SUPPLEMENTAL INFORMATION

Phylogenetic modeling of enhancer shifts in African mole-rats reveals regulatory changes associated with tissue-specific traits
Parey et al. 2023

SUPPLEMENTAL DATA

Supplemental Table S1: Species and tissue samples used in this study (related to Methods, Figure 1 and Figure S1) **[Supplemental_Table_S1.xlsx file]**

Supplemental Table S2. Simulations to calibrate statistical thresholds for the branch-shift test of the phylogenetic modeling approach (related to Figure 2) **[Supplemental_Table_S2.xlsx file]**

Supplemental Table S3. Genes associated to enhancers with an activity shift in the ancestral mole-rat branch (related to Figure 2) **[Supplemental_Table_S3.xlsx file]**

Supplemental Table S4. Gene Ontology enrichments (TOP 100) for enhancers with an activity shift in the ancestral mole-rat branch (related to Figures 3 and 4) **[Supplemental_Table_S4.xlsx file]**

Supplemental Table S5. C2 pathway enrichments for enhancers with an activity shift in the ancestral mole-rat branch (related to Figures 3 and 4) **[Supplemental_Table_S5.xlsx file]**

Supplemental Table S6. Gene Ontology enrichments for promoters with an activity shift in the ancestral mole-rat branch (related to Figures 3 and 4) **[Supplemental_Table_S6.xlsx file]**

Supplemental Table S7. Gene Ontology enrichments (TOP 100) for primed enhancers with gain and losses of activity as defined by parsimony in the ancestral mole-rat branch (related to Figure 3) **[Supplemental_Table_S7.xlsx file]**

Supplemental Table S8. Gene Ontology enrichments (TOP 100) for enhancers with an activity shift in the guinea pig branch (related to Figure 3) (related to Figures 3) **[Supplemental_Table_S8.xlsx file]**

Supplemental Table S9. Transcription factor binding site enrichments for enhancers with an activity shift in the ancestral mole-rat branch (related to Figures 3) **[Supplemental_Table_S9.xlsx file]**

Supplemental Table S10. Clustering of enriched gene ontologies in ancestral and single-species mole-rat branches (related to Figure 4). **[Supplemental_Table_S10.xlsx file]**

Supplemental Table S11. Transcription factor binding sites associated with specific repetitive elements in nonalignable mole-rat enhancers (related to Figure 5) **[Supplemental_Table_S11.xlsx file]**

Table S12. Primer sequences used for 4C-seq and RT-QPCR (related to Figures S6 and S7). **[Supplemental_Table_S12.xlsx file]**

Supplemental datasets:

The following datasets have been uploaded as Supplemental Material, and deposited to Zenodo (https://doi.org/10.5281/zenodo.7442105), along with the code, inputs and environment to reproduce the presented analyses (prediction of regulatory elements from histone marks peaks, identification of orthologous regions, normalization of reads densities,

phylogenetic modeling of regulatory activity shifts and functional enrichment tests for mole-rats regulatory elements).

Dataset S1: Promoters, enhancers and primed enhancers predicted in each species before cross-mapping (.bed files).

Dataset S2: Orthologous promoters, enhancers and primed enhancers given in each species genomic coordinates system (.bed files), and normalized H3K27ac reads densities at orthologous promoters and enhancers.

Dataset S3: UP and DOWN promoters and enhancers in mole-rats (.bed files), and regions to genes association tables.

Dataset S4: Nonalignable promoters and enhancers in each species (.bed files), *de novo* repeats annotation for mole-rats (.bed and consensus .fasta files), tables with nonalignable mole-rats enhancers enriched in specific repeats.

SUPPLEMENTAL METHODS

Source and detail of tissues

All animal experiments conducted in the UK were in accordance with the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 and performed under the terms and conditions of the UK Home Office licenses P51E67724 (D.V.) and P7EBFC1B1 (E.J.S.). Damaraland mole-rat tissues were obtained from the Kuruman River Reserve, Kalahari Research Trust following local ethical approval (export permit FAUNA 0718/2/2016) and imported to the UK (DEFRA import authorization ITIMP16/0374).

At least two independent biological replicates from different animals were performed for each species and antibody. Wherever possible, tissues from young adult males were used. Tissues were prepared immediately post-mortem (typically within an hour) to maximize experimental quality. Post-mortem tissues were kept on ice until processed to minimize potential loss of protein-DNA interactions during post-mortem time. Sample allocations to experimental batches were randomised to ensure unbiased distributions of species, tissue, individual and sex, using the R/Bioconductor package OSAT (Yan et al. 2012).

Chromatin immunoprecipitation and high-throughput sequencing

Tissues were prepared by direct perfusion of the liver with PBS, followed by dicing the whole organs (liver and heart) in small pieces around $1cm^3$. Blood clots within the heart ventricles were removed. Cross-linking of the diced tissue was performed in 1% formaldehyde solution for 20 min, addition of 250 mM glycine and incubation for a further 10 min to neutralize the formaldehyde. After homogenization of cross-linked tissues in a dounce tissue grinder, samples were washed twice with PBS and lysed according to published protocols (Schmidt et al. 2009) to solubilize DNA-protein complexes. Chromatin was fragmented to 300 bp average size by sonication on a Misonix sonicator 3000 with a 418 tip (1/16 inch diameter). Chromatin from 50-200 mg of dounced tissue was used for each ChIP experiment using antibodies against H3K4me3 (millipore 05-1339), H3K27ac (abcam ab4729), H3K4me1 (abcam ab8895), TBX5 (Insight Biotechnology sc-515536), RXRA (sc-553), FOXA1 (ab70382), HNF4A (ARP31946), CEBPA (sc-9314) and ONECUT1 (sc-13050). Illumina sequencing libraries were

prepared from ChIP-enriched DNA using ThruPLEX DNA-seq library preparation kit (Takara Bio, for histone modifications) or NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, for transcription factors) with up to 10ng of input DNA and 8-15 PCR cycles. After PCR, libraries were pooled in equimolar concentrations and sequenced on Illumina HiSeq 4000 or NovaSeq instruments.

Computational analysis of ChIP-seq data and definition of regulatory regions

Basic alignment and peak calling: Aligned BAM files were obtained with BWA 0.7.17 and the Ensembl v99 assemblies HetGla_1.0 (Naked mole-rat), DMR_v1.0 (Damaraland mole-rat), Cavpor3.0 (guinea pig) and GRCm38 (mouse). We used MACS2 (Zhang et al. 2008) to call peaks for each ChiP-seq replicate, using default parameters and "--keep-dup all" to retain duplicate reads. Before peak-calling, multi-mapping reads were removed and read-depth adjusted to 20 million uniquely mapped reads (or all available reads for low-depth libraries).

Definition of regulatory regions from ChIP-seq peaks: We first constructed sets of reproducible peaks for each combination of histone mark, tissue and species by merging peaks identified across a minimum of two biological replicates (with minimum 50% length overlap). In a second step, we used these sets of reproducible peaks to identify promoters, enhancers and primed enhancers independently for each species and tissue, according to the following criteria: all H3K4me3 peaks overlapping an H3K27ac peak (minimum 50% bases overlap) were predicted to be promoters, H3K27ac peaks not overlapping promoters were predicted as enhancers, and H3K4me1 peaks not overlapping any H3K4me3 or H3K27ac peaks were as predicted primed enhancers.

Definition of orthologous regulatory regions: we used the liftOver local software (Kuhn et al. 2013) to map genomic coordinates of the regulatory elements identified in naked mole-rat, Damaraland mole-rat and guinea pig to the mouse genome. LastZ pairwise whole-genome alignments with mouse were downloaded from Ensembl Compara v99 and converted from Ensembl maf format to the UCSC chain format using UCSC tools (including mafToPsl, pslToChain and chainSwap, v357 for all software). We validated the correct implementation

of the coordinates conversion step using liftOver with the generated chain alignment files by comparing the resulting coordinates with those obtained from queries directly performed through the Ensembl API(Cunningham et al. 2022).

We defined a set of “high-confidence orthologous regions” as 4-way orthologous regions for which we required robust liftOver mapping of regulatory elements across the four genomes. This was achieved in two steps, involving (i) the definition of the set of regulatory regions, expressed in mouse genome coordinates, that can be aligned from each of the other species to mouse; and (ii) a filtering step to retain only (sub)regions with a strict reciprocal liftOver mapping in each genome (e.g. a naked mole-rat region mapping to mouse, and the mouse coordinates of this region mapping back to each of the other three genomes). In this filtering step, the final criteria for regions to be defined as high-confidence 4-way orthologs were: a reciprocal liftOver mapping from mouse to the three other genomes, with a maximum of 70% gaps allowed (-minMatch 0.30) and similar lengths of the resulting regions across the four genomes (the difference in length between the mouse region and regions in any other species must be less than 15% of the largest region).

Homogenization of regulatory region type: for cases where orthologous regions were defined as different types of regulatory elements across species, we used a majority rule system to homogenize regulatory element types. For instance, a region defined as a promoter in mouse, guinea pig and naked mole-rat but as an enhancer in Damaraland mole-rat was re-defined as a promoter across all species. In case of ties, regions are arbitrarily assigned to the “highest-level” regulatory type (i.e. promoters have priority over enhancers and primed enhancers, and enhancers have priority over primed enhancers).

Exclusion of greylisted regions: H3K27ac signal read density normalization with input ChIP is critical for quantitative analyses of regulatory activity. We therefore established greylists of regions with an unusually elevated signal in input ChIP experiments to tag them for exclusion. We took advantage of the approach implemented in the GreyListChIP R package (Team 2021), which flags regions with elevated signal in the input. In practice, for all pairs of input –

H3K27ac sample ChIP experiments, we ran chipseq-greylist v1.0.2 with 100 bootstrap (--bootstraps 100), a simpler python implementation of GreyListChIP (available from https://github.com/roryk/chipseq-greylist). Two greylists were computed, one for each tissue, and including all regions flagged by chipseq-greylist in at least one input ChIP in any species. Finally, any region from the set of "high-confidence orthologous elements" overlapping with a greylisted region was removed before phylogenetic modeling. Across the different sets, the fraction of alignable elements that we exclude as greylisted ranged between 0.5% and 2%.

4C-seq of chromatin 3D contacts at candidate loci

Crosslinked tissues were homogenised with a douncer and 100mg of homogenised tissue was resuspended as single cells in 5 ml isolation buffer (10% FBS in PBS). Cells were pelleted and lysed in 1 ml lysis buffer (50mM Tris-HCl pH 7.5, 0.5% NP-40, 1% TX-100, 150mM NaCl, 5mM EDTA and protease inhibitors) for 20 minutes on ice. Pelleted nuclei were washed and resuspended in 0.5 ml of 1.2X first restriction enzyme buffer, then samples were incubated with 0.3% SDS for 1 hour at 37 °C with shaking and later incubated with 2.5% Triton X-100 for 1 hour. 100U of appropriate first restriction enzyme were added and incubated 3h at 37 °C, then additional 100U were added and incubated overnight at 37 °C. Digestion was stopped by incubating for 20 min at 65 °C. Samples were then diluted in a final volume of 6 ml with ligation buffer (10X ligation buffer: 660mM Tris-HCl, pH 7.5, 50mM MgCl2, 50mM DTT, and 10mM ATP) and 50 Weiss unit of T4 DNA ligase were added, following overnight incubation at 16ºC. Reversal of crosslinks was performed by overnight incubation with 10 mg/ml proteinase K at 65ºC and templates were purified with AppMag PCR Clean Up Beads. Next, 50U of appropriate second restriction enzyme were added and incubated overnight at 37ºC with shaking. Second ligation was performed with 50 Weiss units T4 DNA ligase and overnight incubation at 16ºC. Then, 4C template was purified with AppMag PCR Clean Up Beads and four parallel PCR reactions per experiment were performed with 200ng of template. PCR products were pooled and cleaned-up with AMPure XP beads. Sequencing libraries were generated by using NEBNext Ultra II DNA Library Prep Kit for Illumina and sequenced in a

NextSeq 2000 instrument. Sequencing data was processed with the pipe4C pipeline (Krijger et al. 2020) to generate normalized 4C coverage tracks. Coverage was normalized to sum up to 10,000 reads over the captured loci scaffold and plotted as a running average across 21 fragment ends, clipped at the bait to maximize the range of non-bait signal. We called peaks using peakC (Geeven et al. 2018) with default parameters, to identify regions in significant contact with the promoter.

RNA isolation and RT-QPCR

Total RNA was extracted with RNA Universal Mini Kit (Qiagen) and treated with DNase (Ambion, AM1907) to remove contaminating genomic DNA. RNA (1 ug) was retrotranscribed using High-capacity cDNA synthesis kit (Thermo Fisher Scientific), and cDNA was diluted 1/10 for qPCRs using KAPA SYBR FAST (Sigma-Aldrich, KK4610). RT–qPCR was carried out on a Roche LC480 for 40 cycles, using relative quantification relative to *Tbp* (internal control). Primer sequences in Supplemental Table S12 were designed against canonical transcripts in each species to target different exons, and forward or reverse primer sequences targeting multiple species were selected from a multiple alignment obtained with ClustalW (Thompson et al. 1994).

Detection of regulatory activity shifts using parsimony

For comparison with phylogenetic modeling, we identified up and down promoters and enhancers in mole-rats from the same set of high-confidence orthologous elements with a complementary parsimony-based approach. To achieve this, we applied the dollo parsimony criterion, which states that a trait – here the regulatory activity of a genomic region - can be gained only once but can be lost multiple times. Under this criterion, and using the ancestral mole-rat branch as an example, Up enhancers are the enhancers detected as active in both mole-rats (i.e. overlapping a MACS2 H3K27ac peak corresponding to an enhancer in both species) and inactive in both guinea pig and mouse. Respectively, Down elements are detected as inactive in both mole-rats and as active in both guinea pig and mouse. Therefore,

and in contrast to phylogenetic modeling, parsimony-based regulatory shifts are based on binary (presence or absence of peaks) rather than continuous data (normalised reads).

Detection of regulatory activity shifts using differential binding analysis

For comparison with phylogenetic modeling, we implemented the procedure described in the DiffBind R package (Ross-Innes et al. 2012) to identify regulatory regions differentially active in mole-rats compared to guinea pig and mouse. For each regulatory region, we measured regulatory activity as the raw number of H3K27ac ChIP-seq reads minus the raw number of input ChIP-seq reads. We then normalized these counts across samples using the median of ratio method from the DESeq2 package (Love et al. 2014). Finally, we ran the DESeq2 differential analysis procedure, contrasting a first group containing all samples from both mole-rats against a second group with samples from guinea pig and mouse. We filtered DESeq2 results to retain differentially active elements with FDR < 0.01 and abs(log_2 foldchange) > 2.

Gene ontologies, pathways and genes enrichment tests

We downloaded the mouse gene ontologies ("GO") data from the MGI database on the 20th of April 2022, as well as C2 pathway annotations from MgDB on https://bioinf.wehi.edu.au/software/MSigDB/. We filtered GO data to retain only GO of the Biological Process ("BP") domain, and C2 pathways to retain only 1,397 pathways (mostly retaining REACTOME, KEGG and BIOCARTA pathways). We transferred mouse Gene Ontology and pathway annotations to the naked mole-rat and Damaraland mole-rat using gene orthologies from Ensembl compara version 102, extracted with Ensembl BioMart (Cunningham et al. 2022). We re-implemented the region-based gene ontologies tests implemented in GREAT (McLean et al. 2010), allowing us to use custom genomes (the naked-mole rat and Damaraland mole-rat genomes) and the more recent Gene Ontology annotations downloaded from MGI.

We used GREAT's default gene to regions association rule, establishing gene basal regulatory domains (5kb upstream, 1kp downstream) with distal extension up to 1,000kb (or to the nearest next basal domain if it is closer). We validated these gene to regions association using

published mouse HiC data (Chapski et al. 2018), confirming that a majority of predicted associations occur within the same TAD (85% in liver, 80% in heart). GREAT can perform three GO enrichment tests: a binomial test over regions against the genome as background (a), a hypergeometric test over genes against the genome as background (b) and a hypergeometric test over regions using a carefully selected set of genomic regions as background (c). We implemented all three tests, using similar rules for GO propagation and filters as in GREAT. We propagate GO annotations by associating genes annotated to a specific GO to all of this GO term's parents. To increase statistical power and alleviate redundancy, we filter GO to only test for the most specific GO terms amongst all GO associated to the exact same foreground genes. We also validated the recovered GO enrichments are robust to the choice of alpha threshold. For Gene Ontology and pathways enrichment tests, we used GREAT recommended tests (a) and (b) (i.e. tests against the whole genome as background). We retained only terms found enriched with both tests, ranked them according to BH adjusted p-value < 0.05 of test (a), and selected the top 100 GO terms and top 20 C2 pathways. To identify single instances of significantly-enriched genes we used test (c), using the complete regulatory regions set of a given category as background (for instance all orthologous heart enhancers as background for nmrdmr up heart enhancers). Lastly, we validated our implementation by verifying that highly similar terms were found when using the GREAT web-server with the mouse genome and older Gene Ontology data from MGI. For regulatory elements with an activity shift in the ancestral mole-rat branch, we used the Damaraland mole-rat genome, which has a higher contiguity, to perform functional enrichment tests. For single-species shifts and elements enriched in repeats, we used this species genome for the tests.

De novo repeat annotation in mole-rat species

We constructed *de novo* repeat libraries for naked mole-rat and Damaraland mole-rat using RepeatModeler version 2.0.2a (Flynn et al. 2020) with default parameters. For guinea pig and mouse, we downloaded pre-computed *de novo* repeat libraries from the Dfam database

(Storer et al. 2021). We next ran RepeatMasker version 4.1.2-p1 to annotate the location of repeats in mole-rat, guinea pig and mouse genomes, in sensitive mode (-s) and skipping bacterial insertion check (-no_is), with rmblastn version 2.10.0+. The average 2-parameters Kimura distance of TE sequences to their consensus were computed using the calcDivergenceFromAlign.pl script from RepeatMasker, which corrects for elevated mutation rates in CpG loci. Finally, landscape plots were drawn using the createRepeatLandscape.pl script from RepeatMasker.

Overlaps between regulatory elements and repeats

We used BEDTools version v2.29.2 (Quinlan and Hall 2010), across “orthologous”, “nonalignable” and “genome” elements sets. Specifically, we used BEDTools intersect to compute overlaps between elements and repeats; and BEDTools jaccard to obtain the corresponding jaccard index (ratio of the intersection to the union of the datasets, in number of bases). “Genome” sets were constructed from n=10,000 random permutations of the corresponding “orthologous” or “nonalignable” set, on the whole genome. Following the approach from the jaccard test of the GenomTriCorr R package (Favorov et al. 2012), we tested for significant differences in overlaps with repeats across sets. Here, for comparisons with “genome” elements, the n=10,000 permutation-based Jaccard indexes form the null distribution to which we compare the Jaccard index obtained on the corresponding “orthologous” (or “nonalignable”) set. For comparisons between “orthologous” and “nonalignable” sets, we compare the Jaccard index obtained on a “nonalignable” set with n=m elements to the null distribution obtained from 100 random samplings of n=m elements drawn from the “orthologous” and “nonalignable” sets combined. P-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) procedure.

We again used the Jaccard test to define significantly enriched repeat families. To do this, we first selected repeat families with a minimum of 100 instances overlapping an element in a “nonalignable” set. Second, for each repeat family, we computed the Jaccard index of the intersection between regulatory elements of a “nonalignable” set and the repeats in a family.

Third, we constructed a null distribution of n=100 Jaccard indexes computed from the random permutations of “nonalignable” elements over the genome. Finally, we retained repeat families as significantly enriched in a “nonalignable” set when BH adjusted p-values were < 0.05.

SUPPLEMENTAL REFERENCES

Chapski DJ, Rosa-Garrido M, Hua N, Alber F, Vondriska TM. 2018. Spatial Principles of Chromatin Architecture Associated With Organ-Specific Gene Regulation. *Front Cardiovasc Med* **5**: 186.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R et al. 2022. Ensembl 2022. *Nucleic Acids Res* **50**: D988-D995.

Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput Biol* **8**: e1002529.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* **117**: 9451-9457.

Geeven G, Teunissen H, de Laat W, de Wit E. 2018. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res* **46**: e91.

Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 2020. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**: 17-32.

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Brief Bioinform* **14**: 144-161.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495-501.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**: 389-393.

Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**: 240-248.

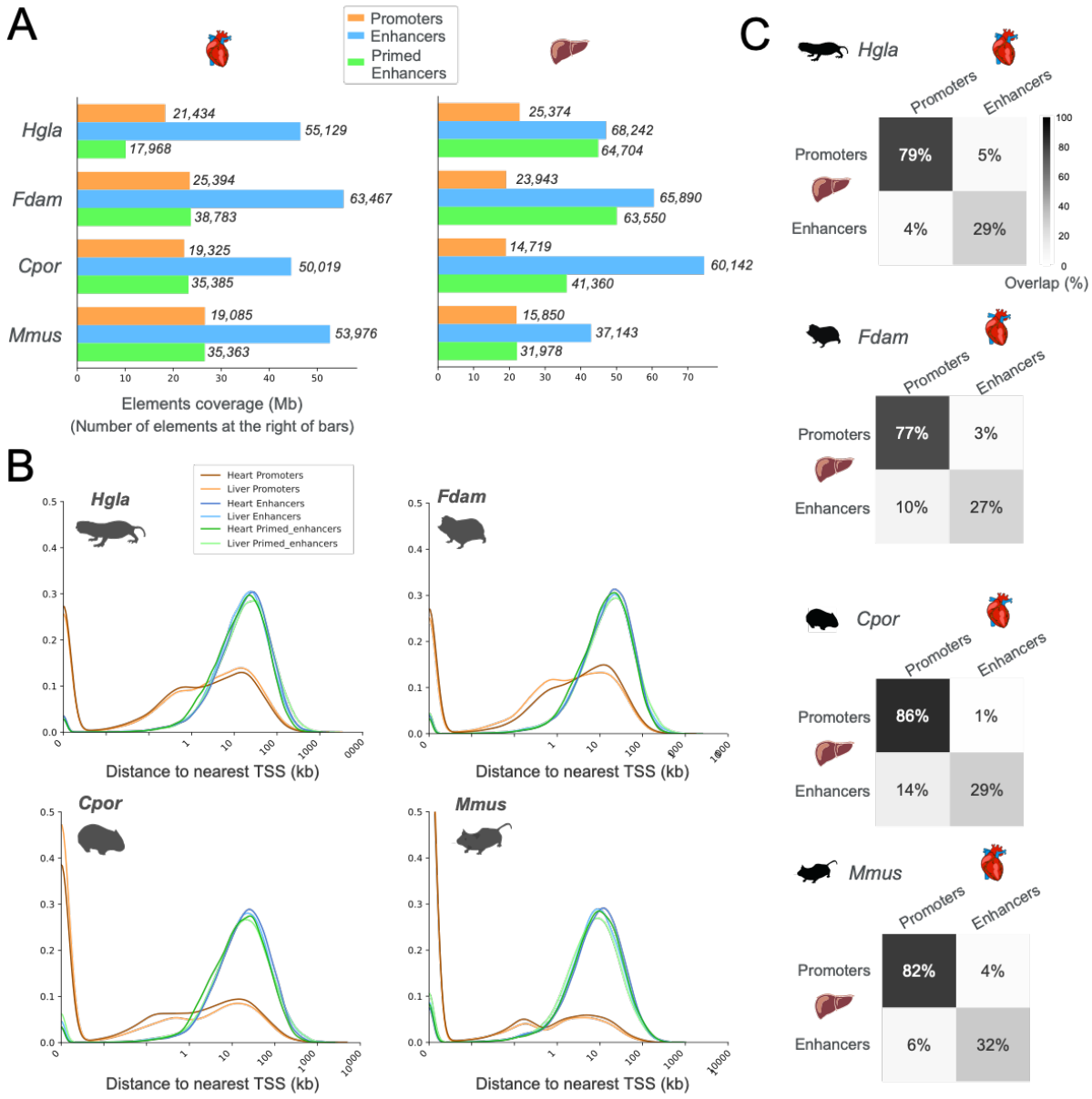
Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA* **12**: 2.

Team RC. 2021. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.

Yan L, Ma C, Wang D, Hu Q, Qin M, Conroy JM, Sucheston LE, Ambrosone CB, Johnson CS, Wang J et al. 2012. OSAT: a tool for sample-to-batch allocations in genomics experiments. *BMC Genomics* **13**: 689.

SUPPLEMENTAL FIGURES

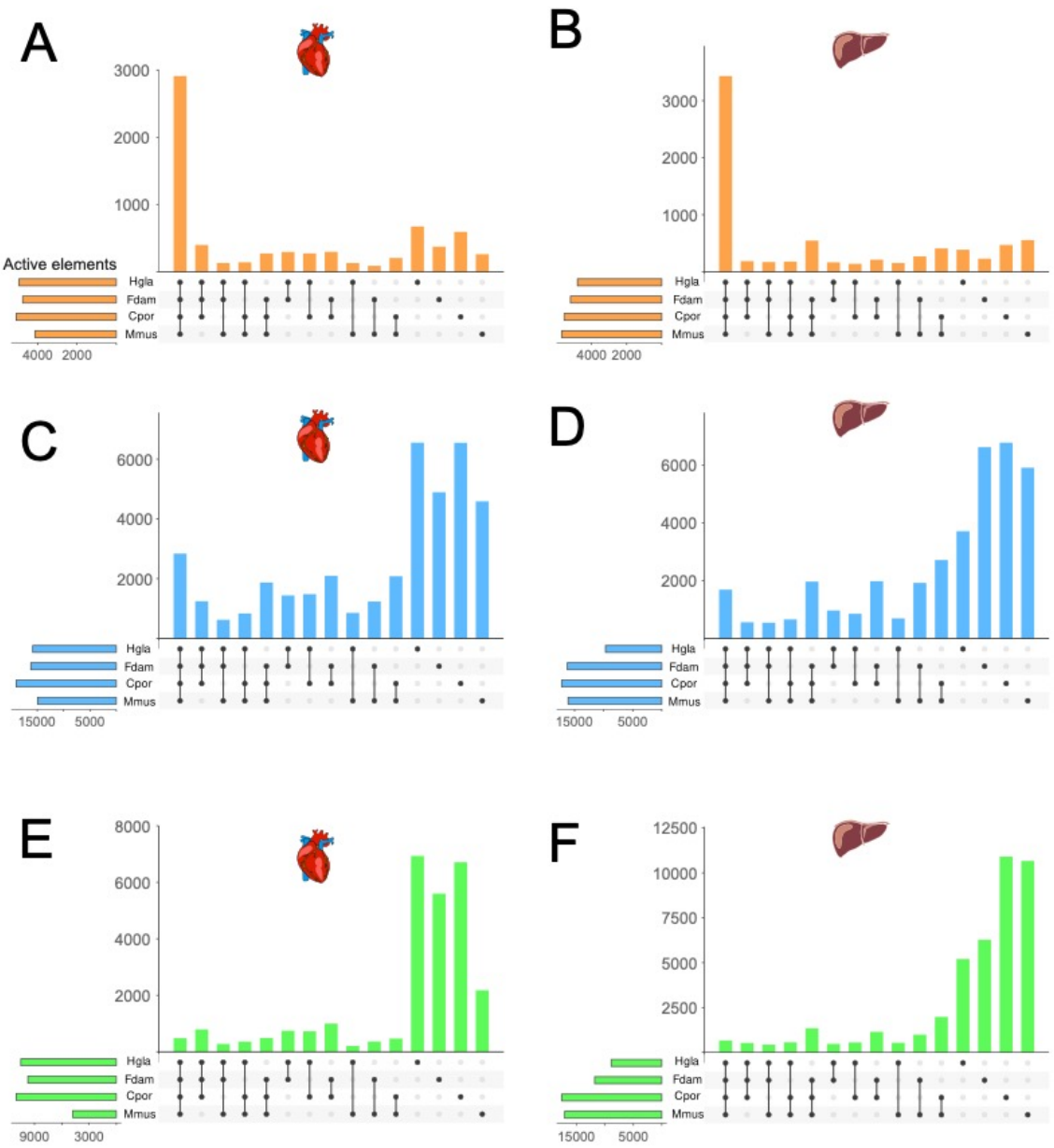


Supplemental Figure S1. Basic properties of promoters, enhancers and primed enhancers identified in heart and liver tissues across two mole-rats and two outgroup rodents (related to Figure 1)

A. Number of promoters, enhancers and primed enhancers identified in each species. Bars correspond to total genomic coverage of each set of elements. Number of elements are indicated at the right of bars.

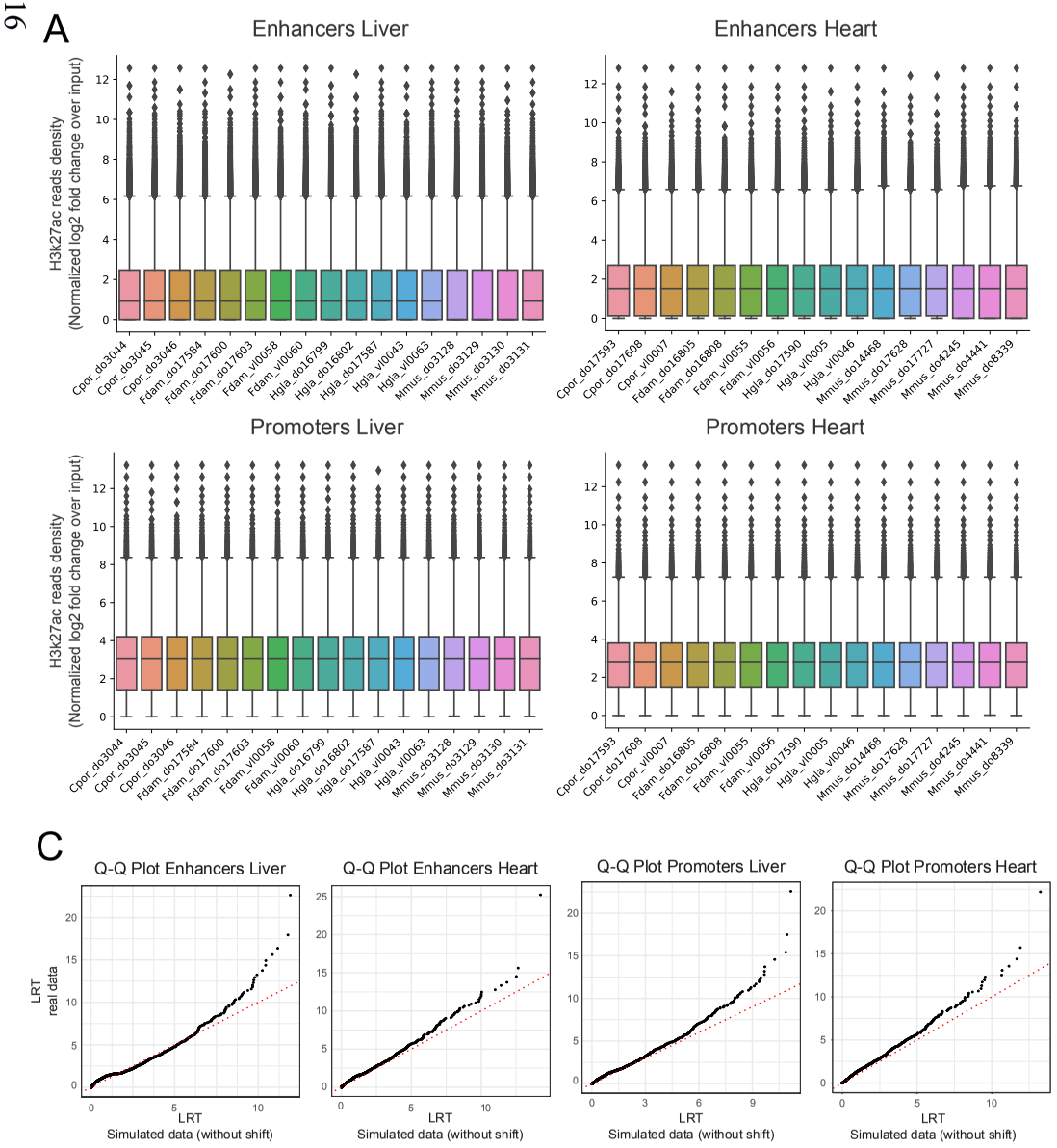
B. Distribution of distances to the nearest transcription start site (TSS) for promoters, enhancers and primed enhancers across the two tissues and four species. As expected, identified promoters are typically close to annotated genes, while enhancers are distal regulatory elements.

C. Overlap between promoters and enhancers across liver and heart tissues, in each of the four species.



Supplemental Figure S2. Cross-mapping of promoters, enhancers and primed enhancers across the four study species (related to Figure 1)

A-F. Activity of orthologous promoters, enhancers and primed enhancers across species. Upset plots indicate the number of orthologous promoters (orange), enhancers (blue) and primed enhancers (green) active across several species or species-specific. In agreement with previous observations, promoter epigenomic activity is highly conserved across species while enhancers and primed enhancers are mostly species-specific.

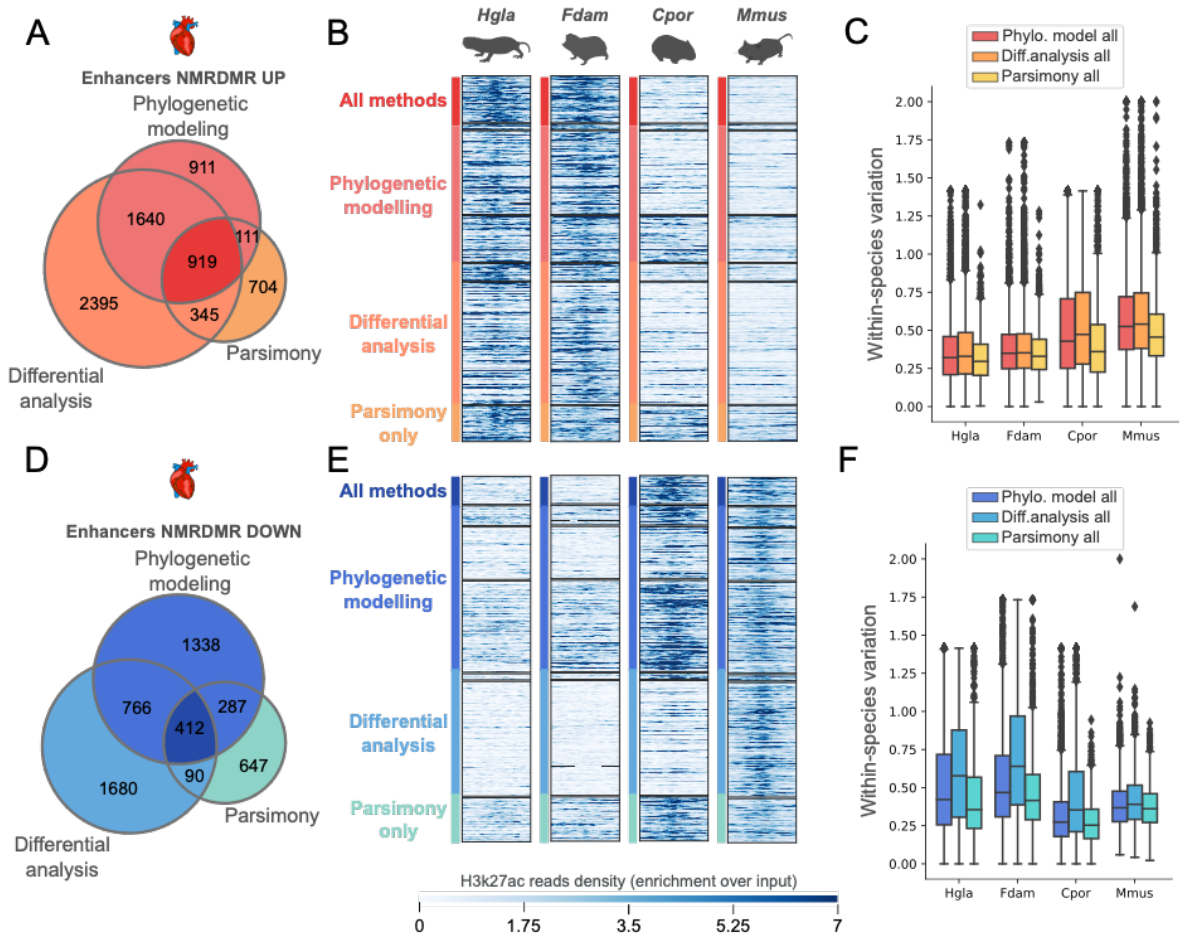

A
Enhancers Liver
Enhancers Heart
Promoters Liver
Promoters Heart
H3k27ac reads density
(Normalized log2 fold change over input)
Cpor_do3044
Cpor_do3045
Cpor_do3046
Fdam_do17584
Fdam_do17600
Fdam_do17603
Fdam_vl0058
Fdam_vl0060
Hgla_do16799
Hgla_do16802
Hgla_do17587
Hgla_vl0043
Hgla_vl0063
Mmus_do3128
Mmus_do3129
Mmus_do3130
Mmus_do3131
Cpor_do17593
Cpor_do17608
Cpor_vl0007
Fdam_do16805
Fdam_do16808
Fdam_vl0055
Fdam_vl0056
Hgla_do17590
Hgla_vl0005
Hgla_vl0046
Mmus_do14468
Mmus_do17628
Mmus_do17727
Mmus_do4245
Mmus_do4441
Mmus_do8339
B
Enhancers Liver H3K27ac
Enhancers Heart H3K27ac
Promoters Liver H3K27ac
Promoters Heart H3K27ac
Promoters Heart H3K4me3
Promoters Heart H3K4me3
Height
Bootstrap, 1000 replicates
hclust (*, "average")
C
Q-Q Plot Enhancers Liver
Q-Q Plot Enhancers Heart
Q-Q Plot Promoters Liver
Q-Q Plot Promoters Heart
LRT
real data
Simulated data (without shift)

Supplemental Figure S3. Data normalisation and parameter estimation for phylogenetic modeling (related to Figure 2)

A. Distribution of normalized H3K27 reads densities at orthologous enhancers and promoters. Reads densities were expressed as $\log_2$ fold-change of signal over input ChIP and normalized with quantile normalization.

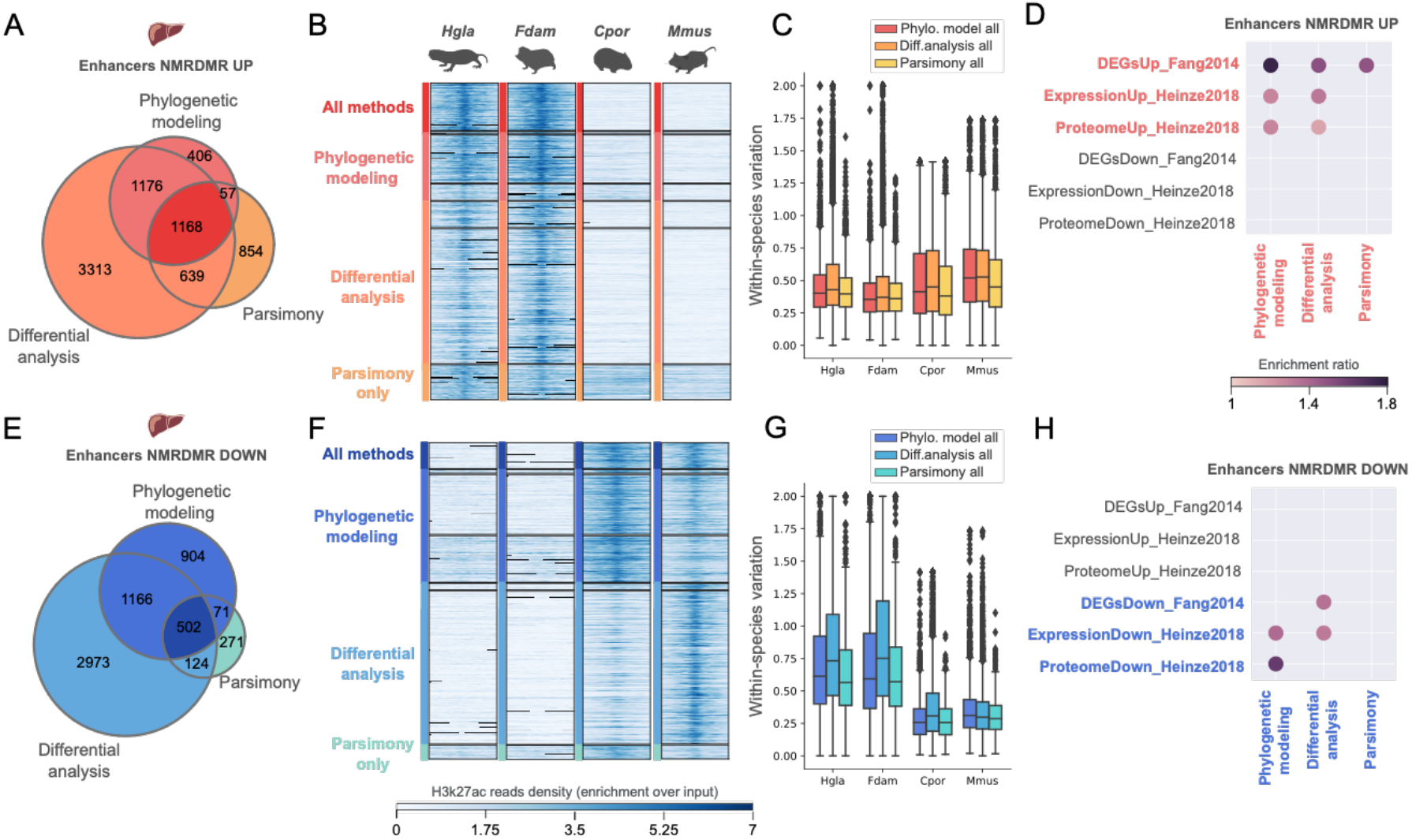
B. Hierarchical clustering of normalized reads densities at orthologous enhancers and promoters. Hierarchical clustering was performed using the euclidean distance and average linkage, with bootstrap support computed from resampling using the shipunov R package. Clustering with normalized H3K27ac read densities recapitulates the species phylogeny with high support. Conversely, H3K4me3 normalized read densities lack phylogenetic signal and are not suitable for phylogenetic modeling with EVE.

C. Q-Q plots comparing distributions of Likelihood Ratio Tests (LRT) statistics from the EVE branch-shift test. LRT were computed from data simulated under the null (no shift of regulatory activity, x-axis) and observed data (y-axis). LRTs from observed and simulated data follow the same distribution for low LRT values with a departure from the null for observed data at high LRT values, showing that simulated LRTs can be used to compute empirical p-values.



Supplemental Figure S4. Comparison of phylogenetic modeling results in heart with regulatory shifts inferred from parsimony and differential binding analysis (related to Figure 2)

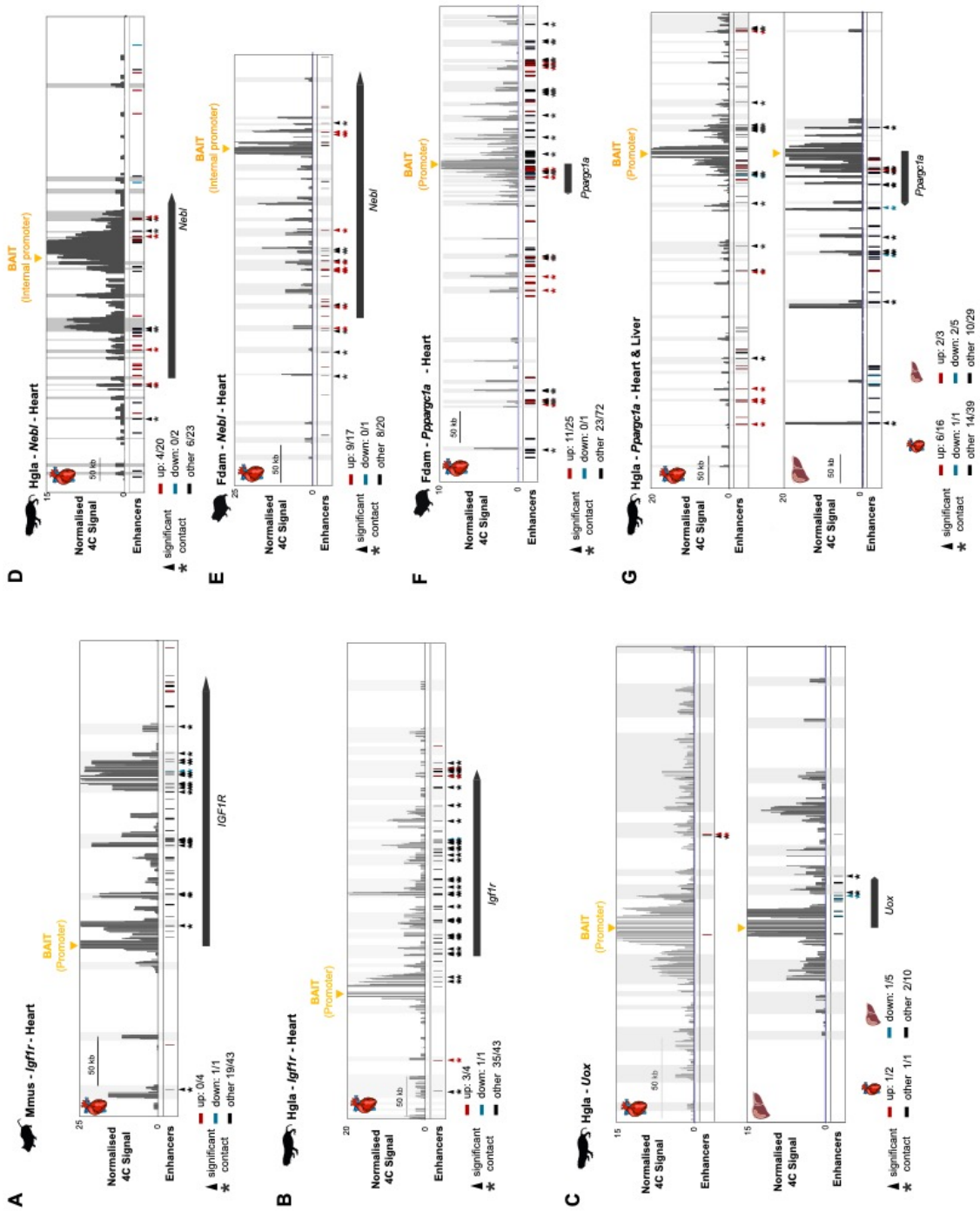
A-C. Comparison of heart enhancers identified by phylogenetic modeling, differential analysis and parsimony as up-regulated in both mole-rats (Up enhancers, ancestral branch) **A.** Venn diagram showing the overlap between elements identified by each method. **B.** Read density heatmaps for elements identified by the different approaches; black lines indicate overlapping subsets across approaches (as denoted in A.). H3K27ac read density is presented as fold enrichment over input, averaged across biological replicates. Colors on y-axis correspond to category combinations highlighted on the Venn diagram in A. **C.** Within-species variation (coefficient of variation) in read density for enhancers identified with the different methods. Up elements identified with differential analysis show systematically higher variation, suggesting they could contain a higher fraction of false positives. **D-F.** Comparison of heart enhancers identified by phylogenetic modeling, differential analysis and parsimony, as down-regulated in both mole-rats (Down enhancers, ancestral branch). Representation as in A-C.


A
Enhancers NMRDMR UP
Phylogenetic modeling
406
1176
57
1168
854
3313
639
Parsimony
Differential analysis
B
Hgla
Fdam
Cpor
Mmus
All methods
Phylogenetic modeling
Differential analysis
Parsimony only
C
Phylo. model all
Diff.analysis all
Parsimony all
Within-species variation
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Hgla
Fdam
Cpor
Mmus
D
Enhancers NMRDMR UP
DEGsUp_Fang2014
ExpressionUp_Heinze2018
ProteomeUp_Heinze2018
DEGsDown_Fang2014
ExpressionDown_Heinze2018
ProteomeDown_Heinze2018
Phylogenetic modeling
Differential analysis
Parsimony
Enrichment ratio
1
1.4
1.8
E
Enhancers NMRDMR DOWN
Phylogenetic modeling
904
1166
71
502
271
2973
124
Parsimony
Differential analysis
F
All methods
Phylogenetic modeling
Differential analysis
Parsimony only
H3k27ac reads density (enrichment over input)
0
1.75
3.5
5.25
7
G
Phylo. model all
Diff.analysis all
Parsimony all
Within-species variation
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Hgla
Fdam
Cpor
Mmus
H
Enhancers NMRDMR DOWN
DEGsUp_Fang2014
ExpressionUp_Heinze2018
ProteomeUp_Heinze2018
DEGsDown_Fang2014
ExpressionDown_Heinze2018
ProteomeDown_Heinze2018
Phylogenetic modeling
Differential analysis
Parsimony

Supplemental Figure S5. Comparison of phylogenetic modeling results in liver with regulatory shifts inferred from parsimony and differential binding analyses (related to Figure 2)

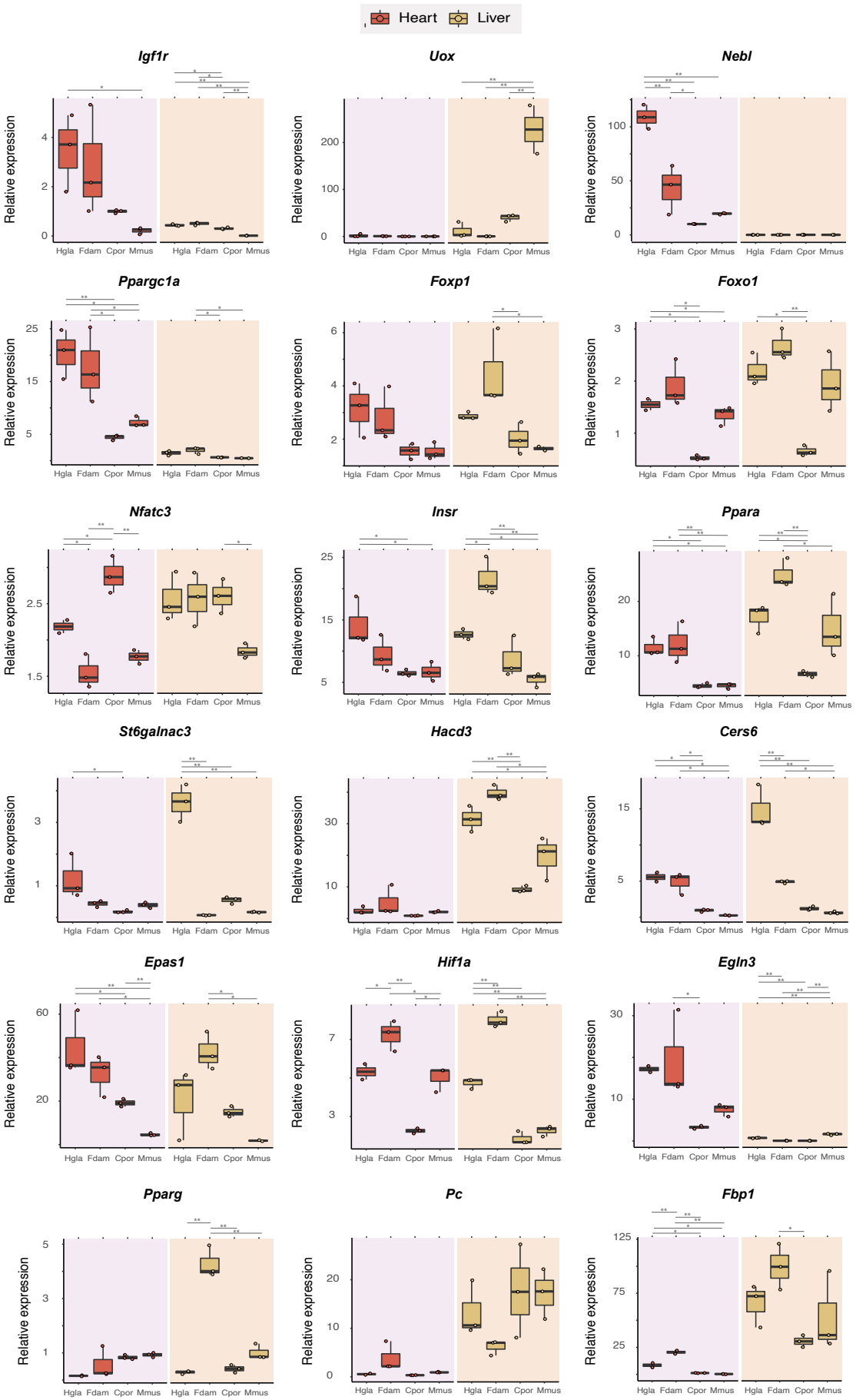
A-D. Comparison of liver enhancers identified by phylogenetic modeling, differential analysis and parsimony as up-regulated in both mole-rats (Up enhancers, ancestral branch). **A-C** as in Fig. S4. **D.** Association of identified Up elements with previously reported differentially-expressed genes (DEGs) in mole rats (see GREAT enrichment tests, Methods). Phylogenetic modeling recovers elements significantly associated with DEGs, and with higher enrichment ratios than other methods.

E-H. Comparison of liver enhancers identified by phylogenetic modeling, differential analysis and parsimony, as down-regulated in both mole-rats (Down enhancers, ancestral branch). Representation as in A-D.



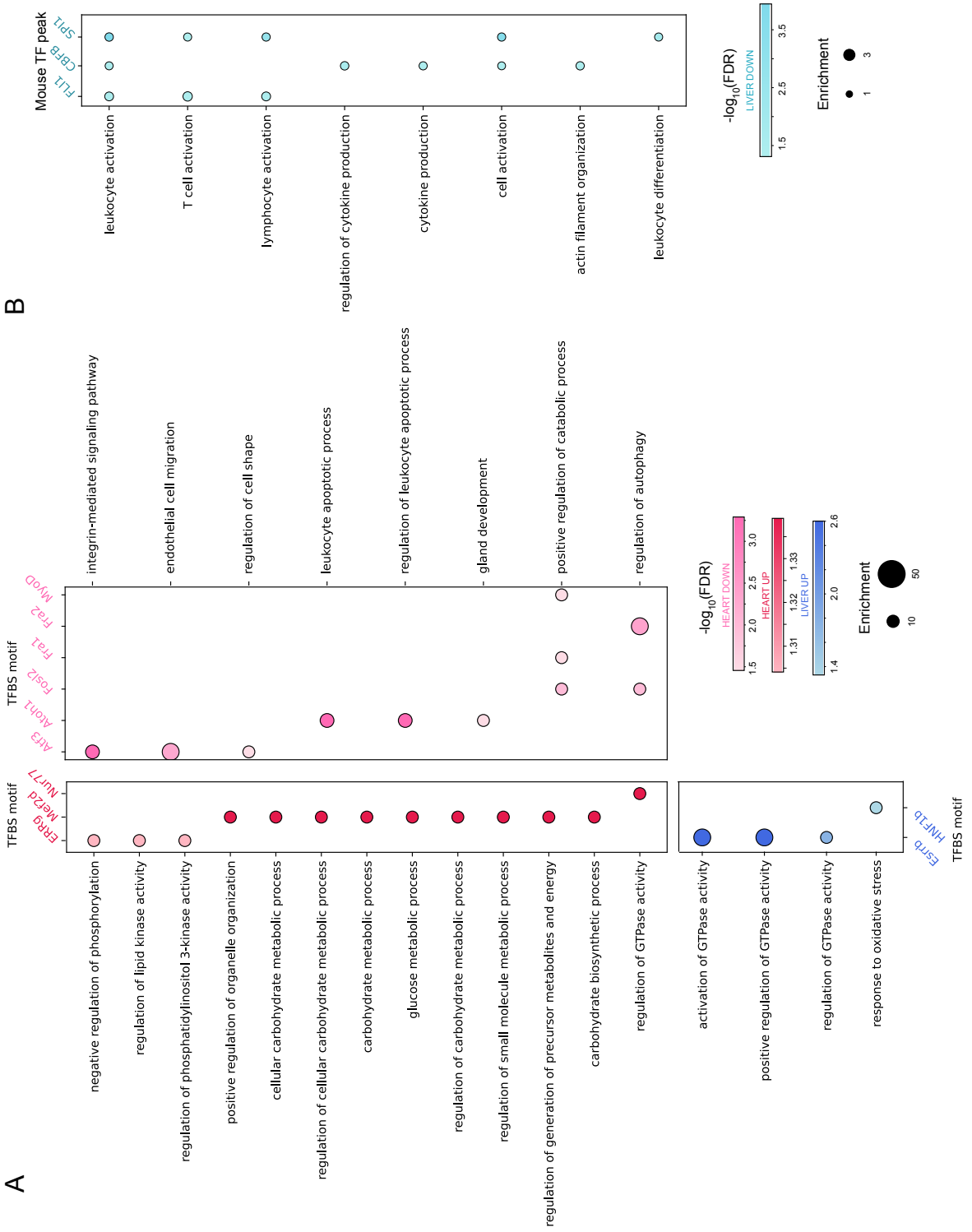
Supplemental Figure S6. 4C-seq validation of chromatin interactions between promoters on candidate loci and enhancer shifts in mole-rats (related to Figures 2, 3, and 4).

A-G. For each locus, tissue and species, top panels (4C-signal) show normalized interaction signals between promoter baits in each locus (Bait, orange arrow) and other genomic regions. Grey bars denote significant interactions (Methods). Enhancer panels below show candidate enhancer elements as defined in the study (red for “Up”, blue for “Down” and black for other). Significant contacts between candidate enhancers and 4C chromatin interactions are denoted with black arrows and asterisks.


Heart
Liver
Igf1r
Uox
Nebl
Ppargc1a
Foxp1
Foxo1
Nfatc3
Insr
Ppara
St6galnac3
Hacd3
Cers6
Epas1
Hif1a
Egln3
Pparg
Pc
Fbp1
Relative expression
Hgla Fdam Cpor Mmus

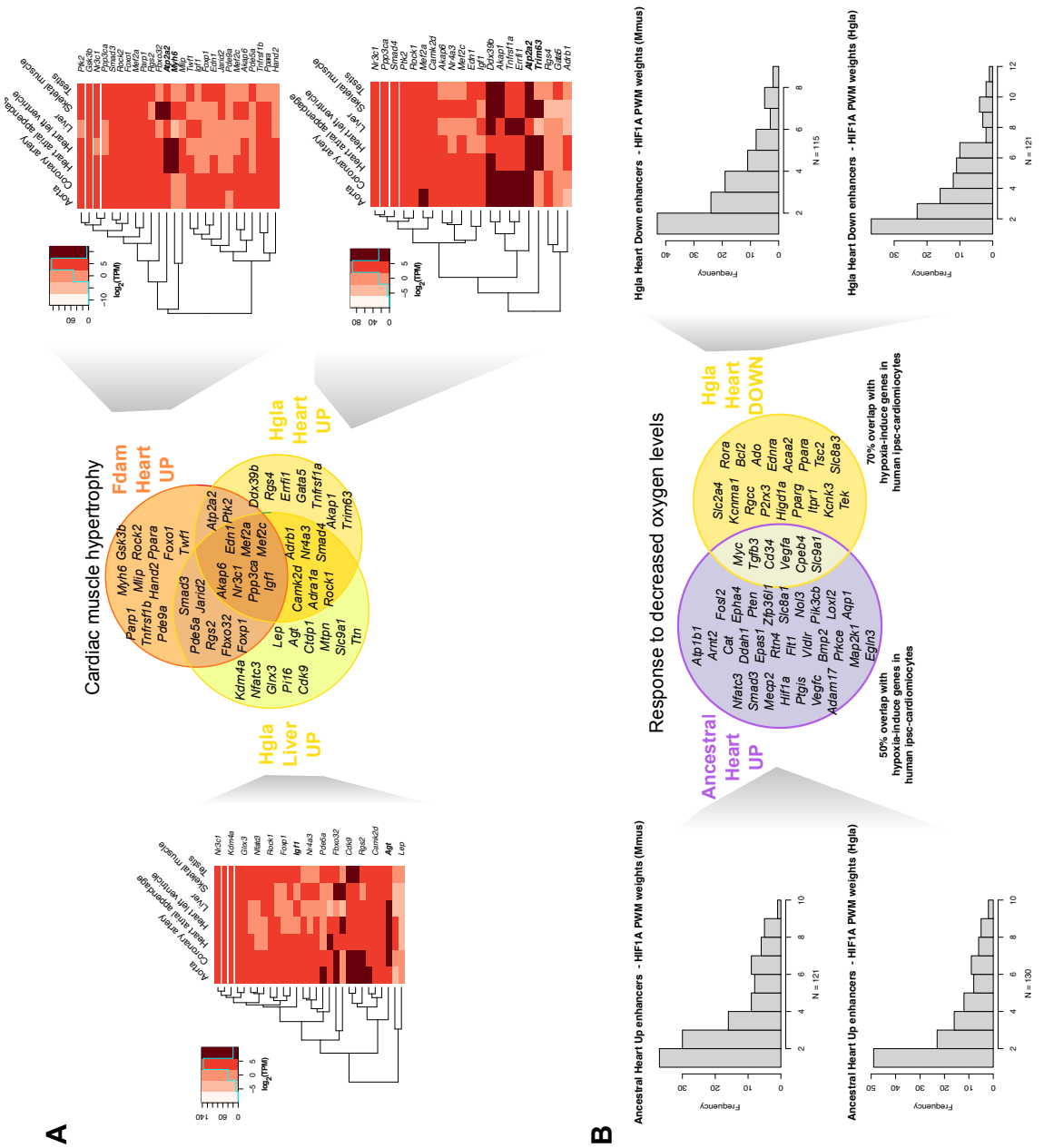
Supplemental Figure S7. Validation of mRNA expression levels for candidate genes associated with enhancer shifts in mole-rats (related to Figures 2, 3, 4 and 5)

Measurement of mRNA expression levels for candidate loci across the study. Each panel indicates relative gene expression across heart and liver samples from the four study species (N = 2-3 per species and tissue; Hlga, naked mole-rat; Fdam, Damaraland mole-rat; Cpor, guinea pig; and Mmus, mouse). * $p < 0.05$ ** $p < 0.001$ (one-way ANOVA, Tukey post-hoc correction). Candidate loci relate to biological processes highlighted in the study as follows: *Igf1r* (response to insulin, Figures 2 and 4), *Uox* (purine catabolism, Figure 2), *Nebl* (heart contraction, Figure 3), *Ppargc1a* (energy homeostasis, Figure 3); *Foxp1*, *Foxo1* and *Nfatc3* (cardiac hypertrophy, Figure 4); *Insr* and *Ppara* (response to insulin, Figure 4); *St6galnac3*, *Hacd3* and *Cers6* (sphingolipid biosynthesis, Figure 4); *Epas1*, *Hif1a* and *Egln3* (response to hypoxia, Figure 4); *Ppara* and *Pparg* (fatty acid oxidation, Figure 5); *Pc* and *Fbp1* (gluconeogenesis, related to Figure 3).



Supplemental Figure S8. Enriched transcription factor binding motifs and binding sites associated with specific Gene Ontology terms in heart and liver. (related to Figure 3)

A. Enrichment of TFBS motifs -GO pairs for Up and Down enhancers in heart, and Up enhancers in liver. Circle plots denote enrichment of TFBS – GO pairs overlap compared to background (circle size) and FDR statistical significance (circle shade: pink scale for Down enhancers in heart, red scale for Up heart and blue scale for Up in liver; permutation-based p-values were corrected for multiple testing with the BY procedure, Methods).
B. Enrichment of ChIP-seq mouse TF binding signals – GO pairs for Down enhancers in liver. Representation as in A.

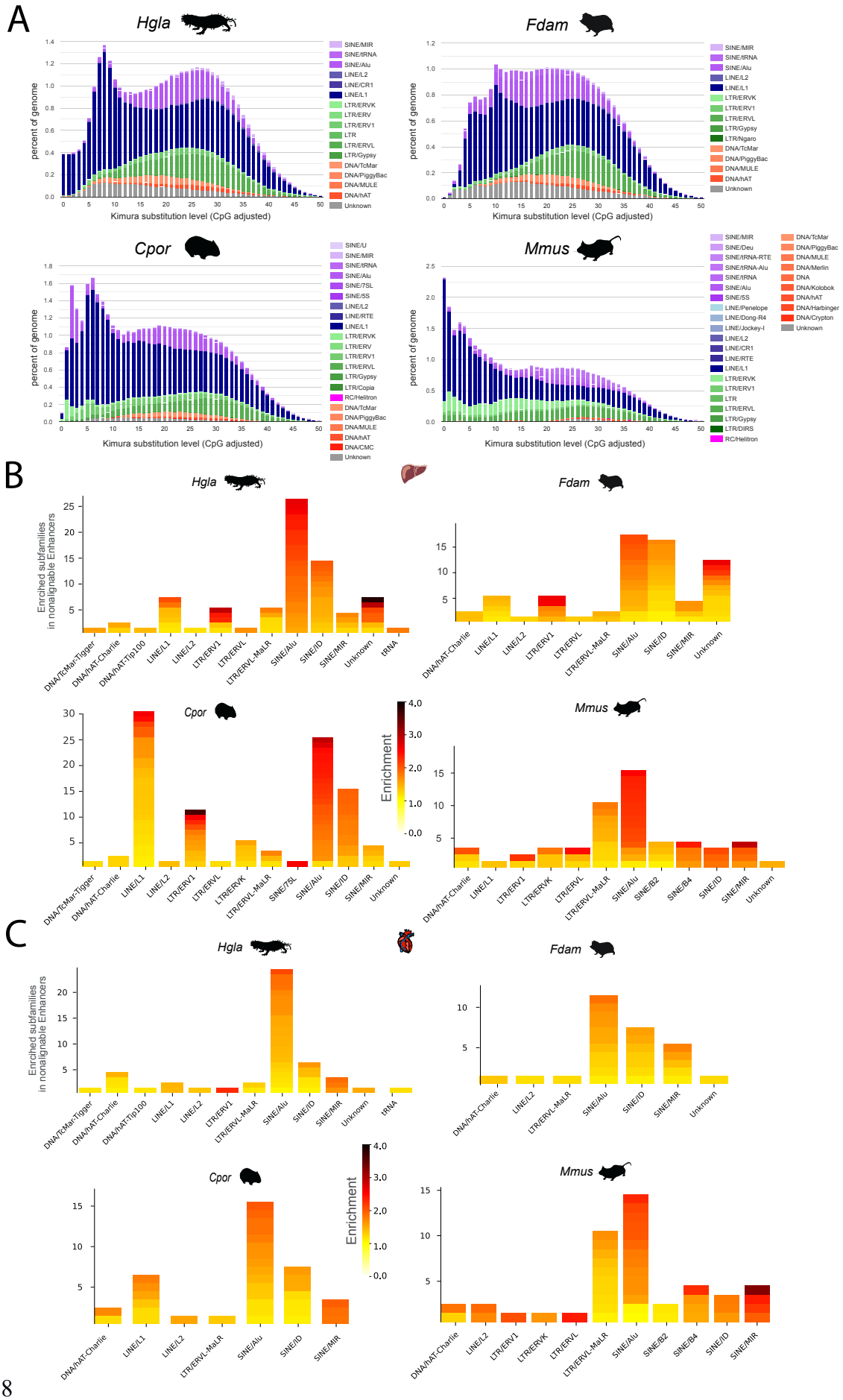


465

Supplemental Figure S9. Additional properties of ancestral and single-species enhancer shifts in selected ontology categories (related to Figure 4 and Discussion)

A. Gene expression heatmaps showing human gene expression levels across selected tissues from GTEX v8 data. Represented genes correspond to those associated to the
470 Cardiac hypertrophy ontology term and Up liver enhancers in the naked mole-rat branch (left), Up heart enhancers in the Damaraland mole-rat branch (upper right), and Up heart enhancers in the naked mole-rat branch (lower right).

B. Comparison of HIF binding affinities for mole-rat enhancer shifts proximal to genes in the Response to hypoxia ontology category, specifically for Up heart enhancers in the ancestral mole-rat branch (left) and Down heart enhancers in the naked mole-rat branch (right). For each, histograms show the distribution of predicted HIF binding affinities for orthologous locations of each elements set in mouse (top) and naked mole-rat (bottom), using the HIF1A position-weight matrix from JASPAR database (Core Vertebrate collection). The mouse and naked mole-rat HIF binding affinity score distributions are not significantly different for Up heart enhancers in the ancestral branch (Kolmogorov-Smirnov test, p=0.82). However, for Down heart enhancers in the naked mole-rat branch, predicted HIF binding affinities tend to be higher in mouse orthologous sequences compared to naked mole-rat sequences (Kolmogorov-Smirnov test, p=0.05).


A
Hgla
Fdam
Cpor
Mmus
percent of genome
Kimura substitution level (CpG adjusted)
SINE/MIR
SINE/tRNA
SINE/Alu
LINE/L2
LINE/CR1
LINE/L1
LTR/ERVK
LTR/ERV
LTR/ERV1
LTR
LTR/ERVL
LTR/Gypsy
DNA/TcMar
DNA/PiggyBac
DNA/MULE
DNA/hAT
Unknown
SINE/U
SINE/7SL
SINE/5S
LINE/RTE
LTR/Copia
RC/Helitron
DNA/CMC
SINE/Deu
SINE/tRNA-RTE
SINE/tRNA-Alu
LINE/Penelope
LINE/Dong-R4
LINE/Jockey-I
LINE/CR1
LTR/DIRS
DNA/Kolobok
DNA/Merlin
DNA
DNA/Harbinger
DNA/Crypton
RC/Helitron
B
Enriched subfamilies in nonalignable Enhancers
DNA/TcMar-Tigger
DNA/hAT-Charlie
DNA/hAT-Tip100
LINE/L1
LINE/L2
LTR/ERV1
LTR/ERVL
LTR/ERVL-MaLR
SINE/Alu
SINE/ID
SINE/MIR
Unknown
tRNA
DNA/hAT-Charlie
SINE/7SL
SINE/B2
SINE/B4
Enrichment
C

Supplemental Figure S10. Additional properties of repetitive elements in mole-rat genomes and their association with nonalignable enhancers (related to Figure 5)

A. Landscape plots for repeat annotations in each of the four rodents. For mole-rats, we generated *de novo* repeat libraries with RepeatModeller. For mouse and guinea pig, we re-used publicly available *de novo* repeat libraries from Dfam. Landscape plots were drawn using scripts from the RepeatMasker suite.

B-C. Repeat families enriched in nonalignable enhancers in liver (B) and heart (C) for each of the four rodents (permutation tests, Methods, corrected p-values <0.01). For each family, the number of enriched sub-family is indicated by the height of bars, colours for each sub-family are proportional to enrichment ratios.