The logic of transcriptional regulator recruitment architecture at *cis*-regulatory modules controlling liver functions

Julie Dubois-Chevalier,[1] Vanessa Dubois,[1] Hélène Dehondt,[1] Parisa Mazrooei,[2] Claire Mazuy,[1] Aurélien A. Sérandour,[3] Céline Gheeraert,[1] Penderia Guillaume,[1] Eric Baugé,[1] Bruno Derudas,[1] Nathalie Hennuyer,[1] Réjane Paumelle,[1] Guillemette Marot,[4] Jason S. Carroll,[3] Mathieu Lupien,[2] Bart Staels,[1] Philippe Lefebvre,[1,5] and Jérôme Eeckhoute[1,5]

[1]*Université Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1011-EGID, F-59000 Lille, France;* [2]*The Princess Margaret Cancer Centre, University Health Network, Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada;* [3]*Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, United Kingdom;* [4]*Université Lille, MODAL Team, Inria Lille-Nord Europe, 59650 Villeneuve-d'Ascq, France*

Control of gene transcription relies on concomitant regulation by multiple transcriptional regulators (TRs). However, how recruitment of a myriad of TRs is orchestrated at *cis*-regulatory modules (CRMs) to account for coregulation of specific biological pathways is only partially understood. Here, we have used mouse liver CRMs involved in regulatory activities of the hepatic TR, NR1H4 (FXR; farnesoid X receptor), as our model system to tackle this question. Using integrative cistromic, epigenomic, transcriptomic, and interactomic analyses, we reveal a logical organization where *trans*-regulatory modules (TRMs), which consist of subsets of preferentially and coordinately corecruited TRs, assemble into hierarchical combinations at hepatic CRMs. Different combinations of TRMs add to a core TRM, broadly found across the whole landscape of CRMs, to discriminate promoters from enhancers. These combinations also specify distinct sets of CRM differentially organized along the genome and involved in regulation of either housekeeping/cellular maintenance genes or liver-specific functions. In addition to these TRMs which we define as obligatory, we show that facultative TRMs, such as one comprising core circadian TRs, are further recruited to selective subsets of CRMs to modulate their activities. TRMs transcend TR classification into ubiquitous versus liver-identity factors, as well as TR grouping into functional families. Hence, hierarchical superimpositions of obligatory and facultative TRMs bring about independent transcriptional regulatory inputs defining different sets of CRMs with logical connection to regulation of specific gene sets and biological pathways. Altogether, our study reveals novel principles of concerted transcriptional regulation by multiple TRs at CRMs.

[Supplemental material is available for this article.]

Regulation of gene transcription allows for the definition and maintenance of multiple cell and tissue phenotypes in higher eukaryotes as well as their ability to respond and adapt to changing environmental conditions. While active TP53-recruiting *cis*-regulatory modules (CRMs) were shown to harbor an unsophisticated organization and very low complexity (Verfaillie et al. 2016), other studies have demonstrated that active CRMs corecruit numerous transcriptional regulators (TRs) (Kittler et al. 2013; Liu et al. 2014; Siersbæk et al. 2014). This has led to envisioning CRMs as genomic nexus sites where the activities of a large set of TRs are integrated into transcriptional regulatory output signals. However, how recruitment of a myriad of TRs is orchestrated at CRMs and accounts for regulation of selective biological pathways is only partially understood.

In this context, we have used modulation of hepatic gene expression by the nuclear receptor (NR) family member NR1H4, commonly known as FXR (farnesoid X receptor), as our model system to decipher the transcriptional regulatory logic operating at CRMs. Indeed, the liver has instrumental roles in multiple homeostatic processes involving tight control of its transcriptome. Moreover, NR1H4, a nuclear receptor for bile acids (BAs), is highly expressed in the liver where it exerts broad regulatory activities. In addition to being a central node coordinating liver metabolic functions (cholesterol, BAs, lipid, and glucose homeostasis) (Lefebvre et al. 2009), NR1H4 also exerts hepatoprotective activities (Wang et al. 2008). For instance, NR1H4 promotes liver regeneration after partial hepatectomy (Huang et al. 2006) linked to its ability to promote hepatocyte survival and proliferation (Huang et al. 2006; Meng et al. 2010). Moreover, mutations in the *NR1H4* gene in humans are linked to neonatal cholestasis with rapid progression to

[5]**Joint senior authors**
Corresponding author: jerome.eeckhoute@inserm.fr
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.217075.116.

end-stage liver disease, and to vitamin K-independent coagulopathy (Gomez-Ospina et al. 2016).

NR1H4 binds DNA as a heterodimer with the retinoid X receptor (RXR) and modulates transcription through interaction with cofactors (Mazuy et al. 2015). Transcriptional regulators that may coregulate NR1H4 target genes have often been investigated at the level of single genes and CRMs (i.e., promoters or enhancers). Only a few TRs individually collaborating with NR1H4 on a genomic scale have been described, including NR5A2 (LRH1; liver receptor homolog) (Chong et al. 2012) and HNF4A (Thomas et al. 2013).

In this context, we have implemented integrative functional genomic analyses which allowed us to characterize how TR recruitment is organized at hepatic CRMs and to decipher the logical connection to regulation of specific biological pathways/functions.

Results

A large set of TRs share their recruitment sites with NRIH4 in the mouse liver genome, defining distinct classes of CRMs

In order to define TRs interconnected with NR1H4, we used our chromatin immunoprecipitation-high throughput sequencing (ChIP-seq) data from Lien et al. (2014) to define the NR1H4 cistrome in the mouse liver and compared it with that of 47 other TRs (Supplemental Table S1). We found that TRs exhibited varying levels of cistrome overlap with NR1H4 (Fig. 1A). Remarkably, all but one TR showed greater overlap with the NR1H4 cistrome than the unrelated control REST (Fig. 1A), a transcriptional repressor of neuronal genes in nonneuronal cells (Supplemental Fig. S1; Chen et al. 1998). Therefore, 45 out of the 47 analyzed TRs may combine with NR1H4 at CRMs. In order to define how these TRs are organized at NR1H4-bound CRMs, we performed integrative cistromic analyses using self-organizing maps (SOMs) (Fig. 1B; Xie et al. 2013). The map resulting from these analyses shows individual nodes grouping together NR1H4-bound CRMs with similar TR binding patterns (Supplemental Fig. S2A). These CRMs were mostly shorter than 2 kilobases (kb) (Supplemental Fig. S2B) and recruited two to 45 TRs (Supplemental Fig. S2C). In order to better define

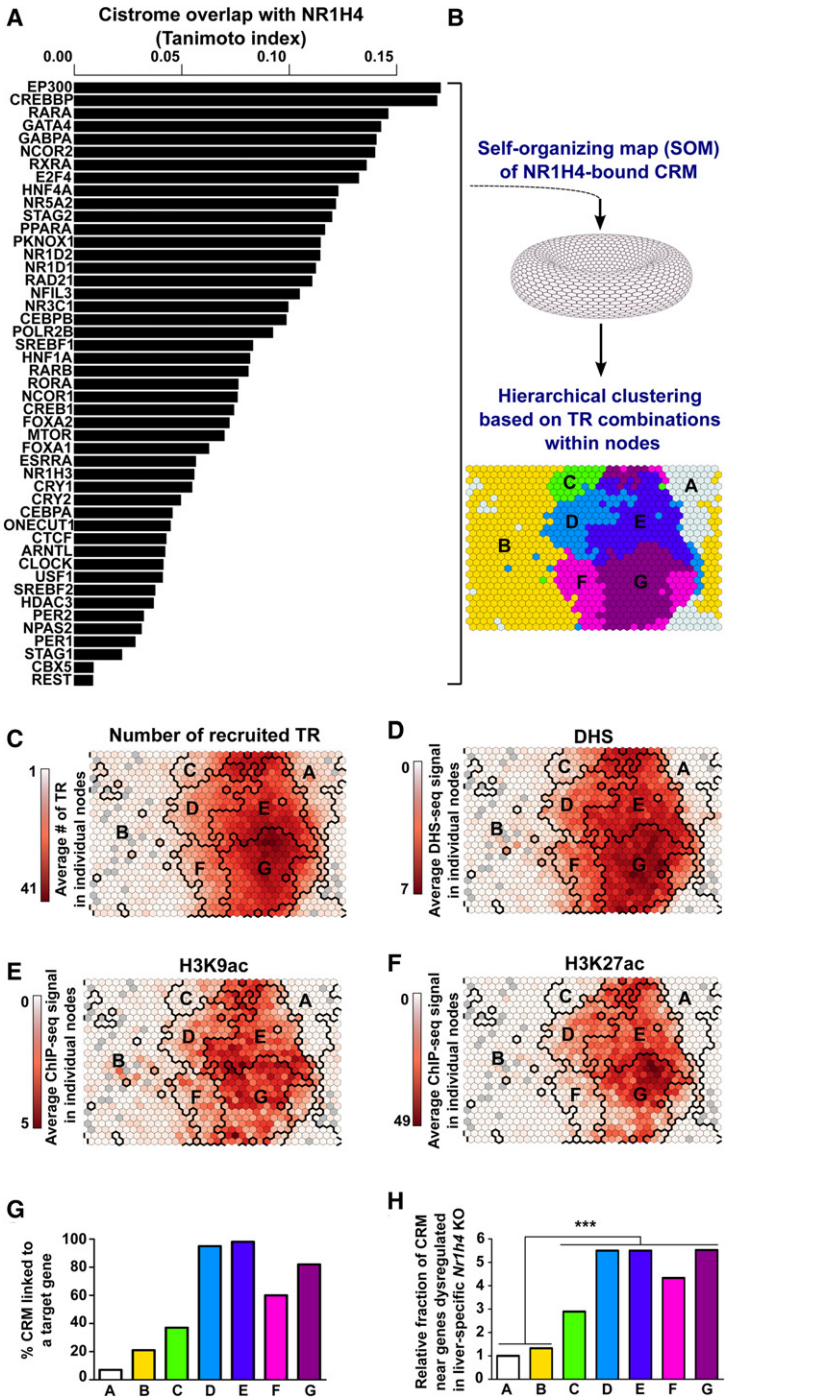

Figure 1. Integrative cistromics identifies the active subset of NR1H4-bound CRMs which consists of distinct classes of TRs recruiting CRMs. (*A*) Individual comparison of the NR1H4 cistrome in mouse liver with that of the 47 indicated TRs. (*B*) NR1H4-bound CRMs from the mouse liver genome were classified using a self-organizing map (SOM) based on their pattern of TR recruitment. Hierarchical clustering was subsequently used to identify seven main classes of CRMs which are indicated on the planar view of the toroidal map using different colors and denoted A to G. (*C*–*F*) The map issued from *B* was used to indicate the average number of binding TRs (*C*), the average DHS (*D*), H3K9ac (*E*), or H3K27ac (*F*) levels at CRMs contained in each node. Bold black lines indicate the borders of the clusters. (*G*) Percentage of CRMs from classes A–G potentially involved in gene transcriptional regulation. (*H*) Relative number of CRM from classes A–G found within 25 kilobases (kb) of the transcriptional start site (TSS) of genes whose expression is dysregulated in the liver of liver-specific *Nr1h4* KO mice. This window allows the capture of a large fraction of distal sites able to influence gene expression (Akhtar et al. 2013). Fisher's exact test with Benjamini–Hochberg correction was used to define statistically significant differences between classes; (***) $P < 0.001$.

the general features of the main subsets of NR1H4-bound CRMs, we further grouped the nodes issued from the SOM into seven classes (labeled from A to G) using hierarchical clustering based on the representative TR combination of each node (Fig. 1B; Supplemental Fig. S2D). We checked that the obtained CRM classification did not result from the clustering of TRs analyzed within the same study and therefore that preferential colocalization of TRs from the same data set can be ruled out as a major confounding effect (Supplemental Fig. S3). We then plotted the average number of TRs recruited to these CRMs (Fig. 1C) together with the average levels of DNase I hypersensitivity (DHS) (Fig. 1D) and histone H3 lysine 9 and 27 acetylation (H3K9ac and H3K27ac) used as chromatin markers of active CRMs (Fig. 1E,F). We observed that CRMs from classes E and G, and to a lesser extent those from classes D and F, showed hallmarks of active CRMs bound by multiple TRs, i.e., strong DHS and histone acethylation levels (Fig. 1C–F). In line with this, most of the CRMs from classes D to G were successfully ascribed target genes using a model correlating cross-tissue CRM activities based on histone acetylation to the transcriptional expression of surrounding genes (Fig. 1G; O'Connor and Bailey 2014). Moreover and in line with these data, CRMs from classes D to G were significantly enriched in the vicinity of genes dysregulated in the liver of *Nr1h4* liver-specific knock-out (KO) mice (Fig. 1H). Hence, we focused further analyses on CRMs defining fully active transcriptional regulatory elements from classes D to G (see Supplemental Results 1; Supplemental Figs. S4, S5 for additional details).

NR1H4-bound CRMs with distinct TR compositions are associated with regulation of cellular housekeeping and liver-specific functions

The TR recruitment pattern characterizing each CRM class (D to G) was defined by multidimensional scaling (MDS) analyses based on the frequency of co-occurrence of all TRs relative to NR1H4 and to one another. This approach clearly indicated that a subset of TRs preferentially co-occurred with NR1H4 for each CRM class (Supplemental Figs. S6–S9). Therefore, we focused our analyses on TRs which were the most strongly interconnected with NR1H4 (Tanimoto index >0.7) (Fig. 2A–D). Classes D and F accommodated fewer TRs (Fig. 2A–D), which were all found in classes E and G, respectively (Fig. 2E). Together with data from Figure 1 and Supplemental Results 2 (Supplemental Fig. S10), this indicates that CRMs from classes E and G are variants of CRMs from classes D and F, respectively, characterized by stronger activity and additional TR binding complexity. Hence, CRMs from classes D and E (hereafter called CRMs$^{D\text{-}E}$), on the one hand, and from classes F and G (hereafter called CRMs$^{F\text{-}G}$), on the other hand, were grouped together for subsequent analyses. Comparing TRs recruited at CRMs$^{D\text{-}E}$ with those recruited at CRMs$^{F\text{-}G}$ revealed a set of common factors (Fig. 2E,F) (TRs depicted in black and hereafter called TRs^{shared}) comprising NRs including HNF4A, PPARA, NR1D2 (also known as REV-ERB beta), and the NR1H4 dimerization partner RXR, as well as PKNOX1 (also known as PREP1), CEBPB, and GATA4. TRs^{shared} also include RNA polymerase II, cofactors such as CREBBP (also known as CBP) and EP300 (also known as P300), and members of the cohesin complex, all known to be broadly associated with active CRMs (Fig. 2E,F). TRs preferentially found at CRMs$^{D\text{-}E}$ or CRMs$^{F\text{-}G}$ also emerged from these data (TRs depicted in blue or violet and hereafter called TRs$^{D\text{-}E}$ and TRs$^{F\text{-}G}$, respectively). This included NR family members [NR5A2 and RARA for TRs$^{D\text{-}E}$ and NR3C1 (also known as GR), NR1D1 (also known as REV-ERBalpha), and RORA for TRs$^{F\text{-}G}$] as well as E2F and ETS family members (E2F4 and GABPA) for TRs$^{D\text{-}E}$ and the corepressor NCOR1, the FOXA family members FOXA1 and FOXA2, the PAR-bZIP factor NFIL3 (also known as E4BP4), and the homeobox HNF1A for TRs$^{F\text{-}G}$ (Fig. 2E,F; Supplemental Fig. S11). Several of these TRs identified as the main factors interconnected with NR1H4 through our integrative cistromic approach were recovered from analysis of NR1H4-bound complexes in the mouse liver using rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) (Fig. 2G; Supplemental Table S2; Mohammed et al. 2016), indicating that at least a fraction of the TRs highlighted by our analyses directly coregulates transcription with NR1H4.

We next sought to define whether differential TR recruitment to CRMs$^{D\text{-}E}$ and CRMs$^{F\text{-}G}$ could contribute to regulation of different biological pathways in the liver. With this aim, we first interrogated whether genes encoding TRs$^{D\text{-}E}$ or TRs$^{F\text{-}G}$ were genetically linked to distinct mouse phenotypes (MPs) using the mammalian phenotype ontology database (Smith et al. 2005). While both were linked to MPs related to liver morphology/functions, TRs$^{F\text{-}G}$ were more specifically linked to altered metabolic homeostasis, while TRs$^{D\text{-}E}$ were more specifically associated with developmental defects (Fig. 2H). In line with this, striking differences between CRMs$^{D\text{-}E}$ and CRMs$^{F\text{-}G}$ were also found when genes assigned to those CRMs as described previously (Fig. 1G) were used to perform gene ontology (GO) enrichment analyses. Indeed, genes linked to CRMs$^{D\text{-}E}$ were enriched for cellular housekeeping/maintenance functions, while those linked to CRMs$^{F\text{-}G}$ were mainly identified as involved in energy metabolism, detoxification, and coagulation (Fig. 2I; Supplemental Fig. S12; see Supplemental Results 3 and Supplemental Fig. S13 for specific gene examples). Moreover, while genes linked to CRMs$^{D\text{-}E}$ were similarly expressed over a wide range of mouse tissues, those linked to CRMs$^{F\text{-}G}$ exhibited preferential expression in the liver (Fig. 2J; Supplemental Fig. S12). This was also linked to greater changes in expression of genes linked to CRMs$^{F\text{-}G}$ in the liver of *Nr1h4* KO mice (Supplemental Fig. S14).

CRMs$^{D\text{-}E}$ and CRMs$^{F\text{-}G}$ show coordinated differences in their pattern of activities across tissues and genomic organization

Expression profiles of genes linked to CRMs$^{D\text{-}E}$ and CRMs$^{F\text{-}G}$ (Fig. 2J) led us to investigate whether these CRMs exhibited differential activation status across tissues using DHS. In line with expression data from Figure 2J, we found that CRMs$^{D\text{-}E}$ were identified as ubiquitous DHS, while CRMs$^{F\text{-}G}$ were all identified as DHS only in the liver (Fig. 3A). Analysis of H3K4 methylation levels showed that CRMs$^{D\text{-}E}$ exhibited preferential enrichment for H3K4me3 over H3K4me1 (Fig. 3B,C), an epigenetic pattern associated with promoters (Heintzman et al. 2007; Lupien et al. 2008). Indeed, a comparison with GENCODE transcriptional start sites (TSSs) indicated that CRMs$^{D\text{-}E}$ almost exclusively (89%) correspond to promoter-proximal CRMs (hereafter called CRMs$^{D\text{-}E}$ promoters) (Fig. 3D). Conversely, preferential enrichment for H3K4me1 over H3K4me3 was consistent with CRMs$^{F\text{-}G}$ mostly comprising enhancers (70%) (Fig. 3B–D). Moreover, CRMs$^{F\text{-}G}$ form clusters along the genome since they were significantly associated with blocks of regulation defined as genomic regions comprising active CRMs marked with H3K27ac within 12.5 kb of one another (Fig. 3E; Whyte et al. 2013). Among those regulatory blocks, 22% comprised NR1H4-bound CRMs both at gene promoters and enhancers, as exemplified by the *Nr0b2* or *Fmo3* genes (Supplemental Fig. S13G,H).

Transcriptional regulation is spatially constrained within topologically associating domains (TADs), whose borders are mostly

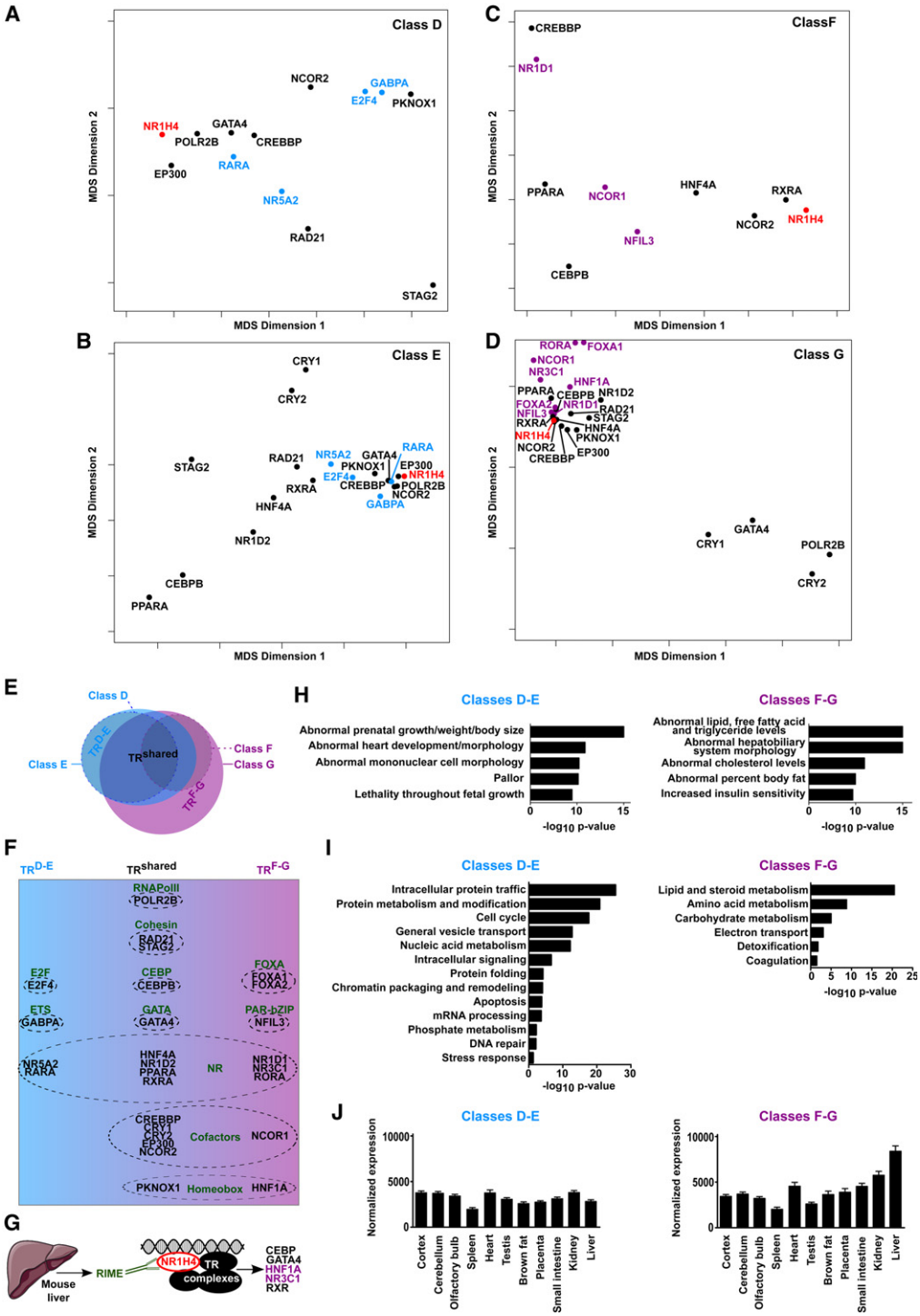

Figure 2. NR1H4-bound CRMs comprise two main classes which relate to the regulation of different gene sets and biological functions. (*A–D*) Multidimensional scaling (MDS) was performed as described in the Methods section using CRM from classes D, E, F, or G as indicated. These panels show TRs which are the most strongly interconnected with NR1H4 in each class (Tanimoto index > 0.7). NR1H4 is depicted in red while TRs^{D-E} and TRs^{F-G} specifically found in $CRMs^{D-E}$ or $CRMs^{F-G}$ are depicted in blue or violet, respectively. TRs^{shared} are depicted in black and were found both in $CRMs^{D-E}$ and $CRMs^{F-G}$. (*E*) Venn diagram summarizing the overlaps between TRs found at CRMs from classes D to G in panels *A–D*. (*F*) Overview of TRs comprising the TR^{shared}, TR^{D-E}, and TR^{F-G} subsets. TRs were grouped according to their function or affiliation to larger families, which are indicated in bold. (*G*) TRs highlighted in Figure 2F, which could be identified in complexes with NR1H4 in RIME experiments, are indicated. As expected, NR1H4 was retrieved from these analyses but is not reported. (*H*) The main TRs found at $CRMs^{D-E}$ or $CRMs^{F-G}$ from panels *A–D* were used in ToppCluster to identify associated mouse phenotypes (Kaimal et al. 2010). Bonferroni-corrected *P*-values ($-\log_{10}$) are shown. (*I*) Gene ontology (GO) enrichment analyses were performed using DAVID (Huang et al. 2009) and genes associated with $CRMs^{D-E}$ or $CRMs^{F-G}$. Bonferroni-corrected *P*-values ($-\log_{10}$) are shown. (*J*) Average normalized mRNA expression levels of genes associated with $CRMs^{D-E}$ or $CRMs^{F-G}$ across indicated mouse tissues were obtained using BioGPS data (Wu et al. 2009). Results are means ± S.E.M.

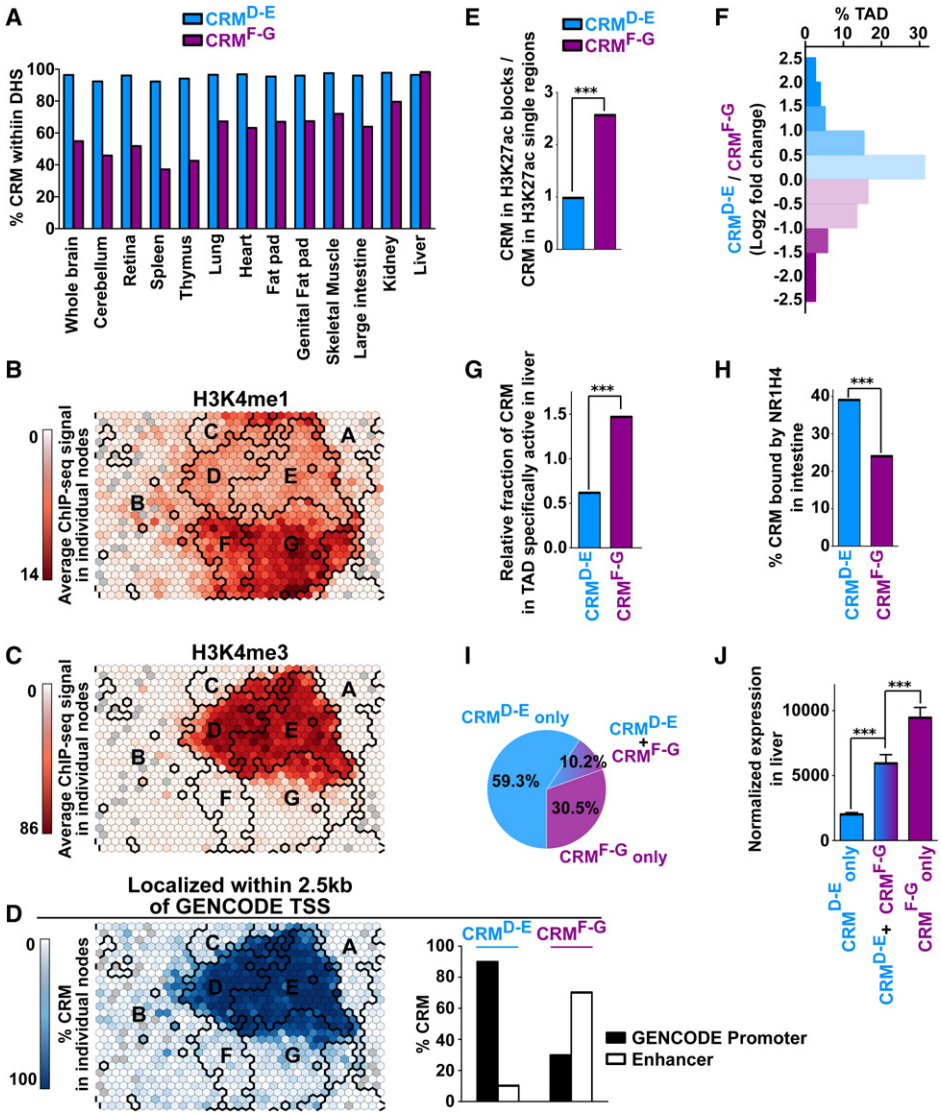

Figure 3. CRM from classes D-E and F-G differ in their identity, activity across tissues, and genomic distribution. (*A*) CRMs^{D-E} or CRMs^{F-G} were intersected with DHS sites identified in the indicated mouse tissues by the ENCODE Consortium (Vierstra et al. 2014). (*B–D*) The map issued from Figure 1B was used to indicate the H3K4me1 (*B*) or H3K4me3 (*C*) ChIP-seq levels as well as the percentage of CRMs localized within 2.5 kb of a GENCODE TSS (*D*) in each node. Bold black lines indicate the borders of the clusters. The bar graph in *D* summarizes the percentage of CRMs^{D-E} and CRMs^{F-G} labeled as promoters or enhancers. (*E*) Active CRMs were defined as enriched for H3K27ac in the mouse liver genome using data from Yue et al. (2014) and were grouped into blocks when separated by less than 12.5 kb. The bar graph shows the number of CRMs found into clusters, i.e., comprised within the aforementioned blocks, relative to those found outside clusters, i.e., single regions. Fisher's exact test was used to define statistically significant differences between CRMs^{D-E} and CRMs^{F-G}; (***) $P < 0.001$. (*F*) CRMs^{D-E} and CRMs^{F-G} found in active mouse TADs identified in Zhao et al. (2013) were counted and normalized to the respective total number of CRMs. The bar plot shows the frequency distribution of TADs with a different ratio of CRMs^{D-E} relative to CRMs^{F-G}. (*G*) A similar analysis was performed using TADs specifically active in the mouse liver (Zhao et al. 2013), and results were plotted as a bar graph. Fisher's exact test was used to define statistically significant differences between CRMs^{D-E} and CRMs^{F-G}; (***) $P < 0.001$. (*H*) CRMs^{D-E} and CRMs^{F-G} also identified as NR1H4 binding sites in the mouse liver in Thomas et al. (2010) were compared to intestine NR1H4 binding sites from the same study. Fisher's exact test was used to define statistically significant differences between CRMs^{D-E} and CRMs^{F-G}; (***) $P < 0.001$. (*I*) Percentage of genes uniquely associated with CRMs^{D-E} (CRMs^{D-E} only) or CRMs^{F-G} (CRMs^{F-G} only) or associated with both (CRMs^{D-E} + CRMs^{F-G}). (*J*) Average-normalized mRNA expression levels of genes uniquely associated with CRMs^{D-E} (CRMs^{D-E} only) or CRMs^{F-G} (CRMs^{F-G} only) or associated with both (CRMs^{D-E} + CRMs^{F-G}) in the mouse liver was obtained using BioGPS data (Wu et al. 2009). Results are means ± S.E.M. One-way ANOVA with Bonferroni's multiple comparison test was used to define statistically significant differences; (***) $P < 0.001$.

invariant across cell types (Dixon et al. 2012). More than one third of active mouse TADs defined in Zhao et al. (2013) showed greater than twofold differences in enrichment for CRMs^{D-E} relative to CRMs^{F-G} (Fig. 3F). Moreover, the presence of CRMs^{F-G} was significantly enriched within TADs specifically active in the liver (Fig. 3G), and NR1H4 binding to these CRMs was less conserved across tissues as revealed by comparison with NR1H4 binding sites from the mouse intestine (Fig. 3H). This points to a genomic compartment level organization segregating CRMs^{D-E} from CRMs^{F-G}. Nevertheless, this compartmentalization is not strict, and we found that a limited subset of genes were associated with both CRMs^{D-E} and CRMs^{F-G} (almost exclusively CRM^{D-E} promoters

together with $CRM^{F\text{-}G}$ enhancers) (Fig. 3I; Supplemental Table S3). Interestingly, this subset of genes comprise housekeeping genes with higher expression levels in the mouse liver compared to genes uniquely associated with $CRMs^{D\text{-}E}$ (Fig. 3J; Supplemental Fig. S15).

Altered transcriptome in the liver of TR knock-out mice relates to their involvement in distinct combinatorial *trans*-regulatory modules at NR1H4-bound promoters and enhancers

Since $CRMs^{F\text{-}G}$ comprised both enhancers and promoters, we considered the possibility that part of the complexity of co-occurring TRs was masked in our previous analyses. Therefore, we plotted the occurrence of each individual TR at promoters versus enhancers from $CRMs^{F\text{-}G}$. We observed that TRs^{shared} and $TRs^{F\text{-}G}$ were the most frequently found factors both at promoters and enhancers (Fig. 4A). Hence, we defined TR^{shared} as a core *trans*-regulatory module (TRM) (shared between $CRM^{D\text{-}E}$ and both $CRM^{F\text{-}G}$ promoters and enhancers) and $TR^{F\text{-}G}$ as a liver-specific functions control TRM (obligatory module specific to $CRMs^{F\text{-}G}$). Additionally, a large fraction of the NR1H4-bound $CRM^{F\text{-}G}$ promoters were also characterized by the presence of $TR^{D\text{-}E}$ which was found to characterize $CRM^{D\text{-}E}$ in Figure 2 (Fig. 4A,B; Supplemental Fig. S16; also cf. E2F4, GABPA, NR5A2, and RARA profiles in Supplemental Fig. S11 and Fig. 3B–D). Preferential recruitment to promoters of the $TR^{D\text{-}E}$ E2F4 and GABPA was linked to a greater enrichment of their binding motifs compared to enhancers, in sharp contrast with motifs recognized by the $TR^{F\text{-}G}$ FOXA1-2, NFIL3, and HNF1A (Fig. 4C). A direct connection between motif enrichment and recruitment was not evidenced for other factors including NR (Fig. 4C; Supplemental Fig. S17A). Altogether, these results indicate that $TR^{D\text{-}E}$ actually represents a specific TRM operating at NR1H4-bound promoters.

To functionally validate these findings, we interrogated the liver transcriptome from TR knock-out mice (Supplemental Table S1). In line with the predictions made from the cistromic analyses, dysregulated genes in the liver of *Ncor1* and *Hnf1a* KO mice preferentially belonged to the subset of NR1H4-regulated genes associated with $CRMs^{F\text{-}G}$ (Fig. 4D; Supplemental Fig. S18). Conversely, dysregulated genes in the liver of *Ppara* and *Nr5a2* KO mice similarly fell within the subsets of NR1H4-regulated genes linked to $CRMs^{D\text{-}E}$ and $CRMs^{F\text{-}G}$ (Fig. 4D; Supplemental Fig. S18). Altogether, these data verify that TRs from different TRMs are differentially involved in the regulation of genes linked to $CRMs^{D\text{-}E}$ and $CRMs^{F\text{-}G}$ and therefore provide functional support for the different TRMs identified by our cistromic analyses.

Hierarchical combinations of obligatory and facultative TRMs further specify activities of different subsets of NR1H4-bound CRMs

CRMs from class G ($CRMs^{G}$) are the most densely bound sites (Fig. 1C). Therefore, we focused on this class of CRMs to test the hypothesis that additional TRMs may characterize and specify the activities of a limited subset of CRMs. With this aim, we analyzed the co-occurrence of each pair of TRs at $CRMs^{G}$ and subsequently organized the TRs based on hierarchical clustering (Fig. 5A). Further supporting our previous findings, this analysis was able to retrieve a main cluster largely composed of TRs^{shared}/core TRMs and $TRs^{F\text{-}G}$/liver-specific functions control TRMs as well as a cluster corresponding to $TR^{D\text{-}E}$/promoter TRMs (Fig. 5A). Interestingly, an additional cluster comprised of the core circadian TRs [ARNTL (also known as BMAL1), CLOCK, PER1, PER2, CRY1, CRY2, and NPAS2] (Zhang and Kay 2010) was evidenced as binding to a substantial fraction of both CRM^{G} promoters and enhancers (Fig. 5A). In addition to a lack of co-occurrence of $TRs^{F\text{-}G}$, clustering of these circadian TRs was not as obvious at CRMs from class E (Supplemental Fig. S19).

To better define how the newly discovered circadian TRMs distribute at $CRMs^{G}$ relative to other TRs, we next monitored the occurrence of TRs in $CRMs^{G}$ after they were divided into CRMs devoid of (bound by 0–2 circadian TRs; hereafter denoted as CRM^{G} -circadian TRMs) or bound by the circadian TRMs (bound by all seven core circadian TRs; hereafter denoted as CRM^{G} + circadian TRMs). The latter showed relative enrichment for E-box motifs (BMAL1_MOUSE.H10MO.C; Bonferroni-corrected *P*-value = 3.36×10^{-72}), which mediate core circadian transcription factor (ARNTL, CLOCK, and NPAS2) recruitment (Supplemental Fig. S20). Promoters and enhancers were separately considered in these analyses since module combinations differ at these CRMs. These analyses revealed no large differences in TR occurrence at promoters and enhancers with or without the circadian TRM (Fig. 5B,C). Therefore, the circadian TRM does not represent an alternative module at a subset of CRMs but rather a facultative module which adds to the obligatory core and liver-specific functions control TRM as well as the promoter TRM at selective enhancers and promoters.

In line with these observations, we found that a greater fraction of genes linked to CRM^{G} + circadian TRMs was dysregulated in the liver of *Per2* KO mice (Fig. 5D). We hypothesized that these CRMs would also show greater association with circadian gene regulation. Indeed, we found that genes linked to CRM^{G} + circadian TRMs had a significantly greater chance to display circadian transcription in the mouse liver (Fig. 5E). In line with this, CRM^{G} + circadian TRMs labeled as enhancers had a significantly greater chance to display circadian eRNA transcription and therefore circadian activity (Fig. 5F; Fang et al. 2014). Altogether, these data indicate that the presence of the circadian TRM allows reinforcing circadian regulation of a specific subset of NR1H4 target genes.

Finally, in order to lend independent support to our conclusions, we used intra-genomic replicate (IGR) analyses to monitor how changes in chromatin recruitment are coordinated among TRs in the mouse liver. The IGR tool predicts the impact of single nucleotide variants (SNVs) on TR chromatin binding (Cowper-Sal·lari et al. 2012; Bailey et al. 2016). IGR analyses were performed to define the effect of SNVs localized within $CRMs^{G}$ on binding of all TRs, and results were subsequently mined using either principal component analyses (PCA) or hierarchical clustering (Fig. 5G; Supplemental Fig. S21, respectively). We found that TRs could be arranged into the same three main groups described in Figure 5A (Fig. 5G; Supplemental Fig. S21). A subcluster containing most $TRs^{F\text{-}G}$ was also evidenced by these analyses (Supplemental Fig. S21). To rule out a confounding effect of TRs preferential cobinding on these results, IGR data were further analyzed by using pairwise comparisons to define how frequently a given SNV impacts on the binding of two TRs by strictly focusing on CRMs where these two TRs are corecruited. Importantly, hierarchical clustering of these data produced similar results (Fig. 5H). Altogether, these analyses therefore indicate that TRMs previously identified are independently evidenced by selective and coordinated modulation of TR chromatin recruitment.

Discussion

Our study, leveraging combinations of integrative cistromic, epigenomic, transcriptomic, and interactomic analyses, allowed the

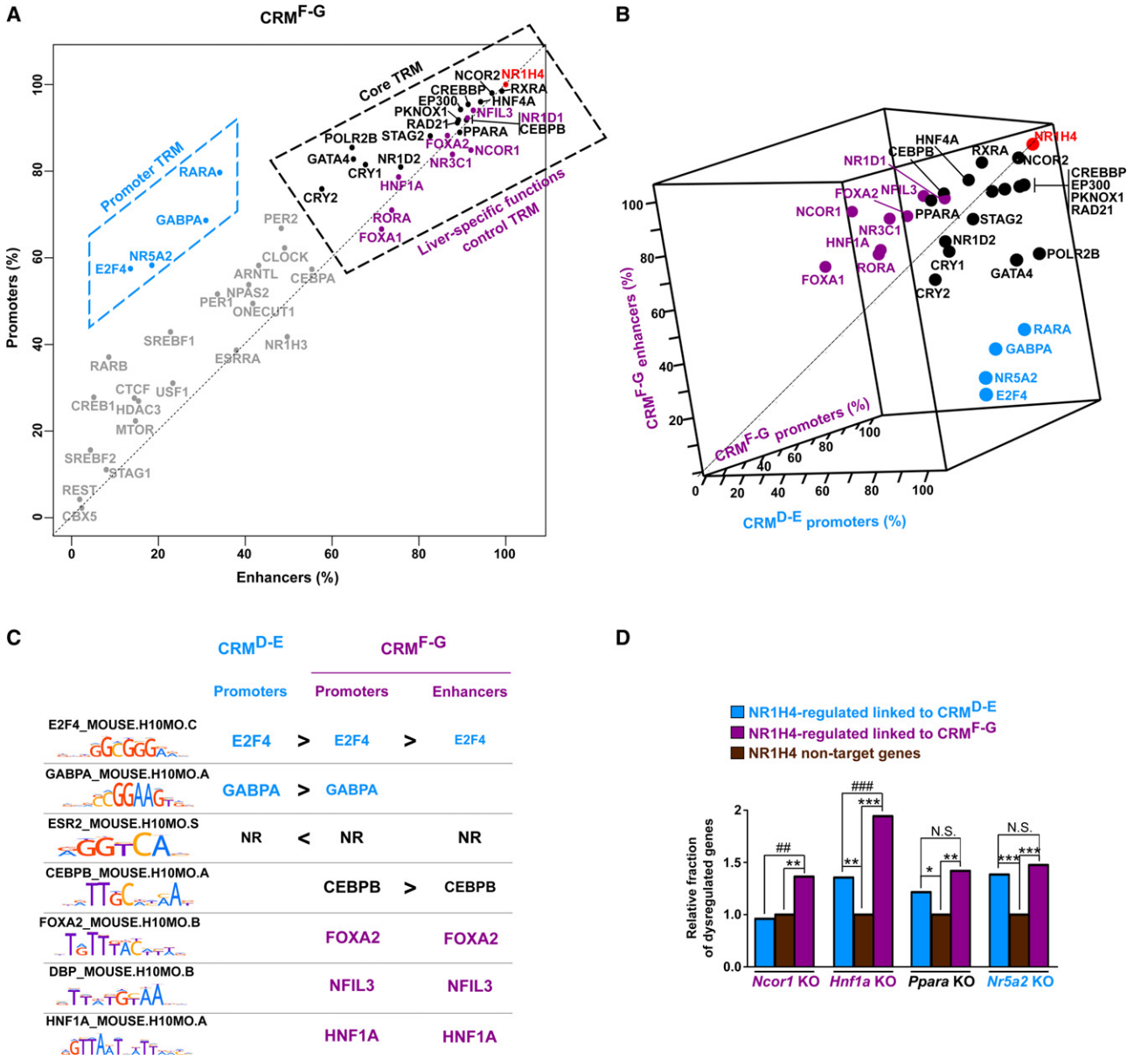

Figure 4. Main TRMs occurring at NR1H4-bound CRMs and functional validation of their role using dysregulated expression in the liver of mice KO for representative TRs. (*A*) Plot showing the occurrence of each TR at promoters versus enhancers from CRMs$^{F\text{-}G}$. TRs^{shared}, TRs$^{D\text{-}E}$, and TRs$^{F\text{-}G}$ are depicted in black, blue, and violet according to Figure 2. Other TRs were displayed in gray. TRs^{shared} and TRs$^{F\text{-}G}$ defining the core and liver-specific functions control TRMs on one hand and TRs$^{D\text{-}E}$ defining the promoter TRMs on the other hand are highlighted into dashed boxes. (*B*) Three-dimensional plot showing the occurrence of each TR at CRM$^{D\text{-}E}$ and CRM$^{F\text{-}G}$ promoters as well as CRM$^{F\text{-}G}$ enhancers. TRs^{shared}, TRs$^{D\text{-}E}$ and TRs$^{F\text{-}G}$ are depicted in black, blue, and violet according to Figure 2. (*C*) DNA binding motifs enriched in CRM$^{D\text{-}E}$ and CRM$^{F\text{-}G}$ promoters and in CRM$^{F\text{-}G}$ enhancers (defined using regions from class A as controls) are indicated using the name of the recognizing transcription factor. Moreover, < and > were used to indicate significant differential enrichment within distinct sets of CRMs. (*D*) Fraction of NR1H4 target genes dysregulated in the liver of mice KO for the indicated TR. Genes exclusively associated with CRMs$^{D\text{-}E}$ or CRMs$^{F\text{-}G}$ and whose expression is modified in the liver of *Nr1h4* KO mice were used for these analyses. Genes which are not linked to NR1H4-bound CRMs and whose expression is not altered in the liver of *Nr1h4* KO mice (NR1H4 nontarget genes) served as the reference (arbitrarily set to 1). Fisher's exact test with Benjamini–Hochberg correction was used to define statistically significant differences with NR1H4 nontarget genes ([*] $P < 0.05$, [**] $P < 0.01$, and [***] $P < 0.001$) or between NR1H4-regulated genes linked to CRMs$^{D\text{-}E}$ and CRMs$^{F\text{-}G}$ ([##] $P < 0.01$, [###] $P < 0.001$, [N.S.] not significant).

revelation of the transcriptional regulatory logic underlying NR1H4 activities in the liver. A central finding of our study is that TRs co-occurring at NR1H4-bound CRMs are organized into obligatory and facultative TRMs which work in a combinatorial manner to define distinct subsets of CRMs and target genes (Fig. 6). Importantly, transcriptomic data of liver from TR KO mice validated the TRM organization identified by the cistromic analyses and point to additive transcriptional regulatory inputs by the different TRMs. This organization into modules that are differentially and hierarchically found at CRMs transcends both (1) the definition of liver-identity TRs based on privileged labeling of their encoding gene with broad H3K4me3 in the liver (Supplemental

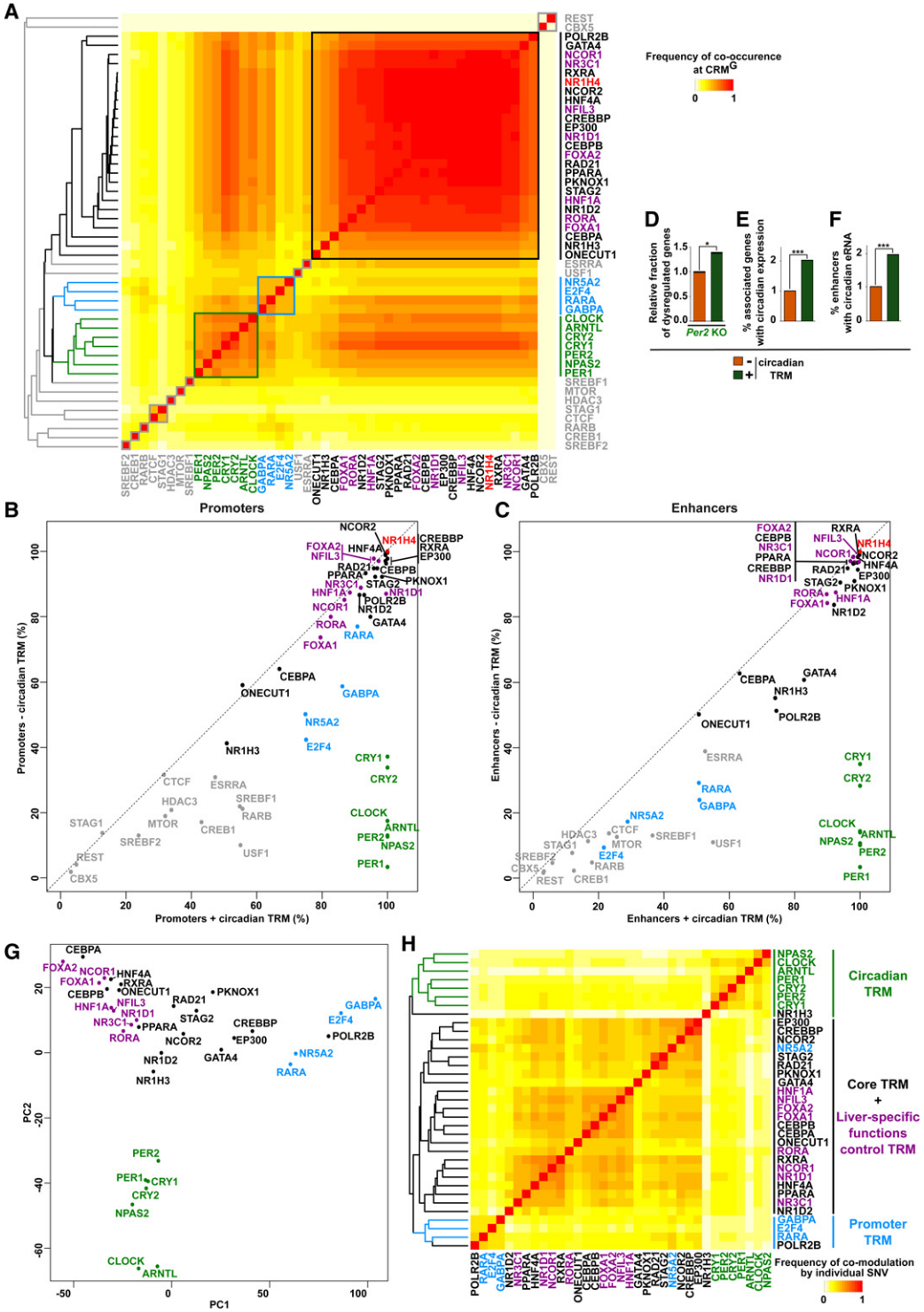

Figure 5. Identification and functional validation of hierarchical combinations of TRMs at NR1H4-bound CRMs. (*A*) Heat map showing TR co-occurrence at CRMs^G defined using a Tanimoto index. TRs were organized based on hierarchical clustering, and main clusters were framed. The hierarchical clustering tree is shown on the *left*. (*B*,*C*) Plots showing the occurrence of each TR at CRM^G promoters (*B*) or enhancers (*C*) characterized by the presence of 0–2 (−circadian TRMs) or all seven core circadian TRs (+ circadian TRMs). (*D*) Analyses were performed as in Figure 4D using transcriptomic data from the liver of *Per2* KO mice. (*E*) The fraction of genes exclusively associated with CRM^G −/+ circadian TRMs displaying circadian expression in the mouse liver was defined using genes with circadian transcription identified using global run-on sequencing (GRO-seq) (Fang et al. 2014). Fisher's exact test was used to define statistically significant differences between CRM^G −/+ circadian TRMs; (***) $P < 0.001$. (*F*) The fraction of CRM^G enhancers −/+ circadian TRMs displaying circadian eRNA transcription in the mouse liver was defined using data from Fang et al. (2014). Fisher's exact test was used to define statistically significant differences between CRM^G enhancers −/+ circadian TRMs; (***) $P < 0.001$. (*G*) The IGR tool was used to predict the impact of 63,968 SNVs on binding to CRMs^G of the indicated TRs, and data were then mined using PCA. Fold change was set to 0 when the modulatory effect of a SNV did not reach statistical significance (Benjamini–Hochberg corrected P-value > 0.05) or when it relates to weak TR binding (i.e., binding not called by MACS2 in our previous analyses). (*H*) The IGR tool was used to predict the impact of SNVs localized within CRMs^G on chromatin binding of the indicated TRs, and pairwise comparisons were subsequently performed. Only SNVs localized within CRMs^G corecruiting the two TRs and significantly modulating the binding of one of these two TRs (Benjamini–Hochberg corrected P-value < 0.05) were considered. The frequency of comodulation by individual SNPs was calculated using a Tanimoto index. The hierarchical clustering tree is shown on the *left*.

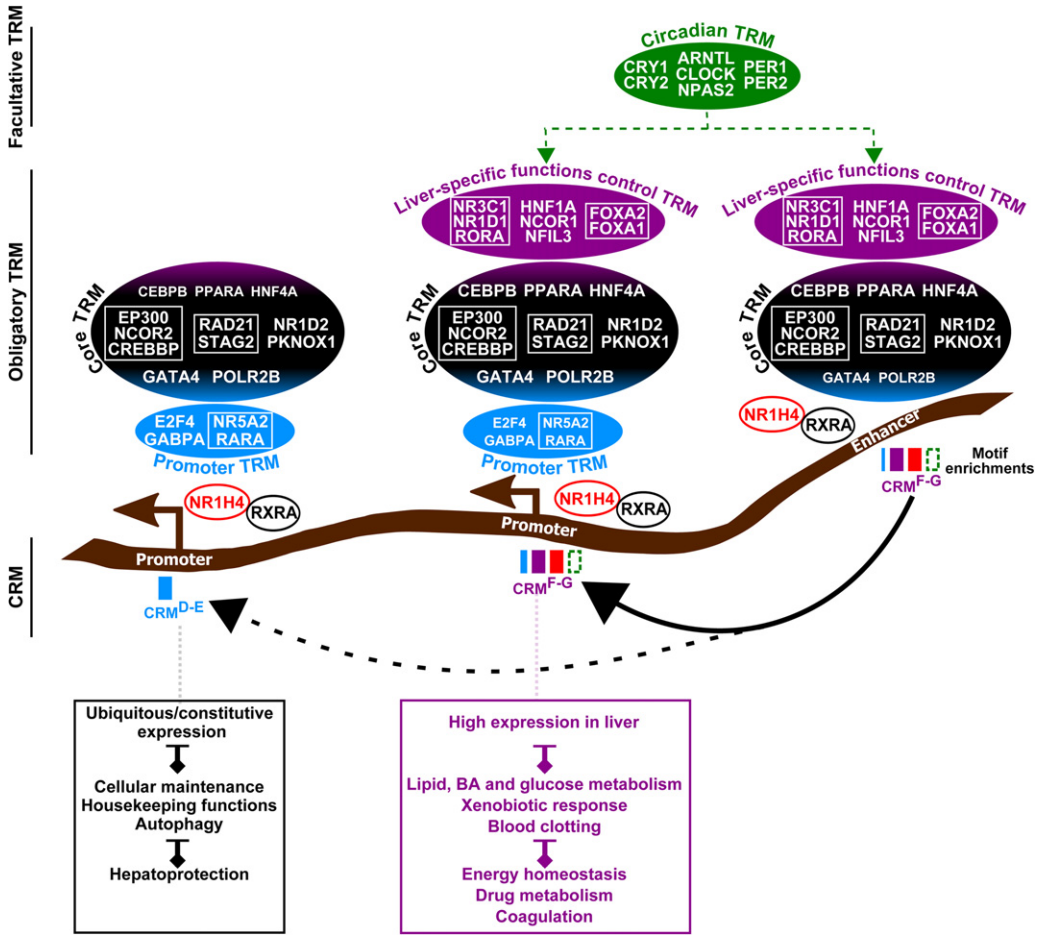

Figure 6. Hierarchical and combinatorial TRM recruitment discriminate NR1H4-bound promoters and enhancers involved in control of cellular maintenance and liver-specific functions. The transcriptional regulatory logic defined in this study is summarized. The size of the TR and TRM *above* the chromatin and of the boxes *below* the chromatin correlates with the occurrence of binding and DNA motif at the distinct classes of CRM, respectively. Refer to Discussion for details.

Fig. S22), a recently discovered feature of cell-identity genes (Benayoun et al. 2014; Chen et al. 2015), and (2) the previously documented hierarchical activities of selective TRs including pioneer factors (CEBP, FOXA, GATA4) (Magnani et al. 2011; Zaret and Carroll 2011), since TR from these groups are present in the different TRMs. Hence, our results reveal novel principles of concerted transcriptional regulation by multiple TRs at CRMs.

Our study points to intrinsic differences in transcriptional regulation of housekeeping and liver-specific genes by NR1H4, the former being mainly promoter-based while the latter involves recruitment to both promoters and distal enhancers (Fig. 6). This is in line with conclusions from studies conducted in other biological systems (Ernst et al. 2011; Tong et al. 2016) and is further supported by the observation that $CRM^{D\text{-}E}$ promoters are half as connected to distal enhancers when compared to $CRM^{F\text{-}G}$ promoters (Fisher's exact test $P = 1.7 \times 10^{-68}$). Nevertheless, a subset of housekeeping genes displaying stronger expression levels in the liver was linked to regulation by distal $CRM^{F\text{-}G}$ enhancers. This is reminiscent of our previous findings, which indicated that the cell-type–specific regulator PPARG uses distal enhancers to modulate housekeeping gene expression during adipocyte differentiation (Oger et al. 2014). Moreover, our data are also consistent with a recent study showing that, in addition to being critical for transcription of cell-type–specific genes, tissue-specific distal enhancers also play an additive role regarding expression of a subset of housekeeping genes (Beck et al. 2014). Overall, our findings are consistent with the described hepatic functions of several TRs and suggest that selected TRs, including NR1H4, may coordinately serve as a nexus for concerted regulation of housekeeping/cellular maintenance genes and liver-specific metabolic functions (see Supplemental Discussion for details).

Recent studies indicate that promoters and enhancers actually share similar genomic architectures and unifying functionalities (Kim and Shiekhattar 2015; Nguyen et al. 2016). In this context, promoter-proximal CRMs could behave as TSS-proximal enhancers, allowing for autonomous transcription of housekeeping genes (Arnold et al. 2016). Alternatively, these CRMs could bear both enhancer and promoter activities thanks to the specific presence of selective TRs conferring strong promoter activities to CRMs (Nguyen et al. 2016). This is consistent with our findings that promoters and enhancers share similar TRMs, except for the additional presence of a specific set of TRs (E2F4, GABPA, NR5A2, and RARA) at the former ones.

At least a fraction of the TRs we have identified pertains to chromatin-bound NR1H4 complexes in the liver, and IGR data further indicate that TRMs represent functional units with

coordinated recruitment to chromatin. This may rely on several mechanisms involving sharing of DNA motifs within a family, such as NRs or hierarchical or reciprocally facilitated chromatin recruitment (Magnani et al. 2011; Zaret and Carroll 2011; Madsen et al. 2014), as well as formation of protein complexes which are important drivers of TR genomic colocalization (Xie et al. 2013; Liu et al. 2014). In this context, the core TRM comprises general cofactors indirectly recruited by CRM-bound transcription factors (Fig. 6), most probably including $TR^{D\text{-}E}$ and $TR^{F\text{-}G}$ (Fig. 5A,G,H; Supplemental Figs. S19, S21). TR tethering may also mediate recruitment of NR1H4 and several TRs^{shared}, such as CEBPB, HNF4A, and PPARA to $CRM^{D\text{-}E}$ promoters since these CRMs lack enrichment for their DNA recognition motifs. This is concomitant with a lower binding intensity of these TRs at $CRM^{D\text{-}E}$ promoters (Supplemental Fig. S17B). Alternatively, binding through yet uncharacterized recognition motifs cannot entirely be ruled out (Neph et al. 2012; Jolma et al. 2015). Interestingly, $CRM^{D\text{-}E}$ promoters show a greater coverage by CpG islands (CGI) (Supplemental Fig. S17C), which has been described as a distinctive feature of promoters harboring transcriptionally permissive chromatin states and which could provide specific regulatory mechanisms to housekeeping genes (Deaton and Bird 2011; Beck et al. 2014).

While our study has allowed us to draw general transcriptional regulatory principles using genomic-level analyses, many subtle variations of co-occuring TRs exist at NR1H4-bound CRMs, and whether and how these more subtle variations impact target gene regulation by the identified TRMs remains to be defined. This focused level of analysis may, however, be more impacted by heterogeneity of the original data sets including differences in mouse feeding and handling as well as ChIP efficiency and potential confounding effects due to usage of independent data sets. Additionally, further studies are required to define the precise functional outputs of the connection between TRs of the different TRMs and NR1H4 (see Supplemental Results 4; Supplemental Fig. S23 for additional details).

While focusing on NR1H4-bound CRMs allowed us to leverage its known functions and target genes in the mouse liver to validate and interpret our findings, we have further shown that the identified hierarchical combinations of TRMs extend to a broader CRM landscape (Supplemental Results 5; Supplemental Fig. S24). This indicates that this level of organization revealed by our study is most probably a general principle operating at CRMs.

Methods

Public functional genomics data recovery and TR ChIP-seq data processing

Public functional genomics data used in this study were downloaded from public databases and are listed in Supplemental Table S1. All raw data were processed similarly including FastQC analysis (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), read mapping to mm10 using Bowtie (version 1.0.0) (Langmead et al. 2009), peak calling using model-based analysis of ChIP-seq version 2 (MACS2) (Zhang et al. 2008), and visual inspection of called peaks using the Integrated Genome Browser (IGB) (Nicol et al. 2009). Replicates were managed using the irreproducibility discovery rate (IDR) (Li et al. 2011), and false positive calls repeatedly identified in inputs and IgG ChIP-seq were removed from all data sets. CRMs used for the SOM analysis were identified as genomic regions with co-occurring recruitment of at least two different TRs. Details and parameters which were used are provided in the Supplemental Material.

Self-organizing maps analyses

The self-organizing maps (SOMs) were generated using the R package "kohonen2" (Wehrens and Buydens 2007). The input vectors from CRMs, optimal number of nodes, and parameters to train the SOMs were defined according to Xie et al. (2013) and are detailed in the Supplemental Material. Nodes were further grouped into classes based on hierarchical clustering performed using the hclust function of the R package "Stats" (R Core Team 2015). A planar projection of the toroidal map was used for data visualization.

Multidimensional scaling and hierarchical clustering analyses of TR co-occurrence

TR co-occurence at CRMs from classes D, E, F, or G was used to calculate Tanimoto distance matrices, which were used for MDS and hierarchical clustering using the R packages "Stats" (R Core Team 2015) and "gplots" (https://cran.r-project.org/web/packages/gplots/), respectively. Details are provided in the Supplemental Material.

Gene ontology, mouse phenotype, and gene set enrichment analyses

GO enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7) (Huang et al. 2009). ToppCluster was used to link TRs to MPs (Kaimal et al. 2010). Gene set enrichment analysis (GSEA) was performed using the GSEA software developed at the Broad Institute (Subramanian et al. 2005). Details for used parameters are provided in the Supplemental Material.

CRM target gene assignment and transcriptomic data analyses

CRM localized within 2.5 kb of a GENCODE gene TSS were assigned to this gene. Target gene assignment for distal CRMs was performed using a model correlating cross-tissue CRM activities based on histone acetylation to gene transcriptional expression (O'Connor and Bailey 2014).

Raw transcriptomic data from Affymetrix microarrays were normalized using the Partek Genomics Suite or the R package "oligo" (Carvalho and Irizarry 2010). The average normalized expression of genes (averaged by Gene Symbol) were then used to perform the differential expression analyses using limma (Smyth 2004; Ritchie et al. 2015). For the *Nr1h4* KO data, a meta-analysis was performed using the metaMA package (Marot et al. 2009). Details are provided in the Supplemental Material.

Intragenomic replicates

The functional impact of SNV on TR binding within mouse liver DHS was predicted using the IGR tool as previously described (Cowper-Sal·lari et al. 2012). Details are provided in the Supplemental Material. PCA and hierarchical clustering were performed using the R packages "FactoMineR" (Le et al. 2008) and "Stats" (R Core Team 2015), respectively.

Transcription factor recognition motif enrichment analyses

NR1H4 binding motif enrichments were determined using CENTDIST (Zhang et al. 2011). Differential transcription factor motif enrichment and motif scanning were performed using the MEME suite (McLeay and Bailey 2010) as detailed in the Supplemental Material.

Animals and liver gene expression analyses

Nr1h4 and *Ppara* KO mice have been described previously (Porez et al. 2013; Berrabah et al. 2014; Pawlak et al. 2015). Animal studies were performed in compliance with European Community specifications regarding the use of laboratory animals and approved by the Nord-Pas de Calais Ethical Committee for animal use.

RNA extraction, reverse transcription (RT), and real-time quantitative PCR (qPCR) were performed as previously described (Dubois-Chevalier et al. 2014). MoGene-2_0-st Affymetrix arrays were used for transcriptomic analyses. Details are provided in the Supplemental Material.

Rapid immunoprecipitation mass spectrometry of endogenous proteins

Livers from wild-type mice were processed for double cross-linking as described in the Supplemental Material and then used for RIME as described in Mohammed et al. (2016). Mass spectrometry was performed by the proteomic core facility at Cancer Research UK. TRs detected in IgG samples were discarded.

Statistical analyses

Statistical analyses were performed using the Prism software (GraphPad) and R (R Core Team 2015). The specific tests and corrections for multiple testing that were used are indicated in the figure legends.

Data access

Raw and processed data sets from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE87566 and GSE87567.

Acknowledgments

The authors thank Dr. D'Santos C. and the proteomic core facility at Cancer Research UK (Cambridge, UK) for processing RIME samples. This work was supported by grants from the Fondation pour la Recherche Médicale (Equipe labellisée, DEQ20150331724), "European Genomic Institute for Diabetes" (E.G.I.D., Agence Nationale de la Recherche, ANR-10-LABX-46), and European Commission. B.S. is a member of the Institut Universitaire de France and is supported by the European Research Council (ERC Grant Immunobile, contract 694717).

References

Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LFA, van Lohuizen M, van Steensel B. 2013. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154:** 914–927.

Arnold C, Zabidi M, Pagani M, Rath M, Schernhuber K, Kazmar T, Stark A. 2016. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat Biotechnol* **35:** 136–144.

Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, Dowar M, Thu KL, Cescon DW, Silvester J, et al. 2016. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nat Genet* **48:** 1260–1266.

Beck S, Lee B, Rhee C, Song J, Woo A, Kim J. 2014. CpG island-mediated global gene regulatory modes in mouse embryonic stem cells. *Nat Commun* **5:** 5490.

Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E, et al. 2014. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158:** 673–688.

Berrabah W, Aumercier P, Gheeraert C, Dehondt H, Bouchaert E, Alexandre J, Ploton M, Mazuy C, Caron S, Tailleux A, et al. 2014. The glucose sensing O-GlcNacylation pathway regulates the nuclear bile acid receptor FXR. *Hepatology* **59:** 2022–2033.

Carvalho BS, Irizarry RA. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26:** 2363–2367.

Chen ZF, Paquette AJ, Anderson DJ. 1998. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet* **20:** 136–142.

Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, Rodriguez B, Xi Y, Xia Z, Chen X, et al. 2015. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet* **47:** 1149–1157.

Chong HK, Biesinger J, Seo Y, Xie X, Osborne TF. 2012. Genome-wide analysis of hepatic LRH-1 reveals a promoter binding preference and suggests a role in regulating genes of lipid metabolism in concert with FXR. *BMC Genomics* **13:** 51.

Cowper-Sal·lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, Moore JH, Lupien M. 2012. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* **44:** 1191–1198.

Deaton A, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* **25:** 1010–1022.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380.

Dubois-Chevalier J, Oger F, Dehondt H, Firmin FF, Gheeraert C, Staels B, Lefebvre P, Eeckhoute J. 2014. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res* **42:** 10943–10959.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49.

Fang B, Everett LJ, Jager J, Briggs E, Armour SM, Feng D, Roy A, Gerhart-Hines Z, Sun Z, Lazar MA. 2014. Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. *Cell* **159:** 1140–1152.

Gomez-Ospina N, Potter CJ, Xiao R, Manickam K, Kim M, Kim KH, Shneider BL, Picarsic JL, Jacobson TA, Zhang J, et al. 2016. Mutations in the nuclear bile acid receptor FXR cause progressive familial intrahepatic cholestasis. *Nat Commun* **7:** 10713.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 311–318.

Huang W, Ma K, Zhang J, Qatanani M, Cuvillier J, Liu J, Dong B, Huang X, Moore DD. 2006. Nuclear receptor-dependent bile acid signaling is required for normal liver regeneration. *Science* **312:** 233–236.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57.

Jolma A, Yin Y, Nitta K, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527:** 384–388.

Kaimal V, Bardes EE, Tabar SC, Jegga AG, Aronow BJ. 2010. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* **38:** W96–W102.

Kim T, Shiekhattar R. 2015. Architectural and functional commonalities between enhancers and promoters. *Cell* **162:** 948–959.

Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, Cheng C, Gerstein M, White KP. 2013. A comprehensive nuclear receptor network for breast cancer cells. *Cell Rep* **3:** 538–551.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Le S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. *J Stat Softw* **25:** 1–18.

Lefebvre P, Cariou B, Lien F, Kuipers F, Staels B. 2009. Role of bile acids and bile acid receptors in metabolic regulation. *Physiol Rev* **89:** 147–191.

Li Q, Brown J, Huang H, Bickel P. 2011. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* **5:** 1752–1779.

Lien F, Berthier A, Bouchaert E, Gheeraert C, Alexandre J, Porez G, Prawitt J, Dehondt H, Ploton M, Colin S, et al. 2014. Metformin interferes with bile acid homeostasis through AMPK-FXR crosstalk. *J Clin Invest* **124:** 1037–1051.

Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman MJ, Song X, Zhang F, Ma Q, Ohgi KA, et al. 2014. Enhancer activation requires *trans*-recruitment of a mega transcription factor complex. *Cell* **159:** 358–373.

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132:** 958–970.

Madsen MS, Siersbæk R, Boergesen M, Nielsen R, Mandrup S. 2014. Peroxisome proliferator-activated receptor γ and C/EBPα synergistically activate key metabolic adipocyte genes by assisted loading. *Mol Cell Biol* **34:** 939–954.

Magnani L, Eeckhoute J, Lupien M. 2011. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet* **27:** 465–474.

Marot G, Foulley J, Mayer C, Jaffrézic F. 2009. Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics* **25:** 2692–2699.

Mazuy C, Helleboid A, Staels B, Lefebvre P. 2015. Nuclear bile acid signaling through the farnesoid X receptor. *Cell Mol Life Sci* **72:** 1631–1650.

McLeay R, Bailey T. 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11:** 165.

Meng Z, Wang Y, Wang L, Jin W, Liu N, Pan H, Liu L, Wagman L, Forman BM, Huang W. 2010. FXR regulates liver repair after CCl4-induced toxic injury. *Mol Endocrinol* **24:** 886–897.

Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. 2016. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc* **11:** 316–326.

Neph S, Vierstra J, Stergachis A, Reynolds A, Haugen E, Vernot B, Thurman R, John S, Sandstrom R, Johnson A, et al. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489:** 83–90.

Nguyen TA, Jones RD, Snavely AR, Pfenning AR, Kirchner R, Hemberg M, Gray JM. 2016. High-throughput functional comparison of promoter and enhancer activities. *Genome Res* **26:** 1023–1033.

Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25:** 2730–2731.

O'Connor TR, Bailey TL. 2014. Creating and validating *cis*-regulatory maps of tissue-specific gene expression regulation. *Nucleic Acids Res* **42:** 11000–11010.

Oger F, Dubois-Chevalier J, Gheeraert C, Avner S, Durand E, Froguel P, Salbert G, Staels B, Lefebvre P, Eeckhoute J. 2014. Peroxisome proliferator-activated receptor γ (PPARγ) regulates genes involved in insulin/IGF signalling and lipid metabolism during adipogenesis through functionally distinct enhancer classes. *J Biol Chem* **289:** 708–722.

Pawlak M, Baugé E, Lalloyer F, Lefebvre P, Staels B. 2015. Ketone body therapy protects from lipotoxicity and acute liver failure upon Pparα deficiency. *Mol Endocrinol* **29:** 1134–1143.

Porez G, Gross B, Prawitt J, Gheeraert C, Berrabah W, Alexandre J, Staels B, Lefebvre P. 2013. The hepatic orosomucoid/α1-acid glycoprotein gene cluster is regulated by the nuclear bile acid receptor FXR. *Endocrinology* **154:** 3690–3701.

R Core Team. 2015. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43:** e47.

Siersbæk R, Rabiee A, Nielsen R, Sidoli S, Traynor S, Loft A, La Cour Poulsen L, Rogowska-Wrzesinska A, Jensen ON, Mandrup S. 2014. Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep* **7:** 1443–1455.

Smith CL, Goldsmith CW, Eppig JT. 2005. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* **6:** R7.

Smyth G. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**. doi: 10.2202/1544-6115.1027.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102:** 15545–15550.

Thomas AM, Hart SN, Kong B, Fang J, Zhong X, Guo GL. 2010. Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine. *Hepatology* **51:** 1410–1419.

Thomas AM, Hart SN, Li G, Lu H, Fang Y, Fang J, Zhong X, Guo GL. 2013. Hepatocyte nuclear factor 4 α and farnesoid X receptor co-regulates gene transcription in mouse livers on a genome-wide scale. *Pharm Res* **30:** 2188–2198.

Tong A, Liu X, Thomas BJ, Lissner MM, Baker MR, Senagolage MD, Allred AL, Barish GD, Smale ST. 2016. A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* **165:** 165–179.

Verfaillie A, Svetlichnyy D, Imrichova H, Davie K, Fiers M, Kalender Atak Z, Hulselmans G, Christiaens V, Aerts S. 2016. Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res* **26:** 882–895.

Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, et al. 2014. Mouse regulatory DNA landscapes reveal global principles of *cis*-regulatory evolution. *Science* **346:** 1007–1012.

Wang Y, Chen W, Moore DD, Huang W. 2008. FXR: a metabolic regulator and cell protector. *Cell Res* **18:** 1087–1095.

Wehrens R, Buydens L. 2007. Self- and super-organising maps in R: the kohonen package. *J Stat Softw* **21:** 1–19.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153:** 307–319.

Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW Jr, et al. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* **10:** R130.

Xie D, Boyle AP, Wu L, Zhai J, Kawli T, Snyder M. 2013. Dynamic *trans*-acting factor colocalization in human cells. *Cell* **155:** 713–724.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515:** 355–364.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25:** 2227–2241.

Zhang EE, Kay SA. 2010. Clocks not winding down: unravelling circadian networks. *Nat Rev Mol Cell Biol* **11:** 764–776.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

Zhang Z, Chang CW, Goh WL, Sung W, Cheung E. 2011. CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res* **39:** W391–W399.

Zhao J, Shi H, Ahituv N. 2013. Classification of topological domains based on gene expression and regulation. *Genome* **56:** 415–423.

Received October 18, 2016; accepted in revised form April 5, 2017.

The logic of transcriptional regulator recruitment architecture at *cis* -regulatory modules controlling liver functions

Julie Dubois-Chevalier, Vanessa Dubois, Hélène Dehondt, et al.

Genome Res. 2017 27: 985-996 originally published online April 11, 2017
Access the most recent version at doi:10.1101/gr.217075.116