since *Lgr5* depletion *in vivo* results in tissue fibrosis[55] we evaluated the levels of *Lgr5* in our mutant mice and found reduced expression and less hydroxymethylation of *Lgr5* loci in *Tet1$^{hypo/hypo}$* mice (Extended Data Figure 8i).

To discriminate whether the defects on liver regeneration observed were caused by the lack of TET1 expression in the adult ductal compartment, we generated a ductal-specific TET1 mutant mouse by crossing the *Tet1$^{flx/flx}$* allele with the ductal specific driver *Prom1Cre^{ERT2}* (Extended Data Figure 9a and[56,57]). To visualise and trace recombination events, we further combined this mouse with the *Rosa^{lslZsGreen}* reporter to generate the *Prom1Cre^{ERT2}/Rosa^{lslZsGreen}/Tet1$^{flx/flx}$*, referred here as *Prom1$^{\Delta Tet1/ZsGreen}$* in contrast to the TET1 WT, named here as *Prom1$^{Tet1WT/ZsGreen}$* mice. We confirmed the reliability of the ZsGreen to reflect TET1 levels after recombination. No ZsGreen induction was found without tamoxifen treatment (Extended Data Figure 9b-d).

To assess the role of TET1 in ductal-mediated liver regeneration, we used a recently established liver damage model where hepatocyte proliferation is inhibited by p21-over-expression[9] and fed the mice DDC for 3 weeks (Figure 8a and Extended Data Figure 9e-f). We observed a massive expansion of ductal cells (OPN^{+}/ZsGreen^{+}) in *Prom1$^{Tet1WT/ZsGreen}$* mice while *Prom1$^{\Delta Tet1/ZsGreen}$* mice exhibited a significant reduction (Figure 8b-c), in agreement with our Tet1 hypomorphic model (see Figure 7a-b). Notably, when we examined the contribution of TET1 depleted ductal cells to hepatocyte regeneration, we observed a dramatic reduction in the size of hepatocyte

clusters in the *Prom1*$^{\Delta Tet1/ZsGreen}$ mice, with most clusters formed by 1-2 cells only, while *Prom1*$^{Tet1WT/ZsGreen}$ mice readily generated hepatocyte clusters from 1 to 156 cells (Figure 8d-e).

Molecular analysis of TET1-null ductal cells upon damage indicated that also *in vivo* TET1 binds to the TSS and regulates the expression of some genes from the pro-regenerative YAP/Hippo and ErbB/MAPK signalling pathways (namely *Egfr*, *Gadd45b*, *Wwtr1/Taz* and *Tead1)* (Extended Data Figure 9g-h), in line with our organoid data (see Figure 6).

Altogether, our studies demonstrate that TET1 plays a crucial role in ductal-driven liver regeneration, at least in part, through the direct activation of the YAP/Hippo and ErbB/MAPK signalling pathways.

Discussion

Many adult epithelial tissues exhibit cellular plasticity not associated with unrelated fates, but with contribution to tissue repair (see[58] for extended details). Under homeostasis a unipotent population of hepatocytes maintain the tissue[45,59,60]. Following hepatocyte injury, the lost tissue is repaired by remaining hepatocytes[61]. However, upon severe or chronic liver damage, mature cholangiocytes activate a regenerative response to restore both themselves and hepatocytes[5,9,62,63]. Yet, the molecular mechanisms behind the activation of this cellular plasticity on liver resident ductal cells remain largely unknown. This knowledge is critical to understand human liver diseases characterized by prominent ductal proliferation and hepatic fibrosis[64,65].

Here we demonstrate that upon damage and during organoid formation resident ductal cells undergo genome-wide changes in their transcriptional landscape and a significant remodelling of their DNA methylome and hydroxymethylome. We identify demethylation and TET1-mediated hydroxymethylation as an epigenetic mechanism required for ductal cell activation *in vitro* and *in vivo,* after damage (Figure 8f). The acquisition of the cellular plasticity that endows differentiated ductal cells with regenerative capacity *in vivo*, might occur through a progenitor state, as our organoid data imply. However, whether *in vivo*, new cells are provided through a direct division of differentiated cells, via de-differentiation to a progenitor state, by direct trans-differentiation or a combination of all these[66], remains unknown and will require further and more extensive investigations.

Cancer cell lines and liver cancer, exhibit relatively low levels of 5hmC[67,68]. In contrast, our results, indicate that transient high levels of 5hmC are required to induce ductal cells to activate the regenerative program, similar to what has been reported in pluripotent cells[39]. TET enzymes have been shown to promote genome integrity in mouse ES cells[69]. Hence, it is tempting to speculate that transient *Tet1* induction during liver damage might be a mechanism for activating the regenerative program in ductal cells while preserving genome integrity in the regenerating cell.

Interestingly, our analyses indicate that the mechanism by which TET1 facilitates the acquisition of cellular plasticity and subsequent pro-regenerative effect is, at least in part, through the direct regulation of ErbB, MAPK and YAP/Hippo regenerative pathways[40,50-53]. Whether other genes transcriptionally activated/repressed by TET1 are involved in the process requires further investigations.

Notably, the rewiring of the transcriptome and DNA methylome and hydroxymethylome occurs prior to proliferation, as a response to tissue damage and in the absence of any ectopic genetic manipulation. This mechanism resembles embryonic reprogramming, where genome-wide methylation erasure is essential to reset the epigenome for pluripotency[28]. Our observations might represent a more general mechanism by which adult committed cells initiate the regenerative response to damage.

Acknowledgements

M.H. is a Wellcome Trust Sir Henry Dale Fellow and is jointly funded by the Wellcome Trust and the Royal Society (104151/Z/14/Z). A.H.B. was funded by Wellcome Trust Senior Investigator Award 103792 and Royal Society Darwin Trust Research Professorship. S.J.F. is supported by a MRC grant: MR/P016839/1. L.A. was supported by a Marie-Curie Postdoctoral fellowship (Grant RG81823_H2020-MSCA-IF-2016) and a NC3Rs grant awarded to M.H (NC/R001162/1). M.A.M. is supported by a Medical Research Council (MRC) doctoral training grant (MR/K50127X/1). L.C-E is jointly funded by a Wellcome Trust Four-Year PhD Studentship with the Stem Cell Biology and Medicine Programme and by a Wellcome Cambridge Trust Scholarship. J.v.d.A. was supported by EMBO Long-term Fellowship ALTF 1600_2014 and Wellcome Trust Postdoctoral Training Fellowship for Clinicians 105839. F.A. is supported by ERC advanced research grant to M.Z.G. M.Z.G. is a Wellcome Trust Senior Research Fellow. This work was partially funded by a H2020 LSMF4LIFE (ECH2020-668350) awarded to M.H. and partially funded by ERC advanced grant to M.Z.G and a Wellcome Trust Senior Investigator Award awarded to E.A.M. (104640/Z/14/Z, 092096/Z/10/Z) and a Cancer Research Programme Grant awarded to E.A.M. (C13474/A18583, C6946/A14492). G.V. would like to thank Wolfson College at University of Cambridge and the Genetics Society, London for financial help. The authors acknowledge core funding to the Gurdon Institute from the Wellcome Trust (092096) and CRUK (C6946/A14492). The authors thank Dr Robert Krautz and Dr Walter Sanseverino for advice on bionformatic analyses; Mr Robert Arnes-Benito and Andrew A. Malcom, for histological and immunostaining assistance; Dr Wolf Reik and Dr José Silva for sharing TET1 plasmids; Mr Richard Butler for developing macro scripts to quantify 5hmC stainings; Mr Kay Harnish and Dr Charles Bradshaw of the Gurdon Institute genomic and bioinformatic facility for high-throughput sequencing; The Gurdon Institute facilities for assistance with imaging, animal care and bioinformatics analysis and Dr Andy Riddell and Dr Maike Paramor (Cambridge Stem Cell Institute) for assistance with FACS sorting and library preparation, respectively; CRUK CI genomic facility for sequencing of WGBS and RRHP libraries; Margaret Keighren (MRC Human Genetics Unit, University of Edinburgh) for technical support. Life Science editors for assistance during manuscript preparation. M.H. would like to

thank Prof Benjamin Simons and Prof Hans Clevers for critical comments on the manuscript.

Author contributions

M.H. and L.A. conceived and designed the project and interpreted the results. L.A., M.A.M., L.C-E., G.B., G.V., N.A., J.v.d.A., A.R. and MH designed and performed experiments and interpreted results.. L.A. designed and performed the *in vitro* experiments, M.A.M., designed and performed the *in vivo* experiments, L.C-E., the hydroxymethylation and EdU stainings, G.B, experiments with small molecule inhibitors. G.V. and E.A.M. prepared and analysed WGBS and RRHP libraries, analysed RNAseq and interpreted corresponding bionformatic analyses related. N.A., A.R. and S.J.F. performed experiments with β1 integrin model and interpreted results of the p21 models. J.v.d.A. and A.H.B. performed DamID-seq experiments. B.F-C helped on the *in vivo* analysis. R.A.C. helped on bioinformatics analyses. R.L.M. provided the *R26Fucci2a* line. F.A. and M.Z.G. performed the live imaging of ductal cells. L.A. and M.H. wrote the manuscript. All authors commented on the manuscript.

Author information

Authors declare no competing financial interests.

Figure Legends:

Fig. 1: G1/G0 arrested liver ductal cells require ~48h to start cell proliferation and initiate liver organoids cultures

R26Fucci2a mice constitutively express a bi-cistronic cell-cycle reporter that allows discriminating between G1/G0 [Cherry-hCdt1+ (30/120), red] and S/G2/M [Venus-hGem+, (1/110) green] phases of the cell cycle. **a**, Experimental approach **b**, $EpCAM^+$ liver ductal cells from R26Fucci2a mice were FACS-sorted according to the expression of mCherry-hCdt1 (C) and/or mVenus-hGem (V). The graph represents percentage of $EpCAM^+$ cells positive for mCherry and/or mVenus. Each dot represents an independent experiment from an independent mouse (n=3). Graph is presented as mean±SD of 3 independent experiments. **c**, Representative bright field images of 500 C^+/V^- $EpCAM^+$ and C^-/V^- $EpCAM^+$ cells cultured for 6 days as liver organoids. The graph represents mean±SD of organoid formation efficiency (n=3 experiments). *p*-value was calculated using Student's two tailed t-test. **, p=0.001413095. **d,** Still images from a representative movie of C^+/V^- $EpCAM^+$ ductal cells monitored for 72h using a spinning-disk confocal microscope. Scale bars, 10μm. **e**, Graph represents G0/G1 length for the first (I) and second (II) cell cycles since t=0h (isolation) of n=34 cells, pooled from 3 independent experiments). Global mean

of G0/G1 length is shown (G0/G1 I = 37.97 h, hours; G0/G1 II = 10.20 h, hours). h, hours.

Fig. 2: Liver ductal cells undergo genome-wide changes in their transcriptional landscape during organoid initiation and *in vivo* upon damage

a-e, Expression analysis of ductal cells during organoid initiation. **a,** Experimental Scheme. Graph represents DE genes (pairwise approach with Wald test performed using Sleuth. Threshold FDR <0.1) **b**, Hierarchical clustering of all 7580 DE genes. Heatmap represents averaged TPM values of biological replicates scaled per gene (Z-score). Number in bold, cluster. n, number of genes/cluster. **c**, GO and statistical analyses were performed using DAVID 6.8. Red, cluster containing DE genes at 12h and 24h. **d,** Heatmaps representing averaged Z-score of indicated genes. **e,** Graphs represent mean±SD of n=6 independent RT-qPCR experiments. Independent experimental data are listed in Source Data. Data are presented as fold-change compared to t=0h. *p*-value is calculated using two-way ANOVA combined with Tukey HSD test. *p*-value of comparisons *vs* t=0 are shown. **, p<0.01; ***, p<0.001. Exact *p-values* are provided in Source Data. **f-i,** Expression analysis of ductal cells following liver damage by supplementing the diet with 0.1% DDC (see methods). **f,** Experimental scheme. **g,** Immunofluorescence analysis of ductal cell proliferation upon damage. Representative images are shown (3 experiments). Scale bar, 50µm. Graph represents mean±SD of proliferating ductal cells (undamaged n=3 mice, DDC d2 n=3 mice, d3 n=4 mice, d5 n=4 mice). *p*-values were calculated *vs* undamaged using pairwise comparisons with Wilcoxon rank sum test (DDC d3 p= 0.01201; DDC d5 p= $7.6E^{-05}$). *, p<0.05; ***, p<0.001. **h,** RNA sequencing analysis of sorted $EpCAM^+$ ductal cells isolated from undamaged or DDC-treated livers (2 livers have been assessed per time point). Venn diagram, overlap between DE genes *in vitro* and *in vivo*. *p*-value is calculated using normal approximation of the hypergeometric probability. Table indicates the GO analysis (top 3 significant categories) of the 7 clusters identified in **i** (Cluster 1 n=183; Cluster 2 n=276; Cluster 3 n=260; Cluster 4 n=69; Cluster 5 n= 76; Cluster 6 n= 154; Cluster 7 n=90) and their p-values obtained with DAVID 6.8. **i**, Heatmap (averaged Z score) of the hierarchical clustering of the 1108 DE genes based on the *in vitro* expression profile. Number in bold, cluster. n, number of genes/cluster.

Fig. 3: TET1 catalytic activity is required for liver organoid initiation and maintenance

a, FACS-sorted $EpCAM^+$ ductal cells freshly isolated from WT undamaged livers were transfected with a pool of siRNAs, each of them targeting specifically a selected epigenetic modifier, and organoid formation efficiency was evaluated 10 days later. Results are shown as percentage of organoid formation efficiency compared to mock transfected cells. The graph represents mean±SD of n=3 independent experiments (dots). *p*-values were calculated using one-way ANOVA in conjunction with Tukey's HSD test by comparison to siCtrl. *, *p*= 0.01031057 siTet1 *vs* siCtrl,. **b,** FACS-sorted $EpCAM^+$ ductal cells derived from *RosaCreERT2 x* $Tet1^{flx/flx}$ mouse livers were plated in organoid isolation medium supplemented with 5µM hydroxytamoxifen or vehicle and organoid formation efficiency was evaluated 6 days later. Representative bright field images are shown. Data are reported as percentage of organoid formation compared to Cre^-Tam^- cells. Graphs represent mean±SD of n=3 independent experiments. *p*-value was calculated using Student's two-tailed t-test *vs* Cre^-Tam^- (*, Cre^-Tam^+ *p*=0.03781815; ***, Cre^+Tam^+ *p*=$4.812E^{-05}$). **c-e,** $EpCAM^+$ ductal cells

isolated from Tet1 hypomorphic mice were used to generate liver organoids (*Tet1*$^{hypo/hypo}$, blue) or were transfected with a hTET1 full length cDNA (hypo-OE organoids, red) or catalytically inactive hTET1 H1671Y/D1673A (hypo-OEcat.mut. organoids, turquoise). Organoids derived from WT littermates were used as controls (black). **c,** Scheme indicates the lines generated. **d,** Western blot analysis of TET1 protein levels. The graph represents TET1 levels. Complete blot is shown in data source. Results are presented as mean±SD of n=3 independent experiments (dot). **, *p*-value calculated using Student's two-tailed t-test vs WT (*Tet1*$^{hypo/hypo}$ p=0.006779543). **e,** Representative bright field images of *WT* (2 line), *Tet1*$^{hypo/hypo}$(4 line) hypo-OE (1 line) and hypo-OEcat.mut. (1 line) organoid lines at passage 3. Graph indicates passage number.

Fig. 4: Liver ductal cells undergo global remodelling of DNA methylation and hydroxymethylation landscapes *in vivo* upon damage

a-l, gDNA from undamaged or DDC-damaged livers was split in two fractions and prepared for WGBS (**a-f**) or RRHP (**g-l**) (2 mice per time point). **a,** Experimental design. **b,** Graph shows the percentage of modified CpG (mCpG) sites according to different level categories (average of replicates). **c,** Number of differentially methylated/hydroxymethylated regions (DMRs) present in the n=2 biological replicates. DMR were called based on a modification difference ≥25%, p<0.05 (see methods). **d-e**, Graphs (mean±95%CI) represent percentage of modified cytosines at TSS for all n=337 up-regulated genes (**d**) or selected ones (**e**) showing decreased mCpG levels at d3 (average of replicates). *p*-value was obtained by Kruskal Wallis test with Dunns multiple comparison. ****, undamaged *vs* d3 p<0.0001, ***, undamaged *vs* d5 p=0.0003, d3 *vs* d5 p=0.0004. TET1 targets (see Figure 5) are represented in bold red. **f**, Graph represents all n=349 up-regulated genes after damage presenting increased mCpG level at TSS (mean ±95% CI). *p*-value was obtained by Kruskal Wallis test with Dunns multiple comparison. ****, undamaged *vs* d3 p<0.0001, undamaged *vs* d5 p=0.3773, d3 *vs* d5 p<0.0001. **g,** Distribution of total 5hmC sites identified. **h,** Number of genes showing ≥4 5hmC sites around their TSS. **i,** Graph represents median±IQR of 5hmC counts from the n=3581 genes differentially hydroxymethylated. *p*-value was obtained using Kruskal Wallis test coupled with Dunn's multiple comparison. All p-values are <0.0001. ****, p<0.0001 **j,** The heatmap represents the z-score values of 5hmC absolute count. 5hmC levels were classified into 6 clusters. n, number of genes/cluster. Graphs (median±IQR) represent the number of 5hmC counts of differentially hydroxymethylated genes. *p*-value was obtained by Kruskal Wallis test with Dunns multiple comparison. All p-values correspond to p<0.0001 (****), except for ***, p=0.0009. **k**, Heatmap represents Z-score of the 154 overlapping genes. **l**, Graph represents the levels of mCpG from the 154 genes identified in **k** averaged for the 2 biological replicates. In **k-l,** TET1 targets (see Figure 5) are represented in bold red.

Fig. 5: TET1 regulates the activation of genes involved in organoid formation and liver regeneration

a-b, TET1-DamID analyses were performed in 3 independent experiments. **a,** Heatmaps of TET1-DamID (left) and H3K4me3 (right) binding at the TSS Venn diagram indicates the overlap between the DamID-seq TET1 and H3K4me3 target genes identified by ChIP-seq. **b,** Genome tracks of TET1 (Dam-ID) and H3K4me3 (ChIP) peaks on selected genes. Graphs show TET1-Dam/Dam only ratio (blue) and H3K4me3 number of reads (green). **c,** Sorted EpCAM^{+} cells from WT undamaged

livers were cultured as organoids and analysed at the indicated time points (n3 experiments). Upper panels: hMeDIP (dots, green) and MeDIP (squares, red) levels in the indicated genomic region upstream and downstream of the *Lgr5* TSS. Lower panels: TET1 (blue), TCF4 (brown) and H3K4me3 (purple) ChIP-qPCR at the TSS. mRNA expression is shown in black. *p*-value was obtained using Student's two-tailed t-test. Statistical analyses were performed *vs* t=0h. (Upstream 5hmC 12h p=0.004305136, 18h, p=3.26345E^{-05}, 48h p=8.36527E^{-06}; 5mC 12h p=0.009532377, 18h, p=0.001130234, 48h p=0.001564496; TSS 5hmC 12h p=0.011044339, 18h, p=0.005230947, 48h p=0.000485153; Downstream 5hmC 18h, p=*0*.004305136, 48h p=3.26345E^{-05}; 5mC 48h p=8.36527E^{-06}; *Lgr5* mRNA 48h p=0.001991489; TET1 ChIP 12h p=0.005403182, 18h, p=0.003789515, 48h p=0.000119801; H3K4me3 ChIP 48h p= 0.000774002). *, p<0.05; **, p <0.01***; p <0.001. **d,** Overlap between the n=1108 DE genes identified in Fig. 2h-i and TET1 targets identified by DamID-seq. *p*-value of the overlap is calculated using normal approximation of the hypergeometric probability. The heatmap (TPM, z-scored) presents the expression profile of the 216 TET1 targets DE *in vivo* and *in vitro*. Graphs show the gene expression levels of n=216 genes (median±95% CI) as ln(TPM +1). *p*-values are obtained with one-way ANOVA followed by Tukey's multiple comparisons test. 0h *vs* 48h p=0.0379, 0h *vs* Org p=0.0039; Und *vs* d3 p=0.0013, Und *vs* d5 p<0.0001.*, p< 0.05; **, p <0.01; ***, p <0.001.

Fig. 6: Tet1 regulates YAP/Hippo and ErbB, MAPK signalling pathways

a, KEGG pathway enrichment and statistical analyses on the genes identified as TET1-DamID targets in liver organoids (n=3) and showing differential levels of 5hmC *in vivo* from RRHP using DAVID 6.8. **b**, TET1 ChIP-qPCRs in liver organoids. Data are reported as percentage of input. Graph represents mean ±SD of n=3 independent experiments. **c**, mRNA expression levels of selected TET1 targets in WT or *RosaCreERT2 x Tet1*$^{flx/flx}$ organoids both treated with 5µM tamoxifen for 24hrs. Cells were harvested 24hrs after tamoxifen treatment. Data are reported as fold change compared to Ctrl. Graph represents mean±SD of n=3 independent experiments. *p*-value obtained using Student's two tailed t-test upon comparison to Ctrl. *Egfr*, p= 0.000479886; *Foxo3*, p= 0.031392276; *Jun*, p= 0.004319905; *Gadd45b*, p= 0.023554286; *Ctgf*, p= 0.005333732; *Wwtr1*, p= 0.000230442; *Tead1*, p= 0.002322422. *, p<0.05; **, p<0.01; ***, p<0.001. **d**, mRNA expression levels of YAP/Hippo TET1 targets in Tet1$^{hypo/hypo}$ organoids and TET1$^{hypo\text{-}OE}$ organoids. Graph represents mean±SD of n=3 independent experiments. *p*-value obtained using Student's two tailed t-test upon comparison to WT. *Gadd45b*, Tet1$^{hypo/hypo}$ p= 6.00424E^{-05}; TET1$^{hypo\text{-}OE}$ p= 6.24089E^{-05}. *Ctgf*, Tet1$^{hypo/hypo}$ p= 0.000677729; TET1$^{hypo\text{-}OE}$ p= 0.001247481. *Wwtr1,* Tet1$^{hypo/hypo}$ p= 0.002222631; TET1$^{hypo\text{-}OE}$ p= 0.010861863. *Tead1*, Tet1$^{hypo/hypo}$ p= 0.009343297; TET1$^{hypo\text{-}OE}$ p= 0.013645094. *, p<0.05; **, p<0.01; ***, p<0.001 **e**, TET1 ChIP-qPCRs in EpCAM^{+} FACS-sorted cells grown in organoid conditions for 18hrs. Data are reported as percentage of input. Graph represents mean ±SD of n=3 independent experiments. **f**, EpCAM^{+} ductal cells freshly isolated from undamaged livers were treated at 0-18hrs or 18-48hrs with the small molecule inhibitors as indicated. Organoid formation was quantified at day 6. Graph represents organoid formation efficiency and indicates mean ±SD of n=6 independent experiments. Statistical analyses were performed with two-ways ANOVA with Bonferroni's multiple compared test *vs* DMSO control group. 18-48hrs Gefitinib, p<0.0001; PD0325901 p<0.0001; Verteporfin, p=0.0039. **, p<0.01; ***, p<0.001. Representative pictures of organoids are shown.

Fig. 7: Tet1 hypomorphic mice exhibit reduced ductal regeneration and extensive fibrosis upon damage

a-b, *WT* (grey) and *Tet1*$^{hypo/hypo}$ mice (blue) were fed normal chow or a chow supplemented with 0.1% DDC for 5 days. **a**, Representative images of immunofluorescence staining for the ductal marker OPN (red) and the proliferation marker Ki67 (white). Scale bar, 25µm. PV, portal vein. Graphs represent the percentage of proliferating (Ki67$^+$) ductal cells (OPN$^+$) (median±IQR) obtained from 55 FOV for *WT* (n=3) and 56 FOV for *Tet1*$^{hypo/hypo}$ mice (n=3) at day 0 (undamaged), and 253 FOV for *WT* (n=7) and 169 FOV for *Tet1*$^{hypo/hypo}$ (n=6) at day5 of DDC damage. Data are represented a boxplots showing the median, IQR and overall range. Grey dots, outliers from a single counted FOV defined as >1.5 IQR above or below the median. Red squares, median level corresponding to each independent mouse. *p*-values were obtained using two-sided Kolmogorov-Smirnov test. ***, p< 2.2x10^{-16}. **b,** Histogram showing the population distribution of proliferating ductal cells (OPN$^+$, Ki67$^+$) by plotting frequency density of counts across the sample range (bar) and the kernel density estimate line. Dashed lines show median values. **c-d,** *WT* (grey) and *Tet1*$^{hypo/hypo}$ (blue) mice were fed normal chow or a chow supplemented with 0.1% DDC for 5 days for 8 consecutive cycles as described in the scheme and methods. Liver tissues were collected 3 months after the last cycle and PicroSirius red staining was performed to analyse the levels of fibrosis (collagen deposition). **c**, Representative images of PicroSirius red staining (red) (3 mice per time point). Scale bar, 200µm. **d,** Graph represents mean±95% CI of the area of collagen deposition per FOV (n=3 mice per time point per genotype). Statistical analysis was performed on the 3 mean values per genotype compared to undamaged using Student's two-tailed t-test. *, p<0.05.

Fig. 8: Ductal specific TET1 depletion results in impaired hepatocyte regeneration

a, Experimental Scheme. **b,** Representative images of 10µm liver sections showing ZsGreen$^+$ ductal cells (OPN$^+$) (n=9 per genotype). Scale bar, 50µm **c,** Graph showing median±IQR of average OPN$^+$ cells per FOV for each individual mouse (n=9 per genotype). Global median level is highlighted in red. *p*-value was calculated using Wilcoxon rank sum test. *, p= 0.03768. **d,** Representative images of 50µm frozen liver sections showing regenerative clusters of ZsGreen$^+$ hepatocytes (HNF4a$^+$) and ductal cells (OPN$^+$). Scale bar, 50µm. **e,** Cumulative relative frequency plots (top graph) and corresponding box plots (bottom graph) showing median (red), upper and lower quartiles and the range (dots represent outliers) of ZsGreen$^+$ hepatocyte cluster size of *Prom1*$^{TET1WT/ZsGreen}$ (n=3) and *Prom1*$^{\Delta Tet1/ZsGreen}$ (n=6) mice. *p*-value was determined by two sided Kolomogorov-Smirnov test. ***, p< 2.2x10^{-16}. **f**, Experimental model.

Extended Data Figure 1: Non-proliferative EpCAM$^+$ ductal cells initiate organoid cultures

a, EpCAM$^+$ ductal cells were isolated from WT livers by FACS using a sequential gating strategy as follows: cells were gated for FSC and SSC and subsequently singlets were gated using FSC/Pulse width. Then, cells were negatively selected for PE/Cy7 (to exclude CD11b$^+$, CD31$^+$ and CD45$^+$ cells) and positively selected for APC (EpCAM$^+$) to obtain CD11b$^-$/CD31$^-$/CD45$^-$/EpCAM$^+$ ductal cells (EpCAM$^+$ cells). These cells give rise to proliferative organoids with ~15% efficiency. Representative bright field pictures of 500 EpCAM$^+$ and EpCAM$^-$ cells 6 days after

seeding. Graph represents mean ±SD of n=3 independent experiments. **b,** RT-qPCR analysis of gene expression of the proliferation marker *mKi67* (left) and stem-cell (*Lgr5*) and ductal (*Epcam* and *Sox9*) markers (right) at the indicated time points after seeding. Graphs represent the mean of n=3 independent experiments. *p*-value obtained using Student's two tailed t-test upon comparison to t= 0h. *, p<0.05; ***, p<0.001. **c,** Proliferation analysis. EdU (10µM) was incorporated to sorted EpCAM^{+} ductal cells at different intervals after seeding (0h, 24h and 48h, arrows) and evaluated by immunofluorescence analysis 24h after each incorporation. Representative images are shown. Scale bar, 10µm. Graph represents the percentage of EdU+ cells. Results are expressed as mean±SD cells from n=3 independent experiments. Student's two tailed t-test statistical analyses were performed *vs* t=24h. *, p<0.05; **, p<0.01; ***, p<0.001

Extended Data Figure 2: Transcriptional changes in ductal cells *in vitro* during liver organoid formation and *in vivo* upon damage

a-e, RNA-seq analysis of ductal cells isolated from adult livers (0h) and at different time points after culture. For DE genes, a pairwise approach with Wald test was performed on each gene using Sleuth. FDR <0.1 was selected as threshold. **a**, Graphs represent the number of significantly DE genes for each comparison. **b**, Hierarchical clustering analysis of epigenetic regulators found DE (383 out of 698 published in ref 49), in at least one comparison. Heatmap represents averaged TPM values scaled per gene. Results are presented as the averaged gene expression of the biological replicates. n, number of replicates. **c-e**, RNA-seq analysis of ductal cells isolated from adult livers (0h) and at day 3 and day 5 after liver damage (2 mice were assessed per time point). The heatmap shows the 1552 DE genes at least in one comparison (TPM>5, FDR<0.1, |b|>0.58). Clustering analysis identified 5 different clusters (Clusters 1-5) according to the expression profile (Cluster 1 n=835; Cluster 2 n=185; Cluster 3 n=503; Cluster 4 n=20; Cluster 5 n=9). Number of genes in each cluster is indicated in brackets. Results are presented as average of the at least 3 biological replicates. **d**, Graph represents the number of significant DE genes in the different comparisons. **e**, GO and statistical analyses of the 3 main clusters identified in **c** were performed using DAVID 6.8.

Extended Data Figure 3: TET1 catalytic activity is required for liver organoid formation and maintenance

a, *Tet1* and *Lgr5* mRNA levels (n=3 mice). Student's two-tailed t-test statistical analyses were performed *vs* undamaged. **b**, *Tet1* mRNA levels (24h after transfection) and organoid formation efficiency 10 days after Tet1 siRNA knock-down using 4 independent Tet1 siRNAs. Data is presented as percentage relative to siCtrl. Graph indicates mean±SD of n=3 independent experiments. Student's two-tailed t-test statistical analyses were performed *vs* siCtrl. **c**, Scheme of the two different *Tet1* alleles used. **d,** *Tet1* mRNA levels in *WT*, *Tet1*$^{hypo/+}$ and *Tet1*$^{hypo/hypo}$ and *Tet1* conditional knock-out (cKO) organoids presented as mean±SD of n=3 experiments. **e,** Representative Western blot image showing TET1 protein levels in *WT*, *Tet1*$^{hypo/+}$ and *Tet1*$^{hypo/hypo}$ organoids (3 independent experiments). **f**, Organoid formation efficiency from FACS-sorted EpCAM^{+} cells derived from *RosaCre*^{ERT2} x *Tet1* $^{flx/flx}$ livers treated with 5µM hydroxytamoxifen (mean±SD of n=3 independent experiments). Student's two-tailed t-test statistical analyses were performed *vs* non-induced control. **g,** Whole mount immunofluorescence staining of 5hmC (green) on *WT*, *Tet1*$^{hypo/hypo}$, *hypo-OE* and *hypo-OE*$^{cat.mut.}$ organoids. Representative images are shown (2 experiments).

Scale bar, 50 μm. **h**, Graph represents organoid size at the indicated passages (mean±SD of n=3 independent experiments). Student's two tailed t-test statistical analyses were performed *vs* WT. **i**, Growth curves. **j,** Organoid formation efficiency at the indicated passage expressed as a percentage of organoids. Graphs represent mean±SD of n=3 independent experiments. Student's two tailed t-test statistical analyses were performed *vs* WT. **k**, Representative confocal images of Cleaved Caspase 3 whole mount immunostaining on *WT*, *Tet1*$^{hypo/hypo}$, *hypo-OE* and *hypo-OE*$^{cat.mut.}$ organoids (2 independent experiments). Scale bar, 25μm.

Extended Data Figure 4: WGBS of ductal cells upon damage uncovers a global epigenetic remodelling of the DNA methylome

a, Number of WGBS unique mapped reads in the different biological replicates. **b**, Bisulfite conversion rate. **c-h**, WGBS analyses were performed in merged biological replicates per time point (n=2). Only CpG sites with ≥3 reads were further analysed. **c**, CpG counts in merged biological replicates per time point. **d**, Genome-wide Spearman's correlation score at the time points analysed shows dynamic CpG modifications. **e**, Functional localisation of DMRs. DMRs were called if the difference in cytosine modification between samples was ≥25% with a p-value of <0.05, using DSS software. **f**, Violin plot of the DMR length distribution (in base pairs) identified in the n=2 biological replicates. Lines and numbers, median. **g**, Density plot indicating the difference in mCpG levels for loss/gain DMRs for each comparison. **h**, Venn diagram showing the overlap between TET1 targets (see Figure 5) that are transcriptionally up-regulated and genes showing either loss (left) or gain (right) of mCpG at the TSS according to the WGBS analyses. Hierarchical clustering analyses of the overlapping genes are presented as heatmaps of TPMs scaled per gene (Z-score).

Extended Data Figure 5: 5hmC levels increase in ductal cells *in vitro* and *in vivo* upon damage

a-c, EpCAM^{+} ductal cells sorted from 0.1% DDC livers (**a**), β1 integrin mutant mice fed with normal chow (undamaged) or DDC **(b)** or WT undamaged livers and grown as organoids **(c).** 5hmC fluorescence intensity was normalised to DAPI. Data are presented as violin plots of the ratio 5hmC/DAPI. Each dot represents the median value (shown in red) of cells counted/mouse. **a**, 353 cells from n=4 undamaged mice, 231 cells from n=5 mice after 3 days of DDC, and 392 cells from n=5 mice at DDC d5; **b,** 138 cells from undamaged, 119 cells at day 1, 247 at day 7 and 125 at day 14 after returning the mice to normal chow (recovery) pooled from 2 livers isolated independently from 2 mice were analysed; **c,** 2500 (0h), 900 (24h) and 2000 (48h) cells from n=3 independent experiments were analysed. *p*-values were calculated using pairwise comparisons with Wilcoxon rank sum test. **a**, d3 *vs* d0 $p= 1x10^{-13}$; d5 *vs* d0 $p< 2.2x10^{-16}$. **c,** 0h *vs* 24h $p< 2.2x10^{-16}$; 48h *vs* 0h $p< 2.2x10^{-16}$. Scale bar, 10μm. **d**, All 5hmC sites identified by RRHP. **e,** Number of genes associated to TSS showing differential 5hmC levels. The number of CpG sites (n) with unique gain of hydroxymethylation is shown. **f**, Graphs represent distribution of percentage of mCpG identified by WGBS in CGI outside TSS (n=32673) using the average of the 2 independent samples (violin plots, black lines median, left) and number of 5hmC counts (median±IQR) in CGI outside TSS (n= 25579) (right). **g,** GO and statistical analyses of the clusters identified in Fig. 4j (Cluster 2 n=347; Cluster 3 n=1659; Cluster 4 n=1424; Cluster 6 n=140) were performed using DAVID 6.8. Heatmap

shows the expression profile of the 84 overlapping genes and is presented as averaged Z score of the 2 biological replicates.

Extended Data Figure 6: TET1 regulates actively transcribed genes in liver organoids

a-d, DamID-sequencing was performed in EpCAM$^+$ sorted ductal cells derived from already established liver organoids (3 independent experiments). Only TET1-Dam peaks identified in all 3 experiments were considered for further analyses. **a**, Scheme of DamID-seq protocol. **b**, Heatmaps showing TET1 peaks identified by DamID-seq (left panels) and H3K4me3 peaks identified by ChIP-seq (right panels). Heatmaps are centred in the middle of the peak (0) and show a genomic window of ±10kb. Top heatmaps represent common peaks between TET1 and H3K4me3 (2848 peaks) while bottom heatmaps represent TET1-specific peaks (2254 peaks). **c**, Pie-chart indicates the percentage of genomic distribution of TET1-Dam peaks. **d**, GO and statistical analyses of biological processes among TET1-Dam targets in liver organoids were performed using DAVID 6.8. n, number of genes. **e,** 5hmC and 5mC levels determined by MeDIP and hMeDIP followed by qPCR on the indicated genomic region surrounding *Lgr5* TSS in *WT* (black), *Tet1*$^{hypo/hypo}$ (blue) and *hypo-OE* (red) organoids. Graphs represent mean of n=3 independent experiments. Student's two tailed was performed comparing samples to WT. *, p<0.05; ** =p <0.01 **f,** TET1 ChIP-qPCR at *Lgr5* TSS (left panel) and *Lgr5* mRNA levels (right panel) in *WT*, *Tet1*$^{hypo/hypo}$ and *hypo-OE* organoids. Graphs represent mean±SD of n=3 independent experiments. Student's two tailed t-test statistical analyses were performed *vs* WT. **, p <0.01 **g**, Sorted EpCAM$^+$ cells from WT livers were cultured in organoid medium and harvested for DNA, chromatin and mRNA expression analyses at the indicated time points. Graphs represent mean of n=3 independent experiments. Student's two tailed t-test analyses were performed *vs* t=0h *, p<0.05; ** =p <0.01; *** =p <0.001

Extended Data Figure 7: Treatment with Rapamycin impairs organoid formation

a, EpCAM$^+$ ductal cells freshly isolated from the undamaged liver were treated at 0-18hrs or 18-48hrs with the indicated small molecule inhibitors. Organoid formation was quantified at day 6. Graph represents organoid formation efficiency and indicates mean ±SD of n=3 independent experiments. Statistical analyses were performed with two-ways ANOVA with Bonferroni's multiple compared test (*vs* DMSO control group). DMSO control quantifications are shown in Fig. 6f. Representative pictures of organoids treated with the inhibitors at 18-48hrs are shown.

Extended Data Figure 8: TET1 hypomorphic mice present a significantly impaired ductal regeneration upon damage.

a, Graph represents mean ±SD of mouse weight of *WT* (n=21 mice), *Tet1*$^{hypo/+}$ (n=13 mice) and *Tet1*$^{hypo/hypo}$ (n=27 mice) littermates. Student's two tailed t-test statistical analyses were performed. **b**, Relative mouse weight of *WT* (n=5), *Tet1*$^{hypo/+}$ (n=1) and *Tet1*$^{hypo/hypo}$ (n=5) mice. **c**, Representative H&E stainings (3 experiments) of intestines from 50 week old *WT* and *Tet1*$^{hypo/hypo}$ mice. Scale bar, 100µm. **d**, Representative H&E stainings (3 experiments) of small intestine from 10 week old *WT* and *Tet1*$^{hypo/hypo}$ mice treated with DDC for 5 days. Scale bar, 100µm. **e-f**, Box-and-whisker plots showing median and IQR of proliferating ductal cells (OPN$^+$/Ki67$^+$) during recovery (n=3 *WT* and n=4 *Tet1*$^{hypo/hypo}$ mice) (**e**) or total ductal cells (OPN$^+$) at the different time points indicated (**f**) (Undamaged, n=3 *WT* and n=3

Tet1$^{hypo/hypo}$ mice ; DDC, n=7 *WT* and n=6 *Tet1*$^{hypo/hypo}$ mice; Recovery, n=3 *WT* and n=4 *Tet1*$^{hypo/hypo}$ mice). Grey dots, outliers from a single counted FOV defined as >1.5 IQR above or below the median. Red squares, median level corresponding to each independent mice. *p*-values obtained by two-sided Kolmogorov-Smirnov test. **g,** Population distribution of the total number of ductal cells (OPN^{+}) Dashed lines show median values obtained from 55 FOV for *WT* (3 mice) and 56 FOV for *Tet1*$^{hypo/hypo}$ (3mice) at day 0 (undamaged) and 110 FOV for *WT* (3 mice) and 153 FOV for *Tet1*$^{hypo/hypo}$ (4 mice) at day 12 (recovery). **h,** PCK immunohistochemistry (3 experiments) from *WT* (left) and *Tet1*$^{hypo/hypo}$ (right) undamaged or in recovery after DDC (day 12) livers. Nucleus, Haematoxylin. Scale bar, 100µm. **i,** *Lgr5* and *Tet1* mRNA levels, TET1 ChIP and hMedIP on *Lgr5* TSS were analysed in undamaged and DDC treated livers. Graphs represent mean±SD of values obtained from n=3 independent biological replicates (dot). *p*-value was calculated using Student's two-tailed t-test.

Extended Data Figure 9: Ductal specific Tet1 conditional deletion impairs duct-mediated liver regeneration

a, Schematic of the *Prom1Cre*^{ERT2}/*Rosa*^{lslZsGreen}/*Tet1*$^{flx/flx}$ mouse model. **b,** Representative immunofluorescence analysis (OPN^{+} red, ZsGreen^{+}, green) of *Prom1*$^{\Delta Tet1/ZsGreen}$ and *Prom1*$^{Tet1WT/ZsGreen}$ upon tamoxifen treatment and injection of AAV8-TBG p21 (2 mice per genotype). Nucleus, Hoechst. Scale bar, 50 µm **c**, Representative immunofluorescence analysis of livers from *Prom1*$^{Tet1WT/ZsGreen}$ mice injected with AAV8-TBG p21 not receiving tamoxifen treatment (2 mice per genotype). Scale bar, 100 µm. **d**, *Tet1* expression in EpCAM^{+}/ZsGreen^{+} ductal cells isolated by FACS from *Prom1*$^{\Delta Tet1/ZsGreen}$ (n=4) or *Prom1*$^{Tet/ZsGreen}$ (n=4) livers derived from mice treated for 3-cycles of DDC and collected 12 days after damage. Graph represents mean±SD of *Tet1* expression expressed as a fold change compared to *Prom1*^{Tet1WT}. Student's two tailed t-test statistical analyses were performed. ***, p<0.001. **e**, Representative pictures of P21 immunohistochemistry analyses. Scale bar, 200 µm. **f**, Weight curves of mice undergoing AAV8-TBG-p21 injection followed by DDC treatment (mean± 95%CI). **g**, TET1 ChIP-qPCR analyses on target genes in ZsGreen^{+}/EpCAM^{+} ductal cells isolated from *Prom1*$^{Tet1WT/ZsGreen}$ DDC-treated livers for 5 days. Cells isolated from 3 mice littermates were pooled used for each independent experiment (n=2). ND, not detected. **h,** Graph represents mean ±SD of mRNA expression of *Tet1* and selected target genes (fold change *vs* WT undamaged) in EpCAM^{+} ductal cells isolated from undamaged (n=2 per genotype) or day 5 DDC-treated livers (n=3 per genotype) derived from *Prom1*$^{TET1WT/ZsGreen}$ (grey) or *Prom1*$^{\Delta Tet1/ZsGreen}$ (blue) mice. Statistical analysis was performed using Student's two-tailed t-test compared to the *Prom1*$^{TET1WT/ZsGreen}$ value at the corresponding time point.

References:

1 Choi, T. Y., Ninov, N., Stainier, D. Y. & Shin, D. Extensive conversion of hepatic biliary epithelial cells to hepatocytes after near total loss of hepatocytes in zebrafish. *Gastroenterology* **146**, 776-788, doi:10.1053/j.gastro.2013.10.019 (2014).

2 Russell, J. O. *et al.* Hepatocyte-specific beta-catenin deletion during severe liver injury provokes cholangiocytes to differentiate into hepatocytes. *Hepatology*, doi:10.1002/hep.30270 (2018).

3 Espanol-Suner, R. *et al.* Liver progenitor cells yield functional hepatocytes in response to chronic liver injury in mice. *Gastroenterology* **143**, 1564-1575 e1567, doi:10.1053/j.gastro.2012.08.024 (2012).

4 Huch, M. *et al.* In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247-250, doi:10.1038/nature11826 (2013).

5 Lu, W. Y. *et al.* Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat Cell Biol* **17**, 971-983, doi:10.1038/ncb3203 (2015).

6 Sackett, S. D. *et al.* Foxl1 is a marker of bipotential hepatic progenitor cells in mice. *Hepatology* **49**, 920-929, doi:10.1002/hep.22705 (2009).

7 Shin, S. *et al.* Foxl1-Cre-marked adult hepatic progenitors have clonogenic and bilineage differentiation potential. *Genes Dev* **25**, 1185-1192, doi:10.1101/gad.2027811 (2011).

8 Deng, X. *et al.* Chronic Liver Injury Induces Conversion of Biliary Epithelial Cells into Hepatocytes. *Cell stem cell* **23**, 114-122 e113, doi:10.1016/j.stem.2018.05.022 (2018).

9 Raven, A. *et al.* Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. *Nature* **547**, 350-354, doi:10.1038/nature23015 (2017).

10 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).

11 Prior, N. *et al.* Lgr5(+) stem and progenitor cells reside at the apex of a heterogeneous embryonic hepatoblast pool. *Development* **146**, doi:10.1242/dev.174557 (2019).

12 Okabe, M. *et al.* Potential hepatic stem cells reside in EpCAM+ cells of normal and injured mouse liver. *Development* **136**, 1951-1960, doi:10.1242/dev.031369 (2009).

13 Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299-312, doi:10.1016/j.cell.2014.11.050 (2015).

14 Li, B. *et al.* Adult Mouse Liver Contains Two Distinct Populations of Cholangiocytes. *Stem Cell Reports*, doi:10.1016/j.stemcr.2017.06.003 (2017).

15 Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**, 812-828, doi:10.1101/gad.234294.113 (2014).

16 Iurlaro, M., von Meyenn, F. & Reik, W. DNA methylation homeostasis in human and mouse development. *Curr Opin Genet Dev* **43**, 101-109, doi:10.1016/j.gde.2017.02.003 (2017).
17 Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-220, doi:10.1038/nrg3354 (2013).
18 Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21, doi:10.1101/gad.947102 (2002).
19 Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133, doi:10.1101/cshperspect.a019133 (2014).
20 Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* **10**, 192-206, doi:10.1038/nrm2640 (2009).
21 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).
22 Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**, 341-356, doi:10.1038/nrm3589 (2013).
23 Hill, P. W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324-333, doi:10.1016/j.ygeno.2014.08.012 (2014).
24 Hahn, M. A., Szabo, P. E. & Pfeifer, G. P. 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* **104**, 314-323, doi:10.1016/j.ygeno.2014.08.015 (2014).
25 Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* **13**, 7-13, doi:10.1038/nrg3080 (2011).
26 Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460-464, doi:10.1038/nature12805 (2013).
27 Hill, P. W. S. *et al.* Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* **555**, 392-396, doi:10.1038/nature25964 (2018).
28 Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402, doi:10.1038/nature10008 (2011).
29 Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370-374, doi:10.1038/nature11925 (2013).
30 Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* **30**, 733-750, doi:10.1101/gad.276568.115 (2016).
31 Kim, R., Sheaffer, K. L., Choi, I., Won, K. J. & Kaestner, K. H. Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation. *Genes Dev* **30**, 2433-2442, doi:10.1101/gad.288035.116 (2016).
32 Reizel, Y. *et al.* Postnatal DNA demethylation and its role in tissue maturation. *Nature communications* **9**, 2040, doi:10.1038/s41467-018-04456-6 (2018).

33 Tarlow, B. D., Finegold, M. J. & Grompe, M. Clonal tracing of Sox9+ liver progenitors in mouse oval cell injury. *Hepatology* **60**, 278-289, doi:10.1002/hep.27084 (2014).
34 Dorrell, C. *et al.* Prospective isolation of a bipotential clonogenic liver progenitor cell in adult mice. *Genes Dev* **25**, 1193-1203, doi:10.1101/gad.2029411 (2011).
35 Mort, R. L. *et al.* Fucci2a: a bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. *Cell Cycle* **13**, 2681-2696, doi:10.4161/15384101.2015.945381 (2014).
36 Duncan, A. W., Dorrell, C. & Grompe, M. Stem cells and liver regeneration. *Gastroenterology* **137**, 466-481, doi:10.1053/j.gastro.2009.05.044 (2009).
37 Medvedeva, Y. A. *et al.* EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* **2015**, bav067, doi:10.1093/database/bav067 (2015).
38 Huch, M. & Koo, B. K. Modeling mouse and human development using organoid cultures. *Development* **142**, 3113-3125, doi:10.1242/dev.118570 (2015).
39 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).
40 Natarajan, A., Wagner, B. & Sibilia, M. The EGF receptor is required for efficient liver regeneration. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17081-17086, doi:10.1073/pnas.0704126104 (2007).
41 Yang, L. *et al.* A single-cell transcriptomic analysis reveals precise pathways and regulatory mechanisms underlying hepatoblast differentiation. *Hepatology* **66**, 1387-1401, doi:10.1002/hep.29353 (2017).
42 Marshall, O. J., Southall, T. D., Cheetham, S. W. & Brand, A. H. Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat Protoc* **11**, 1586-1598, doi:10.1038/nprot.2016.084 (2016).
43 Cheetham, S. W. *et al.* Targeted DamID reveals differential binding of mammalian pluripotency factors. *Development* **145**, doi:10.1242/dev.170209 (2018).
44 Hu, M. *et al.* Wnt/beta-catenin signaling in murine hepatic transit amplifying progenitor cells. *Gastroenterology* **133**, 1579-1591, doi:10.1053/j.gastro.2007.08.036 (2007).
45 Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2(+) cells fuel homeostatic renewal of the liver. *Nature* **524**, 180-185, doi:10.1038/nature14863 (2015).
46 Jensen, K. B. *et al.* Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. *Cell stem cell* **4**, 427-439, doi:10.1016/j.stem.2009.04.014 (2009).
47 Chrysanthou, S. *et al.* A Critical Role of TET1/2 Proteins in Cell-Cycle Progression of Trophoblast Stem Cells. *Stem Cell Reports* **10**, 1355-1368, doi:10.1016/j.stemcr.2018.02.014 (2018).

48 Boj, S. F. *et al.* Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* **151**, 1595-1607, doi:10.1016/j.cell.2012.10.053 (2012).
49 Fouraschen, S. M. *et al.* mTOR signaling in liver regeneration: Rapamycin combined with growth factor treatment. *World journal of transplantation* **3**, 36-47, doi:10.5500/wjt.v3.i3.36 (2013).
50 Planas-Paz, L. *et al.* YAP, but Not RSPO-LGR4/5, Signaling in Biliary Epithelial Cells Promotes a Ductular Reaction in Response to Liver Injury. *Cell stem cell* **25**, 39-53 e10, doi:10.1016/j.stem.2019.04.005 (2019).
51 Talarmin, H. *et al.* The mitogen-activated protein kinase kinase/extracellular signal-regulated kinase cascade activation is a key signalling pathway involved in the regulation of G(1) phase progression in proliferating hepatocytes. *Molecular and cellular biology* **19**, 6003-6011, doi:10.1128/mcb.19.9.6003 (1999).
52 Pepe-Mooney, B. J. *et al.* Single-Cell Analysis of the Liver Epithelium Reveals Dynamic Heterogeneity and an Essential Role for YAP in Homeostasis and Regeneration. *Cell stem cell* **25**, 23-38 e28, doi:10.1016/j.stem.2019.04.004 (2019).
53 Yimlamai, D. *et al.* Hippo pathway activity influences liver cell fate. *Cell* **157**, 1324-1338, doi:10.1016/j.cell.2014.03.060 (2014).
54 Serra, D. *et al.* Self-organization and symmetry breaking in intestinal organoid development. *Nature* **569**, 66-72, doi:10.1038/s41586-019-1146-y (2019).
55 Lin, Y. *et al.* HGF/R-spondin1 rescues liver dysfunction through the induction of Lgr5(+) liver stem cells. *Nature communications* **8**, 1175, doi:10.1038/s41467-017-01341-6 (2017).
56 Kamimoto, K. *et al.* Heterogeneity and stochastic growth regulation of biliary epithelial cells dictate dynamic epithelial tissue remodeling. *eLife* **5**, doi:10.7554/eLife.15034 (2016).
57 Zhu, L. *et al.* Multi-organ Mapping of Cancer Risk. *Cell* **166**, 1132-1146 e1137, doi:10.1016/j.cell.2016.07.045 (2016).
58 Blanpain, C. & Fuchs, E. Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science* **344**, 1242281, doi:10.1126/science.1242281 (2014).
59 Lin, S. *et al.* Distributed hepatocytes expressing telomerase repopulate the liver in homeostasis and injury. *Nature* **556**, 244-248, doi:10.1038/s41586-018-0004-7 (2018).
60 Font-Burgada, J. *et al.* Hybrid Periportal Hepatocytes Regenerate the Injured Liver without Giving Rise to Cancer. *Cell* **162**, 766-779, doi:10.1016/j.cell.2015.07.026 (2015).
61 Huch, M. & Dolle, L. The plastic cellular states of liver cells: Are EpCAM and Lgr5 fit for purpose? *Hepatology* **64**, 652-662, doi:10.1002/hep.28469 (2016).
62 Michalopoulos, G. K. The liver is a peculiar organ when it comes to stem cells. *Am J Pathol* **184**, 1263-1267, doi:10.1016/j.ajpath.2014.02.020 (2014).
63 Forbes, S. J. & Rosenthal, N. Preparing the ground for tissue regeneration: from mechanism to therapy. *Nat Med* **20**, 857-869, doi:10.1038/nm.3653 (2014).

64 Hall, C. *et al.* Regulators of Cholangiocyte Proliferation. *Gene Expr* **17**, 155-171, doi:10.3727/105221616X692568 (2017).
65 Lazaridis, K. N. & LaRusso, N. F. The Cholangiopathies. *Mayo Clin Proc* **90**, 791-800, doi:10.1016/j.mayocp.2015.03.017 (2015).
66 Tanaka, E. M. & Reddien, P. W. The cellular basis for animal regeneration. *Developmental cell* **21**, 172-185, doi:10.1016/j.devcel.2011.06.016 (2011).
67 Jin, S. G. *et al.* 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer research* **71**, 7360-7365, doi:10.1158/0008-5472.CAN-11-2023 (2011).
68 Thomson, J. P. *et al.* Loss of Tet1-Associated 5-Hydroxymethylcytosine Is Concomitant with Aberrant Promoter Hypermethylation in Liver Cancer. *Cancer research* **76**, 3097-3108, doi:10.1158/0008-5472.CAN-15-1910 (2016).
69 Kafer, G. R. *et al.* 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *Cell reports* **14**, 1283-1292, doi:10.1016/j.celrep.2016.01.035 (2016).

Methods:

Liver isolation, FACS sorting, culture and transfection

Undamaged or DDC damaged livers were isolated from 8-12 weeks old mice and digested using collagenase/dispase (0.125mg/ml in DMEM/F12) as previously published[4,70]. To obtain a population of ductal cells, cells were stained with CD11b-PE/Cy7 (BD Biosciences), CD45-PE/Cy7 (BD Biosciences), CD31-PE/Cy7 (Abcam) and EpCAM-APC antibody (eBioscience) (Supplementary Dataset 5) and FACS-sorted using a MoFlo cell sorter. The following sorting strategy was pursued: single cells were sequentially gated based on cell size (forward scatter, FSC, versus side scatter, SSC) and singlets (pulse width vs FSC) and then ductal cells were selected based on EpCAM positivity after excluding macrophages (CD11+), blood cells (CD45+) and endothelial cells (CD31+), hence obtaining a pure population of single $CD11b^-/CD31^-/CD45^-/EpCAM^+$, named $EpCAM^+$ from here on (see Supplementary Dataset 5 for antibody list). In order to determine the phase of the cell-cycle, $EpCAM^+$ cells derived from *R26Fucci2a* mice[35] were further gated for mCherry-hCtd1(30/120) (G1/G0) and mVenus-hGem(1/110) (S/G2/M).

Sorted cells were seeded in matrigel and cultured in Advanced DMEM/F12 (Gibco) supplemented with Penicillin/Streptomycin, Glutamax and HEPES (all from Gibco), and 1xN2 (Gibco), 1xB27 (Gibco), 500nM n-Acetylcysteine (Sigma), 10nM Gastrin (Sigma), 50ng/ml EGF (Peprotech), 100ng/ml FGF10 (Peprotech), 50ng/ml HGF (Peprotech), 10mM Nicotinamide (Sigma), 10% R-spondin1 conditioned medium

(home-made), 25ng/ml Noggin (Peprotech), 30% Wnt conditioned medium (home-made) and 10uM of Rock inhibitor Y-27632 (Sigma), as previously reported[4,70]. After 48h, Wnt conditioned medium, Noggin and Rock inhibitor were removed from the culture medium.

Organoid formation efficiency from ductal sorted cells was determined by seeding single cells at 500 cells/well and counting organoid numbers/well 6 days later. Organoid formation efficiency from organoid cultures was determined by dissociating organoids into single cells following TrypleE (Gibco) incubation for 10min. Cells were then seeded at 500 cells/well and number of organoids/well was counted after 6 days.

For experiments with small molecule inhibitors, $EpCAM^+$ ductal cells were grown for the time and concentration indicated in the legend with Verteporfin (Tocris, 5305), Gefinitib (Stratech Scientific, S1025), PD0325901 (Sigma, PZ0162), Rapamycin (Sigma, R8781) and PD173074 (Tocris, 3044).

For EdU studies, a 24h pulse of EdU was performed in FACS-sorted $EpCAM^+$ cells at 0-24h, 24-48h, 48-72h after seeding. Cells were then cytospun, using to Click-iT EdU Alexa Fluor 594 Imaging Kit (Molecular Probes) according to manufacturer's instructions.

For treatment of Cre^{ERT2} positive cells *in vitro*, 5mM of (Z)-4-Hydroxytamoxifen (Sigma) was added o/n to the medium.

To perform siRNA experiments, $1x10^4$ $EpCAM^+$ cells freshly isolated from undamaged livers were transfected either with a pool of 4 ON-Targetplus siRNA (Dharmacon) for each candidate gene (screen) or with 4 independent *Tet1* siRNA, using Lipofectamine RNAimax (Life Technologies) according to manufacturer's instructions. Briefly, cells and Lipofectamine-RNA mix were spun at 600g at 32°C or 45min and then incubated 4h at 37C. Cell suspension was then collected and seeded in matrigel in Isolation medium. Organoid formation efficiency was assessed 6 days later. siRNAs used are listed in Supplementary Dataset 5.

To generate stable organoid lines, ectopically expressing full-length hTET1 cDNA (TET1wt) or catalytically inactive TET1 (H1671Y, D1673A) (TET1 cat.mut.) reported in[29,39]. Cells were transfected into $5x10^4$ $CD11b^-/CD31^-/CD45^-/EpCAM^+$ freshly isolated cells using Lipofectime 2000 (Life Technologies) according to manufacturer's instructions. Briefly, cells and Lipofectamine-DNA mix were spun at 600g at 32°C for 45min and then incubated 4h at 37C. Cell suspension was then

collected, spun for 5min at 300g and seeded in matrigel in Isolation medium for 48h and then switched to expansion medium. Blasticidin (2µg/ml) was added 48h after transfection in expansion medium in order to maintain stable expression of the transgenes. Organoid formation efficiency was assessed 6 days later.

hTET1 cDNA was cloned into a Mammalian Targeted DamID vector with a CAG-promoter downstream of LT3-Dam[43] to create a Dam-hTET1 fusion protein using Gibson assembly. Organoids were transiently transfected as described above with either a Dam-only or Dam-hTET1 fusion construct, together with pCAG-Venus at a 3:1 ratio. Around 5×10^3, 2.5×10^4 and 4.5×10^4 $Venus^+$ cells were FACS-sorted 72h later and processed independently for each the 3 biological replicates respectively.

5hmC/EdU immunocytochemistry

EdU/5hmC staining was performed in FACS-sorted cells fixed in 4% paraformaldehyde (PFA) at time (0h) or embedded in matrigel and cultured in Isolation Medium for 24h and 48h. For the latter two conditions, EdU was added to the medium at 10µM for a 24h pulse (namely from 0-24h and from 24-48h). Cells were fixed with 4% PFA within the matrigel bubble and extracted by washing with cold Advanced DMEM/F12. Cells were cytospun onto SuperFrost Plus slides (VWR) and stained using the Click-iT EdU Alexa Fluor 594 Imaging Kit (Molecular Probes) according to manufacturer's instructions. Cells were permeabilised and blocked simultaneously with PBS containing 1% Triton, 1% DMSO, 1% BSA and 2% donkey serum for 25min, after which they were treated with 2N HCl at RT for 20min. Following thorough washes with PBS, the cells were incubated with the 5hmC primary antibody (Active Motif, 39769, Supplementary Dataset 5) at 1:1000 dilution in permeabilisation/blocking buffer pre-diluted 1:100 in PBS. After washing with PBS, cells were incubated for 90min with an anti-rabbit Alexa 488 secondary antibody at a 1:250 dilution in PBS containing 0.05% BSA. Nuclei were counterstained with DAPI at 0.5µg/ml in water. Cover slips were mounted with Vectashield (Vector Laboratories) and sealed with nailpolish. The quantification of the intensity of 5hmC levels was performed with an in-house designed macro for the Fiji software [71]; EdU+ cells were counted manually.

Confocal images were captured on a Leica SP5 inverted confocal microscope, (LAS AF) and processed with Volocity 6.3 (PerkinElmer). The quantification of the

intensity of 5mC levels was performed with an in-house designed macro for the Fiji software.

Whole Genome Bisulfite Sequencing (WGBS) library preparation and analysis

Ductal cells were isolated from 2 independent 8-12 weeks old mice per time point and split for libraries preparation. High-molecular-weight genomic DNA (HMW gDNA) was extracted by first lysing the cells in lysis buffer [20mM Tris pH8, 4mM EDTA, 20mM NaCl, 1%SDS] and proteinase K [ThermoFisher, #EO0491] for 90min at 56ºC), then by treating lysate with RNase A (#EN0531, ThermoFisher) for 5min at RT. gDNA was extracted using Phenol/chloroform (Phenol:Chloroform:Isoamyl Alcohol 25:24:1, Invitrogen #15593031). HMW gDNA was then sonicated to fragment size of 300-400bp (Covaris, E220), following manufacturer's instructions. Fragments were then purified using PureLink PCR Purification kit (ThermoFisher, #K310001) and purity and length were determined using NanoDrop, Qubit and Tapestation (Agilent). To estimate bisulfite conversion efficiency, un-methylated lambda phage cl857 Sam7 DNA was used as spike-in (0.5% of total DNA amount). Typically, NGS libraries were prepped using 200ng of sonicated fragments using NEBNext Ultra II DNA Library Prep, following manufacturer's instructions (New England BioLabs, E7645S). Briefly, blunt fragments are first end-repaired and A-tailed using T4 DNA polymerase and Klenow Fragment. They are then ligated on both flanks with Illumina methylated adaptors (NEB, E7535S). Adapted fragments were then purified with Agencourt AMPure Beads at a 0.8x ratio (Beckman Coulter, Inc). Libraries (~50ng) were then treated with sodium bisulfite according to the protocol (Imprint DNA Modification Kit; Sigma, MOD50) and then barcoded (NEBNext Multiplex Oligos for Illumina, NEB E7335S) and amplified by PCR (8 PCR cycles) with KAPA HiFi HS Uracil+ RM (KAPA Biosystems). Indexed libraries were sequenced on HiSeq 4000 (High Output mode, v.4 SBS chemistry, at CRUK, Cambridge Institute, UK) to generate 150bp paired reads.

Quality of sequenced read pairs was determined, and adaptor sequences and low quality reads removed using TrimGalore `--paired --fastqc –illumina` (v0.4.4_dev, Babraham Inst.). Adaptor-trimmed paired reads were aligned to the mouse assembly GRCm38.p6 and to the lambda genome (used to determine bisulfite non-conversion rate) using Bismark[27] (v0.19.0). Single-end reads from un-aligned paired-end reads were also mapped. Alignment parameters were: 1 mapping mismatch

allowed with a maximum insert size for valid paired-end read alignments of 500bp (`-N 1 –X 500`, respectively). Clonal reads were removed using `deduplicate_bismark`. Methylation at CpG sites was called using `bismark_methylation_extractor –no_overlap --ignore_r2 2` and `methylpy call-methylation-state` using merged paired end and single end reads. Biological duplicates were merged for overall methylome profiling and only CpG sites with ≥3 unique mapped reads were used for analyses. DMRs (default parameters with DNA methylation differences of ≥25% between groups) were generated using R package DSS (v2.26.0) and CGI for GRCm38.p6 were predicted as previously published[72]. For analysis of cytosine modifications (5mC and 5hmC) on up-regulated genes *in vivo*, levels of modification at CpG context (mCpG) were averaged over TSS regions (+/- 500bp) of genes (given as percentage of mCpG). WGBS analyses, including gene expression correlation and Spearman correlation, were done using DSS (v2.26.0), methylpy (v1.2.9), R (v3.4.4), using custom scripts. Samtools (v1.5) and bedtools (v2.26.0) were used to generate and analyse mapped reads. R packages ggplot2 (v2.3.0) and pheatmap (v1.0.10) were used to visualise data.

Reduced Representation of Hydroxymethylation Profile (RRHP) library preparation and analysis

Genomic DNA from was extracted using phenol/chloroform as described in WGBS section. Quality and purity of gDNA was assessed using Nanodrop and Tapestation. 500ng of gDNA was then used to produce RRHP data according to manufacter's instructions (Zymo Research D5450). RRHP libraries were multiplexed and sequenced on HiSeq 4000 at Gurdon Institute (single end 50bp).

Adaptor sequences in sequenced reads and low-quality reads were removed using trimGalore (0.4.4_dev, options: `--rrbs --fastqc --illumina`). Trimmed reads were then mapped to GRCm38.p6 mouse genome using bowtie2 (version 2.3.3.1, options: `--end-to-end`). Then, only reads with the 5′-CCGG tag were further analysed. Unique CCGG sites were counted genome-wide and only 5hmC sites present in both biological replicates were analysed. To identify differential hydroxymethylation levels, TSS regions (+2,-1 kbp around TSS) showing ≥4 unique 5hmC sites at any time points were analysed. Heatmaps of 5hmC show scaled values of absolute 5hmC count at TSS present in both biological replicates for each group. The non-parametric Kruskal-Wallis test by ranks was used to compare 5hmC levels

between groups, followed by Dunn's test to correct for multiple testing (adjusted p-values of <0.05 were considered significant). Statistical tests were performed using Prism (v8.0).

Time-lapse microscopy and Image analysis

Time-lapse images of single FACS-sorted ductal cells were acquired on an inverted spinning disk confocal microscope (3i Intelligent Imaging Innovations) with a Zeiss LD C-Apochromat 40x (1.1 numerical aperture, NA) immersion objective. Cells were imaged in a humidified chamber with 5% CO_2. Images were taken at intervals of 60min with a Z-step of 2.4µm. Time-lapse acquisition were processed with the Slidebook6 software and analysed by Fiji image processing software, as previously described[73].

RNA-sequencing and analyses

$EpCAM^+$ freshly isolated cells were isolated from undamaged or DDC-damaged livers (day 3 and 5) for RNA extraction or embedded in matrigel and collected at different time points after culture (time 12h, 24h, 48h and 6 days, the later named as organoids). The starting time point (0h) was collected after seeding in matrigel but prior to adding any medium. Total RNA was extracted using PicoPure RNA isolation kit (ThermoFisher Scientific) according to manufacturer's instructions. RNA libraries were prepared by using Smartseq2[74]. RNA sequencing was performed using Illumina Hiseq sequencer at the Gurdon Institute. Quality of sequenced read pairs was determined, adaptor sequences and low quality reads removed using `TrimGalore --paired --fastqc --illumina` (v0.4.4). Reads were mapped and quantified (TPM) using kallisto v0.43.1 (`kallisto quant --bias --single -b 100 -l 500 -s 80 -t 1`) with the mouse assembly GRCm38.p6. Differential gene expression was performed using sleuth (v0.29.0; sleuth_lrt: likelihood ratio test) with FDR <0.1, with sequencing batch effect adjustment. Only DE genes with a maximal TPM of >5 at >1 time point (T0, D3 or D5) and showing considerable expression level difference (|b|>0.58) were analysed. Principal component analysis (centered and scaled) were produced with build-in R programme `prcomp`. Mean TPM values between biological replicates of each group were used for downstream analysis. Graphs and heatmaps were produced with R packages ggplot2, rgl and pheatmap. Heatmap of gene

expression are scaled TPM values (z score). The list of epigenetic regulators was found in the database Epifactors[37].

Organoid whole mount immunostaining

Organoids removal from matrigel was performed by mechanically disrupting the matrigel by gently pipetting 5 times and then incubated for 10min in cold Advanced DMEM/F12 medium. Cells were fixed with PFA 4% for 30min on ice and incubated with blocking solution (PBS 1%Triton 1%BSA in PBS) for 1.5 hours. For 5hmC staining, organoids were treated for 20min with 2N HCl before incubation with the primary antibody followed by 3 washes in 1% blocking solution. Primary antibodies were incubated o/n at 4°C Fluorophore-conjugated secondary antibodies were then incubated for 2h. Nuclei were stained with DAPI. Confocal images were captured on a Leica SP5 inverted confocal microscope and processed with Leica LasX software.

Quantitative PCR (qPCR)

Total RNA was extracted from freshly isolated $EpCAM^+$ cells or cultured in isolation or expansion medium (organoid cultures) using PicoPure RNA isolation kit (ThermoFisher Scientific) according to manufacturer instructions. cDNA was synthesized using 50-250ng of total RNA and a M-MLV Reverse Transcriptase kit (Promega). cDNA was amplified with iTaq™ Universal SYBR Green Supermix (BioRad) and specific primers (see Supplementary Dataset 5). All targets were amplified (40 cycles) on a CFX96 Real-Time qPCR Detection System (Biorad). Ct values were analyzed using BioRad CFX manager. Expression levels were normalized to the expression of the housekeeping gene *Hprt.*

5mC and 5hmC DNA immunoprecipitation (MeDIP and hMeDIP) and chromatin immunoprecipitation (ChIP)

Genomic DNA was extracted by incubating either sorted $EpCAM^+$ cells (freshly isolated or cultured in matrigel in isolation or expansion medium) or dissociated organoids with lysis buffer (50mM Tris HCl PH8.0, 50mM NaCl, 5mM EDTA, 1.0% SDS) for 10min at 4°C Genomic DNA was sheared by sonication (Bioruptor, Diagenode) to 200-500bp average fragments and purified using MinElute PCR purification micro kit (Qiagen). The following immunoprecipitation buffer was used: 50mM Tris HCl PH8, 250mM NaCl, 5mM EDTA, 0,5% Triton, 0,10% SDS. TET1

(Millipore 09-872 and ABE1034) and H3K4me3 (Diagenode C1541003) primary antibodies (Supplementary Dataset 5) were incubated o/n and immunocomplexes were recovered with DiaMag Protein A or G coated magnetic beads (Diagenode). ChIP-qPCR experiments were then performed on genomic regions of interest and data were normalised to IgG and expressed as percentage of input material (see Supplementary Dataset 5 for primer list). Chromatin Immunoprecipitation (ChIP) experiments from cells freshly isolated from liver tissue were performed using LowCell ChIP kit (Diagenode) followed by iPure kit v2 (Diagenode) according to manufacture's instructions. ChIP experiments for cells grown *in vitro* as organoid were performed as previously described[75]. DNA derived from 2 independent biological H3K4me3 ChIP experiments in liver organoids was pulled and processed using Thruplex DNA seq kit according to manufacturer's instructions. The library was sequenced in-house using an Illumina Hiseq 1500. H3K4me3 ChIP-seq reads were mapped to GRCM38/mm10 with bowtie2 (v2.2.9). Peak calling on H3K4me3 ChIP-seq was done using MACS2 (v2.1.0) (broad, q<0.01) on individual bam-files, with input as a control. Genes were called from peaks using GREAT (v3.0.0)[76] (single nearest gene, +/-2kb from TSS). Plots were generated through SeqPlots (v1.12.1)[77].

DamID sequencing

Cells obtained by trypsin-mediated dissociation of mouse liver organoids were transfected with hTET1-Dam or Dam-only vector together with Venus-enconding plasmid (see above). DamID-seq on Venus$^+$ FACS-sorted cells was performed as previously described[42]. All sequencing experiments were performed as single-end 50bp reads generated by the Gurdon Institute NGS Core using an Illumina HiSeq 1500. DamID sequencing data from three paired replicates were mapped to GRCM38/mm10 and processed using the damidseq_pipeline script[78], with default settings apart from a 300 bin-width. Peak-calling on DamID samples was done using MACS2 (v2.1.0) (broad, q<0.01) on individual bam-files, with Dam-only as control. Peak files from all replicates were merged and intersected with every single replicate with bedtools (v2.25.0) to obtain those peak-regions, which were only present in all three replicates with a q-value <0.01. Genes were called from peaks using GREAT (v3.0.0)[76] (single nearest gene, +/-2kb from TSS). Plots were generated through SeqPlots (v1.12.1)[77]. Peaks were annotated to overlapping genomic features with the

ChIPseeker-package (v1.18.0) in R using annotations and gene IDs from TxDb.Mmusculus.UCSC.mm10.knownGene and org.Mm.eg.db (v3.4.4). DamID data are presented as Dam-hTETt1/Dam ratios with the midline at 1.

Western blot assay

Cell lysates were prepared using RIPA buffer (10mM Tris-Hcl PH 8.0m 1mM EDTA, 1% Triton X-100, 0.1% SDS. 150mM NaCl) supplemented with proteinase inhibitor cocktail (Roche) and sonicated for 5min using a Bioruptor sonicator (Diagenode). Lysates were cleared by centrifugation at 13,000rpm for 15min. Samples were loaded on Precast Mini Protean TGX gels (Biorad) and transferred on nitrocellulose membrane (Biorad), which was blocked in 5% milk and incubated O.N. with TET1 (Millipore 09-872) (1:1000) or actin (Abcam ab3280) (1:2000) (Supplementary Dataset 5). Then, anti-rabbit or mouse horseradish peroxidase (HRP) conjugated secondary antibodies were used and antibody-protein complexes were visualised using ECL (GE-Healthcare). Bands intensities were quantified using Fiji software. All antibodies used are listed in Supplementary Dataset 5.

Mouse line generation and maintenance

All mouse experiments have been regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). All animal experiments have been performed in adult (>8 weeks-old) mice. Both female and male mice were used. The *R26Fucci2a* mouse line was generated from a cross between the previously described *R26Fucci2aR*[35] and Cre745 mouse lines (a kind gift from DJ Kleinjan, University of Edinburgh). Cre745 mice ubiquitously express Cre-recombinase under the control of the CAAG promoter[79], the resulting progeny of this cross therefore had permanently excised the STOP cassette in the *R26Fucci2aR* locus resulting in ubiquitous expression of Fucci2a. The CAAG-Cre transgene was crossed away in subsequent matings. The "knock-out first" *Tet1*$^{tm1(KOMP)Wtsi}$ mouse line (named as *Tet1*$^{hypo/hypo}$ line) was obtained from the International Knockout Mouse consortium (IKMC). To ensure mouse fertility the line was initially crossed with MF1 mice to generate a C57/Bl6xMF1 mixed background. All further generated mice were inbred within this line to maintain this mixed background. In order to generate the *Tet1* conditional (*Tet1*$^{flx/flx}$) mouse line, *Tet1*$^{tm1(KOMP)Wtsi}$ mice were bred with a

ubiquitously expressing Flipase recombinase (*Rosa26FLPe),* taking advantage of built-in *frt* sites, leaving only *loxP* sites flanking exon 4 of *Tet1* (NM_001253857). For TET1 conditional deletion, *Tet1*$^{flx/flx}$ mice were bred with either the ubiquitous *Rosa26Cre*^{ERT2} (JAX lab, ROSA26Sor$^{tm1(cre/ERT2)Tyj}$/J) or the ductal specific *Prom1Cre*^{ERT2} Cre drivers[57]. Deletion of exon 4 was induced in 8-10 weeks old mice by 3 intraperitoneal injections of tamoxifen, performed at 48h intervals, at a dose of either 4mg or 0.2 mg/g (see specific experimental scheme for dosage) diluted in sunflower oil.

DDC treatment

For Acute DDC treatment, adult mice were transferred to wheat-free cages and fed with food supplemented with 0.1% DDC (Custom Animal diets, LLC). The diet was provided *ad libitum* for the duration of the experiment (2-5 days) or switched back to normal chow after 5 days. Mice were allowed to recover for 7 days.
Chronic DDC experiments were performed by supplementing the diet with 0.1% DDC for 5-days for eight cycles, with a 3-day interval of normal diet between cycles. All mice were euthanized by exposure to CO_2, and, when required, blood was collected by cardiac puncture. Serum was then submitted for analysis at the Department of Veterinary Medicine, University of Cambridge.

P21 overexpression and DDC treatment

Lineage tracing and *Tet1* deletion in *Prom1Cre*^{ERT2}/Rosa^{IslZsGreen}/*Tet1*$^{flx\text{-}flx}$ mice was induced as described above. After one week, $7.5x10^{11}$ viral particles of AAV8-TBG-p21 were injected intravenously (by tail vein injection) as previously described[9]. After a week of wash out (allowing for a combined two-week wash out from tamoxifen injections), the mice were fed chow supplemented 0.1% DDC *ad libitum* for 5 days in three cycles each with 3 days of recovery in between DDC treatments. After 2 weeks recovery time after the final DDC dose, mice were culled and analysed.

Histology and immunohistochemistry

Livers and small intestines were removed and fixed in 10% neutral buffered formalin (Sigma-Aldrich) overnight at room temperature. After fixation tissues were frozen in O.C.T compound (VWR chemicals) or embedded in paraffin (Thermo Scientific) and

processed for analysis. For Haematoxylin and Eosin (H&E) and pan-cytokeratin staining, 5μm paraffin embedded sections were deparaffinised by washing in Xylene followed by descending concentrations of ethanol (100%, 95%, 70% 50%). Slides were then either stained for standard H&E analysis or immunostained with a pan-cytokeratin (PCK) antibody (Supplementary Dataset 5). For the PCK immunostaining, following de-paraffination and hydration, endogenous peroxidase activity was blocked by incubating sections in 3% H_2O_2:MeOH for 15min at room temperature. Antigen retrieval was performed by incubating the sections in 800 units/ul Proteinase K diluted in Tris-EDTA (pH 8) and 0.5% triton X-100 for 10min at 37°C. Sections were allowed to cool to room temperate before being blocked in blocking buffer [2% Normal goat serum, 1% Bovine serum albumin, 0.1% triton X-100 in TBS] for 1h at room temperature. Primary antibody was diluted in 1:100 dilution of the blocking buffer and incubated o/n at 4C. Antibody detection was carried out using the Bright-DAB system (ImmunoLogic) following manufacturers' protocol. Sections were counterstained in haematoxylin and mounted in DPX mounting medium.

When stated, 5um paraffin sections were stained with Picro-sirius Red according to manufacturer instructions (Abcam, ab150681). Fibrotic area was calculated using a Fiji Image macro developed in house.

Immunofluorescence

Liver sections (50-100μm thick) were cut from O.C.T. embedded samples and washed in PBS twice. Sections were blocked in 1% Triton X-100 (Sigma-Aldrich), 5% DMSO (Sigma-Aldrich), 2% Donkey Serum (Sigma-Aldrich) for 16h at 4C and incubated with the primary antibody (see Supplementary Dataset 5) diluted in 0.5% Triton X-100, 1% DMSO, 2% Donkey Serum for 72h at 4C. Sections were washed thoroughly over 24h with 0.5% Triton X-100 and 1% DMSO. Appropriate fluorophore conjugated secondary antibodies were diluted in 0.5% Triton X-100, 1% DMSO and 2% Donkey Serum and incubated on sections for 48h at 4C. The sections were washed in PBS and incubated with Hoechst 33342 diluted 1:1000 in PBS for 1h at room temperature. Finally, sections were incubated in ascending glycerol concentrations (10%, 30%, 50%, 70% 90%) for 1h each and then mounted in Vectashield (Vector Laboratories). Stained sections were imaged using a Leica SP5 confocal microscopy and analysed in Fiji.

Code availability:

All codes used are available upon request.

Statistics and Reproducibility

Statistical analyses are described in detail for each panel. Briefly, statistical analyses of ChIP-qPCR, RT-qPCR, cell-culture experiments were performed using Prism 6 software. Student two tailed t-test or two ways ANOVA combined with Tukey HSD test were used according to the experiment. Statistical analyses used for identification of DMRs based on WGBS we re identified using DSS. RRHP and immunostainings were performed using R (v3.4.4). Kruskal Wallis test with Dunns multiple comparisons was used a statistical test for RRHP. Statistical analyses of immunofluorescence data were performed using Wilcoxon rank sum test. Population distributions of the proliferative ductal populations ($Ki67^{+}/OPN^{+}$ cells) as well as the total ductal populations (OPN^{+}) were compared between genotypes using two sided Kolmogorov-Smirnov test. Distributions were then visualised using box and whisker plots and histograms with kernel density estimate values overlaid. Histogram bin sizes were determined by splitting the total data range into 30 bins of equal size. Frequency density was then calculated by taking the number of counts within each bin and dividing it by bin size and total number of counts in the group.

For peak calling of DamID-sequencing and ChIP-sequencing experiments statistics were performed using MACS2. DE genes in the RNA-sequencing were called using Sleuth with Wald test. Statistical analyses of expression of TET1 targets were performed with one-way ANOVA followed by Tukey's multiple comparisons test. GO analyses were perfomed using DAVID 6.8. n size of the samples was indicated for each panel. No data points were removed. All experiments presented were reproducible.

Data availability:

RNA, ChIP, DamID, WGBS and RRHP sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE123133.

All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Supplementary datasets:

Movie1: Time lapse movie of EpCAM+ ductal cells FACS-sorted from undamaged R26Fucci2a mouse embedded in matrigel and grown in organoid culture conditions for 72h

Supplementary Dataset 1: RNA-sequencing data

Supplementary Dataset 2: WGBS and RRHP data

Supplementary Dataset 3: TET1-DamID sequencing and H3K4me3 ChIP-sequencing data

Supplementary Dataset 4: List of DE genes *in vivo* and merge with TET1 targets, WGBS and RRHP

Supplementary Dataset 5: List of antibodies, primers and siRNA sequenc

References:

70 Broutier L. *et al.* Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat Protoc* **11**, 1724-1743, doi:10.1038/nprot.2016.097 (2016).

71 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

72 Wu, H., Caffo, B., Jaffee, H. A., Irizarry, R. A. & Feinberg, A. P. Redefining CpG islands using hidden Markov models. *Biostatistics* **11**, 499-514, doi:10.1093/biostatistics/kxq005 (2010).

73 Antonica, F., Orietti, L. C., Mort, R. L. & Zernicka-Goetz, M. Concerted cell divisions in embryonic visceral endoderm guide anterior visceral endoderm migration. *Developmental biology* **450**, 132-140, doi:10.1016/j.ydbio.2019.03.016 (2019).

74 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

75 Aloia, L. *et al.* Zrf1 is required to establish and maintain neural progenitor identity. *Genes Dev* **28**, 182-197, doi:10.1101/gad.228510.113 (2014).

76 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

77 Stempor, P. & Ahringer, J. SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res* **1**, 14, doi:10.12688/wellcomeopenres.10004.1 (2016).

78 Marshall, O. J. & Brand, A. H. damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371-3373, doi:10.1093/bioinformatics/btv386 (2015).

79 Kleinjan, D. A. *et al.* Long-range downstream enhancers are essential for Pax6 expression. *Developmental biology* **299**, 563-581, doi:10.1016/j.ydbio.2006.08.060 (2006).

Figure 1_Aloia et al.

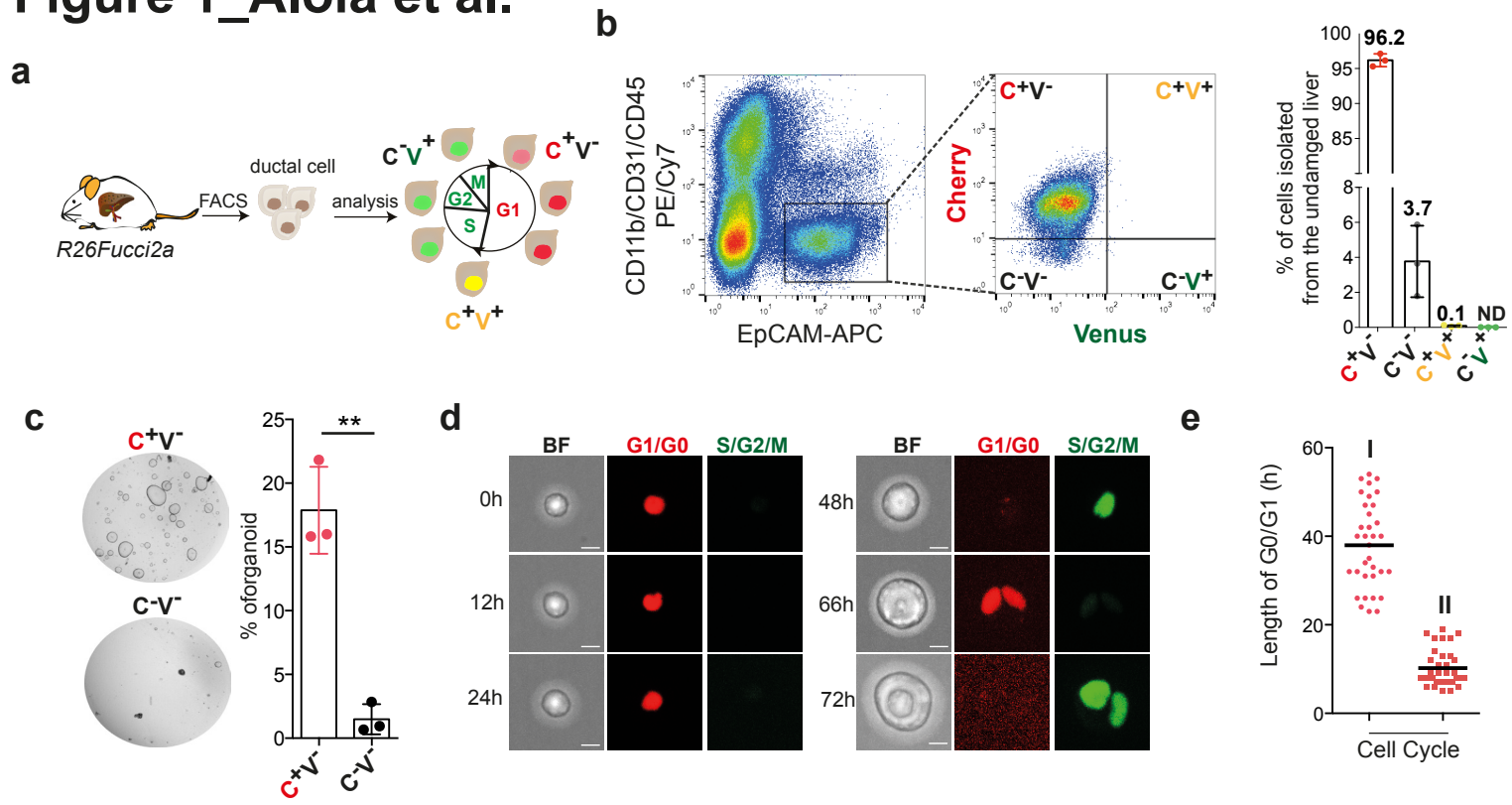


Figure 2_Aloia et al.

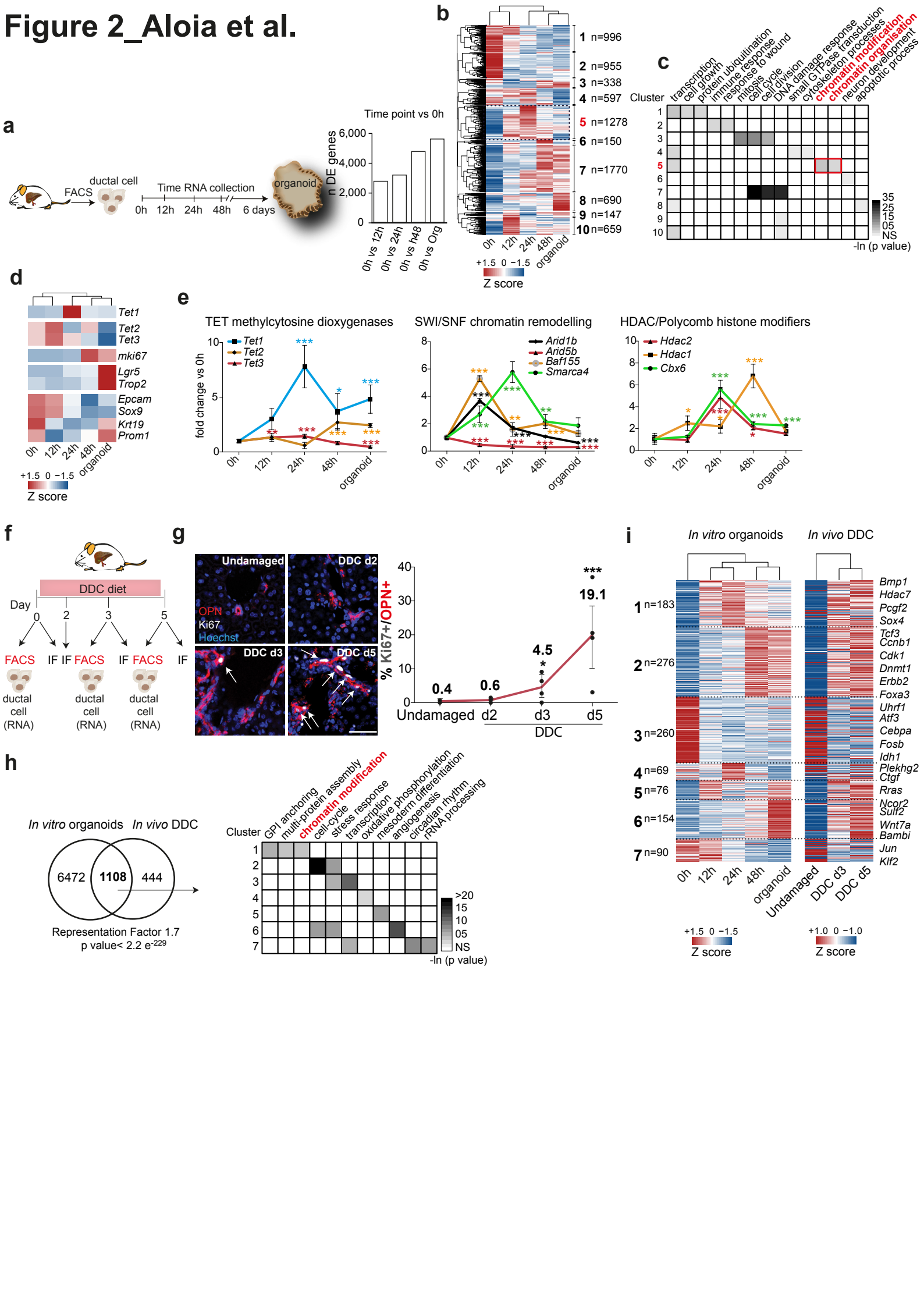


Figure 3_Aloia et al.

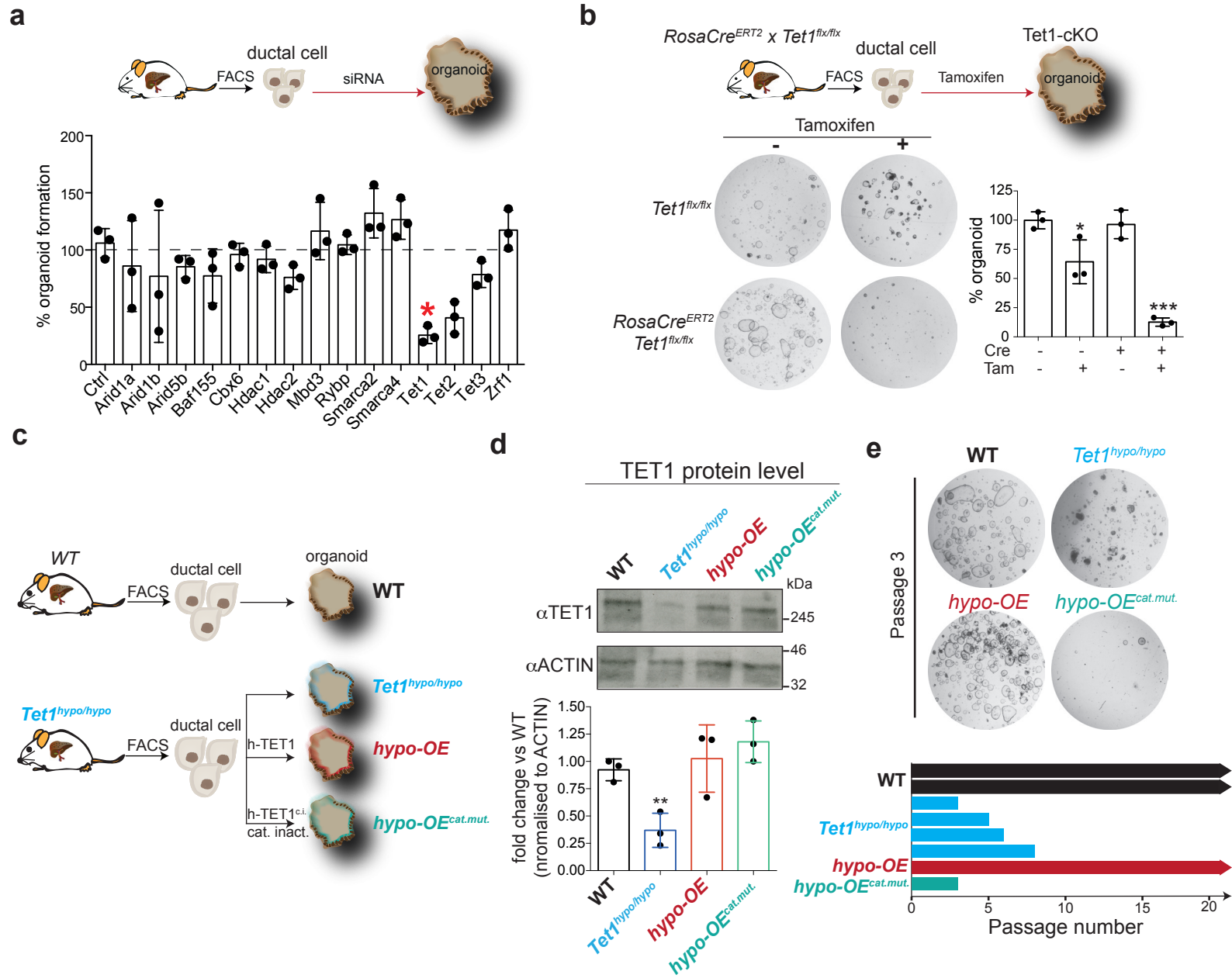


Figure 4_Aloia et al.

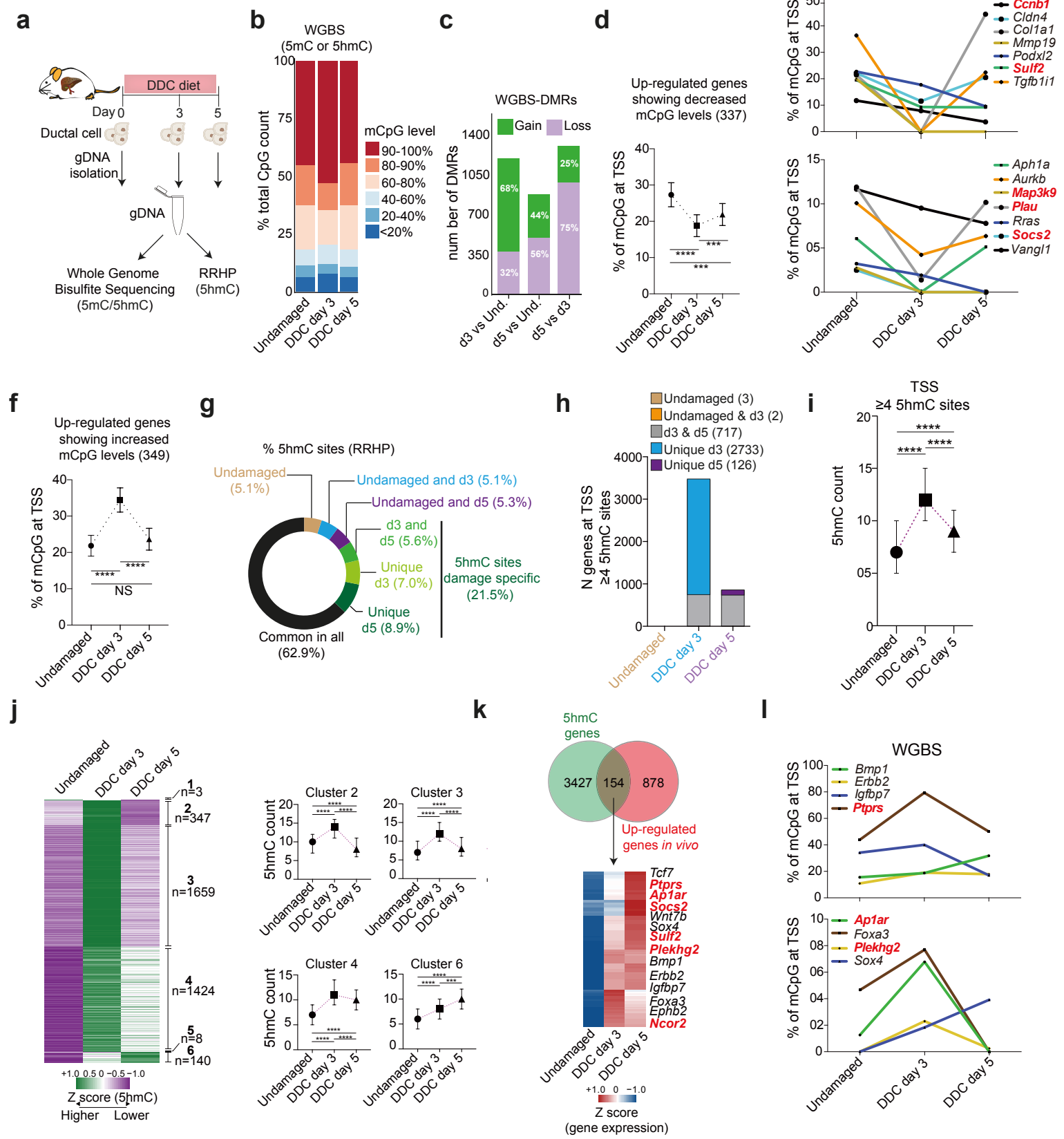


Figure 5_Aloia et al.

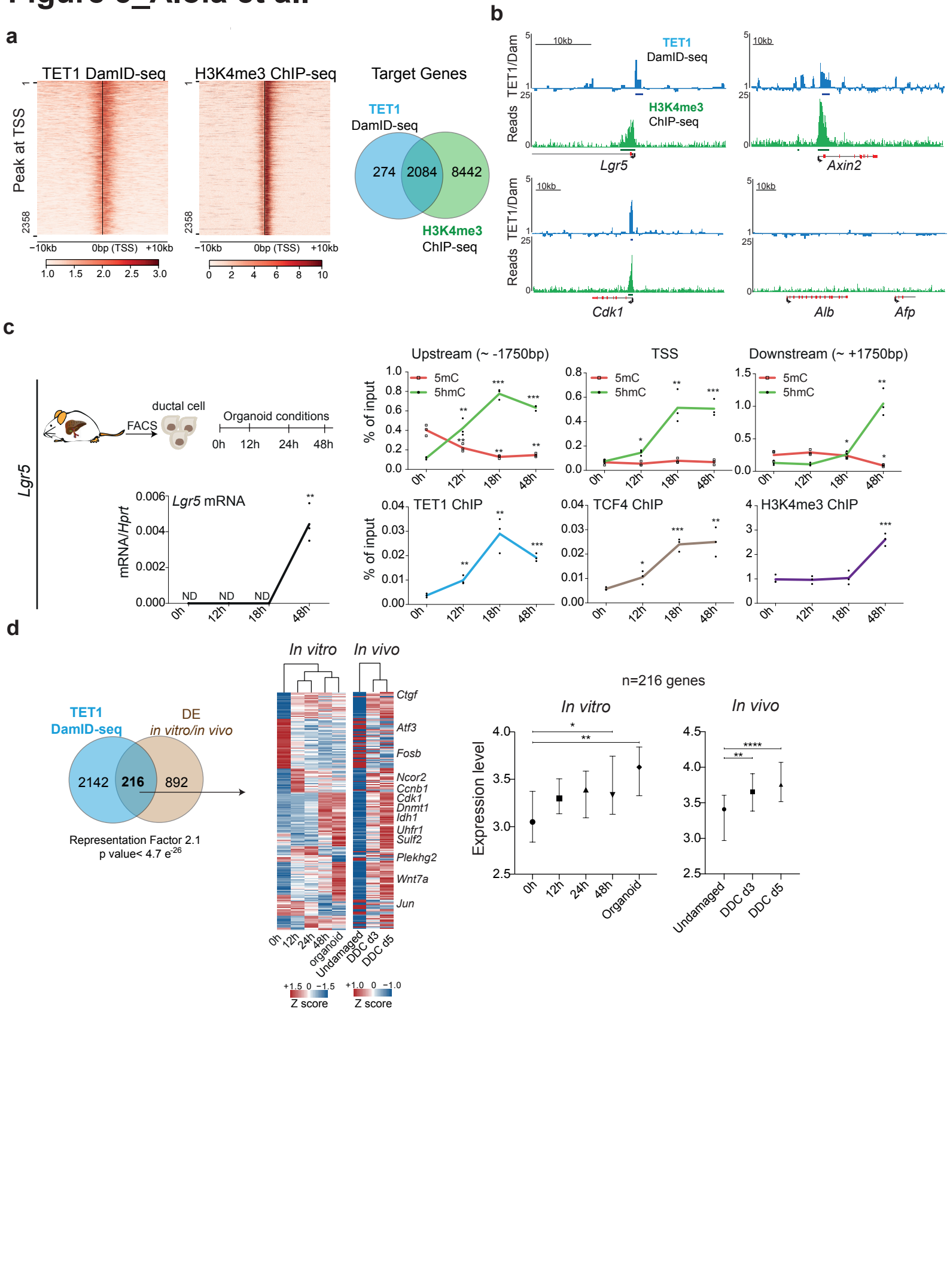


Figure 6_Aloia et al.

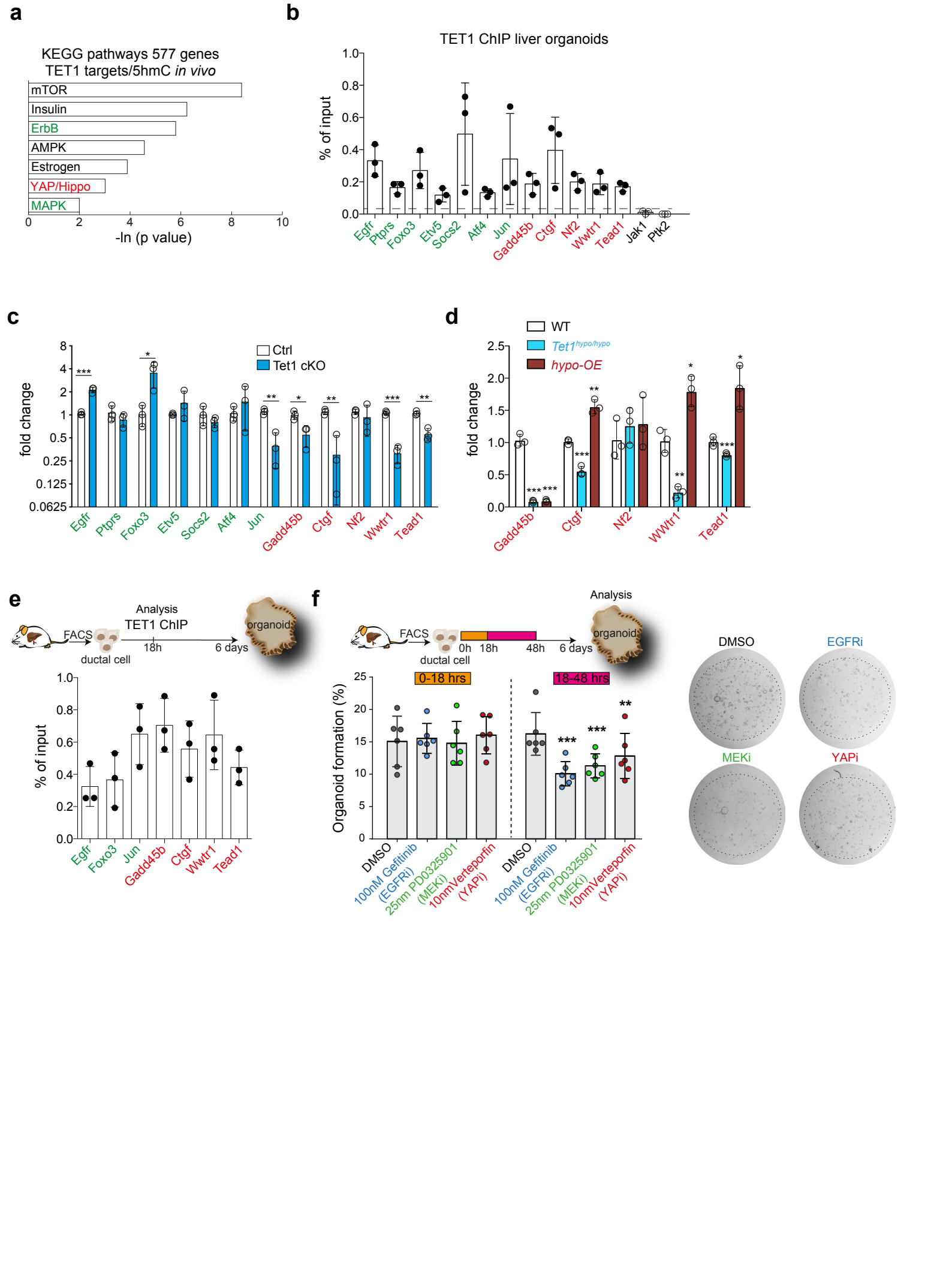


Figure 7_Aloia et al.

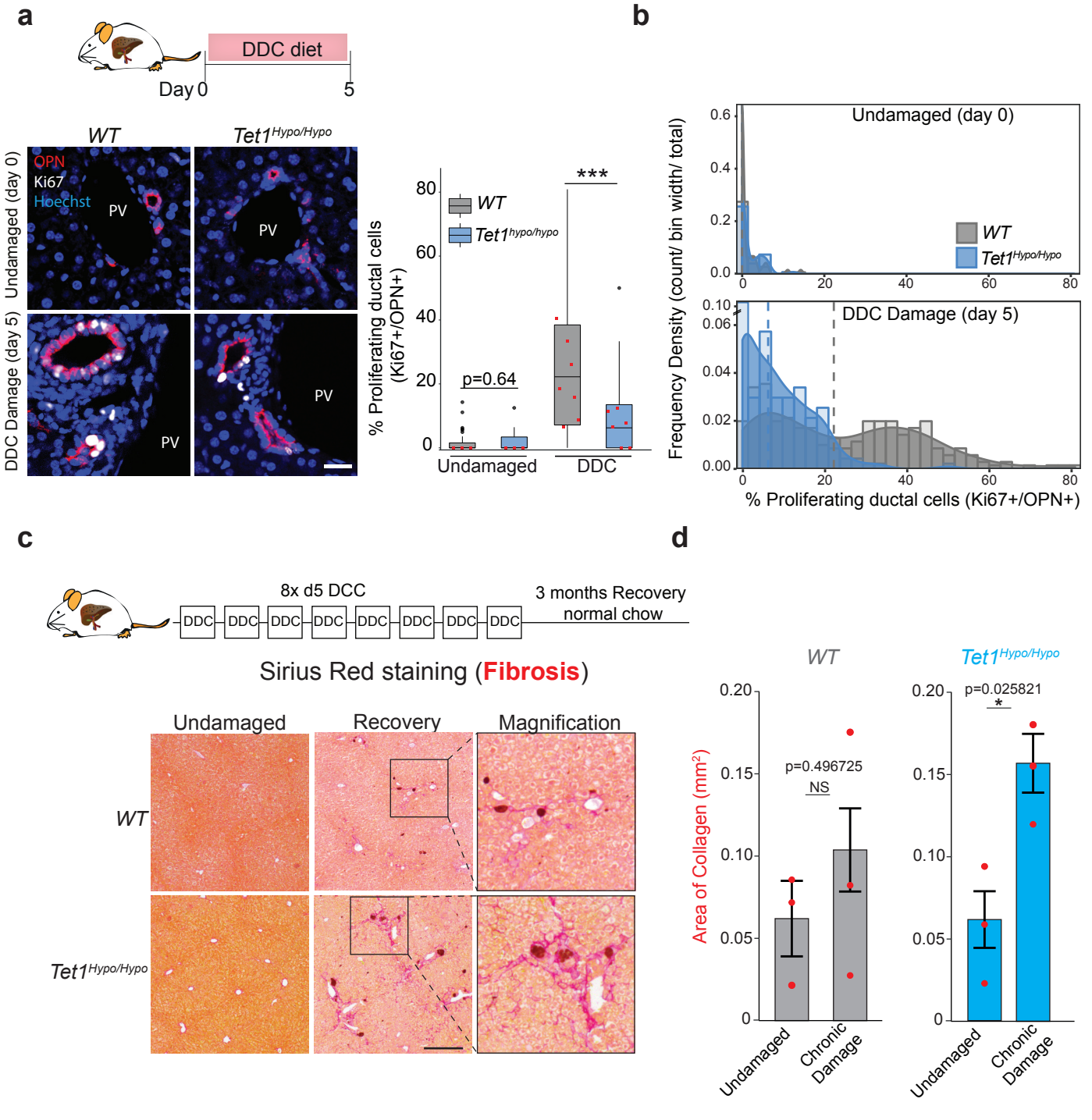
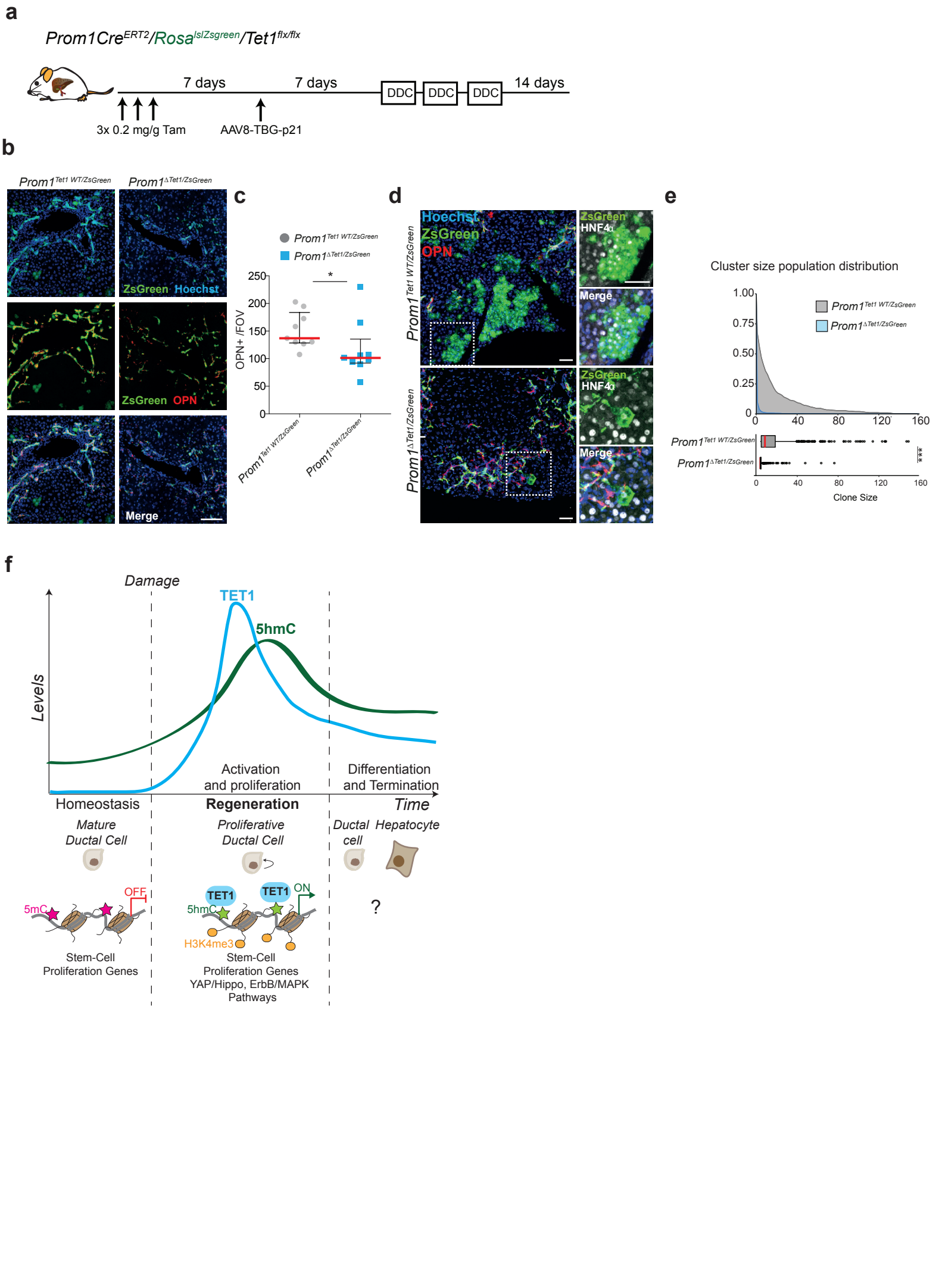
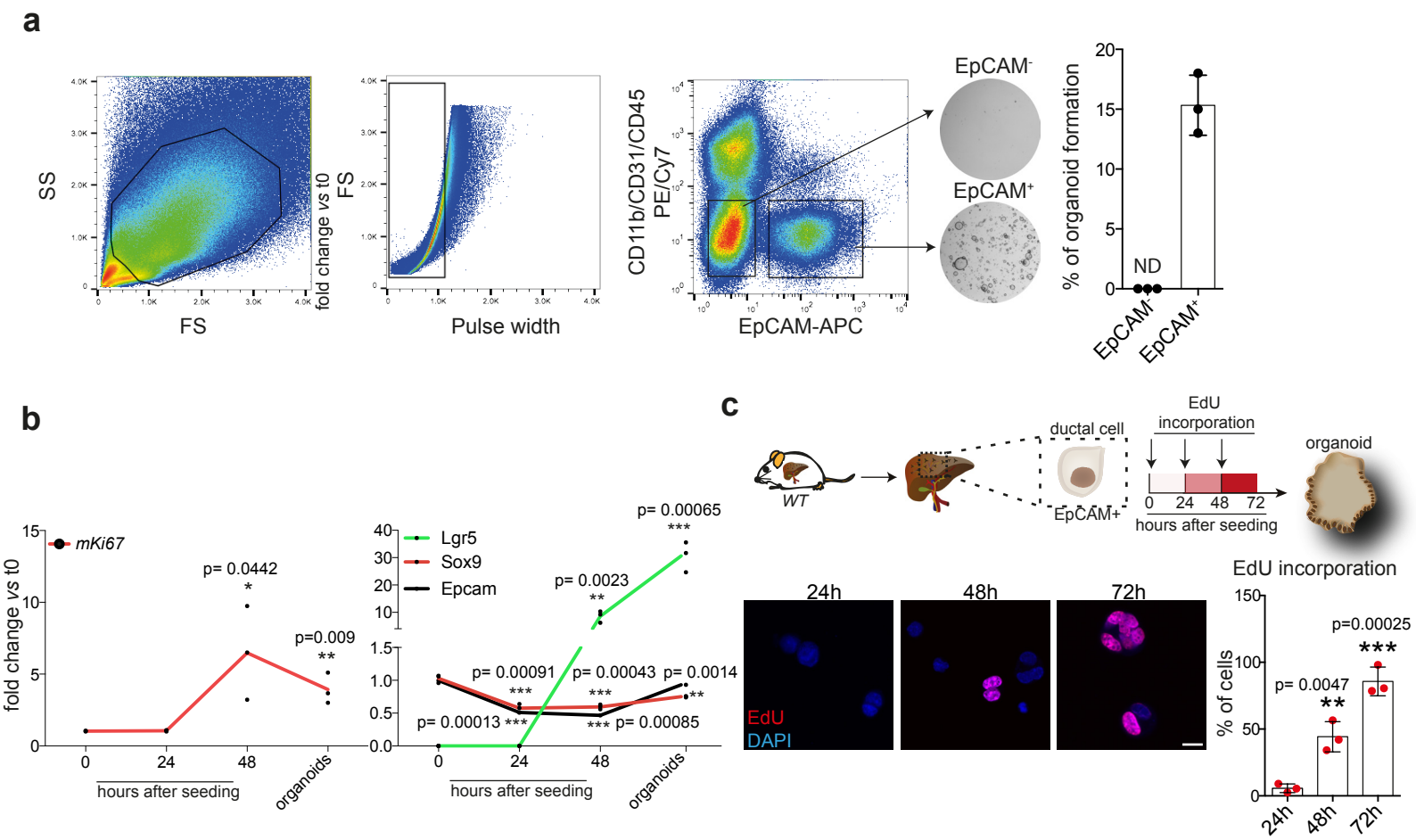


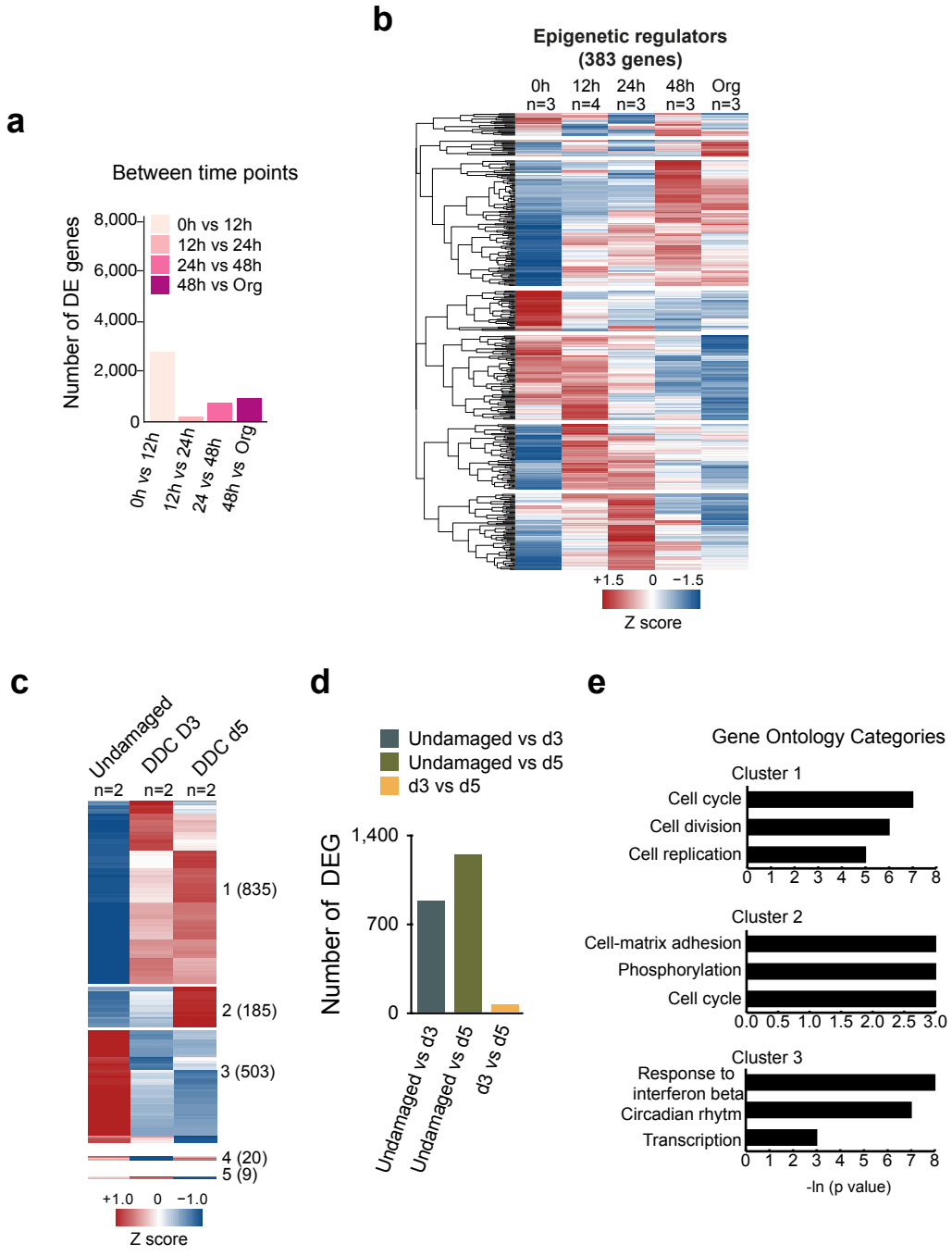
Figure 8_Aloia et al.



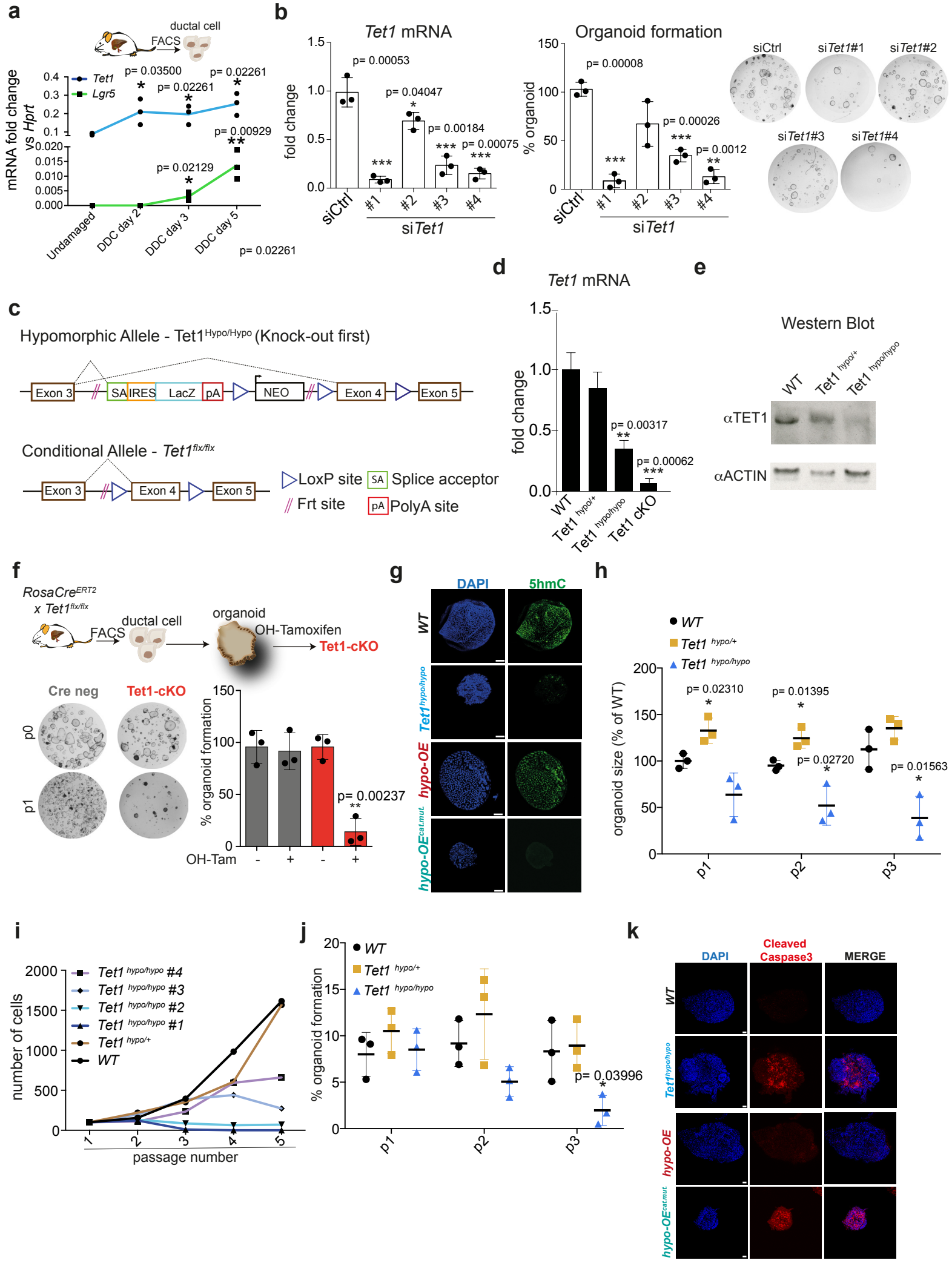
Aloia_Extended Data Figure 01



Aloia_Extended Data Figure 02

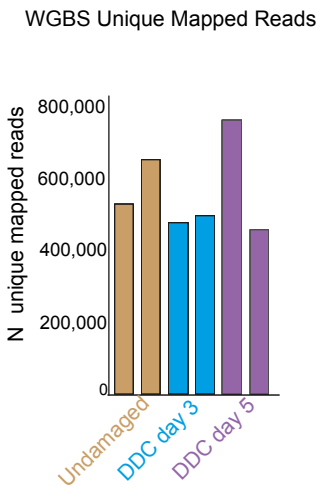


Aloia_Extended Data Figure 03

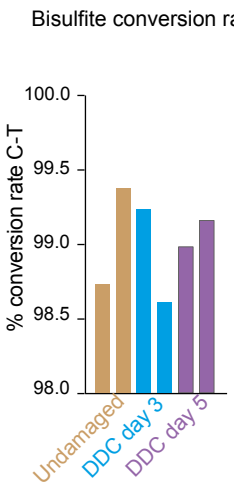


Aloia_Extended Data Figure 04

a



b

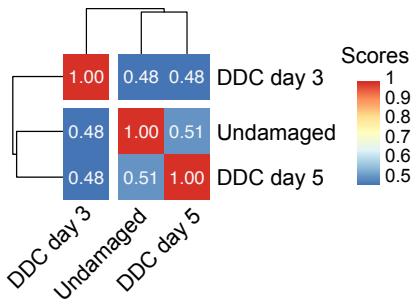


c

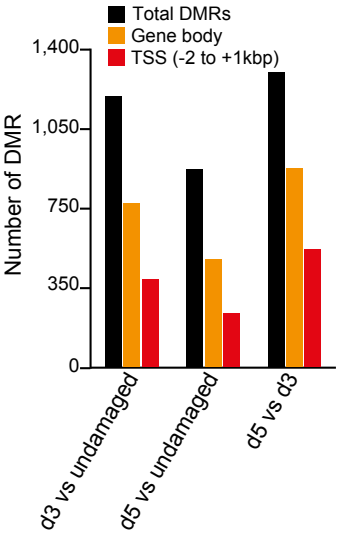
CpG count (merged n=2)

	Und. (n=2)	d3 (n=2)	d5 (n=2)	Shared
All	29,328,807	28,317,933	29,776,416	
≥3 cov	22,347,108	18,723,512	23,844,822	18,456,635

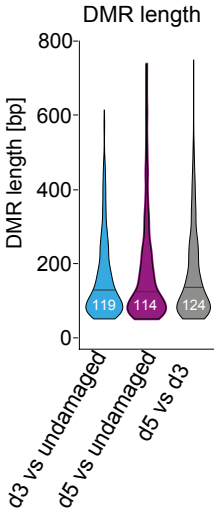
d



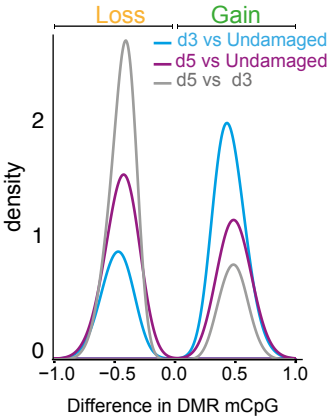
e



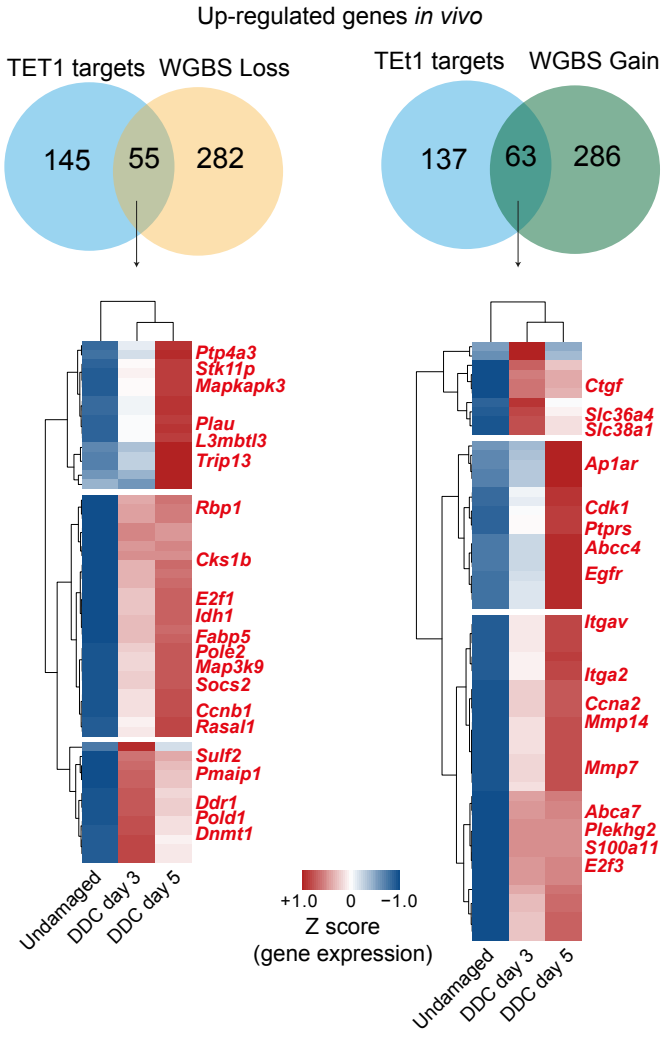
f



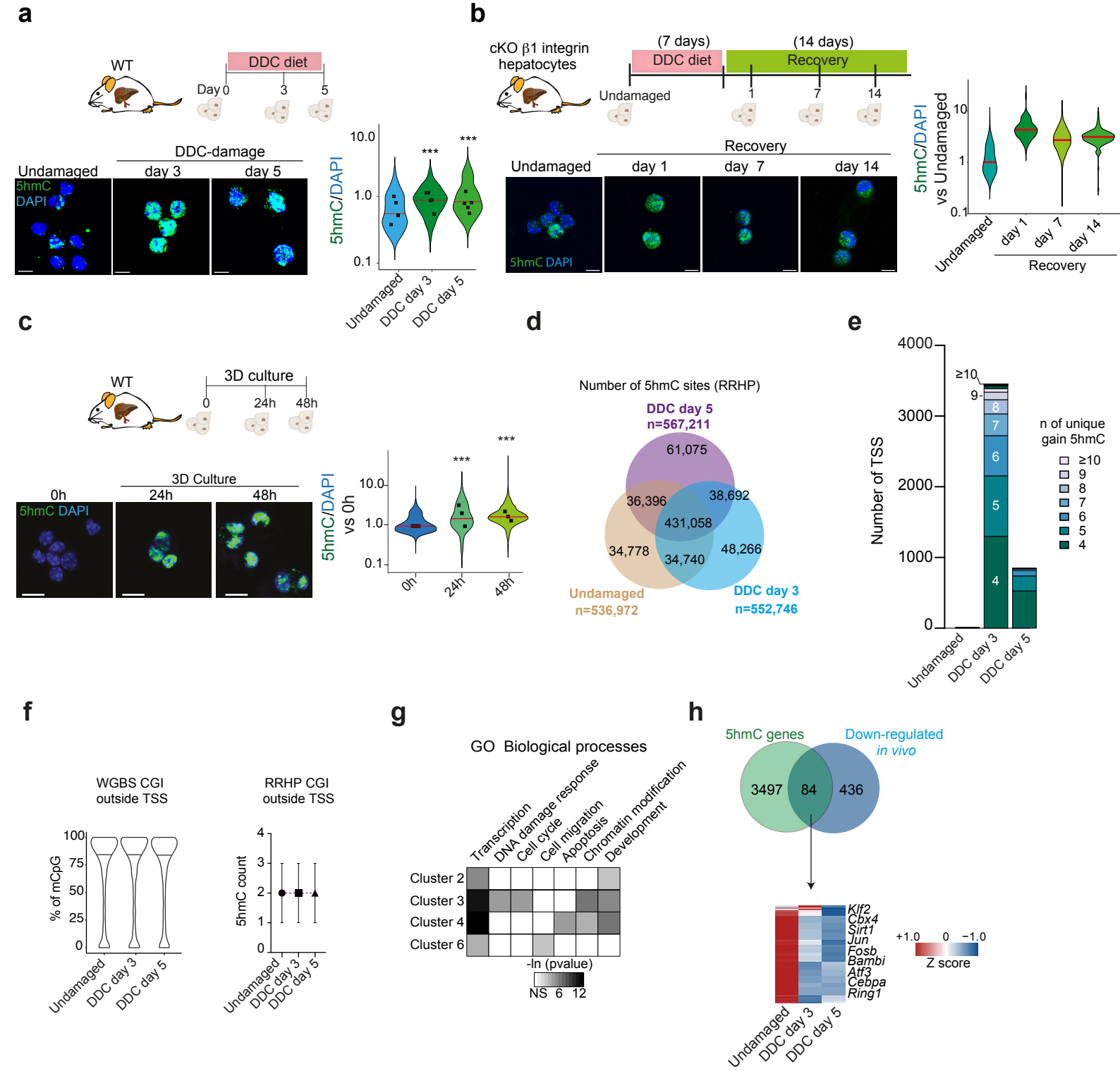
g



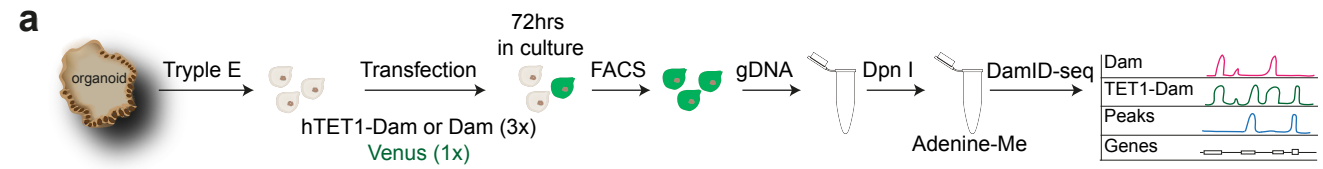
h

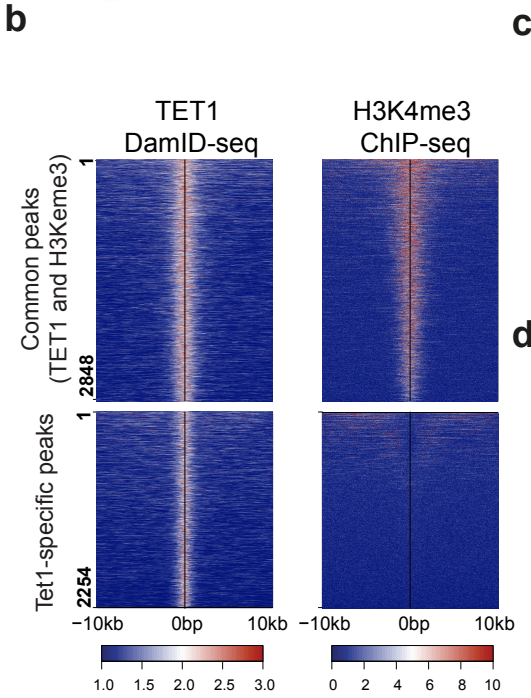


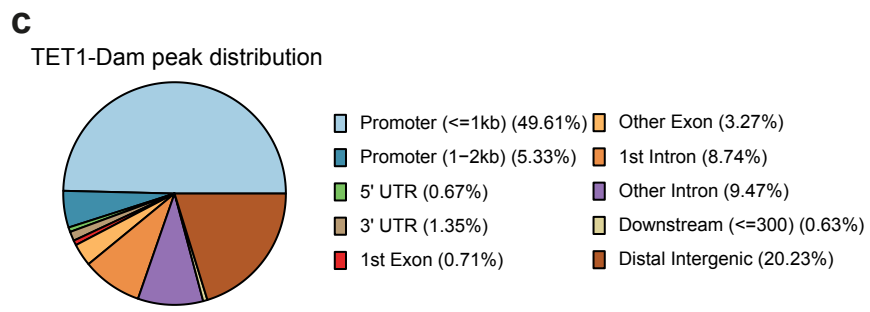
Aloia_Extended Data Figure 05

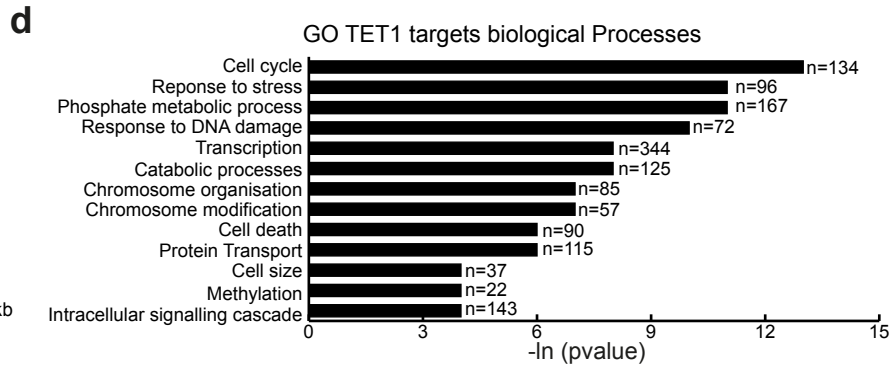


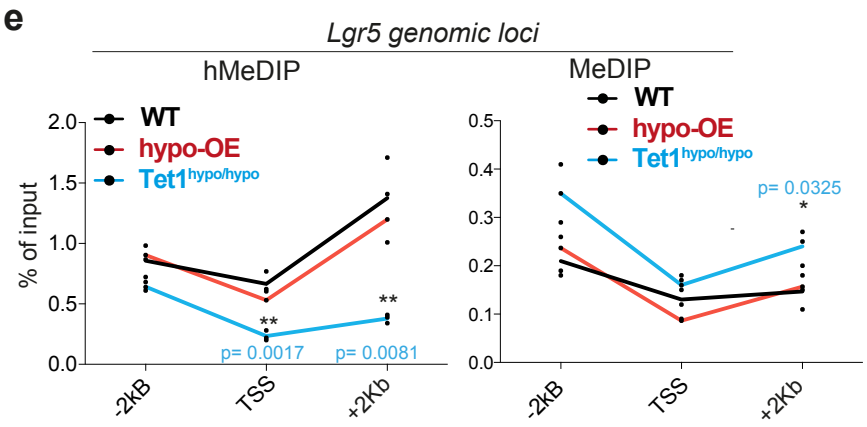
Aloia_Extended Data Figure 06

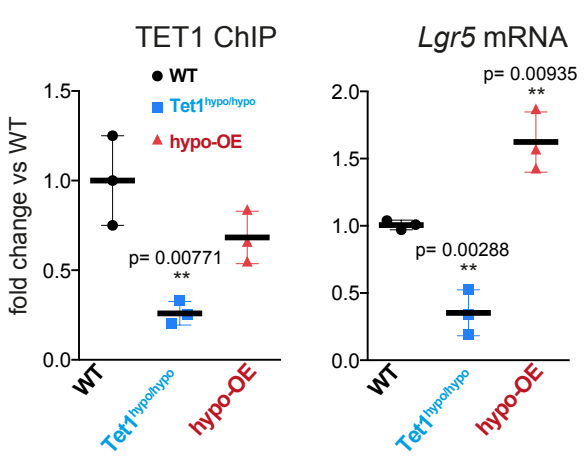


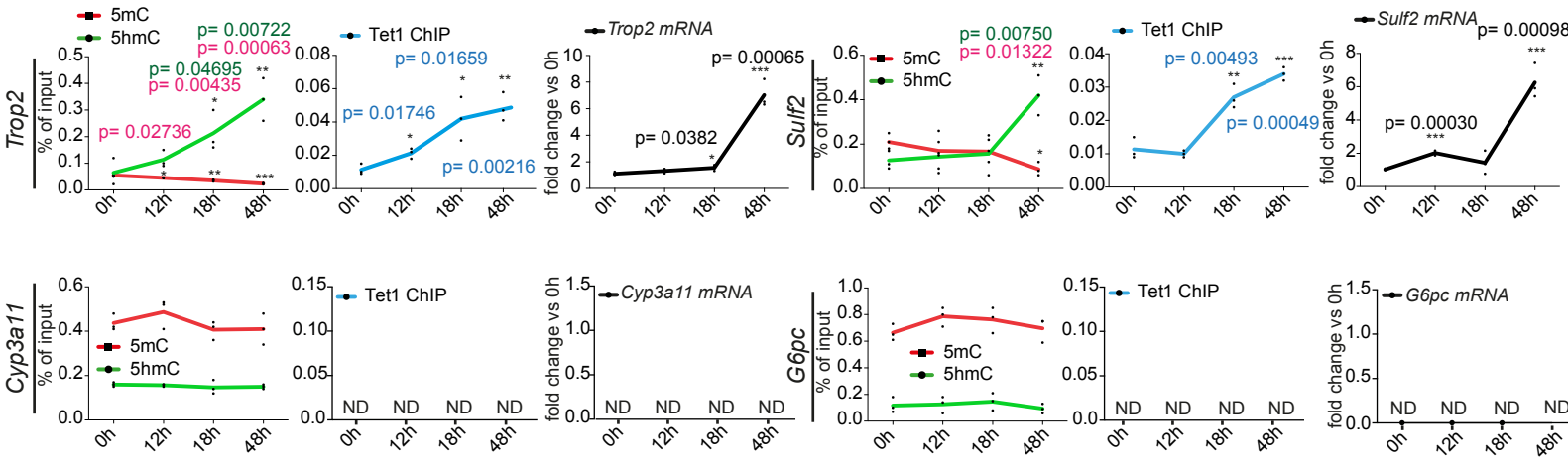






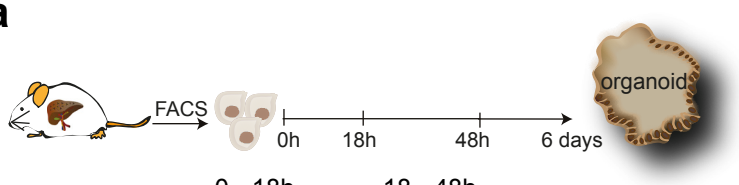


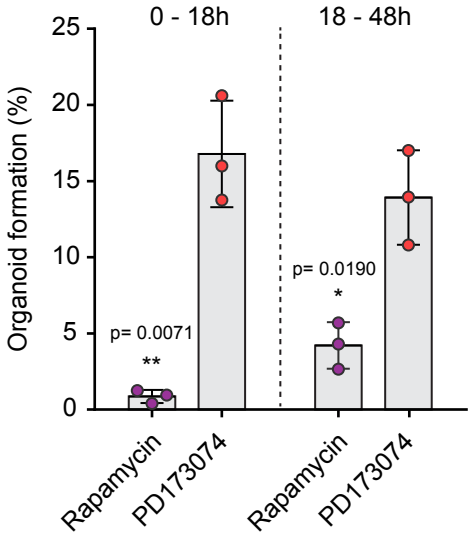


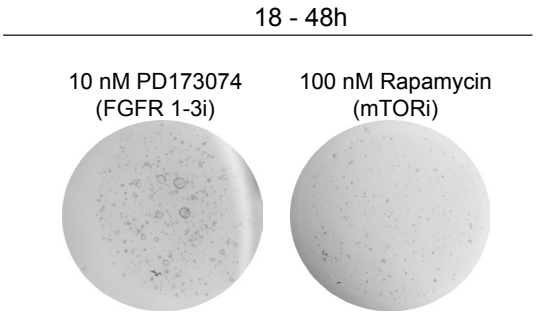


Aloia_Extended Data Figure 07

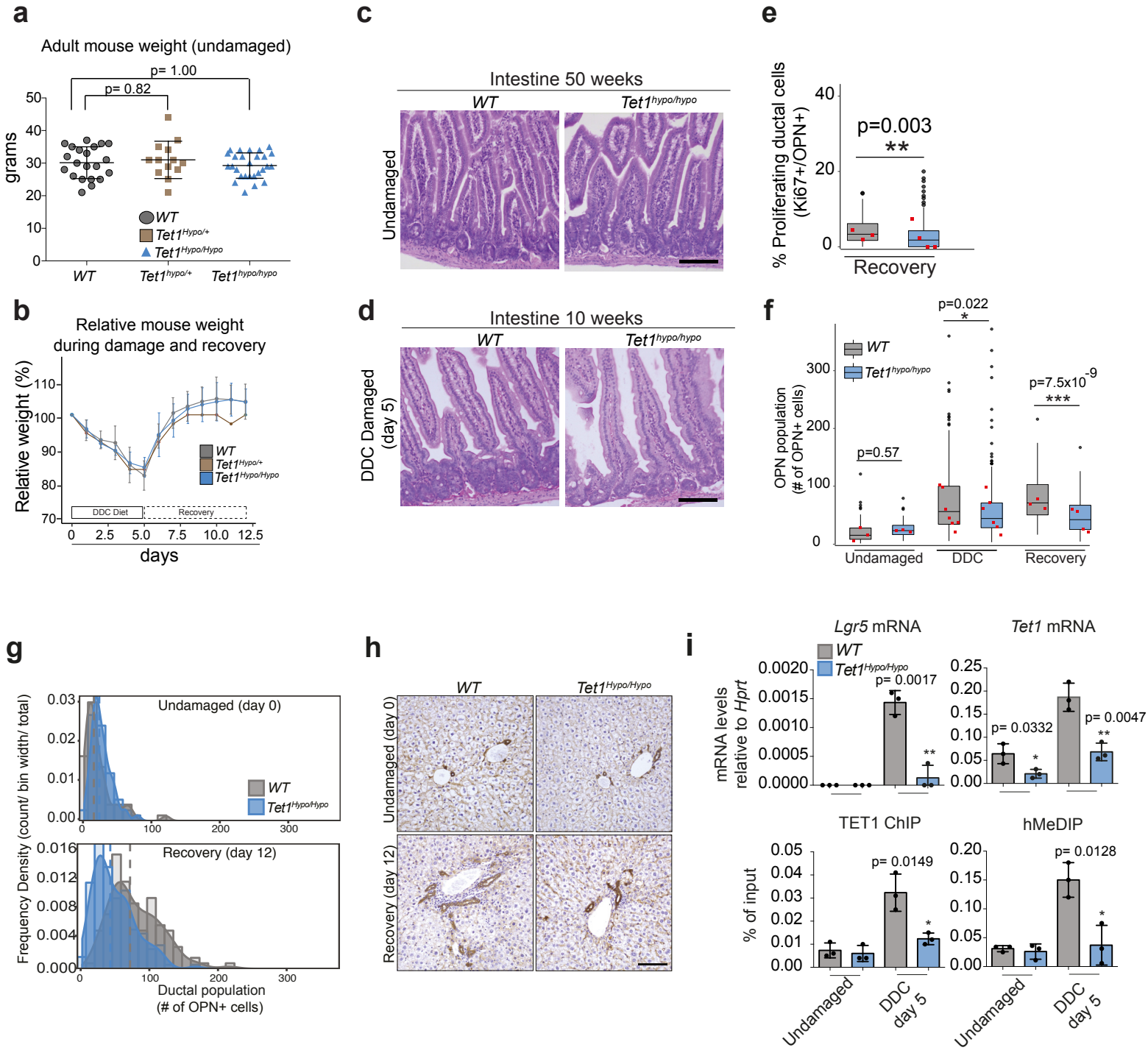
a







Aloia_Extended Data Figure 08



Aloia_Extended Data Figure 09