An extended transcription factor regulatory network involving stabilization loops controls hepatocyte identity
Julie Dubois-Chevalier, Céline Gheeraert, Alexandre Berthier, Clémence Boulet, Vanessa Dubois, Loïc Guille, Marie Fourcot, Guillemette Marot, Karine Gauthier, Laurent Dubuquoy, et al.

To cite this version:
Julie Dubois-Chevalier, Céline Gheeraert, Alexandre Berthier, Clémence Boulet, Vanessa Dubois, et al.. An extended transcription factor regulatory network involving stabilization loops controls hepatocyte identity. 2023. hal-04137074

HAL Id: hal-04137074
https://hal.science/hal-04137074
Preprint submitted on 22 Jun 2023
An extended transcription factor regulatory network involving stabilization loops controls hepatocyte identity

Running title: Hepatocyte identity stabilization loops

Julie Dubois-Chevalier 1, Céline Gheeraert 1, Alexandre Berthier 1, Clémence Boulet 1, Vanessa Dubois 1,2, Loïc Guille 1, Marie Fourcot 3, Guillemette Marot 4, Karine Gauthier 5, Laurent Dubuquoy 6, Bart Staels 1, Philippe Lefebvre 1,* Jérôme Eeckhoute 1,*,#

1 Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1011-EGID, F-59000 Lille, France
2 Basic and Translational Endocrinology (BaTE), Department of Basic and Applied Medical Sciences, Ghent University, B-9000, Ghent, Belgium
3 Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41 - UAR 2014 - PLBS, F-59000 Lille, France
4 Univ. Lille, Inria, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France
5 Institut de Génomique Fonctionnelle de Lyon (IGFL), CNRS UMR 5242, INRAE USC 1370, École Normale Supérieure de Lyon, Lyon, France
6 Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000, Lille, France

* Contributed equally
#

Corresponding author:

1
Dr. Jérôme Eeckhoute
INSERM UMR 1011-Bâtiment J&K
Université de Lille, Faculté de Médecine de Lille-Pôle Recherche
Boulevard du Professeur Leclerc
59045 Lille cedex
France
Tel : 33(0)3 20 97 42 20 / Fax : 33(0)3 20 97 42 01
E-mail: jerome.eeckhoute@inserm.fr
Cell identity is specified by a core transcriptional regulatory circuitry (CoRC), typically limited to a small set of interconnected cell-specific transcription factors (TFs). By mining global hepatic TF regulons, we provide a more complex organization of the transcriptional regulatory network controlling identity. We have identified that tight functional interconnections controlling hepatocyte identity extend to non-cell-specific TFs beyond the CoRC, which we call hepatocyte identity (Hep-ID)\textsuperscript{CONNECT} TFs. Besides controlling identity effector genes, Hep-ID\textsuperscript{CONNECT} TFs also engage in reciprocal transcriptional regulation with TFs of the CoRC. In homeostatic basal condition, this translates into Hep-ID\textsuperscript{CONNECT} TFs being involved in fine tuning CoRC TF expression including their rhythmic expression patterns. Importantly, a role for Hep-ID\textsuperscript{CONNECT} TFs in the control of hepatocyte identity was revealed in dedifferentiated hepatocytes where Hep-ID\textsuperscript{CONNECT} TFs were able to reset CoRC TF expression. This was observed upon activation of NR1H3 or THRBI in hepatocarcinoma or hepatocytes subjected to inflammation-induced loss of identity. Hence, our study establishes that hepatocyte identity is controlled by an extended array of TFs beyond the CoRC uncovering identity stabilization loops.

**Keywords:** transcription factors / cell identity / core regulatory network / hepatocyte dedifferentiation / liver disease.
INTRODUCTION

Multicellular organisms are built upon collaborative activities of phenotypically and functionally distinct cell-types. Individual cell identities and functions are acquired thanks to the activity of cell-specific transcription factors (TFs). Indeed, cell-specific TFs directly control the expression of non-TF genes exerting activities/functions characterizing individual cell-types, alternatively known as identity effector genes (Almeida et al., 2021; Arendt et al., 2016). A limited set of identity TFs engaging into auto- and cross-regulatory loops typically defines the core transcriptional regulatory circuitry (CoRC) (Almeida et al., 2021; Arendt et al., 2016). CoRCs are considered cornerstones for establishment and maintenance of cell identities (Almeida et al., 2021; Arendt et al., 2016). Indeed, CoRCs allows to self-sustain high expression of identity TFs and their target effector genes in addition to being involved in the continuous modulation of the hepatic transcriptome in response to environmental stimuli (Almeida et al., 2021; Arendt et al., 2016; Boyer et al., 2005; Wilkinson et al., 2017).

In line, hepatocyte identity is typically defined as relying on a handful of interconnected hepatocyte-specific CoRC TFs (hereafter called Hep-ID TFs) including HNF4A, FOXA or NR1H4 (FXR) (Tachmatzidi et al., 2021). However, many additional TFs beyond these hepatic identity CoRC TFs have been ascribed with roles in the control of hepatocyte activities (Bideyan et al., 2021). However, whether and how these additional TFs exert their functions in concert with the hepatic CoRC is not fully understood. Moreover, considering that sustained expression of Hep-ID TFs has a positive impact on hepatic pathological conditions (Berasain et al., 2022), better defining how CoRC TFs are functionally linked to additional TFs is of pathophysiological interest. Indeed, breakdown in liver functions in advanced stages of liver injuries and in cancer results from hepatocyte loss of identity stemming from compromised expression of the hepatic CoRC TFs (Berasain et al., 2022; Dubois et al., 2020a).
As outlined hereabove, apprehending the transcriptional regulation of cell identity through the lens of CoRCs, by definition, only leads to consider the role of a very limited subset of TFs. In this context, the real architecture and outline of the hepatocyte identity TF network has not been clearly established beyond Hep-ID TFs. We postulated that focusing on TF interconnections at the genomic-scale would allow to define how the Hep-ID TF network spreads beyond the CoRC and would refine our understanding of hepatocyte identity transcriptional control.

RESULTS

Hepatocyte identity TFs extensively co-bind TF-encoding gene promoters beyond the CoRC

We defined the hepatocyte CoRC as a set of 13 Hep-ID TFs commonly defined in different studies as interconnected hepatocyte-specific TFs (D’Alessio et al., 2015; Dubois et al., 2020a; Kyrmizi et al., 2006; Zhou et al., 2017) (Table S1). Hep-ID TFs comprise the well-accepted and thoroughly experimentally-verified drivers of hepatocyte identity (Reizel et al., 2020; Tachmatzidi et al., 2021). As expected from our previous studies (Dubois-Chevalier et al., 2020; Dubois-Chevalier et al., 2017b; Dubois et al., 2020a), monitoring the mouse liver cistromes of 8 Hep-ID TFs (CEBPA, FOXA2, HNF4A, NR1H4, NR5A2, ONECUT1, PPARA, PROX1; Table S2) pointed to their extensive co-recruitment at the Hep-ID TF-encoding gene promoters when compared to a control set of non-Hep-ID TF-encoding gene promoters (Fig.1A and Fig.S1A-B). This led us to define a strategy called Promoter-centric TF network analysis (ProTFnet) (Fig.1B) with the aim to establish global TF interconnections through promoter binding patterns. Indeed, this approach consists in mining TF cistromes to identify TF binding to all TF-encoding gene promoters. Here, hepatic TF cistromes (n=49; including those of Hep-ID TFs and extending to transcriptional cofactors) (Dubois-Chevalier
et al., 2017b; Dubois et al., 2020a) were used to monitor the binding to all TF-encoding gene promoters active in the mouse liver (n=925; see Materials and Methods for details). These promoters were grouped together based on their TF binding pattern similarity, i.e. cistromic-based classification, using a self-organizing map (SOM) (Fig.S2) (Dubois-Chevalier et al, 2017a). Hierarchical clustering was next performed and identified 7 main clusters of TF-encoding gene promoters (Fig.1C and Fig.S3A). We annotated these clusters A to G based on progressive TF co-recruitment pattern complexity (Fig.1C-D, Fig.S3B-C and Fig.S4). Promoters from cluster G were also overall the most strongly active as revealed by chromatin accessibility (DHS), hepatic histone acetylation (H3K27ac) or mRNA expression levels of associated TF-encoding genes (Fig.1E and Fig.S5). Moreover, Hep-ID TF binding was also most pronounced at promoters from cluster G, both when considered individually (Fig.S3D-E and Fig.S4G) or collectively, i.e. Hep-ID TF co-recruitment (Fig.1F). Therefore, these data indicated that cluster G captured most of the TF-encoding gene promoters preferentially targeted by Hep-ID TFs.

In order to further characterize TF-encoding genes from cluster G, we prepared an additional classification of TF-encoding gene promoters based on their levels of hepatic activity and specificity [hereafter referred to as promoter hepatic activity score; see Materials and Methods for a full description of how this score was obtained]. This second classification, which was used to discriminate 7 promoter classes (denoted 1 to 7) displaying a progressive increase in the promoter hepatic activity score (Fig.S6; Table S1), was compared with the previously obtained cistromic-based classification. This analysis confirmed that cluster G overall comprised promoters with the highest hepatic activity scores (Fig.1G) including, in particular, a group of promoters shared with class 7 (G-7 promoters, n=13; Fig.S7A-C). G-7 promoters both displayed enrichment for Hep-ID TF-encoding genes (Fig.1H) and were strongly bound by these Hep-ID TFs (Fig.1I and Fig.S7D-F). Hence, G-7 is the prominent
subset of promoters capturing hepatic CoRC auto-/cross-binding. However, in addition to Hep-ID TFs, cluster G also comprised a majority of TF promoters with lower hepatic activity scores (G-4 to G-6 promoters, n=117) (Fig.1G and Fig.S7A-D; Table S1). To further characterize the TF-encoding gene heterogeneity in cluster G, we performed a meta-analysis of transcriptomic data from primary human and mouse cells (n=126 and 39 different cell-types including hepatocytes, respectively) to define 3 groups of TFs displaying cell-type specific (CTS), cell-type enriched (CTE) or ubiquitous (UBQ) expression patterns (Fig.1J and Fig.S8; see Materials and Methods). This further highlighted that, besides Hep-ID TFs and an additional set of cell-specific TFs with well characterized hepatic functions (Nr1i3/Car, Rorc/Rorg, Hnf1a and Gata4) (Tachmatzidi et al., 2021), the majority of TFs from cluster G belonged to CTE or UBQ TFs (Fig.1K and Table S1).

Altogether, these analyses indicated that extensive co-recruitment of Hep-ID TFs is not limited to their own highly hepatocyte-specific genes within the CoRC but extends to an additional set of TFs.

**Hep-ID**\(^{\text{CONNECT}}\) **TFs are expressed under the control of and collaborate with Hep-ID TFs in the regulation of hepatocyte identity effector genes**

In order to further characterize the CTE/UBQ TFs from cluster G, we refined our analysis of their individual expression patterns by defining how their levels of expression in mouse primary hepatocytes (MPH) compare to those in other primary cells (n=38 different cell-types). Interestingly, we found that a subset of the CTE/UBQ TFs from cluster G, hereafter called Hep-ID**\(^{\text{CONNECT}}\) **TFs, was nevertheless characterized by privileged hepatocyte expression (i.e. MPH ranked among the top 10 expressing cells and expression in MPH above average expression in other cells) (Fig.2A and Fig.S9). These data suggested that Hep-ID**\(^{\text{CONNECT}}\) **TFs might be directly dependent upon Hep-ID TFs for their enhanced hepatocyte expression. In line, expression of Hep-ID**\(^{\text{CONNECT}}\) **TFs was decreased in transcriptomic data
obtained from MPH of adult mice with hepatocyte-specific deletion of the Hep-ID TF Hnf4a (Fig.2B). Moreover, reminiscent of Hep-ID TF-encoding genes, expression of Hep-IDCONNECT TFs was increased in the final hepatocyte differentiation stage occurring during liver postnatal maturation (Fig.2C). We and others have established that compromised hepatocyte identity due to decreased Hep-ID TF gene expression is commonly found in severe liver injuries in both mouse models and humans (Argemi et al., 2019; Berasain et al., 2022; Bou Saleh et al., 2021; Dubois et al., 2020a; Gunewardena et al., 2022; Hyun et al., 2020; Loft et al., 2021). Therefore, we further mined Hep-IDCONNECT TF expression in a meta-analysis of transcriptomic data obtained from mouse liver injury models triggering hepatocyte dedifferentiation (Fig.S10). Interestingly, we found that Hep-IDCONNECT TFs showed reduced expression levels in the livers of these mouse models (Fig.2D). Similar observations were made using transcriptomic data obtained from microdissected hepatocytes from alcohol-related human liver cirrhosis (Bou Saleh et al., 2021) (Fig.2E). Overall, observed changes in expression of Hep-IDCONNECT TFs were reminiscent of those of Hep-ID TFs, albeit often with a lower amplitude, and different from those of other TFs from cluster G. A similar observation was made when interrogating the breadth of H3K4me3 labelling at Hep-IDCONNECT TF gene promoters, a feature positively linked to the role of TFs in the control of cell identity (Benayoun et al., 2014; Chen et al., 2015; Pekowska et al., 2010). Indeed, the breadth of H3K4me3 labelling at the promoters of Hep-IDCONNECT TF genes was also comprised in between that observed for the Hep-ID TF and other TF-encoding genes (Fig.2F). Moreover, H3K4me3 labelling at Hep-IDCONNECT TF genes displayed intermediate tissue-specificity when compared to Hep-ID TF genes, on the one hand, and remaining TF-encoding genes from cluster G, on the other hand (Fig S11). Altogether, these data point to the potential importance of Hep-IDCONNECT TFs in fully differentiated hepatocytes.
Indeed, many Hep-ID\textsuperscript{CONNEC}T TFs have individually, i.e. without an identified unifying rationale, been ascribed, at least to some extent, a role in the control of hepatic metabolic activities (Fig.S12). In line, mining genes associated with phenotypes from the Mammalian Phenotype Ontology resource (Chen \textit{et al}, 2007; Smith \textit{et al}, 2005) revealed that, albeit less significant when compared to Hep-ID TFs, Hep-ID\textsuperscript{CONNEC}T TF-encoding genes are linked to abnormal liver phenotypes and metabolic functions (Fig.2G). Mining transcriptomic data obtained from the livers of mice with hepatocyte-specific deletion of individual Hep-ID\textsuperscript{CONNEC}T TF-encoding genes (hep/- mice) indicated that this is functionally underlain by a preferential dysregulation of downstream hepatocyte identity effector genes, i.e. non-TF genes with liver-specific broad H3K4me3 domains defined in (Dubois \textit{et al}., 2020a) (Fig.2H). While such a bias was also observed in transcriptomic data obtained from the livers of mice deficient for individual Hep-ID TF genes, deletion of the general chromatin organizer Ctcf did not give rise to preferential dysregulation of hepatocyte identity effector genes (Fig.2H). In line, similar to Hep-ID TFs and different from CTCF, binding of most Hep-ID\textsuperscript{CONNEC}T TFs was stronger at hepatocyte identity effector gene promoters (Fig.2I). We found that preferential binding remained observable when using a more stringent control, i.e. promoters with comparable accessibility and activity (Activity matched control promoters; see Materials and Methods and Fig.S13 for further details). This ruled out that chromatin accessibility alone explains preferential binding to Hep-ID TF gene promoters (Fig.2I). Finally, binding of Hep-ID and Hep-ID\textsuperscript{CONNEC}T TFs correlated at the promoter of these genes pointing to combinatorial transcriptional regulation (Fig.2J).

Altogether, these data identified that Hep-ID TFs of the hepatic CoRC target Hep-ID\textsuperscript{CONNEC}T TF genes, and that these 2 sets of TFs further collaborate to control identity effector genes.
**Hep-ID**CONNCT TFs are reciprocally targeting Hep-ID TF-encoding genes to finely tune their expression in homeostatic conditions

We hypothesized that the intimate connection between Hep-ID and Hep-ID**CONNCT** TFs identified so far could further extend to reciprocal transcriptional regulation where Hep-ID**CONNCT** TFs would bind and control expression of Hep-ID TF-encoding genes. Mining Hep-ID**CONNCT** TFs binding in mouse liver indicated greater recruitment to Hep-ID TF gene promoters when compared to control activity-matched TF-encoding gene promoters, reminiscent of data obtained for Hep-ID TFs (Fig.3A and Fig.S14). However, this only translated into moderate transcriptional changes in expression of Hep-ID TF-encoding genes. Indeed, only the deletion of a subset of the Hep-ID TF-encoding genes, and not that of Hep-ID**CONNCT** TF genes, individually triggers global down-regulation of the Hep-ID TF gene expression (Fig.3B). This most probably relates to the intrinsic role of identity TF networks, which are built-up to allow robustness of individual cell-type transcriptional programs (Almeida et al., 2021; Kyrmizi et al., 2006).

Severeral Hep-ID**CONNCT** TFs are central to establishment of rhythmic hepatocyte gene expression (Fig.S15) including both circadian (BHLHE40, CLOCK, DBP, NFIL3, NR1D1/2) (Mukherji et al, 2019) and ultradian (XBP1) (Meng et al, 2020; Pan et al, 2020) rhythms. Hence, rather than being instrumental for regulating steady-state expression levels of Hep-ID TF genes, we envisioned that Hep-ID**CONNCT** TFs might be involved in proper rhythmic expression of Hep-ID TF encoding genes. In line, mining data from (Meng et al., 2020) revealed that a majority of Hep-ID TF genes (9 out of 13), including for example *Mlxipl*, were characterized by XBP-1 dependent ultradian expression patterns (Fig.3C-D and Table S3).

**Activation of Hep-ID**CONNCT TFs resets CoRC TF expression and hepatocyte identity in dedifferentiated hepatocytes
Recent studies have found that individual TF activities may only be fully revealed in non-basal conditions and that TFs may be important to (re-)establish rather than to maintain cell-specific transcriptomes (Hunter et al., 2020; Lo et al., 2022). Therefore, we envisioned that, at least for some Hep-ID\textsuperscript{CONNECT} TFs, a role in controlling the Hep-ID TF network might be revealed in a biological context where hepatocyte identity is challenged such as cancer. Mining transcriptomics of human hepatocarcinoma (HCC) from the cancer genome atlas (TCGA; n=366) indicated that expression of Hep-ID TF genes exhibited a trend towards positive correlation with that of many Hep-ID\textsuperscript{CONNECT} TFs (Fig.4A). Moreover, binding of Hep-ID\textsuperscript{CONNECT} TFs to Hep-ID TF genes was also observed in hepatic cancer cells (Fig.4B) and overexpression or activation of several Hep-ID\textsuperscript{CONNECT} TFs, GATA binding protein 6 (GATA6), krüppel-like factor 9 (KLF9), T-box transcription factor 3 (TBX3) and the nuclear receptors liver X receptor alpha (NR1H3 also known as LXR alpha) and thyroid hormone receptor beta (THRB), has been shown to oppose hepatocarcinogenesis (Fig4A) (Kowalik et al., 2020; Liang et al., 2021; Lin et al., 2020; Sun et al., 2014; Tan et al., 2019). Mechanisms involved have largely remained undefined and we hypothesized that these HCC suppressive effects may be linked to a resetting of Hep-ID TF expression. Therefore, we leveraged transcriptomic analyses performed on HCC cell-lines treated with ligands activating NR1H3 (Huh7 cells treated with GW3965) (Vazquez Salgado et al., 2022) or THRB (HepG2 cells treated with 3,3′,5-triiodo-L-thyronine (T3) or GC-1) (Yuan et al., 2012). In line with our hypothesis, we found that treatment with GW3965, T3 or GC-1 induced a global increase in Hep-ID TF gene expression in HCC cell-lines (Fig.4C). We were able to reproduce induction of Hep-ID TF gene expression by T3 and GC-1 in HepG2 cells, induction of \textit{DIO1} being used as a positive control in these assays (Fig.4D-E). In addition, THRB activation also increased the ratio of HNF4A adult promoter P1- to embryonic promoter P2-derived isoforms (Fig.4F), a phenomenon known to promote hepatocyte differentiation and function (Dubois et al.,
Western blot assays confirmed T3-induced changes in Hep-ID TF levels (Fig.4G). Similar data were obtained using a non-cancerous human immortalized hepatic cell-line (IHH) used as another model of partially dedifferentiated hepatocytes (Schippers et al., 1997a) (Fig.S16). Moreover, interrogating the transcriptomic data obtained in a rat model of hepatocarcinogenesis (Kowalik et al., 2020) revealed that T3 was able to counteract the cancer-driven loss of expression of Hep-ID TFs (Fig.4H). Indeed, T3 treatment led to robust transcriptional induction of Hep-ID TF-encoding genes whose expression was most strongly down-regulated in the cancerous nodules (Fig.4I). This was in contrast with T3 not triggering a global rise in Hep-ID TF expression in healthy mice (Fig.S17) (Singh et al., 2018) further substantiating a role for THRB in controlling Hep-ID TF genes only in context of dedifferentiated hepatocytes. T3-mediated re-expression of Hep-ID TFs further translated into re-expression of the most strongly downregulated identity effector genes consistent with concomitant re-establishment of a transcriptional program closer to that of differentiated hepatocytes (Fig.4J).

Beyond cancer, as stated earlier, we and others have reported that hepatocyte partial loss of identity is a hallmark of severe liver injuries [recently reviewed in (Berasain et al., 2022)]. In particular, inflammation is a well-established common driver of hepatocyte dysfunction (Del Campo et al., 2018). For instance, tumor necrosis factor or interleukin 1 beta (IL1B) have been shown to trigger loss of hepatocyte identity (Hyun et al., 2020). In this context, we set-up an experimental model where mice were injected with IL1B for 3h before livers were collected. IL1B treatment indeed triggered loss of hepatocyte identity including decreased expression of several Hep-ID TF genes as assessed through transcriptomic analyses (Fig.S18A-B). We next determined whether T3 treatment promoted the recovery of Hep-ID TF gene expression in this model, i.e. mice were first injected with IL1B and 3h latter with or without T3 for an additional 3h (Fig.5A). This acute setting allowed us to investigate the
effect of T3 independent of its previously described pro-regenerative activities through modulation of hepatocyte proliferation (Tang et al., 2022). Inbred laboratory mice raised in standardized conditions nevertheless keep displaying high phenotypic trait variability (Gartner, 1990; Tuttle et al., 2018) and are prone to polyphenism, i.e. several discrete phenotypes on the same genetic background (Dalgaard et al., 2016; Yang et al., 2022). Here, mice were subdivided into 3 groups (low, intermediate and high responsiveness groups) based on their response to T3 as judged using stimulation of the Dio1 and Hectd3 gene expression, which are previously identified THRB hepatic target genes (Fig.5B) (Paquette et al., 2011). Importantly, these differences were due to variability in the T3 response only, since IL1B-mediated induction of pro-inflammatory genes was similar between the 3 groups and no negative correlation between T3 and IL1B responses was detected (Fig.5C and Fig.S17C). Interestingly, the IL1B+T3 high responsiveness group showed expression recovery of Hep-ID TF genes whose levels were higher in this group when compared to those in the IL1B only and/or IL1B+T3 low responsiveness groups (Fig.5D).

Altogether, these data have identified the Hep-ID\textsuperscript{CONNECT} TF THRβ as an actionable target to reset expression of hepatic identity TF genes in partially dedifferentiated hepatocytes.

**DISCUSSION**

Our promoter-centric and multi-omics approach called ProTFnet allowed us to unveil the complexity of the TF network controlling hepatocyte identity and functions. Using this approach, we identified that Hep-ID TFs of the CoRC are tightly connected with an additional layer of TFs through reciprocal co-recruitment to their promoters. The CoRC, in addition to being instrumental for cell identity establishment and maintenance, comprises TFs which are also key for controlling and adapting mature cell functions to homeostatic requirements. In
In this context, the tight connection between Hep-ID\textsuperscript{CONNECT} TFs and the CoRC provides an additional layer of regulation to finely tune Hep-ID TF expression. For instance, Hep-ID\textsuperscript{CONNECT} TFs comprise regulators of rhythmic gene expression including XBP1. Other Hep-ID\textsuperscript{CONNECT} TFs, TBX3 and TCF7L2, have been identified as regulators of zonated hepatocyte transcriptional programs (González-Blas\textit{et al.}, 2022) pointing to a role of Hep-ID\textsuperscript{CONNECT} TFs in specifying proper hepatic gene expression in both space and time. Importantly, rhythmic and zonated hepatic gene expression are crucial to maintain appropriate liver metabolic and non-metabolic activities (Meng\textit{et al.}, 2020; Mukherji\textit{et al.}, 2019; Pan\textit{et al.}, 2020; Paris & Henderson, 2022).

While not being required to sustain Hep-ID TF gene expression and cell identity in basal condition, our data obtained with NR1H3 and THR in dedifferentiated hepatocytes revealed that Hep-ID\textsuperscript{CONNECT} TF activation can be leveraged to reset the CoRC and hepatocyte identity. These findings are consistent with other recent observations where TF binding to cell-specific target genes translates into a more critical role in (re)establishment and dynamic regulation rather than maintenance of steady-state gene expression levels (Lo\textit{et al.}, 2022). Our study is also consistent with recent insights into the understanding of cell fate choice, which have suggested a more distributed form of control relying on combinatorial activities of dozens of TFs rather than an organization strictly dominated by a few master TFs (Chubb\textit{et al.}, 2021; Mittnenzweig\textit{et al.}, 2021). In this context, we propose that Hep-ID\textsuperscript{CONNECT} TFs are involved in identity stabilization loops, whose functional importance with regards to cell identity is revealed in pathophysiological situations of dedifferentiation (Fig.6). Context-dependent activities is an intrinsic property of TFs and precisely defining how Hep-ID\textsuperscript{CONNECT} TFs sense the cellular state to adapt their functions will require additional studies.

Loss of hepatocyte identity is a main feature of tumorigenesis, but is now ascribed a broader pathophysiologival relevance including loss of activity of severely injured livers.
(Berasain et al., 2022). In both situations, resetting hepatocyte identity through re-expression of CoRC TFs is considered of potential therapeutic interest (Berasain et al., 2022; Chao et al., 2020). By identifying the Hep-ID\textsubscript{CONNECT} TF-controlled identity stabilization loops, our study has identified new avenues to achieve this goal. In particular, our data with agonists of the nuclear receptors NR1H3 and THRBP are particularly relevant as several of those are being considered for treatment of liver diseases (Hatziagelaki et al., 2022; Russo-Savage & Schulman, 2021). With regards to HCC, our data provide explanation for the reported beneficial effects exerted by NR1H3 and THRBP. Many Hep-ID\textsubscript{CONNECT} TFs, including NR1H3 and THRBP, do not display typically used features to define hepatocyte identity TFs such as hepatocyte-specific expression and association with super-enhancers (Dubois et al., 2020a; Hnisz et al., 2013) (Table S4). Hence, our study has allowed to uncover an unforeseen role for an extended transcriptional regulatory network beyond the CoRC in the control of hepatocyte identity (Fig.6).

MATERIALS AND METHODS

Data retrieval

Public functional genomics data used in this study were downloaded from the Gene Expression Omnibus (GEO) (Barrett et al., 2013), ArrayExpress (Sarkans et al., 2021), ENCODE (Yue et al. 2014), the UCSC Genome Browser (Raney et al. 2011), FANTOM5 (Lizio et al., 2015) or from BioGPS (Wu et al. 2009) and are listed in Table S2.

Hep-ID TFs were defined as hepatocyte-identity TFs retrieved in at least 2 out of 3 independent studies (D’Alessio et al., 2015; Dubois et al., 2020a; Zhou et al., 2017). The top 20 TFs were used for (D’Alessio et al., 2015).

Mouse liver active promoters were defined based on Precision nuclear Run-On Sequencing (PRO-Seq) data (Wang et al, 2018). HepG2 active promoters were defined as the
transcriptional start sites (TSSs) from Gencode v32 (Frankish et al., 2018) with H3K27ac ChIP-seq signal (ENCODE, Table S2) with at least 2-fold enrichment over control within a 1 kilobase (+/- 500 bp of TSS) window. Only the most active TSS per gene, i.e. highest H3K27ac ChIP-seq signal, were kept.

The IDR thresholded peaks issued from ChIP-seq analyses of flag-tagged Hep-IDCONNECT TFs (n=14) in HepG2 cells were obtained from ENCODE (Table S2).

Tau indexes of tissue-specific expression were retrieved from (Kryuchkova-Mostacci & Robinson-Rechavi, 2017).

**Self-Organizing Maps (SOM) analyses**

Active promoters in the mouse liver were defined as *cis*-regulatory modules, i.e. genomic regions bound by at least 2 different transcriptional regulators (Fig.S2A-B) as in (Dubois-Chevalier et al., 2017a), which overlap active TSSs identified as PRO-seq data summits in (Wang et al., 2018). Non-unique associations between TSSs and genes in the PRO-seq data were discarded. Only transcription factor (TF) encoding genes, retrieved using a manually curated gene list originally obtained from the AnimalTFDB2.0 database (Zhang et al., 2015), were considered. Transcriptional regulator co-recruitment analyses made use of mouse liver ChIP-seq data for 49 factors (Table S2), which were uniformly processed including quality control checks as described in (Dubois-Chevalier et al., 2017a). All these data were obtained using the liver of untreated adult mice. In line with the vast majority of the liver chromatin stemming from hepatocytes (which are the prominent cell-type in the liver and are, moreover, mostly polyploid cells - up to 85% in mice with mainly tetraploid hepatocytes) (Duncan et al., 2010), hepatic TF cistromes can be faithfully inferred from mouse liver ChIP-seq data (Dubois-Chevalier et al., 2017a; Dubois et al., 2020a; Kyrmizi et al., 2006; Schmidt et al., 2010; Sommars et al., 2019).
The SOM were generated using the R package kohonen2 (Wehrens & Buydens, 2007) as described in (Dubois-Chevalier et al., 2017a). Cells containing promoters with similar transcriptional regulator binding patterns (Fig.S2C-D) were further grouped into clusters (denoted A-G) based on hierarchical clustering performed using the hclust function of the R package Stats (R_Core_Team, 2015). We used the Ward agglomeration method and the best representative transcriptional regulator combination (prototype) for each individual cell. The number of clusters was chosen according to homogeneity analyses (http://lastresortsoftware.blogspot.fr/2010/08/homogeneity-analysis-of-hierarchical.html) (Bedward et al, 1992) and biological significance. A planar projection of the toroidal map was used for data visualization.

**Analyzes of transcriptional regulator co-recruitment patterns in SOM-derived clusters**

Transcriptional regulator co-occurrence in clusters A-G was used to calculate Tanimoto distance matrices, which were used to draw heatmaps and perform hierarchical clustering or perform multi-dimensionnal scaling (MDS) analyses using R (R_Core_Team, 2015) as described in (Dubois-Chevalier et al., 2017a). To analyze the combinations of transcriptional regulators bound at the different promoters of a given cluster, a “frequent itemsets” search (combinations of 2 to 49 regulators) was performed using the arules R package (Hahsler et al, 2005). Only itemsets occurring in at least 50% of promoters of a given cluster were considered and defined as the core co-recruitment nodes. Finally, the percentage of occurrence of each transcriptional regulator in these nodes was retrieved.

**Transcriptomic data analyses**

*Processing of raw data and differential gene expression analyses*
Raw transcriptomic data from Affymetrix microarrays were normalized on a local instance of Galaxy (Afgan et al., 2018) using the GIANT APTtool (Affymetrix Power Tools; www.thermofisher.com/fr/fr/home/life-science/microarray-analysis/microarray-analysis-partners-programs/affymetrix-developers-network/affymetrix-power-tools.html) (with options: gc correction when available, scale intensity and rma at probeset level) from the GIANT toolbox (Vandel et al., 2020). Normalized expressions were averaged per Gene Symbol (NetAffx Annotation Release 36, July, 2016). Raw transcriptomic data from Agilent microarrays were processed with the limma R package (Ritchie et al., 2015) used to normalize the data through the “backgroundCorrect” function (parameters used were method :"normexp" and normexp.method : "rma") and to filter out low expressed probes. Normalized data from Illumina Bead Chips were retrieved from the Gene Expression Omnibus database and annotated using GPL6101_Illumina_RatRef-12_V1_0_R1_11222119. Orthologous mouse genes were retrieved using Ensembl annotations (release 105) (Howe et al., 2021). Differential expression analyses were performed with GIANT using the limma tool (FDR cutoff set at 0.15).

RNA-seq data were analyzed using the Galaxy web platform (Afgan et al., 2018). Mapping of reads on mm10 was performed with HISAT2 (version 2.21; options: default) (Kim et al., 2019). Mapped reads mapping to exons were subsequently retrieved and merged by gene_id with Htseq-count (version 0.9.1) (Anders et al., 2015) using the mm10 annotation of Ensembl (release 102) (Yates et al., 2020) (with the following options, mode: union, minimum alignment quality: 10, “do not count non uniquely or ambiguously mapped reads”, stranded: no). Normalization and differential analyses were performed with EdgeR (version 3.36.0) (Liu et al., 2015; Robinson et al., 2010) (options: lowly expressed genes filtered out, cutoff: <1CPM in n samples - with n corresponding to number of biological replicates in one condition, FDR<0.15, normalization method: TMM, Robust settings: True). Normalized gene
expressions were obtained by averaging data per Gene Symbol using the Ensembl annotation (release 105) (Howe et al., 2021). Dot plots displaying dysregulation in transcriptomic data were performed with the ggplot2 R package (Wickham, 2016).

**Mining of CAGE-seq data to define CTS (cell-type specific), CTE (cell-type enriched) and UBQ (ubiquitous) TFs**

CAGE-seq data from FANTOM5 (Lizio et al., 2015) were used to define CTS, CTE and UBQ TF-encoding genes based on their patterns of expression in primary human and mouse cells. Human and mouse orthologous genes were defined using the Ensembl annotation (version 105) (Howe et al., 2021). Only TF-encoding genes retrieved in both the human and mouse CAGE-seq data (n=1009) were considered. Their expression levels in mouse and human cells were stacked in a unique matrix centered and scaled, which was used to perform a Multiple Factor Analysis (MFA) using the MFA function of the R package FactoMineR (Lê et al., 2008) considering expression in mouse and human cells as two distinct groups of variables and TFs as individuals (parameters were set as follows, type: c(‘s’,’s’), ncp: 5). The 2 first components of this MFA accounted for approximately 84% of the variability of the two combined datasets. A hierarchical clustering was performed on these 5 first components using the HCPC function of the FactoMineR package (parameters were set as follows, nbclust: 3, metric: “euclidean”, method: “ward”). This identified 3 clusters of TFs characterized by an increasing number of cell-types displaying high expression from CTS to UBQ TFs (Fig.S8).

**Mining of transcriptomic data to monitor hepatic dedifferentiation**

To monitor hepatocyte dedifferentiation in mouse liver injury models, all liver injury transcriptomic data were batch corrected against the liver development transcriptomic data used as a reference. This was performed with the ComBat function of the R package SVA (parameters were set as follows, mean.only: T, par.prior: T, control samples for the batch
correction were the adult livers for the differentiation study and the livers from wild-type or untreated mice for the injury studies) (Leek et al, 2022). Then a principal component analysis (PCA) was computed on the liver differentiation study using the R package FactoMineR (scale.unit set to F) (Lê et al., 2008). Since the first component, representing 64% of the dataset variability, allowed to separate the different stages of liver differentiation, it was used to project the liver injury studies. Fold changes between injured and control livers were computed on the batch corrected data and the median log₂ fold change for individual genes across the different studies was recovered and plotted.

*Mining transcriptional changes in identity genes versus control gene sets*

Transcriptional modulations of identity effector and TF-encoding genes were compared to those in control genes matched for mouse liver promoter activity. A hierarchical clustering was performed based on promoter DHS-seq and H3K27ac ChIP-seq signals together with mouse liver gene expression levels considering all active non-TF genes (Fig.S13) or only TF-encoding genes (Fig.S14). Individual signals were scaled and clustering was performed using log₂-transformed data. This allowed to define 4 clusters of promoters with different activity levels among both non-TF and TF-encoding genes. Groups of control genes were defined by randomly selecting an equivalent number of non-identity genes matching the distribution of identity effector or TF genes in these clusters. This was performed 1000 times (without replacement) to compare the identity genes to the individual control gene sets (Fig.S13 and Fig.S14). Results of Wilcoxon rank sum tests were recorded and the mode of the p-value distribution was used to select a representative control gene set among the 1000 subsamples. Selected representative control groups were in main figures.

*Gene expression correlation in human HCC*

The cbBioPortal (https://www.cbioportal.org/) (Gao et al, 2013) was used to retrieved spearman correlations between the log₂ mRNA expression of individual Hep-ID\(^{\text{CONNECT}}\) TF
genes with that of Hep-ID TF genes in HCC within "Liver Hepatocellular Carcinoma (TCGA, Firehose Legacy)" (n=366 samples).

**DHS-seq and histone ChIP-seq data analyses**

Uniform reprocessing of the data including peak calling and genome-wide average signal has been described in (Dubois-Chevalier et al., 2017a). ChIP-seq signals were also alternatively used as enrichments over input, which were obtained as follows. First, input datasets from several mouse liver ChIP-seq studies (Table S2) were merged into a “meta-input” file after removal of duplicated reads and false-positives regions identified by ENCODE (blacklisted regions v1) (Amemiya et al, 2019) or defined as repeatedly enriched in inputs and IgG ChIP-seq in our previous study (Dubois-Chevalier et al., 2017a). Then, the Bam file for each TF ChIP-seq dataset was used to run MACS2 callpeak (-g=1.89e9, -q=0.05, --keep-dup=all, --scale-to=small, B) using the “meta-input” Bam file as control. Finally, the two BedGraph files issued from MACS2 callpeak were used in MACS 2 bdgcmp (m=logFE) to obtain the genome-wide log2 fold enrichment track files over input for each TF (MACS2 version 2.2.7.1) (Zhang et al, 2008). DHS-seq or ChIP-seq signal at a given promoter was defined as the maximum signal within TF binding sites (i.e. cis-regulatory modules as defined hereabove) encompassing this promoter retrieved using the extract bed function of bwtool version 1.0 (Pohl & Beato, 2014). Comparison of individual TF binding to identity and control gene promoters involved reiterative comparisons with independent matched control groups as described hereabove for transcriptomics analyses. Combinatorial binding of Hep-ID TFs, Hep-ID\textsuperscript{CONNECT} and CTCF on effector gene promoters was compared using the distance correlation provided by the dcor function of the Rfast package in R (Papadakis et al, 2022).

Visualization of ChIP-seq data at selected genes was performed using the Integrated Genome Browser (IGB 9.1.4) (Freese et al, 2016)
H3K4me3 ENCODE ChIP-seq data from mouse tissues (Shen et al., 2012) were used to call H3K4me3-enriched regions using the broad peak calling option of MACS2 as described in (Chen et al., 2015). Genes were assigned to H3K4me3-enriched regions using the closestBed function (with parameters -t all, -k 1, -d, -mdb) of the BEDTools version 2.3.0 (Quinlan & Hall, 2010), i.e. active gene TSS from the mouse liver PRO-Seq data (Wang et al., 2018) were assigned H3K4me3-enriched regions when distance was 0. Calling of super-enhancers (SE) and target gene assignation, essentially performed as in (Loven et al., 2013; Whyte et al., 2013), was previously described (Dubois-Chevalier et al., 2020).

**Classification of TFs based on their promoter hepatic activity score**

In order to monitor liver specificity, the DHS-seq or H3K27ac ChIP-seq signals were extracted considering a region spanning +/- 500 base pairs around active promoters (summits of PRO-seq data) (Wang et al., 2018). The DHS and H3K27ac signals for each tissue were normalized using the Spark method (Nielsen et al., 2012), which takes into account signal median and variance on all genes in a tissue. Liver specificity indexes were computed by dividing the normalized signals in the liver by the average normalized signals in other tissues (Table S2). A gene expression liver specificity index was similarly computed using normalized gene expressions extracted from the BioGPS Mouse MOE430 Gene Atlas (Wu et al., 2016) (Table S2). Log2-transformed signals, gene expressions and liver-specificity indexes were combined in a matrix, which was centered and scaled and further used for hierarchical clustering using the hclust function of R (parameters were set as follows, method: “Ward.D2” on Euclidean distance matrix) (R_Core_Team, 2015). This allowed to define 7 TF-encoding gene promoter classes characterized by increasing liver activity and specificity together referred to as hepatic activity score (Fig.S6). Heatmaps were plotted with heatmap.2 from the gplots R package (Warnes et al., 2022).
H3K27ac ChIP-seq data from purified mouse primary hepatocytes versus non-hepatocytes (Roh et al., 2017) were further used to characterize the obtained clusters. CAGE-seq data from FANTOM5 (Table S2) were also leveraged to calculate gene expression in hepatocytes when compared to average expression in other primary cell-types.

TF-encoding gene promoters were grouped according to both clustering based on transcriptional regulator co-recruitment (clusters A-G) and based on hepatic activity and specificity (classes 1-7).

**Hep-ID\textsuperscript{CONNECT} TF binding to Hep-ID TF encoding genes in HepG2 cells**

**Mouse phenotype ontology and literature mining**

Enrichments within the mouse phenotype ontology or biological pathway data were defined using ToppCluster (Chen et al., 2007; Kaimal et al., 2010). Default parameters were used and only terms linked to >5 TF-encoding genes were considered.

References to Hep-ID\textsuperscript{CONNECT} TFs in the scientific literature related to liver/hepatocyte metabolic functions were retrieved using the easyPubMed package (https://www.data-pulse.com/dev_site/easypubmed/) in R. Articles with co-occurrence of a given Hep-ID\textsuperscript{CONNECT} TF gene name and (“hepatocyte” or “liver”) and “metabolism” in their title or abstract were considered. The number of retrieved manuscripts for each Hep-ID\textsuperscript{CONNECT} TF was visualized as a heatmap prepared using the heatmap.2 function of the R package gplots.

**Cell culture and treatments**

The human cell line HepG2 (ATCC, HB-8065) was cultured in minimum essential medium (MEM) (Gibco, 11095080) supplemented with 10% fetal bovine serum (FBS;
Dutscher, 500105H1), MEM non-essential amino acids (Gibco, 11140035), 1 mM sodium pyruvate (Gibco, 11360070) and 100 U/mL penicillin-streptomycin (Gibco, 15140). Immortalized human hepatocytes (IHH) (Schippers et al, 1997b) were cultured in William’s E medium (Gibco, 22551022) supplemented with 10% FBS (Dutscher, SV30160.03), 20 mU/mL bovin insulin (Sigma, I5500), 50 nM dexamethasone (Sigma, D1756) and 100 U/mL penicillin-streptomycin (Gibco, 15140). Cell culture was performed in a humidified atmosphere of 5% CO2 in a 37 °C incubator. Cells were grown in media containing 10% dextran-coated charcoal stripped serum for 48h and then treated with 10 nM T3 (T6397, Sigma-Aldrich) or 10 nM GC-1 (SML1900, Sigma-Aldrich) for 24 or 96h before being harvested for RNA or protein extraction. For serum stripping, activated charcoal solution (C9157, Sigma-Aldrich) was washed 3 times and prepared in ultra-pure water at final concentration 5%. Dextran T70 (2.5g ; 31390, Sigma-Aldrich) was added to the pellet and charcoal-coated dextran was mixed with 500 mL of serum and incubated overnight at 4°C. After centrifugation (20 min, 4000g, room temperature), the serum was filtered using a 0.2 μm filter and heat inactivated at 56°C for 45 min.

**Mouse experiments**

Young adult male C57BL/6J wild-type (WT) mice (7-11 weeks old) were purchased from Charles River and housed in standard cages in a temperature-controlled room (22–24°C) with a 12-h dark–light cycle. They had *ad libitum* access to tap water and standard chow and were allowed to acclimate for 2 weeks prior to initiation of the experimental protocol. Acute liver inflammation was induced by intraperitoneal injection of recombinant mouse IL1B using 0.5 μg/ mouse (575102, BioLegend) or vehicle (PBS) for 3h followed by T3 treatment 0.2 mpk (T6397, Sigma-Aldrich). Livers were collected 3h later. Mice with hepatocyte-specific deletion of *Thrb* have been described in (Billon et al, 2014).
All animal studies were performed in compliance with EU specifications regarding the use of laboratory animals and have been approved by the Nord-Pas de Calais Ethical Committee (APAFIS#30322-202102221656794 v4).

RNA expression analyses

Tissues were homogenized using Minilys and 1.4 mm ceramic beads (Bertin Technologies). Total RNA was extracted from cell lines and tissues using the Nucleospin® RNA kit (Macherey-Nagel) according to the manufacturer’s protocol.

RNA was reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystem). Quantitative PCR (qPCR) was performed on an Fast Applied (Applied Biosystem, Life Technologies, Cergy Pontoise, France) using the Takyon kit (Agilent Technologies). The specificity of the amplification was checked by recording the dissociation curves, and the efficiency was verified to be above 95% for each primer pair. mRNA levels were normalized to the expression of housekeeping genes and the fold induction was calculated using the cycle threshold (ΔΔCT) method. The sequences of primers used are listed in Table S5.

For transcriptomic analyses, RNA integrity and quantity were evaluated using the Agilent 2100 Bioanalyser (Agilent Technologies). RNA was then processed for transcriptomic analysis using Affymetrix GeneChip arrays (MoGene 2.0) or high-throughput sequencing (RNA-seq) as previously described (Bobowski-Gerard et al., 2022).

Protein extraction

Proteins from the chromatin fraction were prepared as in (Dubois et al., 2020a). Cells were scraped in ice-cold Phosphate-Buffered Saline (PBS), pelleted by centrifugation at 400 g for 5 min, lysed in Buffer A (50 mM HEPES pH 7.5, 10 mM KCl, 1.5 mM MgCl2, 340 mM...
sucrose, 10% glycerol, 1 mM DTT and protease inhibitor cocktail from Roche) and incubated for 10 min at 4°C. Samples were centrifuged at 1,300 g for 5 min at 4°C and supernatants were discarded. Nuclear pellets were washed with Buffer A and subsequently lysed in solution B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, and protease inhibitor cocktail). After incubation for 30 min at 4°C, samples were centrifuged at 1,700 g for 5 min at 4°C and supernatants were discarded. Chromatin pellets were washed with solution B, resuspended in Buffer C (50 mM Tris–HCl pH 8.0, 1 mM MgCl2, and 83 U/µl benzonase) and incubated for 20 min at 4°C. Laemmli buffer 6× was added before loading for Western immunoblotting.

**Western blot assays**

Protein extraction and western blotting were performed as described in (Dubois et al., 2020a). Protein concentrations were determined using the Pierce™ BCA protein assay kit (Thermo scientific). Then, 20 µg of proteins were separated by 10% SDS-PAGE and immunodetected by Western immunoblotting using the primary antibodies listed in the Table S6. Primary antibodies were detected using HRP-conjugated secondary antibodies (Sigma-Aldrich). Images were acquired using the iBright™ CL1500 Imaging System (Thermo Fisher Scientific).

**Statistical analyses**

Statistical analyses were performed using the Prism software (GraphPad, San Diego, CA) and R (R_Core_Team, 2015). The specific tests and corrections for multiple testing which were used as well as the number of mice or independent biological replicates are indicated in the figure legends. Two-sided tests were used unless specified in the figure legends. All bar graphs show means ± SD (standard deviations). Box plots are composed of a box from the 25th to the 75th percentile with the median as a line. Unless specified in the
figure legend, whiskers extent to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

ACKNOWLEDGMENTS

The authors thank Drs Wang and Liu (Vanderbilt University Medical Center, Nashville, TN, USA) for providing the summit coordinates of active mouse liver promoters and Drs Ninon Very N and Marie Bobowski-Gerard (Inserm U1011, Univ. Lille) for technical assistance.

Financial support: This work was supported by the Agence Nationale de la Recherche (ANR) grants “HSCreg” (ANR-21-CE14-0032), “MEdicAL” (ANR-21-CE17-0016) and “European Genomic Institute for Diabetes” E.G.I.D (ANR-10-LABX-0046), a French State fund managed by ANR under the frame program Investissements d’Avenir I-SITE ULNE / ANR-16-IDEX-0004 ULNE as well as by Fondation Recherche Médicale (FRM grants EQU202203014645).

DISCLOSURES

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: JD-C, CG, KG, LD, BS, PL, JE; validation: JD-C, CG; formal analysis: JD-C, CG, JE; investigation: JD-C, CG, AB, CB, VD, LG, MF, GM; resources: KG, LD; data curation: JD-C, CG, JE; writing—review & editing: JD-C, CG and JE with inputs from co-authors; visualization: JD-C, CG, JE; project administration: BS, PL, JE; funding acquisition: BS, PL, JE.
DATA AVAILABILITY

Transcriptomics data generated in this study have been deposited into the Gene Expression Omnibus (GEO) database and are available under accession numbers GSE216278 and GSE218724. The following secure tokens allow review of these GEO records: olclmskcbdsxdij and gfwfakaezxoppsr, respectively.

REFERENCES


Hatzigelaki E, Paschou SA, Schon M, Psaltopoulou T, Roden M (2022) NAFLD and thyroid function: pathophysiological and therapeutic considerations. Trends Endocrinol Metab


Tachmatzidi EC, Galanopoulou O, Talianidis I (2021) Transcription Control of Liver Development. Cells 10


Figure legends:

**Fig.1. Promoter-centric mining of the hepatic TF network using ProTFnet.**

A) The cistromes of 8 Hep-ID TFs (CEBPA, FOXA2, HNF4A, NR1H4, NR5A2, ONECUT1, PPARA, PROX1; Table S2) were used to define the percentage of those TFs binding to Hep-ID TF (n=13) or non-Hep-ID TF (Control TF genes; n=13) encoding gene promoters. The control group used was selected for providing data representative of those obtained with 1000 reiterations of this analysis. Wilcoxon rank sum test with continuity correction was used to assess statistical significance. *p<0.05. B) Overview of the ProTFnet strategy implemented in this study where (identity) TF binding to TF-encoding gene promoters is monitored and subsequently used to define distinct clusters of promoters through SOM and hierarchical clustering. Clusters are subsequently characterized using multi-omics (cistromic, epigenomic and transcriptomic) data mining to explore the complexity of the identity TF network. C) Planar view of the toroidal map issued from the SOM analysis was used here to display clusters A-G and hereafter to visualize different features of these clusters (panels D-E and Fig.S3E). The dendrogram issued from the hierarchical clustering analysis is shown on the right. D) The map issued from the SOM analyses was used to show the average number of co-recruited TFs at gene promoters contained in individual cells. Bold orange lines indicate the borders of clusters A-G. E) The map issued from the SOM analyses was used to show the average ChIP-seq signal for mouse liver H3K27ac at gene promoters contained in individual cells. Bold black lines indicate the borders of clusters A-G. F) Heatmap showing the occurrence (percentage) of individual transcriptional regulators in the core co-recruitment nodes of clusters A-G, i.e. binding combinations found in at least 50% of the promoters of a given cluster. G) Heatmap displaying the distribution (percentage) of promoters for each one of clusters A-G into clusters 1-7, where promoters were grouped according to increasing hepatic activity score (Fig.S6). H) Enrichment for Hep-ID TF genes (Table S1) within the
different groups of TF gene promoters. Here, and in all subsequent analyses of the clusters A-G and 1-7 cross-comparison, groups containing less than 10 promoters were not displayed. I) The normalized cistromes of 8 Hep-ID TFs (see A) were used to define the median Hep-ID TF ChIP-seq signal at promoters found in the different groups. J) Principal component analysis (PCA) of TF gene expression in mouse (n=39) and human (n=126) primary cell-types (see Materials and Methods). Individual TFs are displayed as dots projected on the first 2 components and the 3 main clusters issued from hierarchical clustering are shown and labelled as UBQ (ubiquitous), CTE (cell-type enriched) and CTS (cell-type specific) (Fig.S8). K) The data from panel J were used to selectively display TF genes from cluster G.
Fig.2. Identification and characterization of Hep-ID\textsuperscript{CONNECT} TFs.

A) The CTE/UBQ TFs from cluster G (Fig.1K) were plotted based on their expression in mouse primary hepatocytes (MPH) when compared to other primary mouse cell-types (n=38). For each individual TF, cells were first ranked according to decreasing gene expression and the rank of MPH was plotted on the x axis (i.e. rank 1 indicates highest expression is in MPH). Second, expression in MPH was divided by the average expression in all other primary cells and plotted on the y axis as log\textsubscript{2} fold difference. Hep-ID\textsuperscript{CONNECT} TF-encoding genes were defined as those preferentially expressed in MPH (rank ≤10 and FC>0). For comparison, Hep-ID TF genes were plotted in an additional box on the right of the one highlighting Hep-ID\textsuperscript{CONNECT} TFs. B-E) Expression of Hep-ID, Hep-ID\textsuperscript{CONNECT} and remaining TF-encoding genes from cluster G (Others) was monitored in indicated transcriptomic data (Table S2). Box plots show log\textsubscript{2} fold changes between MPH of \textit{Hnf4a}\textsuperscript{-/-} (\textit{Hnf4a} KO) versus wild-type mice (B), adult versus newborn mouse livers (C), a meta-analysis of severe mouse liver injuries versus control livers (D; see Materials and Methods and Fig.S10) and microdissected hepatocytes from alcohol-related human liver cirrhosis (alcoholic steatohepatitis) versus control livers (E). Statistical significance was assessed using one-sided one-sample t-test with Benjamini-Hochberg correction for multiple testing to determine if the mean log\textsubscript{2} FC was statistically lower (B, D, E) or higher (C) than 0. *q<0.05. F) Distribution of H3K4me3 domain length at the TSS of Hep-ID, Hep-ID\textsuperscript{CONNECT} and remaining TF-encoding genes from cluster G (Others) as defined through broad peak calling on mouse liver H3K4me3 ChIP-seq data. Statistical difference between groups was defined using Kruskal-Wallis with Wilcoxon pairwise comparison tests followed by Benjamini-Hochberg correction for multiple testing correction. *q<0.05. G) Mouse phenotypes associated with Hep-ID, Hep-ID\textsuperscript{CONNECT} and remaining TF-encoding genes from cluster G (Others) were defined using ToppCluster. Dendrograms of hierarchical clustering are shown. ToppCluster uses
hypergeometric tests and Bonferroni correction. **H)** Enrichment for identity effector genes among the top 1000 transcriptionally dysregulated genes in the MPH/livers of indicated genetically deficient mice. Log₂ odds ratios were computed to compare the proportion of dysregulated versus non-dysregulated genes within identity effector genes or control non-TF-encoding genes. Then a two-sided Fisher exact test was performed to assess if the proportion of dysregulated genes was significatively different within the identity and control gene groups with Benjamini-Hochberg correction. *q<0.05. **I)** Binding of indicated Hep-ID and Hep-ID\(^{\text{CONNECT}}\) TFs to the promoter of identity effector genes and a control group of non-TF-encoding genes of similar size (n=424) was monitored using mouse liver ChIP-seq data. A control group of non-TF-encoding genes (n=424) matched for their promoter mouse liver activity was also used (Fig.S13). The control groups used were selected for providing data representative of those obtained with 1000 reiterations of this analysis (see Materials and Methods; Fig.S13). The distribution of ChIP-seq signals is shown using box plots. Pairwise Wilcoxon Rank Sum Tests with Benjamini-Hochberg correction was used to define whether the binding at identity effector genes versus control genes was significantly different for each analyzed TF recruitment. *q<0.05. **J)** Correlation between Hep-ID TFs, Hep-ID\(^{\text{CONNECT}}\) TFs and CTCF recruitment, as judged through the mining of mouse liver ChIP-seq data, to identity effector genes. The dendrogram is issued from hierarchical clustering analysis.
Fig.3. Hep-ID\textsuperscript{CONNECT} TF binding to and regulation of Hep-ID TF encoding genes in basal conditions.

A) Binding of indicated Hep-ID and Hep-ID\textsuperscript{CONNECT} TFs to the promoter of Hep-ID TF genes and a control group of non-Hep-ID TF-encoding genes matched for their promoter activity (Fig.S14) of similar size (n=13) was monitored using mouse liver ChIP-seq data. The control group used was selected for providing data representative of those obtained with 1000 reiterations of this analysis (see Materials and Methods; Fig.S14). The distribution of ChIP-seq signals is shown using box plots. One-sided Wilcoxon rank sum tests with Benjamini-Hochberg correction was used to define whether the binding on Hep-ID TF gene promoters was greater than on control genes for each individual TF ChIP-seq dataset. *q<0.05. B) Transcriptional modulation of Hep-ID TFs and a control group of non-Hep-ID TF-encoding genes matched for their promoter activity (Fig.S14) of similar size (n=13) in mouse liver/MPH of mice deleted for the indicated Hep-ID or Hep-ID\textsuperscript{CONNECT} TF genes. The control group used was selected for providing data representative of those obtained with 1000 reiterations of this analysis (see Materials and Methods; Fig.S14). The distribution of log\textsubscript{2} fold changes is shown using box plots. Wilcoxon rank sum tests with Benjamini-Hochberg correction was used to define whether transcriptional regulation of Hep-ID TF genes was different from that of the control genes for each individual transcriptomic dataset. *q<0.05. C) 12hr gene expression oscillation analyses in the mouse liver performed by Meng et al. (Meng \textit{et al.}, 2020) from WT and \textit{XBP1}\textsuperscript{hep/-} animals were used to identify XBP1-dependent oscillating expression patterns for Hep-ID TF genes (Table S3). D) Average gene expression levels of \textit{Mlxipl} in the livers of WT and \textit{XBP1}\textsuperscript{hep/-} mice across circadian time.

Fig.4. Resetting Hep-ID TF gene expression through Hep-ID\textsuperscript{CONNECT} TF activation in hepatocellular carcinoma.
A) Distribution of the spearman correlations between Hep-ID TF genes and individual Hep-ID\textsuperscript{CONNECT} TF genes in human HCC shown using box plots. Hep-ID\textsuperscript{CONNECT} TFs which have been ascribed with HCC suppressive functions are indicated on the right. B) Cistromes for 13 Hep-ID\textsuperscript{CONNECT} TFs in HepG2 cells were mined for binding to Hep-ID TF encoding genes. Peaks localized +/- 10 kilobases from transcriptional start sites were considered in these analyses. C) Transcriptional modulation of Hep-ID TFs and a control group of non-Hep-ID TF-encoding genes matched for their promoter activity (Fig.S14) of similar size (n=13) in Huh7 or HepG2 cells treated with GW3965, T3 or GC-1, respectively. The control group used was selected for providing data representative of those obtained with 1000 reiterations of this analysis (see Materials and Methods; Fig.S14). The distribution of log\textsubscript{2} fold changes is shown using box plots. Wilcoxon rank sum tests with Benjamini-Hochberg correction was used to define whether transcriptional regulation of Hep-ID TF genes was different from that of the control genes for each individual transcriptomic dataset. *q<0.05. D-F) mRNA expression of the indicated genes was monitored using RT-qPCR in HepG2 cells treated with T3 or GC-1 for 24h or 96h. Bar graphs show mean ±SD of log\textsubscript{2} fold changes in treated versus untreated HepG2 cells. For Hnf4a, the log\textsubscript{2} fold change in the ratio of P1 over P2 promoter-derived isoforms is also shown. Gray dots show the results obtained from 3 independent biological replicates, each performed in technical triplicates. One-sample t-test with Benjamini-Hochberg correction for multiple testing was used to determine if the mean log2 FC was statistically different from 0. *q<0.05. G) Western blots assays performed using antibodies against the indicated proteins on extracts from HepG2 cells treated or not with T3 for 24h. Rep#1-3 indicates the 3 independent biological replicates analyzed. H) Modulation of Hep-ID TF gene expression in precancerous nodules compared to control rat livers and in liver nodules of rats treated with T3 compared to nodules of non-treated rats. A control group of non-Hep-ID TF-encoding genes matched for their promoter activity (Fig.S14) of similar size
(n=13) is also shown. The control group used was selected for providing data representative of those obtained with 1000 reiterations of this analysis (see Materials and Methods; Fig.S14). Box plots show the log2 fold changes. Wilcoxon rank sum tests with Benjamini-Hochberg correction was used to define whether transcriptional regulation of Hep-ID TF genes was different from that of the control genes for each individual transcriptomic dataset. *q<0.05. I) Dot plots showing the transcriptional regulation of individual Hep-ID TF gene expression in precancerous nodules compared to control rat livers (left; Nodules/Control) and in liver nodules of rats treated with T3 compared to nodules of non-treated rats (right; Nodules+T3/Nodules). No data were recovered for Onecut2, Proxl and Mlxipl. J) Identity effector genes significantly downregulated in Nodules/Control (q<0.05) were split in 3 groups according to their log2 fold changes (i.e. low, intermediate and high repression; pink boxes) and then monitored for induction in Nodules+T3/Nodules (green boxes). Statistical differences between the High repression and the other groups regarding the Nodules/Control comparison, on the one hand, or the Nodules+T3/Nodules comparison, on the other hand, were defined using Kruskal-Wallis with Wilcoxon pairwise comparison tests followed by Benjamini-Hochberg correction for multiple testing correction. *q<0.05.

Fig.5. Resetting Hep-ID TF gene expression through THRB activation in inflammation-induced hepatocyte dedifferentiation.

A) Experimental protocol for acute inflammation-induced loss of hepatocyte identity in vivo. Mice were injected with IL1B (IL1B; n=17), IL1B followed by T3 (IL1B+T3; n=30) or a control group (PBS; n=14). All livers were collected 6h after the initial injection. 

B-D) mRNA expression of the indicated genes was monitored in mouse livers using RT-qPCR. Mice treated with IL1B+T3 were subdivided into tertiles (n=10) based on the mean expression of the Dio1 and Hectd3 genes and defined as low, intermediate or high T3
responsiveness groups (Low resp., Int. resp. and High resp., respectively). Fold change relative to the mean of the control (PBS) group is shown using box plots whose whiskers extend to the maximum and minimum values. Statistical differences between the High resp. and the other IL1B-treated groups were defined using Kruskal-Wallis with Wilcoxon pairwise comparison tests followed by Benjamini-Hochberg correction for multiple testing correction. *q<0.05.

Figure 6. Proposed model for the control of hepatocyte identity by an extended transcription factor network. Schematic summarizing the main findings of our study pointing to an extended hepatic TF identity network which includes THRB. See discussion for greater details.
I-C - significant abnormal hepatobiliary system morphology

B.

MPH from Hnf4a KO

Expression relative to control

Mouse liver post-natal maturation

Expression relative to control

Severe mouse liver injury models

Expression relative to control

Hepatocytes from human cirrhotic livers

Expression relative to control

Mouse liver

G.

Mouse phenotype term enrichment

liver/biliary system phenotype
abnormal hepatobiliary system physiology
abnormal triglyceride level
abnormal liver morphology
abnormal liver triglyceride level
abnormal liver physiology
abnormal hepatobiliary system morphology
abnormal circulating hormone level
abnormal lymphopoiesis
abnormal leukopoiesis

H.

Enrichment for identity effector in dysregulated genes (log Odds Ratio)

Identity effector genes
Control promoters
All

J.

TF binding correlation

Hep-Id

Hep-Id Connect TFs

Hep-Id Connect TFs

Hep-Id Connect TFs

Hep-Id Connect TFs

Hep-Id Connect TFs
Fig. 3

**A**

Hep-ID TF genes

Activity-matched control TF genes

**B**

Hep-ID TF genes

Activity-matched control TF genes

**C**

Hep-ID TF 12hr oscillation

**D**

Minigene normalized RNA expression

Time (h)
Fig. 4

A. Correlation with Hep-ID TF gene expression in HCC

HCC suppressor

-0.5  0.0  0.5

-26x552

Atf4
Esrra
Hef6
Hsa4
Crem
Cebpa
Nod3
Tcf712
Dbp
Xbp1
Repil
Znf595
Bhlhe40
Atf5
Nrtl
ld2
Zbtb44
Tbx3
Gala6
Clock
KIF9
Klf12
Tnr12
Thrb
Rreb1

B. Number of bound Hep-ID cistromes (n=14)

Hep-ID TF cistromes

0  2  4  6  8  10  12

C. Hep-ID TF genes Activity-matched control TF genes

GW3965  T3  GC-1

D. mRNA expression

log$_2$ fold change / control

24h  96h

E. NR1H4 ONECUT1

log$_2$ fold change / control

24h  96h

F. HNF4A

P1 isoforms

log$_2$ fold change / control

24h  96h

P2 isoforms

log$_2$ fold change / control

24h  96h

P1/P2

log$_2$ fold change / control

24h  96h

G. NR1H4 HNF4A-p1 LAMIN A/C

55 kDa

55 kDa

55 kDa

55 kDa

T3 Rep#1 Rep#2 Rep#3

H. Hep-ID TF genes Activity-matched control TF genes

24h  96h

I. q-value

0  1  2  3

Absolute FC (log$_2$)

J. Identify effector gene

log$_2$ fold change

Repression in nodules

Low  Int.  High

Nodules/Control

Nodules+T3/Nodules

Nodules/Control

Nodules+T3/Nodules

Cepba
Foxa2
Hnfr
Hnfr4
Nr0b2
Nrl1
Nrl12
Nrl5a2
OneCut1
Ppara

0.075

Abs

P

0.05

Int.

0.15

Low

0.75

Int.

High

-0.15

Repression

Up-regulation
Fig. 5

A) Liver collection

PBS → PBS → PBS group
IL1B → PBS → IL1B group
IL1B → T3 → IL1B+T3 group

B) Dio1

C) Ccl2

Hectd3

Icam1

D) Nr1h4

Onecut1

Mxipl

Foxa2

Ppara

Onecut2

* p<0.05 compared to PBS group
Cell-specific expression/activity

Hep-ID TFs

CoRC

Hep-ID CONNECT TFs

↑ hepatic expression

Hepatocyte identity effector genes

Identity stabilization loop (dedifferentiated hepatocytes)

Fine-tuning (basal homeostatic conditions)

NRF2

THRB

Fig. 6