



Organizing combinatorial transcription factor recruitment at cis -regulatory modules

Julie Dubois Dubois-Chevalier, Parisa Mazrooei, Mathieu Lupien, Bart Staels,
Philippe Lefebvre, Jérôme Eeckhoutte

► To cite this version:

Julie Dubois Dubois-Chevalier, Parisa Mazrooei, Mathieu Lupien, Bart Staels, Philippe Lefebvre, et al.. Organizing combinatorial transcription factor recruitment at cis -regulatory modules. Transcription, 2018, 9 (4), pp.233-239. 10.1080/21541264.2017.1394424 . inserm-02153895

HAL Id: inserm-02153895

<https://inserm.hal.science/inserm-02153895>

Submitted on 12 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Organizing combinatorial transcription factor recruitment at *cis*-regulatory modules

Julie Dubois-Chevalier ¹, Parisa Mazrooei ², Mathieu Lupien ², Bart Staels ¹, Philippe Lefebvre ^{1,§}, Jérôme Eeckhoutte ^{1,§,*}

¹ Univ. 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, ON M5G 1L7, Canada

§ Joint senior authors

* Corresponding author:

Dr. Jérôme Eeckhoutte

INSERM UMR 1011-Bâtiment J&K

Université Lille 2; Faculté de Médecine de Lille-Pôle Recherche

Boulevard du Professeur Leclerc

59045 Lille cedex

France

Tel: 33(0)3 20 97 42 20 / Fax: 33(0)3 20 97 42 01

E-mail: jerome.eeckhoutte@inserm.fr

Keywords

Chromatin binding, *cis*-regulatory modules, DNA recognition elements, functional genomics, liver, transcription factors, *trans*-regulatory modules, single nucleotide variants

Abstract

Gene transcriptional regulation relies on *cis*-regulatory DNA modules (CRMs), which serve as nexus sites for integration of multiple transcription factor (TF) activities. Here, we provide evidence and discuss recent literature indicating that TF recruitment to CRMs is organized into combinations of *trans*-regulatory protein modules (TRMs). We propose that TRMs are functional entities composed of TFs displaying the most highly interdependent chromatin binding which are, in addition, able to modulate their recruitment to CRMs through inter-TRM effects. These findings shed light on the architectural organization of TF recruitment encoded by their recognition motifs within CRMs.

Gene expression regulation involves combinatorial recruitment of multiple transcription factors (TFs) organized into modules at transcriptional regulatory regions

Transcriptional regulation is one of the main control steps involved in specifying appropriate gene expression. Cells rely on an array of transcription factors (TFs), which total several hundreds in higher Eukaryotes ¹, to finely tune gene expression. This is achieved through concomitant regulation of a given gene by multiple TFs, which combine activities at *cis*-regulatory modules (CRMs) including promoters and enhancers ². Combinatorial regulation of gene expression by multiple TFs allows cells to handle the challenging task of controlling their transcriptome with specificity, robustness and adaptability to environmental signals. For instance, the concentration of TF binding sites or their DNA recognition motifs [hereafter called response elements (REs)] is highly predictive of strong enhancer activity ³. In this context, understanding how recruitment and activities of multiple TFs are functionally orchestrated at CRMs constitutes fundamental questions in the field of transcriptional regulation.

It has been well established, in different biological systems and using a compendium of various technical approaches, that TF co-recruitment to CRMs often does not lead to a mere additive effect but actually implies functional interferences between TF activities. These interferences can be positive or negative and are exerted through direct or indirect mechanisms ^{4,5}. Direct effects are mediated through TF-TF interactions, which can modulate TF functions (cofactor recruitment, DNA binding) through steric hindrance or induced conformational changes ⁶. Competition for REs has also been reported ⁷. Indirect functional interference between TFs often involves modulation of their binding to CRMs through chromatin-based mechanisms ^{8,9}. Indeed, by remodelling the local chromatin structure through nucleosome positioning and/or epigenetic modifications (of DNA and/or histones), a given TF can indirectly modulate recruitment to CRMs of another TF. In this context, hierarchical

recruitment of TF to chromatin has been reported, where pioneer TFs bind to and remodel condensed chromatin to allow subsequent binding of subordinate TFs ^{10,11}. In line with TF interdependent chromatin binding, it has been reported that allelic imbalance in occupancy or changes in a given TF genomic recruitment across evolution can not be entirely explained by genetic variation occurring directly in its RE ^{12,13}.

Understanding how RE organization within CRMs translates into concerted TF recruitment has relevance not only for our basic knowledge of CRM evolution and transcriptional regulatory activities, but is also directly relevant to human diseases. Indeed, changes in TF binding to chromatin and CRM activities trigger abnormal transcriptomes such as those linked to enhancer reprogramming in cancer ¹⁴. Moreover, many single nucleotide variants (SNVs) which modulate traits or disease occurrence have been shown or predicted to impact REs, TF DNA binding domains and/or TF chromatin binding ^{15–17}.

Using mouse liver gene transcriptional regulation as a model system, we have recently described that different subsets of TFs (and cofactors) show preferential and interdependent co-binding to CRMs, thereby defining *trans*-regulatory modules (TRMs) ¹⁸. Modular combinatorial binding of TFs has also been inferred from hundreds of TF binding data from the K562 cell-line ¹⁹. In the liver, in addition to a core set of TFs broadly bound to liver CRMs, we found that liver-specific gene transcriptional regulation involves the combinatorial recruitment of 3 types of TRMs comprising a promoter TRM, a TRM specifically and broadly controlling hepatic functions and a facultative circadian TRM (Fig.1A). We also confirmed that the presence of REs directly recognized by TFs from a given TRM was, in most instances, instructive for its presence at CRMs.

Interdependent chromatin recruitment of TFs within liver TRMs

In our study, organization of TF chromatin recruitment into combinations of TRMs was further indicated by results from intra-genomic replicate (IGR) analyses^{20,21}. The IGR tool predicts the impact of SNVs on TF chromatin binding by comparing the average TF ChIP-seq signal intensity across the genomic loci that contains the reference or the variant allele of each SNV (within underlying sequences of 7-mers)^{20,21}. This allowed to define that SNVs concomitantly and preferentially impact chromatin binding of TFs comprised within the same TRM¹⁸. To define which REs were impacted by the analyzed SNVs, we used PERFECTOS-APE²². This tool first identified REs overlapping our SNVs of interest and then predicted how these SNVs may alter TF binding affinity to these overlapping REs. The obtained results were then used to define how changes in RE affinity correlate with changes in TF binding to chromatin previously defined by IGR (Fig.1B). Hierarchical clustering of the data identified clusters of REs positively correlated with clusters of TFs, which we define as blocks (Fig.1C). For instance, blocks linking TFs from the promoter TRM to E2F/EGR/ETS motifs, the TRM specifically controlling hepatic functions to FOX/CEBP/PAR-bZIP/Homeobox (HNF6) motifs and the circadian TRM to E-boxes were highlighted by these analyses (framed in Fig.1C; see Table S1 for a full list of motifs). These analyses predicted that, in addition to being required for recruitment of their cognate TFs, individual REs are also often important for the co-recruitment of additional TFs within TRMs. This suggests a lack of strong hierarchy in TF recruitment within TRMs where many individual TFs can modulate recruitment of the broad set of TFs these modules comprise. An additional block corresponding to many, but not all, nuclear receptors (NRs) and their REs was also identified in Fig.1C. This NR block comprised several NRs from the core set of TFs which are essential regulators of liver functions (RXRA, HNF4A, PPARA, NR1D2/REV-ERB beta). Of note, some other NRs showed an opposite correlation between predicted SNV-induced changes in NR RE affinity and their chromatin binding (NR1H3/LXR alpha and NR1H4/FXR). These

results are in line with a previous study showing that PPARA and NR1H3/LXR alpha bind to overlapping sites in a mutually exclusive manner in the mouse liver ⁷. This also further indicates extensive NR crosstalk at shared liver CRMs where common REs fine tune chromatin binding of many NR family members. NR5A2/LRH-1 and RARA, which displayed poor correlation with NR RE, may be recruited to promoters through other means such as tethering ¹⁸.

Liver TRMs show different levels of mutual interference in chromatin recruitment

A lack of correlation (Fig.1C) does not however imply that SNVs modulating RE affinity do not impact on TF chromatin binding because milder effects may be more difficult to call with statistical significance or context-specific effects could lead to divergent consequences on TF binding at different CRMs. Therefore, we next compiled SNVs modulating RE affinity from each cluster (1 to 11) and monitored the fraction predicted to modulate binding of each TF. This revealed that most clusters of motifs from Fig.1C were predicted to modulate recruitment of TFs outside from the TRM comprising their cognate TF (Fig.1D). For example, SNVs modulating NR RE affinity (clusters 10-11) broadly modulate chromatin binding of other TFs while NR chromatin binding is in return affected by changes in RE affinity from most other clusters (Fig.1D). The lack of correlation (Fig.1C) is most probably to be ascribed to weaker and more complex indirect effects when SNVs influence chromatin binding of TFs from a different TRM. Importantly, reciprocal influences are not always symmetrical and are even absent in some instances. For example, SNVs in motifs from cluster 1 modulate binding of TFs defining the promoter TRM (RARA, NR5A2, GABPA, E2F4), but not that of TFs from the liver-specific functions control TRM (HNF1A, FOXA2, RORA, NFIL3, FOXA1), while the opposite was true for SNVs in motifs from

clusters 6 to 9 (Fig.1D). This suggests that recruitment of these 2 TRMs is functionally independent at liver CRMs.

Discussion and perspectives

Altogether, the data presented in ¹⁸ and herein point to a hierarchical model of TF interdependent chromatin recruitment where the most highly interdependent TFs define TRMs which can secondarily impact on each other (Fig.2).

This suggests that TRMs consist of sets of TFs whose chromatin binding is influenced as a bulk by individual REs. Indeed, while not all show the same importance, our data indicate a broad array of REs/TFs is able to influence a given TRM recruitment. A hierarchy with a few specific REs/TFs, such as pioneer TFs, controlling recruitment of other TFs from the same TRM could not be clearly identified. Recent studies have indicated that pioneer TF chromatin binding can be modulated by surrogate TFs ^{23,24}. For instance, during pluripotency reprogramming by Oct4, Sox2, Klf4, and c-Myc (O,S,K and M), OSK act as pioneer TFs for M which, in return, stabilizes OSK chromatin recruitment ²³. Hence, the hierarchy of TF recruitment enabling CRM activation during cell differentiation or reprogramming is most probably largely lost once cells have acquired a new stable state where CRM activation is maintained through acquired additional cooperative TF binding (Fig.2).

Beyond intra-TRM regulation of TF chromatin binding, how TRMs mutually influence their recruitment to CRMs (inter-TRM modulations) (Fig.2) will be a challenging and important question to tackle. This may require to better define the RE grammar which governs their arrangements at CRMs and how this specifies TRM recruitment and crosstalk. Going deeper into understanding CRM organization will undoubtedly benefit from a better knowledge of the actual diversity of REs involved in TF recruitment ⁵. Indeed, many TF REs are unknown and recent studies suggest that the spectrum of RE recognized by a given TF

may extend well beyond what has been defined so far ²⁵⁻²⁷. In addition, within the nucleus, CRMs co-ordinately regulating common gene(s) are found in close proximity, involving for example long-range contacts between promoters and enhancers ²⁸. In this context, how TRMs at a given CRM influence TRM binding at another functionally interconnected CRM constitutes another important question. Recent advances in genomic editing technologies including those using CRISPR/Cas9 as well as novel approaches allowing for analysis of the proteome bound to single CRM ²⁹ will undoubtedly contribute to better define how TRM recruitment is orchestrated at CRMs in the near future.

In summary, we envision TRMs as entities allowing for the integration of activities of highly interdependent and interconnected TFs, which bring specific transcriptional regulatory inputs secondarily integrated altogether to provide genes with coordinated transcriptional regulatory signals. In this context, it will be instrumental to define how the identified TF interdependent binding within and across TRM evolves in different (patho)physiological conditions. Finally, and most importantly, this knowledge will need to be leveraged to better understand how TF interdependent chromatin binding functionally translates into transcriptional regulatory outputs ^{4,18}.

Materials and Methods

Analyses were performed using CRM involved in liver-specific gene regulation which we have previously identified, i.e. CRM^G from ¹⁸.

The intra-genomic replicate (IGR) tool ^{20,21} was used to predict the impact of single nucleotide variants (SNVs) found within CRM^G as described in ¹⁸.

PERFECTOS-APE ²² version 2.0.3 was used with the HOCOMOCOv10 mononucleotide PWMs and their pre-computed thresholds ³⁰ to define how these SNVs impact on transcription factor response elements (REs) with parameters.

Fold changes provided by IGR and PERFECTOS-APE were retrieved when significant effects ($FDR < 0.05$) were found. Fold changes were set to 0 when statistical significance was not reached.

Correlation between IGR and PERFECTOS-APE data was performed using the “cor” function of the R package “stat”³¹.

Heatmaps and hierarchical clusterings were obtained using the heatmap.2 function of the R package “gplots”³² using euclidean distance and aggregation method “ward.D2”

Similar motifs were clustered based on matrix similarities using RSAT matrix-clustering with parameters : `hclust_method = average`, `metric_build_tree = NCor`, `calc = sum`, `lth w = 5`, `lth cor = 0.6`, `lth NCor = 0.4` and `quick = true`³³

Disclosure of potential conflicts of interest

The authors have no potential conflicts of interest to disclose.

Funding

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., ANR-10-LABX-46) and European Commission. B.S. is supported by the European Research Council (ERC Grant Immunobile, contract 694717).

Legend to figures

Figure 1. Definition of TF interdependent binding to CRMs through comparison of SNV-induced changes in RE affinity and TF chromatin binding

A) Summary of the main conclusions drawn from¹⁸ showing that TFs (and cofactors) bind to mouse liver CRMs as TRMs consisting of factors with preferential interdependent co-

recruitment. **B)** Schematic representation of the strategy used to compare SNVs-mediated changes in RE affinity and TF chromatin binding. Impact of SNVs on REs was defined using PERFECTOS-APE²² while their impact on TF chromatin binding was predicted through mining liver ChIP-seq data using IGR^{20,21}. Details are provided in the Materials and Methods section. **C)** Correlations between SNV-induced changes in RE affinity and TF chromatin binding were defined using cor function of R package “stat” as detailed in the Materials and Methods section. The heatmap shows how changes in affinity of individual REs correlate with that in TF chromatin binding. The main RE families from clusters 1 to 11 are indicated at the bottom (see Table S1 for the full list of REs contained in each cluster). TFs and cofactors are colored according to Fig.1A. Hierarchical clustering trees are shown on top and on the left. **D)** SNVs were assigned the best predicted impacted RE by PERFECTOS-APE, which were then used to define the fraction of REs from a given cluster which had been predicted to impact TF/cofactor chromatin binding by IGR. The heatmap shows the percent of SNVs impacting motifs comprised within clusters 1 to 11 (defined in Fig.1C) which were concomitantly predicted to impact a given TF/cofactor chromatin binding. Hierarchical clustering trees are shown on top and on the left.

Figure 2. Proposed model of combinatorial TF recruitment organization at liver CRMs.

See text for details.

References

1. Zhang H-M, Chen H, Liu W, Liu H, Gong J, Wang H, Guo A-Y. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* 2012; 40:D144–9.
2. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012; 489:91–100.

3. Santiago-Algarra D, Dao LTM, Pradel L, España A, Spicuglia S. Recent advances in high-throughput approaches to dissect enhancer function. *F1000Research* 2017; 6:939.
4. Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, Tewhey R, Isakova A, Deplancke B, Bernstein BE, et al. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci U S A* 2017; 114:E1291–300.
5. Weingarten-Gabbay S, Segal E. The grammar of transcriptional regulation. *Hum Genet* 2014; 133:701–11.
6. Weikum ER, Knuesel MT, Ortlund EA, Yamamoto KR. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nat Rev Mol Cell Biol* 2017; 18:159–74.
7. Boergesen M, Pedersen TÅ, Gross B, van Heeringen SJ, Hagenbeek D, Bindesbøll C, Caron S, Lalloyer F, Steffensen KR, Nebb HI, et al. Genome-wide profiling of liver X receptor, retinoid X receptor, and peroxisome proliferator-activated receptor α in mouse liver reveals extensive sharing of binding sites. *Mol Cell Biol* 2012; 32:852–67.
8. Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK. Effect of natural genetic variation on enhancer selection and function. *Nature* 2013; 503:487–92.
9. Pham T-H, Minderjahn J, Schmidl C, Hoffmeister H, Schmidhofer S, Chen W, Längst G, Benner C, Rehli M. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res* 2013; 41:6391–402.
10. Magnani L, Eeckhoutte J, Lupien M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet* 2011; 27:465–74.
11. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 2011; 25:2227–41.
12. Reddy TE, Gertz J, Pauli F, Kucera KS, Varley KE, Newberry KM, Marinov GK, Mortazavi A, Williams BA, Song L, et al. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res* 2012; 22:860–9.
13. Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams DJ, Talianidis I, Marioni JC, et al. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 2013; 154:530–40.
14. Roe J-S, Hwang C-I, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriack H, Young CM, Miyabayashi K, et al. Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* 2017; 170:875–888.e20.
15. Barrera LA, Vedenko A, Kurland JV, Rogers JM, Gisselbrecht SS, Rossin EJ, Woodard J, Mariani L, Kock KH, Inukai S, et al. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* 2016; 351:1450–4.

16. Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, Stamatoyannopoulos JA. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat Genet* 2015; 47:1393–401.
17. Zhou S, Treloar AE, Lupien M. Emergence of the Noncoding Cancer Genome: A Target of Genetic and Epigenetic Alterations. *Cancer Discov* 2016; 6:1215–29.
18. Dubois-Chevalier J, Dubois V, Dehondt H, Mazrooei P, Mazuy C, Sérandour AA, Gheeraert C, Guillaume P, Baugé E, Derudas B, et al. The logic of transcriptional regulator recruitment architecture at cis-regulatory modules controlling liver functions. *Genome Res* 2017; 27:985–96.
19. Guo Y, Gifford DK. Modular combinatorial binding among human trans-acting factors reveals direct and indirect factor binding. *BMC Genomics* 2017; 18:45.
20. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, Moore JH, Lupien M. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 2012; 44:1191–8.
21. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, Dowar M, Thu KL, Cescon DW, Silvester J, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nat Genet* 2016; 48:1260–6.
22. Vorontsov IE, Kulakovskiy IV, Khimulya G, Nikolaeva DD, Makeev VJ. PERFECTOS-APE - Predicting Regulatory Functional Effect of SNPs by Approximate P-value Estimation [Internet]. 2017 [cited 2017 Aug 10]. page 102–8. Available from: <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005189301020108>
23. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 2012; 151:994–1004.
24. Swinstead EE, Miranda TB, Paakinaho V, Baek S, Goldstein I, Hawkins M, Karpova TS, Ball D, Mazza D, Lavis LD, et al. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* 2016; 165:593–605.
25. Isakova A, Groux R, Imbeault M, Rainer P, Alpern D, Dainese R, Ambrosini G, Trono D, Bucher P, Deplancke B. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat Methods* 2017; 14:316–22.
26. Jolma A, Yin Y, Nitta K, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* 2015; 527:384–8.
27. Neph S, Vierstra J, Stergachis A, Reynolds A, Haugen E, Vernot B, Thurman R, John S, Sandstrom R, Johnson A, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 2012; 489:83–90.
28. de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 2013; 502:499–506.

29. Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* 2017; 170:1028–1043.e19.
30. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res* 2016; 44:D116-125.
31. R Core Team. R: A language and environment for statistical computing. R Found Stat Comput [Internet] 2015; Available from: <http://www.R-project.org>.
32. Warnes G, Bolker B, Bonebakker L, Gentleman R, Liaw W, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, et al. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. 2016; Available from: <https://CRAN.R-project.org/package=gplots>
33. Castro-Mondragon JA, Jaeger S, Thieffry D, Thomas-Chollier M, van Helden J. RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res* 2017; 45:e119.

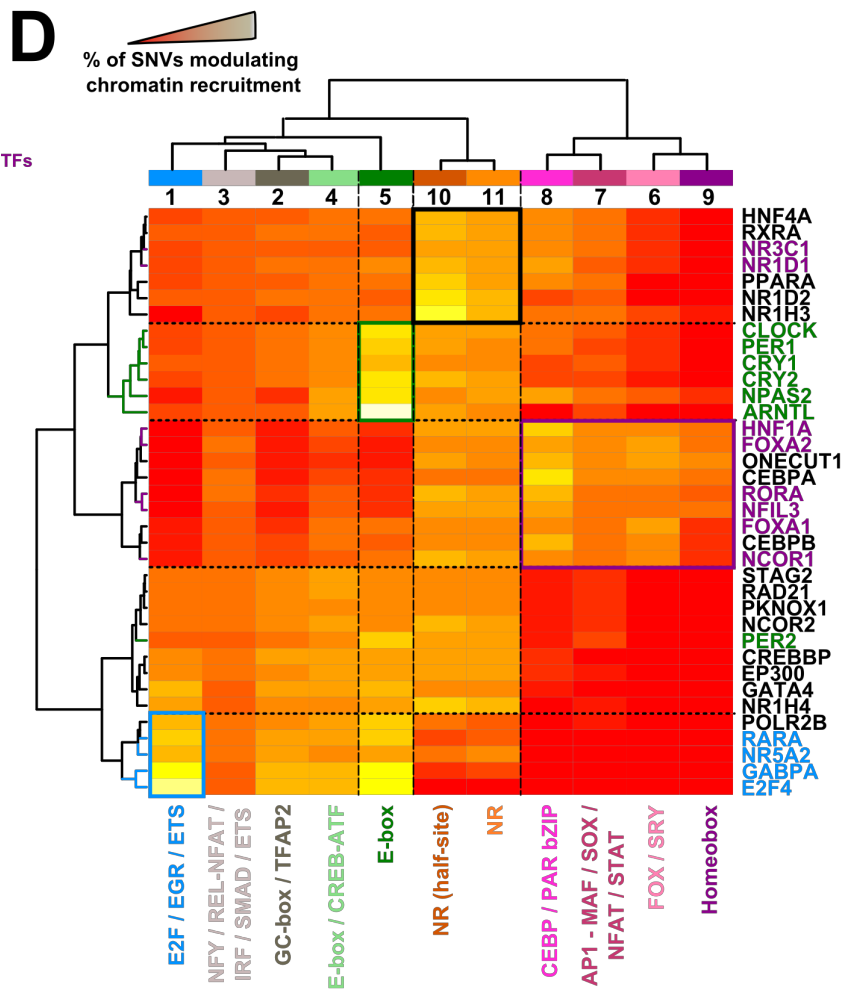
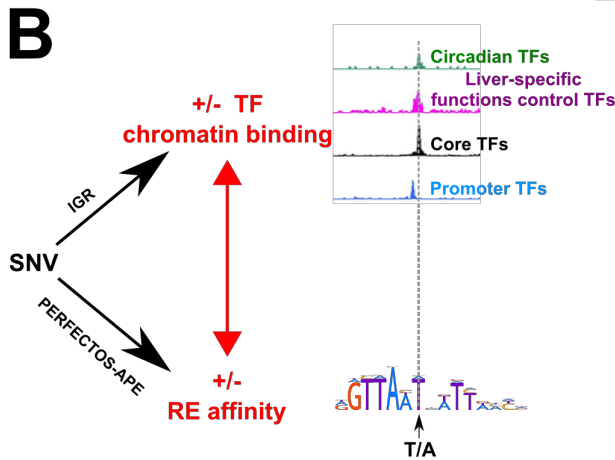
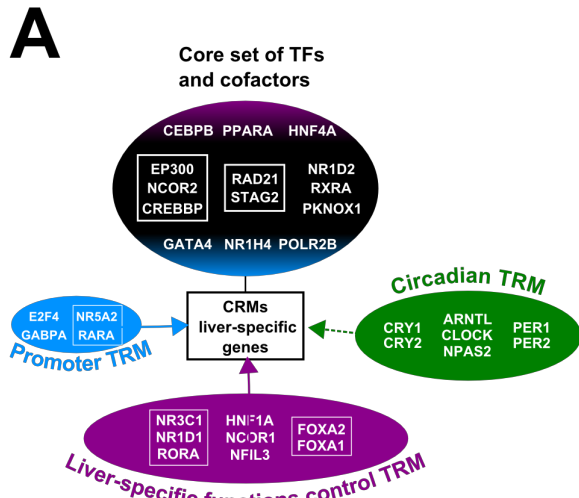
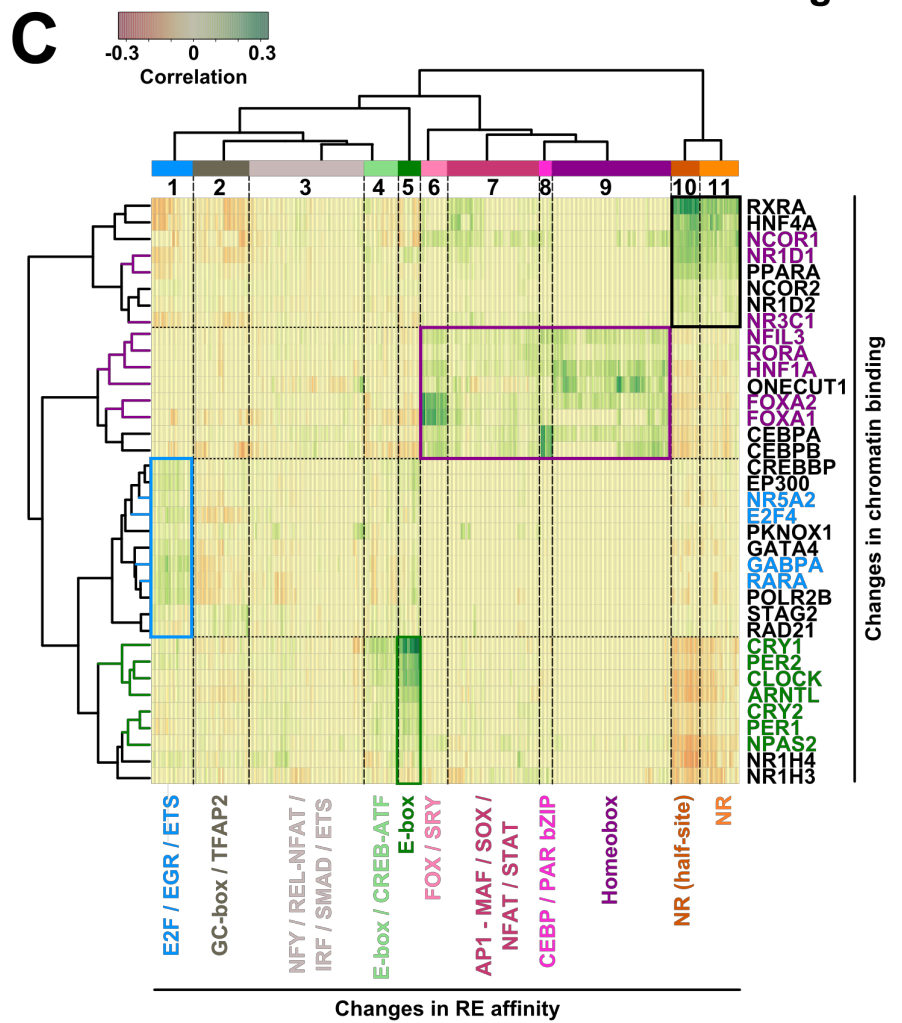
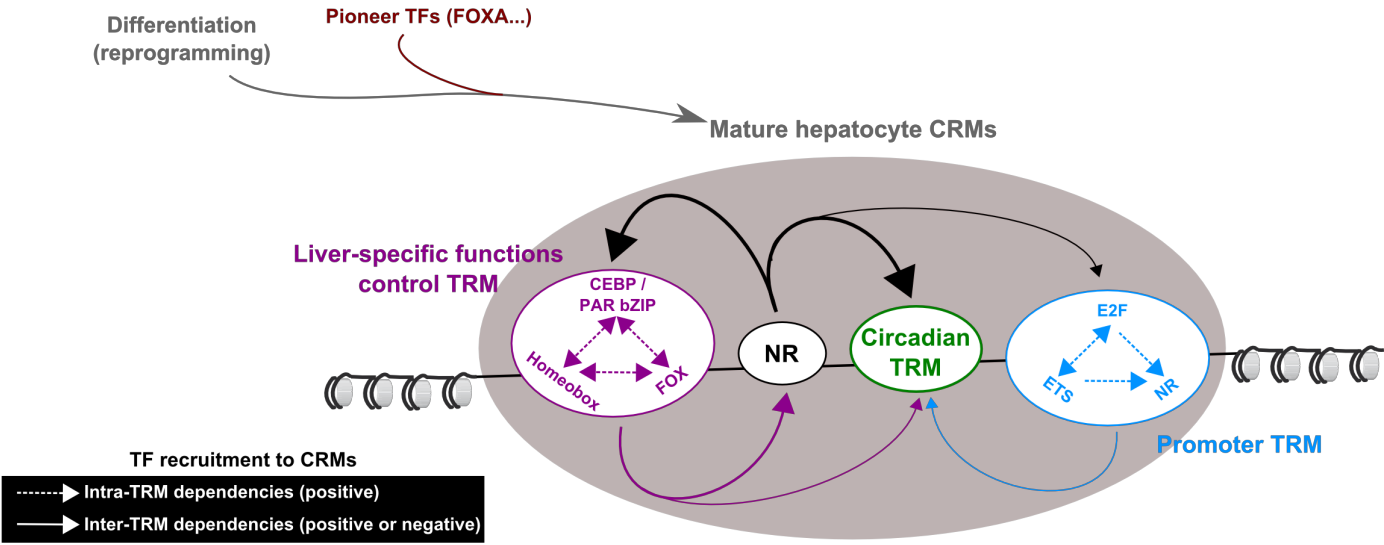


Fig.2



Heatmap cluster 1	Cluster 1	Collection_1_m27_SP2_MOUSE.H10MO.C.Collection_1_m30_WT1_MOUSE.H10MO.D.Collection_1_m10_EGR1_MOUSE.H10MO.A.Collection_1_m11_EGR2_MOUSE.H10MO.C.Collection_1_m21_PNPFP_MOUSE.H10MO.C.Collection_1_m18_FU1_MOUSE.H10MO.A.Collection_1_m20_GABPA_MOUSE.H10MO.A.Collection_1_m17_ET52_MOUSE.H10MO.C.Collection_1_m19_GABP1_MOUSE.H10MO.C.Collection_1_m29_TDP1_MOUSE.H10MO.D.Collection_1_m2_EZF1_MOUSE.H10MO.A.Collection_1_m6_EZF4_MOUSE.H10MO.C.Collie
	Cluster 2	Collection_1_m4_EPF2_MOUSE.H10MO.B.Collection_1_m5_EPF3_MOUSE.H10MO.B.Collection_1_m7_EZF5_MOUSE.H10MO.B.Collection_1_m6_EZF7_MOUSE.H10MO.D.Collection_1_m10_SLI2_MOUSE.H10MO.B.Collection_1_m11_GLI3_MOUSE.H10MO.B.Collection_1_m6_GLI1_MOUSE.H10MO.C.C
	Cluster 3	Collection_1_m1_AHR_MOUSE.H10MO.B.Collection_1_m2_AHR2_MOUSE.H10MO.D.Collection_1_m3_ZFX_MOUSE.H10MO.C
	Cluster 4	Collection_1_m1_AHR_MOUSE.H10MO.B.Collection_1_m2_AHR2_MOUSE.H10MO.D.Collection_1_m3_ZFX_MOUSE.H10MO.C
	Cluster 5	Collection_1_m1_AHR_MOUSE.H10MO.B.Collection_1_m2_AHR2_MOUSE.H10MO.D.Collection_1_m3_ZFX_MOUSE.H10MO.C
	Cluster 6	Collection_1_m28_SPA_MOUSE.H10MO.D
	Cluster 7	Collection_1_m25_PAX5_MOUSE.H10MO.S
	Cluster 8	Collection_1_m26_RFX1_MOUSE.H10MO.C
Heatmap cluster 2	Cluster 1	Collection_1_m27_PLA61_MOUSE.H10MO.S.Collection_1_m20_KLF6_MOUSE.H10MO.D.Collection_1_m31_SP1_MOUSE.H10MO.S.Collection_1_m7_EGR4_MOUSE.H10MO.D.Collection_1_m10_SLI2_MOUSE.H10MO.B.Collection_1_m11_GLI3_MOUSE.H10MO.B.Collection_1_m6_GLI1_MOUSE.H10MO.C.C
	Cluster 2	Collection_1_m16_KLF15_MOUSE.H10MO.D.Collection_1_m28_PURA_MOUSE.H10MO.D.Collection_1_m22_MAZ_MOUSE.H10MO.D.Collection_1_m39_ZN149_MOUSE.H10MO.A.Collection_1_m32_SP1_MOUSE.H10MO.A.Collection_1_m32_SP3_MOUSE.H10MO.B
	Cluster 3	Collection_1_m2_COE1_MOUSE.H10MO.C.Collection_1_m19_NF5M1_MOUSE.H10MO.C.Collection_1_m1_AP2A_MOUSE.H10MO.A.Collection_1_m17_ET52_MOUSE.H10MO.C.Collection_1_m19_GABP1_MOUSE.H10MO.C.Collection_1_m29_TDP1_MOUSE.H10MO.D.Collection_1_m2_EZF1_MOUSE.H10MO.A.Collection_1_m6_EZF4_MOUSE.H10MO.C.Collie
	Cluster 4	Collection_1_m14_RZF1_MOUSE.H10MO.C.Collection_1_m25_NF81_MOUSE.H10MO.D.Collection_1_m2_AP2B_MOUSE.H10MO.D
	Cluster 5	Collection_1_m24_NFA_MOUSE.H10MO.S.Collection_1_m34_VBX1_MOUSE.H10MO.D
	Cluster 6	Collection_1_m24_NFA_MOUSE.H10MO.S.Collection_1_m34_VBX1_MOUSE.H10MO.D
	Cluster 7	Collection_1_m13_HCT_MOUSE.H10MO.C
	Cluster 8	Collection_1_m26_PLA61_MOUSE.H10MO.D
Heatmap cluster 3	Cluster 1	Collection_1_m30_NFPA_MOUSE.H10MO.D.Collection_1_m30_PNAX_MOUSE.H10MO.D.Collection_1_m3_CEBPZ_MOUSE.H10MO.D.Collection_1_m10_P35_MOUSE.H10MO.D.Collection_1_m11_FOXH_MOUSE.H10MO.A.Collection_1_m48_PBX3_MOUSE.H10MO.A.Collection_1_m36_NFYC_MOUSE.H10M
	Cluster 2	Collection_1_m3_ZEP2_MOUSE.H10MO.D.Collection_1_m42_ZEP1_MOUSE.H10MO.D.Collection_1_m32_NF4B1_MOUSE.H10MO.B.Collection_1_m72_TF66_MOUSE.H10MO.S.Collection_1_m13_RELB_MOUSE.H10MO.C.Collection_1_m4_REL_MOUSE.H10MO.C.Collection_1_m30_NFAC1_MOUSE.H10M
	Cluster 3	Collection_1_m16_HSF1_MOUSE.H10MO.A.Collection_1_m17_HSF2_MOUSE.H10MO.A
	Cluster 4	Collection_1_m2_HSF1_MOUSE.H10MO.D.Collection_1_m2_HSF2_MOUSE.H10MO.D
	Cluster 5	Collection_1_m59_SMA1_MOUSE.H10MO.D.Collection_1_m79_USBP1_MOUSE.H10MO.C.Collection_1_m20_IRF3_MOUSE.H10MO.C.Collection_1_m23_IRF7_MOUSE.H10MO.C.Collection_1_m18_IRF1_MOUSE.H10MO.C.Collection_1_m25_IRF9_MOUSE.H10MO.C.C
	Cluster 6	Collection_1_m20_ZB118_MOUSE.H10MO.D.Collection_1_m2_ATOH1_MOUSE.H10MO.C.Collection_1_m74_TFE2_MOUSE.H10MO.S
	Cluster 7	Collection_1_m58_RJMK3_MOUSE.H10MO.B.Collection_1_m8_PEBB_MOUSE.H10MO.C.Collection_1_m57_RJMK2_MOUSE.H10MO.B
	Cluster 8	Collection_1_m4_SPA_MOUSE.H10MO.B.Collection_1_m4_EHF_MOUSE.H10MO.A.Collection_1_m6_ELF3_MOUSE.H10MO.D.Collection_1_m5_ETV4_MOUSE.H10MO.B.Collection_1_m8_STAT3_MOUSE.H10MO.B.Collection_1_m7_ELK3_MOUSE.H10MO.D.Collie
Heatmap cluster 4	Cluster 1	Collection_1_m2_P5_MOUSE.H10MO.D.Collection_1_m48_P5_MOUSE.H10MO.D
	Cluster 2	Collection_1_m2_P5_MOUSE.H10MO.D.Collection_1_m48_P5_MOUSE.H10MO.D
	Cluster 3	Collection_1_m39_NK32_MOUSE.H10MO.C.Collection_1_m37_NK321_MOUSE.H10MO.D.Collection_1_m38_NK328_MOUSE.H10MO.C
	Cluster 4	Collection_1_m46_PAX6_MOUSE.H10MO.D.Collection_1_m47_PAX8_MOUSE.H10MO.D.Collection_1_m44_PAX2_MOUSE.H10MO.D.Collection_1_m45_PAX5_MOUSE.H10MO.D
	Cluster 5	Collection_1_m71_TF2L1_MOUSE.H10MO.C
	Cluster 6	Collection_1_m71_TF2L1_MOUSE.H10MO.C
	Cluster 7	Collection_1_m76_TLX1_MOUSE.H10MO.D.Collection_1_m77_TLX1_MOUSE.H10MO.S
	Cluster 8	Collection_1_m15_HLTF_MOUSE.H10MO.D
Heatmap cluster 5	Cluster 1	Collection_1_m26_KASO_MOUSE.H10MO.B
	Cluster 2	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 3	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 4	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 5	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 6	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 7	Collection_1_m73_TCF2_MOUSE.H10MO.D
	Cluster 8	Collection_1_m73_TCF2_MOUSE.H10MO.D
Heatmap cluster 6	Cluster 1	Collection_1_m9_TBX20_MOUSE.H10MO.C
	Cluster 2	Collection_1_m75_TGIF1_MOUSE.H10MO.D
	Cluster 3	Collection_1_m72_FUBP1_MOUSE.H10MO.D
	Cluster 4	Collection_1_m14_TF2L1_MOUSE.H10MO.C
	Cluster 5	Collection_1_m14_TF2L1_MOUSE.H10MO.C
	Cluster 6	Collection_1_m14_TF2L1_MOUSE.H10MO.C
	Cluster 7	Collection_1_m14_TF2L1_MOUSE.H10MO.C
	Cluster 8	Collection_1_m14_TF2L1_MOUSE.H10MO.C
Heatmap cluster 7	Cluster 1	Collection_1_m83_SOX4_MOUSE.H10MO.C
	Cluster 2	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 3	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 4	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 5	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 6	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 7	Collection_1_m7 AP2C_MOUSE.H10MO.A
	Cluster 8	Collection_1_m7 AP2C_MOUSE.H10MO.A
Heatmap cluster 8	Cluster 1	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 2	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 3	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 4	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 5	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 6	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 7	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
	Cluster 8	Collection_1_m81_ZBTB6_MOUSE.H10MO.D
Heatmap cluster 9	Cluster 1	Collection_1_m4_MYOD1_MOUSE.H10MO.C.Collection_1_m15_MYOG_MOUSE.H10MO.C.Collection_1_m3_CREB1_MOUSE.H10MO.B.Collection_1_m1_ATF1_MOUSE.H10MO.B.Collection_1_m2_CREB1_MOUSE.H10MO.B
	Cluster 2	Collection_1_m5_E4F1_MOUSE.H10MO.D.Collection_1_m9_SREB1_MOUSE.H10MO.B
	Cluster 3	Collection_1_m16_PAX2_MOUSE.H10MO.S
	Cluster 4	Collection_1_m16_PAX2_MOUSE.H10MO.S
	Cluster 5	Collection_1_m11_MLXPL_MOUSE.H10MO.D
	Cluster 6	Collection_1_m7_HESX1_MOUSE.H10MO.D
	Cluster 7	Collection_1_m7_HESX1_MOUSE.H10MO.D
	Cluster 8	Collection_1_m20_SHA_MOUSE.H10MO.C
Heatmap cluster 10	Cluster 1	Collection_1_m4_CXXC1_MOUSE.H10MO.D
	Cluster 2	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 3	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 4	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 5	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 6	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 7	Collection_1_m12_MTF1_MOUSE.H10MO.C
	Cluster 8	Collection_1_m12_MTF1_MOUSE.H10MO.C
Heatmap cluster 11	Cluster 1	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 2	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 3	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 4	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 5	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 6	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 7	Collection_1_m28_MCR_MOUSE.H10MO.D
	Cluster 8	Collection_1_m28_MCR_MOUSE.H10MO.D

Table S1. Full list of REs from clusters 1 to 11