Generation of functional hepatocytes by forward programming with nuclear receptors

Rute A. Tomaz[1,2], Ekaterini D. Zacharis[1,2], Fabian Bachinger[1,2], Annabelle Wurmser[1,2], Daniel Yamamoto[1,2], Sandra Petrus-Reurer[2], Carola M. Morell[1,2], Dominika Dziedzicka[1,2], Brandon T. Wesley,2[1], Imbisaat Geti[1,2], Charis-Patricia Segeritz[1,2], Miguel Cardoso de Brito[1,2], Mariya Chhatriwala[2], Daniel Ortmann[1,2], Kourosh Saeb-Parsy[2], Ludovic Vallier[1,2,3*]

1. Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
2. Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, Cambridgeshire, UK
3. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

*Corresponding author

Abstract

Production of large quantities of hepatocytes remains a major challenge for a number of clinical applications in the biomedical field. Directed differentiation of human pluripotent stem cells (hPSC) into hepatocyte-like cells (HLCs) provides an advantageous solution and a number of protocols have been developed for this purpose. However, these methods usually follow different steps of liver development *in vitro* which is time consuming and requires complex culture conditions. In addition, HLCs lack the full repertoire of functionalities characterising primary hepatocytes. Here, we explore the interest of forward programming to generate hepatocytes from hPSCs and to bypass these limitations. This approach relies on the overexpression of 3 hepatocyte nuclear factors (*HNF1A*, *HNF6* and *FOXA3*) in combination with different nuclear receptors expressed in the adult liver using the OPTi-OX platform. Forward programming allows for the rapid production of hepatocytes (FoP-Heps) with functional characteristics using a simplified process. We also uncovered that the overexpression of nuclear receptors such as RORc can enhance specific functionalities of FoP-Heps thereby validating its role in lipid/glucose metabolism. Together, our results show that forward programming could offer a versatile alternative to direct differentiation for generating hepatocytes *in vitro*.

Introduction

Hepatocytes are the main cell type of the liver, comprising 80% of its volume and performing a vast array of vital functions including lipid metabolism, storage of macronutrients, secretion of plasma proteins and xenobiotic detoxification (Gordillo et al., 2015; Si-Tayeb et al., 2010; Trefts et al., 2017). Diseases affecting these functions are life threatening and end stage forms require liver transplantation. However, only a limited number of patients can benefit from this therapy due to scarcity of donors and the side effects of immunosuppression. Cell-based therapy using primary hepatocytes has already been found to be an attractive therapeutic alternative to whole organ transplants (Dhawan et al., 2020). However, primary human hepatocytes (PHHs) are in short supply as they can only be obtained from suboptimal livers unsuitable for transplantation. Furthermore, they display a short life, absence of proliferation and rapid loss of functionality *in vitro* (Mitry et al., 2002). Similarly, the development of new platforms for drug development and toxicology screens is greatly affected by the lack of robust sources of hepatocytes. For all these reasons, alternative sources of hepatocytes are urgently needed. Producing hepatocytes from human pluripotent stem cells (hPSCs) using directed differentiation protocols has been shown to be an advantageous alternative to PHHs (Palakkan et al., 2017; Szkolnicka & Hay, 2016). These protocols commonly follow key stages of liver development *in vitro* and allow the production of hepatocyte-like cells (HLCs) which exhibit key hepatic functions including Albumin secretion, lipid metabolism, glycogen storage, and urea cycle activity. However, HLCs systematically present an immature/foetal-like phenotype lacking the full repertoire of functions of mature hepatocytes (Baxter et al., 2015; Grandy et al., 2019; Yiangou et al., 2018). The development of fully functional hepatocytes *in vitro* is challenging due to the lack of detailed knowledge concerning the molecular mechanisms driving functional maturation *in vivo*. Indeed, maturation is a process that also occurs progressively *in vivo*. Mimicking this timeline and the associated

combination of metabolic changes, exposure to oxygen, nutrition and microbiome constitutes a major challenge for direct differentiation protocols (Chen et al., 2011). As an alternative, overexpression of transcription factors has been explored as a way to improve functionality of *in vitro* generated hepatocytes (Boon et al., 2020; Nakamori et al., 2016; Zhao et al., 2013). Moreover, transdifferentiation of somatic cells into liver cells has been achieved by overexpression of liver-enriched transcription factors (LETFs) in mouse and human fibroblasts (Rombaut et al., 2021). Importantly, these LETFs comprise the HNF1, HNF3 (FOXA), HNF4 and HNF6 (ONECUT) families all of which play key roles in coordinating liver development (Gordillo et al., 2015; Lau et al., 2018; Schrem et al., 2002). However, direct cell conversion from somatic cell types has a low efficiency/yield due to the strong epigenetic restrictions present in fully differentiated cells. Furthermore, somatic cells have restricted capacity of proliferation which limits large-scale production of hepatocytes without the use of oncogenic manipulation (Du et al., 2014; Huang et al., 2014). Forward programming by direct overexpression of transcription factors in hPSCs could bypass these limitations. Indeed, the epigenetic state of hPSCs is more permissive to direct cellular conversion, while their capacity of proliferation is far more superior to somatic cells. Accordingly, this approach has been successfully used to generate neurons, skeletal myocytes, and oligodendrocytes by taking advantage of the OPTi-OX system (Pawlowski et al., 2017). Here, we decided to exploit the same platform to produce hepatocytes by forward programming. We first tested different combinations of LETFs and identified a cocktail of 3 factors sufficient to drive the conversion into immature hepatocytes (FoP-Heps). We then performed transcriptomic and epigenetic comparisons between HLCs and PHHs to identify additional transcription factors which could further increase the functional maturation of hepatocytes. This comparison revealed that a number of nuclear receptors are expressed in adult hepatocytes and thus are likely to be inducers of functionality and maturation *in vivo*. A

selection of these factors was combined with LETFs and we identified that the 4TFs HNF1A-HNF6-FOXA3-RORc were the most efficient cocktail to generate FoP-Heps displaying features of mature hepatocytes including CYP3A4 activity, protein secretion and hepatotoxic response. Thus, forward programming offers an alternative to direct differentiation, bypassing the need for complex culture conditions and lengthy timelines. Moreover, FoP-Heps display a level of functionally relevant for regenerative medicine, as well as disease modelling or drug screening.

Results

Liver-enriched transcription factors allow forward programming into cells with hepatocyte identity

The first step to develop a forward programming method consists in identifying a cocktail of transcription factors which can recreate the transcriptional network characterising the target cell type. However, this step is challenging for hepatocytes as liver development is not initiated by a single and specific master regulator, and the factors driving functional maturation of hepatocytes remain to be fully uncovered. To bypass these limitations, we decided to focus on the LETFs which are known to control the induction of the hepatic program during foetal development and have been tested in somatic cell conversion (Rombaut et al., 2021). The coding sequence of 4 LETFs (*HNF4A*, *HNF1A*, *HNF6* and *FOXA3*) was cloned into the OPTi-OX system (Figure 1A) and the resulting inducible cassette was targeted into the AAVS1 gene safe harbour (Bertero et al., 2016; Pawlowski et al., 2017). After selection, individual sublines were picked, expanded and genotyped before further characterisation. Addition of doxycycline (dox) for 24h was sufficient to induce homogenous and robust expression of each LETF in the selected hESCs (Figure 1B, C) confirming the efficacy of the OPTi-OX system in inducing transgene expression.

Importantly, this induction was not associated with differentiation into liver cells (data not shown) suggesting that LETFs alone are not sufficient to impose an hepatocytic identity. Thus, we decided to screen culture conditions which could sustain both the survival and differentiation of hepatocytes (data not shown) and found that after the initial 24h in E6 medium, the cells acquired a hepatocyte-like morphology when cultured in Hepatozyme complete medium for 14 days (Figure 1D). Interestingly, the resulting cells expressed hepatocyte markers such as Albumin (*ALB*), Alpha-1 Antitrypsin (A1AT or *SERPINA1*) and Alpha-Fetoprotein (*AFP*) (Figure 1E) and displayed CYP3A4 activity levels comparable to HLCs generated by direct differentiation (Figure 1F). Next, we asked whether all the 4 LETFs were necessary to achieve this hepatocyte-like phenotype. For that, we removed each factor to generate hESCs sublines expressing combinations of 3 factors (Figure 1-figure supplement 1A). Robust and homogeneous expression at the protein level was again confirmed after 24h of dox induction (Figure 1-figure supplement 1B). Induction of each combination of 3 LETFs in culture conditions identified above showed that *HNF1A*, *HNF6* or *FOXA3* were necessary to generate cells expressing hepatocytes markers such *ALB,* albeit with heterogeneity at the protein level (Figure 1G, Figure 1-figure supplement 2A,B). *HNF4A* overexpression seemed to be dispensable as cells generated by overexpression of the 3 remaining LETFs (*HNF1A*, *HNF6*, *FOXA3*) acquired a cobblestone-like morphology, and expressed high levels of *ALB*, *SERPINA1* and *AFP* (Figure 1G, H, I, Figure 1–figure supplement 2). Strikingly, hepatocytes generated using these 3TFs (3TF FoP-Heps) achieved the highest levels of CYP3A4 activity suggesting overexpression of *HNF4A* itself is unnecessary to acquire the hepatocyte characteristics screened, as its expression might be induced by one of the 3 LETFs (Figure 1J). Altogether, these results showed that overexpression of *HNF1A*, *HNF6* and *FOXA3*, is sufficient to forward program hPSCs towards hepatocyte-like cells.

HLCs generated by direct differentiation lack the expression of specific nuclear receptors

Following these encouraging results, we aimed to increase the functionality of 3TF FoP-Heps by adding TFs which could play a role in promoting hepatic maturation and functionality. However, identifying these factors proved to be challenging as there is little information about the mechanisms driving functional maturation of hepatocytes, especially following birth when adult hepatic functions are established. To bypass this limitation, we decided to compare the transcriptome profile of adult PHHs to the transcriptome of HLCs generated from hPSCs by direct differentiation. Indeed, HLCs represent a foetal state which has been broadly characterised (Baxter et al., 2015), while 3TFs FoP-Heps are likely to be less relevant for natural development. For this comparison, we used a state-of-the-art protocol (Hannan et al., 2013; Touboul et al., 2010) which has been used for modelling liver disease (Rashid et al., 2010; Segeritz et al., 2018) and as proof of concept for cell-based therapy applications (Yusa et al., 2011). This protocol starts by the production of endoderm cells expressing *SOX17,* followed by the specification of foregut expressing *HHEX*, after which cells transit through a hepatoblast-like state marked by *TBX3* (Figure 2-figure supplement 1A). Interestingly, LETFs were expressed during this differentiation protocol at levels comparable to PHHs (Figure 2-figure supplement 1B,C) confirming that these first steps follow a natural path of development. The resulting progenitors undergo a final stage of differentiation into HLCs expressing functional markers such as *ALB* and *SERPINA1* (Figure 2A, Figure 1-figure supplement 1A,D). Despite displaying key hepatic functions (Baxter et al., 2015; Grandy et al., 2019; Yiangou et al., 2018) HLCs represent a “foetal” state as shown by the expression of *AFP* (Figure 2-figure supplement 1A,D) or by the limited activity/expression of *CYP3A4*, *CYP2A6* or *CYP2C9* (Figure 1B, Figure 2-figure supplement 1E). RNA-sequencing (RNA-

seq) performed on HLCs generated from either human induced pluripotent stem cells (hiPSCs) or human embryonic stem cells (hESCs), and from PHHs freshly harvested (fPHHs) or cultured *in vitro* as monolayer (pPHHs) reinforced these observations. Principal component analysis (PCA) of the most variable 500 genes showed a clear distinction between the 3 cell types, with HLCs clustering in between undifferentiated hiPSCs and PHHs confirming their intermediate state of differentiation (PC1: 52%, Figure 2C). In order to further explore the differences between HLCs and PHHs, we combined Differential Gene Expression (DGE) and Gene Ontology (GO) analyses to identify genes and biological functions specific to each cell type (Figure 2D). Genes uniquely expressed in PHHs (cluster 1) were associated with adult liver functions such as response to xenobiotic and xenobiotic metabolism, inflammatory response and complement activation (Figure 2E). Genes expressed in both HLCs and PHHs (Heps; cluster 3) were involved in liver development, fatty acid metabolism or broad cellular functions (Figure 2F). Of note, genes specifically up-regulated in HLCs were associated with extracellular matrix organization and varied developmental functions which could originate from their *in vitro* environment (Figure 2-figure supplement 1F). We then decided to focus specifically on transcription factors (TFs) using a previously curated list (Lambert et al., 2018) (Figure 2G,H) and identified 36 TFs highly expressed in PHHs vs HLCs ($p < 0.05$, log2 fold change > 2). Interestingly, reactome pathway analysis grouped these TFs into two main pathways: the NFI family and a cohort of 8 nuclear receptors (Figure 2H). Nuclear receptors are known to be involved in key liver functions including the metabolism of lipids and glucose, bile acid clearance, xenobiotic sensing and regeneration (Rudraiah et al., 2016). Thus, we hypothesise that nuclear receptors could be the most promising candidates to improve hepatocyte functionality in our culture system. Taken together these observations show that the immature state of HLCs is associated with the

absence of several nuclear receptors thereby suggesting that these factors could be necessary to drive functional maturation of hepatocytes.

Epigenetic characterisation of HLCs suggests a role for nuclear receptors RORc, AR and ERα

To further refine the list of nuclear receptors identified by our transcriptomic analyses, we decided to compare the epigenetic landscape of HLCs vs PHHs. Indeed, we hypothesised that nuclear receptors binding regulatory regions in PHHs could have a key function in maturation and thus we aimed to identify the most important factors by screening underlying motifs in such regions. ChIP-sequencing (ChIP-seq) was performed on histone marks including H3K27ac (active regulatory regions), H3K4me1 (active or primed regulatory regions), and H3K27me3 (silenced genes) (Creyghton et al., 2010; Wang et al., 2015). These marks were profiled in HLCs derived from both hiPSCs and hESCs, and PHHs while undifferentiated hiPSCs were used as control. As expected, PCA analyses showed a marked divergence between the epigenetic profile of HLCs and hiPSCs independently of the mark analysed (Figure 3A). Interestingly, HLCs and PHHs clustered in close proximity suggesting that these cell types share an important part of their epigenetic profile despite their transcriptomic differences. The profiles of H3K27ac and H3K27me3 showed the highest variance between HLCs/PHHs and hiPSCs, confirming the importance of these marks for establishing cellular identity (Figure 3A). We then performed differential peak calling to identify regulatory regions uniquely enriched and active in PHHs versus HLCs ("PHH-specific"), and vice versa ("HLC-specific") (Figure 3-figure supplement 1A). We also profiled H3K4me1 and H3K27me3 in either PHH or HLC-specific regions. This analysis revealed that H3K4me1 was absent at "PHH-specific" regions in HLCs and seems to be broadly replaced by spread of H3K27me3 deposition instead (Figure 3B). Interestingly, discrete portions of regulatory

208 regions lacked H3K27ac in genes downregulated in HLCs (see for example *CYP3A4* and
209 *UGT1A*, Figure 3C, Figure 3-figure supplement 1B). Ontology of genes associated with each
210 set of regions highlighted several adult liver metabolic processes in the "PHH-specific" set
211 such as steroid, lipid and xenobiotic metabolism, and range of neural functions for "HLC-
212 unique" regions (Figure 3-figure supplement 1C), in agreement with the transcriptomic
213 analyses. Overall, these results suggested that a subset of genes involved in adult liver
214 functions lack H3K4me1 priming as well as full H3K27ac deposition in HLCs. These
215 regions lacking H3K27ac in HLCs, appeared to display repressive marks such as H3K27me3.
216 In addition, HLCs displayed active histone marks in regions including genes which are not
217 associated with liver differentiation confirming that cells generated from hPSCs also present
218 an epigenetic signature specific to their *in vitro* state (Figure 3-figure supplement 1C). Taken
219 together these observations suggested that HLCs and PHHs broadly share the same
220 epigenetic identity when compared to hiPSCs. However, the activation of a limited and
221 specific set of regulatory regions is missing in HLCs, which potentially explains the absence
222 of expression of adult liver genes and lack of functional maturation. To identify the nuclear
223 receptors potentially involved in regulating these regions, we performed motif enrichment
224 analysis in the "PHH-specific" regions marked by H3K27ac. Interestingly, this analysis
225 identified a significant enrichment for the androgen (AR) and Estrogen (ERα) response
226 elements, as well as RORc motifs (Figure 3D), which were among the top differentially
227 expressed nuclear receptors in our transcriptomic analyses. We then decided to further
228 investigate the importance of these nuclear receptors throughout development using mouse
229 RNA-seq datasets obtained at different stages of liver organogenesis (E12.5, E16.5, P0, 8
230 week and 10 week adults (Figure 3–figure supplement 1D). Interestingly, the expression of
231 these 3 nuclear receptors was found to be upregulated specifically in the adult liver (Figure
232 3–figure supplement 1E,F). Altogether, these observations suggested that the nuclear

receptors AR, ERα and RORc could play a role in establishing or maintaining a transcriptional network characterising mature hepatocytes.

Overexpression of RORc increases the functionality of hepatocytes generated by forward programming

We next tested the capacity of RORc (*RORγ*), AR, and ERα (*ESR1*), to further improve the functionality of FoP-Heps generated using 3 LETFS. For that, we generated hESC lines inducible for the expression of HNF1A, HNF6 and FOXA3 (3TFs) in combination with each of the nuclear receptors identified above (Figure 4–figure supplement 1A). The homogeneous induction of the 4TFs was validated using immunostaining. Interestingly, these analyses showed that the overexpressed nuclear receptors were located in the nucleus, and in the case of AR in both cytoplasm and nucleus (Figure 4–figure supplement 1B,C,D). We then induced forward programming using the culture conditions identified above and observed the production of polyploid cells with a cobblestone morphology (Figure 4A). The hepatocytic identity of these cells was confirmed by the expression of Albumin, *SERPINA1*/A1AT and AFP in all lines (Figure 4B,C,D,E, Figure 4–figure supplement 1E). RORc overexpression resulted in a higher number of cells expressing Albumin, which were still heterogeneous across the cell population, potentially representing different subpopulations of hepatocyte-like cells (Figure 4B). Moreover, these cells also secreted higher levels of Albumin, although the mRNA expression of this marker was not significantly upregulated (Figure 4C, D). In addition, these cells tend to secret lower levels of AFP, although the lower mRNA levels identified were not statistically significant when compared to 3TFs alone (Figure 4D, E, Figure 4–figure supplement 1E). Notably, CYP3A4 activity levels were significantly higher in cells generated in the presence of RORc, as compared with cells reprogrammed with only 3TFs (Figure 4F). We next tested whether stimulation with exogenous ligands specific for

each nuclear receptor could further induce functional maturation as measured by CYP3A4 activity (desmosterol for RORc, β-estradiol for ERα and testosterone for AR). Interestingly, only β-estradiol treatment resulted in a 3-fold increase in CYP3A4 activity, whereas testosterone treatment significantly decreased CYP3A4 activity and desmosterol had no effect (Figure 4G). This increase was not observed with 3TFs Fop-Heps thereby suggesting the effect of these ligands was linked to the overexpression of their receptor.

RORc generated FoP-Heps appeared to have the highest level of functionality and thus, we decided to validate the potential of this combination of factors in an alternative pluripotent stem cell line. We generated Opti-OX hiPSC with the 3TFs or the 3TFs+RORc FoP system, validated the upregulation of these factors after 24h of dox treatment (Figure 4–figure supplement 2A) and then induced differentiation following the protocol established above. FoP-Heps derived from hiPSC also displayed cobblestone morphology (Figure 4–figure supplement 2B) and expressed Albumin, AFP and *SERPINA1*/A1AT (Figure 4–figure supplement 2C,D,E,F,G). Concordantly, a higher number of Albumin positive cells were detected at day 20, as well as higher levels of secreted Albumin and transcript, when RORc was overexpressed (Figure 4–figure supplement 2C,F,G). In particular, *ALB* transcript further increased by day 30, suggesting that this cell background might require additional time to achieve maturation. In addition, the presence of RORc significantly increased basal CYP3A4 activity thereby confirming the positive effect of this factor on functional maturity of FoP-Heps (Figure 4–figure supplement 2H). Overall, these results showed that overexpression of specific nuclear receptors was compatible with the generation of FoP-hepatocytes. In particular, overexpression of RORc could improve the functionality of hepatocytes generated by overexpression of 3 LETFs confirming the role of this nuclear receptor in hepatocyte maturation.

283 **4TF FoP-Heps display functional characteristics *in vitro***

284 Next, we sought to further characterise the functionality of the 4TF (HNF1A, HNF6, FOXA3
285 and RORc) FoP-Heps derived from either hESC (eFoP-Heps) or hiPSCs (iFoP-Heps) in
286 comparison with HLCs generated by the direct differentiation and PHHs. CYP3A4 activity
287 was significantly higher in 4TF FoP-Heps forward programmed after 20 days than those
288 achieved by HLCs after 30 days of directed differentiation (Figure 5A). In addition, we
289 analysed the expression of markers associated with hepatic metabolic functions such as phase
290 I (cytochrome P450 enzymes) and phase II (UGTs) biotransformation, gluconeogenesis
291 (*G6PC* and *PCK1*) and lipid (*PPARα*, *PPARγ*, *FASN* and *APOA1*) and bile acid (*NR1H4*)
292 metabolism. FoP-Heps expressed a range of these functional markers confirming the
293 acquisition of hepatic functionality (Figure 5B,C, Figure 5-figure supplement 1A). Overall,
294 the levels of expression achieved by forward programming were equivalent to those achieved
295 by direct differentiation with the exception of gluconeogenesis genes which were increased in
296 FoP-Heps (Figure 5C). Interestingly, expression of gluconeogenesis and lipid metabolism
297 genes was comparable between FoP-Heps and PHHs. However, induction of cytochrome
298 P450 genes remains challenging, indicating that the acquisition of this specific hepatic
299 function could need further refinement of our protocol (Figure 5B,). Of note, in FoP-Heps,
300 the 4TFs remained expressed at physiological levels at the end of our protocol (Figure 5B,
301 Figure 5–figure supplement 1B). Next, we sought to further characterise the transcriptome of
302 day 20 4TF FoP-Heps by RNA-seq by comparing the transcriptome of eFoP-Heps with
303 undifferentiated hiPSCs, HLCs (derived from direct differentiation of hESC and hiPSCs),
304 adult PHHs as well as foetal liver cells. As suggested by the marker expression pattern
305 identified by qPCR, the transcriptome of eFoP-Heps closely clustered with HLCs derived
306 from both hESCs and hiPSCs (Figure 5D). Interestingly, both eFoP-Heps and HLCs clustered
307 separately from foetal liver cells, indicating that despite these cells not acquiring a fully

mature phenotype, these also don't fully resemble a foetal liver stage. We further interrogated the ontology of genes differentially expressed between eFoP and the different groups of samples (Figure 5–figure supplement 2). As expected, genes that gain expression in eFoP-Heps compared to undifferentiated hiPSCs are strongly associated with several liver functions such as hormone metabolism, lipid localization and transport, blood coagulation (Figure 5E), further confirming that cells generated by forward programming acquire a hepatocyte identity. Indeed, eFoP-Heps, as well as hiPSC-derived HLCs, expressed a range of adult liver genes that were found to be highly expressed in adult PHHs as identified in Figure 2D (Figure 2-source data 2), including genes that were not expressed in the foetal stage (Figure 5F). In addition to expressing mature hepatocyte markers, both eFoP-Heps and iFoP-Heps were also able to uptake LDL from the culture medium confirming their capacity to transport lipids (Figure 6A). In order to further explore their capacity to metabolise lipids, FoP-Heps were grown in 3D for an additional 5 (D20) or 10 (D30) days as we recently observed that such culture conditions facilitate lipid accumulation in HLCs (Carola M. Morell, personal communication, (Tilson et al., 2021). We first confirmed that FoP-Heps grown in 3D retained the expression of hepatocyte markers (Figure 6B). Interestingly, *SERPINA1* or *UGT1A6* expression increased in these conditions suggesting an increase in functional maturation promoted in 3D (Figure 6B). We then tested the capacity of both eFoP and iFoP-Heps to respond to fatty acids by treating these cells with both oleic acid (OA) and palmitic acid (PA) which are known to induce steatosis and lipotoxicity respectively (Ricchi et al., 2009). In line with their known effect on hepatocytes, OA treatment induced a strong accumulation of lipids as shown by BODIPY staining (Figure 6C) while PA treatment induced a reduction in cell viability consistent with lipotoxicity (Figure 6D). Thus, FoP-Heps appear to react to fatty acids similarly to their primary counterpart. Finally, we explored the interest of FoP-Heps for modelling the hepatotoxic effect of paracetamol/acetaminophen. For

that, Fop-Heps were grown in the presence of an acetaminophen (APAP) dose known to induce liver failure. This treatment resulted in a 50% reduction in cell viability (Figure 6E) suggesting that FoP-Heps could be used for cytotoxic studies. In summary, these results showed that 4TF FoP-Heps derived from either hESC or hiPSCs display characteristics of functional hepatocytes such as expression of genes involved in drug, lipid, glucose and bile acid metabolism, capacity to uptake LDL and fatty acids from the culture medium, as well as response to hepatotoxic factors, demonstrating their potential interest for modelling liver disease *in vitro* and toxicology screening.

Discussion

In this study, we have established a method to forward program hPSCs into hepatocytes by taking advantage of the OPTi-OX platform (Pawlowski et al., 2017). The success of this approach depends on the selection of TFs, combining factors controlling early liver development and regulators of adult hepatic functions. Nonetheless, most forward programming methods rely on a master regulator to convert hPSCs into a specific cell type. As an example, neurons and muscle cells can be generated by the simple overexpression of NGN2 and MYOD respectively (Pawlowski et al., 2017). Our results show that production of hepatocytes requires a more complex process involving 3 transcription factors but also a culture media supporting primary hepatocytes. Furthermore, our best LETFs combination did not include HNF4A, which is known to be a key regulator of hepatocyte function in the adult liver. On the contrary, removing HNF4A significantly improved the identity of the hepatocytes generated. Similar observations were recently reported for the direct reprogramming of human umbilical vein endothelial cells into bipotent hepatocyte progenitor cells where HNF4A was found to be detrimental (Inada et al., 2020). HNF4A is essential not only in adult liver but also during development, especially in the establishment of the liver

bud (Gordillo et al., 2015). Thus, HNF4A might also have a role in preserving foetal liver cells such as hepatoblasts and its overexpression during forward programming could block the acquisition of an adult hepatocytic identity. This example illustrates the challenges to identify factors which are uniquely expressed in the adult liver.

Importantly, FoP-Heps generated by LETF overexpression acquired an hepatocytic identity with reduced adult functions, suggesting that this cocktail of transcription factors might only convert hiPSCs into foetal-like cells. Thus, we decided to add factors which could direct functional maturation of the liver. This latest category of factors was identified by performing a transcriptomic and epigenetic comparison of PHHs and HLC generated by direct differentiation. The focus on HLCs was based on their well characterised foetal state and also the broad experience with the cells. These analyses identified a subset of nuclear receptors that were exclusively expressed in PHHs and in the adult liver thereby confirming the relevance of our approach. Of particular interest, RORc, ERα and AR were identified as key candidates for controlling functional maturation in hepatocytes. Importantly, nuclear receptors are well known to control diverse liver functions including lipid and glucose homeostasis, bile acid clearance, xenobiotic sensing and regeneration (Rudraiah et al., 2016). Both steroid hormonal receptors ERα and AR have been shown to have roles in the regulation of energy homeostasis in the liver (Shen & Shi, 2015). Moreover, ERα is involved in cholesterol clearance (Zhu et al., 2018) and has also been associated with liver regeneration (Kao et al., 2018) and bilirubin metabolism through CYP2A6 (Kao et al., 2017). RORc is a nuclear receptor expressed in peripheral tissues including liver, muscle and adipose tissue and has been proposed to function as an intermediary between the circadian clock and glucose/lipid metabolism (Cook et al., 2015). Moreover, *RORγ*-deficient mice exhibit insulin sensitivity and reduced expression of gluconeogenesis, lipid metabolic markers, and a subset of phase I enzymes involved in bile acid synthesis and phase II

383 enzymes (Kang et al., 2007; Takeda, Kang, Freudenberg, et al., 2014; Takeda, Kang, Lih, et
384 al., 2014). Based on these previous reports, we propose that the overexpression of RORc and
385 other nuclear receptors could improve specific functions in FoP-Heps by activating a subset
386 of target genes in the hepatic context induced by the LETFs overexpression. Importantly,
387 hepatocyte functionality is spatially different across the liver lobule, being influenced by the
388 gradient of oxygen, nutrients and signalling (Trefts et al., 2017). This hepatic zonation drives
389 different metabolic processes in regards to glucose, lipids, iron, or even xenobiotics, which
390 are under the control of different transcriptomic programs (Halpern et al., 2017). Thus, we
391 expect that different combinations of nuclear factors in the background induced by LETFs
392 overexpression could enable the production of hepatocytes with a distinct repertoire of
393 functions.
394 FoP-Heps generated with the overexpression of the 4TFs (HNF1A, HNF6, FOXA3 and
395 RORc) displayed functional features of adult hepatocytes including Albumin and A1AT
396 secretion, basal CYP3A4 activity, expression of PhaseI/PhaseII enzymes, gluconeogenesis
397 and lipid metabolism markers, capacity to uptake LDL and fatty acids as well as response to
398 toxic compounds. Nonetheless, *CYP3A4* expression remains limited and this gene remains
399 difficult to induce *in vitro*. Thus, additional TFs could be necessary to generate FoP-Heps
400 exhibiting the full spectrum of functional activities displayed by PHHs. Similarly, culture
401 conditions could be further improved to support key hepatic functions. Indeed, the basal
402 medium used in our protocol does not prevent dedifferentiation of PHHs and thus might not
403 be compatible with the production of fully functional cells by forward programming.
404 Nonetheless, the forward programming method established here presents several advantages
405 over conventional directed differentiation protocols. This is a robust two-steps method which
406 bypasses the need for multi-step differentiations which are often associated with batch-to-
407 batch variability. Furthermore, forward programming is faster, generating functional cells in

20 days, as opposed to 30-35 days for direct differentiation. Finally, the yield of cells seems favourable and compatible with large-scale production. Indeed, we observed that forward programming was associated with an 6-8 fold increase in cell number during differentiation while the yield of direct differentiation is lower (data not shown). Moreover, the phenotype achieved is stable even in the absence of doxycycline which allows this method to be applicable in cell therapy and drug discovery (data not shown).

Taken together, our results describe the first method for generating hepatocytes using forward programming. This approach represents the first step towards the high-throughput and large-scale production of specialized hepatocytes displaying a spectrum of functions relevant for different applications in disease modelling and drug screening.

Methods

hPSC culture

The human ESC H9 (WiCell) and IPSC A1ATD$^{R/R}$ (Yusa et al., 2011) lines were used in this project. Human iPSC line was derived as previously described, under approval by the regional research ethics committee (REC 08/H0311/201). Both hPSCs were cultured on vitronectin XFTM (10µg/mL, StemCell Technologies)-coated plates and in Essential 8 (E8) chemically defined medium consisting of DMEM/F12 (Gibco), L-ascorbic acid 2-phosphate (1%), insulin-transferrin-selenium solution (2%, Life Technologies), sodium bicarbonate (0.7%), and Penicillin/Streptomycin (1%), freshly supplemented with TGFβ (10ng/ml, R&D) and FGF2 (12ng/ml, Qkine) (Chen et al., 2011). For routine dissociation, cells were incubated with 0.5µM EDTA (ThermoFisher Scientific) for 3 minutes at 37°C seeded in small clumps. Cells were maintained at 37°C in 20% O2, 5% CO2 and medium was replenished every 24h.

Authentication of hPSCs was achieved by confirming the expression of pluripotency genes. Cells were routinely confirmed to be mycoplasma free using broth and PCR-based assays. The cell lines are not on the list of commonly misidentified cell lines (International Cell Line Authentication Committee).

Gene Targeting

Inducible hESC and hiPSC lines were generated using the OPTi-OX system as previously described (Bertero et al., 2016; Pawlowski et al., 2017). Briefly, two gene safe harbours were targeted (GSH). The hROSA26 locus was targeted with a constitutively expressed transactivator (rtTA) and the AAVS1 locus with the transgenes of interest under a TET responsive element (TRE). Different combinations of transcription factors and/or nuclear receptors as stated throughout the manuscript were cloned. Template cDNA sequences were obtained either from Dharmacon: *HNF6* (MHS6278-213244170), *HNF1A* (MHS6278-202857902), *RORy* (MHS6278-202800991) and *ESR1* (MHS6278- 211691051); or amplified from human primary liver cDNA: *HNF4A*, *FOXA3* and *AR*. Sequences were amplified using the KAPA HiFi HotStart ReadyMix (Roche). The primers used to amplify and clone the sequences into the backbone vector contained upstream and downstream overhangs in order to generate a GSG (Gly-Ser-Gly) linker and a different 2A peptide as listed in table 1. The different vectors were constructed by Gibson Assembly (New England Biolabs) using a 1:3 pmol ratio of vector to insert. For targeting, hPSCs were dissociated into single cells with STEMpro accutase (Thermo Fisher) for 5 minutes, and 1 million cells were transfected with 2µg of donor vector and 2µg of each AAVS1 ZFN expression plasmids using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza). Cells were seeded in E8 medium supplemented with 10µM ROCK Inhibitor Y-27632 (Selleckchem). After 5-7 days, colonies were selected with

1µg/ml puromycin (Sigma Aldrich) for at least 2 days, after which they were individually picked and genotyped as previously described (Bertero et al., 2016; Pawlowski et al., 2017).

Hepatocyte Direct Differentiation

hPSCs were dissociated into single cells following incubation with StemPro Accutase (Thermo Fisher) for 5 minutes at 37°C and seeded at a density of 50.000cells/cm^2 in E8 medium supplemented with 10µM ROCK Inhibitor Y-27632 (Selleckchem). Hepatocytes were differentiated 48h after seeding, as previously reported (Hannan et al., 2013) with minor modifications. Following endoderm differentiation, anterior foregut specification was achieved with RPMI-B27 differentiation media supplemented with 50ng/ml Activin A (R&D) for 5 days. Cells at the foregut stage were further differentiated into hepatocytes with Hepatozyme complete medium: HepatoZYME-SFM (Thermo Fisher) supplemented with 2mM L-glutamine (Thermo Fisher), 1% penicillin-streptomycin (Thermo Fisher), 2% non-essential amino acids (Thermo Fisher), 2% chemically defined lipids (Thermo Fisher), 14µg/ml of insulin (Roche), 30µg/ml of transferrin (Roche), 50 ng/ml hepatocyte growth factor (R&D), and 20 ng/ml oncostatin M (R&D), for up to 27 days.

Forward Programming into Hepatocytes

hPSCs were dissociated into single cells following incubation with StemPro Accutase (Thermo Fisher) for 5 minutes at 37°C and seeded at a density of 40-50.000cells/cm^2 in E8 medium supplemented with 10µM ROCK Inhibitor Y-27632 (Selleckchem). E8 medium was replenished the following day. Following 48h, initial induction of the transgenes was achieved by incubation in E6 medium (E8 without growth factors) supplemented with 1mg/ml doxycycline (dox) for 24h. Cells were then maintained in Hepatozyme complete medium supplemented with 1mg/ml dox for the remaining duration of the protocol. Medium

was replenished every day for the next 4 days, and every other day here after. For specific experiments, cell lines were treated with 100nM of desmosterol, testosterone or β-estradiol (E2) from day 2 of forward programming. All ligands were purchased from Sigma-Aldrich and reconstituted in ethanol. For 3D cultures, forward programmed cells were embedded in Matrigel Growth Factor Reduced Basement Membrane Matrix, Phenol Red-free (Corning) at day 15 or day 20 and cultured for 5 days or 10 days, respectively. Cells were dissociated with Hank's based cell dissociation buffer (Gibco) for 20 minutes at 37°C, resuspended in Matrigel and seeded in 40-50 µL domes in Hepatozyme complete medium supplemented with 1mg/ml dox.

Primary Liver Samples

Fresh primary hepatocytes used for RNA-seq were obtained as previously reported (Segeritz et al., 2018). Primary plated hepatocytes from 4 donors (3 males and 1 female) were purchased from Biopredic International (Rennes, France), meting the manufacturer's quality control requirements. Cells were maintained in short-term monolayer cultures in William's E (Gibco) supplemented with 1% Glutamine (Gibco), 1% Penicillin-streptomycin (Gibco), 700nM Insulin (Sigma-Aldrich) and 50µM Hydrocortisone (Sigma). Functional assays such as CYP3A4 activity measurement were performed in Hepatozyme complete medium within 48h of receipt. Bulk foetal tissue was obtained from patients undergoing elective terminations up to the third trimester, under approval by the regional research ethics committee (REC-96/085). The tissue was lysed and RNA harvested as indicated below.

CYP3A4 Assay

Measurement of CYP3A4 enzymatic activity was performed using the P450 Glo kit (Promega). Cells were incubate with 1:1000 luciferin-IPA in Hepatozyme complete for 1h at

37˚C. Supernatant was mixed with detection reagent in a 1:1 ratio and incubated at RT for 20 minutes in Greiner white 96 well microplates (Sigma Aldrich). Luminescence was measured in triplicate on a GloMax plate reader. Hepatozyme complete medium was used as background control. Relative light units were normalised for background, volume and average total number of cells obtained after differentiation.

LDL uptake assay

LDL uptake capacity was measured with the LDL Uptake Assay Kit (Abcam). Cells were incubated with 1:100 human LDL conjugated to DyLight™ 550 in Hepatozyme complete medium for 3 hours at 37°C. Cells were then washed and fixed with 4% PFA for 20 minutes at 4°C.

Fatty Acid Treatments

Forward programmed cells were embedded in 3D from day 20 and cultured for 7 days in Hepatozyme complete medium supplemented with either BSA (control), or oleic acid (0.25mM) or palmitic acid (0.25mM) conjugated with BSA. Intracellular lipid accumulation was detected by incubating cells with 1µl/ml Bodipy (Thermo scientific) for 30 minutes, followed by DAPI (Hoechst) diluted 1:10.000 in PBS for 30 minutes and imaged on a Zeiss LSM 700 confocal microscope

APAP toxicity

The hepatotoxicity of acetaminophen (APAP) was tested by incubating forward programmed cells cultured in 3D from day 15 in Hepatozyme complete medium supplemented with 25 mM acetaminophen (R&D) for 48h hours (day 18 to day 20) after which cell viability was determined.

Cell Viability

Cell viability was determined by incubating cells with 1:10 Presto Blue reagent (Invitrogen) in Hepatozyme complete medium at 37°C for 4 hours. Fluorescence was measured using the EnVision plate reader with an excitation emission of 560nm/590nm.

RT-qPCR

RNA was extracted from either cells or tissues using GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) according to the manufacturer's instructions. 500ng of RNA were reverse transcribed into cDNA using Random Primers and SuperScript II (Invitrogen) according to the manufacturer's instructions. qPCR was performed using the KAPA SYBR FAST qPCR Kit low-ROX (Sigma-Aldrich) with 200 nM of forward and reverse primers (Sigma-Aldrich; primers listed in table 2) on a QuantStudio 5 (Applied Biosystems). qPCRs were performed in technical duplicates and normalised to the average of two housekeeping genes (*RPLP0* and *PBGD*) using the 2–ΔCt method.

Immunofluorescence staining

Cells in monolayer were fixed in 4% PFA for 20 minutes at 4°C and blocked for 30 minutes in 10% donkey serum (BioRad) and 0,1% Triton X-100 (Sigma-Aldrich). Fixed cells were incubated with primary antibodies listed in table 3 in 1% donkey serum and 0,01% Triton X-100 overnight at 4°C. Following washing, cells were incubated with Alexa Fluor 488-, 568- or 647-conjugated secondary antibodies (Life Technologies) for 1h at room temperature diluted in 1% donkey serum and 0,01% Triton X-100. For nuclei visualisation, cells were incubated with adding DAPI/Hoechst 33258 (bis-Benzimide H, Sigma-Aldrich) diluted

1:10.000 in PBS for 10 minutes at room temperature. Cells were imaged either on a Zeiss Axiovert 200M or on a Zeiss LSM 700 confocal microscope.

Secreted protein quantification

Albumin, alpha-fetoprotein and alpha-1-antitrypsin were measured in the cell culture supernatant of monolayer cultures, which were replenished with fresh Hepatozyme complete medium 24h prior to collection. Concentrations were detected by ELISA (performed by core biomedical assay laboratory, Cambridge University Hospitals) and normalised to cell number.

RNA-seq analyses

RNA-seq datasets were generated for undifferentiated hiPSCs (n=3), hESC-derived HLCs (n=2), hiPSC-derived HLCs (n=6), freshly harvested PHHs (fPHHs, n=3), commercially purchased PHHs (pPHHs, n=2), bulk foetal liver samples (FL, n=3) and hESC-derived 4TF FoP-Heps (eFoP, n=3). RNA was extracted from either cells or tissues using GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) according to the manufacturer's instruction. Poly-A library preparation and sequencing were performed by Cambridge Genomic Services (hESC_HLCs; pPHHs), the Wellcome Trust Sanger Institute (hiPSCs, hiPSC_HLCs, fPHHs), and the Cambridge Stem Cell Institute and CRUK (FL, eFoP). Quality of reads was assessed with FastQC. For consistency, fastq reads were split into single-end reads and trimmed to the same length (40bp) using cutadapt version 2.10. Single-end fastq files were mapped and quantified using salmon version 1.2.1 with the following parameters: -l A, -GCbias, -posbias, -validatemappings (Patro et al., 2017). The index used was pre-built from the human GRCh38 cDNA reference sequence from Ensembl (refgenomes.databio.org). Differential gene expression was calculated using DESeq2 (Love

et al., 2014), with the following parameters padj>0.05, basemean > 100 and log2 fold change > 2 or < -2 between groups as depicted in each figure. Gene ontology enrichment was calculated with the clusterProfiler package (Yu et al., 2012). Pathway analysis on significantly misregulated transcription factors was assessed using ReactomePA (Yu & He, 2016). Mouse liver polyA plus RNA-seq was downloaded from ENCODE (Consortium, 2012). Single-end fastq reads were trimmed in both replicates from each dataset to 70bp using cutadapt version 2.10. Fastq were mapped and quantified using salmon version 1.2.1 with the following parameters: -l A, -Gbias, -seqbias, -validatemappings using a pre-build mm10 cDNA reference genome. DeSeq2 was used to generate all plots for visualisation.

Chromatin immunoprecipitation (ChIP)

ChIP was performed as previously reported (Brown et al., 2011). Briefly, chromatin was crosslinked with 1% formaldehyde (Sigma-Aldrich) for 10 minutes at room temperature and quenched with 0.125M glycine (Sigma-Aldrich). Cells and nuclei were subsequently lysed and chromatin was sonicated to fragment DNA to about 200-500 bp on a Bioruptor Pico sonication device (Diagenode). Sonicated chromatin was pre-cleared with same-host IgG and protein G Dynabeads (Thermo Fisher), 100µg of cleared chromatin (protein) was incubated with 2µg of the following antibodies overnight at 4ºC: H3K27ac (Abcam, ab4729), H3K4me1 (Abcam, ab8895), H3K27me3 (Active Motif, 39155) and H3K4me3 (Merk, 05-745R), after which complexes captured with 30µl of protein G Dynabeads (Thermo Fisher). Complexes were washed, RNAse A (Thermo Fisher) and Proteinase K (Sigma-Aldrich) treated and DNA was purified by phenol-chloroform extraction and precipitated with GlycoBlue (Thermo Fisher), sodium acetate (Thermo Fisher) and ethanol (Sigma-Aldrich). A sonicated chromatin sample (1%) was also collected as input for normalisation and 10ng of DNA were used for ChIP-sequencing library preparations.

605

606 **ChIP-seq analyses**

607 Library preparation and sequencing and alignment were performed by the Wellcome Trust
608 Sanger Institute DNA Sequencing Facility (Hinxton, UK). Sequencing was performed on an
609 Illumina HiSeq v4 to obtain paired-end reads with 75bp length. ChIP-seq reads were mapped
610 to human genome assembly GRCh38 with bwa. Aligned data in BAM format was sorted and
611 indexed with samtools. Coverage files were generated using deeptools bamCoverage, with a
612 bin size of 10bp and normalised as RPKM for visualisation in IGV and heatmap
613 representation with deeptools. In order to plot principal component analysis (PCA), average
614 scores were calculated over 1000bp bins. For peak calling, BAM files were converted to
615 SAM and peaks called using homer (Heinz et al., 2010). Both replicates were used for peak
616 calling against input with disabled local filtering invoking the following flags for H3K27ac: -
617 region -L 0. In order to identify regulatory regions specifically active in PHH or HLCs,
618 differentially bound peaks were determined using PHH datasets as target against all HLCs
619 datasets as background, and vice versa, with a fold enrichment over background of 4. A list
620 of genes annotated for each peak dataset can be found in Supplementary Tables 2 and 3. For
621 motif enrichment, peak calling was performed on nucleosome free regions by invoking the
622 flags -L 1 -nfr, in order to determine the "dips" within H3K27ac-rich regions. These sets of
623 regions were overlapped with the differentially bound peaks as above, in order to perform
624 PHH or HLC-specific motif enrichment. Peak annotation and gene ontology enrichment were
625 determined with the clusterProfiler R package (Yu et al., 2012). Undifferentiated hiPSCs
626 ChIP-seq reads aligned to the same genome assembly were downloaded from ENCODE
627 (Consortium, 2012) and treated as above.

628

629 **Data availability**

RNA-seq datasets used in this study are accessible on Array Express under the accession number E-MTAB-10634 and E-MTAB-11852. In addition, 3 of the hiPSC_HLCs data sets have been previously deposited with the accession number E-MTAB-6781 (Segeritz et al., 2018). Mus musculus C57BL/6 liver embryo RNA-seq datasets were obtained from the ENCODE database (Nakamori et al., 2016) (https://www.encodeproject.org/) with the following accession numbers: ENCSR216KLZ (E12.5 liver), ENCSR826HIQ (E16.5 liver), ENCSR096STK (P0 liver), ENCSR000BYS (8 weeks mixed sex adult liver) and ENCSR216KLZ (10 weeks adult liver). ChIP-seq datasets generated in this study have been deposited on Array Express with the accession number E-MTAB-10637, and publicly available datasets for hiPSCs were used from the ENCODE database with the following accession numbers: ENCSR729ENO (H3K27ac), ENCSR249YGG (H3K4me1), ENCSR386RIJ (H3K27me3), ENCSR657DYL (H3K4me3) and ENCSR773IYZ (input).

Statistical analysis

Statistical analyses were conducted using GraphPad 9.0.0 and specific tests are indicated in the figure legends. For each figure, sample size n indicates the number of independent experiments or biological replicates and individual values are represented for every graph. Testing between groups was performed with at least $n \geq 3$ independent experiments and *p* value groups are indicated within the figure where significant.

Acknowledgements

This work was supported by funding from the European Research Council Grant New-Chol, the UK Regenerative Medicine Platform and a core support grant from the Wellcome MRC - Cambridge Stem Cell Institute. We thank Anna Osnato and Pedro Madrigal for bioinformatics support, and Stephanie Brown for technical advice. We acknowledge the

Wellcome Trust Sanger Institute sequencing platform, the ENCODE Consortium and the ENCODE production laboratories in generating the particular datasets used in this manuscript.

Competing Interests

L.V. is a founder and shareholder of DefiniGEN, Aculive Therapeutics and Billitech. F.B. is a PhD student sponsored by bit.bio. The remaining authors declare no competing interests.

664 **References**

665

666 Baxter, M., Withey, S., Harrison, S., Segeritz, C. P., Zhang, F., Atkinson-Dell, R., Rowe, C.,
667 Gerrard, D. T., Sison-Young, R., Jenkins, R., Henry, J., Berry, A. A., Mohamet, L., Best, M.,
668 Fenwick, S. W., Malik, H., Kitteringham, N. R., Goldring, C. E., Piper Hanley, K., Vallier,
669 L., & Hanley, N. A. (2015, Mar). Phenotypic and functional analyses show stem cell-derived
670 hepatocyte-like cells better mimic fetal rather than adult hepatocytes. *J Hepatol, 62*(3), 581-
671 589. https://doi.org/10.1016/j.jhep.2014.10.016

672

673 Bertero, A., Pawlowski, M., Ortmann, D., Snijders, K., Yiangou, L., Cardoso de Brito, M.,
674 Brown, S., Bernard, W. G., Cooper, J. D., Giacomelli, E., Gambardella, L., Hannan, N. R.,
675 Iyer, D., Sampaziotis, F., Serrano, F., Zonneveld, M. C., Sinha, S., Kotter, M., & Vallier, L.
676 (2016, Dec 1). Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies
677 of human development using hPSCs. *Development, 143*(23), 4405-4418.
678 https://doi.org/10.1242/dev.138081

679

680 Boon, R., Kumar, M., Tricot, T., Elia, I., Ordovas, L., Jacobs, F., One, J., De Smedt, J.,
681 Eelen, G., Bird, M., Roelandt, P., Doglioni, G., Vriens, K., Rossi, M., Vazquez, M. A.,
682 Vanwelden, T., Chesnais, F., El Taghdouini, A., Najimi, M., Sokal, E., Cassiman, D.,
683 Snoeys, J., Monshouwer, M., Hu, W. S., Lange, C., Carmeliet, P., Fendt, S. M., & Verfaillie,
684 C. M. (2020, Mar 13). Amino acid levels determine metabolism and CYP450 function of
685 hepatocytes and hepatoma cell lines. *Nat Commun, 11*(1), 1393.
686 https://doi.org/10.1038/s41467-020-15058-6

687

688 Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C. H., Lim, B., Vardy, L., Dunn, N. R.,
689 Trotter, M., Pedersen, R., & Vallier, L. (2011, Aug). Activin/Nodal signaling controls
690 divergent transcriptional networks in human embryonic stem cells and in endoderm
691 progenitors. *Stem Cells, 29*(8), 1176-1185. https://doi.org/10.1002/stem.666

692

693 Chen, G., Gulbranson, D. R., Hou, Z., Bolin, J. M., Ruotti, V., Probasco, M. D., Smuga-Otto,
694 K., Howden, S. E., Diol, N. R., Propson, N. E., Wagner, R., Lee, G. O., Antosiewicz-
695 Bourget, J., Teng, J. M., & Thomson, J. A. (2011, May). Chemically defined conditions for
696 human iPSC derivation and culture. *Nat Methods, 8*(5), 424-429.
697 https://doi.org/10.1038/nmeth.1593

Consortium, E. P. (2012, Sep 6). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. https://doi.org/10.1038/nature11247

Cook, D. N., Kang, H. S., & Jetten, A. M. (2015). Retinoic Acid-Related Orphan Receptors (RORs): Regulatory Functions in Immunity, Development, Circadian Rhythm, and Metabolism. *Nucl Receptor Res, 2*. https://doi.org/10.11131/2015/101185

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010, Dec 14). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A, 107*(50), 21931-21936. https://doi.org/10.1073/pnas.1016071107

Dhawan, A., Chaijitraruch, N., Fitzpatrick, E., Bansal, S., Filippi, C., Lehec, S. C., Heaton, N. D., Kane, P., Verma, A., Hughes, R. D., & Mitry, R. R. (2020, May). Alginate microencapsulated human hepatocytes for the treatment of acute liver failure in children. *J Hepatol, 72*(5), 877-884. https://doi.org/10.1016/j.jhep.2019.12.002

Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T., Zhao, D., Sun, Y., Shu, J., Guo, Q., Yin, M., Sun, D., Lu, S., Shi, Y., & Deng, H. (2014, Mar 6). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. *Cell Stem Cell, 14*(3), 394-403. https://doi.org/10.1016/j.stem.2014.01.008

Gordillo, M., Evans, T., & Gouon-Evans, V. (2015, Jun 15). Orchestrating liver development. *Development, 142*(12), 2094-2108. https://doi.org/10.1242/dev.114215

Grandy, R., Tomaz, R. A., & Vallier, L. (2019, Jan 24). Modeling Disease with Human Inducible Pluripotent Stem Cells. *Annu Rev Pathol, 14*, 449-468. https://doi.org/10.1146/annurev-pathol-020117-043634

Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E. E., Baydatch, S., Landen, S., Moor, A. E., Brandis, A., Giladi, A., Avihail, A. S., David, E., Amit, I., & Itzkovitz, S. (2017, Feb 16). Single-cell spatial reconstruction reveals

global division of labour in the mammalian liver. *Nature, 542*(7641), 352-356. https://doi.org/10.1038/nature21065

Hannan, N. R., Segeritz, C. P., Touboul, T., & Vallier, L. (2013, Feb). Production of hepatocyte-like cells from human pluripotent stem cells. *Nat Protoc, 8*(2), 430-437. https://doi.org/10.1038/nprot.2012.153

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010, May 28). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell, 38*(4), 576-589. https://doi.org/10.1016/j.molcel.2010.05.004

Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., Lai, D., Hu, Z., Chen, L., Zhang, Y., Cheng, X., Ma, X., Pan, G., Wang, X., & Hui, L. (2014, Mar 6). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell, 14*(3), 370-384. https://doi.org/10.1016/j.stem.2014.01.003

Inada, H., Udono, M., Matsuda-Ito, K., Horisawa, K., Ohkawa, Y., Miura, S., Goya, T., Yamamoto, J., Nagasaki, M., Ueno, K., Saitou, D., Suyama, M., Maehara, Y., Kumamaru, W., Ogawa, Y., Sekiya, S., & Suzuki, A. (2020, Oct 21). Direct reprogramming of human umbilical vein- and peripheral blood-derived endothelial cells into hepatic progenitor cells. *Nat Commun, 11*(1), 5292. https://doi.org/10.1038/s41467-020-19041-z

Kang, H. S., Angers, M., Beak, J. Y., Wu, X., Gimble, J. M., Wada, T., Xie, W., Collins, J. B., Grissom, S. F., & Jetten, A. M. (2007, Oct 22). Gene expression profiling reveals a regulatory role for ROR alpha and ROR gamma in phase I and phase II metabolism. *Physiol Genomics, 31*(2), 281-294. https://doi.org/10.1152/physiolgenomics.00098.2007

Kao, T. L., Chen, Y. L., Kuan, Y. P., Chang, W. C., Ho, Y. C., Yeh, S., Jeng, L. B., & Ma, W. L. (2017, Nov). Estrogen-Estrogen Receptor alpha Signaling Facilitates Bilirubin Metabolism in Regenerating Liver Through Regulating Cytochrome P450 2A6 Expression. *Cell Transplant, 26*(11), 1822-1829. https://doi.org/10.1177/0963689717738258

765 Kao, T. L., Kuan, Y. P., Cheng, W. C., Chang, W. C., Jeng, L. B., Yeh, S., & Ma, W. L.
766 (2018). Estrogen receptors orchestrate cell growth and differentiation to facilitate liver
767 regeneration. *Theranostics, 8*(10), 2672-2682. https://doi.org/10.7150/thno.23624
768
769 Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale,
770 J., Hughes, T. R., & Weirauch, M. T. (2018, Oct 4). The Human Transcription Factors. *Cell,*
771 *175*(2), 598-599. https://doi.org/10.1016/j.cell.2018.09.045
772
773 Lau, H. H., Ng, N. H. J., Loo, L. S. W., Jasmen, J. B., & Teo, A. K. K. (2018, May). The
774 molecular functions of hepatocyte nuclear factors - In and beyond the liver. *J Hepatol, 68*(5),
775 1033-1048. https://doi.org/10.1016/j.jhep.2017.11.026
776
777 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
778 dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550.
779 https://doi.org/10.1186/s13059-014-0550-8
780
781 Mitry, R. R., Hughes, R. D., & Dhawan, A. (2002, Dec). Progress in human hepatocytes:
782 isolation, culture & cryopreservation. *Semin Cell Dev Biol, 13*(6), 463-467.
783 https://doi.org/10.1016/s1084952102001350
784
785 Nakamori, D., Takayama, K., Nagamoto, Y., Mitani, S., Sakurai, F., Tachibana, M., &
786 Mizuguchi, H. (2016, Jan 15). Hepatic maturation of human iPS cell-derived hepatocyte-like
787 cells by ATF5, c/EBPalpha, and PROX1 transduction. *Biochem Biophys Res Commun,*
788 *469*(3), 424-429. https://doi.org/10.1016/j.bbrc.2015.12.007
789
790 Palakkan, A. A., Nanda, J., & Ross, J. A. (2017, Apr). Pluripotent stem cells to hepatocytes,
791 the journey so far. *Biomed Rep, 6*(4), 367-373. https://doi.org/10.3892/br.2017.867
792
793 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017, Apr). Salmon
794 provides fast and bias-aware quantification of transcript expression. *Nat Methods, 14*(4), 417-
795 419. https://doi.org/10.1038/nmeth.4197
796
797 Pawlowski, M., Ortmann, D., Bertero, A., Tavares, J. M., Pedersen, R. A., Vallier, L., &
798 Kotter, M. R. N. (2017, Apr 11). Inducible and Deterministic Forward Programming of

799 Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. *Stem*
800 *Cell Reports, 8*(4), 803-812. https://doi.org/10.1016/j.stemcr.2017.02.016
801
802 Rashid, S. T., Corbineau, S., Hannan, N., Marciniak, S. J., Miranda, E., Alexander, G.,
803 Huang-Doran, I., Griffin, J., Ahrlund-Richter, L., Skepper, J., Semple, R., Weber, A., Lomas,
804 D. A., & Vallier, L. (2010, Sep 1). Modeling inherited metabolic disorders of the liver using
805 human induced pluripotent stem cells. *J Clin Invest, 120*(9), 3127-3136. https://doi.org/43122
806 [pii]
807 10.1172/JCI43122
808
809 Ricchi, M., Odoardi, M. R., Carulli, L., Anzivino, C., Ballestri, S., Pinetti, A., Fantoni, L. I.,
810 Marra, F., Bertolotti, M., Banni, S., Lonardo, A., Carulli, N., & Loria, P. (2009, May).
811 Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in cultured
812 hepatocytes. *J Gastroenterol Hepatol, 24*(5), 830-840. https://doi.org/10.1111/j.1440-
813 1746.2008.05733.x
814
815 Rombaut, M., Boeckmans, J., Rodrigues, R. M., van Grunsven, L. A., Vanhaecke, T., & De
816 Kock, J. (2021, May 11). Direct reprogramming of somatic cells into induced hepatocytes:
817 cracking the Enigma code. *J Hepatol*. https://doi.org/10.1016/j.jhep.2021.04.048
818
819 Rudraiah, S., Zhang, X., & Wang, L. (2016). Nuclear Receptors as Therapeutic Targets in
820 Liver Disease: Are We There Yet? *Annu Rev Pharmacol Toxicol, 56*, 605-626.
821 https://doi.org/10.1146/annurev-pharmtox-010715-103209
822
823 Schrem, H., Klempnauer, J., & Borlak, J. (2002, Mar). Liver-enriched transcription factors in
824 liver function and development. Part I: the hepatocyte nuclear factor network and liver-
825 specific gene expression. *Pharmacol Rev, 54*(1), 129-158. https://doi.org/10.1124/pr.54.1.129
826
827 Segeritz, C. P., Rashid, S. T., de Brito, M. C., Serra, M. P., Ordonez, A., Morell, C. M.,
828 Kaserman, J. E., Madrigal, P., Hannan, N. R. F., Gatto, L., Tan, L., Wilson, A. A., Lilley, K.,
829 Marciniak, S. J., Gooptu, B., Lomas, D. A., & Vallier, L. (2018, Oct). hiPSC hepatocyte
830 model demonstrates the role of unfolded protein response and inflammatory networks in
831 alpha1-antitrypsin deficiency. *J Hepatol, 69*(4), 851-860.
832 https://doi.org/10.1016/j.jhep.2018.05.028

Shen, M., & Shi, H. (2015). Sex Hormones and Their Receptors Regulate Liver Energy Homeostasis. *Int J Endocrinol, 2015*, 294278. https://doi.org/10.1155/2015/294278

Si-Tayeb, K., Lemaigre, F. P., & Duncan, S. A. (2010, Feb 16). Organogenesis and development of the liver. *Dev Cell, 18*(2), 175-189. https://doi.org/10.1016/j.devcel.2010.01.011

Szkolnicka, D., & Hay, D. C. (2016, Jun). Concise Review: Advances in Generating Hepatocytes from Pluripotent Stem Cells for Translational Medicine. *Stem Cells, 34*(6), 1421-1426. https://doi.org/10.1002/stem.2368

Takeda, Y., Kang, H. S., Freudenberg, J., DeGraff, L. M., Jothi, R., & Jetten, A. M. (2014). Retinoic acid-related orphan receptor gamma (RORgamma): a novel participant in the diurnal regulation of hepatic gluconeogenesis and insulin sensitivity. *PLoS Genet, 10*(5), e1004331. https://doi.org/10.1371/journal.pgen.1004331

Takeda, Y., Kang, H. S., Lih, F. B., Jiang, H., Blaner, W. S., & Jetten, A. M. (2014). Retinoid acid-related orphan receptor gamma, RORgamma, participates in diurnal transcriptional regulation of lipid metabolic genes. *Nucleic Acids Res, 42*(16), 10448-10459. https://doi.org/10.1093/nar/gku766

Tilson, S. G., Morell, C. M., Lenaerts, A. S., Park, S. B., Hu, Z., Jenkins, B., Koulman, A., Liang, T. J., & Vallier, L. (2021, Dec). Modeling PNPLA3-Associated NAFLD Using Human-Induced Pluripotent Stem Cells. *Hepatology, 74*(6), 2998-3017. https://doi.org/10.1002/hep.32063

Touboul, T., Hannan, N. R., Corbineau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J., Weber, A., & Vallier, L. (2010, May). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology, 51*(5), 1754-1765. https://doi.org/10.1002/hep.23506

Trefts, E., Gannon, M., & Wasserman, D. H. (2017, Nov 6). The liver. *Curr Biol, 27*(21), R1147-R1151. https://doi.org/10.1016/j.cub.2017.09.019

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., Muth, K., Palmer, J., Qiu, Y., Wang, J., Lam, D. K., Raum, J. C., Stoffers, D. A., Ren, B., & Sander, M. (2015, Apr 2). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell, 16*(4), 386-399. https://doi.org/10.1016/j.stem.2015.02.013

Yiangou, L., Ross, A. D. B., Goh, K. J., & Vallier, L. (2018, Apr 5). Human Pluripotent Stem Cell-Derived Endoderm for Modeling Development and Clinical Applications. *Cell Stem Cell, 22*(4), 485-499. https://doi.org/10.1016/j.stem.2018.03.016

Yu, G., & He, Q. Y. (2016, Feb). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst, 12*(2), 477-479. https://doi.org/10.1039/c5mb00663e

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012, May). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287. https://doi.org/10.1089/omi.2011.0118

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordonez, A., Hannan, N. R., Rouhani, F. J., Darche, S., Alexander, G., Marciniak, S. J., Fusaki, N., Hasegawa, M., Holmes, M. C., Di Santo, J. P., Lomas, D. A., Bradley, A., & Vallier, L. (2011, Oct 20). Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature, 478*(7369), 391-394. https://doi.org/nature10424 [pii] 10.1038/nature10424

Zhao, D., Chen, S., Duo, S., Xiang, C., Jia, J., Guo, M., Lai, W., Lu, S., & Deng, H. (2013, Jan). Promotion of the efficient metabolic maturation of human pluripotent stem cell-derived hepatocytes by correcting specification defects. *Cell Res, 23*(1), 157-161. https://doi.org/10.1038/cr.2012.144

Zhu, L., Shi, J., Luu, T. N., Neuman, J. C., Trefts, E., Yu, S., Palmisano, B. T., Wasserman, D. H., Linton, M. F., & Stafford, J. M. (2018, Feb). Hepatocyte estrogen receptor alpha mediates estrogen action to promote reverse cholesterol transport during Western-type diet feeding. *Mol Metab, 8*, 106-116. https://doi.org/10.1016/j.molmet.2017.12.012

Tables

Table 1. Sequences of primers used for cloning.

Gene	**Primer**	**Sequence 5' - 3'** (GSG linker sequence underlined)
HNF1A	Start_F	CAC TTT TGT CTT ATA CTT ACT AGT GCC ACC ATG GTT TCT AAA CTG AGC CAG CTG CAG
HNF1A	P2A_R	TTC CAC GTC TCC TGC TTG CTT TAA CAG AGA GAA GTT CGT GGC TCC GGA GCC CTG GGA GGA AGA GGC CAT CTG G
HNF4A	E2A_F	TAT GCT CTC TTG AAA TTG GCT GGA GAT GTT GAG AGC AAC CCT GGA CCT GTC AGC GTG AAC GCG CCC CT
HNF4A	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA GAT AAC TTC CTG CTT GGT GAT GGT CG
HNF4A	P2A_F	TCT CTG TTA AAG CAA GCA GGA GAC GTG GAA GAA AAC CCC GGT CCT GTC AGC GTG AAC GCG CCC CT
HNF4A	T2A_R	CTC CTC CAC GTC ACC GCA TGT TAG AAG ACT TCC TCT GCC CTC TCC GGA GCC GAT AAC TTC CTG CTT GGT GAT GGT CG
HNF4A	Start_F	CAC TTT TGT CTT ATA CTT ACT AGT GCC ACC ATG GTC AGC GTG AAC GCG CCC
HNF4A	P2A_R	TTC CAC GTC TCC TGC TTG CTT TAA CAG AGA GAA GTT CGT GGC TCC GGA GCC GAT AAC TTC CTG CTT GGT GAT GGT CG
FOXA3	T2A_F	AGT CTT CTA ACA TGC GGT GAC GTG GAG GAG AAT CCC GGC CCT CTG GGC TCA GTG AAG ATG GAG GC
FOXA3	E2A_R	CTC AAC ATC TCC AGC CAA TTT CAA GAG AGC ATA ATT AGT ACA CTG TCC GGA GCC GGA TGC ATT AAG CAA AGA GCG GGA ATA G
FOXA3	P2A_F	TCT CTG TTA AAG CAA GCA GGA GAC GTG GAA GAA AAC CCC GGT CCT CTG GGC TCA GTG AAG ATG GAG GC
FOXA3	T2A_R	CTC CTC CAC GTC ACC GCA TGT TAG AAG ACT TCC TCT GCC CTC TCC GGA GCC GGA TGC ATT AAG CAA AGA GCG GGA ATA G
FOXA3	T2A_F	AGT CTT CTA ACA TGC GGT GAC GTG GAG GAG AAT CCC GGC CCT CTG GGC TCA GTG AAG ATG GAG GC
FOXA3	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA GGA TGC ATT AAG CAA AGA GCG GGA ATA G
HNF6	P2A_F	TCT CTG TTA AAG CAA GCA GGA GAC GTG GAA GAA AAC CCC GGT CCT AAC GCG CAG CTG ACC ATG GAA GC
HNF6	T2A_R	CTC CTC CAC GTC ACC GCA TGT TAG AAG ACT TCC TCT GCC CTC TCC GGA GCC TGC TTT GGT ACA AGT GCT TGA TGA AGA AGA T
HNF6	T2A_F	AGT CTT CTA ACA TGC GGT GAC GTG GAG GAG AAT CCC GGC CCT AAC GCG CAG CTG ACC ATG GAA GC
HNF6	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA TGC TTT GGT ACA AGT GCT TGA TGA AGA AGA T
RORγ	E2A_F	TAT GCT CTC TTG AAA TTG GCT GGA GAT GTT GAG AGC AAC CCT GGA CCT GAC AGG GCC CCA CAG AGA CAG
RORγ	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA CTT GGA CAG CCC CAC AGG TGA C
ESR1	E2A_F	TAT GCT CTC TTG AAA TTG GCT GGA GAT GTT GAG AGC AAC CCT GGA CCT ACC ATG ACC CTC CAC ACC AAA GCA
ESR1	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA GAC CGT GGC AGG GAA ACC CTC
AR	E2A_F	TAT GCT CTC TTG AAA TTG GCT GGA GAT GTT GAG AGC AAC CCT GGA CCT GAA GTG CAG TTA GGG CTG GGA AG
AR	Stop_R	AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CTA CTG GGT GTG GAA ATA GAT GGG CTT G

912 **Table 2.** Sequences of primers used for qPCR.

Gene	Forward	Reverse
AFP	TGCGGCCTCTTCCAGAAACT	TAATGTCAGCCGCTCCCTCG
ALB	CCTTTGGCACAATGAAGTGGGTAACC	CAGCAGTCAGCCATTTCACCATAG
APOA1	AGACAGCGGCAGAGACTATG	CCAGTTGTCAAGGAGCTTTAGG
CYP2A6	CAGCACTTCCTGAATGAG	AGGTGACTGGGAGGACTTGAGGC
CYP2C8	CATTACTGACTTCCGTGCTACAT	CTCCTGCACAAATTCGTTTTCC
CYP2C9	GCCGGCATGGAGCTGTTTTTAT	GCCAGGCCATCTGCTCTTCTT
CYP3A4	TGTGCCTGAGAACACCAGAG	GTGGTGGAAATAGTCCCGTG
FASN	GCAAGCTGAAGGACCTGTCT	AATCTGGGTTGATGCCTCCG
FOXA3	TGGGCTCAGTGAAGATGGAG	GGGGATAGGGAGAGCTTAGAG
G6PC	GTGTCCGTGATCGCAGACC	GACGAGGTTGAGCCAGTCTC
HHEX	GCCCTTTTACATCGAGGACA	AGGGCGAACATTGAGAGCTA
HNF1A	TGGCCATGGACACGTACAG	GCTGCTTGAGGGTACTTCTG
HNF4A	CATGGCCAAGATTGACAACCT	TTCCCATATGTTCCTGCATCAG
HNF6	GTGTTGCCTCTATCCTTCCCAT	CGCTCCGCTTAGCAGCAT
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG
NR1H4	ACTGAACTCACCCCAGATCAA	TGGTTGCCATTTCCGTCAAA
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT
PCK1	ACACAGTGCCCATCCCCAAA	GGTGCGACCTTTCATGCACC
POU5F1	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
PPARa	CCCTCCTCGGTGACTTATCC	CGGTCGCACTTGTCATACAC
PPARy	GAGCCTGCATCTCCACCTTAT	AGAAACCCTTGCATCCTTCACA
RORy	CTACGGCAGCCCCAGTTT	GCTGGCATGTCTCCCTGTA
RPLP0	GGCGTCCTCGTGGAAGTGAC	GCCTTGCGCATCATGGTGTT
SERPINA1	CCACCGCCATCTTCTTCCTGCCTGA	GAGCTTCAGGGGTGCCTCCTCTG
SOX17	CGCACGGAATTTGAACAGTA	GGA TCAGGGACCTGTCACAC
TBX3	TGGAGCCCGAAGAAGAGGTG	TTCGCCTTCCCGACTTGGTA
UGT1A1	TGATCCCAGTGGATGGCAGC	CAACGAGGCGTCAGGTGCTA
UGT1A6	GGAGCCCTGTGATTTGGAGAGT	GACCCCGGTCACTGAGAACC

913

914

915

916

917

918

919

920

921

922 **Table 3.** Lisf of primary antibodies.

Protein	Supplier	Catalog number	Host	Concentration
Albumin	Bethyl Laboratories	A80-229A	goat	1:100
Alpha-1 antitrypsin	Dako	A0012	rabbit	1:100
Alpha-Fetoprotein	Dako	A0008	rabbit	1:100
HNF4A	Abcam	ab92378	rabbit	1:100
HNF1A	Santa cruz	sc-135939	mouse	1:50
HNF6	Santa cruz	sc-13050	rabbit	1:100
FOXA3	Santa cruz	sc-166703	mouse	1:50
RORc	Abcam	ab221359	rabbit	1:100
ERα	Abcam	ab32063	rabbit	1:100
AR	Abcam	ab108341	rabbit	1:100

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941 **Figure Legends**

942

943 **Figure 1.** Forward programming of hPSCs into hepatocytes with 4 and 3 liver-enriched
944 transcription factors (LETFs). (A) Schematic representation of the two sequentially targeted
945 loci. The human ROSA26 was targeted with a constitutively expressed reverse tetracycline
946 transactivator (rtTA). The AAVS1 locus was targeted with the 4 LETFs *HNF1A*, *HNF6*,
947 *FOXA3* and *HNF4A* downstream of a Tet-responsive element (TET). (B) mRNA induction
948 levels of the 4 factors in targeted hESCs (Targ) relative to untargeted (Untarg) hESCs,
949 stimulated with dox for 24h (n=3). Data is shown relative to the untargeted control. (C)
950 Immunofluorescence staining of the 4 LETFs in targeted and untargeted hESCs after 24h of
951 inducible overexpression (iOX) with dox confirming transgene induction. Nuclei were
952 counterstained with DAPI (blue). Scale bar, 200µm. (D) Schematic representation of the iOX
953 culture conditions for forward programming. Phase contrast images of hESCs targeted with
954 the 4 LETFs after 10 and 15 days of forward programming. (E) mRNA levels of hepatocyte
955 markers (*ALB*, *SERPINA1* and *AFP*) in hESCs targeted with the 4 LETFs after 10 and 15
956 days of forward programming. Untargeted hESCs treated with the same protocol as in (D)
957 were used as control (n=4). Statistical difference was calculated with unpaired t-test against
958 untargeted. (F) CYP3A4 activity levels normalised per cell number (millions) in untargeted
959 and targeted hESCs with the 4 LETFs after 15 days of forward programming (n=5) and
960 HLCs generated by direct differentiation (n=6). Statistical difference between targeted and
961 untargeted cells was calculated with unpaired t-test. (G,H,I) mRNA levels of hepatocyte
962 markers (*ALB*, *SERPINA1* and *AFP*) in hESCs targeted with the 4 LETFs and with
963 combinations of 3 LETFs (n=4). The factor removed from each construct is indicated.
964 Expression levels were determined after 10, 15, 20 and 25 days of forward programming.
965 Statistical differences were calculated with one-way ANOVA, corrected for multiple

966 comparisons compared to 4 LETFs. All mRNA levels were normalised to the average of 2
967 housekeeping genes (*PBGD* and *RPLP0*). (J) CYP3A4 activity levels normalised per cell
968 number (millions) in hESCs targeted with the 4 LETFs and combinations of 3 LETFs after
969 10, 15, 20 and 25 days of forward programming (n=3-5). Statistical differences were
970 calculated with one-way ANOVA, corrected for multiple comparisons compared to 4 LETFs.
971 In all plots, bars represent mean with SD, and individual datapoints are shown for all
972 biological replicates. Significant p-values are shown at each comparison and indicated as *p<
973 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.
974 The following supplements are available for Figure 1:
975 **Figure supplement 1**. Validation of inducible overexpression (iOX).
976 **Figure supplement 2**. Characterisation of the phenotype of hPSCs forward programmed into
977 hepatocytes with 3 LETFs (*HNF1A*, *FOXA3* and *HNF6*).
978 **Source data 1.** Individual measurements and statistical tests related to Figure 1.
979
980 **Figure 1 – figure supplement 1**. Validation of inducible overexpression (iOX). (A)
981 Schematic representation of the combinations of 3 LETFs cloned into the AAVS1 safe
982 harbour. (B) Immunofluorescence staining of all 4 LETFs in hESCs targeted with the
983 constructs shown in (A) after 24h of iOX with dox confirming transgene induction. Nuclei
984 were counterstained with DAPI (blue). Scale bars, 200µm.
985
986 **Figure 1–figure supplement 2**. Characterisation of the phenotype of hPSCs forward
987 programmed into hepatocytes with 3 LETFs. (A) Phase contrast images and (B)
988 immunofluorescence staining of hepatocyte markers Albumin (yellow), A1AT (green) and
989 AFP (cyan) in hESCs forward programmed with either 4 or 3 LETFs for 20 days. Untargeted

990 hESCs treated with the same protocol for 20 days were used as control. Nuclei were
991 counterstained with DAPI (blue). Scale bar, 200µm.
992
993 **Figure 2.** HLCs and PHHs display transcriptomic differences associated with their state of
994 maturation. (A) Immunofluorescence staining of Albumin (yellow) and HNF4A (red) in
995 HLCs differentiated for 30 days. Nuclei were counterstained with DAPI (blue). Scale bar,
996 100µm. (B) CYP3A4 activity levels normalised per cell number (millions) in HLCs
997 differentiated for 30 days (n=6) and PHHs (n=4). Bars represent mean with SD, and
998 individual datapoints represent the different biological replicates. Statistical difference was
999 calculated with unpaired t-test. (C) PCA of undifferentiated hiPSCs, HLCs derived from
1000 hESC (hESC_HLCs) and hiPSC (hiPSC_HLCs), freshly harvested PHHs (fPHHs) or plated
1001 PHHs (pPHHs). (D) Heatmap showing the proportion of genes differentially expressed in
1002 each cell type (cluster 1 - PHHs, cluster 2 - HLCs, cluster 4 - hiPSCs) as well as in Heps
1003 (HLCs and PHHs) against undifferentiated hiPSCs (cluster 3). (E, F) Dotplot showing the top
1004 15 hits on gene ontology enrichment analysis on genes associated to cluster 1 and cluster 3 as
1005 shown in (D). The size of each dot represents number of genes associated to each term and
1006 the colours represents the adjusted p-value. (G) Heatmap showing the differential gene
1007 expression of transcription factors between PHHs (fresh or plated) and HLCs (hESC and
1008 hiPSC derived). (H) Reactome pathway enrichment analysis on transcription factors
1009 identified in (G). Differential gene expression was calculated with log2(fold change) higher
1010 than 2 and adjusted p-value <0.05. Hierarchical clustering on samples was generated by
1011 Euclidean distance.
1012 The following supplements are available for Figure 2:
1013 **Figure supplement 1**. Characterisation of transcriptional differences between HLCs and
1014 PHHs.

Source data 1. Individual measurements and statistical tests related to Figure 2 and corresponding supplements.

Source data 2. List of genes differentially expressed in the 4 clusters in Figure 2D.

Figure 2-figure supplement 1. Characterisation of transcriptional differences between HLCs and PHHs. (A) mRNA levels measured by qPCR throughout 30 days of hiPSC differentiation into HLCs. Representative markers of each stage of differentiation are shown: pluripotency (*NANOG*, *POU5F1*), endoderm (*SOX17*), foregut (*HHEX*), hepatoblast (*TBX3*), foetal (*AFP*) and mature (*ALB*) hepatocytes. (B) mRNA levels of the 4 LETFs tested in forward programming quantified throughout 30 days of hiPSC direct differentiation into HLCs. (C) Comparison of the mRNA levels of the 4 LETFs in differentiated HLCs in (B) with PHHs. (D) mRNA levels of hepatocyte markers *ALB*, *SERPINA1* and *AFP* in undifferentiated hiPSCs, 30-day differentiated HLCs and PHHs. (E) mRNA levels of hepatocyte markers *CYP3A4*, *CYP2A6* and *CYP2C9* in undifferentiated hiPSCs, HLCs and PHHs. All mRNA levels were normalised to the average of 2 housekeeping genes (*PBGD* and *RPLP0*). Bars represent mean with SD, and individual datapoints are shown for n=4 biological replicates. Statistical difference was calculated with one-way ANOVA, corrected for multiple comparisons compared to PHHs. P-values are indicated as *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001. (F,G) Dotplot showing the top 15 hits on gene ontology enrichment analysis on genes associated to cluster 2 and cluster 4 as shown in Figure 2D. The size of each dot represents number of genes associated to each term and the colours represents the adjusted p-value.

Figure 3. Epigenetic status of regulatory regions differs between states of maturation in HLCs and PHHs. (A) PCA of the global enrichment profile of H3K27ac, H3K4me1 and

H3K27me3 across 2 replicates of undifferentiated hiPSCs, hESC and hiPSC-derived HLCs, and PHHs. Average scores were computed for genomic regions of 1000bp for the entire genome. (B) Average density plots and heatmaps showing enrichment levels for H3K27ac, H3K4me1 and H3K27me3 within a 10Kb window centred at H3K27ac PHH-unique (blue) or HLC-unique (green) regions. Scales are adjusted to maximum peak intensity for each dataset. (C) Enrichment profiles of H3K27ac, H3K4me1 and H3K27me3 across the *UGT1A* locus. Profiles are shown for one replicate of undifferentiated hiPSCs, hESC and hiPSC-derived HLCs, and PHHs. Red bars represent PHH-unique H3K27ac peaks. (D) Nuclear receptor motifs identified as overrepresented binding sites at H3K27ac PHH-unique regions.

The following supplements are available for Figure 3:

Figure supplement 1. Comparison of the epigenetic profile of HLCs and PHHs.

Source data 1. Peak annotation results for PHH-unique H3K27ac regions

Source data 2. Peak annotation results for HLC-unique H3K27ac regions

Source data 3. Motif enrichment results for PHH-unique H3K27ac regions

Figure 3-figure supplement 1. Comparison of the epigenetic profile of HLCs and PHHs. (A) Distance to TSS of H3K27ac peaks identified in the PHH and HLC datasets. (B) Enrichment profiles of H3K27ac, H3K4me1 and H3K27me3 across the *CYP3A4* locus. Profiles are shown for one replicate of undifferentiated hiPSCs, hESC and hiPSC-derived HLCs, and PHHs. Red bars represent PHH-unique H3K27ac peaks. (C) Bar plot showing the top 15 hits on gene ontology enrichment analysis on genes annotated to PHH or HLC-unique H3K27ac positive regions. The colour gradient represents the adjusted p-value. (D) PCA of the expression profiles in mouse livers harvested at E12.5, E16.5, P0 and adult stages (8 weeks and 10 weeks). Two biological replicates of each timepoint are shown. (E) Heatmap showing the relative expression of genes annotated as nuclear receptors across different

stages of liver development, ordered left to right. Hierarchical clustering on gene expression levels was generated by Euclidean distance. Top clusters represent genes with highest expression changes between foetal and adult stages. (F) Normalised counts for 3 nuclear receptors (*Rory*, *Esr1* and *Ar*) across different stages of liver development, showing expression increasing in the adult stage.

Figure 4. Forward programming of hESCs into hepatocytes with nuclear receptors. (A) Phase contrast images and (B) immunofluorescence staining for Albumin (yellow) and (C) A1AT (green) in hESCs forward programmed for 20 days with 3TFs alone or in combination with the nuclear receptors RORc, ER α and AR. Nuclei were counterstained with DAPI (blue). Scale bars, 200µm. (D) mRNA levels of hepatocyte markers (*ALB*, *SERPINA1* and *AFP*) in FoP-Heps generated with 3TFs alone or in combination with nuclear receptors for 20 and 30 days (n=4). Expression data was normalised to the average of 2 housekeeping genes (*PBGD* and *RPLP0*). (E) Protein secretion levels of Albumin, A1AT and AFP in hESC derived FoP-Heps generated with 3TFs alone or in combination with nuclear receptors for 20 days (n=4). Data was normalised per total cell number (millions). (F) CYP3A4 activity levels normalised per cell number (millions) in FoP-Heps targeted with 3TFs with or without nuclear receptors, after 20 and 30 days of forward programming (n=3-6). Statistical differences were calculated with one-way ANOVA, corrected for multiple comparisons compared to 3TFs (day 20). (G) CYP3A4 fold induction levels in FoP-Heps treated with 100nM of the ligands as indicated from day 2. Data is normalised to untreated control at day 20 of forward programming (n=3). Statistical differences were calculated with paired t-test. In all plots, bars represent mean with SD, and individual datapoints are shown for all biological replicates. Significant p-values are shown at each comparison and indicated as *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.

The following supplements are available for Figure 4:

Figure supplement 1. Validation of iOX of combinations of 3TFs and nuclear receptors.

Figure supplement 2. Forward programming of hiPSCs into hepatocytes with 4TFs.

Source data 1. Individual measurements and statistical tests related to Figure 4 and supplements.

Figure 4-figure supplement 1. Validation of iOX of combinations of 3TFs and nuclear receptors . (A) Schematic representation of the combinations of factors cloned into the AAVS1 locus. (B) Immunofluorescence staining of the 3 TFs and nuclear receptors in hESCs targeted with the constructs shown in (A) after 24h of iOX with dox confirming transgene induction. (C) Immunofluorescence staining of nuclear receptors in hESCs targeted with 3TFs alone, as negative control. Scale bars, 200µm. (D) Zoomed images of immunofluorescence staining of nuclear receptors showing nuclear localisation. Scale bars, 50µm. Nuclei were counterstained with DAPI (blue). (E) Immunofluorescence staining for AFP (cyan) in hESCs forward programmed for 20 days with 3TFs alone or in combination with the nuclear receptors RORc, ER α and AR. Nuclei were counterstained with DAPI (blue). Scale bars, 200µm.

Figure 4–supplement 2. Forward programming of hiPSCs into hepatocytes with 4TFs. (A) Immunofluorescence staining of the 3TFs in hiPSCs targeted with 3TFs alone or with RORc after 24h of iOX with dox confirming transgene induction. Scale bars, 200µm. (B) Phase contrast images and (C) immunofluorescence staining for Albumin (yellow), (D) A1AT (green) and (E) AFP (cyan) in hiPSCs forward programmed for 20 days with 3TFs alone or with RORc. Nuclei were counterstained with DAPI (blue). Scale bars, 200µm. (F) mRNA

levels of hepatocyte markers (*ALB*, *SERPINA1* and *AFP*) in hiPSC derived FoP-Heps generated with 3TFs alone or with RORc for 20 and 30 days (n=4). Statistical differences were calculated with unpaired t-test. All expression data was normalised to the average of 2 housekeeping genes (*PBGD* and *RPLP0*). (G) Protein secretion levels of Albumin, A1AT and AFP in hiPSC derived FoP-Heps generated with 3TFs alone or with RORc for 20 days (n=4). Data was normalised per total cell number (millions). (H) CYP3A4 activity levels normalised per cell number (millions) in hiPSC FoP-Heps targeted with 3TFs with or without RORc after 20 days of forward programming (n=6). Statistical differences were calculated with one-way ANOVA, corrected for multiple comparisons compared to 3TFs (day 20). In all plots, bars represent mean with SD, and individual datapoints are shown for all biological replicates. Significant p-values are shown at each comparison and indicated as *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.

Figure 5. 4TF FoP-Heps are transcriptionally equivalent to HLCs. (A) CYP3A4 activity levels in eFoP-Heps at day 20 (n=6) and day 30 (n=6), iFoP-Heps at day 22 (n=6) and day 30 (n=3), against direct differentiation HLCs (n=6) and PHHs (n=4). Statistical differences were tested with one-way ANOVA, corrected for multiple comparisons, between eFoP-Heps group and HLCs. (Statistical test of HLC vs PHHs can be found in Figure 2B) (B) mRNA levels of phase I (*CYP3A4*, *CYP2A6* and *CYP2C8*) and phase II (*UGT1A6*) biotransformation enzymes in 4TF FoP-Heps, HLCs and PHHs (n=4). (C) mRNA level of *ALB*, gluconeogenesis (*G6PC* and *PCK1*), and lipid (*PPARγ*) metabolism in FoP-Heps, HLCs and PHHs (n=4). Statistical differences were calculated with one-way ANOVA, corrected for multiple comparisons, and all samples compared to HLCs. P-values are indicated as *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001. (D) PCA of undifferentiated hiPSCs, HLCs derived from hESC (hESC_HLCs) and hiPSC (hiPSC_HLCs), freshly harvested PHHs (fPHHs) or plated PHHs

(pPHHs), foetal livers (FL) and 4TF hESC-derived FoP-Heps (eFoP_Heps). (E) Dotplot showing the top 15 hits on gene ontology enrichment analysis on genes associated to genes differentially expressed between eFoP-Heps and undifferentiated hiPSCs. The size of each dot represents number of genes associated to each term and the colours represents the adjusted p-values. (F) Heatmap showing expression of genes associated with adult hepatocyte functions found to be expressed in hiPSC-derived HLCs (hiPSC_HLCs) and 4TF hESC-derived FoP-Heps (eFoP_Heps), as compared to undifferentiated hiPSCs. Two clusters where separated as genes expressed (bottom) and not expressed (top) in foetal liver samples. (Foetal_Liver).

The following supplements are available for Figure 5:

Figure supplement 1. Characterisation of 4TF FoP-Heps.

Figure supplement 2. Characterisation of the transcriptome of 4TF FoP-Heps.

Source data 1. Individual measurements and statistical tests related to Figure 5 and corresponding supplements.

Figure 5-figure supplement 1. Characterisation of 4TF FoP-Heps. (A) mRNA levels of phase II (*UGT1A1*) biotransformation enzymes, lipid (*APOA1*, *FASN, PPARα*) and bile acid metabolism (*NR1H4*), and (B) 4TFs (*HNF1A*, *FOXA3*, *HNF6 RORγ*) in hESC (eFoP) and hiPSC (iFoP) derived FoP-Heps targeted with the 4TFs and forward programmed for 20 days (n=4), HLCs and PHHs (n=4). Statistical differences were calculated with one-way ANOVA, corrected for multiple comparisons, compared to HLCs. P-values are indicated as *p < 0.05, **p < 0.01, ***p< 0.001, ****p< 0.0001.

Figure 5-supplement 2. Characterisation of the transcriptome of 4TF eFoP-Heps. Dotplot showing the top 15 hits on gene ontology enrichment analysis on genes differentially

1165 expressed between eFoP and foetal liver (A, B), adult PHHs (C, D), hESC-derived HLCs (E,
1166 F), and hiPSC-derived HLCs (G, H). Differential gene expression was calculated with
1167 log2(fold change) higher than 2 and adjusted p-value <0.05.
1168
1169 **Figure 6**. RORc promotes functionality of 4TF FoP-Heps. (A) Immunofluorescence staining
1170 for LDL in FoP-Heps at day 20 of forward programming. Scale bars, 200µm. Nuclei were
1171 counterstained with DAPI (blue). (B) Comparison of mRNA levels of *SERPINA1, ALB*, *AFP*,
1172 *UGT1A6* , *G6PC* and *APOA1* in FoP-Heps cultured in 2D and 3D for up to 20 or 30 days of
1173 forward programming (n=4). Statistical difference between 2D and 3D were calculated with
1174 unpaired t-test. All expression data was normalised to the average of 2 housekeeping genes
1175 (*PBGD* and *RPLP0*). In all plots, bars represent mean with SD, and individual datapoints are
1176 shown for all biological replicates. P-values are indicated as *p < 0.05, **p < 0.01, ***p<
1177 0.001, ****p< 0.0001. (C) BODIPY staining of FoP-Heps cultured in 3D from day 20 of
1178 forward programming and treated with fatty acids (oleic acid [OA], palmitic acid [PA] or
1179 BSA [Ctr]) as indicated for 7 days. Scale bars, 200µm. Nuclei were counterstained with
1180 DAPI (blue). (D) Cell viability in FoP-Heps treated with the fatty acids as indicated,
1181 normalised against FoP-Heps treated with BSA as control (n=4). (E) Cell viability in FoP-
1182 Heps treated with 25mM of acetaminophen (APAP) for 48h in 3D cultures, normalised
1183 against untreated FoP-Heps (n=4). Significant differences were determined with paired t-test.
1184 In all plots, bars represent mean with SD, and individual datapoints are shown for all
1185 biological replicates. P-values are indicated as *p< 0.05, **p< 0.01, ***p< 0.001, ****p<
1186 0.0001.
1187
1188

Figure 1

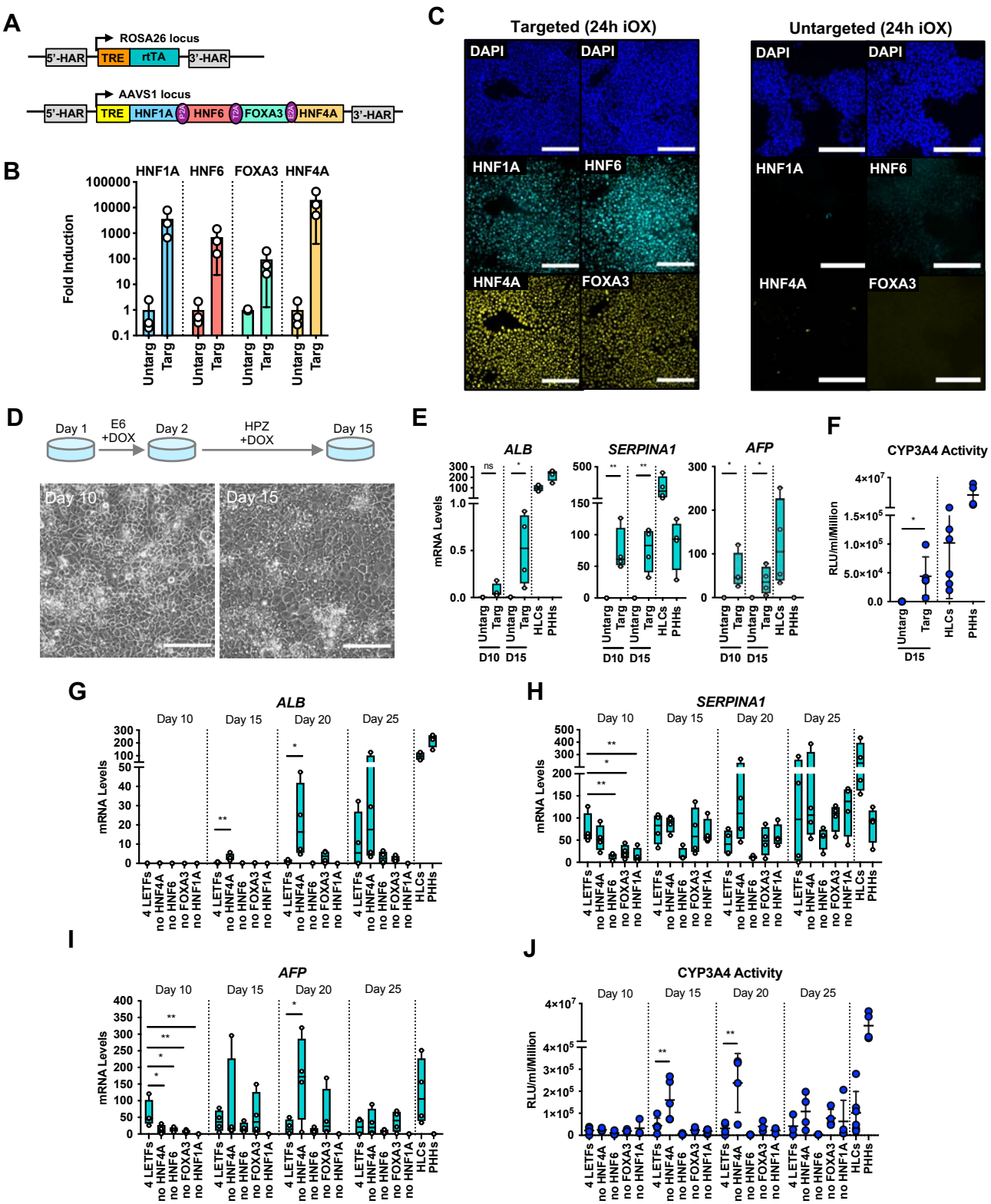

Figure 1 – figure supplement 1

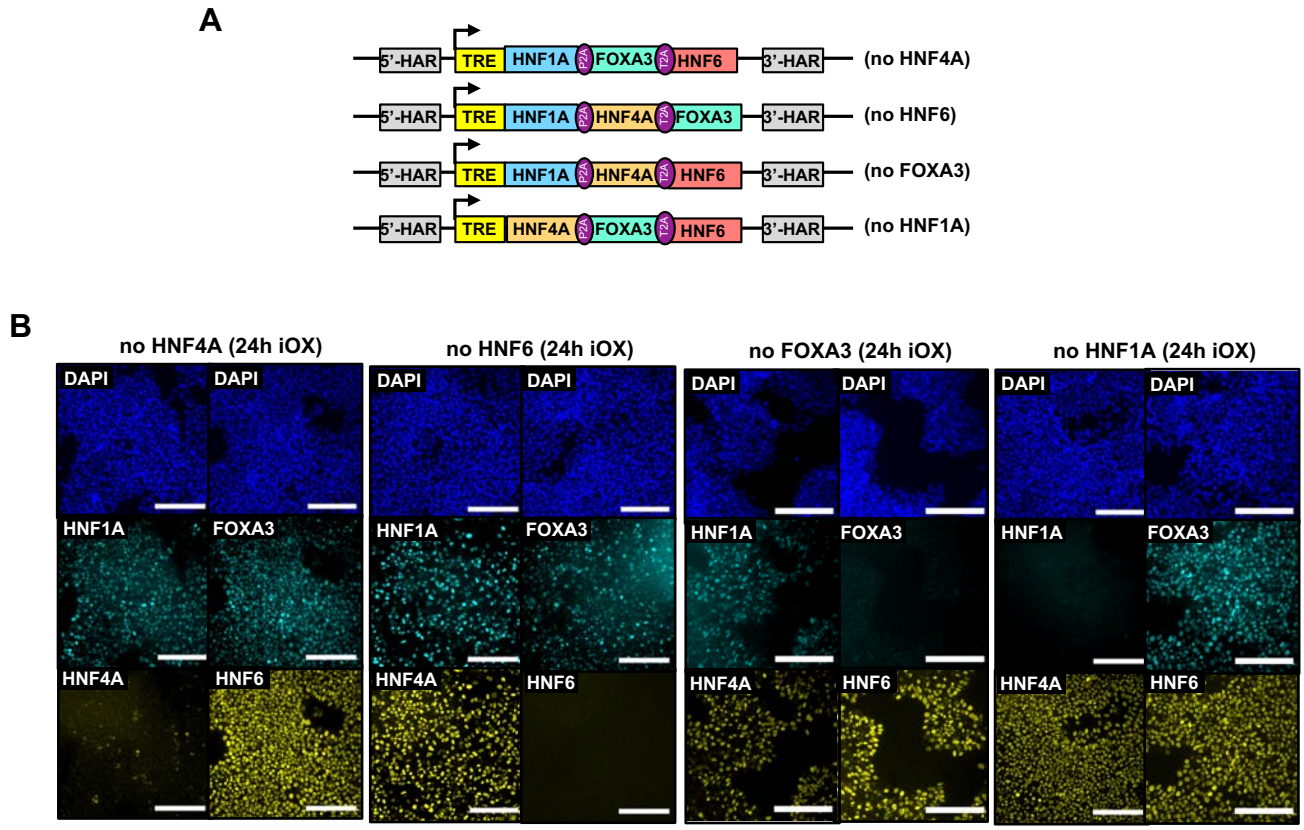

Figure 1 – figure supplement 2

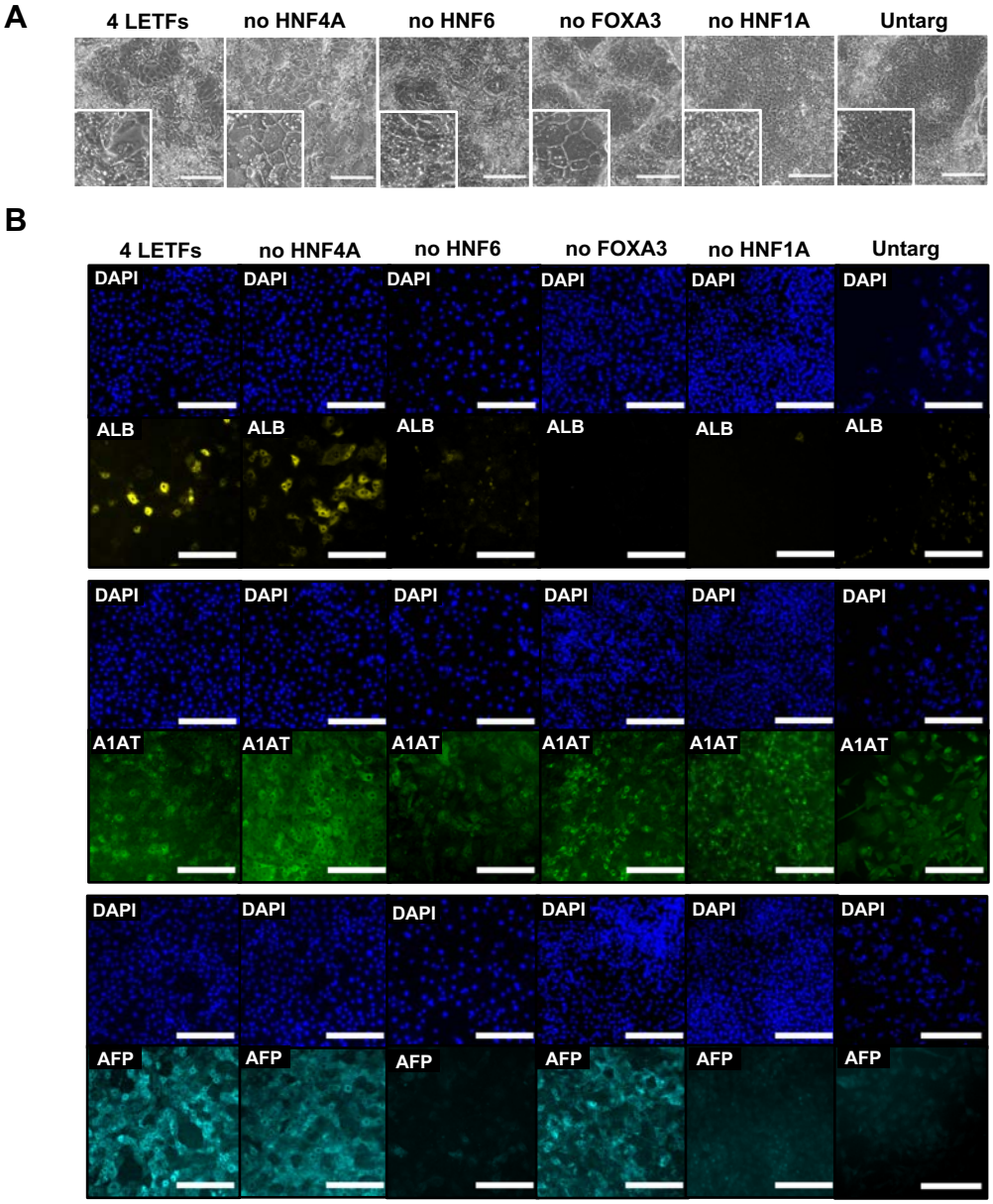

Figure 2

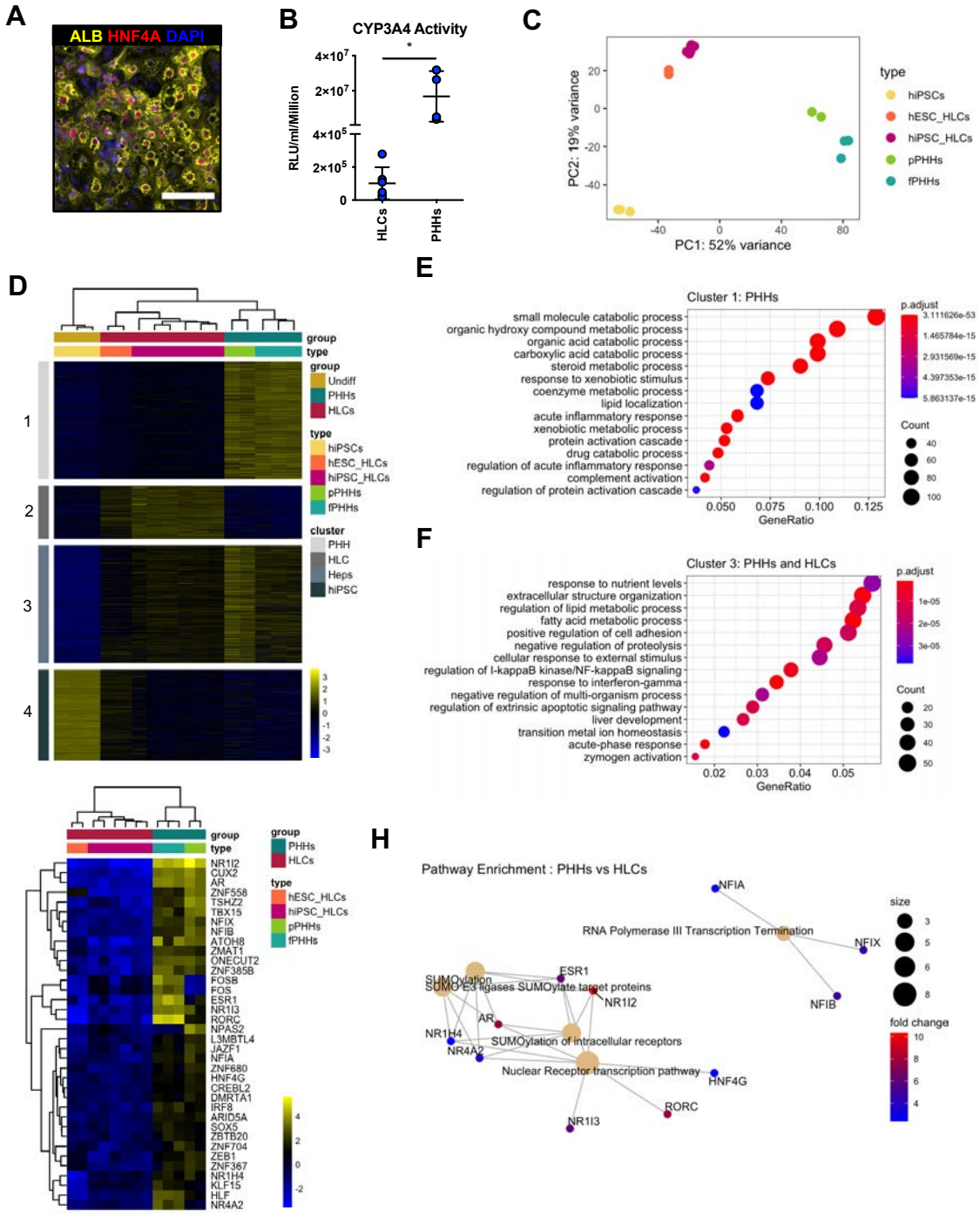

Figure 2 – figure supplement 1

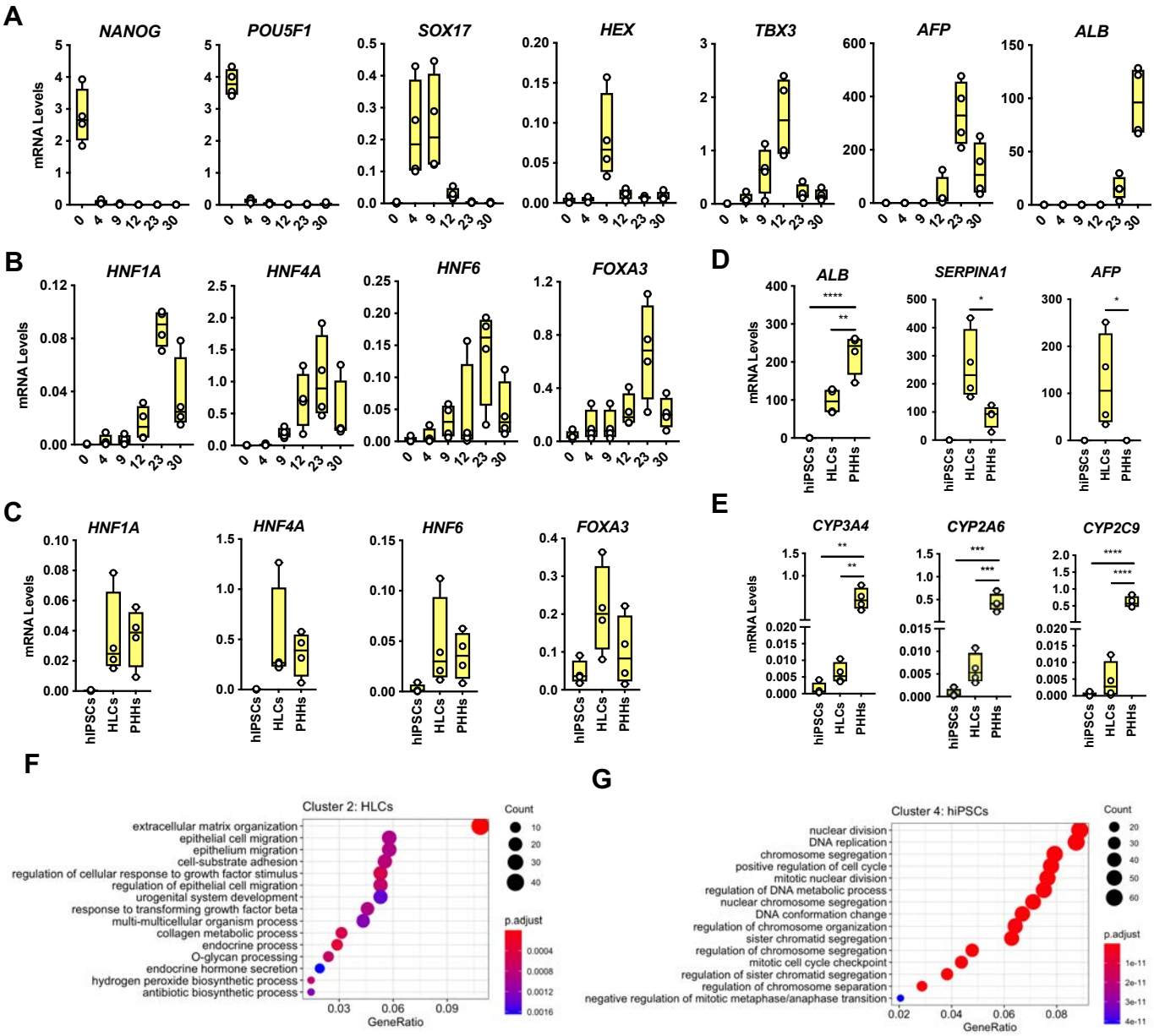

Figure 3

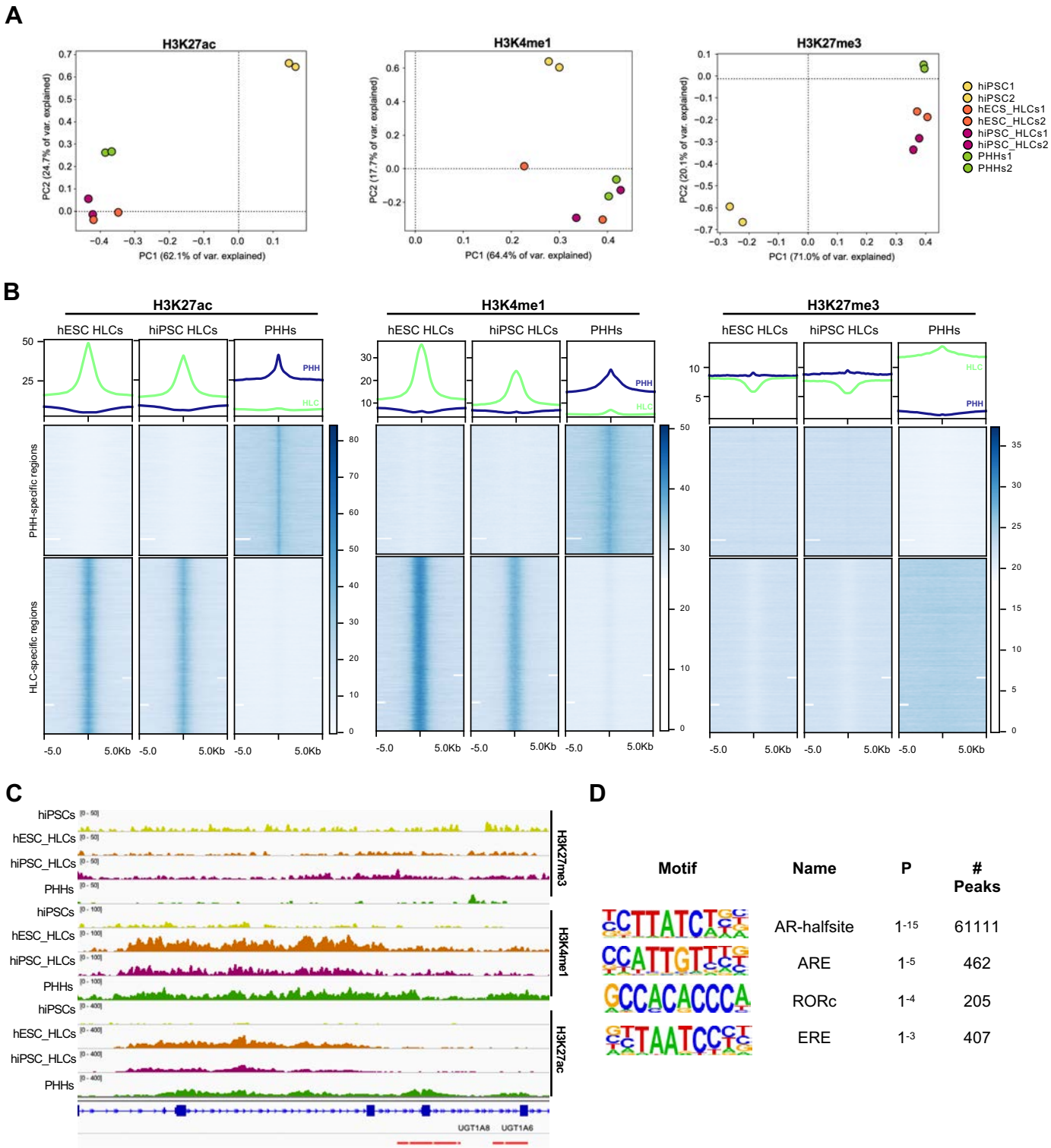

Figure 3 – figure supplement 1

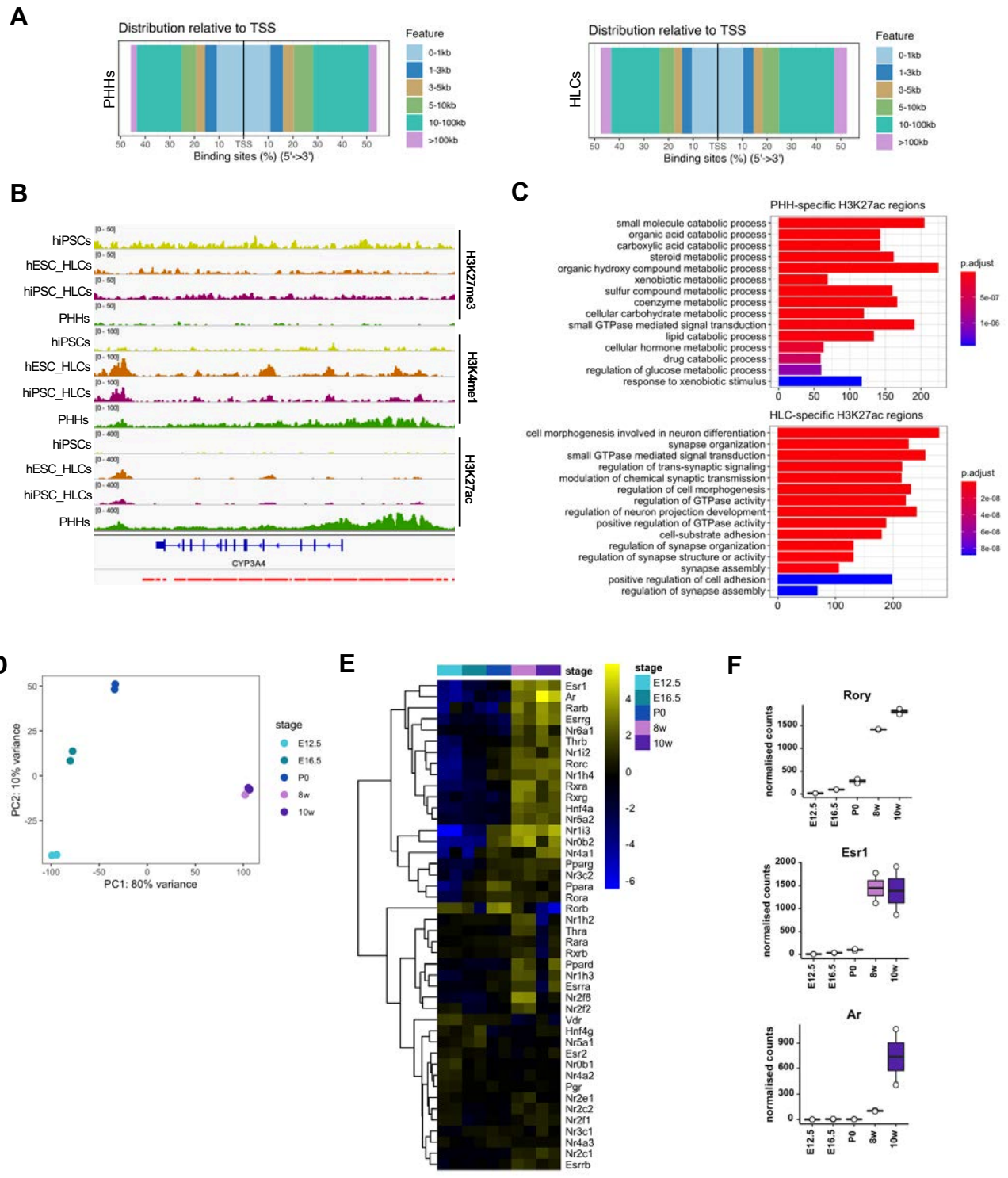

Figure 4

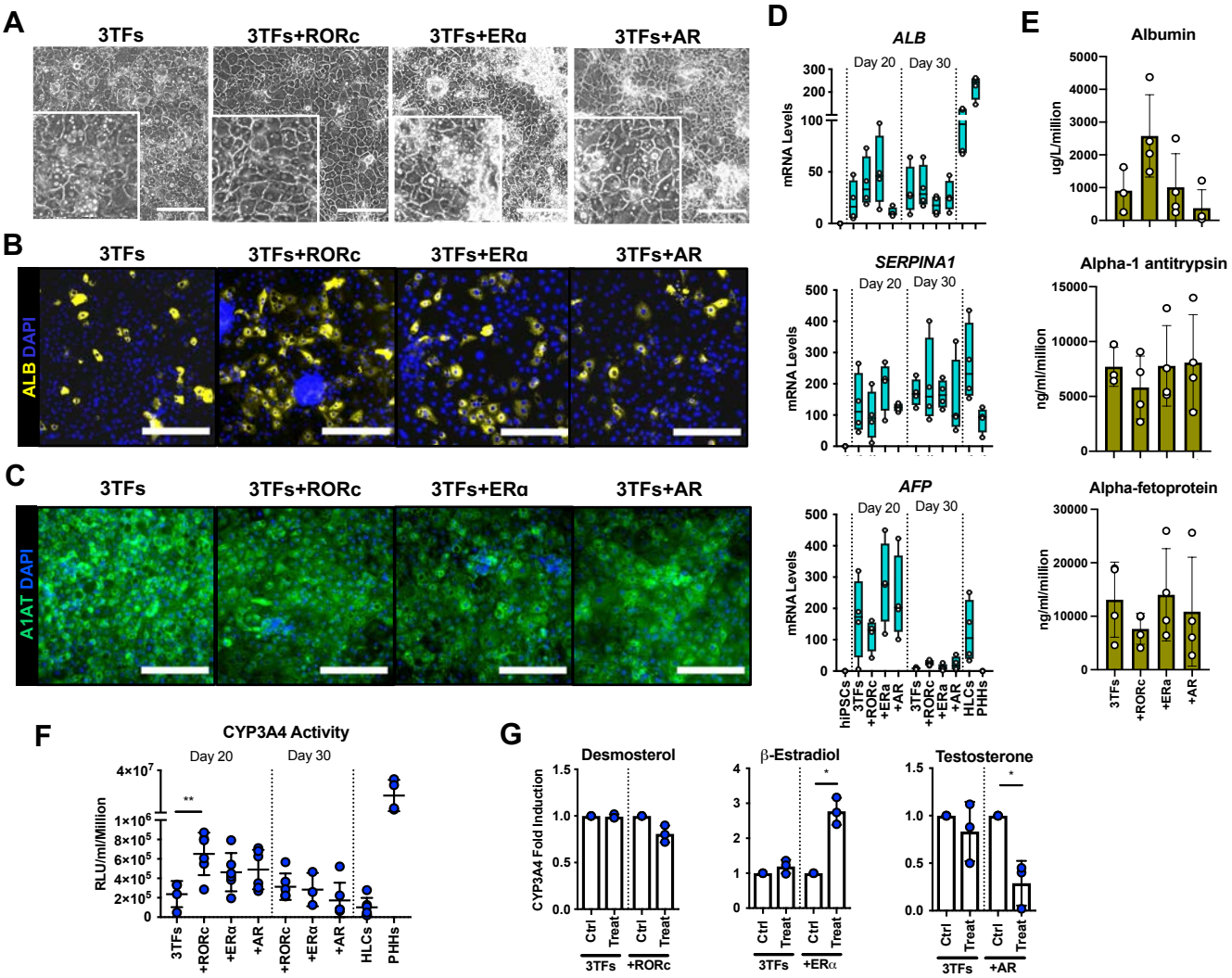

Figure 4 – figure supplement 1

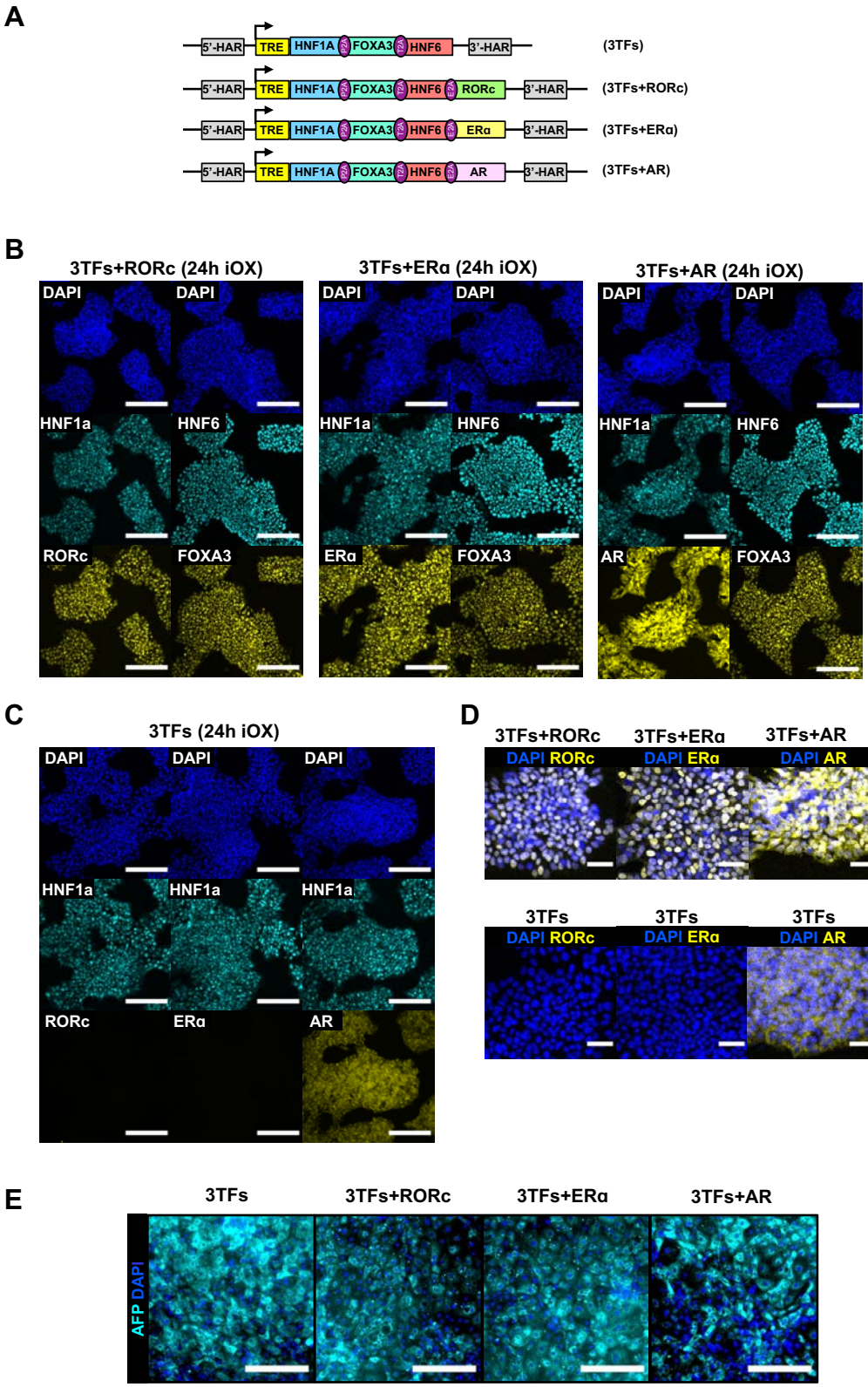

Figure 4 – figure supplement 2

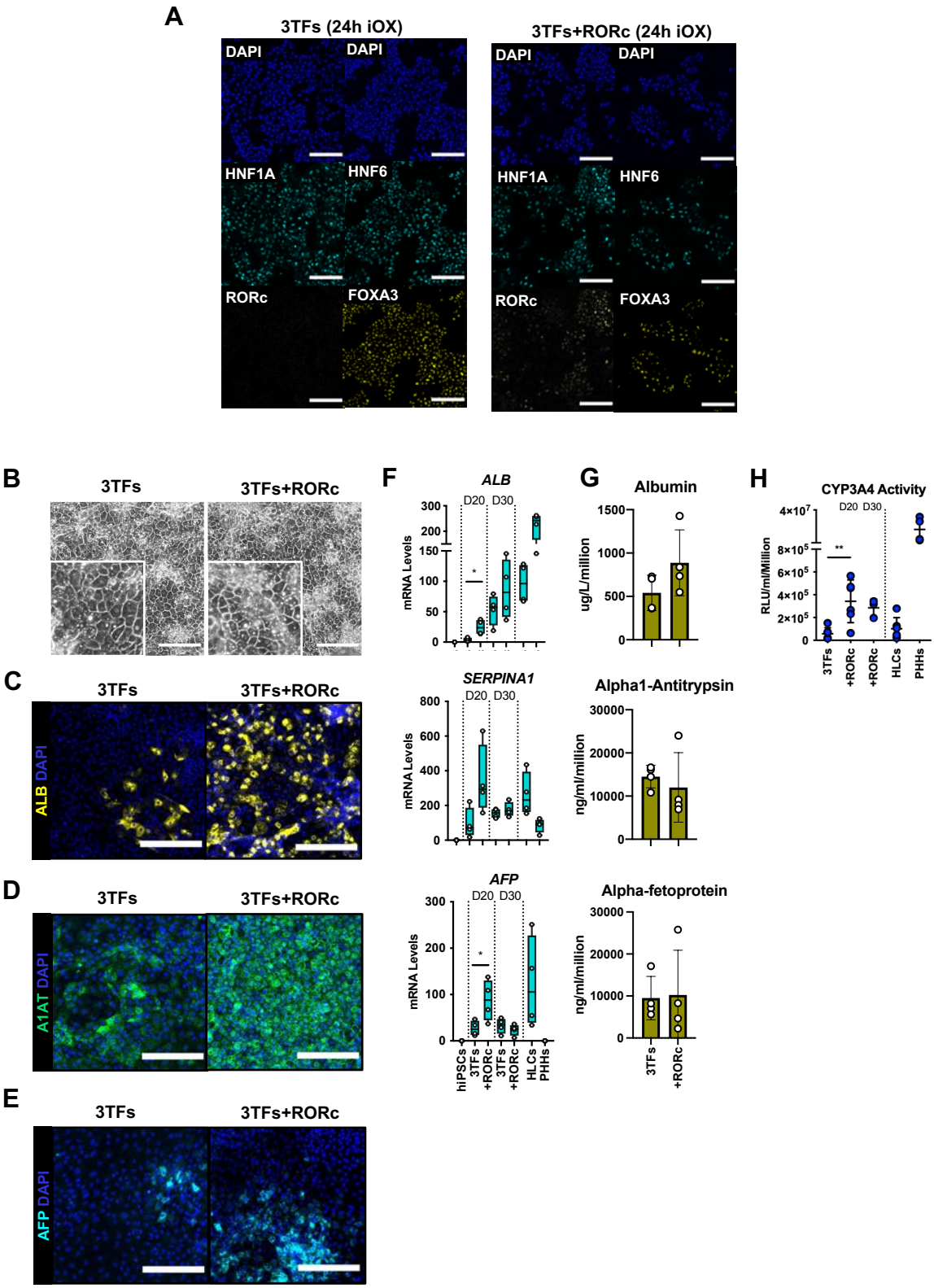

Figure 5

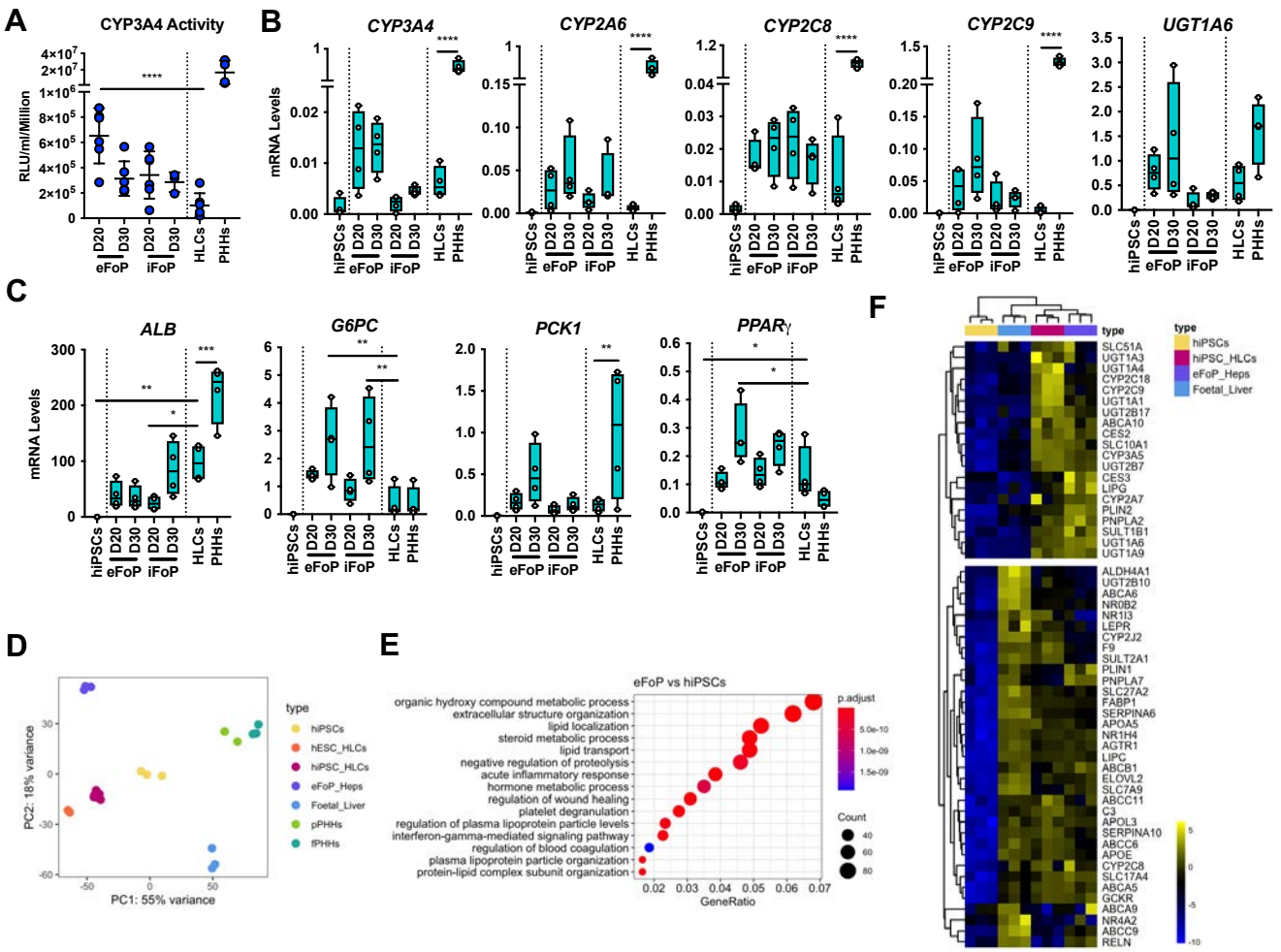

Figure 5 – figure supplement 1

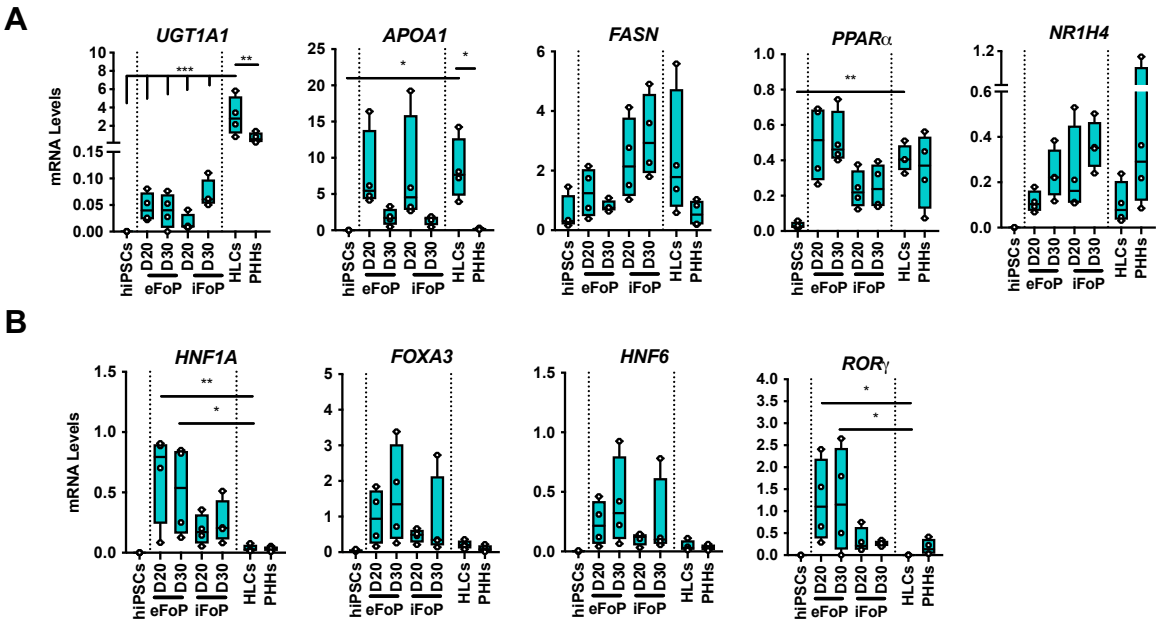

Figure 5 – figure supplement 2

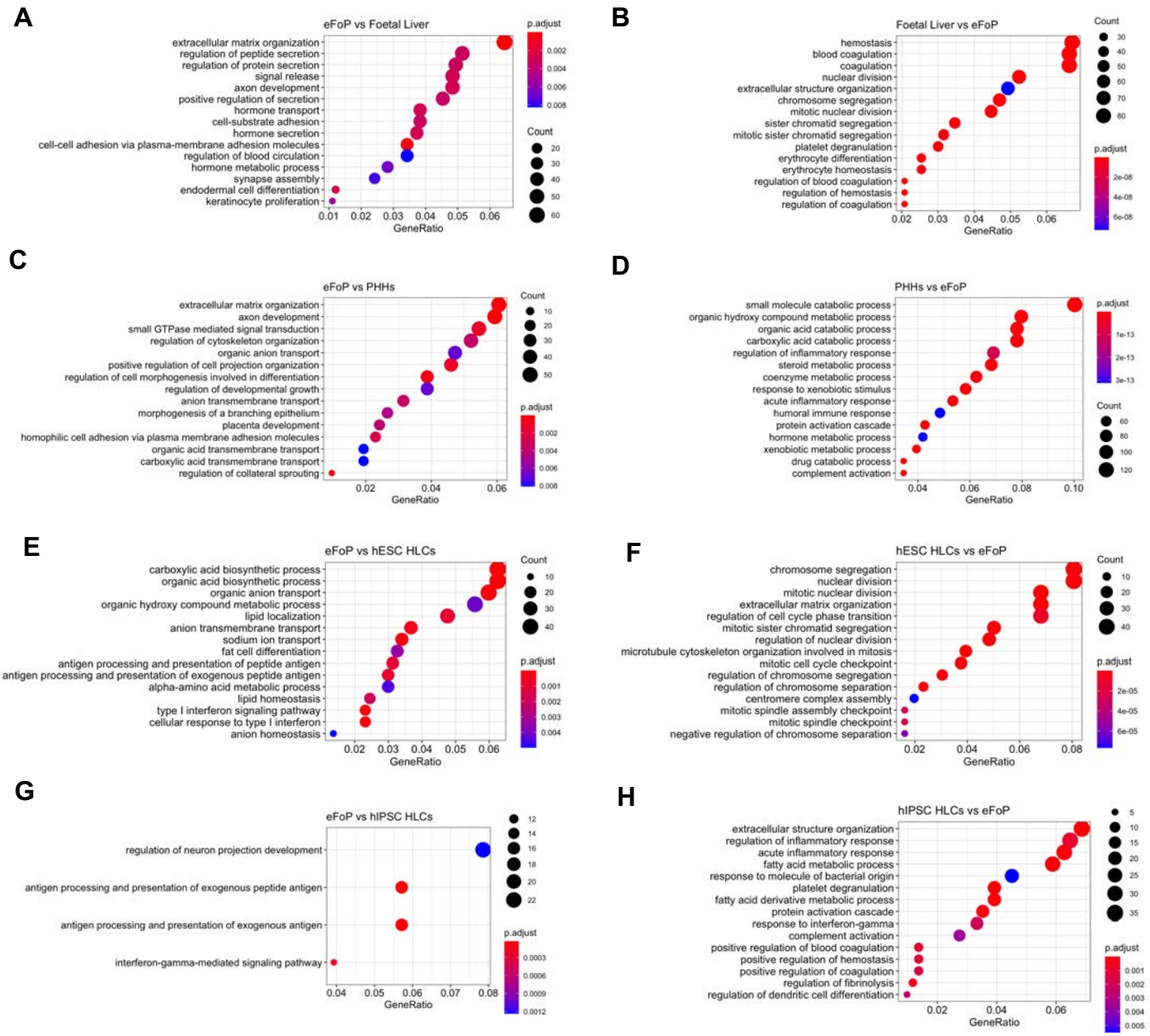

Figure 6