SUPPLEMENTARY MATERIAL TO :

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

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

SUPPLEMENTARY RESULTS

Supplementary Results 1

Classes A and B were discarded from subsequent analyses since they showed low complexity recruitment patterns (Fig.1C and S4A), essentially lacked chromatin features associated with active CRM (Fig.1D-F) and were poorly linked to gene regulation (Fig.1G-H). Class A, despite comparable NR1H4 peak calling scores (Fig.S4B), also lacked enrichment for the canonical NR1H4 DNA binding sequence (Fig.S4C) and phylogenic conservation (Fig.S4D) and therefore most probably comprised unfiltered false positive regions. Non-functional genomic recruitment revealed by class B may stem from spurious binding reflecting genome scanning or non-functional evolving CRM (Lickwar et al. 2012; Spivakov 2014). Class C behaved as an intermediate for most of the analyzed criteria (Fig.1C-D and G-H), but was overall poorly marked with H3K9ac and H3K27ac (Fig.1E-F). We found that these CRM mostly bound CTCF and members of the cohesin complex such as RAD21 (Fig.S5), which characterize regions involved in the three-dimensional organization of the chromatin (Dixon et al. 2012; Rao et al. 2014). Although of potential interest, we focused further analyses on CRM defining fully active transcriptional regulatory elements from classes D to G.

Supplementary Results 2

The distribution of the number of TR co-localizing to individual CRM revealed a bimodal distribution, which was modelled assuming a mixture of an exponential and a normal distribution (Fig.S10A). Estimating the parameters of these distributions using an Expectation-Maximization algorithm allowed to predict that 75% of CRM are under the exponential distribution while the remaining 25% CRM are under the normal distribution.

Using the 3rd quartile of the number of TR (19) as a cut-off (Fig.S10A), we found that the subset of CRM under the normal distribution was particularly enriched in CRM from classes E and G (Fig.S10B).

Supplementary Results 3

Genes linked to CRM^{D-E} comprised recently identified NR1H4 target genes involved in ubiquitous processes such as autophagy (Lee et al. 2014; Seok et al. 2014). This is illustrated in Fig.S13A-C which show NR1H4, TR^{D-E} and TR^{F-G} ChIP-seq profiles at the *Bnip3*, *Map1lc3b* (*Lc3b*) and *Sesn2* genes. Other examples include the *Ero1lb* gene which can regulate susceptibility to endoplasmic reticulum stress (Khoo et al. 2011) and *Btg2* which encodes a cell-cycle regulator whose expression is linked to liver regeneration and hepatocarcinogenesis (Zhang et al. 2009) (Fig.S13D-E and Fig.S13K). Conversely, genes linked to CRM^{F-G} comprised NR1H4 target genes involved in BA metabolism such as *Slc10a1* and *Nr0b2* (*Shp*) (Lefebvre et al. 2009) (Fig.13F-G), the drug-metabolizing enzyme encoding gene *Fmo3* (Bennett et al. 2013) (Fig.S13H), the *Thrsp* (*Spot14*) gene involved in control of lipid metabolism (Duran-Sandoval et al. 2005) (Fig.S13I) and the *Fgg* gene encoding the blood clotting protein Fibrinogen gamma chain (Fig. S13J).

Supplementary Results 4

As an example of cross-talk between NR1H4 and a TR analyzed in our study, we compared hepatic gene regulations induced by NR1H4 and PPARA selective agonists. We found that while genes associated to CRM^{D-E} tend to be similarly regulated by these NR, all combinations exist for genes associated to CRM^{F-G} (Fig.S23). This is consistent with recent large-scale studies of TR activities which indicated context-dependent functions is a common feature (Stampfel et al. 2015). Hence, focused analyses of a limited number of well-defined genes such as described in (Tong et al. 2016) may help better understand the context-dependent functional relationship between TR.

Supplementary Results 5

Analyses of the entire landscape of liver CRM indicates the hierarchical combinations of TRM is a general organizational feature

In order to define whether the logical TRM organization discovered applies beyond NR1H4-bound CRM, we prepared a SOM using all mouse liver CRM, i.e. genomic regions bound at least by 2 out of the 48 analyzed TR. Again, we built on this analysis by further grouping the CRM into 6 classes using hierarchical clustering. This allowed to retrieve CRM corresponding to promoters, enhancers as well as CTCF/cohesin recruiting sites and non-functional/false-positives regions (Fig.S24A-G). In this context, NR1H4 binding was spread over most nodes comprising promoters and a limited fraction of nodes comprising active enhancers (Fig.S24H). Next, in order to monitor how the core, liver-specific functions control, promoter and circadian TRM identified in our study focusing on NR1H4-bound CRM were distributed over the entire CRM landscape, we divided each node into 4 equal compartments which we filled according to the proportion of CRM comprising at least 75% of the TR of a given TRM. We observed that the core TRM was found in a majority of nodes comprising active promoters and enhancers while the promoter TRM was restricted to promoters (Fig.S24I). The liver-specific functions control TRM was found in a large fraction of nodes corresponding to enhancers and in a subset of nodes corresponding to promoters. The circadian TRM was found in a subset of nodes corresponding to both active promoters and enhancers (Fig.S24I). Therefore, these data indicate that organization of CRM into hierarchical combinations of TRM extend to the entire mouse liver CRM landscape.

SUPPLEMENTARY DISCUSSION

In addition to their well-recognized metabolic control functions, NR5A2, PPARA and CEBP have also been ascribed with roles in protection from stress or liver regeneration (Anderson et al. 2002; Jakobsen et al. 2013; Jiao et al. 2014; Kersten 2014; Mamrosh et al. 2014). These findings indicate that a specific set of TR is instrumental in regulating both genes involved in widespread and liver-specific activities. In this context, recent findings have indicated that NR1D1 binds to conserved CRM to regulate common genes across tissues while it cooperates with ONECUT1 to regulate genes involved in metabolism in the liver (Zhang et al. 2015). Hence, a subset of NR including NR1H4 and NR1D1 may coordinately serve as nexus for concerted regulation of housekeeping/cellular maintenance genes and liver-specific metabolic functions. Intertwining of NR1H4-controlled biological outputs such as involvement of autophagy in the control of hepatic lipid homeostasis through autophagic lipolysis (Cingolani and Czaja 2016) contributes another layer of coordinated regulation between housekeeping and liver-specific functions.

SUPPLEMENTARY METHODS

Public functional genomics data recovery

Public functional genomics data used in this study were downloaded from the Gene Expression Omnibus (GEO), ArrayExpress, ENCODE (Yue et al. 2014), the UCSC Genome Browser (Raney et al. 2011) or from BioGPS (GNF1M atlas) (Wu et al. 2009) and are listed in Table S1. All data were obtained using the liver of adult C57BL/6 mice and we only used samples corresponding to untreated mice. Hepatocytes approximately constitute 70% of all cells in the liver and are mostly polyploid cells (up to 85% in C57BL/6 mice with mainly tetraploid hepatocytes) (Duncan et al. 2010). Moreover, we have focused on binding sites for the liver-specific transcription factor NR1H4. Hence, the chromatin signals analyzed largely stem from hepatocytes.

Coordinates for CpG islands (CGI) and gene transcription start sites (TSS) from GENCODE VM4 basic as well as 60-way-placental PhyloP conservation scores for mm10 were downloaded from the UCSC Genome Browser (Raney et al. 2011). Genomic coordinate conversions were performed using the liftOver tool from the UCSC Genome Browser.

TR ChIP-seq data processing

Initial pre-processing was performed using a customized local instance of Galaxy (Afgan et al. 2016). The FastQC package was used to ensure sufficient quality of the FASTQ files included in our analyses (Andrews S. 2010; FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc). All raw data were then mapped to the mm10 version of the mouse genome using Bowtie (version 1.0.0) with default parameters (Langmead et al. 2009). Peak calling was performed using model-based analysis of ChIP-seq version 2 (MACS2) (Zhang et al. 2008). Input DNA was used as control when available, duplicate tags were removed and parameters recommended for analysis of transcription factor ChIP-seq data were applied (Feng et al. 2011). A relatively relaxed cut-off set at $p < 0.001$ was used to initially include most real binding sites. When replicates were available, they were compared using Irreproducibility Discovery Rate (IDR) (Li et al. 2011) to identify any replicate which should be discarded and to select binding sites consistently called among replicates. This was performed according to IDR guidelines (https://sites.google.com/site/anshulkundaje/projects/idr) using the "optimal" number of consistent binding sites. When ChIP-seq experiments for a given factor had been performed at several times of the day, all peaks were used and merged into a single file. For all datasets, binding sites from the mitochondrial DNA (chr M) were discarded together with false positive calls identified from inputs and IgG ChIP-seq. Those were defined as the 0.01% regions with the highest tag counts in a pooled dataset of all available inputs and IgG ChIP-seq data (Pickrell et al. 2011).

Identified TR binding sites were then visually inspected using bigWig signal files and the Integrated Genome Browser (IGB) (Nicol et al. 2009) to check they were genuine enrichments. BigWig signal files were prepared by first discarding reads mapping to the false-positive regions identified earlier. Then, reads were extended at their 3' end according to read length predictions made by MACS2, counted within 25 bp windows genome-wide and normalized to the total number of uniquely mapped sequenced reads. Reads from replicates used for peak calling were merged and processed as described. Average ChIP-seq profiles were also obtained using these bigWig files.

Finally, to define genomic regions of interest for self-organizing maps (SOM) analyses, all TR binding sites identified (extended 250 base pairs on each side of the peak center) were intersected using Bedtools (Quinlan and Hall 2010) in order to identify *cis*-

regulatory modules (CRM) characterized by the co-occurrence of at least 2 different TR.

Self-Organizing Maps (SOM) analyses

The SOM were generated using the R package “kohonen2” (Wehrens and Buydens 2007). The input vectors from CRM, optimal number of nodes and parameters to train the SOM were defined according to (Xie et al. 2013). Training was performed using random initialization of the toroid with hexagonal nodes. To verify the defined optimal number of nodes was appropriate, we performed SOM training using increasing numbers of nodes (200 to 5000) and evaluated the clustering quality using quantization error (Kohonen 2001), qM1(Lavrač et al. 2003) and organization score (Flexer 2001). The SOM training using 100 iterations was sufficient to obtain a convergence towards a low and stable quantization error. Finally, the maps selected for further analyses were the best of 100 trials based on lowest quantization error, highest organization score and highest percentage of non-empty nodes having a significant enrichment of co-localization pattern. This last parameter consisted of a binomial test used to define whether the TR combination representative of a given node is specifically enriched at CRM comprised within this node compared to all other CRM (Xie et al. 2013). The seeds for the selected maps were 53 (SOM of FXR-bound CRM) and 20 (SOM of all CRM). Empty nodes were displayed in grey in the final maps.

Nodes were further grouped into classes 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 TR combination (prototype) for each individual node. 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.

Multidimensional scaling (MDS) analyses

TR co-occurence at CRM from classes D, E, F or G was used to calculate tanimoto distance matrices of dimension 48 TR x 48 TR. A mutli-dimensionnal scaling was then performed for each class using the cmdscale function of the R package “Stats” (R Core Team 2015) and the tanimoto distance matrices as input. The two first dimensions were plotted.

Hierarchical clustering analyses of TR co-occurrence

Hierarchical clusterings of TR co-occurrence and associated heatmaps were obtained using their tanimoto distance matrices calculated for the MDS analyses and the heatmap.2 function of the R package “gplots” (Warnes et al. 2016).

Gene ontology (GO) and mouse phenotype (MP) enrichment analyses

GO enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7) (Huang et al. 2009). Panther biological processes with Bonferroni-corrected p-value < 0.05 were considered and GO terms comprising gene lists that were more than 90% identical were merged into a single class. ToppCluster was used to link TR to MP (Kaimal et al. 2010). MP with Bonferroni-corrected p-values < 0.05 were considered and similar MPs were merged.

Gene set enrichment analyses (GSEA)

GSEA was performed using the GSEA software developed at the Broad Institute (Subramanian et al. 2005). We used 1000 gene-set permutations and the following settings: “weighted” as the enrichment statistic and “Signal2Noise” as the metric for ranking genes.

CRM target gene assignment

CRM localized within 2.5 kb of a GENCODE gene TSS were assigned to this gene. Target gene assignment for distal CRM was performed using a model correlating cross-tissue CRM activities based on histone acetylation to gene transcriptional expression (O'Connor and Bailey 2014). Transcriptomic data from BioGPS and H3K27ac enriched regions from ENCODE (lifted to mm10) (Table S1) for 13 tissues (bone marrow, brown fat, cerebellum, cortex, heart, kidney, liver, olfactory bulb, placenta, small intestine, spleen, testis and thymus) were used.

Transcriptomic data analyses

Raw transcriptomic data from Affymetrix microarrays were normalized using the Partek Genomics Suite or the R package "oligo" (Carvalho and Irizarry 2010) using background correction by Robust Multi-array Average (RMA), quantile normalization and summarization via median-polish. The normalized gene expression values for *Per2* KO transcriptomic data from Agilent microarrays were directly downloaded from the GEO database. Principal component analyses (PCA) were used for quality control of the data. The average normalized expression of genes (averaged by Gene Symbol) were then used to perform the differential expression analyses using limma (Ritchie et al. 2015; Smyth 2004). Dysregulated genes were defined using a Benjamini–Hochberg corrected *p*-value cut-off set at 0.15 for all data except for the *Hnf1a* (0.05) and *Per2* (0.01) KO transcriptomic data. This allowed us to define dysregulated genes which were in the same range (between 1200 and 2700) for all datasets and numerous enough for robust downstream analyses.

To determine dysregulated genes in the liver of *Nr1h4* KO mice, genes analyzed in both E-MTAB-1722 and GSE54557 were retrieved based on their Gene Symbol. Then a meta-analysis was performed using the average normalized expression of these genes using the metaMA package (Marot et al. 2009). The pvalcombination function was used together with the following parameters: moderated set as ”limma” and BHth set at 0.15.

All data were used throughout the study to monitor expression of genes assigned to NR1H4-bound CRM as described above.

Intragenomic replicates (IGR)

The functional impact of SNV on TR binding was predicted using the IGR tool as previously described (Cowper-Sal·lari et al. 2012). IGR compares the average ChIP-seq signal intensity of TR across the genomic loci that contains the underlying sequence (7-mers) of the reference or the variant allele of each SNV. To do so, IGR uses a sliding window of size 7 bp such that it contains the reference or the variant allele and finds all occurrences of these 7-mers. The average intensity of the TR of interest is then computed for all 7-mers with the reference and variant allele separately. The 7-mer with the highest average intensity matching the reference allele is tested against the 7-mer with the highest average intensity that matches the variant allele. The genomic locations of all 7-mers were filtered to include only sites corresponding to accessible chromatin and only the SNVs within 50 bp of transcription factor binding peak center, which co-localized within accessible chromatin were tested. This made use of mouse liver DHS sites which were defined from the ENCODE data (Table S1) using MACS2 and IDR as previously described. Significant modulations of transcription factor binding were considered using a p-value cut-off set at 0.05 (Benjamini–Hochberg corrected values). 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 enrichments between 2 sets of CRM were determined using AME (Analysis of Motif Enrichment) using “Total matches” as the scoring method sequence and a motif match threshold set at 1x10^{-4} while motif scanning was performed with FIMO using default parameters (Find Individual Motif Occurences) (Grant et al. 1017-1018), both from the MEME suite (McLeay and Bailey 2010). The HOCOMOCO Mouse (v10) motif database was used for all these analyses.

Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)

Livers from wild-type mice were minced and passed through a 70 µm filter before double cross-linking with 2 mM disuccinimidyl glutarate (30 minutes at room temperature) and 1% v/v formaldehyde (10 minutes at room temperature). Samples were then processed for RIME as described in (Mohammed et al. 2016). Experiments were performed in duplicates using both an antibody directed against NR1H4 and non-immune control IgG (sc13063 and sc2027 from Santa-Cruz biotechnology, respectively). Mass spectrometry was performed by the proteomic core facility at Cancer Research UK. TR detected in any of the IgG samples were discarded.

Broad H3K4me3 domain identification

H3K4me3 ChIP-seq data from the ENCODE consortium were analyzed to call H3K4me3 enriched regions using MACS2 as defined in (Chen et al. 2015). Broad H3K4me3 domains were defined as those spanning more than 3 times the median size of all H3K4me3 enriched regions in a given tissue.

Animal experimentations

Mice were housed in a temperature-controlled room (22–24°C) with a relative humidity of 36%–80%, and 12-hour light/12-hour dark cycles. *Nr1h4* and *Ppara* KO mice have been described previously (Berrabah et al. 2014; Pawlak et al. 2015; Porez et al. 2013). 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.

Real-time PCR analysis of gene expression

RNA extraction, reverser transcription (RT) and real-time quantitative PCR (qPCR) were performed as previously described (Dubois-Chevalier et al. 2014). Gene expression levels were normalized using the *Rplp0* housekeeping gene expression level as an internal control. All primers used for RT-qPCR are listed in Table S4.

Gene expression microarrays

RNA extracted from primary hepatocytes (n=3) or *Ppara* KO and wild-type mice (n=6) was checked for quantity and quality using the Agilent 2100 Bioanalyzer (Agilent Biotechnologies) before being processed for analysis using MoGene-2_0-st Affymetrix arrays according to the manufacturer’s instructions. Data were analyzed as described hereabove.

Mouse primary hepatocyte isolation and treatment

Mouse primary hepatocytes (n=3) were prepared as described in (Bantubungi et al. 2014)**.** Hepatocytes were grown in serum free William's medium and treated for 4h with GW4064 (2µM) or DSMO.

LEGENDS TO SUPPLEMENTARY FIGURES

Figure S1. The mouse liver REST and NR1H4 cistromes are unrelated

A) The Integrated Genome Browser (IGB) was used to show largely inconsistent ChIP-seq profiles for REST and NR1H4 over a large region of the mouse liver genome. **B)** Genomic Regions Enrichment of Annotations Tool (GREAT) was used to associate REST and NR1H4 binding sites to genes (default parameters) and subsequently identify gene set over-representation [8]. Benjamini–Hochberg corrected *p*-values (-$\log_{10}$) of the top 5 GO terms are shown. GO terms comprising gene lists that were more than 90% identical merged into a single class.

Figure S2. Basic features of the CRM and quality assessment of the SOM analysis.

A) The map issued from Fig.1B was used to indicate the number of independent CRM comprised within each individual node. B**)** Bar graph showing the size distribution of all CRM used for the SOM analysis. **C)** Bar graph showing the distribution of the number of TR co-occurring at individual CRM. **D)** The map issued from Fig.1B was used to indicate the average distance between a given node and its neighbours obtained after pairwise comparison of the most representative TR combination of the individual nodes. Bold black lines indicate the borders of the clusters.

Figure S3. Preferential co-localization of TR from the same dataset can be ruled out as a major confounding effect in the SOM analysis.

A) The map issued from Fig.1B was used to indicate the number of independent studies (see Table S1 for details), which are represented in each individual node. At least 1 TR from a given study had to be found in more than 50% of the CRM of an individual node to be considered. **B)** The map issued from Fig.1B was used to indicate the average percent of TR from a given study, which co-localize to single CRM from each individual node. All studies contributing 3 TR or more are shown.

Figure S4. Additional data showing differential activity of the NR1H4-bound CRM clusters

A) The map issued from Fig.1B was used to show the average distance between CRM from a given node obtained from pairwise comparisons of TR combinations at all CRM. Higher average distance correlates with higher number of binding TR shown in Fig.1C and points to a greater number of combinations with subtle differences. **B)** Box plot displaying $–\log_{10}$ *p*-values provided by the MACS2 peak calling algorithm for NR1H4-bound CRM from the different clusters. **C)** Presence of the canonical NR1H4 binding motif (Inverted repeat 1 or IR1) within NR1H4 binding sites from CRM of classes A-G was defined using CENTDIST (Zhang et al. 2011). **D)** The map issued from Fig.1B was used to show the average phylogenetic conservation score of CRM from individual nodes. Bold black lines indicate the borders of the clusters.

Figure S5. Cluster C comprises CRM with strong CTCF and cohesin binding

The map issued from Fig.1B was used to show the percentage of CRM bound by CTCF or RAD21 in each node (top) as well as the average CTCF and RAD21 ChIP-seq levels at CRM from each individual node (bottom).

Figure S6. MDS analysis of TR co-occurrence at CRM from class D.

MDS was performed as described in the Materials and Methods section using CRM from class D. The framed area, which contains TR which are the most strongly interconnected with NR1H4 (Tanimoto index > 0.7), is shown in details in Fig.2.

Figure S7. MDS analysis of TR co-occurrence at CRM from class E.

MDS was performed as described in the Materials and Methods section using CRM from class E. The framed area, which contains TR which are the most strongly interconnected with NR1H4 (Tanimoto index > 0.7), is shown in details in Fig.2.

Figure S8. MDS analysis of TR co-occurrence at CRM from class F.

MDS was performed as described in the Materials and Methods section using CRM from class F. The framed area, which contains TR which are the most strongly interconnected with NR1H4 (Tanimoto index > 0.7), is shown in details in Fig.2.

Figure S9. MDS analysis of TR co-occurrence at CRM from class G.

MDS was performed as described in the Materials and Methods section using CRM from class G. The framed area, which contains TR which are the most strongly interconnected with NR1H4 (Tanimoto index > 0.7), is shown in details in Fig.2.

Figure S10. Bimodal distribution of the number of co-recruited TR at CRM

A) Plot showing the fitting of the modelled exponential (red) and normal (green) distributions on the TR density distribution at CRM. The equation which was used together with estimated parameters are provided on top of the plot. **B)** The map issued from Fig.1B was used to indicate the percentage of CRM from each individual node which is co-bound by at least 19 different TR.

Figure S11. Examples of TR showing differential occurrence at NR1H4-bound CRM.

The map issued from Fig.1B was used to show the percentage of CRM bound by the indicated TR (identified in Fig.2) in each individual node.

Figure S12. CRM from classes D and E on one hand and from classes F and G on the other hand were associated with genes showing identical GO term enrichments and expression profiles across mouse tissues

Analyses were performed as in Fig.2I and J using genes associated with CRM from classes D, E, F or G as indicated.

Figure S13. TR ChIP-seq profiles at example genes linked to CRM from classes D-E or F-G

The Integrated Genome Browser (IGB) was used to visualize ChIP-seq profiles for NR1H4, TR^{D-E} and TR^{F-G} at the indicated genes associated with CRM^{D-E} (panels **A-E**) or CRM^{F-G} (panles **F-J**), which are highlighted into boxes. DHS, H4K4me1 and 3 as well as H3K9ac ChIP-seq levels are also shown at the bottom. **K)** *Ero1lb* and *Btg2* expression is altered in the liver of *Nr1h4* KO mice. RT-qPCR analyses were performed using the liver of whole-body *Nr1h4* KO mice (left; n=5) or liver-specific *Nr1h4* KO mice (right; n=3). An equivalent number of wild-type littermates were used as controls. Student's t-test for unpaired data was used to define statistically significant differences between control and *Nr1h4* KO mice, * $p < 0.05$ and ** $p < 0.01$.

Figure S14. Expression changes of genes linked to CRM^{D-E} or CRM^{F-G} in the liver of *Nr1h4* KO mice.

Box plot showing absolute fold changes of genes linked to CRM^{D-E} or CRM^{F-G} in the liver of liver-specific *Nr1h4* KO mice. A Mann-Whitney test was used to define statistical differences between the 2 groups, *** $p < 0.001$.

Figure S15. Characteristics of genes linked to both CRM^{D-E} and CRM^{F-G}.

A) Gene ontology (GO) enrichment analyses were performed using DAVID (Huang et al. 2009) and genes associated uniquely with CRM^{D-E} (CRM^{D-E} only) or CRM^{F-G} (CRM^{F-G} only) or associated with both CRM classes (CRM^{D-E} + CRM^{F-G}). Bonferroni-corrected p-value (-$\log_{10}$) are shown. **B)** Average normalized mRNA expression levels of genes associated uniquely with CRM^{D-E} (CRM^{D-E} only) or CRM^{F-G} (CRM^{F-G} only) or associated with both CRM classes (CRM^{D-E} + CRM^{F-G}) across indicated mouse tissues were obtained using BioGPS data (Wu et al. 2009). Results are means +/- S.E.M.

Figure S16. Comparison of TR occurrence at CRM^{D-E} and CRM^{F-G} promoters.

Plots showing the occurrence of each TR at CRM^{D-E} and CRM^{F-G} promoters. TR were colored according to Fig.4A.

Figure S17. Additional features discriminating CRM^{D-E} promoters from CRM^{F-G} promoters and enhancers.

A) DNA binding motifs enriched in CRM^{D-E} and CRM^{F-G} promoters and in CRM^{F-G} enhancers (defined using regions from class A as control) are indicated using the name of the recognizing transcription factor. Moreover, < and > were used to indicate significant differential enrichment within distinct sets of CRM. **B)** ChIP-seq signals were recovered from the indicated CRM by restricting the analyses to the specific original peak of a given TR within each CRM. Unbound CRM^{D-E} and CRM^{F-G} were used as a control. Results are means +/- S.E.M. **C)** Plot showing the percentage of the genomic regions encompassed by CRM^{D-E} or CRM^{F-G} promoters which overlaps with CpG islands (CGI).

Figure S18. Fraction of NR1H4 target genes associated with CRM from class D, E, F or G dysregulated in the liver of TR KO mice.

Genes exclusively associated with CRM from class D, E, F or G and whose expression is modified in the liver of liver-specific *Nr1h4* KO mice were used for these analyses. Genes which are not linked to NR1H4-bound CRM and whose expression is not altered in the liver of *Nr1h4* KO mice (NR1H4 non-target 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 non-target genes (* p < 0.05, ** p < 0.01 and *** p < 0.001).

Figure S19. Hierarchical clustering of TR co-occurrence at CRM from class E.

Heatmap showing TR co-occurrence at CRM from class E defined using a Tanimoto index. The hierarchical clustering tree is shown on the left with bars corresponding to circadian TR highlighted in green. Clusters of circadian TR are framed using green lines.

FigS20. Comparison of motif occurrence at CRM^{G} promoters and enhancers +/- circadian TRM.

Plots showing the percentage of CRM harbouring at least one copy of a given DNA binding motif is reported. All motifs from the HOCOMOCO Mouse (v10) database were used (427 motifs). Motifs outside the area delimited by the grey lines show more than 2 fold differences in their occurrence within the 2 sets of CRM being compared. Below are indicated the actual motifs highlighted by the transcription factor names in the plots.

Figure S21. Hierarchical clustering of the impact of SNV on TR binding to CRM^{G} inferred from IGR analyses.

The IGR tool was used to predict the impact of SNV localized within CRM^{G} on chromatin binding of the indicated TR which were grouped based on hierarchical clustering. In order to restrict our analyses to a manageable number of SNV, we selected those modulating NR1H4 binding. The hierarchical clustering tree is shown on the left. 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).

Figure S22. Broad H3K4me3 labelling of genes encoding the indicated TR in the liver and 10 other mouse tissues.

Broad H3K4me3 domains were defined for the mouse liver or for the 10 other tissues also analyzed in Fig.2J as defined in the Supplementary Material. Stars indicate TR whose encoding gene is marked with a broad H3K4me3 domain in the liver and in less than 25% of other analyzed tissues.

Figure S23. Comparison of NR1H4 and PPARA agonist-induced transcriptional regulation of genes linked to CRM$^{D\text{-}E}$ and CRM$^{F\text{-}G}$.

Comparison of the hepatic transcriptomic modulations induced by the PPARA agonist Wy14643 (Rakhshandehroo et al. 2007) and the NR1H4 agonist GW4064 (see the Supplementary Material). We used data obtained after a short time treatment (4-6h) in order to mainly capture primary regulatory events induced by NR1H4 and PPARA. Genes significantly regulated by Wy14643 were defined using a Benjamini–Hochberg corrected *p*-value cut-off set at 0.05. Gene set enrichment analyses (GSEA) (Subramanian et al. 2005) showed a trend towards similar regulation of genes linked to CRM^{D-E}, which was significant for down-regulated genes (**A**). On the other hand, GSEA did not reveal any significant bias for genes linked to CRM^{F-G} and indicated that both similar and opposite transcriptional regulations by NR1H4 and PPARA occur (**B**). The maximal positive and negative enrichment score values were retrieved and the greatest of the 2 was divided by the other (Max ES ratio). When the negative maximal ES value was greater, the ratio was indicated as a negative value. FDR is the false discovery rate provided by the GSEA software.

Figure S24. Hierarchical combinations of TRM extends beyond NR1H4-bound CRM

A) All CRM 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 6 main classes of CRM which are indicated on the planar view of the toroidal map using different colors. Functional identification of CRM comprised within these classes was based on results from panels B-G and is indicated on the top. **B-E)** The map issued from A was used to indicate the average DHS (**B**), H3K4me1 (**C**) H3K4me3 (**D**) and H3K27ac (**E**) levels at CRM contained in each node. Bold black lines indicate the borders of the clusters. **F-H)** The map issued from A was used to show the percentage of CRM bound by CTCF (**F**), RAD21 (**G**) or NR1H4 (**H**) in each node. Bold black lines indicate the borders of the clusters. **I)** Each node was divided into 4 equal compartments which were filled according to the proportion of CRM within that node recruiting the core (black), promoter (blue), liver-specific functions control (violet) and circadian (green) TRM. Only part of the SOM comprising active CRM is shown here.

SUPPLEMENTARY REFERENCES

Afgan, E., Baker, D, van den Beek, M, Blankenberg, D, Bouvier, D, Čech, M, Chilton, J, Clements, D, Coraor, N, Eberhard, Cet al.. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3-W10.

Anderson, S.P., Yoon, L, Richard, E.B, Dunn, C.S, Cattley, R.C and Corton, J.C. 2002. Delayed liver regeneration in peroxisome proliferator-activated receptor-alpha-null mice. *Hepatology* **36**: 544-554.

Bantubungi, K., Hannou, S, Caron-Houde, S, Vallez, E, Baron, M, Lucas, A, Bouchaert, E, Paumelle, R, Tailleux, A and Staels, B. 2014. Cdkn2a/p16Ink4a regulates fasting-induced hepatic gluconeogenesis through the PKA-CREB-PGC1α pathway. *Diabetes* **63**: 3199-3209.

Bedward, M., Keith, D and Pressey, R. 1992. Homogeneity analysis: Assessing the utility of classifications and maps of natural resources.. *Australian Journal of Ecology* **17**: 133-139.

Bennett, B.J., de Aguiar Vallim, T.Q, Wang, Z, Shih, D.M, Meng, Y, Gregory, J, Allayee, H, Lee, R, Graham, M, Crooke, Ret al.. 2013. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab* **17**: 49-60.

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

Carvalho, B.S. and Irizarry, R.A. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**: 2363-2367.

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

Cingolani, F. and Czaja, M.J. 2016. Regulation and Functions of Autophagic Lipolysis. *Trends Endocrinol Metab* **27**: 696-705.

Dixon, J.R., Selvaraj, S, Yue, F, Kim, A, Li, Y, Shen, Y, Hu, M, Liu, J.S and 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, F.F, Gheeraert, C, Staels, B, Lefebvre, P and Eeckhoute, J. 2014. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res* **42**: 10943-10959.

Duncan, A., Taylor, M, Hickey, R, Hanlon Newell, A, Lenzi, M, Olson, S, Finegold, M and Grompe, M. 2010. The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**: 707-710.

Duran-Sandoval, D., Cariou, B, Percevault, F, Hennuyer, N, Grefhorst, A, van Dijk, T.H, Gonzalez, F.J, Fruchart, J, Kuipers, F and Staels, B. 2005. The farnesoid X receptor modulates hepatic carbohydrate metabolism during the fasting-refeeding transition. *J Biol Chem* **280**: 29971-29979.

Flexer, A.. 2001. On the use of self organizing maps for clustering and visualization. *Intelligent Data Analysis* **5**: 373-384.

Grant, C., Bailey, T and Noble, W. 1017-1018. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**.

Huang, D.W., Sherman, B.T and Lempicki, R.A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Jakobsen, J.S., Waage, J, Rapin, N, Bisgaard, H.C, Larsen, F.S and Porse, B.T. 2013. Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. *Genome Res* **23**: 592-603.

Jiao, M., Ren, F, Zhou, L, Zhang, X, Zhang, L, Wen, T, Wei, L, Wang, X, Shi, H, Bai, Let al.. 2014. Peroxisome proliferator-activated receptor α activation attenuates the inflammatory response to protect the liver from acute failure by promoting the autophagy pathway. *Cell Death Dis* **5**: e1397.

Kaimal, V., Bardes, E.E, Tabar, S.C, Jegga, A.G and Aronow, B.J. 2010. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* **38**: W96-102.

Kersten, S.. 2014. Integrated physiology and systems biology of PPARα. *Mol Metab* **3**: 354-371.

Khoo, C., Yang, J, Rajpal, G, Wang, Y, Liu, J, Arvan, P and Stoffers, D.A. 2011. Endoplasmic reticulum oxidoreductin-1-like β (ERO1lβ) regulates susceptibility to endoplasmic reticulum stress and is induced by insulin flux in β-cells. *Endocrinology* **152**: 2599-2608.

Kohonen, T.. 2001. Self Organizing Maps. *Springer*.

Lavrač, N., Gamberger, D, Blockeel, H and Todorovski, L. 2003. *Machine Learning: ECML 2003: 14th European Conference on Machine Learning, Cavtat-Dubrovnik, Croatia, September 22-26, 2003. Proceedings*.

Le, S., Josse, J and Husson, F. 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**: 1-18.

Lee, J.M., Wagner, M, Xiao, R, Kim, K.H, Feng, D, Lazar, M.A and Moore, D.D. 2014. Nutrient-sensing nuclear receptors coordinate autophagy. *Nature* **516**: 112-115.

Lefebvre, P., Cariou, B, Lien, F, Kuipers, F and 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 and Bickel, P. 2011. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**: 1752-1779.

Lickwar, C.R., Mueller, F, Hanlon, S.E, McNally, J.G and Lieb, J.D. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251-255.

Mamrosh, J.L., Lee, J.M, Wagner, M, Stambrook, P.J, Whitby, R.J, Sifers, R.N, Wu, S, Tsai, M, Demayo, F.J and Moore, D.D. 2014. Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution. *Elife* **3**: e01694.

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

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

Mohammed, H., Taylor, C, Brown, G.D, Papachristou, E.K, Carroll, J.S and D'Santos, C.S. 2016. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc* **11**: 316-326.

O'Connor, T.R. and Bailey, T.L. 2014. Creating and validating cis-regulatory maps of tissue-specific gene expression regulation. *Nucleic Acids Res* **42**: 11000-11010.

Pawlak, M., Baugé, E, Lalloyer, F, Lefebvre, P and 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 and 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.*

Raney, B.J., Cline, M.S, Rosenbloom, K.R, Dreszer, T.R, Learned, K, Barber, G.P, Meyer, L.R, Sloan, C.A, Malladi, V.S, Roskin, K.Met al.. 2011. ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res* **39**: D871-5.

Rao, S.S.P., Huntley, M.H, Durand, N.C, Stamenova, E.K, Bochkov, I.D, Robinson, J.T, Sanborn, A.L, Machol, I, Omer, A.D, Lander, E.Set al.. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665-1680.

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

Seok, S., Fu, T, Choi, S, Li, Y, Zhu, R, Kumar, S, Sun, X, Yoon, G, Kang, Y, Zhong, Wet al.. 2014. Transcriptional regulation of autophagy by an FXR-CREB axis. *Nature* **516**: 108-111.

Smyth, G.. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**.

Spivakov, M.. 2014. Spurious transcription factor binding: non-functional or genetically redundant?. *Bioessays* **36**: 798-806.

Stampfel, G., Kazmar, T, Frank, O, Wienerroither, S, Reiter, F and Stark, A. 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**: 147-151.

Subramanian, A., Tamayo, P, Mootha, V.K, Mukherjee, S, Ebert, B.L, Gillette, M.A, Paulovich, A, Pomeroy, S.L, Golub, T.R, Lander, E.Set al.. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**: 15545-15550.

Tong, A., Liu, X, Thomas, B.J, Lissner, M.M, Baker, M.R, Senagolage, M.D, Allred, A.L, Barish, G.D and Smale, S.T. 2016. A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. *Cell* **165**: 165-179.

Warnes, G., Bolker, B, Bonebakker, L, Gentleman, R, Liaw, W, Lumley, T, Maechler, M, Magnusson, A, Moeller, S, Schwartz, Met al.. 2016. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.

Wehrens, R. and Buydens, L. 2007. Self- and Super-organising Maps in R: the kohonen package. *J. Stat. Softw.* **21**.

Wu, C., Orozco, C, Boyer, J, Leglise, M, Goodale, J, Batalov, S, Hodge, C.L, Haase, J, Janes, J, Huss, J.W.3et al.. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* **10**: R130.

Xie, D., Boyle, A.P, Wu, L, Zhai, J, Kawli, T and 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, B.Det al.. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355-364.

Zhang, Y., Fang, B, Emmett, M.J, Damle, M, Sun, Z, Feng, D, Armour, S.M, Remsberg, J.R, Jager, J, Soccio, R.Eet al.. 2015. GENE REGULATION. Discrete functions of nuclear receptor Rev-erbα couple metabolism to the clock. *Science* **348**: 1488-1492.

Zhang, Z., Chang, C.W, Goh, W.L, Sung, W and Cheung, E. 2011. CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res* **39**: W391-9.

Zhang, Z., Wang, G, Chen, C, Yang, Z, Jin, F, San, J, Xu, W, Li, Q, Li, Z and Wang, D. 2009. Rapid induction of PC3/BTG2 gene by hepatopoietin or partial hepatectomy and its mRNA expression in hepatocellular carcinoma. *Hepatobiliary Pancreat Dis Int* **8**: 288-293.

Fig.S1

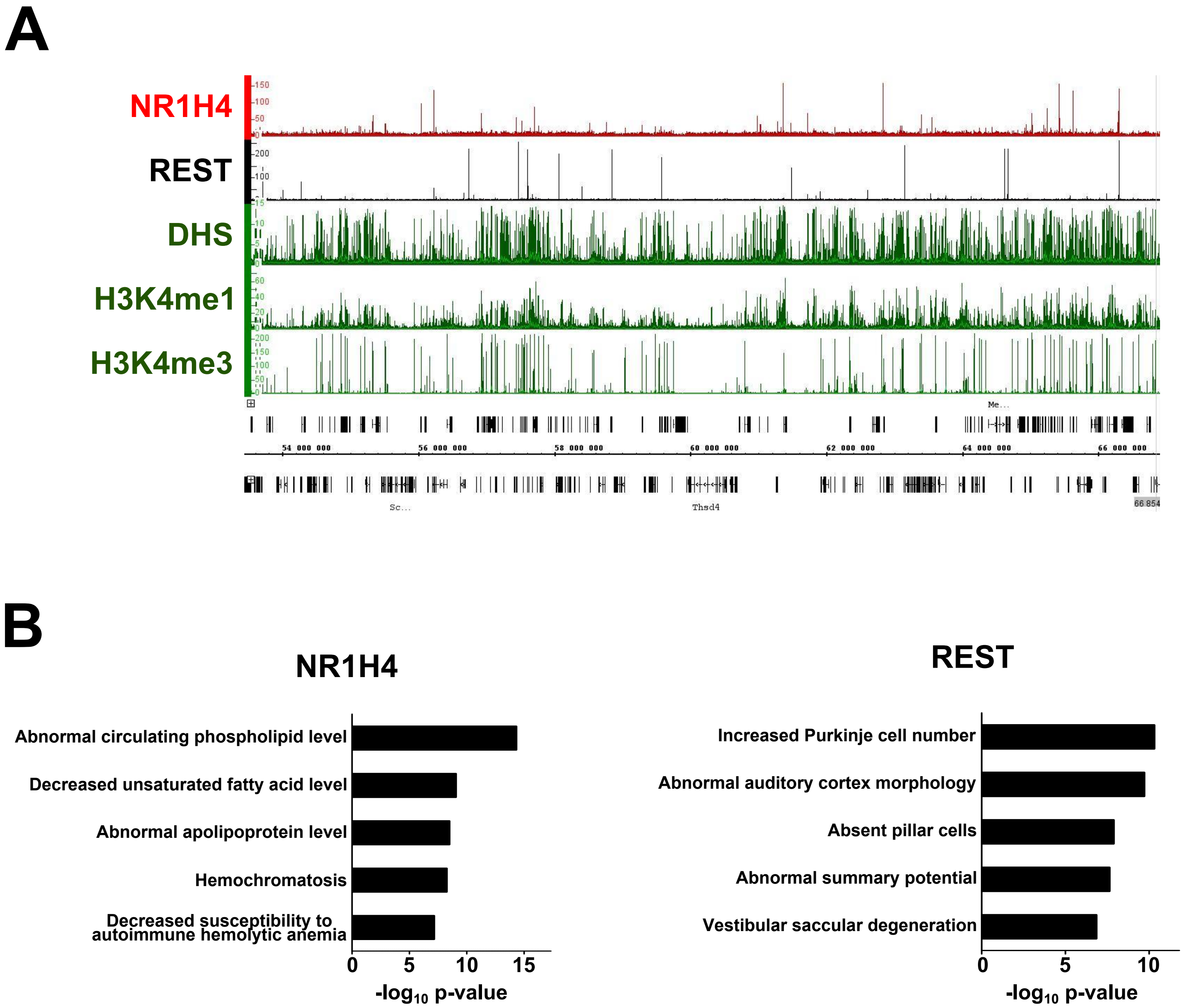

A
NR1H4
REST
DHS
H3K4me1
H3K4me3
150
100
50
0
200
100
15
10
5
0
60
40
20
0
200
150
100
50
0
Me...
54 000 000
56 000 000
58 000 000
60 000 000
62 000 000
64 000 000
66 000 000
Sc...
Thsd4
66.854
B
NR1H4
Abnormal circulating phospholipid level
Decreased unsaturated fatty acid level
Abnormal apolipoprotein level
Hemochromatosis
Decreased susceptibility to autoimmune hemolytic anemia
0
5
10
15
-log10 p-value
REST
Increased Purkinje cell number
Abnormal auditory cortex morphology
Absent pillar cells
Abnormal summary potential
Vestibular saccular degeneration
0
5
10
-log10 p-value

Fig.S2

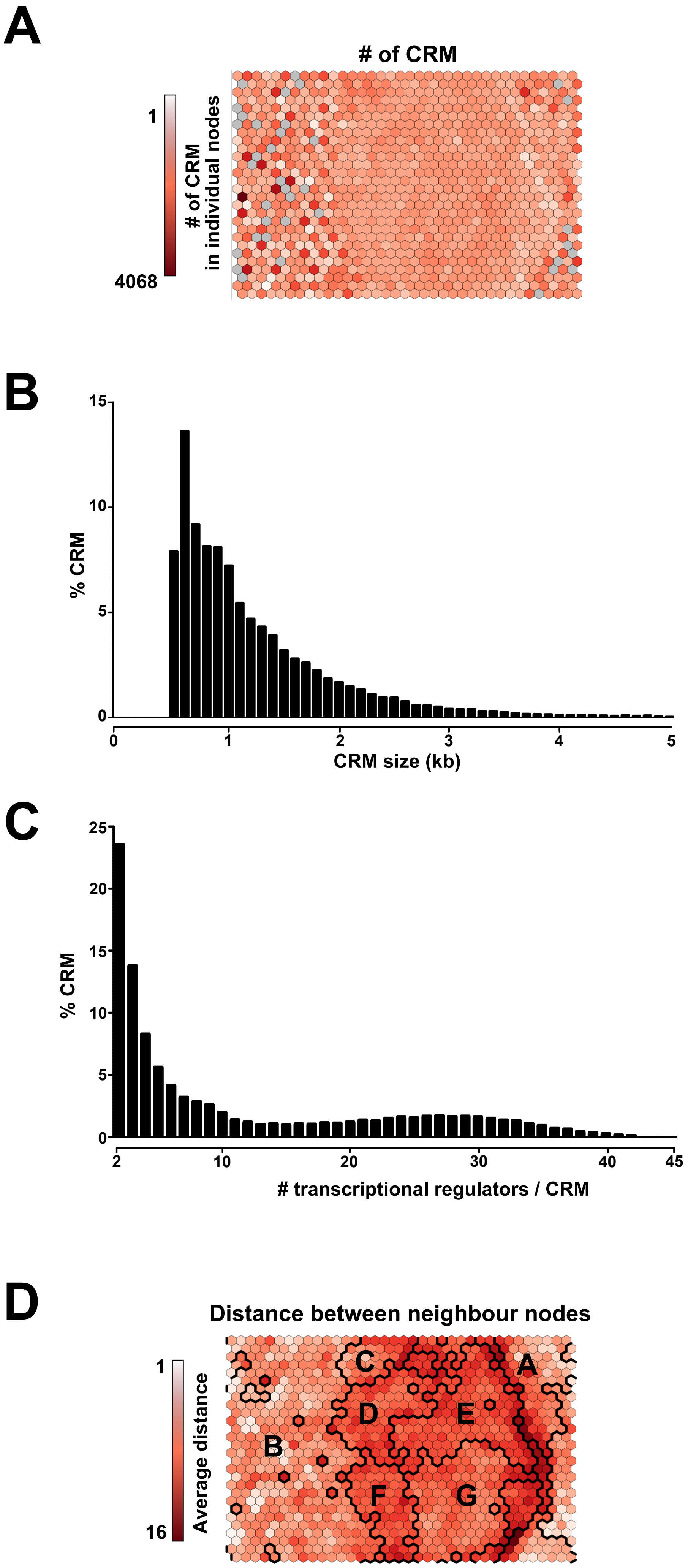

A
# of CRM
1
4068
# of CRM
in individual nodes
B
% CRM
15
10
5
0
0
1
2
3
4
5
CRM size (kb)
C
% CRM
25
20
15
10
5
0
2
10
20
30
40
45
# transcriptional regulators / CRM
D
Distance between neighbour nodes
1
16
Average distance
A
B
C
D
E
F
G

Fig.S3

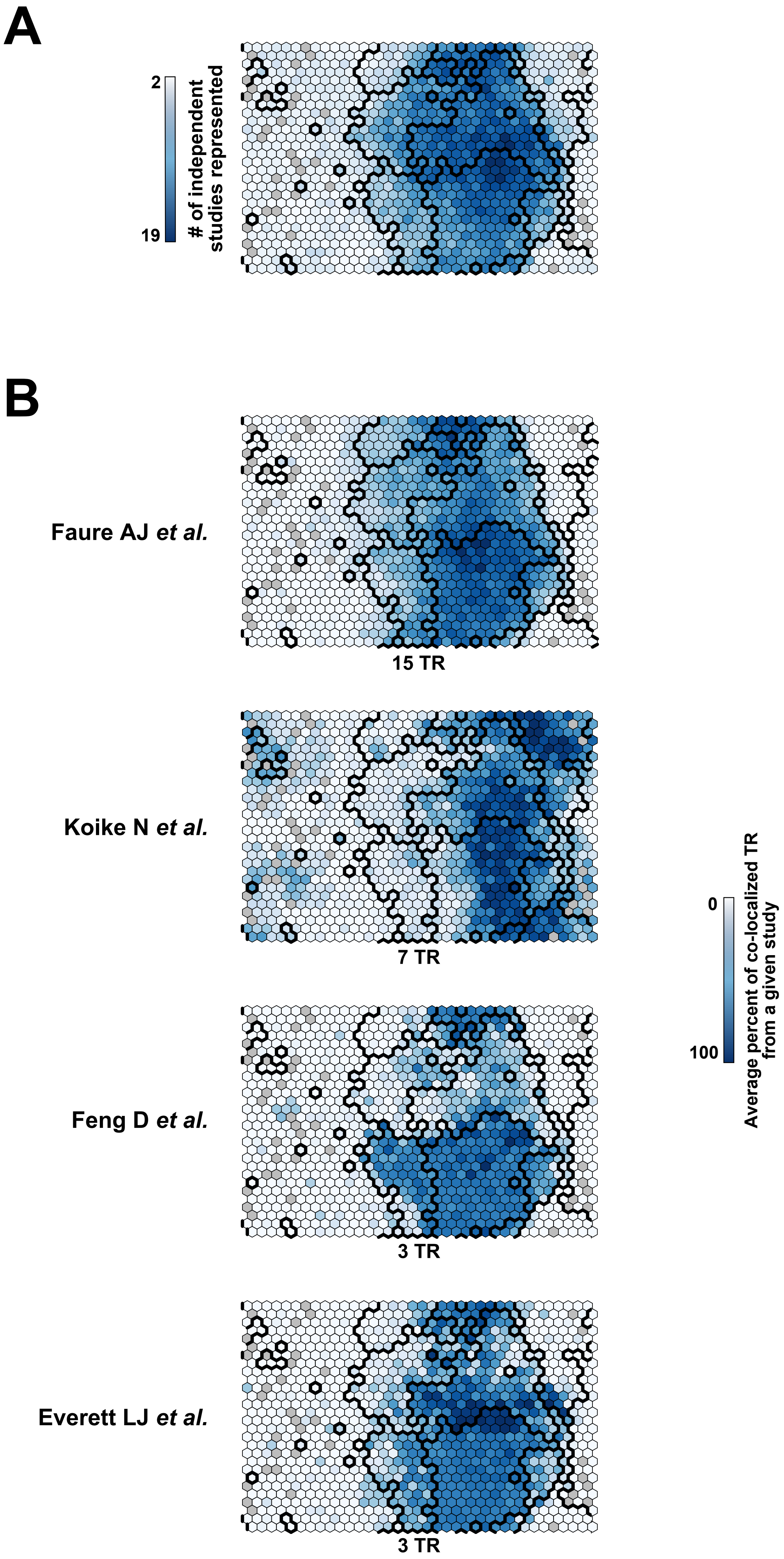

A
2
19
# of independent
studies represented
B
Faure AJ *et al.*
15 TR
Koike N *et al.*
7 TR
Feng D *et al.*
3 TR
Everett LJ *et al.*
3 TR
0
100
Average percent of co-localized TR
from a given study

Fig.S4

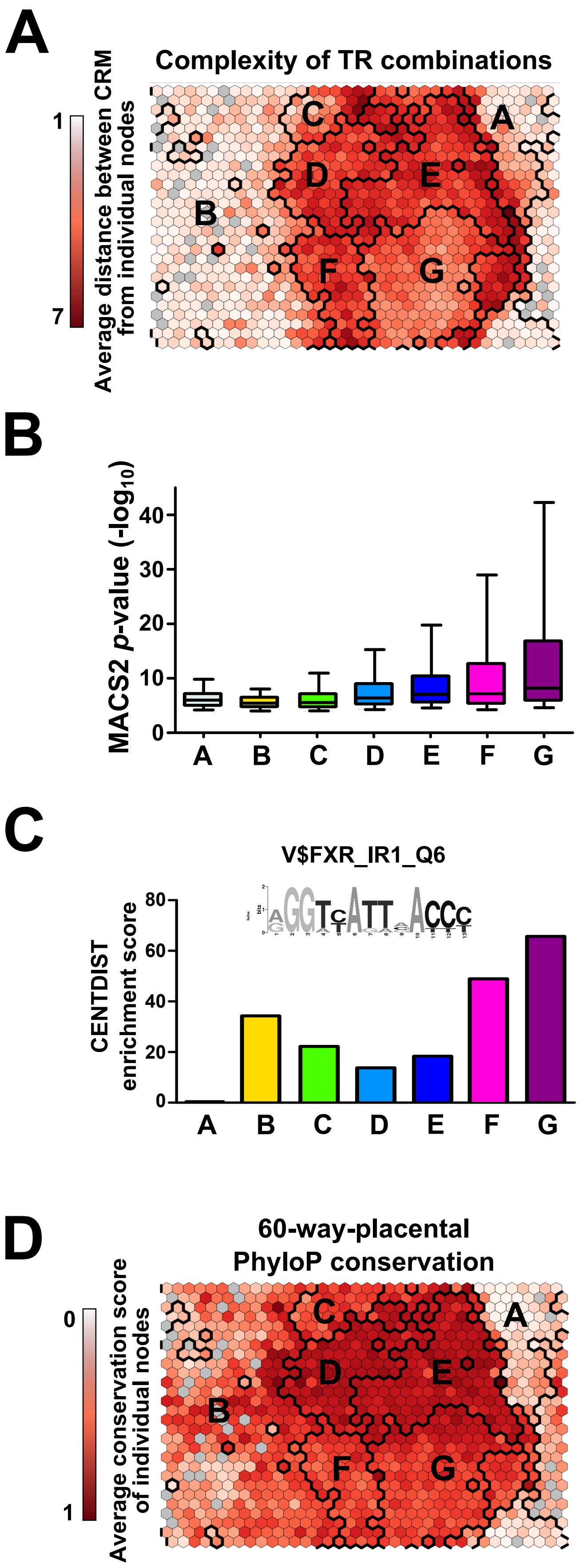
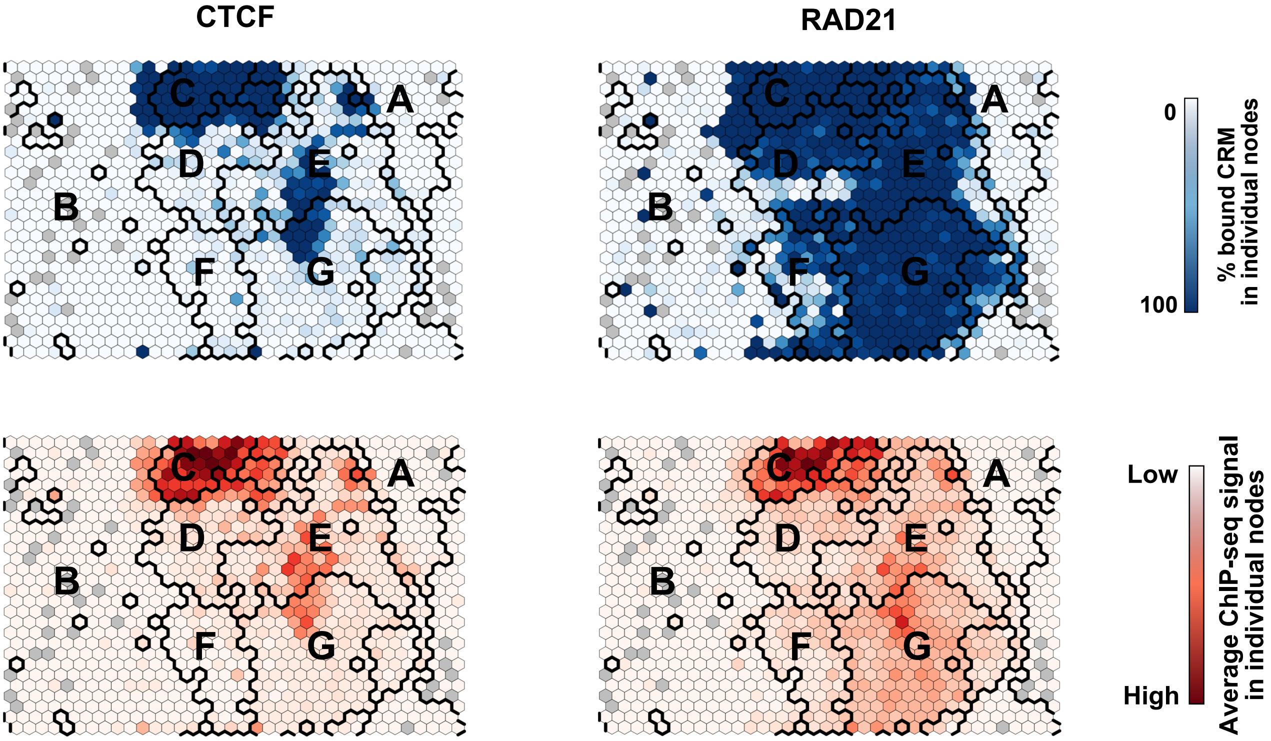

A
Complexity of TR combinations
Average distance between CRM from individual nodes
1
7
A
B
C
D
E
F
G
B
MACS2 p-value (-log10)
0
10
20
30
40
A
B
C
D
E
F
G
C
V$FXR_IR1_Q6
CENTDIST enrichment score
0
20
40
60
80
A
B
C
D
E
F
G
D
60-way-placental PhyloP conservation
Average conservation score of individual nodes
0
1
A
B
C
D
E
F
G

Fig.S5


CTCF
RAD21
C
A
D
E
B
F
G
0
100
% bound CRM
in individual nodes
Low
High
Average ChIP-seq signal
in individual nodes

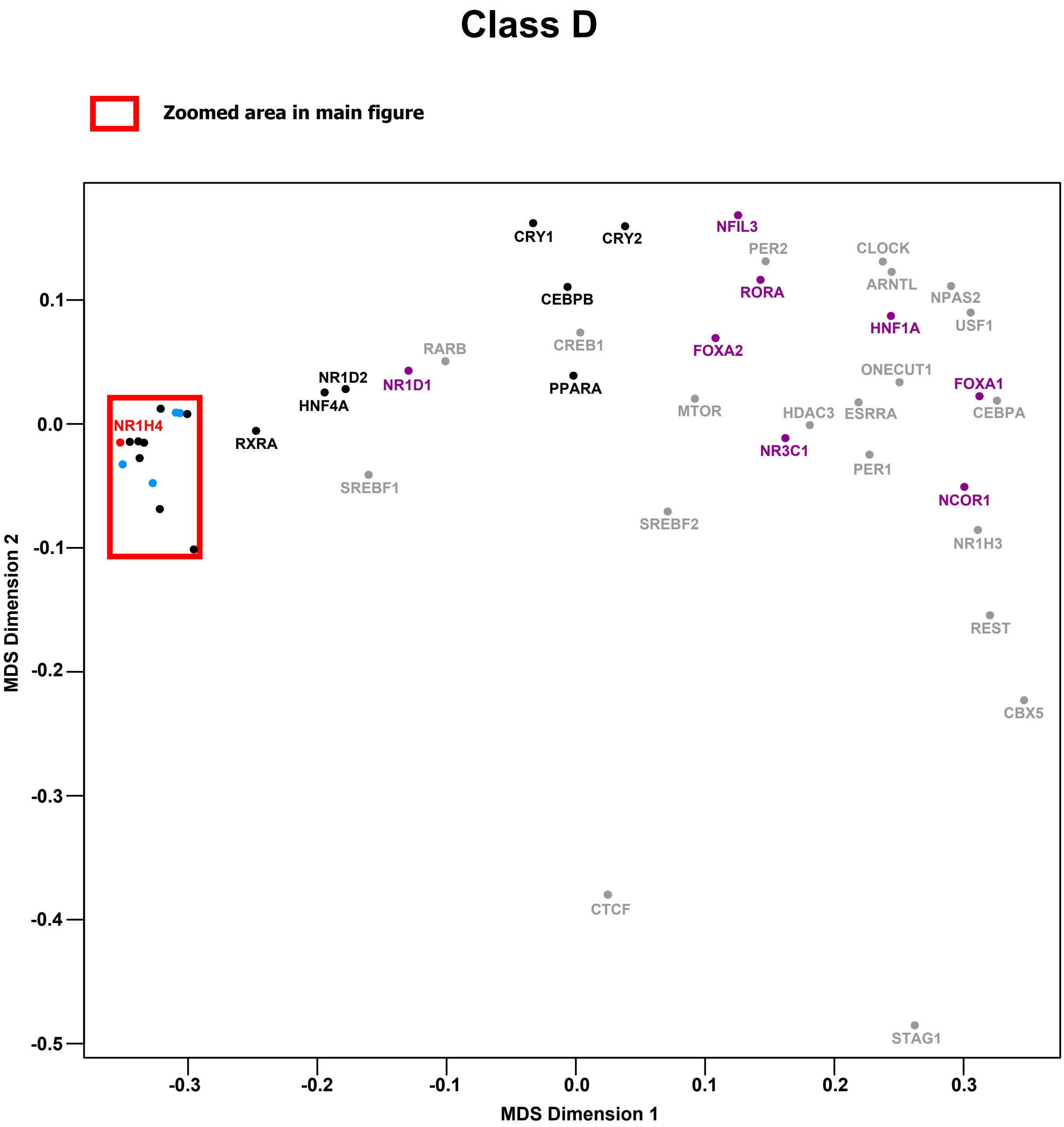

Class D
🟥 Zoomed area in main figure
CRY1
CRY2
NFIL3
PER2
CLOCK
ARNTL
CEBPB
RORA
NPAS2
HNF1A
USF1
RARB
CREB1
FOXA2
NR1D2
NR1D1
ONECUT1
HNF4A
PPARA
FOXA1
MTOR
HDAC3
ESRRA
CEBPA
NR1H4
RXRA
NR3C1
PER1
SREBF1
NCOR1
SREBF2
NR1H3
REST
CBX5
CTCF
STAG1
MDS Dimension 2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
MDS Dimension 1

Fig.S7

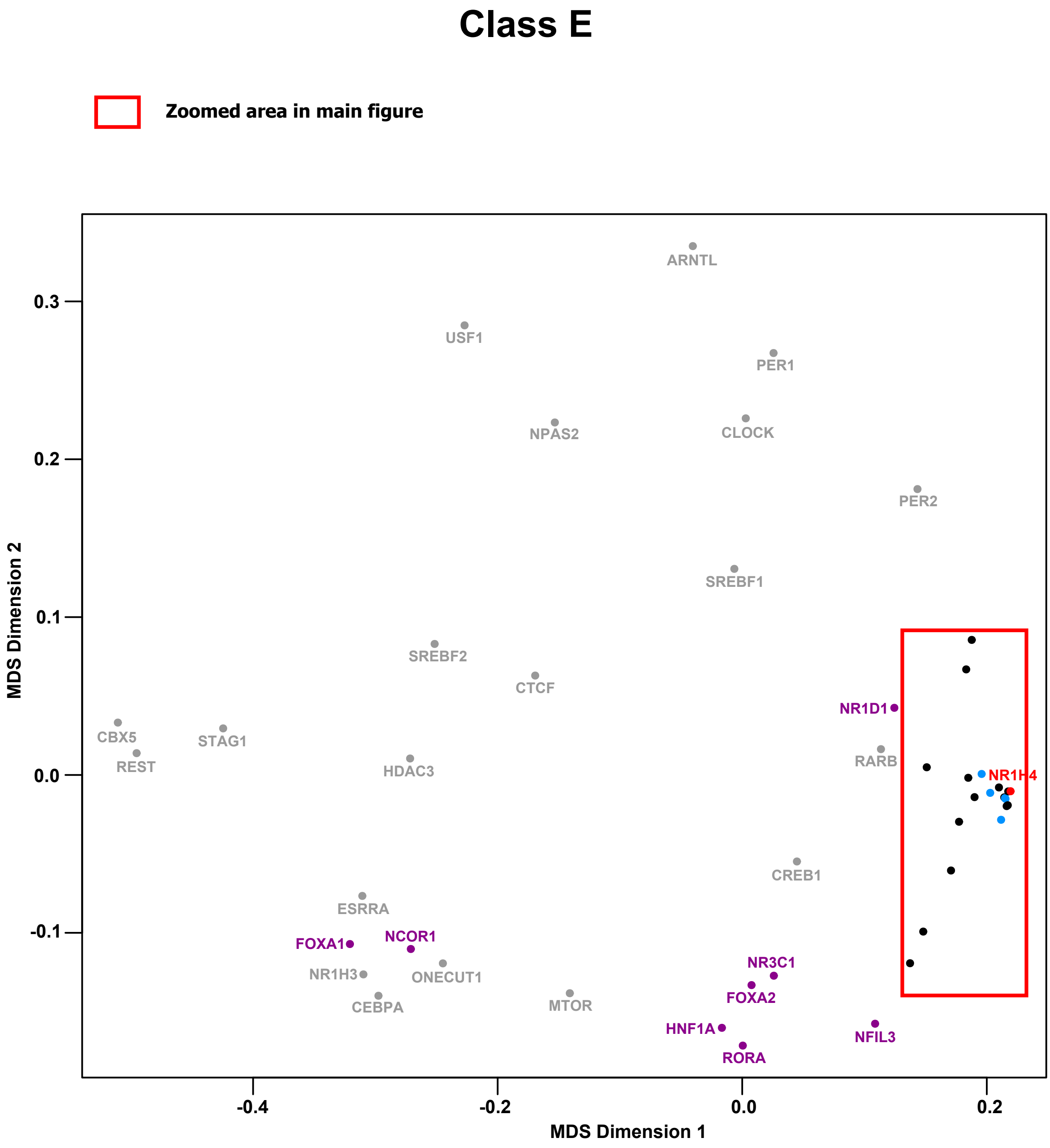

Class E
Zoomed area in main figure
ARNTL
USF1
PER1
NPAS2
CLOCK
PER2
SREBF1
SREBF2
CTCF
NR1D1
CBX5
STAG1
REST
HDAC3
RARB
NR1H4
CREB1
ESRRA
FOXA1
NCOR1
NR1H3
ONECUT1
CEBPA
MTOR
NR3C1
FOXA2
HNF1A
RORA
NFIL3
0.3
0.2
0.1
0.0
-0.1
MDS Dimension 2
-0.4
-0.2
0.0
0.2
MDS Dimension 1

Fig.S8

Class F

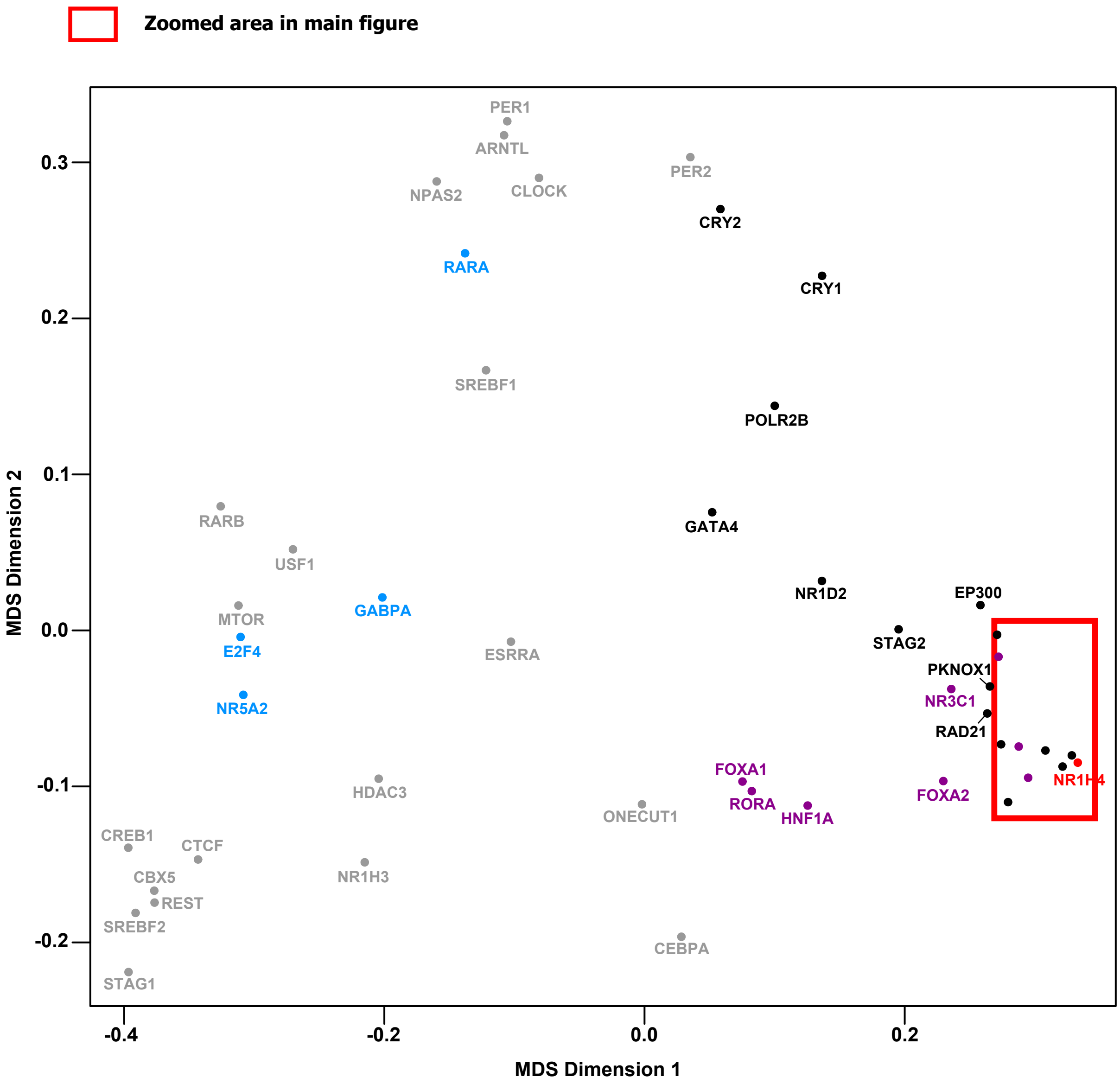


Fig.S9

Class G

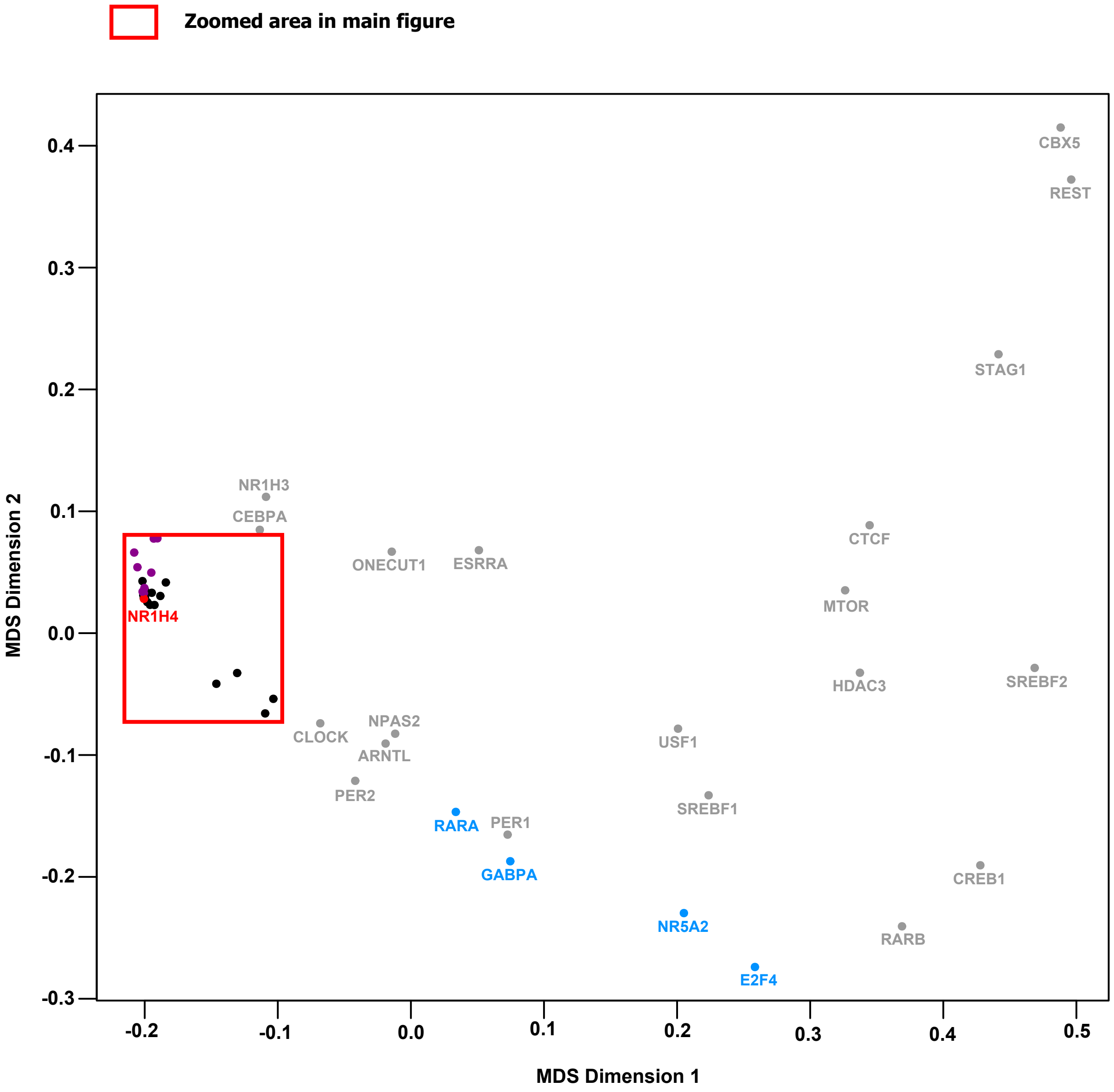


Fig.S10

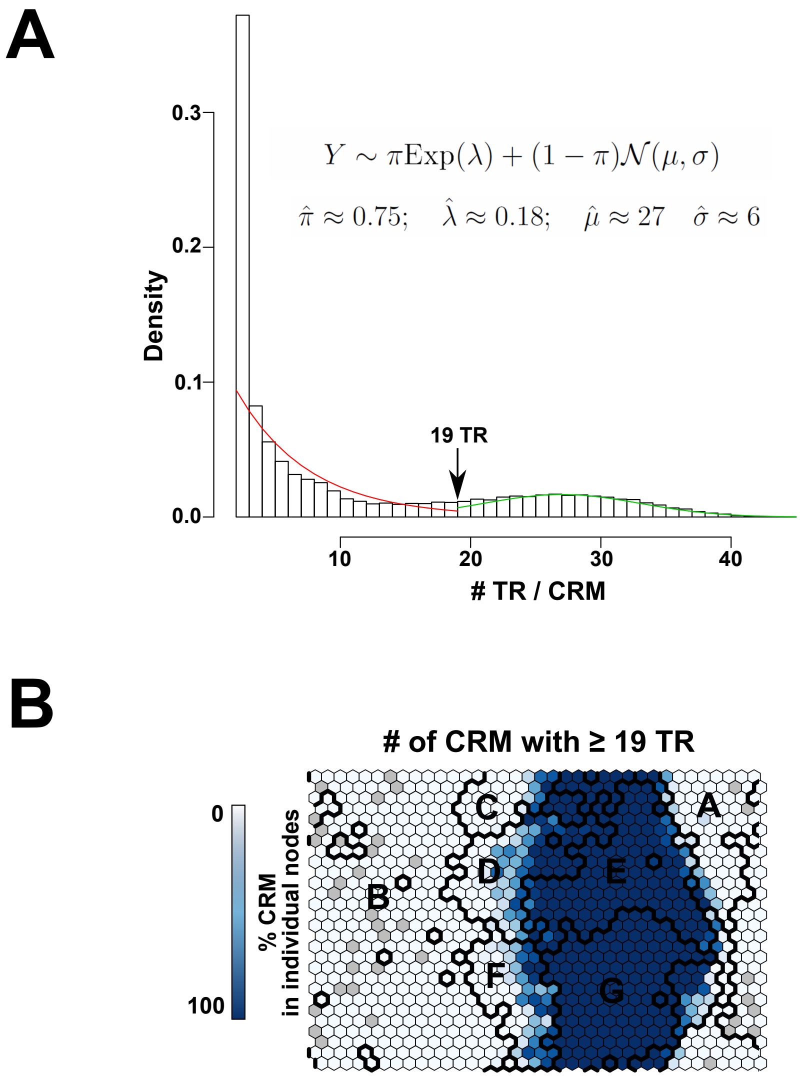

A
0.3
0.2
0.1
0.0
Density
$Y \sim \pi \mathrm{Exp}(\lambda) + (1-\pi)\mathcal{N}(\mu,\sigma)$
$\hat{\pi} \approx 0.75; \quad \hat{\lambda} \approx 0.18; \quad \hat{\mu} \approx 27 \quad \hat{\sigma} \approx 6$
19 TR
10
20
30
40
# TR / CRM
B
# of CRM with ≥ 19 TR
0
100
% CRM
in individual nodes
A
B
C
D
E
F
G

Fig.S11

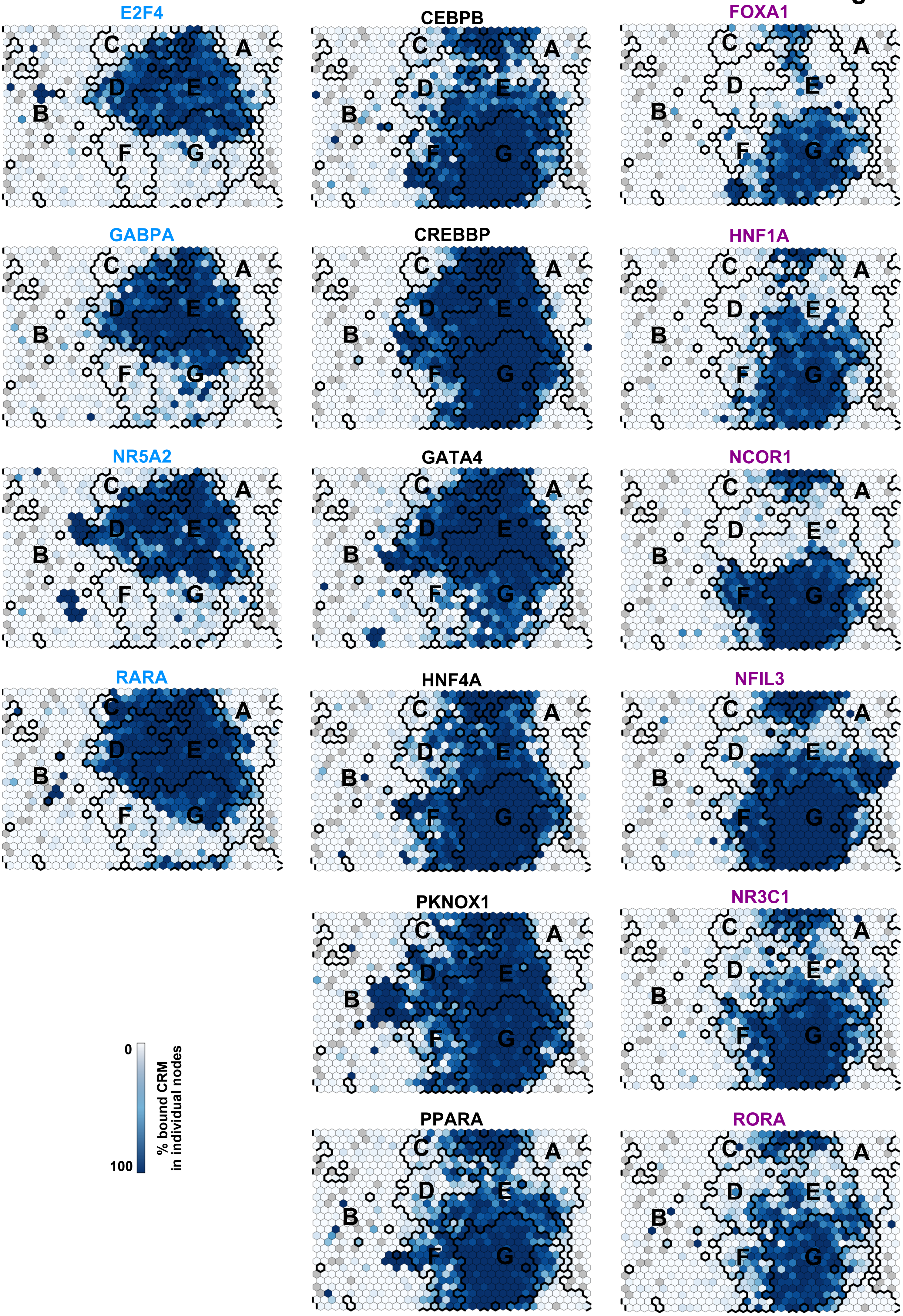

E2F4
CEBPB
FOXA1
GABPA
CREBBP
HNF1A
NR5A2
GATA4
NCOR1
RARA
HNF4A
NFIL3
PKNOX1
NR3C1
PPARA
RORA
A
B
C
D
E
F
G
0
100
% bound CRM
in individual nodes

Fig.S12

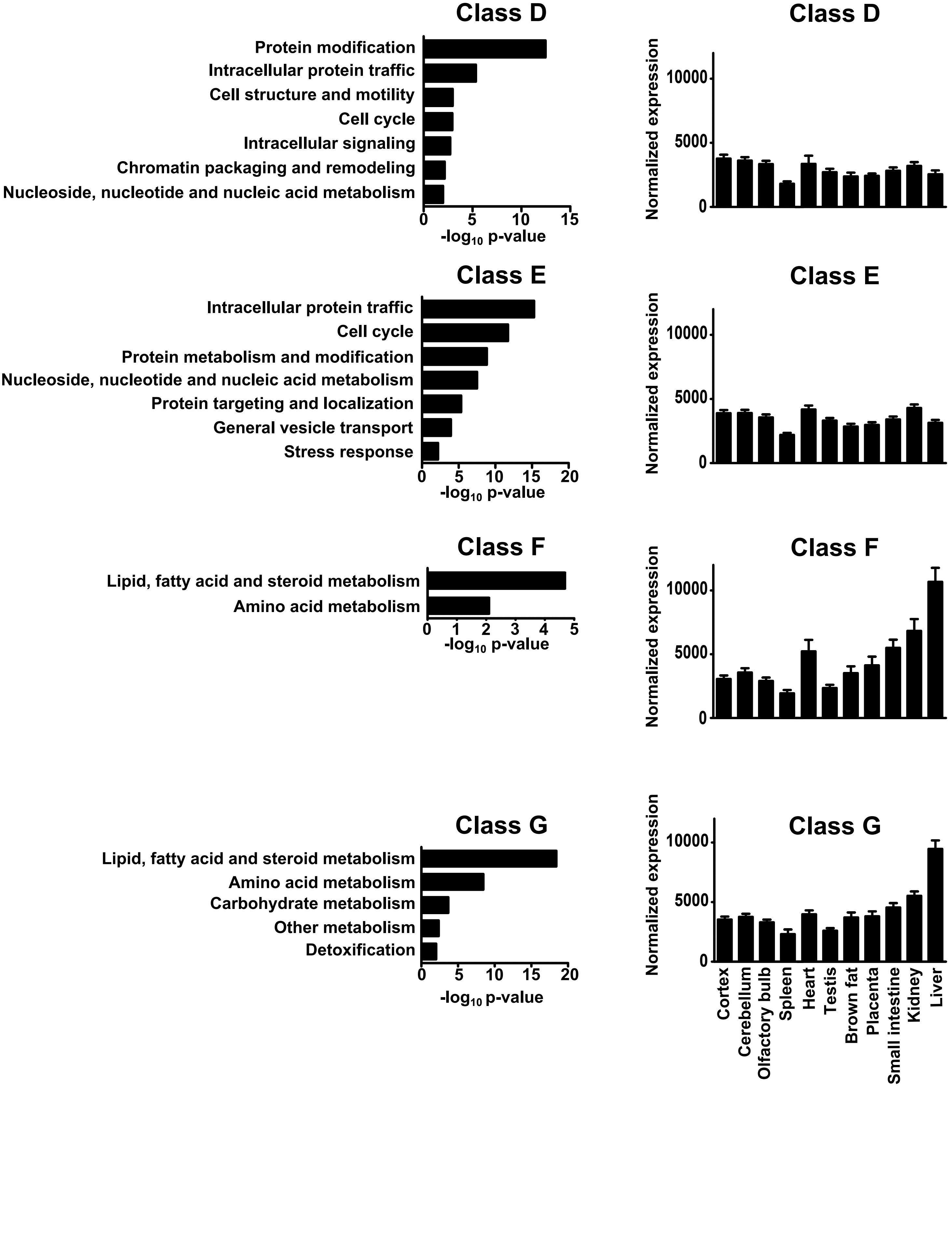

Class D
Protein modification
Intracellular protein traffic
Cell structure and motility
Cell cycle
Intracellular signaling
Chromatin packaging and remodeling
Nucleoside, nucleotide and nucleic acid metabolism
0
5
10
15
$-\log_{10}$ p-value
Class D
Normalized expression
10000
5000
0
Class E
Intracellular protein traffic
Cell cycle
Protein metabolism and modification
Nucleoside, nucleotide and nucleic acid metabolism
Protein targeting and localization
General vesicle transport
Stress response
0
5
10
15
20
$-\log_{10}$ p-value
Class E
Normalized expression
10000
5000
0
Class F
Lipid, fatty acid and steroid metabolism
Amino acid metabolism
0
1
2
3
4
5
$-\log_{10}$ p-value
Class F
Normalized expression
10000
5000
0
Class G
Lipid, fatty acid and steroid metabolism
Amino acid metabolism
Carbohydrate metabolism
Other metabolism
Detoxification
0
5
10
15
20
$-\log_{10}$ p-value
Class G
Normalized expression
10000
5000
0
Cortex
Cerebellum
Olfactory bulb
Spleen
Heart
Testis
Brown fat
Placenta
Small intestine
Kidney
Liver

Fig.S13

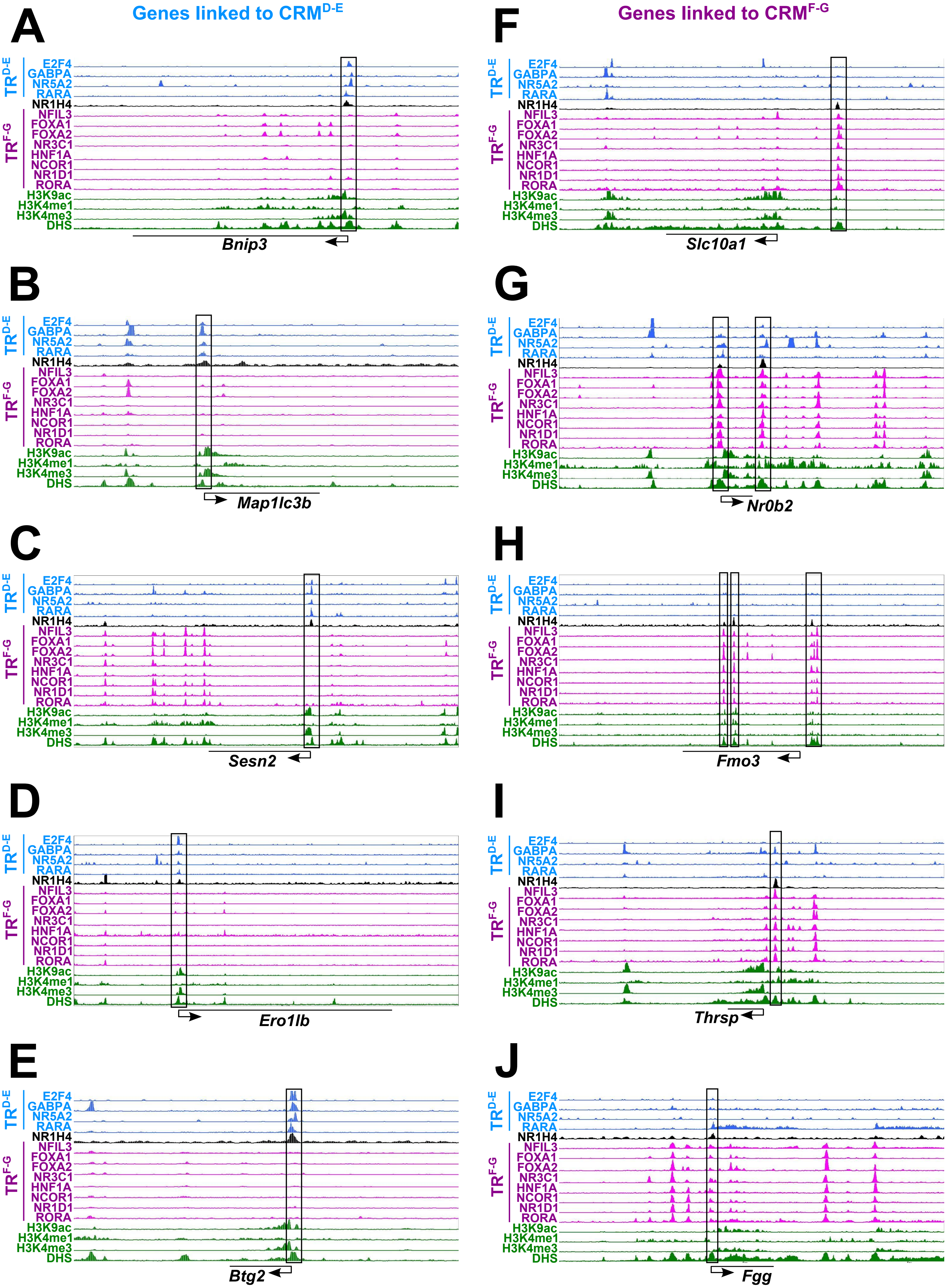

Genes linked to CRM^{D-E}
Genes linked to CRM^{F-G}
A
TR^{D-E}
E2F4
GABPA
NR5A2
RARA
NR1H4
TR^{F-G}
NFIL3
FOXA1
FOXA2
NR3C1
HNF1A
NCOR1
NR1D1
RORA
H3K9ac
H3K4me1
H3K4me3
DHS
Bnip3
B
Map1lc3b
C
Sesn2
D
Ero1lb
E
Btg2
F
Slc10a1
G
Nr0b2
H
Fmo3
I
Thrsp
J
Fgg

Fig.S13

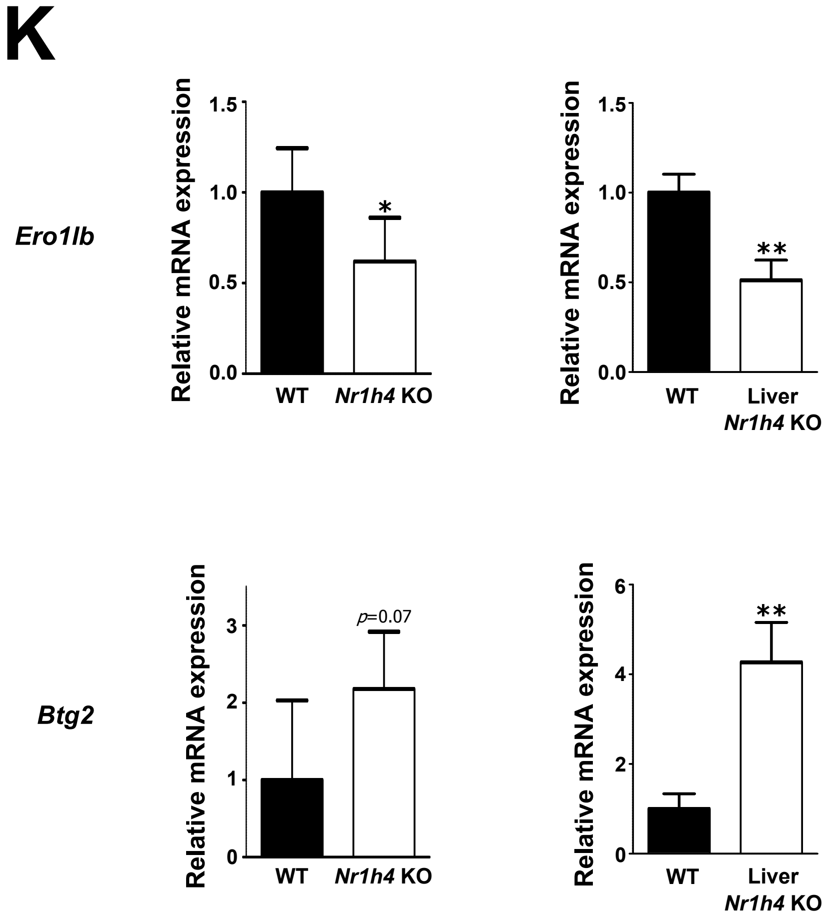

K
Ero1lb
Btg2
Relative mRNA expression
1.5
1.0
0.5
0.0
*
**
WT
Nr1h4 KO
Liver
Nr1h4 KO
3
2
1
0
p=0.07
6
4
2
0
**

Fig.S14

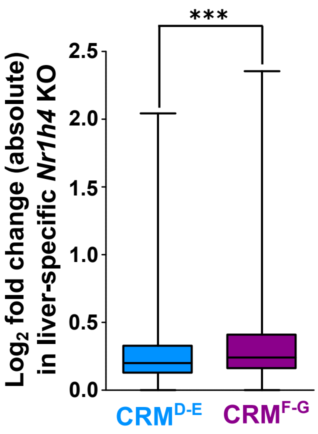

***
2.5
2.0
1.5
1.0
0.5
0.0
Log2 fold change (absolute) in liver-specific Nr1h4 KO
CRMD-E
CRMF-G

Fig.S15

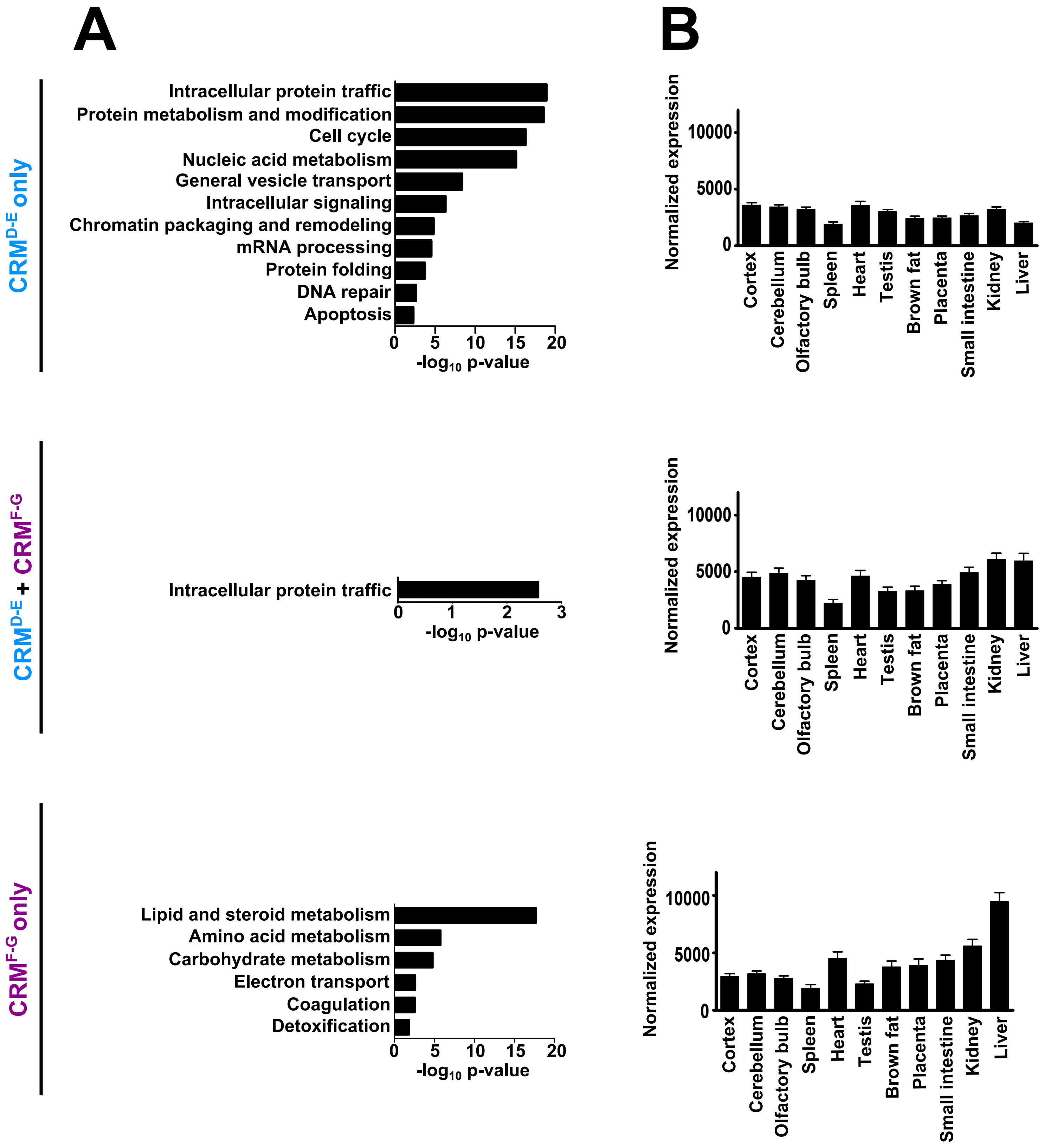
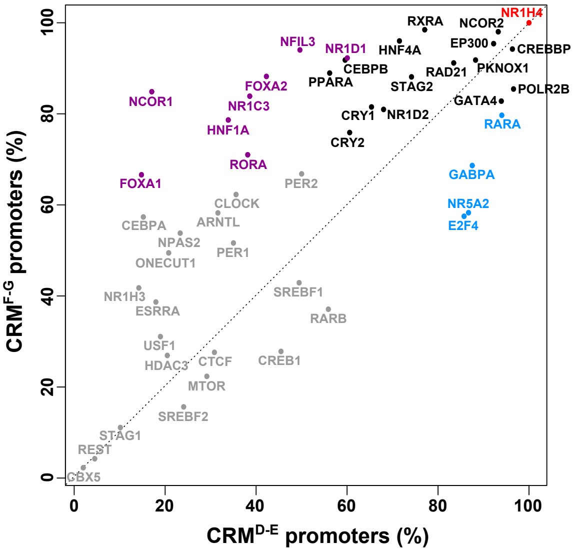

A
B
CRM^{D-E} only
Intracellular protein traffic
Protein metabolism and modification
Cell cycle
Nucleic acid metabolism
General vesicle transport
Intracellular signaling
Chromatin packaging and remodeling
mRNA processing
Protein folding
DNA repair
Apoptosis
0 5 10 15 20
-log$_{10}$ p-value
CRM^{D-E} + CRM^{F-G}
Intracellular protein traffic
0 1 2 3
-log$_{10}$ p-value
CRM^{F-G} only
Lipid and steroid metabolism
Amino acid metabolism
Carbohydrate metabolism
Electron transport
Coagulation
Detoxification
0 5 10 15 20
-log$_{10}$ p-value
Normalized expression
10000
5000
0
Cortex
Cerebellum
Olfactory bulb
Spleen
Heart
Testis
Brown fat
Placenta
Small intestine
Kidney
Liver

Fig.S16



NR1H4
RXRA
NCOR2
EP300
CREBBP
NFIL3
NR1D1
HNF4A
CEBPB
RAD21
PKNOX1
PPARA
STAG2
FOXA2
POLR2B
NCOR1
NR1C3
GATA4
CRY1
NR1D2
HNF1A
RARA
CRY2
RORA
GABPA
PER2
FOXA1
CLOCK
NR5A2
CEBPA
ARNTL
E2F4
NPAS2
PER1
ONECUT1
SREBF1
NR1H3
ESRRA
RARB
USF1
HDAC3
CTCF
CREB1
MTOR
SREBF2
STAG1
REST
CBX5
0
20
40
60
80
100
CRM^{F-G} promoters (%)
0
20
40
60
80
100
CRM^{D-E} promoters (%)

Fig.S17

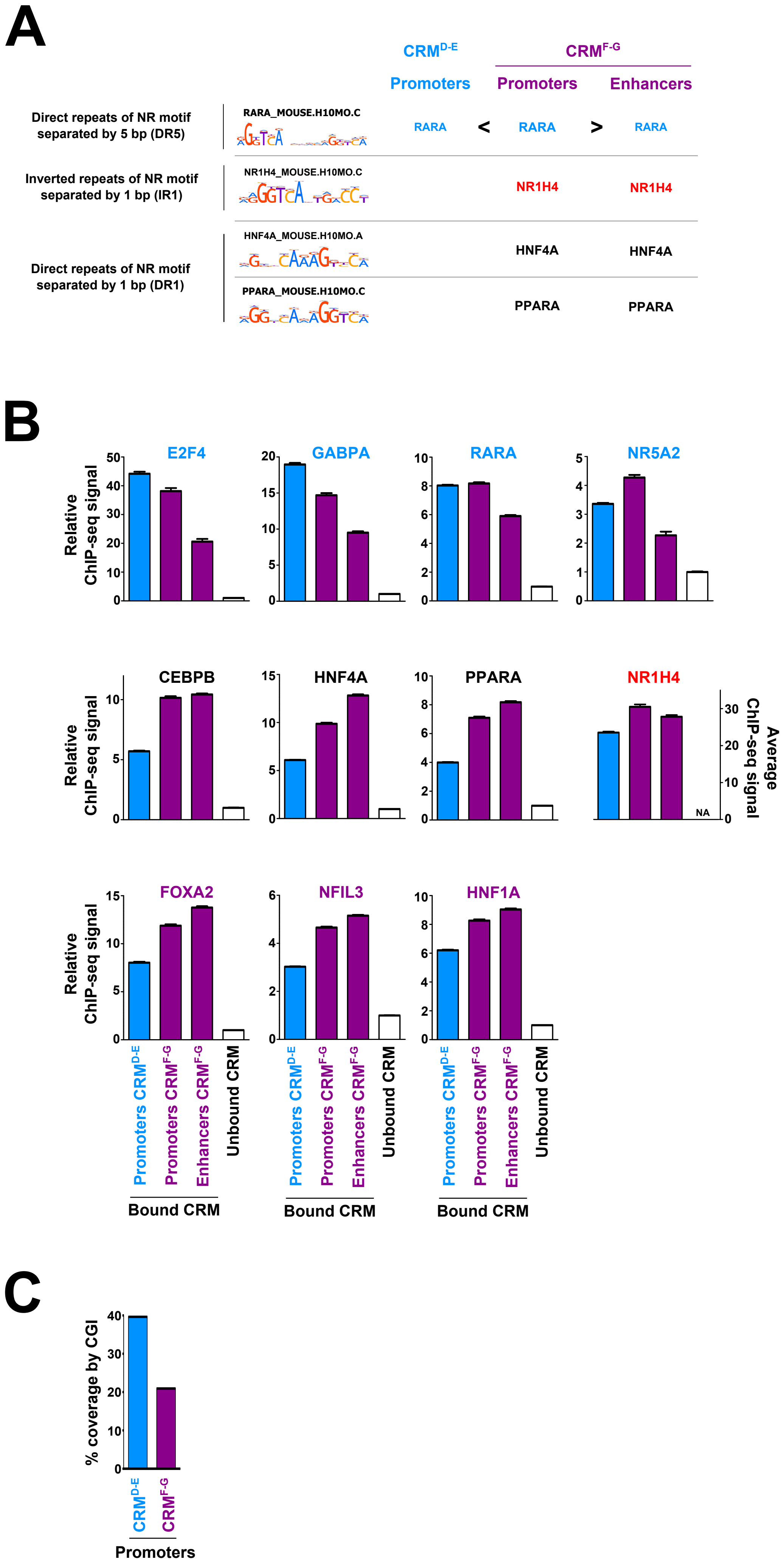

A
CRM^{D-E}
CRM^{F-G}
Promoters
Promoters
Enhancers
Direct repeats of NR motif separated by 5 bp (DR5)
RARA_MOUSE.H10MO.C
RARA < RARA > RARA
Inverted repeats of NR motif separated by 1 bp (IR1)
NR1H4_MOUSE.H10MO.C
NR1H4
NR1H4
Direct repeats of NR motif separated by 1 bp (DR1)
HNF4A_MOUSE.H10MO.A
HNF4A
HNF4A
PPARA_MOUSE.H10MO.C
PPARA
PPARA
B
Relative ChIP-seq signal
E2F4
0
10
20
30
40
50
GABPA
0
5
10
15
20
RARA
0
2
4
6
8
10
NR5A2
0
1
2
3
4
5
Relative ChIP-seq signal
CEBPB
0
5
10
HNF4A
0
5
10
15
PPARA
0
2
4
6
8
10
NR1H4
NA
0
10
20
30
Average ChIP-seq signal
Relative ChIP-seq signal
FOXA2
0
5
10
15
NFIL3
0
2
4
6
HNF1A
0
2
4
6
8
10
Promoters CRM^{D-E}
Promoters CRM^{F-G}
Enhancers CRM^{F-G}
Unbound CRM
Bound CRM
C
% coverage by CGI
0
10
20
30
40
CRM^{D-E}
CRM^{F-G}
Promoters

Fig.S18

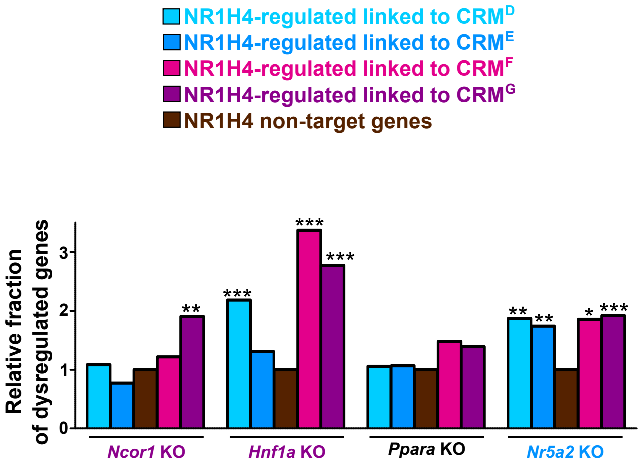

NR1H4-regulated linked to CRM^D
NR1H4-regulated linked to CRM^E
NR1H4-regulated linked to CRM^F
NR1H4-regulated linked to CRM^G
NR1H4 non-target genes
Relative fraction of dysregulated genes
0
1
2
3
**
***
***
***
**
**
*
***
Ncor1 KO
Hnf1a KO
Ppara KO
Nr5a2 KO

Fig.S19

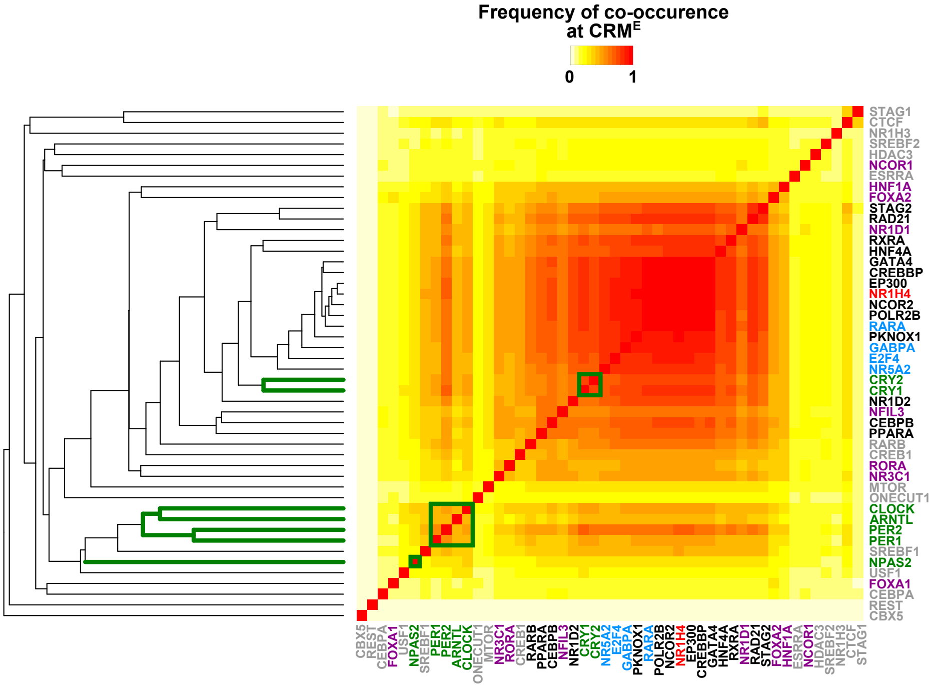
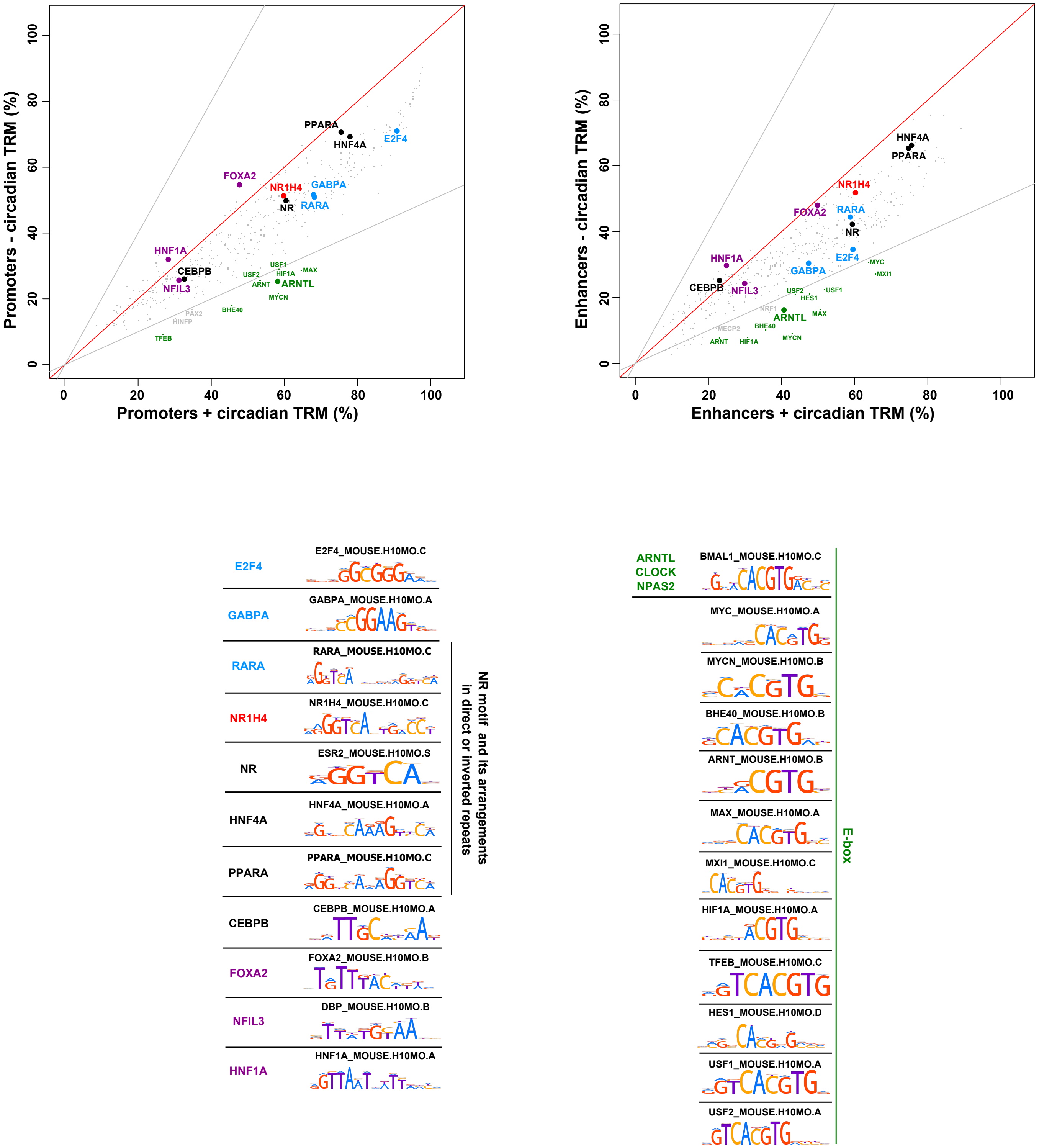

Frequency of co-occurence at CRM^E
0
1
STAG1
CTCF
NR1H3
SREBF2
HDAC3
NCOR1
ESRRA
HNF1A
FOXA2
STAG2
RAD21
NR1D1
RXRA
HNF4A
GATA4
CREBBP
EP300
NR1H4
NCOR2
POLR2B
RARA
PKNOX1
GABPA
E2F4
NR5A2
CRY2
CRY1
NR1D2
NFIL3
CEBPB
PPARA
RARB
CREB1
RORA
NR3C1
MTOR
ONECUT1
CLOCK
ARNTL
PER2
PER1
SREBF1
NPAS2
USF1
FOXA1
CEBPA
REST
CBX5
CBX5
REST
CEBPA
FOXA1
USF1
NPAS2
SREBF1
PER1
PER2
ARNTL
CLOCK
ONECUT1
MTOR
NR3C1
RORA
CREB1
RARB
PPARA
CEBPB
NFIL3
NR1D2
CRY1
CRY2
NR5A2
E2F4
GABPA
PKNOX1
RARA
POLR2B
NCOR2
NR1H4
EP300
CREBBP
GATA4
HNF4A
RXRA
NR1D1
RAD21
STAG2
FOXA2
HNF1A
ESRRA
NCOR1
HDAC3
SREBF2
NR1H3
CTCF
STAG1

Fig.S20



Promoters - circadian TRM (%)
Promoters + circadian TRM (%)
0
20
40
60
80
100
PPARA
HNF4A
E2F4
FOXA2
NR1H4
GABPA
NR
RARA
HNF1A
CEBPB
NFIL3
USF1
USF2
HIF1A
MAX
ARNT
ARNTL
MYCN
BHE40
PAX2
HINFP
TFEB
Enhancers - circadian TRM (%)
Enhancers + circadian TRM (%)
HNF4A
PPARA
NR1H4
FOXA2
RARA
NR
HNF1A
GABPA
E2F4
MYC
MXI1
CEBPB
NFIL3
USF2
HES1
USF1
NRF1
ARNTL
MAX
MECP2
BHE40
ARNT
HIF1A
MYCN
E2F4
E2F4_MOUSE.H10MO.C
GABPA
GABPA_MOUSE.H10MO.A
RARA
RARA_MOUSE.H10MO.C
NR1H4
NR1H4_MOUSE.H10MO.C
NR
ESR2_MOUSE.H10MO.S
HNF4A
HNF4A_MOUSE.H10MO.A
PPARA
PPARA_MOUSE.H10MO.C
CEBPB
CEBPB_MOUSE.H10MO.A
FOXA2
FOXA2_MOUSE.H10MO.B
NFIL3
DBP_MOUSE.H10MO.B
HNF1A
HNF1A_MOUSE.H10MO.A
NR motif and its arrangements in direct or inverted repeats
ARNTL
CLOCK
NPAS2
BMAL1_MOUSE.H10MO.C
MYC_MOUSE.H10MO.A
MYCN_MOUSE.H10MO.B
BHE40_MOUSE.H10MO.B
ARNT_MOUSE.H10MO.B
MAX_MOUSE.H10MO.A
MXI1_MOUSE.H10MO.C
HIF1A_MOUSE.H10MO.A
TFEB_MOUSE.H10MO.C
HES1_MOUSE.H10MO.D
USF1_MOUSE.H10MO.A
USF2_MOUSE.H10MO.A
E-box

Fig.S21

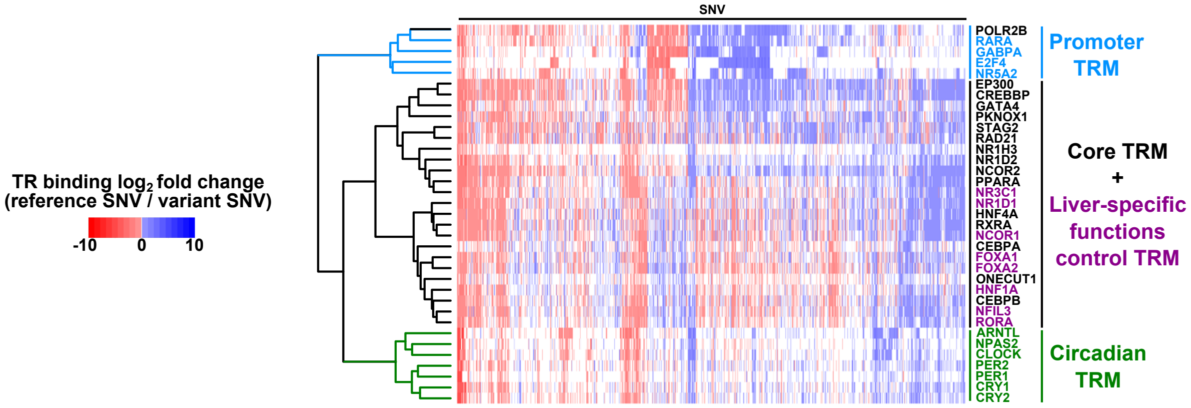

TR binding log$_2$ fold change
(reference SNV / variant SNV)
-10
0
10
SNV
POLR2B
RARA
GABPA
E2F4
NR5A2
EP300
CREBBP
GATA4
PKNOX1
STAG2
RAD21
NR1H3
NR1D2
NCOR2
PPARA
NR3C1
NR1D1
HNF4A
RXRA
NCOR1
CEBPA
FOXA1
FOXA2
ONECUT1
HNF1A
CEBPB
NFIL3
RORA
ARNTL
NPAS2
CLOCK
PER2
PER1
CRY1
CRY2
Promoter TRM
Core TRM
+
Liver-specific functions control TRM
Circadian TRM

Fig.S22

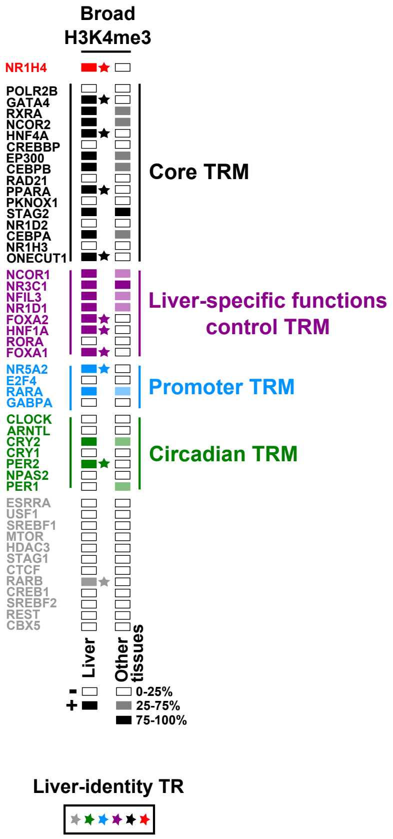

Broad
H3K4me3
NR1H4
POLR2B
GATA4
RXRA
NCOR2
HNF4A
CREBBP
EP300
CEBPB
RAD21
PPARA
PKNOX1
STAG2
NR1D2
CEBPA
NR1H3
ONECUT1
Core TRM
NCOR1
NR3C1
NFIL3
NR1D1
FOXA2
HNF1A
RORA
FOXA1
Liver-specific functions control TRM
NR5A2
E2F4
RARA
GABPA
Promoter TRM
CLOCK
ARNTL
CRY2
CRY1
PER2
NPAS2
PER1
Circadian TRM
ESRRA
USF1
SREBF1
MTOR
HDAC3
STAG1
CTCF
RARB
CREB1
SREBF2
REST
CBX5
Liver
Other tissues
-
+
0-25%
25-75%
75-100%
Liver-identity TR

Fig.S23

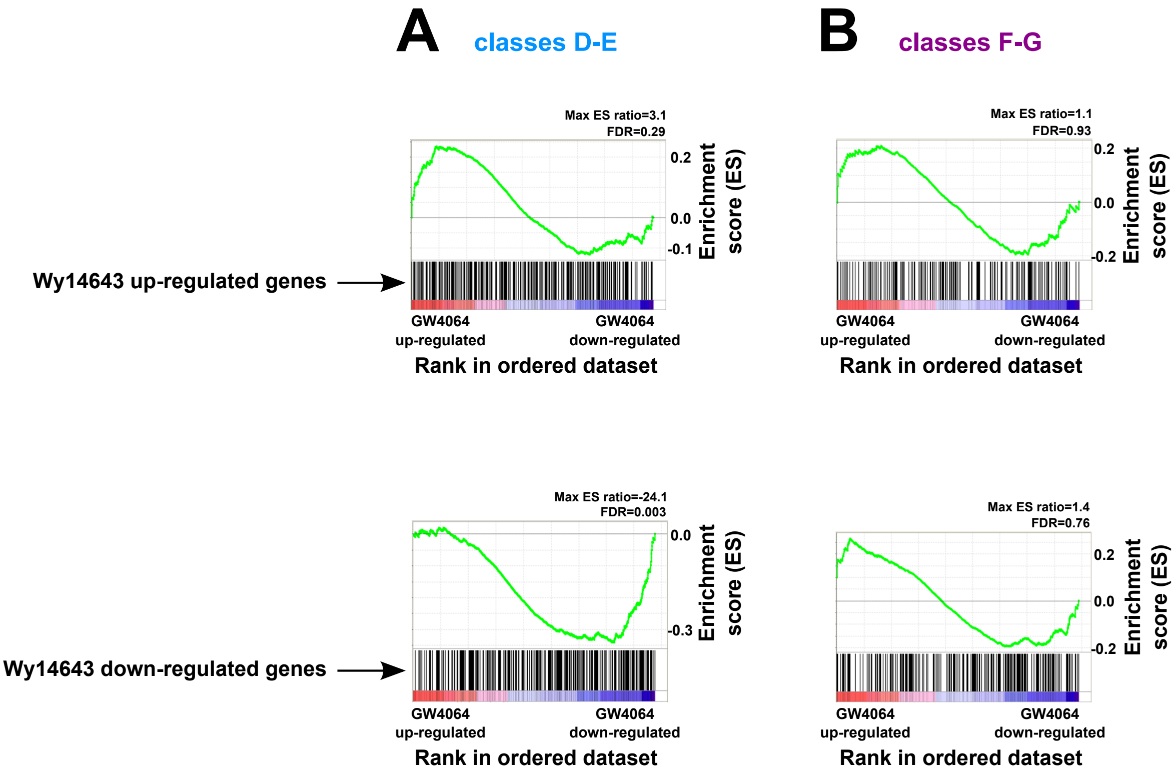

A
classes D-E
B
classes F-G
Max ES ratio=3.1
FDR=0.29
0.2
0.0
-0.1
Enrichment score (ES)
Max ES ratio=1.1
FDR=0.93
0.2
0.0
-0.2
Enrichment score (ES)
Wy14643 up-regulated genes
GW4064 up-regulated
GW4064 down-regulated
Rank in ordered dataset
Max ES ratio=-24.1
FDR=0.003
0.0
-0.3
Enrichment score (ES)
Max ES ratio=1.4
FDR=0.76
0.2
0.0
-0.2
Enrichment score (ES)
Wy14643 down-regulated genes
GW4064 up-regulated
GW4064 down-regulated
Rank in ordered dataset

Fig.S24

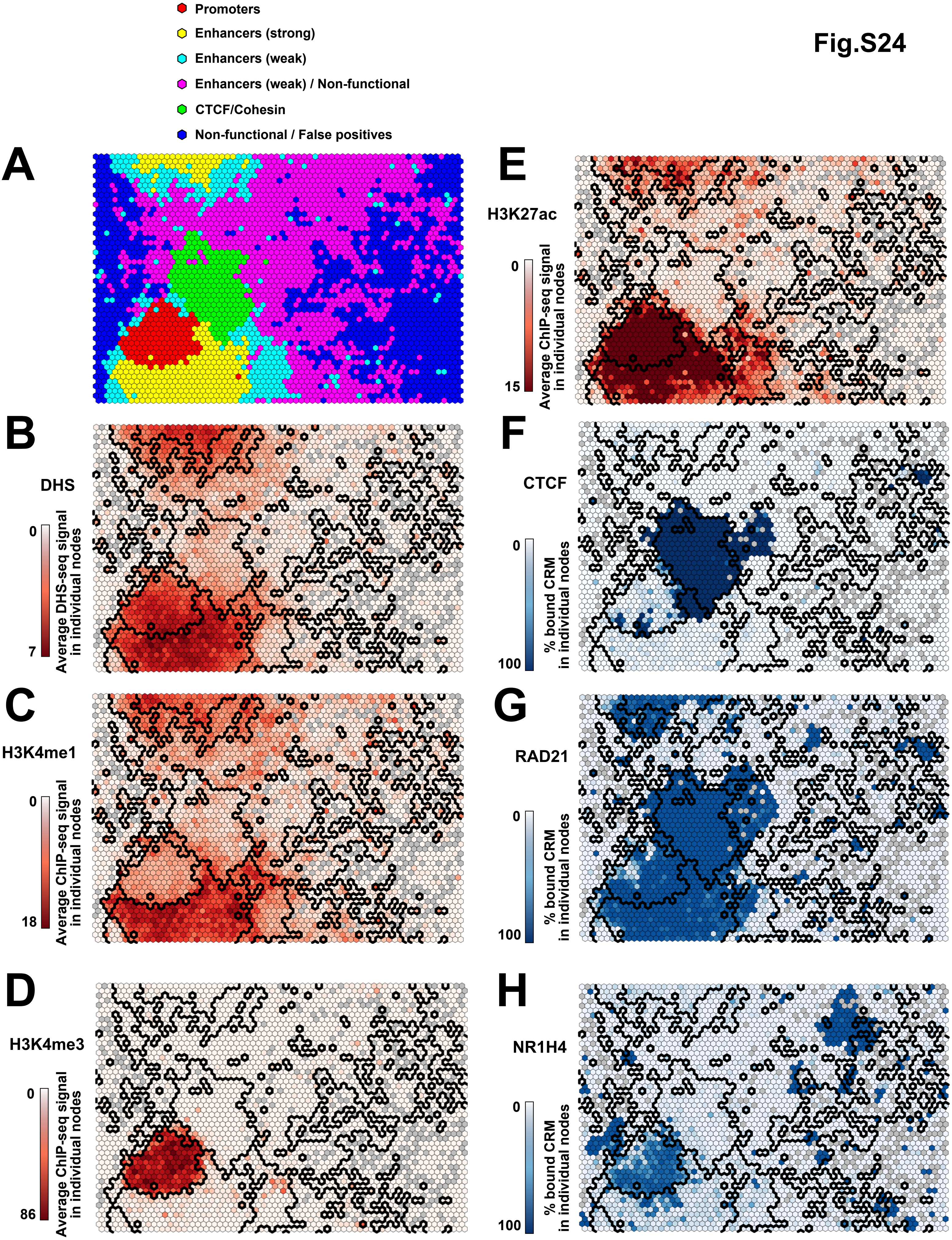
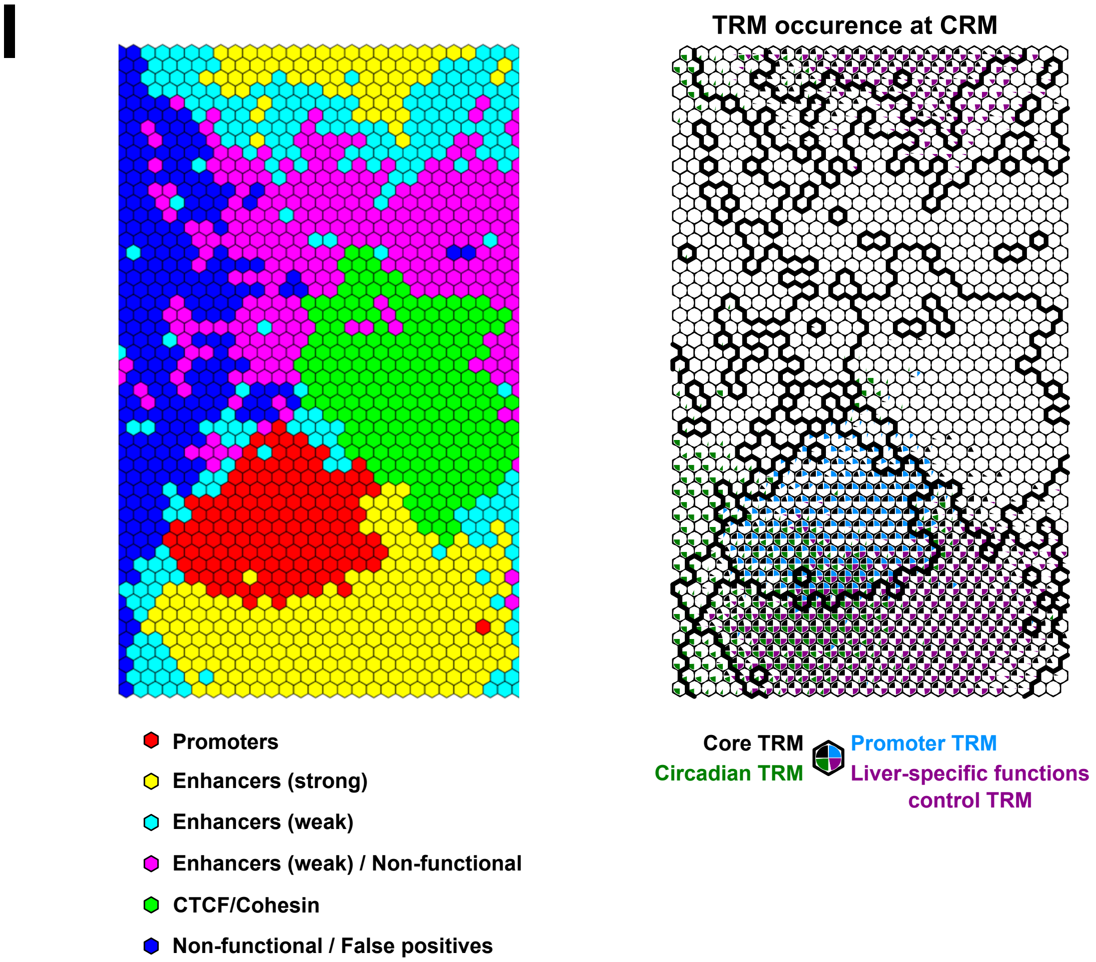

Promoters
Enhancers (strong)
Enhancers (weak)
Enhancers (weak) / Non-functional
CTCF/Cohesin
Non-functional / False positives
A
B
DHS
0
7
Average DHS-seq signal in individual nodes
C
H3K4me1
0
18
Average ChIP-seq signal in individual nodes
D
H3K4me3
0
86
Average ChIP-seq signal in individual nodes
E
H3K27ac
0
15
Average ChIP-seq signal in individual nodes
F
CTCF
0
100
% bound CRM in individual nodes
G
RAD21
0
100
% bound CRM in individual nodes
H
NR1H4
0
100
% bound CRM in individual nodes

Fig.S24


I
TRM occurence at CRM
Promoters
Enhancers (strong)
Enhancers (weak)
Enhancers (weak) / Non-functional
CTCF/Cohesin
Non-functional / False positives
Core TRM
Promoter TRM
Circadian TRM
Liver-specific functions control TRM