SUPPLEMENTAL INFORMATION

Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression

Berthelot, Villar, Horvath, Odom, Flicek

SUPPLEMENTAL DATA

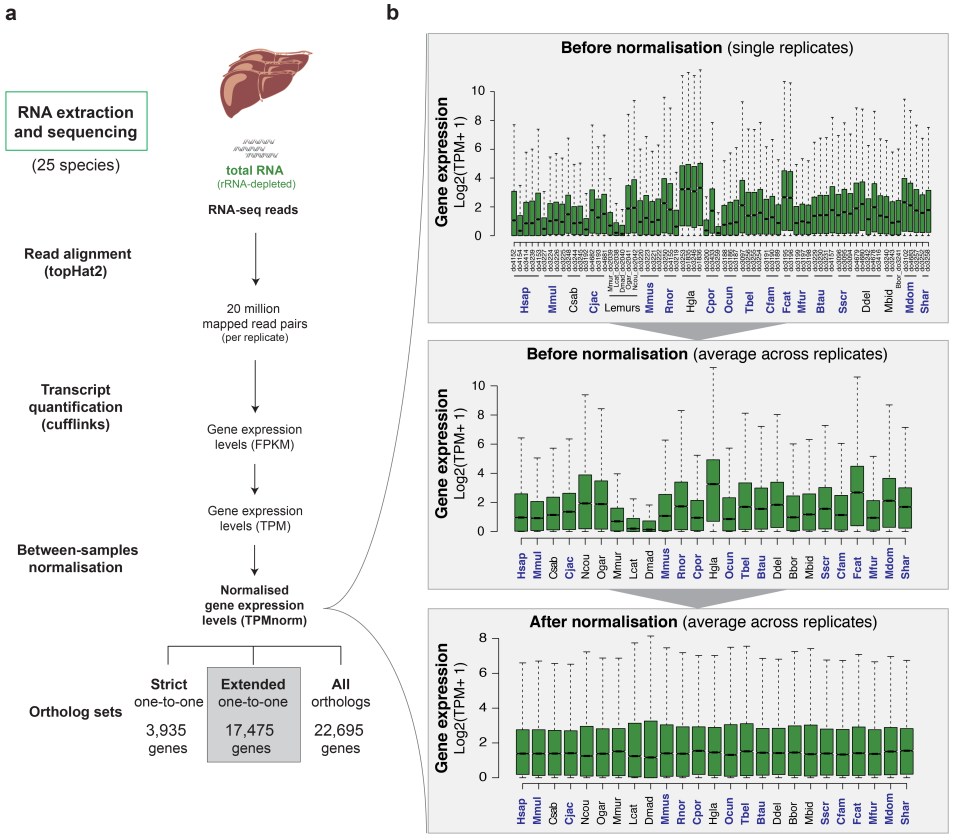
SUPPLEMENTAL DATA

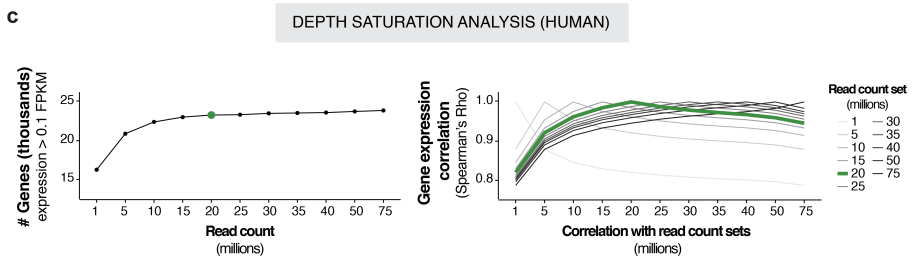
SUPPLEMENTARY TEXT 1.

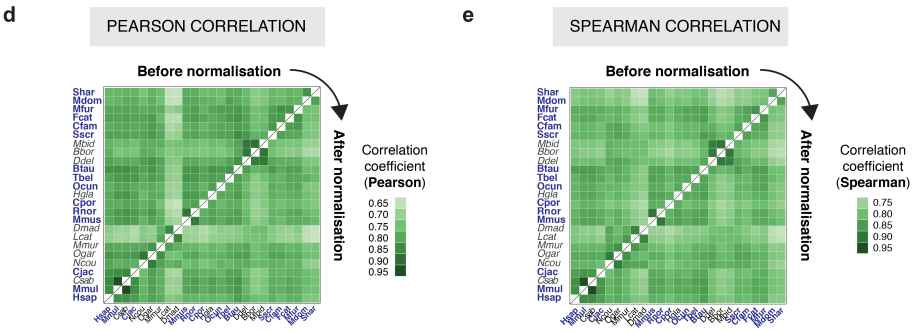
Gene expression correlations are confounded by gene expression level (related to Figure 1 and Figure S2)

In order to connect gene expression evolution with regulatory activity across species, we sought to analyse transcriptional output of gene sets associated to different subsets of regulatory elements. Pairwise correlations of orthologous expression levels have been a widely adopted method for this type of analysis[1-4]. However, comparing correlation coefficients between different gene sets is potentially confounded by expression levels. Indeed, gene sets with lower average expression and/or tighter dynamic ranges may be more sensitive to measurement uncertainty or small variations between species, resulting in noisier expression correlations even in the absence of increased evolutionary divergence. Previous work on gene expression divergence between human and mouse suggests that this may be the case[2].

We therefore investigated whether the evolutionary stability of gene expression between species, as measured by expression correlation, depends on their expression levels. We stratified the set of orthologs by quartiles of average expression across species. Genes in the mostly highly expressed quartile are more strongly correlated across species than are genes of moderate expression; the same is true for genes in the lowly expressed quartiles (Figure S2C). Additionally, each of these subsets taken individually exhibited lower correlation across species than the whole set (Figure S2A). While unsurprising, these results indicate that any attempt to compare the evolution of subsets of genes is potentially confounded by gene expression level. To circumvent this issue, throughout this analysis all gene sets of interest are compared to sets of control genes matched 1-to-1 on expression levels (Methods).


a
RNA extraction and sequencing
(25 species)
total RNA
(rRNA-depleted)
RNA-seq reads
Read alignment (topHat2)
20 million mapped read pairs (per replicate)
Transcript quantification (cufflinks)
Gene expression levels (FPKM)
Gene expression levels (TPM)
Between-samples normalisation
Normalised gene expression levels (TPMnorm)
Ortholog sets
Strict one-to-one 3,935 genes
Extended one-to-one 17,475 genes
All orthologs 22,695 genes
b
Before normalisation (single replicates)
Before normalisation (average across replicates)
After normalisation (average across replicates)
Gene expression Log2(TPM+1)
Lemurs


c
DEPTH SATURATION ANALYSIS (HUMAN)
Genes (thousands) expression > 0.1 FPKM
Read count (millions)
Gene expression correlation (Spearman's Rho)
Correlation with read count sets (millions)
Read count set (millions)


d
PEARSON CORRELATION
e
SPEARMAN CORRELATION
Before normalisation
After normalisation
Correlation coefficient (Pearson)
Correlation coefficient (Spearman)

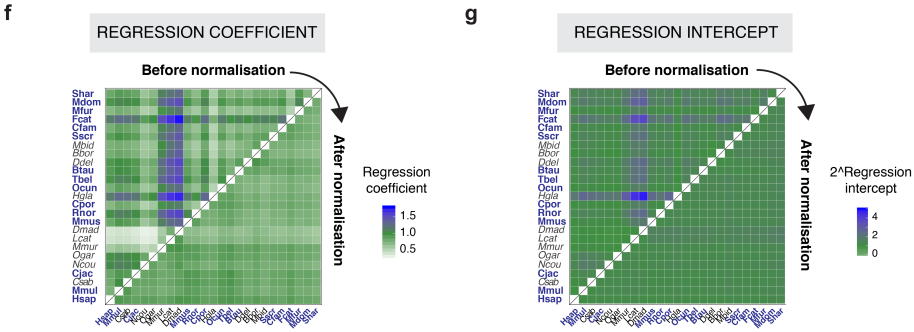

f
REGRESSION COEFFICIENT
g
REGRESSION INTERCEPT
Before normalisation
After normalisation
Regression coefficient
2^Regression intercept

Figure S1. Analysis pipeline and inter-species normalisation of gene expression levels (related to Figures 1 and 2)

(a) RNA-seq analysis pipeline. RNA-sequencing libraries were prepared from total RNA samples (rRNA-depleted) from liver tissue. After read alignment and transcript quantification, gene expression levels were transformed to transcripts per million (TPM) and averaged across all replicates for every species. Gene expression levels were compared for orthologous gene sets, focusing on an "Extended" set of 17,475 one-to-one orthologs comprising all or some of the 25 species compared (in a similar strategy as [5]; see also extended Methods). Orthologous gene expression levels were normalised across species using the median of ratios to the geometric means[6].

(b) Ranges of gene expression (TPM, logarithmic scale) are represented before between-samples normalisation for individual replicates (top) and the average gene expression across replicates from the same species (middle). Between-samples normalisation results in homogeneous ranges of measured gene expression across species (bottom). Species names in bold blue correspond to good quality genome assemblies, with other species depicted in black font.

(c) Number of detected genes (left; measured as fragments per kilobase of exon per million reads mapped (FPKM) > 0.1) and gene expression correlations (right; Spearman) in a representative human sample, with read depth subsampled from one to 75 million mapped reads. For gene expression correlations (right), each line represents the Spearman correlations of expression estimates from a specific read count threshold ("read count sets" legend) with all other estimates ("correlation with read count sets", y-axis). The chosen read count threshold of twenty million reads is highlighted in dark green in both plots.

(d-g) Quality control of the expression normalisation procedure across orthologous genes. Heatmaps compare measures of gene expression correlation across species before and after inter-species normalisation. Pearson correlations **(d)** become more homogeneous after normalisation (rank-based Spearman correlations **(e)** remain unchanged, as expected, as the normalisation is a scaling procedure). Normalisation across samples leads to largely homogeneous regression coefficients (close to one after normalisation, **f**) and regression intercepts (close to zero after normalisation, **g**) on log2-transformed values.

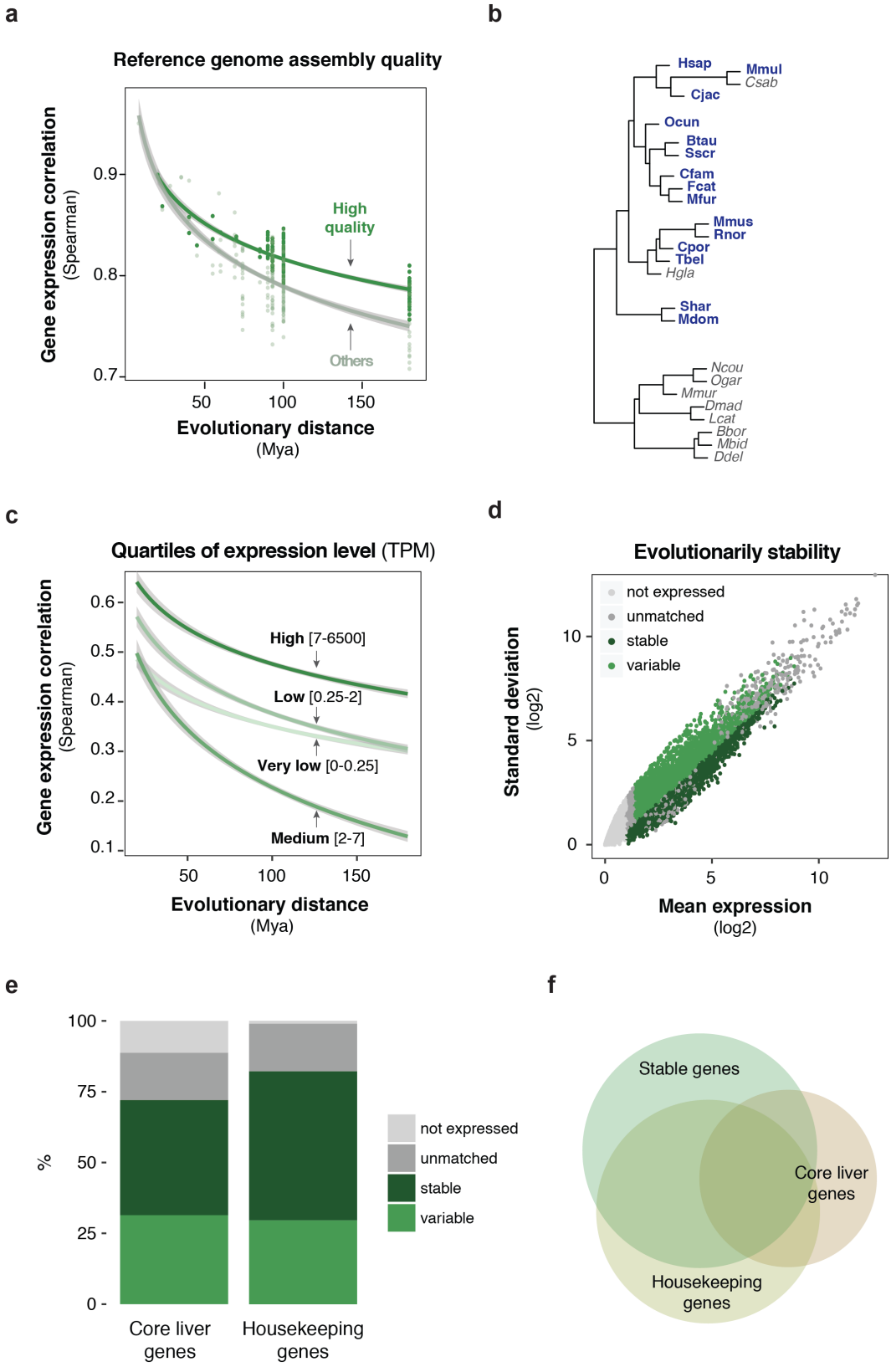


Figure S2. Quality control on measures of gene expression stability (related to Figure 1).

(a) Correlations of gene expression levels between pairs of species decrease with evolutionary distance as expected, though measurements of gene expression

divergence are affected by reference genome quality (correlation coefficient: Spearman's Rho). Pairs of species with higher-quality genomes typically showed higher correlation at similar evolutionary distances (dark green). Solid lines represent a linear regression fit after logarithmic transform of the x-axis, surrounded by grey shading of a 95% confidence interval.

(b) Hierarchical clustering of pairwise gene expression correlations across species recapitulates the mammalian phylogeny for well-annotated genomes (Spearman's Rho, as in (a); species highlighted in blue as in a). Species lacking well-annotated genomes (in grey) largely clustered at the base of the tree regardless of their known phylogenetic relationships, consistent with long-branch attraction. Thus species in grey were excluded from analysis, and the 15 species higlighted in blue were used for the gene expression analyses herein.

(c) Gene expression correlation depends on expression levels. Genes were stratified by expression level (mean across species) into quartiles, and pairwise correlations of gene expression between species were plotted against evolutionary distance. Highly-expressed genes were most correlated between species (dark green, 7-6500 TPM). Genes with medium expression levels (medium green, 2-7 TPM) were the least correlated across species, with low and very low level genes between these two extremes (pale and very pale green, 0.25-2 and 0-0.25 TPM, respectively). For readability only linear regression fits are shown (see (a)).

(d) Expressed genes with evolutionarily stable and variable expression levels were identified based on their coefficient of variation across species (CV; standard deviation normalized by mean expression). Variable genes (top 50% of the CV distribution) are highlighted in light green. Stable genes (bottom 50% of the CV distribution) are highlighted in dark green. Genes in either category with no matched counterpart at the same expression level in the other category were not considered for analysis (in dark grey, "unmatched"). Non-expressed genes are in light grey (mean expression < 1 TPM across species).

(e) Core liver and housekeeping genes were categorised into stable and variable based on the analysis in (d).

(f) Core liver and housekeeping genes account for the majority of evolutionarily stable gene expression, suggesting that evolutionary stability of gene expression associates with core tissue function.

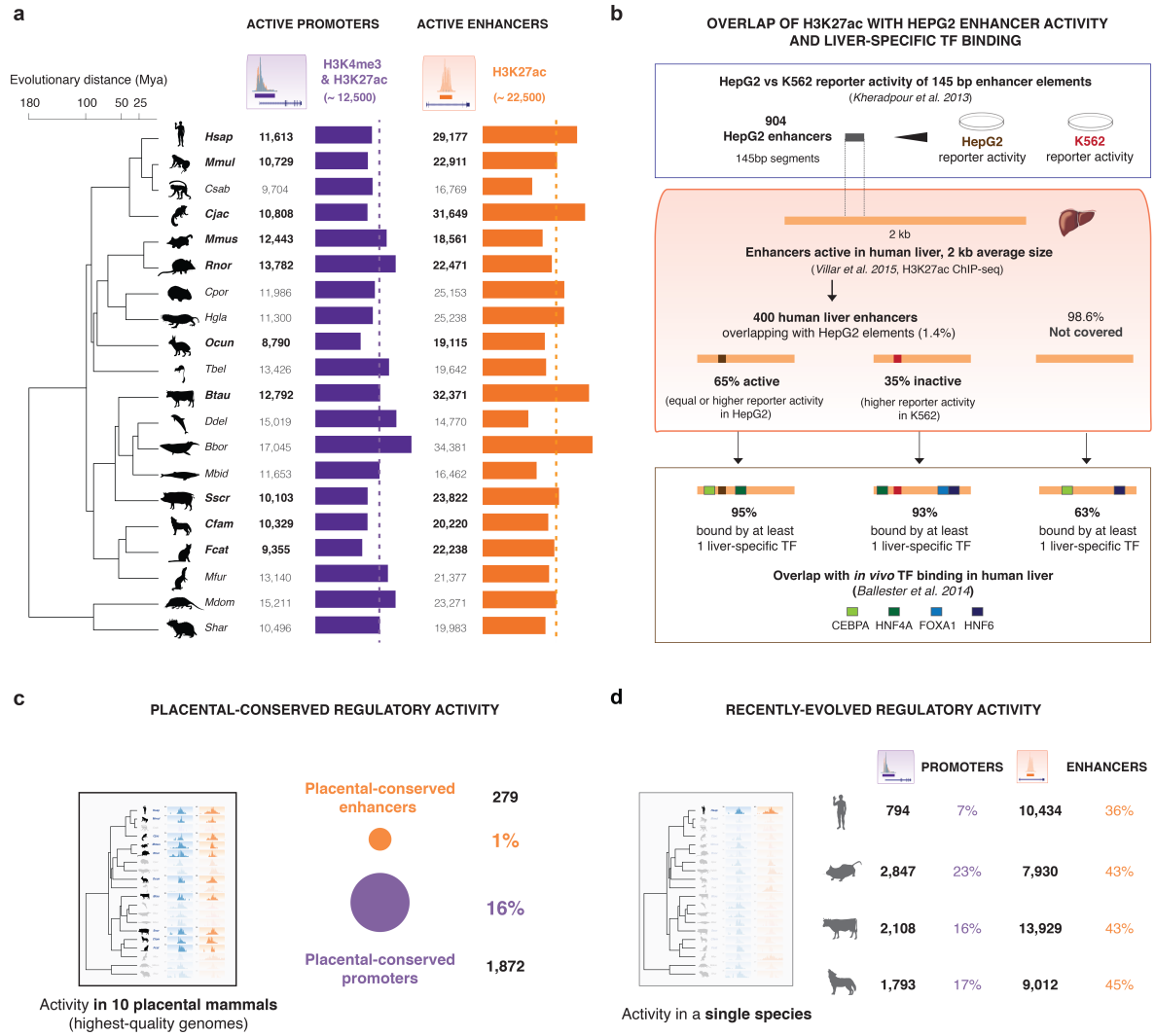


Figure S3. Summary of previously reported regulatory activity across twenty mammalian species (related to Figures 2 to 5)

(a) Summary of chromatin immunoprecipitation sequencing (ChIP-seq) data previously reported for active histone marks H3K4me3 and H3K27ac across twenty mammalian species[7]. Data in each species is categorised as active promoters (purple) for regions enriched in H3K4me3, alone or in combination with H3K27ac (H3K4me3&H3K27ac); and active enhancers (orange) for regions enriched only in H3K27ac. Solid bars indicate number of active promoters or enhancers identified across biological replicates (two or more) in each species, with the exception of *Balaenoptera borealis* (Bbor), where a single individual was profiled. Dashed lines indicate the average number of active promoters or enhancers identified across the twenty species. Species in bold correspond to the ten species with reference genomes aligned against each other in the multiple whole-genome alignment (Methods). Other species are aligned to one or several of these ten species using pairwise whole-genome alignments.

(b) H3K27ac-defined enhancers enrich for regulatory activity: Human liver enhancers identified through H3K27ac ChIP-seq (central inset) were overlapped with 145 bp sequence elements assayed for reporter activity in human liver carcinoma (HepG2) and human erythroleukemia cells (K562) (top inset; [8]). These correspond to enhancer

candidates identified in HepG2 cells and containing motifs for liver-specific transcription factors.

Four hundred human liver enhancers contained at least one 145 bp segment (1.1 segments per enhancer on average). 65% of these enhancers were active based on the reporter activity of the assayed segments, which displayed higher activity in HepG2 compared to K562 cells, or equal activity in both cell lines. The remaining 35% human liver enhancers overlapped segments having higher activity in K562 cells, and were thus classified as inactive in HepG2 cells. Grey inset: Human liver enhancers identified in this study were overlapped with *in vivo* binding locations for four liver-specific transcription factors, as reported independently in human liver samples[9]. Among the 400 enhancers containing segments assayed in Kheradpour et al., 93-95% of them were bound by at least one liver-specific TF, regardless of the reporter activity of their overlapping segments. This suggests that in cases where the overlapping segment was inactive in the reporter assay, the corresponding enhancer may harbour regulatory activity outside the interrogated sequence. Across all liver enhancers in human, 63% are bound by at least one of the four liver-specific transcription factors, in line with previous estimates of functional enhancer activity in H3K27ac-marked regions[10].

(c) Genomic regions with highly-conserved promoter and enhancer activity were identified across placental mammals by comparing ChIP-seq readouts across ten species with high-quality genomes. 279 liver enhancers (orange) showed placental-conserved activity, corresponding to 1% of all enhancers active in human liver samples. In comparison, 1,872 promoters were identified as placental-conserved, accounting for 16% of active liver promoters in human. These elements show increased signals of sequence selection and are enriched in transcription initiation sequences and transcription factor binding sites associated with liver-specific functions[7].

(d) Genomic regions with promoter or enhancer activity exclusive to a single species were identified using one well-assembled reference species in each placental lineage (human, mouse, dog and cow). For example in human, these regulatory elements have likely become active in the 23 million years since human and macaque diverged, and we termed these “recently-evolved” regulatory elements. Recently-evolved promoters (purple) ranged from 794 to 2,847, accounting for 7-23% of all active promoters in each reference species. In comparison, recently-evolved enhancers were pervasive (7,930 to 13,929, representing 36-45% of all active enhancers per species). We note that “recently-evolved” refers to regulatory elements whose activity has been acquired since the split from the closest species in our study phylogeny. Therefore these elements are not fully consistent in terms of their age across our reference species.

Figure modified from [7]

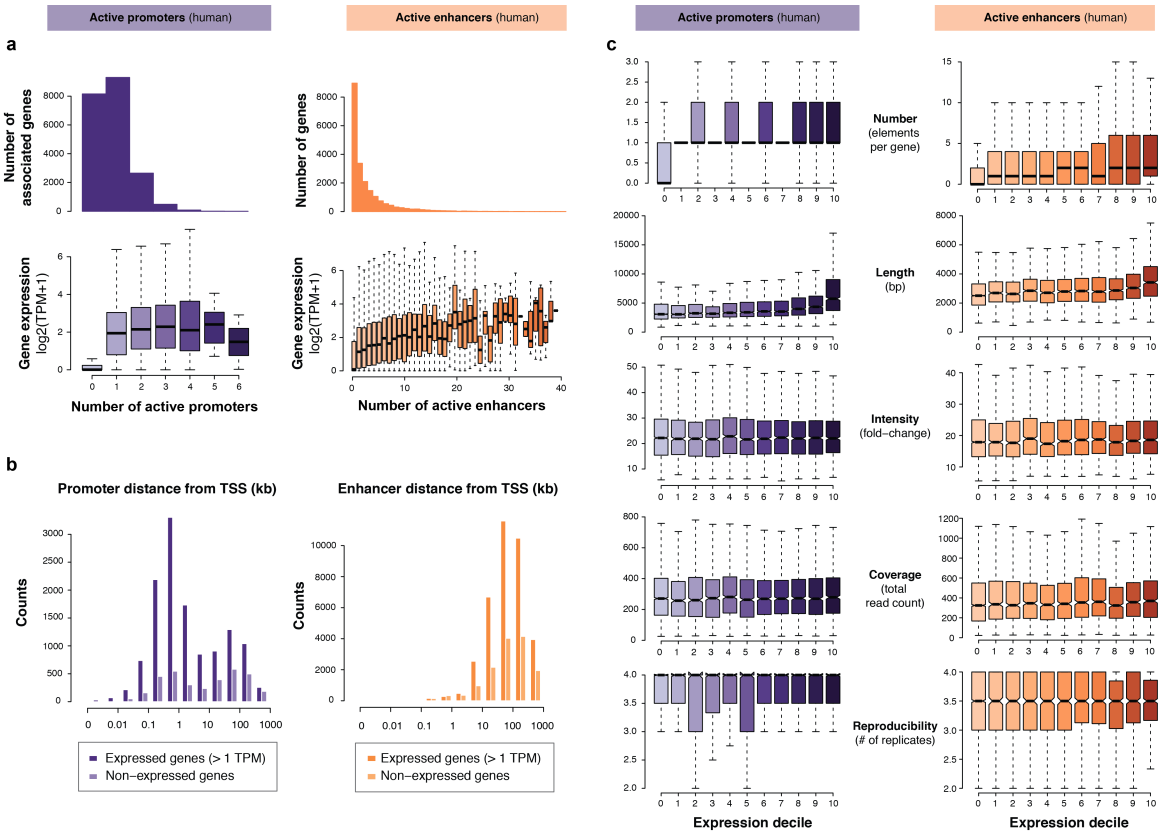


Figure S4. Human regulatory activity correlates with proximal gene expression levels (related to Figure 2)

(a) Number of active promoters (purple) and enhancers (orange) associated to human genes (17,475 one-to-one orthologs; top panels) and gene expression levels (TPM, logarithmic scale) in each category (bottom panels). Increasing numbers of associated enhancers correlate with gradual increases in average gene expression, whereas gene expression remains largely constant with increasing numbers of active promoters.

(b) Genomic distances separating active regulatory regions and the canonical transcription start site (TSS) of their putative target genes in human (promoters: left, purple; enhancers: orange, right). The distance was measured between the midpoint of the regulatory region and the TSS. As expected, promoters are typically proximal to their target genes (within a few kilobases of the predicted target TSS), while enhancers are typically distal. While representing only 57% of the gene set, expressed human genes (> 1 TPM on average in human, dark colours) were much more likely to be associated with active promoters and enhancers than non-expressed genes (light colours), making up 75% of the promoter-gene pairs and 70% of the enhancer-gene pairs.

(c) Experimental characteristics of active promoters (left column) and enhancers (right column) associated to genes with increasingly high expression. Expressed human genes (> 1 TPM) were categorised into deciles of gene expression (x-axis), and the properties of their associated active promoters and enhancers were measured for each decile [7]. Both the length of active promoters (and, to a slight degree, enhancers) and the number of active enhancers associate with higher gene expression levels. ChIP-seq signal intensity (fold-change over input), coverage (reads per peak) or reproducibility across biological replicates remained similar across all deciles of gene expression. The “0” decile groups all genes with expression lower than 1 TPM.

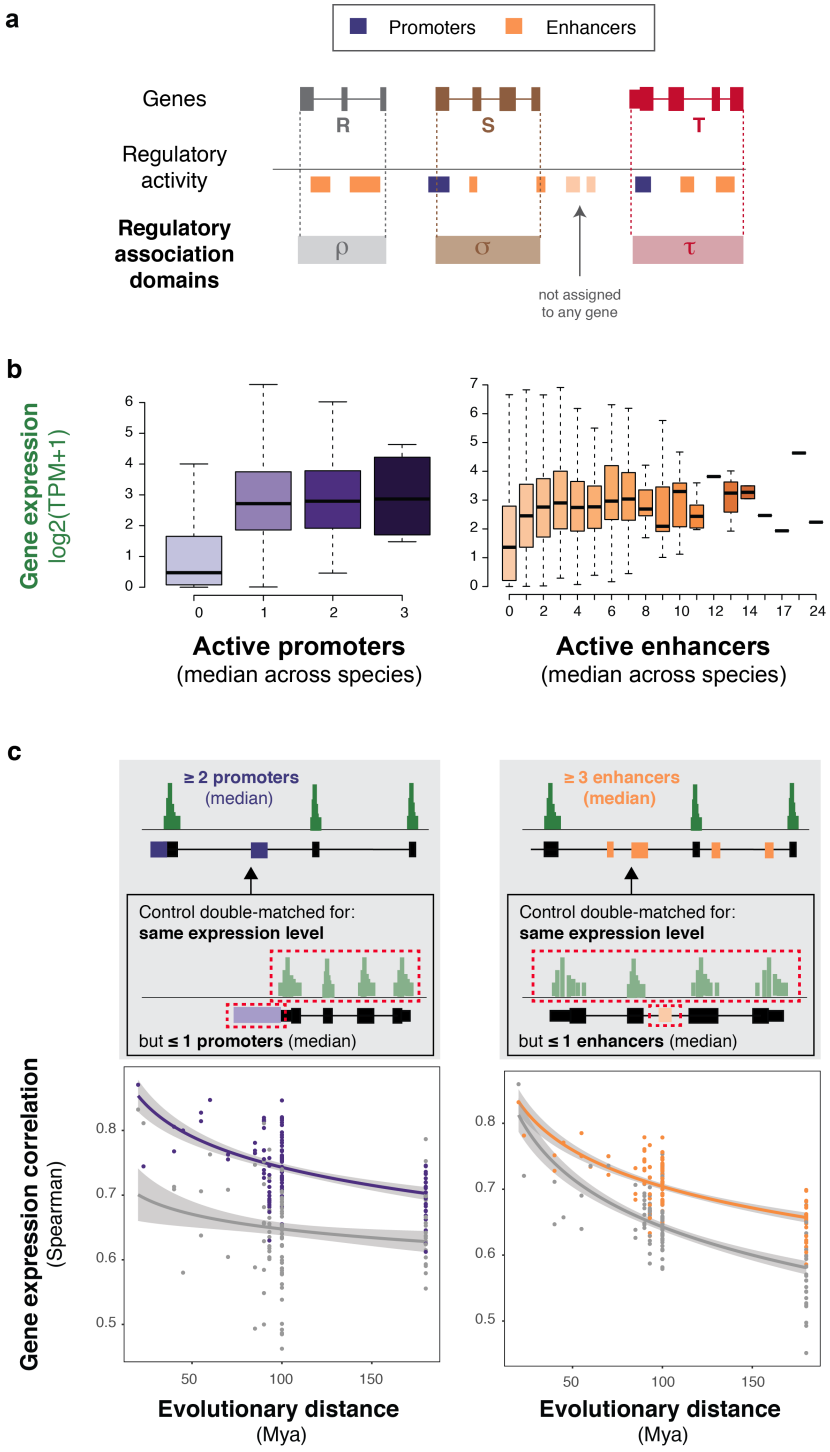


Figure S5. Genes with complex regulatory landscapes are more expressed and more conserved in expression regardless of target assignment method (related to Figure 2)

(a) Genes are associated with all regulatory elements overlapping the gene locus. This target assignment method is more conservative than the one used throughout the manuscript, and only associates a fraction of active promoters or enhancers to a target gene (across all species: 43-93% of promoters and 32-64% of enhancers assigned to one target or more). For comparison, the method used throughout the manuscript assigns > 99% of elements to one target or more, across the vast majority of our study species.

(b) When restricting the regulatory landscapes of genes to active elements that overlap the gene locus as described in **(a)**, we confirm that in an average mammal, gene expression is related to the complexity of its landscape. Shown are the distributions of gene expression (mean across species) across all orthologs, stratified by landscape complexity (median number of active promoters across species, left; or enhancers, right).

(c) When restricting the regulatory landscapes of genes to active elements that overlap the gene locus as described in **(a)**, we confirm that the number of promoters and enhancers associated per gene contributes to evolutionary stability of gene expression. **Grey insets**: Gene expression divergence across species is compared between (i) genes associated to multiple promoters or enhancers (top) and (ii) control genes with the same expression level but associated to few promoters or enhancers (one or none, bottom). **Plots**: Pairwise Spearman correlation coefficients of expression levels between species were plotted against evolutionary distance for genes associated with multiple promoters (left) or enhancers (right), and compared to control gene sets (in grey). In both cases the number of associated promoters or enhancers corresponds to the median number across species. Lines are as described in Figure 1b-c.

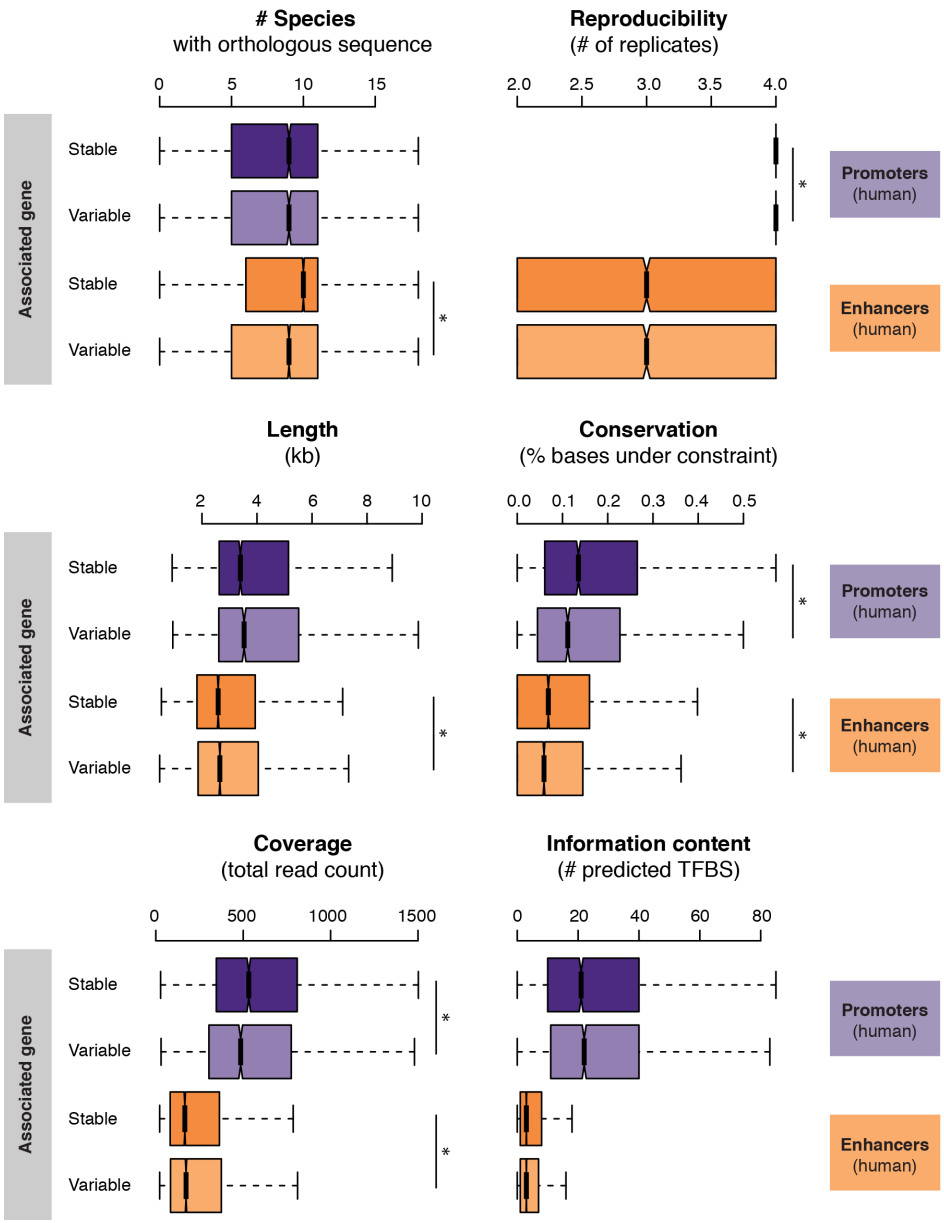


Figure S6. Human regulatory elements associated with genes with either stable or variable transcription are largely similar in experimental and sequence characteristics (related to Figure 2)

Experimental and sequence properties of human active promoters (left) and enhancers (right) associated to genes with either evolutionarily stable or variable expression (Figure S2): number of species with an alignable orthologous sequence; consensus region length; ChIP-seq signal coverage (reads per peak); peak reproducibility across biological replicates; fraction of bases under constraint as defined by GERP[11]; and number of predicted transcription factor binding sites using HOMER[12]. Regulatory elements associated with stable or variable genes showed minute differences especially in length and fraction of sequence under constraint, but the dynamic ranges were largely similar across both categories for all characteristics. Promoters associated with stable genes: n = 6,705; with variable genes: n = 6,166; enhancers associated with stable genes: n = 30,350; with variable genes: n = 30,871; * : $p < 0.05$ (Wilcoxon rank sum test, Bonferroni-corrected).

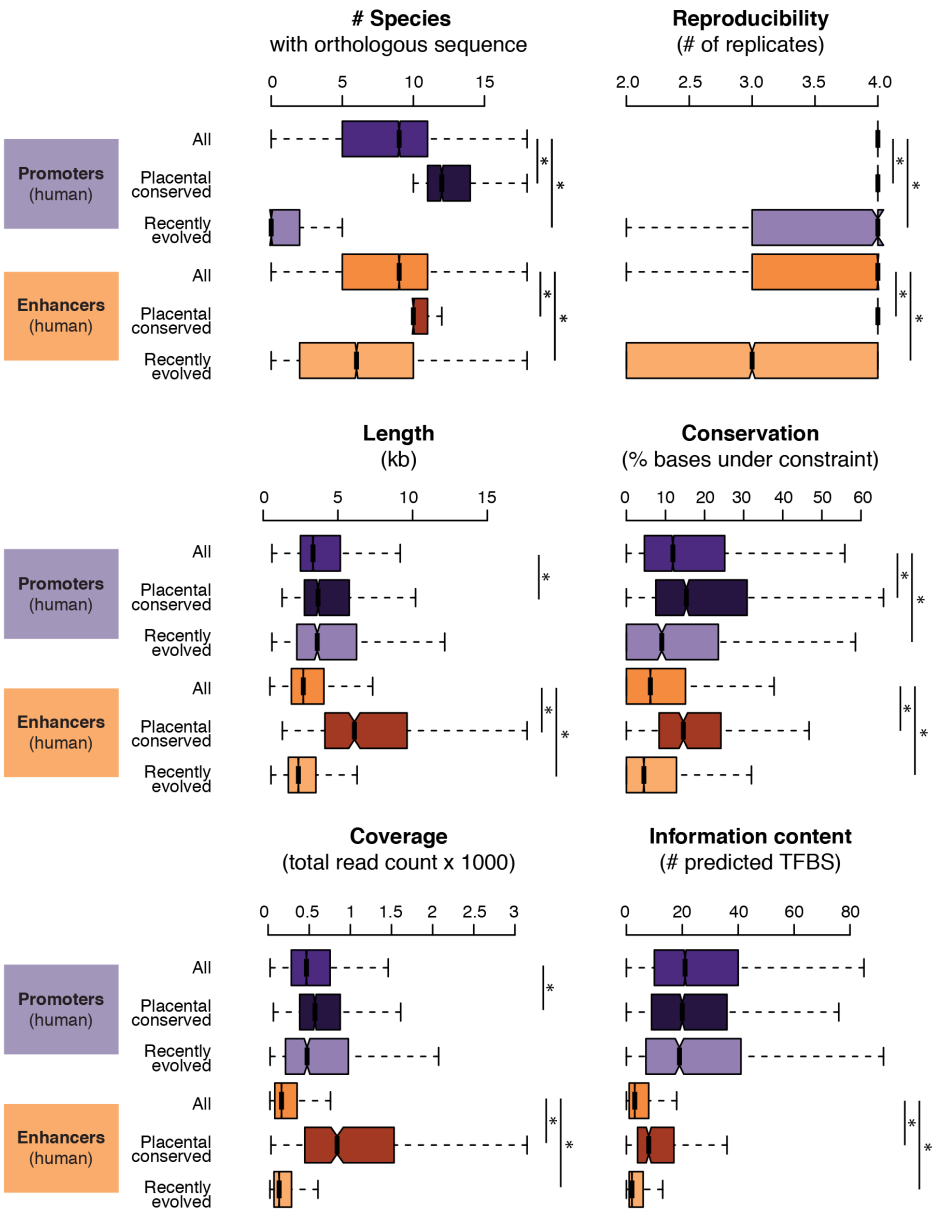


Figure S7. Experimental and sequence characteristics of placental-conserved and recently-evolved regulatory elements in human (related to Figure 3 and 4)

Experimental and sequence properties of human active promoters (left) and enhancers (right) sorted by activity conservation level (all elements, placental-conserved and recently-evolved): number of species with an alignable orthologous sequence; consensus region length; ChIP-seq signal coverage (reads per peak); peak reproducibility across biological replicates; fraction of bases under constraint as defined by GERP[11]; and number of predicted transcription factor binding sites using HOMER[12]. Placental-conserved elements, and especially enhancers, showed evidence of higher functional importance compared to the background. Recently-evolved elements only marginally differed in distribution from the background. Placental-conserved promoters: n = 1,760; recently evolved promoters: n = 787; placental-conserved enhancers: n = 276; recently-evolved enhancers: n = 10,434. * : $p < 0.05$ (Wilcoxon rank sum test, Bonferroni-corrected). Further experimental and sequence properties of conserved and recently-evolved elements were reported in our previous work[7].

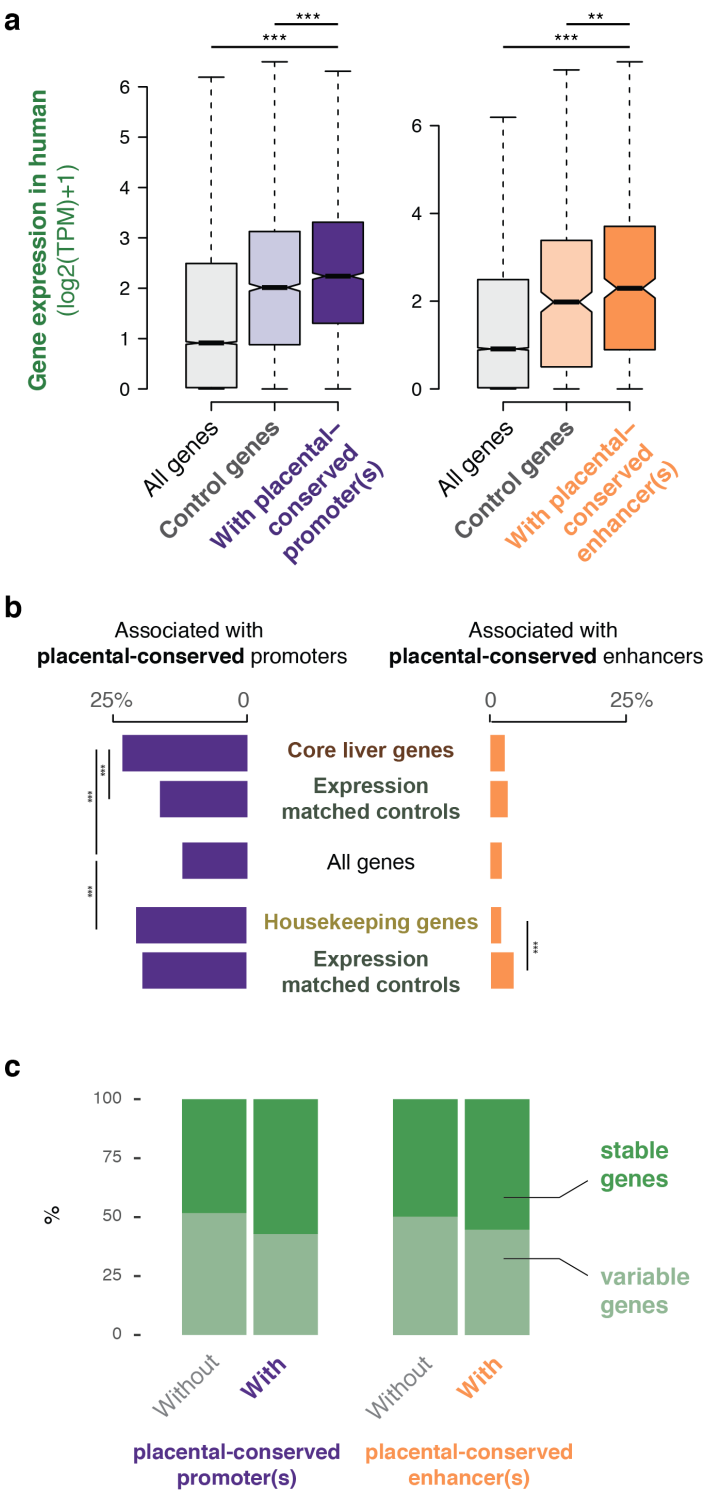


Figure S8. Association of placental-conserved regulatory elements with core liver and housekeeping genes (related to Figure 3)

(a) Genes associated with placental-conserved regulatory elements exhibit higher expression levels in a representative species (human) than either background (all genes) or control genes with the same number of regulatory elements, but none of them placental-conserved, as observed using the mean expression across species (Figure 3). **: $p < 0.01$; ***: $p < 0.001$, Wilcoxon test.

(b) Barplots show the proportion of genes associated with at least one placental-conserved promoter (purple) or enhancer (orange) for core liver genes (top, followed by a control set of genes matched for expression level), all 1-to-1 orthologs (middle), and housekeeping genes (bottom, followed by a control set of genes matched for expression level). Both core liver genes and housekeeping genes are preferentially associated with placental-conserved promoters compared to the entire gene set (23% and 21%, respectively; Chi-squared test: $p = 2*10^{-16}$ and $6*10^{-9}$). Core liver genes are also more likely to be associated with placental-conserved promoters than similarly expressed controls. No association was found between placental-conserved enhancers and core liver genes, while housekeeping genes were less likely to be associated with placental-conserved enhancers than expected based on their expression levels (2%; Chi-squared test: $p = 1*10^{-8}$). ***: $p < 0.001$, Chi-squared test.

(c) Genes associated with placental-conserved promoters (left) are more likely to be classified as stable based on the coefficient of variation of their expression levels across species (odds ratio: 1.33; Chi-squared test: $p = 1*10^{-10}$). Putative targets of placental-conserved enhancers showed a similar but non-significant enrichment (odds ratio: 1.24; $p = 0.10$), possibly due to the small number of such genes that could be classified (253 genes).

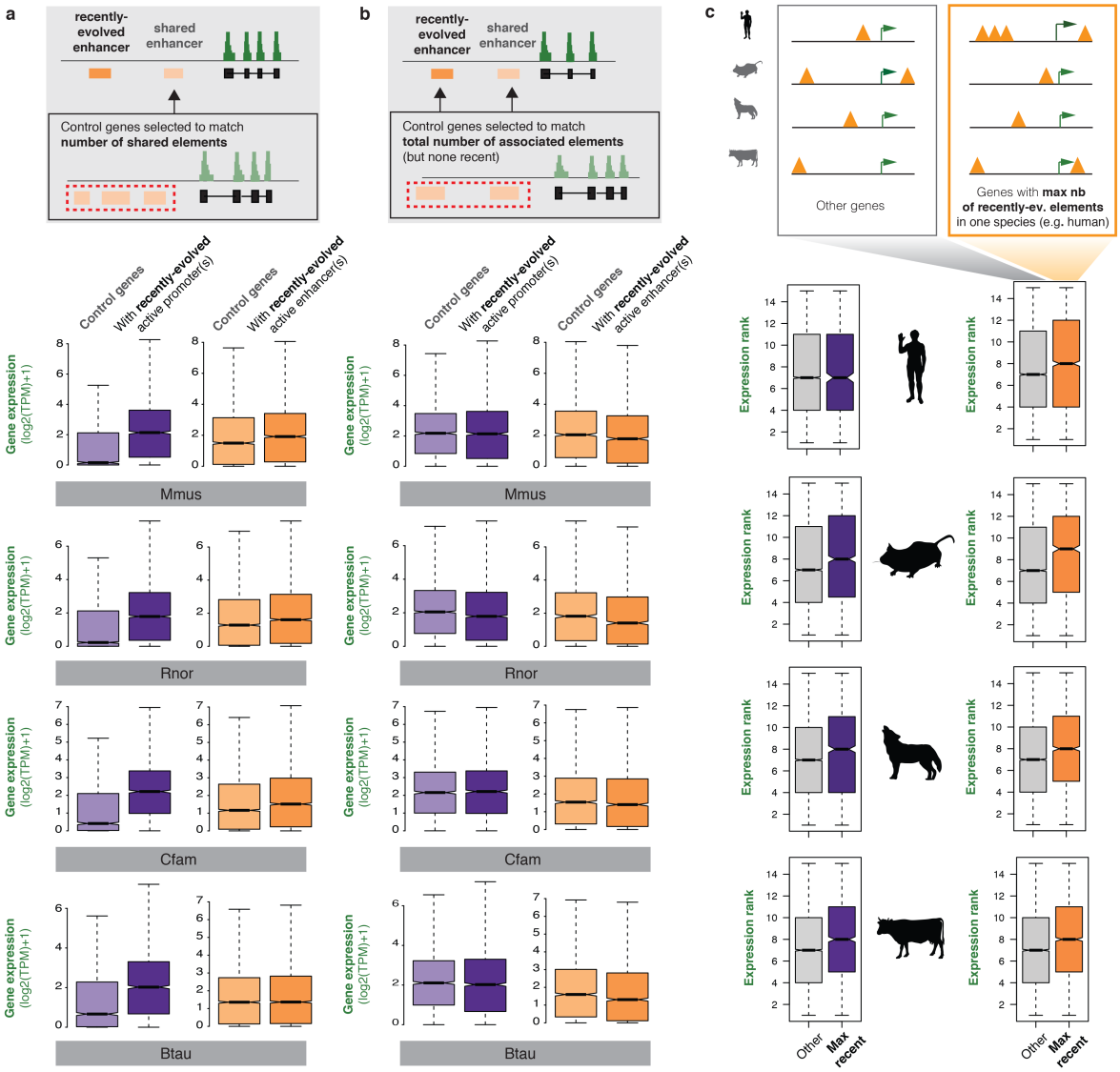


Figure S9. Contribution of recently-evolved regulatory elements to gene expression in additional reference species (related to Figure 4)

For each reference species, gene expression distributions (log2(TMP+1)) are compared for genes associated with recently-evolved promoters (purple) or enhancers (orange) and control genes.

(a) Specific contributions of recently-evolved promoters and enhancers to gene expression levels in mouse, rat, dog and cow. Control genes are matched for the same number of shared elements (top grey inset).

(b) Relative contributions of recently-evolved promoters and enhancers to gene expression levels in mouse, rat, dog and cow. Control genes matched for the same total number of associated elements (top grey inset).

(c) Large numbers of recently-evolved regulatory elements associate with lineage-specific increases in gene expression. For each reference species (human, mouse, dog and cow), we identified genes associated with a larger number of recently-evolved elements (promoters, left, purple; or enhancers, right, orange) in this species than in any other across the ten species where recently-evolved elements could be reliably identified. We

then evaluated where this species (human, mouse, dog or cow) ranked in expression across all fifteen species used for gene expression comparisons (rank 1: species with lowest normalized expression level; rank 15: species with highest normalized expression level). We observed that whichever species had the largest number of recently-evolved elements typically exhibited higher expression than expected (median rank > 7), compared to other species. For other genes, each species displayed the expected distribution of ranks (median rank = 7; i.e. genes were equally likely to be more or less expressed in this species compared to others).

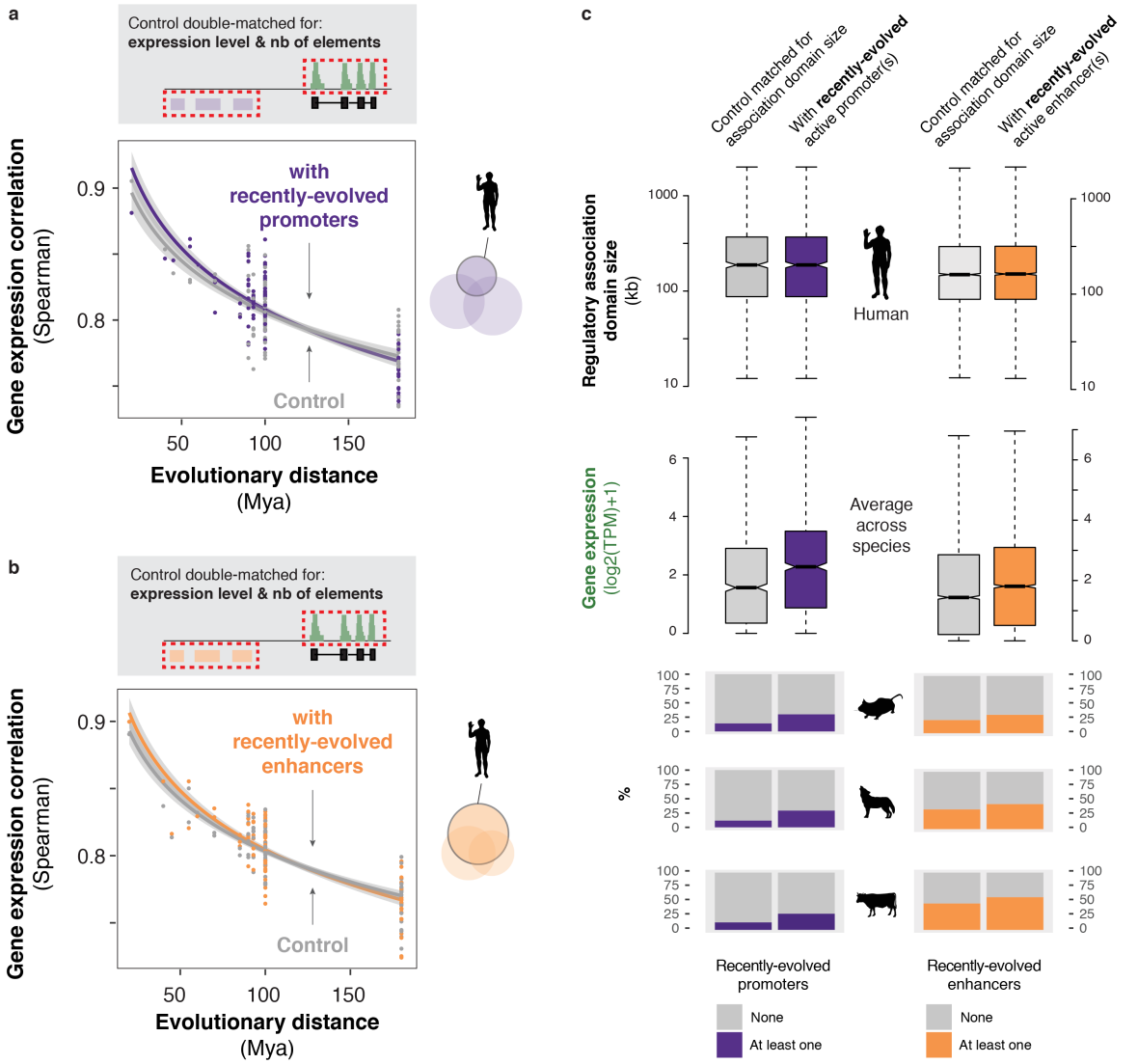


Figure S10. Supplemental analyses on recently-evolved regulatory activity in a single species (related to Figure 5)

(a-b) Expression correlations with increasing evolutionary distance for genes associated to human recent promoters (purple, **a**) or enhancers (orange, **b**). Acquisition of recently-evolved regulatory activity in a single species is not associated with increased gene expression divergence overall, compared to controls matched for gene expression and total number of associated elements.

(c) Top boxplot. Genes associated with recently-evolved promoters (purple, left) or enhancers (orange, right) in human were pair-matched with genes with a regulatory association domain of the same size.

Bottom boxplot. Controlling for the size of regulatory association domains, average gene expression across species remained higher for genes associated with recently-evolved regulatory activity.

Bottom barplots. Controlling for the size of the regulatory association domains, genes associated with recently-evolved promoters (left) or enhancers (right) in human were significantly more likely to be associated with similar recently-evolved elements in other species (mouse, dog and cow shown).

Table S1. Species and tissue samples used in this study (related to Methods, Figure 1 and Figure S1)

Species common name[a]	Species scientific name	Age of sexual maturity / lifespan	Provider	Provider class	Number of replicates	Sex	Age	Age group	Same samples for ChIP-seq?
Human **Hsap**	*Homo sapiens*	12-15 years / 80 years	Addenbrookes Hospital (UK)	Hospital	5	All M	unknown	adult	Yes, for all four ChIP-seq replicates
Macaque **Mmul**	*Macaca mulatta*	4 years / 20 years	Medical Research Council (UK)	Research colony	4	M, M, M, F	unknown, 18, 5, 11 (years)	adult, juvenile	Yes
Vervet **Csab**	*Chlorocebus aethiops sabaeus*	2-5 years / 11-13 years	Vervet Research Colony (US)	Research colony	3	All F	8, 9, 9 (years)	adult	Yes
Marmoset **Cjac**	*Callithrix jacchus*	1.5 years / 12 years	Harlan ltd (UK)	Research colony	4	All M	unknown	adult	Yes, for all three ChIP-seq replicates
Slow loris **Ncou**	*Nycticebus coucang*	1.5 years / 20 years	Duke Lemur Centre	Research colony	1	M	3 years	adult	ChIP-seq data unavailable
Greater galago **Ogar**	*Otolemur Garnettii*	1.5 years / 15 years	Duke Lemur Centre	Research colony	1	M	13 years	adult	Yes (E-MTAB-3933)
Mouse lemur **Mmur**	*Microcebus murinus*	0.5-1.5 years / 15 years	Duke Lemur Centre	Research colony	1	M	8 years	adult	Yes (E-MTAB-3933)
Ring-tailed lemur **Lcat**	*Lemur catta*	2.5-3 years / 15 years	Duke Lemur Centre	Research colony	1	M	14 years	adult	ChIP-seq data unavailable
Aye-aye **Dmad**	*Daubentonia madagascariensis*	2.5 years / 25 years	Duke Lemur Centre	Research colony	1	M	11 years	adult	ChIP-seq data unavailable
Mouse **Mmus**	*Mus musculus domesticus*	6-8 weeks / 1-3 years	Charles river (UK)	Research colony	4	All M	10 weeks	adult	Yes
Rat **Rnor**	*Rattus norvegicus*	5 weeks / 1-3 years	Harlan ltd (UK)	Research colony	3	All M	10 weeks	adult	Only for one of three ChIP-seq replicates
Guinea pig **Cpor**	*Cavia porcellus*	3-5 weeks / 4-8 years	Harlan ltd (UK)	Research colony	3	All M	10 weeks	adult	Yes

Species common name[a]	Species scientific name	Age of sexual maturity / lifespan	Provider	Provider class	Number of replicates	Sex	Age	Age group	Same samples for ChIP-seq?
Naked mole rat **Hgla**	*Heterocephalus glaber*	8-12 months / 30 years	UIC (US)	Research colony	4	All M	1 year	adult	Only for two of three ChIP-seq replicates
Rabbit **Ocun**	*Oryctolagus cuniculus*	5-6 months / 8-12 years	Harlan ltd (UK)	Research colony	3	All M	7,12,12 (months)	juvenile, adults	Yes
Tree shrew **Tbel**	*Tupaia belangeri*	4-5 months / 9-12 years	Cardiff University (UK)	Research colony	4	M, M, F	16, 3, 3, 6 (months)	adult, juveniles	Yes, for all three ChIP-seq replicates
Cow **Btau**	*Bos taurus*	8-12 months / 15 years	B&K ltd (UK)	Commercial	4	All M	2, 1.5, 2, 2 (years)	adult	Yes
Dolphin **Ddel** (short-beaked common dolphin and white-beaked dolphin)	*Delphinus delphis*	12-15 years / 22 years	UK Cetacean Strandings Investigation Programme, Zoological Society of London (UK)	Specialised research programme	3	M, F, F	unknown	adult	Only one ChIP-seq replicate
	Lagenorhynchus albirostris	uknown / 25 years			2	F, M	unknown	adult	Only one ChIP-seq replicate
Sei whale **Bbor**	*Balaenoptera borealis*	8-10 years / 50-70 years			1	F	unknown	juvenile	Yes
Sowerby's beaked whale **Mbid**	*Mesoplodon bidens*	7 years / unknown			2	Both F	unknown	juvenile	Yes
Pig **Sscr**	*Sus scrofa*	6 months / 10-15 years	Harlan ltd (UK)	Research colony	3	All M	2 years	adult	Yes
Dog **Cfam**	*Canis familiaris*	1 year / 12-15 years	Harlan ltd (UK)	Research colony	3	All M	2.5, 1, 1 (years)	adult, juveniles	Yes
Cat **Fcat**	*Felis catus*	5-10 months / 15 years	Isoquimen ltd (Spain)	Research colony	2	Both F	1.5 years	adult	Yes

Species common name[a]	Species scientific name	Age of sexual maturity / lifespan	Provider	Provider class	Number of replicates	Sex	Age	Age group	Same samples for ChIP-seq?
Ferret **Mfur**	*Mustela putorius furo*	6 months / 8 years	B&K ltd (UK)	Research colony	3	All M	8, 6, 6 (months)	adult, juveniles	Yes
Opossum **Mdom**	*Monodelphis domestica*	4-5 months / 4-8 years	MRC National Institute for Medical Research (UK)	Research colony	3	All M	6 months	juveniles	Yes
Tasmanian Devil **Shar**	*Sarcophilus harrisii*	2 years / 5-6 years	Copenhagen Zoo (Denmark)	Zoo	2	F, M	8, 7.5 (years)	adult	Yes

[a] Species abbreviations used in the manuscript are given in bold

Table S2. Coefficients of determination (Pearson R^2) for the linear models fitting gene expression conservation as a function of evolutionary time (related to Figures 2, 3, 4 and 5)

Figure	Test gene set	R^2 - Test	R^2 - Control
Figure 2	Core liver genes	0.54	0.65
	Housekeeping genes	0.61	0.64
Figure 3	Promoters	0.59	0.64
	Enhancers	0.60	0.58
Figure 4	Placental-conserved promoters	0.57	0.60
	Placental-conserved enhancers	0.41	0.36
Figure 5	Recent promoters	0.69	0.65
	Recent enhancers	0.48	0.62
	Recent promoters (2)	0.69	0.65
	Recent enhancers (2)	0.33	0.21

SUPPLEMENTAL REFERENCES

1 Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-348, doi:10.1038/nature10532 (2011).

2 Pervouchine, D. D. *et al.* Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nature communications* **6**, 5903, doi:10.1038/ncomms6903 (2015).

3 Chan, E. T. *et al.* Conservation of core gene expression in vertebrate tissues. *J Biol* **8**, 33, doi:jbiol130 [pii]
10.1186/jbiol130 (2009).

4 Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635-640, doi:10.1038/nature12943 (2014).

5 Fushan, A. A. *et al.* Gene expression defines natural changes in mammalian lifespan. *Aging cell* **14**, 352-365, doi:10.1111/acel.12283 (2015).

6 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

7 Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566, doi:10.1016/j.cell.2015.01.006 (2015).

8 Kheradpour, P. *et al.* Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome research* **23**, 800-811, doi:10.1101/gr.144899.112 (2013).

9 Ballester, B. *et al.* Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *eLife* **3**, e02626, doi:10.7554/eLife.02626 (2014).

10 Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521-1531, doi:10.1016/j.cell.2013.11.033 (2013).

11 Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome research* **15**, 901-913, doi:10.1101/gr.3577405 (2005).

12 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).