Analysis of the transcription factors and their regulatory roles during a step-by-step differentiation of induced pluripotent stem cells into hepatocyte-like cells

Yannick Y. Tauran, Stéphane Poulain, Myriam Lereau-Bernier, Mathieu Danoy, Marie Shinohara, Bertrand-David Segard, Sachi Kato, Taketomo Kido, Atsushi Miyajima, Yasuyuki Sakai, et al.

Analysis of the transcription factors kinetic and of their regulatory roles during a step-by-step differentiation of induced pluripotent stem cells into hepatocyte-like cells

Yannick Tauran[1,2], Stéphane Poulain[3,4], Myriam Lereau-Bernier[1], Mathieu Danoy[1], Marie Shinohara[5], Bertrand-David Segard[1], Sachi Kato[2,3], Taketomo Kido[6],Atsushi Miyajima[6], Yasuyuki Sakai[5], Charles Plessy[2,3], Eric Leclerc[1*]

[1] CNRS UMI 2820; Laboratory for Integrated Micro Mechatronic Systems, Institute of Industrial Science, University of Tokyo; 4-6-1 Komaba; Meguro-ku; Tokyo 153-8505, Japan

[2] Univ Lyon, Université Claude Bernard Lyon 1, Laboratoire des Multimatériaux et Interfaces, UMR CNRS 5615, F-69622 Villeurbanne, France

[3] RIKEN Center for Life Science Technologies, Division of Genomic Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045 Japan.

[4] RIKEN Center for Integrative Medical Sciences, Division of Genomic Medicine, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045 Japan.

[5] CIBIS; Institute of Industrial Science; The University of Tokyo; 4-6-1 Komaba; Meguro-ku; Tokyo 153-8505, Japan.

[6] Laboratory of Stem Cell Therapy, Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

* corresponding author: Eric Leclerc

CNRS UMI 2820; Laboratory for Integrated Micro Mechatronic Systems,

Institute of Industrial Science, University of Tokyo; 4-6-1 Komaba; Meguro-ku;

Tokyo, 153-8505, Japan

Abstract

We investigated the human induced pluripotent stem cells (hiPSCs) during a sequential *in vitro* step-by-step differentiation into hepatocyte-like cells (HLCs) using nanoCAGE, a method for promoters, transcription factors, and transcriptome analysis. Specific genes clusters reflected the different steps of the hepatic differentiation. The proliferation step was characterized by a typical cell cycle and DNA replication. The hepatic endoderm and the HLC steps were marked by a common signature including cell interactions with extracellular matrix (ECM), lipoproteins and hepatic biomarkers (such as albumin and alpha-fetoprotein). The specific HLC profile was characterized by important transcription factors such as HIF1A, JUN, MAF, KLF6, BMP4 and with a larger expression of genes related to Wnt signaling, extracellular matrix, lipid metabolism, urea cycle, drugs, and solute transporters. HLC profile was also characterized by the activation of upstream regulators such as HNF1A, MEIS2, NFIX, WRNIP1, SP4, TAL1. Their regulatory networks highlighted HNF4a as a bridge and linked them to important processes such as EMT-MET transitions, ECM remolding and liver development pathways (HNF3, PPARA signaling, iron metabolism) along the different step of differentiation.

Keywords : NanoCAGE, human induced pluripotent stem cells, hepatocytes, differentiation

1. Introduction

Early studies in liver regenerative medicine explored the transplantation of mature/differentiated cells into the liver to be repaired (Dhawan, 2015). Even though early clinical trials showed some therapeutic effectiveness, this has proven insufficient in most situations. Failures were mainly attributed to the high rate of cell death, to the absence or loss of liver functions, or to the absence of division after implantation with no effect in acute liver failure and very limited effect in patients with hereditary metabolic liver diseases. For these reasons, research is now turning towards the use of other types of cells, mostly engineered from stem cells, and towards the construction of liver tissue, organoids or more complex architectures (Takebe et al., 2017). This has opened a new area of research with extraordinary developments for liver clinical stem cell research (Tsuchiya et al., 2017, Lanzoni et al., 2013).

Important hurdles with the liver reconstruction include the large variety of cell types and architectural complexities, and the need to construct a biliary network – which has not been well addressed–, as well as the regulatory issues to enter clinical application – which needs to evolve rapidly. In 2014, there were also 6700 patients on the waiting list for liver transplant (European report of Journalist Workshop on Organ donation and transplantation, Recent Facts & Figures, held on the 26 November 2014 – Brussels). In parallel, the limited availability of functional liver model for drug testing is reported as a major bottleneck bringing pharmaceutical companies to spend $1 billion/year on liver cells alone (Lin and Khetani, 2016). Providing higher production of functional liver cells supply from human pluripotent stem cells can change this situation.

Hepatocytes are differentiated from human induced pluripotent stem cells emerged as a promising source for *in vitro* liver model (Takahashi et al, 2007). Several studies have shown that human pluripotent cells can be differentiated into hepatocyte-like cells (Touboul et al, 2010, Hay et al., 2008, Cai et al, 2007). The hepatocyte-like cells express major liver phenotypic markers and mimic hepatic metabolism. Although the metabolism of xenobiotics was also described (Rashid et al, 2010, Holmgren et al., 2014), several studies reported that the level of functionality of enzymes such as cytochromes P450 (CYP450) remained weak when compared to mature hepatocytes. In fact, the hepatocyte-like cells still present immature markers such as high levels of AFP expression, which illustrates an incomplete maturation (Rashid et al, 2010; Touboul et al., 2010). In addition, variability and heterogeneous maturations into HLCs are observed from various iPS sources while the same protocol of differentiation is being followed (Godoy et al., 2015). Therefore, HLCs derived from hiPSCs are, at present, encouraging but still controversial model for mimicking the adult hepatocytes.

As a result, additional approaches and investigations have to be integrated in order to increase our knowledge of the differentiation process. In this frame, we used two protocols of cultures based on sequential differentiation leading to albumin productive cells and CYP3A4 positive cells (Lereau Bernier et al., 2019). Although we compared the transcriptome in the HLCs in those two protocols, we did not yet integrate the information resulting from the promoter analysis and the transcriptomic profile from each step of differentiation. Thus, in order to extend the knowledge during this sequential differentiation, we propose in this paper a dedicated sequential step-by-step analysis with the nanoCAGE technology to simultaneously investigate transcriptome profiles and promoter usages in the differentiating cells. We

first characterized gene expression patterns during the pluripotency and self-renewal step (IPSC), then during the first step of differentiation when the iPSCs were derived into definitive endoderm (DE, 5 days of differentiation). The transcriptome of progenitor hepatocytes at the hepatic endoderm (HE, 15 days of differentiation) and final hepatocytes like cells (HLCs, 22 days) were also targeted to give a full overview of expression dynamics and to identify the differentially expressed markers at each step of the protocol. Main promoter activities and related transcription factor binding motifs were further inferred from CAGE data to provide additional information about the specific gene networks sequentially involved in the differentiation process.

2. Materials and methods

2.1 iPS cell source

The iPS cells used in this study (TkDN4-M clones) were provided from the stem cell bank of the Institute of Medical Science of the University of Tokyo (established by Takayama et al., 2010, using a retrovirus harboring four reprogramming factors, *OCT3/4*, *SOX2*, *KLF4*, and *c-MYC*). TkDN4-M human induced pluripotent stem (hiPS) cells were cultivated following previously published protocols for cell differentiation (Si Tayeb et al., 2010).

2.2 iPS differentiation protocol

Stem cells are cultivated on standard TCPS culture dish for initial expansion and all differentiation steps. Culture dishes (6 well plates) are coated with (20mg/mL) Matrigel Matrix suited for iPS cells maintenance (Corning Matrigel hESC-Qualified Matrix, ref. 354277) following manufacturer's instructions (incubation for 1h at 37°C). The exceeding amount of Matrigel is washed with DMEM before cell seeding. TkDN4-M cells are seeded at 10000 cells/cm^2. For maintenance and expansion, cells are cultivated in mTeSR 1 medium (StemCell Technologies, ref. 05850) complemented with the RHO/ROCK pathway inhibitor Y-27632 (StemCell Technologies, ref. 72302) for enhanced cells survival at thawing. After 24h and until the end of the expansion, the medium is renewed daily with pure mTeSR 1. During this initial process, the atmospheric condition is set at 37°C, 20% O_2, and 5%CO_2. The differentiation protocol is initiated when cell confluence reaches 90% (usually after 3 to 4 days).

The differentiation protocol consisted of four steps (here: S1, S2, S3, and S4; S0 corresponding to the proliferation step, supplementary file 1) during which media compositions and atmospheric conditions are specifically defined. During the first step (S1, 5 days, 37°C, 20% O_2, 5% CO_2), hiPSCs are cultivated in RPMI-1640 medium (Gibco) supplemented with B27 (Gibco) and Activin A (R&D Systems) at 100 ng/mL to achieve a DE commitment. Then, as DE cells enter S2 (5 days, 37°C, 5% O_2, 5% CO_2), they are cultivated in RPMI-1640 / B27 supplemented with FGF2 (bFGF, Peprotech) at 10 ng/mL and BMP4 (Humanzyme) at 20 ng/mL. At the end of S2, cells commitment corresponds to hepatoblast progenitors (Early_Pro). During S3 (5 days, 37°C, 5% O_2, 5% CO_2), cells are cultivated in RPMI-1640 / B27 supplemented with HGF at 20 ng/mL to reach the hepatocytes progenitor commitment (or HE). In step 4 (S4, 37°C, 20% O_2, 5% CO_2), cells are cultivated in

hepatocyte culture medium (Lonza HBM, ref. CC-3199 & CC-4182, without hEGF) supplemented with OSM (R&D Systems) at 20 ng/mL for 7 days to differentiate in HLCs.

In this protocol, the medium is renewed every 24 hours.

2.3 RT-qPCR

Total RNAs were isolated and purified from samples with Trizol™ Reagent (Life Technologies, Japan) following the manufacturer's instructions. Concentrations and qualities of extracted RNAs were assessed using a BioSpec-nano (Shimadzu Scientific Instruments, Japan). Reverse-transcription into cDNA was performed from 0.5µg total RNA by using the ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO, Japan). Real-time quantitative PCR was then performed with the THUNDERBIRD SYBR qPCR Mix (TOYOBO, Japan) according to the manufacturer's protocol and StepOnePlus Real-Time PCR system (Applied Biosystems). Primer sequences of OCT3/4, FOXA2, HNF4a, ALB, AFP, CYP3A4, CDH1, CDH2, SNAI1, SNAI2, BMP4 are shown in supplementary file 1. ACTB (β-Actin) was used as reference gene.

2.4 CAGE analysis

CAGE protocols were carried out accordingly to our previous works. Briefly, nanoCAGE libraries were prepared and sequenced as in Poulain et al., 2017. The total RNA was extracted from cell samples conserved in TRIzol (ThermoFisher) with the PureLink RNA mini kit (ThermoFisher). 250 ng from each RNA sample underwent

reverse transcription individually using pseudo-random primers (Arnaud et al., 2016) in order to reduce the amount of sequencing reads mapping to ribosomal DNA. The resulting cDNA samples, tagged by specific index sequences, were then multiplexed to produce a nanoCAGE library that was finally sequenced paired-end (7.53 pM + 10% PhiX) on MiSeq system with the Reagent Kit v3 150-cycles (Illumina).

2.5 CAGE post processing

FASTQ files generated by the sequencer were processed for CAGE scan analysis (Plessy et al., 2010; Kratz et al., 2014) with the MOIRAI pipeline (Hasegawa et al., 2014) OP-WORKFLOW-CAGEscan-short-reads-v2.1. Briefly, sequencing reads were demultiplexed and trimmed to 50 nucleotides with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Then, reads originating from rRNA or oligo-artifacts were removed with TagDust (v2.33) (Lassman, 2015) and the remaining reads were aligned on the human genome (hg19 and hg38) with BWA (v0.7.5a) (Li and Durbin. 2010). Non-proper pairs of reads and PCR duplicates were filtered out with the same tools (v0.1.19) (Li et al., 2009), and properly paired reads belonging to the same molecule were finally clustered with the tools “pairedBamToBed12” (https://github.com/charles-plessy/pairedBamToBed12) version 1.1. Resulting BED files were analyzed with custom R scripts using CAGEr (Haberle et al., 2015) to produce expression tables that were uploaded on the web (http://ge-lab.org:3838/idep/) for differential expression and pathway analysis with iDEP [Steven Xijin Ge, bioRxiv, https://doi.org/10.1101/148411]. iDEP genes set were analysed with FDR equal to 0,1 (adjusted p_value < 0,1). Motif Activity Response Analysis (MARA) was performed online (https://ismara.unibas.ch/fcgi/mara) with ISMARA (Balwierz et al., 2014) and the human genome version hg19. Data sets

were deposited in Zenodo (FASTQ files DOI: 10.5281/zenodo.1014009 and MOIRAI output DOI: 10.5281/zenodo.1017276).

2.6 Albumin measurements

To measure the production of albumin, we performed ELISA sandwich assays in a 96-well plate. The surface activation step was performed using anti-Human Albumin IgG (Bethyl, Japan) diluted at 1/1000 in a Tris buffer (0.1 M, pH = 8.0) overnight at 4°C. The passivation step was realized with gelatin (2 %) diluted in PBS and Tween 20 (0.1 %) and incubated for 1 hour. The samples were incubated overnight at 4°C. After washing, the second anti-Human Albumin IgG coupled with peroxidase (Bethyl, Japan) was incubated overnight at 4°C. The peroxidase activity was revealed by a mixture of H_2O_2 and OPD in citrate buffer (0.05 mM, pH = 7.4). The reaction was stopped with a sulfuric acid solution (4N) and the plate was read at 490nm (iMark Microplate reader, Bio-Rad).

2.7 Immunostaining

Fixation with 4% paraformaldehyde was performed overnight on all prepared samples at 4°C. Samples were then washed and stored at 4°C until the immunostaining started. Permeabilization and blocking of the samples were done with 0.1% Triton X-100 (plusone) and a pork gelatin buffer for 15 minutes and 2 hours respectively. Samples were then incubated with first antibodies (CYP3A4, HNF4a, ALB, AFP, diluted at 1/100, 1/1000, 1/1000 and 1/1000 respectively) overnight at 4°C followed with secondary antibodies diluted at 1/200 for 3 hours at room temperature.

After incubation with antibodies, samples were washed 3 times with PBS. Incubation with DAPI was performed for 15 min, the samples were then washed twice with PBS and kept in PBS for confocal imaging. The antibodies used were Albumin (goat, A80-129A, Bethyl), AFP (goat, sc-8109, Santa Cruz), CYP3A4 (Rabbit, ab135813, abcam), HNF-4Alpha (Rabbit, ab181604, abcam), anti-goat Alexa Fluor 488 (Donkey, ab150129, abcam) and anti-rabbit Alexa Fluor 568 (Donkey, A10042, Thermofisher).

2.8 Statistics

The results from RTqPCR and ELISA were analyzed by Mann and Whitney's tests were for the statistical analysis. Data with * denotes P_value below 5%. iDEP processing of the CAGE dataset was analysed with a false discovery rate of 0,1 in the gene set (adjusted p_value < 0,1). Pathway enrichment analysis was retained with adjusted p_value below 10^{-6}.

3. Results

3.1 Sequential hepatocyte-like cells differentiation process is confirmed by Morphological views, RTqPCR and immunostaining analysis

We confirmed at first the differentiation of the hiPSCs into HLCs as we already established previously this protocol. The morphologies of hiPSCs at the initial proliferation step and at the final HLCs step are given in figures 1A to 1B. Cells were switching from a typical pluripotent stem cells phenotype to a dense tissue made of cells having a polygonal large shape and often a large nucleus. Thus, typical cuboidal hepatocyte phenotype was observed as shown by the arrow in figure 1B.

To confirm the differentiation of iPS cells into HLCs, we performed RTqPCR analysis and followed critical markers such as OCT4, FOXA2, HNF4A, PXR, CYP3A4 and ALB (Figure 1C). We found that OCT4 mRNA level from DE step to HLCs continuously decreased during the differentiation. Both HNF4A and FOXA2 had a peak of mRNA levels in HE. The ALB mRNA level was highly expressed in HE and HLCs steps. Similarly, PXR mRNA was detected in both HE and HLCs steps. Finally, the levels of mRNA of CYP3A4 increased from HE to the HLCs.

To bridge the steps of differentiation with cellular development, we followed the typical mesenchymal-epithelial transition (MET) and typical epithelial-mesenchymal transition (EMT) genes. In DE stage, we found upregulation of BMP4 and CDH2 compared to IPSC stage. At the end of the process, we found an increase of BMP4, SNAI1 and SNAI2 genes whereas CDH2 decrease when compared to DE step (Figure 1D).

Finally, the functionality of hepatic derived cells was confirmed by measuring the levels albumin that increased from 50 to 2500ng/mL/day from HE to HLCs respectively (Figure 1E). The differentiation was also characterized by immunostaining (Figures 1F to 1I) in which the presence of the HNF4A, ALB, and CYP3A4 proteins was demonstrated by the presence of cells in the HLCs step.

3.2 Multivariate analysis of the transcriptome illustrated specific patterns in iPSC cells, definitive endoderm, hepatic endoderm and hepatic like cells.

3.2.1 Hierarchical and K-means clusterings separated the differentiation steps

We have analyzed the transcriptome evolution during the differentiation and the important genes set associated with each step. The hierarchical clustering of the

transcriptomic data performed in iDEP illustrated the step-by-step differentiation process, as shown in figure 2A. The cell development stages were clearly separated which highlighted the iPSCs, DE, HE, and HLCs differentiation steps. Furthermore, the dendrogram grouped the proliferation step and the DE under the same branch of the tree. Later differentiation steps, such as HE and HLCs were grouped under the second branch of the tree.

We extracted six typical clusters of genes by k-means clustering in the overall dataset (Figure 2B and supplemental file 2 for all gene list). The clusters, representing the proliferation step of iPSCs, DE, HE, and HLCs differentiation, were composed of 562, 198, 465 and 510 genes respectively. In addition, to these four clusters, two other clusters were identified. The fifth cluster (176 genes) grouped genes there were common to the proliferation step and to the definitive endoderm. The last cluster (89 genes) represented a common signature for both HE and HLCs.

3.2.2 iPSC is characterized by DNA processing and cell cycle processes

The analysis of the enriched pathways resulting from the different k-means clusters was analyzed with iDEP using the gene ontology annotation (GO biological process), as shown in table 1. The iPSC proliferation step cluster was characterized in the GO biological process by the cell cycle, DNA metabolic process, mitotic cell process (Table 1). The iPSC proliferation step cluster was discriminated by important cell cycle genes such as CCNB1 and CCNA2, DNA processing genes such as DNA2, BRCA2, and PCNA. It also included mitochondrial processing genes such as WDR12 and HDPE1 (cluster gene list is provided in supplemental file 2, the associated gene network is given in figure S1 of supplemental file 1 and the genes involved in the cluster network in supplemental file 3).

3.2.3 DE is characterized by morphogenesis processes

The GO biological process pathways extracted from the DE cluster corresponded to the anatomical structure morphogenesis, the regulation signaling and the intracellular transduction (Table 1). The definitive endoderm step cluster included LGR5, CXCR4 and CER1 genes that are typical DE markers. The DE step was characterized by a network including 155 targets among the 198 genes of the cluster. The hub nodes appeared to be EGF, GNB4, CXCR4, and KIT appeared as central nodes to connect the subnetworks of targets inside the cluster (Figure 3, gene list of this cluster network is given in supplemental file 3). The first subnetwork was centered around EGF and displayed interactions between various FGF compounds (such as FGF8, FGF13). It also involved CDH2. The second important subnetwork was centered around GNB4 and included CXCR4 and TGR5.

3.2.4 A common signature of iPSC and DE highlight the early steps of differentiation

The top GO_biological processes in the joint of iPSCs and DE clusters was related to DNA replication and stem cell maintenance (Table 1). This cluster included NANOG, POU5F1, PHC1, PARP1, DMNT3B, BUB1B and CDC6 genes (the gene list is provided in supplemental file 2, the associated gene network is given in figure S2 in supplemental file 1 and the gene involved in the cluster network in supplemental file 3).

3.2.5 Hepatic endoderm is characterized by important liver markers and differentiation such as HNF4A, FOS, and BMP2.

The top GO biological process of the HE cluster was the organic, oxoacid and carboxylic acid metabolic processes (Table 1). Then, the cluster discriminated the HE through important transcription factors involved in liver development such as HNF4A, PPARA. It also included glutathione metabolism genes such as GPX4, GPX, GSTK1, GSTA2 and drug transporters such as ABCC2 (MRP2), ABCA5, SULT1E1 genes. We also found in this cluster the FOS and BMP2 transcription factors, various genes coding for heat shock proteins and calcium-related genes such as CAMK2D (gene list is provided in supplemental file 2, the associated gene network is given in figure S3 in supplemental file 1 and the gene involved in the cluster network in supplemental file 3).

3.2.6 Hepatic like cells is characterized by specific extracellular markers and hepatic markers

The cluster identified for the HLCs contained targets related to cholesterol metabolisms such as HMGCR, CYP51A1, ABCA1 (cholesterol biosynthesis and transport) and lipid metabolism (FADS1, PPT2, SCD, ACADVL, ACSL1), TGFBR3 and EGFR (receptors) and ECM related compounds (such as collagen, integrin, and laminin). It also included metalloprotein, MMP2, various transcription factors such as HIF1A, JUN, MAF, MAFF, KLF6, BMP4, nuclear receptors (NR3C1, NRIP1), SLC transporters, drug transporters (UGT2B7), hepatic marker (cytokeratin 8, KRT8), urea cycle genes (ODC1, CPS1), Wnt signalling genes (such as CTNND2, SMAD3). The gene list is provided in supplemental file 2, the associated cluster gene network is given in figure 4 in the supplemental file 1 and the gene involved in the network in

supplemental file 3. HLCs cluster was distinguished by anatomical structure morphogenesis and extracellular matrix/structure organization as top GO biological processes (Table 1).

3.2.7 Hepatic endoderm and Hepatocytes-like cells are characterized by specific liver markers such as ALB and AFP

The last cluster for both HE and HLCs was characterized by ECM processing, cellular response to chemical (probably consistently with the HGF and OSM stimulations in those steps that both target the AKT/JAK/STAT signallings, Halfter et al., 2000; Organ et al., 2011) in the top GO_biological processes (Table 1). The gene list included genes encoding for typical liver proteins such as AFP, ALB, various collagen compounds (COL1A1, COL1A2, COL3A1, COL5A1, and COL6A3), and apoproteins (APOM, APOA1, APOA4, APOB), genes of fibrinogen produced in liver FGA, FGB (the gene list is provided in supplemental file 2, the associated gene network is given in figure S4 in supplemental file 1 and the gene involved in the network in supplemental file 3)

3.3 Integration of transcription factors analysis bridged upstream regulators with each specific differentiation steps and their specific biological processes

3.3.1 NFYC, RCOR1, PML, PATZ1, and MXI1 are important transcription factors that bridged iPSC and DE biological processes

The analysis of the upstream regulators in definitive endoderm versus the iPSC steps contributed to extract NFYC, RCOR1, PML, PATZ1, and MXI1 as

important transcription factors between the two steps. In definitive endoderm, we found that PML and PATZ1 were over-activated whereas NFYC, RCOR1, and MXI1 were over activated in hiPSCs (Table 2). The regulatory network proposed using ISMARA analysis is shown in figure 5. The targets of the PML and PATZ1 transcription factors included important endoderm markers such as FOXA2, SOX17, and GATA4. POU5F1 (OCT4), a typical hiPSCs marker was connected to RCOR1 in this network. Furthermore, the DNA processes (such as histones methylation *via* HDAC signaling), cell cycle, were among the top target pathways of the NFYC, RCOR1 and MXI1 transcription factors. Extracellular matrix (ECM) processes were targeted by PML, PATZ1 and RCOR1 genes. Consistently with the genes involved in the mitochondrial protein biogenesis (see section 3.2.2), we also found in our data set that HIF1 signaling was one target of PATZ1. Then, important cascade signaling such as BMP, Wnt, and Toll/TLR were targeted by PATZ1, PML, and MXI1 (Table 2). Finally, the MXI1 transcription factor targets the MAPK (*via* p38) signaling that is an important regulator of the FGF signaling (FGF gene network are extracted in section 3.2.3)

3.4.2 Kinetics of WRNIP1, NFIX, HNF1A, MEIS2 transcription factors bridged important liver functions during the hepatic specification process.

Then, the ISMARA analysis of the HE versus the iPSCs stage demonstrated the overactivity of HNF1A, WNRIP1, MEIS2, NFIX, MECP2, TAF1 in the HE whereas TAF1, SOX2, NFYC, and MIX1_MYC_MYCN were overactivated in hiPSCs (Table 3). The bridges with the definitive endoderm network were done using NFYC and

MIX1_MYC_MYCN that appeared common to both networks. The top targets of HNF1A included important liver transcription factors such as HNF4A and liver proteins such as AFP (alpha-fetoprotein), FGA and FGB (fibrinogen A and B). The proposed regulatory network is provided in figure 6.

3.4.3 Kinetics of WRNIP1, NFIX, HNF1A, MEIS2 transcription factors bridged important liver functions during the hepatic maturation process.

Then, the kinetic analysis with ISMARA revealed transcription factor binding motifs with high activities during the entire process of differentiation. Fourteen motifs showed an associated z-value above 3. The Pearson coefficient for the 14 highest motifs showed a decrease of activity for MYC, E2F7, MTA3, NRF1, ETV5, E2F5, CEBPZ, ZBTB33, TAF1, and SOX2 transcription factor bindings during the differentiation towards hepatocyte-like cells (Table 4). Conversely, the activity of WRNIP1, NFIX, HNF1A, MEIS2 motifs continuously increased over the differentiation (Table 4). The first level regulatory networks of WRNIP1, NFIX, MEIS2, and HNF1A_HNF1B is presented in figure 7. HNF4A appeared as a common node in the 4 regulatory networks (Figure 7). WRNIP1 regulates several motifs important in mature hepatocytes, such as LXRB (NR1H2). Furthermore, NFIX, MEIS2, and HNF1A bridged important MET/EMT genes such as ZEB1, ID4_TCF4_SNAI2 (nb. RTqPCR showed the upregulation of BMP4, SNAI1, and SNAI2 in HE and HLC differentiation steps whereas CDH2 levels were reduced when compared to DE levels in figure 1). The subsequent top targets of the NFIX, MEIS2, WRNIP1, HNF1A transcription factors in our datasets are given in table 5 (in addition to the top 10 upregulated transcription factors). The targets included important liver genes such as HNF4A, and liver proteins such as AFP, FGA (fibrinogen A), APOA4 (lipoprotein),

and cyclin-dependent kinases (CDKN2B, CDKN1C). The targeted pathways of the transcription factors included ECM regulations, MET-EMT related pathway, PPARA signaling, iron metabolism and important liver processes such as the HNF3A and HNF3B pathways (Table 5, supplementary file 2). The integrated network is given in figure 8.

Finally, we checked specifically the activity of the motifs of several well-known key transcription factors such as NANOG, POU5F1, FOXA2, SOX17, HNF4A and PXR that are important in iPSC differentiation into HLC and in mature liver cells. The activity of the motif related to NANOG, OCT4, SOX17 transcription factors decreased from iPSCs and DE to HLCs. On the contrary, FOXA2, HNF4A and PXR motif activity increased in HLC when compared to DE levels. We selectively confirmed those results with the PCR for OCT4, FOXA2, HNF4A, PXR. However, when compared to other transcription factors (such as WRNIP1, NFIX, HNF1A, MEIS2**)**, the ISMARA analysis did not highlight especially those factors as regulators (because there z_values were low). Nevertheless, they appeared as important targets of the selected transcription factors as shown in figure 9.

4. Discussion

We have compared the nanoCAGE profiles of four differentiation steps during the hiPSCs differentiation into hepatocytes like cells. The overall dataset analysis revealed a specific profile for each step. The mRNA levels of typical key transcription factors such as OCT4, HNF4A and key hepatic markers such as PXR, ALB, and CYP3A4 followed a time- and a step-dependant kinetics. Accordingly, OCT4 mRNA levels decreased after the pluripotency and DE steps whereas HNF4A mRNA levels

increased in the hepatic progenitor markers. In addition, the high level of mRNA expression of ALB, PXR, CYP3A4 mRNA was confirmed at the maturation step. Furthermore, the HLCs were characterized by typical hepatocyte phenotypes, positive CYP3A4, and HNF4a cells, and by albumin productions. Moreover, this typical signature appeared in agreement with other literature reports on liver iPSC and stem cells profiling (Ghosheh et al., 2017, Si Tayeb et al., 2010). The process of differentiation and those typical markers extracted from our analysis appeared also in agreement with typical liver organogenesis process and fetal liver development (Lee et al., 2012; Zhao et al., 2009).

Several groups reported transcriptome analysis of hiPSCs differentiation into hepatocytes like cells (Godoy et al., 2015; Matz et al., 2017). Matz and her colleagues reported activation of STAT1, MZF1, KFL4, and SP1 in iPSC step, of ELF5, FEV, INSM1, FOXI1 and STAT1 for DE, LHX3, MIZF, CTCF, NR3C1, and PAX6 for HE, PLAG1, EWSR1-FLI1, IRF2, MEF2A and ELF5 for HLC. In our investigation, the ISMARA processing illustrated that the iPSC to DE transition was mainly explained by the activity of NFYC, RCOR1, MXI1, PATZ1, and PML. Those transcription factors targeted pathways related to DNA methylation and to MAPK cascades. MAPK is reported to be required to improve the formation of endoderm by activin A *through* FGF activation (Sui et al., 2012). This result is consistent with our finding in which the DE gene cluster involved several FGF genes. Furthermore, their target also included important DE genes such as FOXA2, GATA4 and SOX17 observed in hiPSC liver differentiation (Si Tayeb et al., 2010).

The important upregulated transcription factors in our study, concerning the HLC, were HNF1A, NFIX, MEIS2, and WRNIP1. Their respective regulatory networks appeared connected through the HNF4A gene. HNF1A and HNF4A are well-known

hepatic liver transcription factors involved in the hepatic maturation and differentiation. Consistently with our dataset, NFIX was reported to be involved in the repression of AFP expression during liver development (Ren, 2012). Moreover, MEIS2 was also involved in liver development (Biemar et al., 2001; Sanchez Guardado et al., 2011). In our dataset, among the top 10 targets of MEIS2, NFIX, WRNIP1 and HNF1A transcription factors extracted by ISMARA, we detected important liver markers such as HNF4A and FGA (fibrinogen A). Furthermore, their top targets included genes of the ECM (such as COL6A3) and cyclin-dependent kinase inhibitors (important in cell cycle progression). Those targets are able to bridge the transcription factor (ISMARA), the genes (iDEP) and proteins (STRING) network analysis (Fig. 4), as we will discuss below.

At first, the integration of the profiles of our HLC revealed several collagens and ECM related targets (laminin, collagens, vitronectin, integrins, asialoglycoprotein-ASGR1). The positive control of the liver differentiation by ECM is consistent with reports on hepatic differentiation *in vitro* (Liu et al., 1991; Brill et al., 2002 ; Suzuki et al., 2003; Nagaoka et al., 2015; Peters et al., 2016). In that ECM network, our results showed that the DCN gene was involved in the late hepatic profile. Interestingly, the DCN protein was reported to be one potential candidate to accelerate liver regeneration after hepatectomy (Ma et al., 2014). This protein regulates the collagen fibrillogenesis during development and inhibits fibrosis (Baghy et al., 2011).

Then, MMP2 was also found as a central node in the protein interaction networks. Metalloproteins are largely involved in the liver repair, liver injury and fibrosis (Duarte et al., 2015). MMP2 is known to contribute to the extracellular matrix remolding during collagen I accumulation in the fibrotic liver (Theret et al., 1999). MMP2 is also an important EMT transcription factor (Gurzu et al., 2015). MMP

proteins and ECM remolding are reported to be regulated through OSM (OSM is used in our HLC step) via JAK/STAT, which appeared consistent with our results (Korzus et al., 1997).

Furthermore, STRING analysis connected those ECM networks with the BMP4 dependent network. The BMP4 growth factor is involved in the early stages of hepatic differentiation, at the hepatoblast and hepatic specification stages (Si Tayeb et al., 2010). BMP growth factors activate the TGFβ/BMP signaling. This pathway is involved in EMT/MET transitions. In our dataset, we found that the levels of mesenchymal genes such as CDH2 increased in DE and then decreased in the HE and HLC steps. SNA1 and SNAI2 were also over-expressed in HLC. Furthermore, the extracted NFIX, MEIS2 transcription factors were related to cyclin-dependent kinase inhibitors which are important key players during the EMT/MET process (Byers et al, 2013 and Gurzu et al., 2015) that are TGFβ signaling regulators (Nanda et al., 2016). As far as AFP and ALB, both hepatic epithelial markers, were produced in HE and HLC, our result suggests an early EMT followed by the MET during the hepatic maturation. This result was also consistent with recent stem cells finding for hepatocytes (Li et al., 2017).

Finally, important liver pathways such as PPARA signaling, iron metabolism, HNF3A, and HNF3B signaling were activated in our HLC (Table 5, Figure 8). In parallel, we found positive HNF4A and CYP3A4 immunostaining and higher HNF4A and CYP3A4 mRNA levels in the maturation steps when compared to undifferentiated situations. One signaling pathway involving CYP3A4 is regulated through HNF4a *via* PXR activation (Pondugula et al., 2009; Oladimeji et al., 2016). In our protocol, we found a weak PXR overexpression in HLC compared to earlier steps of differentiation. Thus, the kinetic analysis reveals that a refined investigation on the

regulation of the PXR pathway needs to be performed and represents a potential way to explore new strategies to improve hepatic stem cells *in vitro* differentiation. This could be achieved by using specific PXR inducers such as rifampicin or dexamethasone (Negoro et al., 2016; Kondo et al., 2014).

Conclusion

In this study, we have explored the dynamic of the induced pluripotent stem cells differentiation using nanoCAGE technology. The HLCs were characterized by albumin production, CYP3A4 positive immunostaining. The transcription factors analysis revealed a regulatory network involving the HNF1A, MEIS2, WRNIP1, and NFIX as important genes in the HLC step. The transcription factors were bridged through HNF4a. They also targeted important liver pathways such as the PPARA signaling and the HNF3A/HNF3B signaling. Various central nodes in the downstream networks, such as BMP4, DCN, MMP2, were highlighted during the analysis. In addition, the targets of the transcription factors included downstream ECM signaling and TGFbβ signaling. Consistently, we also found that the step-by-step differentiation was characterized by transcriptome profiles illustrating an EMT - MET process. Those nodes and proteins highlighted in this study could appear as potential targets for new strategies of *in vitro* liver tissue maturation.

Acknowledgments

The project was supported by the JSPS Kakenhi 16F16715, the JSPS-CNRS post-doctoral fellowship program of Myriam Lereau-Bernier, P16715, by the iLite ANR-16-RHUS-0005. Bertrand-David Segard was supported by the CNRS, Stephane Poulain by the JSPS Grant-in-aid for Scientific Research (S) 16H06328 and Charles Plessy by a grant to RIKEN CLST (DGT) from the MEXT, Japan.

Conflict of interest

We have no conflict of interest

Ethical issue

All experiments were performed following the ethical rules of CNRS, The University of Tokyo and Riken.

Authors statement

All authors have contributed to this work. Drs Lereau Bernier, Shinohara, Danoy, Segard and Tauran were involved in the biological experiments and cellular assays. Drs Poulain, Kato and Plessy developed the nanoCage technology and the data processing. Drs Tauran and Leclerc analyzed the data set. Profs Sakai, Kido and Miyajima developed and stabilized the iPSC protocol. Drs Tauran and Leclerc wrote the paper. Dr Leclerc designed the project.

References

Arnaud, O., Kato, S., Poulain, S., and Plessy, C. (2016) Targeted reduction of highly abundant transcripts using pseudo-random primers. *Biotechniques*, 60, 169–174.

Baghy, K., Dezso, K., László, V., Fullár, A., Péterfia, B., Paku, S., Nagy, P., Schaff, Z., Iozzo, R.V., and Kovalszky, I. (2011) Ablation of the decorin gene enhances experimental hepatic fibrosis and impairs hepatic healing in mice. *Lab Invest*, 91, 439-51.

Balwierz, P.J., Pachkov, M., Arnold, P., Gruber, A.J., Zavolan, M., and Van Nimwegen, E. (2014) ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res*, 24, 869–884.

Biemar, F., Devos, N., Martial, J.A., Driever, W., and Peers, B. (2001) Cloning and expression of the TALE superclass homeobox Meis2 gene during zebrafish embryonic development. *Mech Dev*, 109, 427–431.

Brill, S., Zvibel, I., Halpern, Z., and Oren, R. (2002) The role of fetal and adult hepatocyte extracellular matrix in the regulation of tissue-specific gene expression in fetal and adult hepatocytes. *Eur J Cell Biol.*, 81, 43-50.

Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., Nilsson, M.B., Gudikote, J., Tran, H., Cardnell, R.J., Bearss, D.J., Warner, S.L., Foulks, J.M., Kanner, S.B., Gandhi, V., Krett, N., Rosen, S.T., Kim, E.S., Herbst, R.S., Blumenschein, G.R., Lee, J.J., Lippman, S.M., Ang, K.K., Mills, G.B., Hong, W.K., Weinstein, J.N., Wistuba, I.I., Coombes, K.R., Minna, J.D., and Heymach, J.V. (2013) An epithelial- mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res*; 19: 279-290.

Cai, J., Zhao, Y., Liu, Y., Ye, F., Song, Z., Qin, H., Meng, S., Chen, Y., Zhou, R., Song, X., Guo, Y., Ding, M., and Deng, H. (2007) Directed differentiation of human embryonic stem cells into functional hepatic cells. Hepatology, 45 :1229-39.

Dhawan, A. (2015) Clinical human hepatocyte transplantation: Current status and challenges. *Liver transpl*, 21, S39–S44.

Duarte, S., Baber, J., Fujii, T. and Coito, A. (2015) Matrix metalloproteinases in liver injury, repair and fibrosis. *Matrix Biol*, 0, 147–156.

Haberle, V., Forrest, A., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015) CAGEr: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res*, 43, 51-60.

Halfter, H., Postert, C., Friedrich, M., Ringelstein, E.B., and Stögbauer, F.(2000) Activation of the Jak-Stat- and MAPK-pathways by oncostatin M is not sufficient to cause growth inhibition of human glioma cells. Brain Res Mol Brain Res, 80, 198-206.

Ghosheh, N., Küppers-Munthe, B., Asplund, A., Edsbagge, J., Ulfenborg, B., Andersson, T., Björquist, P., Andersson, C., Carén, H., Simonsson, S., Sartipy, P., and Synnergren, J. (2017) Comparative transcriptomics of hepatic differentiation of human pluripotent stem cells and adult human liver tissue. *Physiol Genomics*, 49, 430–446.

Godoy, P., Schmidt-Heck, W., Natarajan, K., Lucendo-Villarin, B., Szkolnicka, D., Asplund, A., Björquist, P., Widera, A., Stöber, R., Campos, G., Hammad, S., Sachinidis, A., Chaudhari, U., Damm, G., Weiss, T.S., Nüssler, A., Synnergren, J., Edlund, K., Küppers-Munther, B., Hay, D.C., and Hengstler, J.G. (2015) Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. *J Hepatol*, 63, 934-42.

Gurzu, S., Turdean, S., Kovecsi, A., Contac, A.O., and Jung, I. (2015) Epithelial-mesenchymal, mesenchymal-epithelial, and endothelial-mesenchymal transitions in malignant tumors: An update. *World J Clin Cases,* 3, 393-404.

Hay, D.C., Fletcher, J., Payne, C., Terrace, J.D., Gallagher, R.C., Snoeys, J., Black, J.R., Wojtacha, D., Samuel, K., Hannoun, Z., Pryde, A., Filippi, C., Currie, I.S., Forbes, S.J., Ross, J.A., Newsome, P.N., Iredale, J.P. (2008) Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proc Natl Acad Sci USA*, 105: 12301-6.

Hasegawa, A., Daub, C., Carninci, P., Hayashizaki, Y., and Lassmann, T. (2014) MOIRAI: a compact workflow system for CAGE analysis. *BMC Bioinformatics*, 15, 144-149.

Holmgren, G., Sjögren, A.K., Barragan, I., Sabirsh, A., Sartipy, P., Synnergren, J., Björquist, P., Ingelman-Sundberg, M., Andersson, T., and Edsbagge, J. (2014) Long-term chronic toxicity testing using Human Pluripotent Stem Cell–Derived Hepatocytes. *Drug Metab Dispos*, 42, 1401-1406.

Kondo, Y., Iwao, T., Nakamura, K., Sasaki, T., Takahashi, S., Kamada, N., Matsubara, T., Gonzalez, F.J., Akutsu, H., Miyagawa, Y., Okita, H., Kiyokawa, N., Toyoda, M., Umezawa, A., Nagata, K., Matsunaga, T., and Ohmori, S. (2014) An efficient method for differentiation of human induced pluripotent stem cells into hepatocyte-like cells retaining drug metabolizing activity. *Drug Metab Pharmacokinet,* 29, 237-43.

Kratz, A., Beguin, P., Kaneko, M., Chimura, T., Suzuki, A.M., Matsunaga, A., Kato, S., Bertin, N., Lassmann, T., Vigot, R., et al. (2014) Digital expression profiling of the compartmentalized translatome of Purkinje neurons. *Genome Res*, 24, 1396–1410.

Korzus, E., Nagase, H., Rydell, R., and Travis, L. (1997) The mitogen-activated protein kinase and JAK-STAT signaling pathways are required for an oncostatin M-responsive element-mediated activation of matrix metalloproteinase 1 gene expression. *J Biol Chem*, 272, 1188-96.

Lanzoni, G., Oikawa, T., Wang, Y., Cui, C.B., Carpino, G., Cardinale, V., Gerber, D., Gabriel, M., Bendala, J.D., Furth, M., Gaudio, E., Alvaro, D., Inverardi, L., and Reid, L. (2013) Clinical programs of stem cell therapies for liver and pancreas. *Stem Cells*, 31, 2047-60.

Lassmann, T. (2015) TagDust2: a generic method to extract reads from sequencing data. *BMC Bioinformatics*, 16, 24-28.

Lee, J.S., Ward, W.O., Knapp, G., Ren, H., Vallanat, B., Abbott, B., Ho, K., Karp, S.J., and Corton, J.C. (2012) Transcriptional ontogeny of the developing liver. *BMC Genomics*, 13: 33, doi:10.1186/1471-2164-13-33.

Lereau Bernier, M., Poulain, S., Tauran, Y., Danoy, M., Shinohara, M., Kimura, K., Segard, B.D., Kato, S., Kido, T., Miyajima, A., Sakai, Y., Plessy, C., and Leclerc, E. (2019) Profiling of derived-hepatocyte progenitors from induced pluripotent stem cells using nanoCAGE promoter analysis. *Biochem Eng J*, 142, 7–17.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Li, Q., Hutchins, A., Chen, Y., Li, S., Shan, Y., Liao, A.B., Zheng, D., Shi, X., Li, Y., Chan, W.Y., Pan, G., Wei, S., Shu, X., and Pei, D. (2017) Sequential EMT-MET

mechanism drives the differentiation of human embryonic stem cells towards hepatocytes. *Nat Commun*, 8: 15166.

Lin, C. and Khetani, S.R. (2016) Advances in engineered liver models for investigating drug-induced liver injury. *BioMed Res Int*, Article ID 1829148, 20 pages, DOI:10.1155/2016/1829148.

Liu, J.K., DiPersio, C.M., and Zaret, K. (1991) Extracellular signals that regulate liver transcription factors during hepatic differentiation in vitro. *Mol Cell Biol*, 11, 773-784.

Ma, R., Chen, J., Li, Z., Tang, J., Wang, Y., Cai, X. (2014) Decorin accelerates the liver regeneration after partial hepatectomy in fibrotic mice. *Chin Med J*, 127, 2679-85.

Matz, P., Wruck, W., Fauler, B., Herebian, D., Mielke, T., and Adjaye, J. (2017) Footprint-free human fetal foreskin derived iPSCs: A tool for modeling hepatogenesis associated gene regulatory networks. *Sci Rep*, 7: 6294, DOI:10.1038/s41598-017-06546-9.

Nagaoka, M., Kobayashi, M., Kawai, C., Mallanna, S., Duncan, S. (2015) Design of a vitronectin-based recombinant protein as a defined substrate for differentiation of Human Pluripotent Stem Cells into Hepatocyte-Like Cells. *PLoS One*, 10, e0136350.

Nanda, V., Downing, K.P., Ye, J., Xiao, S., Kojima, Y., Spin, J.M., DiRenzo, D., Nead, K.T., Connolly, A.J., Dandona, S., Perisic, L., Hedin, U., Maegdefessel, L., Dalman, J., Guo, L., Zhao, X., Kolodgie, F.D., Virmani, R., Davis, H.R., Jr, and Leeper, N.J. (2016) CDKN2B regulates TGFβ signaling and smooth muscle cell investment of hypoxic neovessels. *Circ Res*, 118, 230-40.

Negoro, R., Takayama, K., Nagamoto, Y., Sakurai, F., Tachibana, M., and Mizuguchi, H. (2016) Modeling of drug-mediated CYP3A4 induction by using human iPS cell-derived enterocyte-like cells. *Biochem Biophys Res Commun*, 472, 631-6.

Oladimeji, P., Cui, H., Zhang, C., and Chen, T. (2016) Regulation of PXR and CAR by protein-protein interaction and signaling crosstalk. *Expert Opin Drug Metab Toxicol*, 12, 997–1010.

Organ, S.L. and Tsao, M.S. (2011) An overview of the c-MET signaling pathway. *Ther Adv Med Oncol*, 3(1 Suppl), S7–S19.

Peters, D.T., Henderson, C.A., Warren, C.R., Friesen, M., Xia, F., Becker, C.E., Musunuru, K., and Cowan, C.A. (2016) Asialoglycoprotein receptor 1 is a specific cell-surface marker for isolating hepatocytes derived from human pluripotent stem cells. *Development*, 143, 1475-81.

Poulain, S., Kato, S., Arnaud, O., Morlighem, J.É.., Suzuki, M., Plessy, C. and Harbers, M. (2017) NanoCAGE: A method for the analysis of coding and noncoding 5′-capped transcriptomes. In Sara Napoli (ed), Promoter Associated RNA - Methods and Protocols. Humana Press, New York, 1543, 57–109.

Plessy, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., Vitezic, M., Severin, J., Olivarius, S., Lazarevic, D., et al. (2010) Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nat Methods*, 7, 528–534.

Pondugula, S., Dong, H., and Chen, T. (2009) Phosphorylation and protein-protein interactions in PXR-mediated CYP3A repression. *Expert Opin Drug Metab Toxicol*, 5, 861–873.

Rashid, S.T., Corbineau, S., Hannan, N., Marciniak, S.J., Miranda, E., Alexander, G., Huang-Doran, I., Griffin, J., Ahrlund-Richter, L., Skepper, J., Semple, R., Weber, A., Lomas, D.A., and Vallier, L. (2010) Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J Clin Invest*, 120, 3127-3136.

Ren, H. (2012) Regulation of hepatic gene expression during liver development and disease, Theses and Dissertations-Microbiology, Immunology, and Molecular Genetics. 6.

Sánchez-Guardado, L., Irimia, M., Sánchez-Arrones, L., Burguera, D., Rodríguez-Gallardo, L., Garcia-Fernández, J., Puelles, L., Ferran, J., and Hidalgo-Sánchez, M. (2011) Distinct and redundant expression and transcriptional diversity of MEIS gene paralogs during chicken development. *Dev Dynam*, 40, 1475–1492.

Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010) Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology*, 51, 297-305.

Spence, J.R., and Wells, J.M. (2007) Translational embryology: using embryonic principles to generate pancreatic endocrine cells from embryonic stem cells. *Dev Dynam*, 236, 3218-3227.

Sui, L., Mfopou, J.K., Geens, M., Sermon, K., and Bouwens, L. (2012) FGF signaling via MAPK is required early and improves Activin A-induced definitive endoderm formation from human embryonic stem cells. *Biochem Biophys Res Commun*, 426, 380-5.

Suzuki, A., Iwama, A., Miyashita, H., Nakauchi, H., and Taniguchi, H. (2003) Role for growth factors and extracellular matrix in controlling differentiation of prospectively isolated hepatic stem cells. *Development*,130, 2513-24.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131:861-872.

Takebe, T., Sekine, K., Kimura, M., Yoshizawa, E., Ayano, S., Koido, M.,

Funayama, S., Nakanishi, N., Hisai, T., Kobayashi, T., Kasai, T., Kitada, R., Mori, A., Ayabe, H., Ejiri, Y., Amimoto, N., Yamazaki, Y., Ogawa, S., Ishikawa, M., Kiyota, Y., Sato, Y., Nozawa, K., Okamoto, S., Ueno, Y., and Taniguchi, H. (2017) Massive and reproducible production of liver buds entirely from Human Pluripotent Stem Cells. *Cell Rep*, 21, 2661 – 2670.

Theret, N., Lethi, K., Musso, O., and Clement, B. (1999) MMP2 activation by collagen I and concanavalin A in cultured human hepatic stellate cells. *Hepatology*, 30, 462-470.

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., and Vallier, L. (2010), Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology*, 51, 1754-1765.

Tsuchiya, A., Kojima, Y., Ikarashi, S., Seino, S., Watanabe, Y., Kawata, Y., and Terai, S. (2017) Clinical trials using mesenchymal stem cells in liver diseases and inflammatory bowel diseases. *Inflamm Regen*, 37, 16-22.

Zhao, D., Chen, S., Cai, J., Guo, Y., Song, Z., Che, J., Liu, C., Wu, C., Ding, M., and Deng, H. (2009) Derivation and characterization of hepatic progenitor cells from human embryonic stem cells. *PLoS One*, 4: e6468, doi:10.1371/journal.pone.0006468.

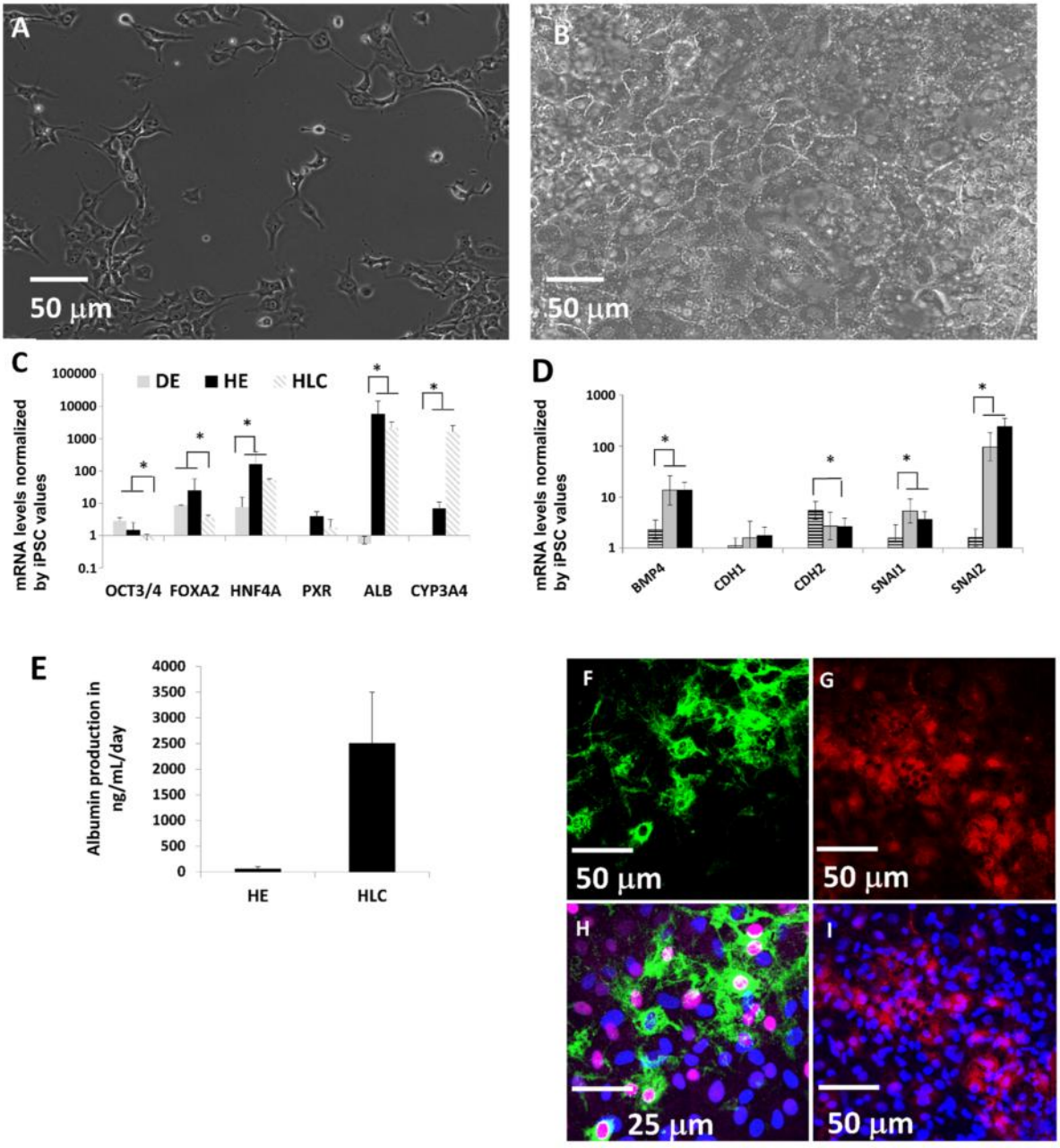

Figure 1: Morphology of iPSCs (A) and HLCs (B). Specific mRNA kinetic of selected hepatic differentiation markers (C) and of EMT/MET transition markers (D) albumin secretion (E) immunostaining of HLCs with nucleus in blue, HNF4 in purple (H), albumin in green (F and H) and CYP3A4 in red (G)

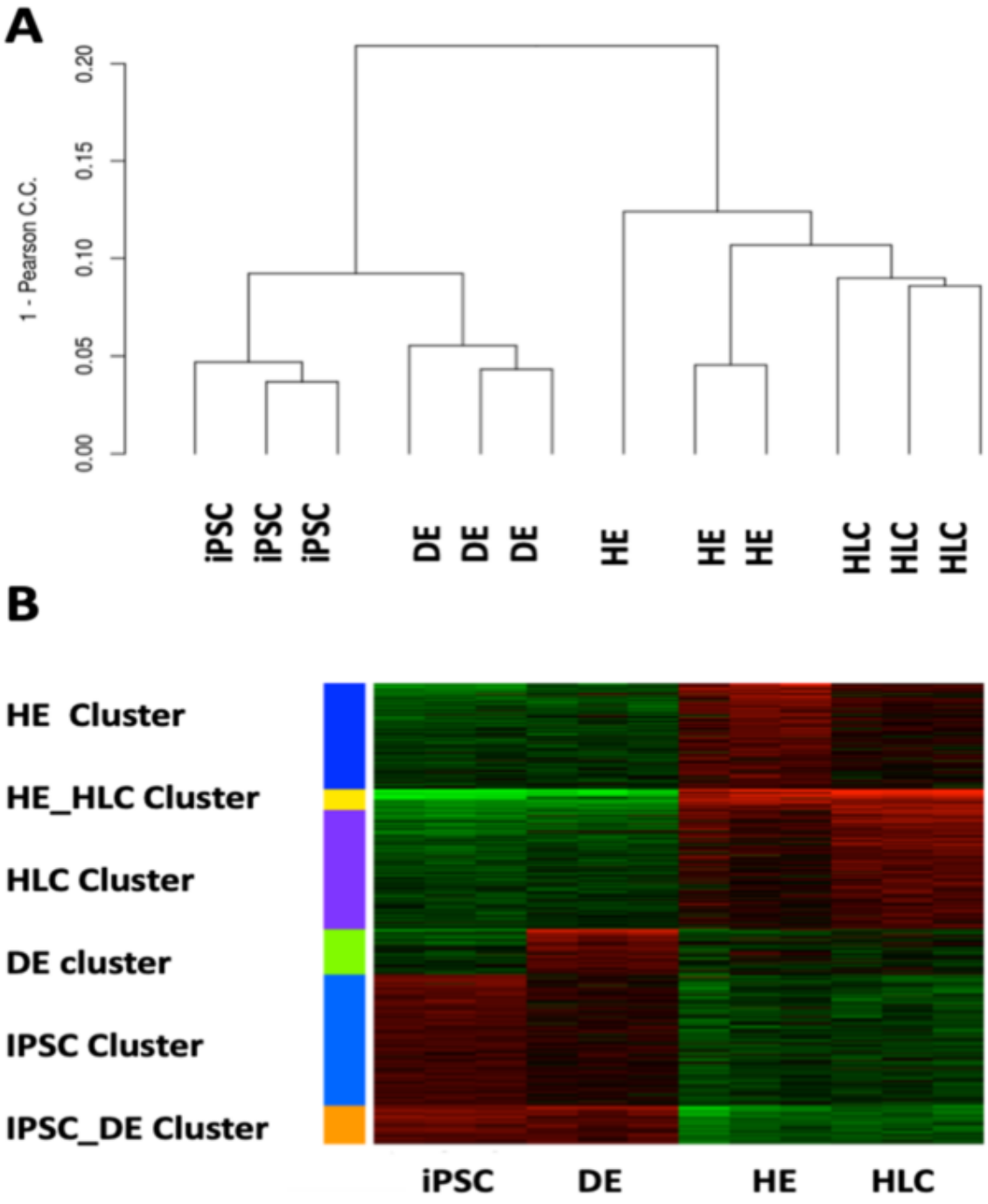

Figure 2: Dendrogram separating the steps of the differentiation process obtained from hierarchical clustering of nanoCAGE sequencing data with iDEP (A) Heatmap of the Kmeans clusters identified upon processing of nanoCAGE sequencing data with iDEP (B)

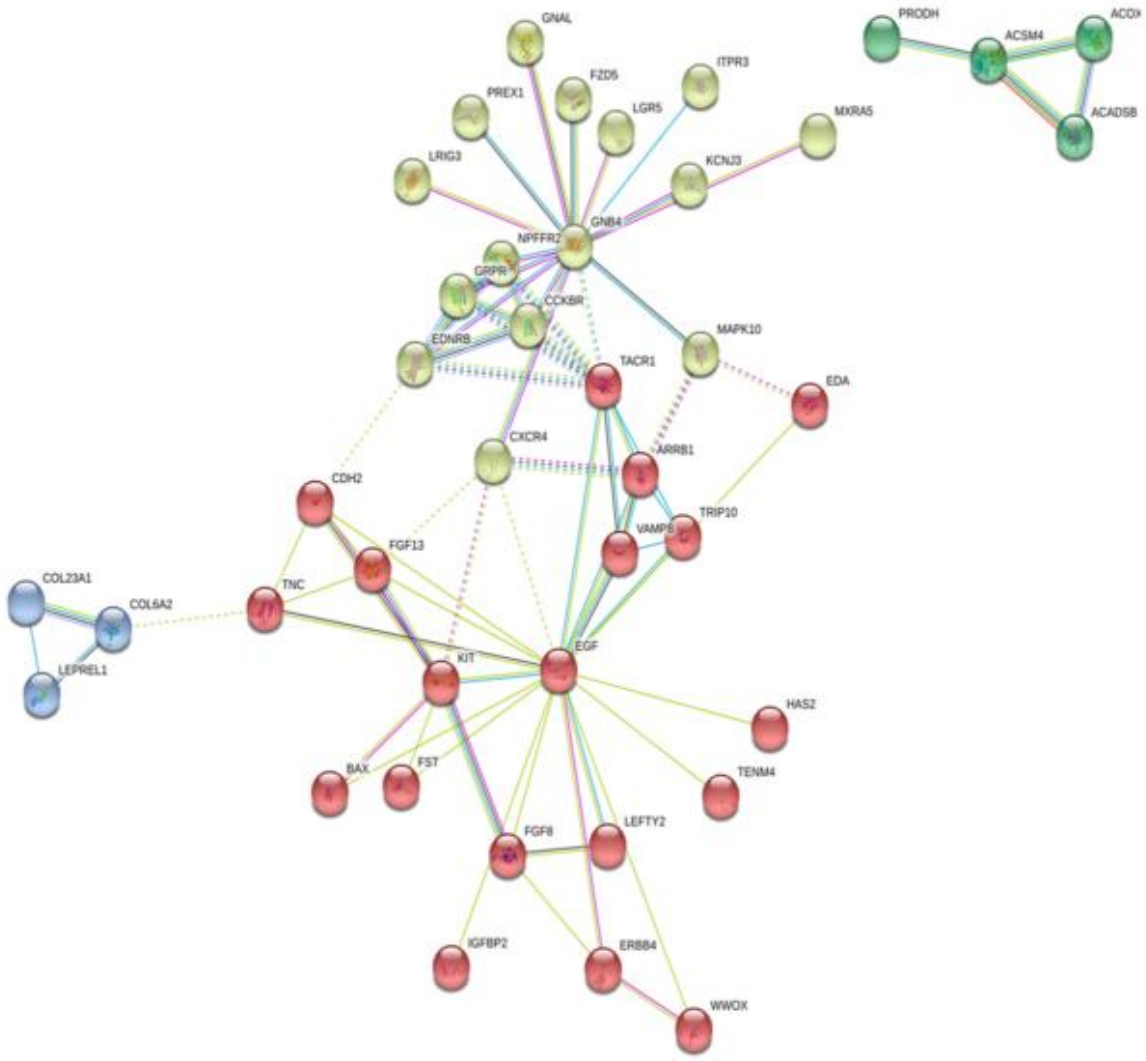

Figure 3: Network of the genes representing the DE cluster

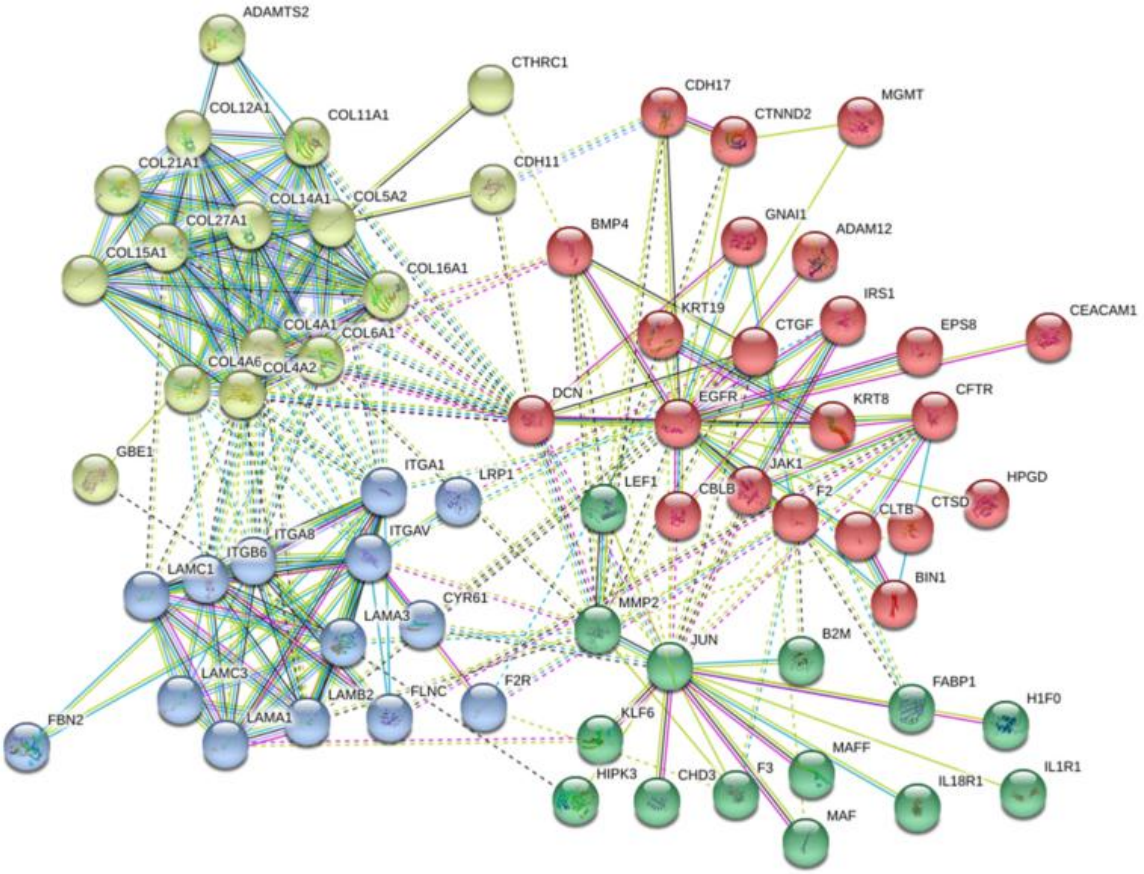

Figure 4: Network of the genes representing the HLC cluster. Gene list involved in the network is given in supplementary file 3

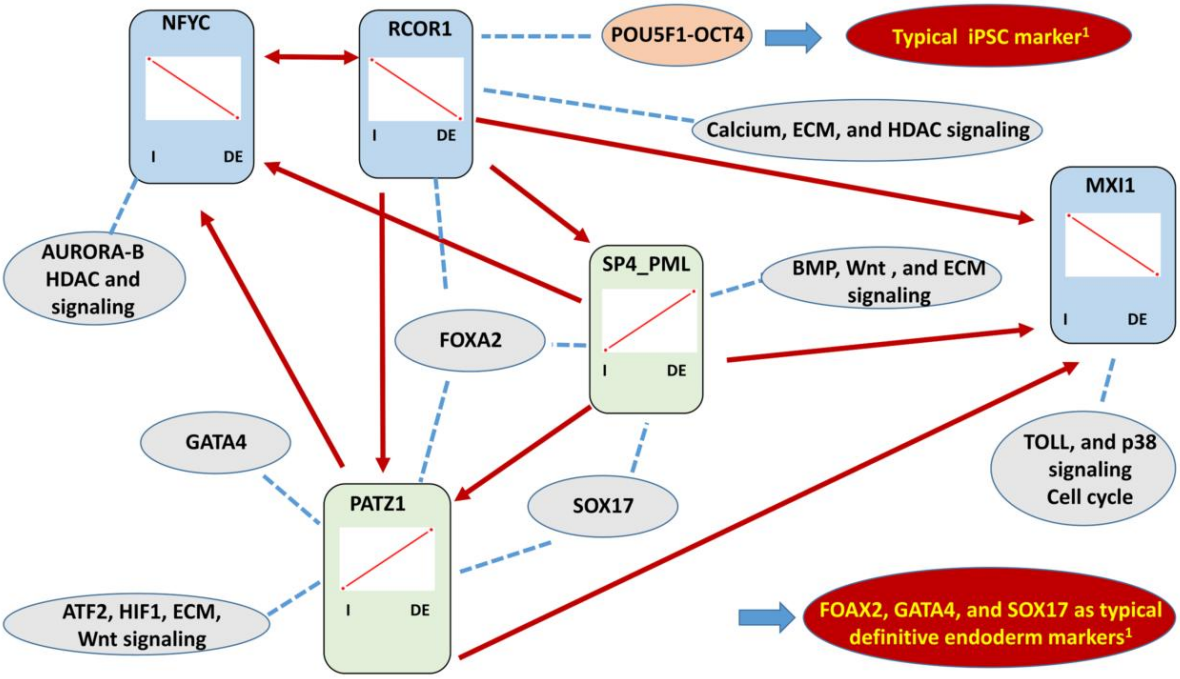

Figure 5: Interaction network between the DE transcription factors and their targeted pathways. Arrows indicate gene regulation extracted from ISMARA regulatory network, dash lines denote targeted pathways and genes extracted from ISMARA and iDEP clustering

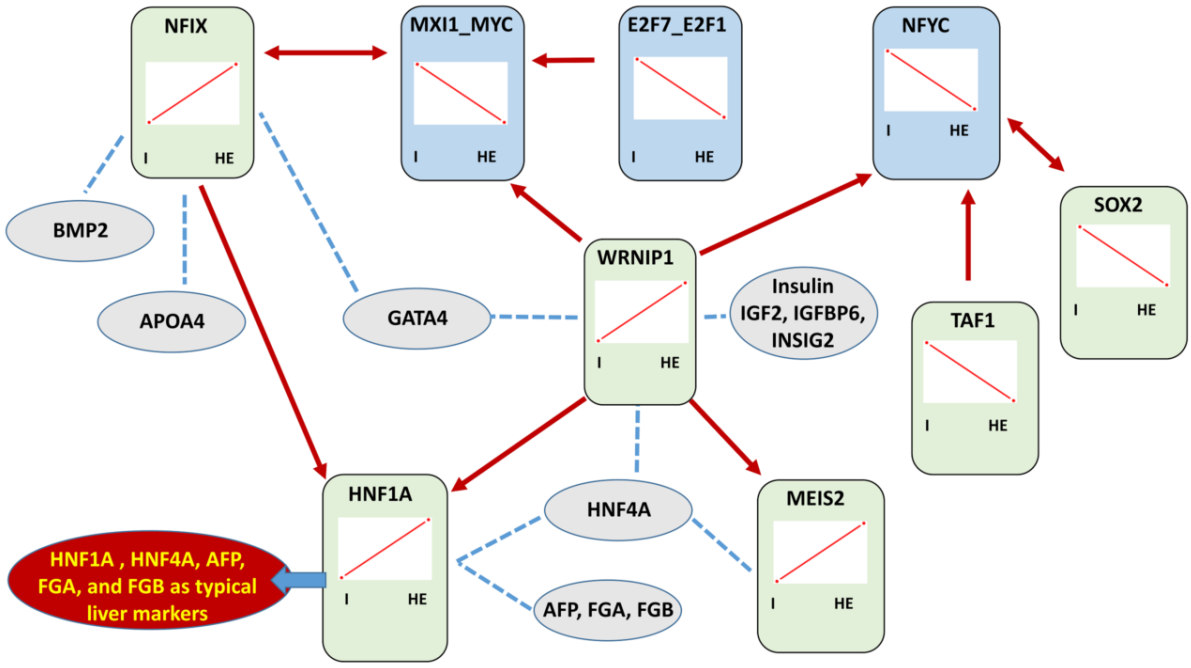

Figure 6: Interaction network between the HE transcription factors and their targeted pathways. Arrows indicate gene regulation extracted from ISMARA regulatory network, dash lines denote targeted pathways and genes extracted from ISMARA and iDEP clustering

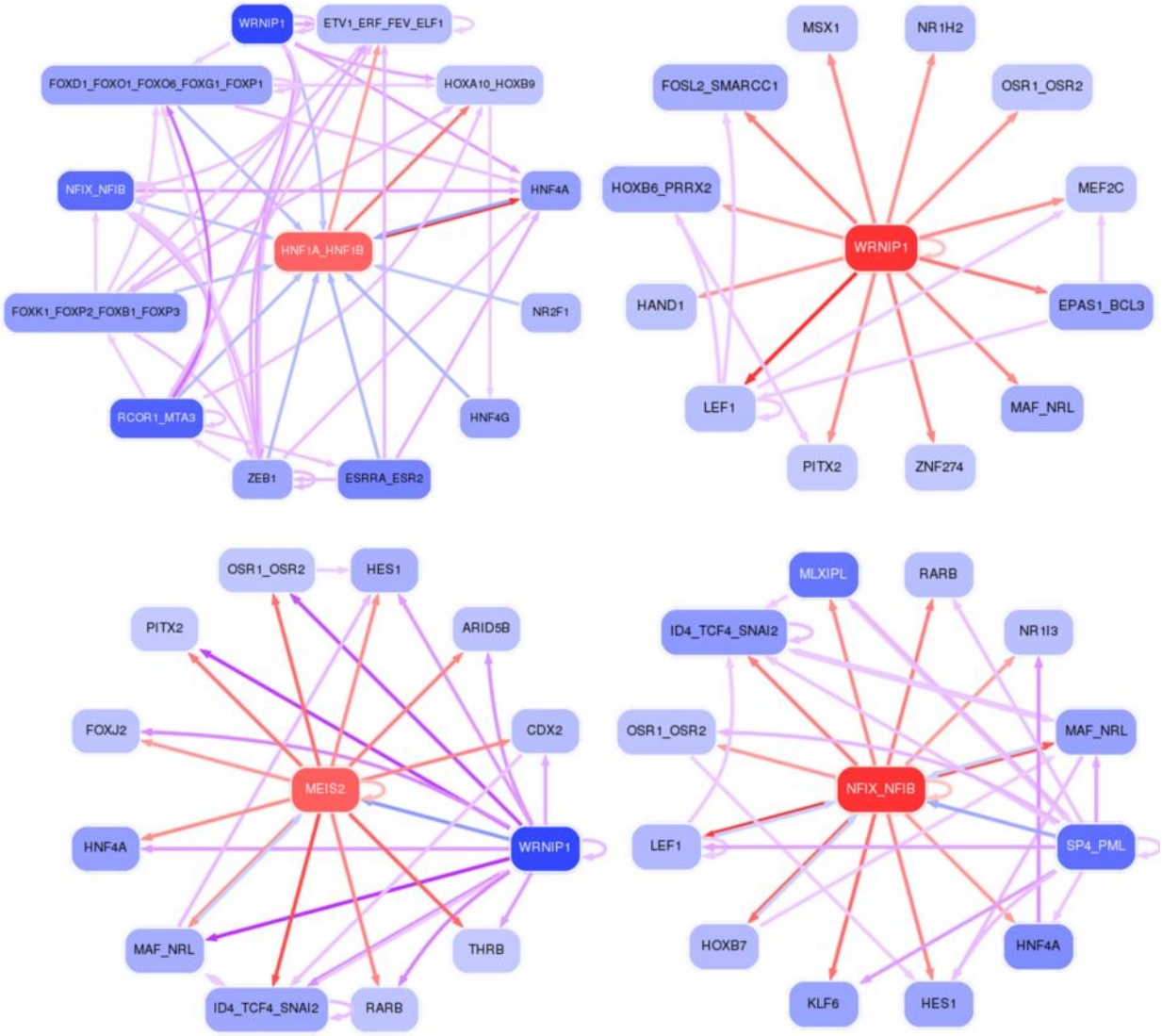

Figure 7: Regulatory network of the top 4 transcription factors, HNF1A, WRNIP1, MEIS2 and NFIX extracted from ISMARA analysis

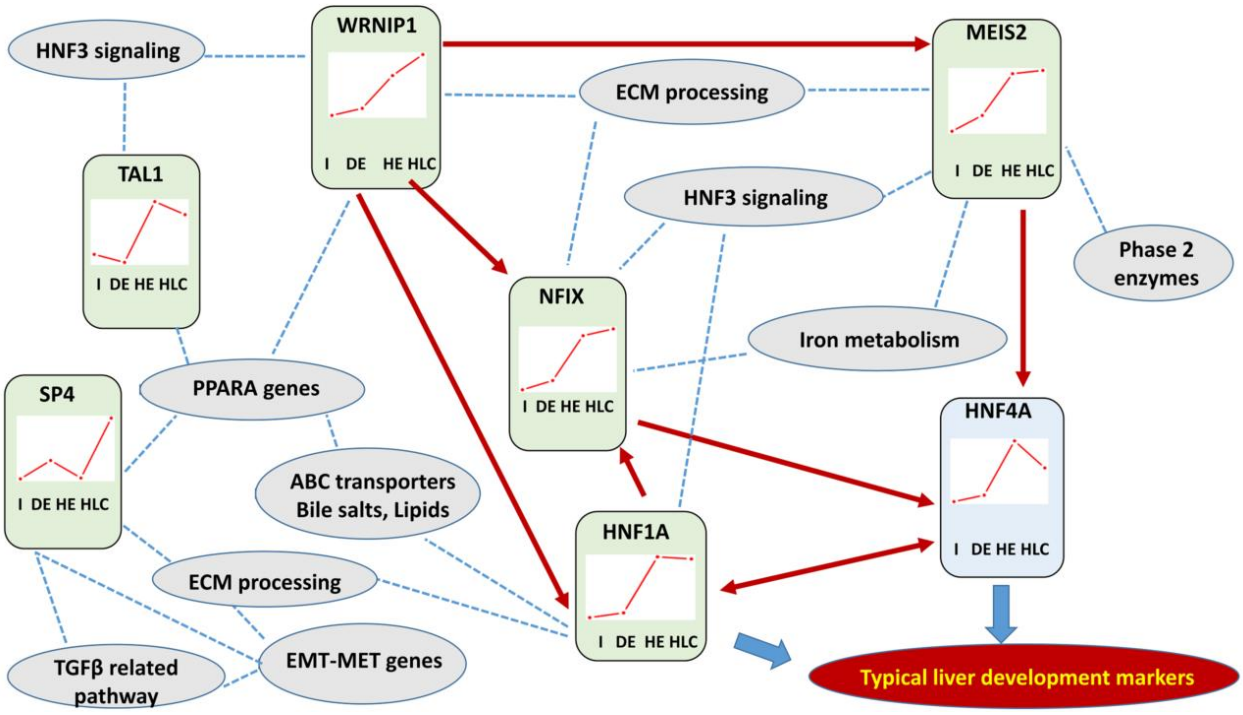

Figure 8: Interaction network displaying the kinetic between iPSC (noted as I) DE, HE and HLC transcription factors and their targeted pathways. Arrows indicate gene regulation extracted from STRING network analysis (detailed in Supplementary File 3) and ISMARA network analysis (Table 3), dash lines denote targeted pathways and extracted from ISMARA and iDEP clustering. HNF4a (z_value = 1.84) was added as a liver typical marker to bridge the table 3 transcription factor list

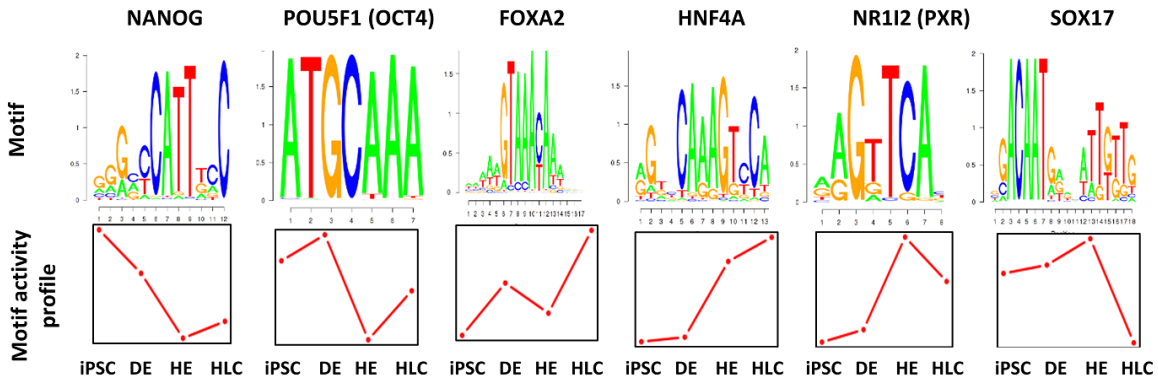

Figure 9: Motifs and motif activity kinetic of selected typical transcription factors during the differentiation process