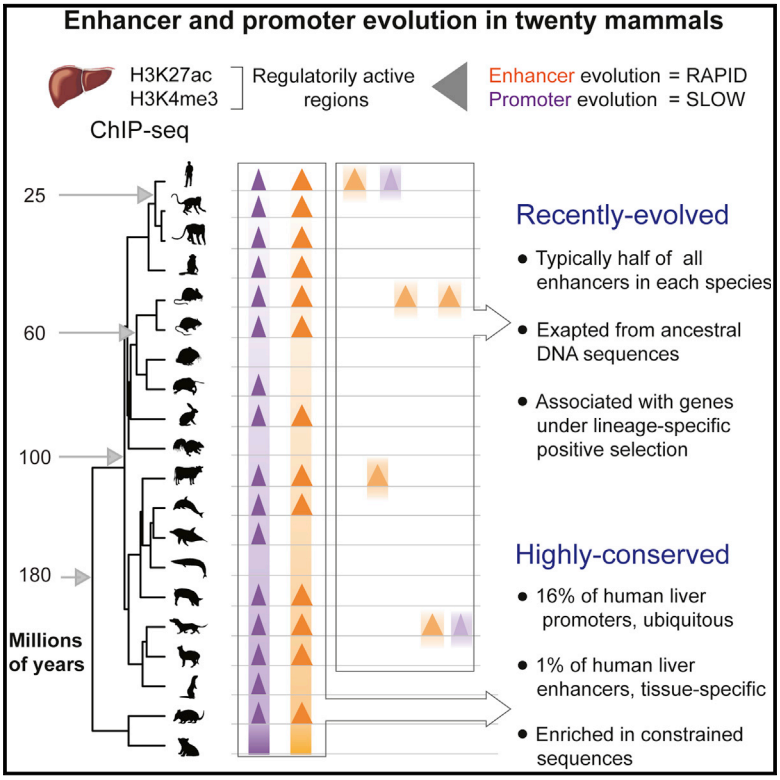
Article

Cell

Enhancer Evolution across 20 Mammalian Species

Graphical Abstract



Highlights

- Rapid enhancer and slow promoter evolution across genomes of 20 mammalian species
- Enhancers are rarely conserved across these mammals
- Recently evolved enhancers dominate mammalian regulatory landscapes
- Unbiased mapping links candidate enhancers with lineage-specific positive selection

Authors

Diego Villar, Camille Berthelot, ..., Paul Flicek, Duncan T. Odom

Correspondence

flicek@ebi.ac.uk (P.F.), duncan.odom@cruk.cam.ac.uk (D.T.O.)

In Brief

Comparative functional genomic analysis in 20 mammalian species reveals distinct features for the evolution of enhancers, in comparison to those of promoters, across 180 million years.

Villar et al., 2015, Cell *160*, 554–566
January 29, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.cell.2015.01.006

CellPress

Enhancer Evolution across 20 Mammalian Species

Diego Villar,[1,11] Camille Berthelot,[2,11] Sarah Aldridge,[1,12] Tim F. Rayner,[1] Margus Lukk,[1] Miguel Pignatelli,[2] Thomas J. Park,[3] Robert Deaville,[4] Jonathan T. Erichsen,[5] Anna J. Jasinska,[6] James M.A. Turner,[7] Mads F. Bertelsen,[8] Elizabeth P. Murchison,[9] Paul Flicek,[2,10,*] and Duncan T. Odom[1,10,*]

[1]University of Cambridge, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, CB2 0RE, UK
[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[3]Department of Biological Sciences, University of Illinois at Chicago (UIC), 845 West Taylor Street, Chicago, IL 60607, USA
[4]UK Cetacean Strandings Investigation Programme (CSIP) and Institute of Zoology, Zoological Society of London, Outer Circle, Regent's Park, London NW1 4RY, UK
[5]School of Optometry and Vision Sciences, Cardiff University, Maindy Road, Cardiff CF24 4HQ, UK
[6]UCLA Center for Neurobehavioral Genetics, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA
[7]Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK
[8]Center for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark
[9]Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK
[10]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[11]Co-first author
[12]Present address: AstraZeneca, Personalised Healthcare and Biomarkers, Molecular Diagnostics, Darwin Building 310, Milton Rd, Cambridge CB4 0WG, UK
*Correspondence: flicek@ebi.ac.uk (P.F.), duncan.odom@cruk.cam.ac.uk (D.T.O.)
http://dx.doi.org/10.1016/j.cell.2015.01.006
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

SUMMARY

The mammalian radiation has corresponded with rapid changes in noncoding regions of the genome, but we lack a comprehensive understanding of regulatory evolution in mammals. Here, we track the evolution of promoters and enhancers active in liver across 20 mammalian species from six diverse orders by profiling genomic enrichment of H3K27 acetylation and H3K4 trimethylation. We report that rapid evolution of enhancers is a universal feature of mammalian genomes. Most of the recently evolved enhancers arise from ancestral DNA exaptation, rather than lineage-specific expansions of repeat elements. In contrast, almost all liver promoters are partially or fully conserved across these species. Our data further reveal that recently evolved enhancers can be associated with genes under positive selection, demonstrating the power of this approach for annotating regulatory adaptations in genomic sequences. These results provide important insight into the functional genetics underpinning mammalian regulatory evolution.

INTRODUCTION

Most mammalian genes are controlled by collections of enhancer regions, often located tens to hundreds of kilobases away from transcription start sites. Recent studies comparing key selected mammals (Cotney et al., 2013; Xiao et al., 2012) have indicated that enhancers may change rapidly during evolution (Degner et al., 2012; Shibata et al., 2012), particularly when compared with evolutionarily stable gene expression patterns (Brawand et al., 2011; Chan et al., 2009; Merkin et al., 2012). Given that most phenotypic differences are hypothesized to largely result from regulatory differences between mammals, it is of profound importance to understand the mechanisms driving enhancer evolution (Villar et al., 2014; Wray, 2007).

Both conserved and recently evolved enhancer sequences have been shown to have important phenotypic consequences. Highly conserved enhancer sequences can regulate fundamental processes, such as embryonic development, and this property has been used to screen for functional regulatory elements (Pennacchio et al., 2006). However, sequence-level changes in enhancer elements can also underlie evolutionary differences between species (Hare et al., 2008; Ludwig et al., 2005), as has now been demonstrated across many organisms (Arnold et al., 2014; Cotney et al., 2013; Degner et al., 2012; McLean et al., 2011; Shibata et al., 2012).

Approaches comparing vertebrate genome sequences, such as those employing 29 mammals, have revealed regulatory regions under sequence constraint (Lindblad-Toh et al., 2011). However, this approach is limited in resolving tissue-specific deployment or regulatory activity directed by small sequence changes, particularly as may be predicted for rapidly evolving enhancer regions (however, see Pollard et al., 2006; Prabhakar et al., 2006). Comparative analysis of mammalian genomes can indicate protein sequence adaptations in particular species or lineages, and infer which coding regions are under positive selection. In contrast, complementary experimental efforts are currently lacking to functionally annotate the many recently sequenced mammalian genomes.

Experimental tools can now empirically identify regulatorily active DNA across entire mammalian genomes. Enhancers can

be identified by mapping regions enriched for acetylated lysine 27 on histone H3 (H3K27ac) via chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Creyghton et al., 2010). Similarly, active gene promoters can be identified as containing both H3K27ac and trimethylated lysine 4 of histone H3 (H3K4me3), which marks sites of transcription initiation (Cain et al., 2011; Santos-Rosa et al., 2002). The usefulness of this approach to map regulatory activity genome-wide has been recently underscored by analysis of H3K27ac dynamics across organ development in mouse (Nord et al., 2013). This study found that most H3K27ac developmental variation occurs distally to transcription start sites and within predicted enhancer elements, most of which could be validated experimentally.

Over 20 sequenced mammalian genomes have been integrated into inter-species alignments within Ensembl (Flicek et al., 2014). Exploiting this computational infrastructure (and related resources in *Drosophila*; Kim et al., 2009), recent studies have dissected how transcription factor (TF) binding has evolved (He et al., 2011; Paris et al., 2013; Schmidt et al., 2010; Stefflova et al., 2013). In addition, enhancer and promoter evolution have been investigated using sets of mammals, where H3K27ac levels have been characterized across tissues and developmental states as a proxy for enhancer function and developmental or tissue-specific gene expression (Cotney et al., 2013; Nord et al., 2013; Xiao et al., 2012).

Here, we report the results of empirically mapping promoter and enhancer evolution across 20 mammals chosen to span the breadth and depth of the class Mammalia, including previously uncharacterized species such as cetaceans and naked mole rat. Our analyses have revealed the tempo and mechanisms underlying enhancer evolution across over 180 million years of mammalian radiation.

RESULTS

Profiling Promoter and Enhancer Regulatory Evolution in Mammalian Liver

We mapped the active promoter and enhancer elements in liver as a representative adult somatic tissue from 20 species of mammals (Figure 1). Study species were selected using three criteria: (1) to capture a substantial fraction of the mammalian phylogenetic tree, (2) to profile the major placental orders in a combination of intra- (6–40 Ma) and inter-lineage (100–180 Ma) evolutionary distances, and (3) to extend our understanding of regulatory evolution to previously uncharacterized mammals whose phenotypes are highly divergent, such as cetaceans, naked mole rat, and Tasmanian devil. Liver from almost all study species was profiled in biological replicates from two or more individuals, except for Sei Whale *(Balaenoptera borealis)*, where only one individual's tissue was available; and for dolphin, for which we combined data from two closely related dolphin species (*Delphinus delphis* and *Lagenorhynchus albirostris*) where a single individual from each species was profiled (Tables S1 and S2, Experimental Procedures).

We quantified using ChIP-seq the genome-wide occurrence of two key histone marks widely used to profile promoters and enhancers: H3K4me3 and H3K27ac (Figure 1) (Creyghton et al., 2010; Santos-Rosa et al., 2002). We identified regions enriched for these histone marks within each mammalian liver genome using only biologically reproducible peaks present in two or more replicates (Figure S1, Experimental Procedures).

A total of 30–45,000 regions per species were enriched in liver, and these separated into H3K27ac, H3K4me3&H3K27ac, and H3K4me3-marked elements (Figures 1C and S1). Our analyses were robust to variability in the genome assembly quality and sample preparation (Experimental Procedures and Figure S2). We confirmed that H3K4me3 often co-occupied the genome with H3K27ac (Heintzman et al., 2009; Zhu et al., 2013), and that most H3K4me3-positive regions occur at transcriptional start sites (Cain et al., 2011; Santos-Rosa et al., 2002), regardless of their H3K27ac enrichment (see Experimental Procedures). In contrast, regions enriched for H3K27ac often were not enriched for H3K4me3, and these often located far from transcriptional start sites (Figure S2).

The regions we identify as enhancers strongly enrich for regulatory activity in liver, consistent with numerous prior studies (Cotney et al., 2013; Creyghton et al., 2010; Nord et al., 2013; Zhu et al., 2013). For over 400 of our human liver enhancers (typically 2 kb in length), the transgenic activities of overlapping 145 bp segments were assayed in liver cancer cells (Kheradpour et al., 2013) (Figure S2). Although each human liver enhancer was on average represented by only a single small sequence element, capturing less than 10% of the enhancer length, over 65% showed activity in transgenic assays in a cancer cell line. Furthermore, over 90% of the enhancers not active in transgenic assays were nevertheless bound in human liver by at least one liver-specific TF (Ballester et al., 2014). In sum, this analysis suggests a sizable majority of our empirically determined enhancers are regulatorily active.

Our data newly demonstrates that the known interplay of H3K4me3 and H3K27ac creates a genomic regulatory landscape that is a uniform feature across mammals (and likely across eumetazoans; Schwaiger et al., 2014). In adult liver, a typical mammalian genome contains on average 12,500 H3K4me3 locations (representing active promoter elements) and 22,500 H3K27ac-enriched regions (representing active enhancers).

Enhancer Evolution Is Appreciably More Rapid Than Proximal Promoter Evolution

We used our genome-wide mapping data in livers from 20 mammals to obtain an empirical and quantitative understanding of evolutionary stability of promoters and enhancers (Figure 2 and Figure S3).

Most non-coding regions in the human genome cannot be mapped across 20 mammals, in large part because the genome structure and regulatory content of complex eukaryotes evolve rapidly (Lynch et al., 2011). We defined the maximum detectable conservation of activity as the number of species in which the DNA could be aligned (Figure 2A). For example, if enhancer activity is highly conserved, then this activity would be detected in all species where the underlying DNA was alignable. In contrast, low conservation would be characterized by the underlying DNA remaining alignable across many species, but without sharing of enhancer activity. Such low conservation could be a

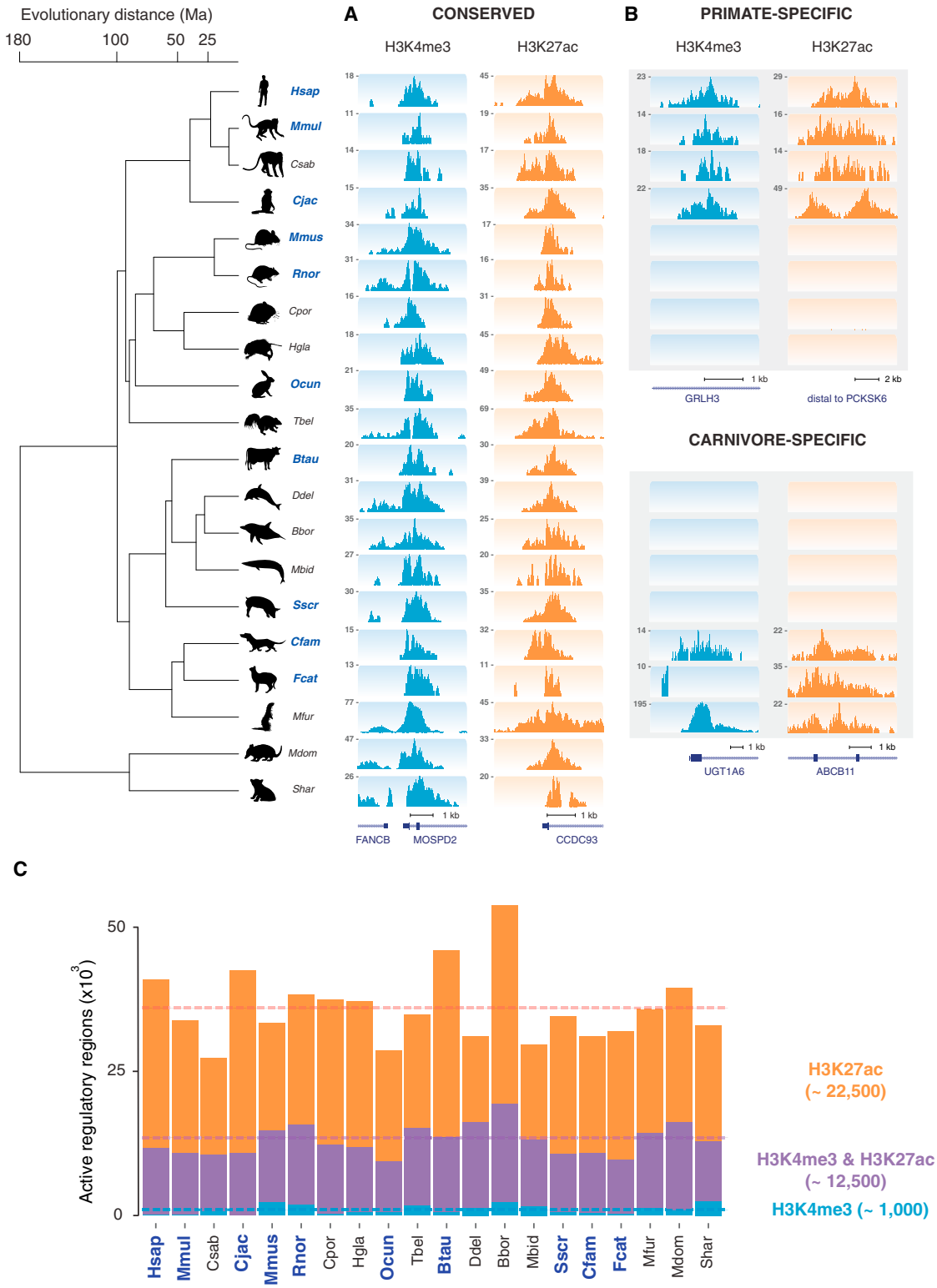


Figure 1. In Vivo Regulatory Activity Assessed in Livers from 20 Mammals

(A and B) Phylogenetic relationships and species divergences are represented by an evolutionary tree, which includes 18 placental species (in four orders) and 2 marsupial species (in two orders). In liver isolated from each species, enhancer activity was globally mapped by identifying genomic regions enriched for acetylation of H3K27 (H3K27ac), and transcription initiation was mapped by identifying genomic regions enriched for tri-methylation of H3K4 (H3K4me3). Shown

(legend continued on next page)

signature of rapid functional evolution or, alternatively, functional neutrality.

Collectively, the DNA sequences used as promoters and the DNA sequences used as enhancers in liver show only slight differences in their alignability across the study species (Figure 2B). This alignability shows a marked increase at approximately 11–13 species, reflecting the contribution to the multiple alignments of the ten highest-quality genomes (Experimental Procedures).

The conservation of active liver promoters tracked remarkably closely with the alignability of the underlying DNA, indicating evolutionarily stable promoter activity (Figure 2C, upper left triangle). In other words, the transcription initiation sites driving gene expression in liver are highly conserved.

We performed a similar analysis for enhancers. Our data reveal that rapid enhancer evolution, often involving exaptation of ancestral DNA, is active and widespread across all the mammalian clades in our study (Figure 2D, orange, and Figure S3), as has been reported in primates (Cotney et al., 2013). Furthermore, the ten highest-quality placental genome sequences contained thousands of cross-alignable regions where enhancer activity was shared in many, but not all, species. These regions are liver enhancers that were likely present in the common placental ancestor and have partially degraded along some lineages. In contrast to promoter sites, enhancer locations evolve rapidly, and comparatively few are deeply conserved (see below). Control analyses show that while promoter conservation may be under-estimated, this is not the case for enhancers (Figure S3).

We asked whether the conservation of liver promoters and enhancers is associated with underlying sequence features (e.g., TF binding sequences, %GC content, sequence constraint), experimental features (reproducibility, occupancy level/intensity, length), or some combination (Figure 3). The best predictor of conservation in promoter regions is the reproducibility and strength of enrichment of H3K4me3 and H3K27ac, with the length of the histone-modified domain and GC content as separate, modest contributors. Thus, experimental features are stronger indicators of the conservation of regulatory activity, and underlying sequence features contribute less to promoter stability. In contrast, the presence of TF binding sites can explain a modest fraction of the conservation of enhancer activity. Nevertheless, as with promoters, the enrichment reproducibility and intensity of signal is the primary predictor of conservation. Collectively, no combination of sequence- and experimental-based features could potentially explain more than a third of the variance in conservation of regulatory activity.

Overall, our data reveal that promoter activity in a representative somatic tissue is highly constrained across mammalian space. In contrast, enhancer evolution is rapid and widespread. Neither enhancer nor promoter activity conservation can be explained purely by underlying sequence elements.

Quantifying the Divergence Rates of Enhancers, Promoters, and TF Binding in a Cross-Section of Mammals

The divergence rate of sequence-specific transcription factor binding (Stefflova et al., 2013) and the extent of regulatory evolution (Cotney et al., 2013; Shibata et al., 2012; Xiao et al., 2012) has been estimated using matched experiments from the same tissues in subsets of typically three to five mammals within a single order. We took a similar approach to calculate how rapidly enhancers and promoters active in liver evolve across 20 mammals.

We first identified, by pairwise analysis of all 20 species, whether regions called as enhancers and promoters were present in the same location between two mammalian genomes (Experimental Procedures, Figure S4). Because this analysis does not use human as the primary reference genome, we could generate multiple independent estimates of how evolutionarily stable enhancers and promoters were for comparable divergence distances. Further, divergence rates could be estimated for evolutionary distances not available from a human-centric analysis. For instance, our data provided multiple comparisons of species separated by 40 to 100 Ma using mouse, cow, or dog as reference that could not be obtained using a human-centric approach (Figure 1).

Inter-species conservation of promoters and enhancers could be plausibly described as a function of time-of-divergence by fitting an exponential decay curve (Experimental Procedures). In liver, promoters diverged at a slower rate than did either enhancers or TF bound regions (Figure 4 and Figure S4). Interestingly, promoters' half-lives are comparable to protein-coding genes' half-lives, at over a billion years (Rands et al., 2014). The higher stability of promoters versus enhancers could be due in part to the intimate functional connection promoters have with the first exon of protein coding genes, which are highly stable features of vertebrate genomes (Lindblad-Toh et al., 2011). Our results are consistent with a model where the increased size and sequence heterogeneity of regions with promoter or enhancer activity could buffer evolutionary changes more robustly than can site-specific TF binding alone (Cotney et al., 2013; Shibata et al., 2012; Xiao et al., 2012).

Highly Conserved Regulatory Regions Are Largely Proximal Promoters

Our mapping of liver enhancer and promoter evolution using mammals spanning both intra-order (6–40 Ma) and inter-order (80–180 Ma) divergence times permits the dissection of conserved (and recently evolved, see below) regulatory regions.

We first quantified how many regions showed strong conservation of activity by defining regions as *highly conserved* if regulatory activity was present in (at a minimum) all ten of the highest-quality placental genomes (Figure 5A). A total of 2,151 genomic regions appeared highly conserved by these criteria, representing 5% of all human regions active in liver. The

are examples of regulatory regions active: (A) across all 20 species (MOSPD2 and CCDC93 loci), and (B) active only in primates (GRLH3 and PCKSK8, top) or active only in carnivores (UGT1A6 and ABCB11, bottom). For order-specific regulatory regions, data from some species are not shown for conciseness.
(C) In liver, a typical mammalian genome contains ~22,500 enhancers enriched for only H3K27ac; ~12,500 promoters enriched for both H3K27ac and H3K4me3 and ~1,000 containing only H3K4me3. Highest quality genomes incorporated into the EPO multiple alignment are labeled in blue (Experimental Procedures).
See also Figures S1 and S2 and Tables S1 and S2.

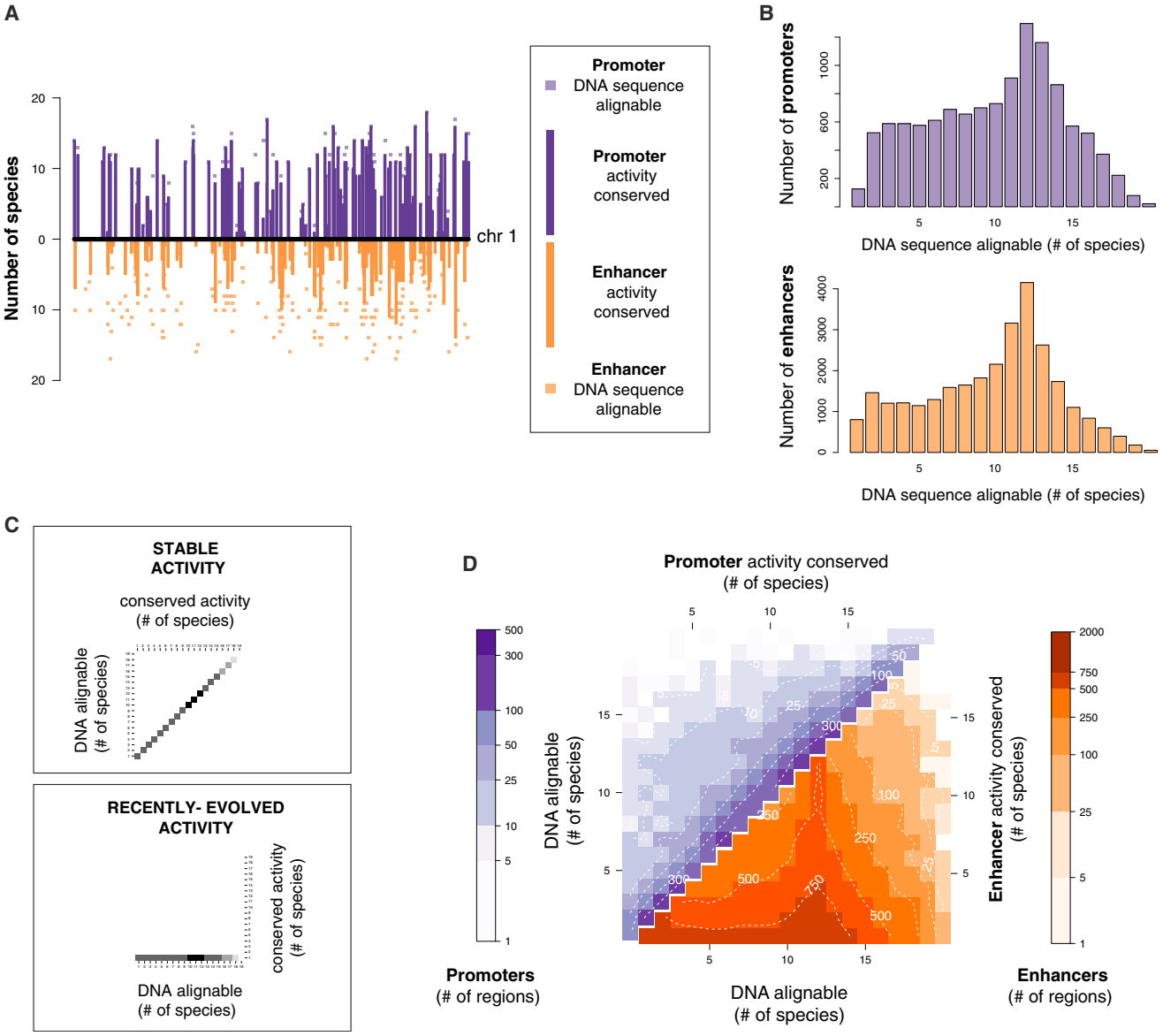


Figure 2. Enhancers Evolve Rapidly; Promoters Are Highly Conserved

(A) For a representative 10 MB region on human chromosome 1, the bar chart on the y axis represents the number of species in which enhancer and promoter elements were active (promoters: top, purple; enhancers: bottom, orange). Squares indicate the number of species where the sequence underlying the active promoter or enhancer was alignable.

(B) The DNA sequences underlying proximal promoters and the DNA sequences underlying enhancers can be aligned to similar numbers of species, suggesting that differences in apparent conservation of activity are not due to differences in alignability.

(C) Schematic diagram showing how the conservation of regulatory activity versus DNA alignability across 20 species of mammals can reveal (top) where DNA function and DNA sequence orthology closely correspond, indicating ancestral activity, and (*bottom*) where pre-existing DNA sequences have been exapted within specific lineages or species, indicating recently evolved activity.

(D) Our data revealed that if the DNA underlying a human-identified proximal promoter region (purple) can be aligned with an orthologous sequence in another species, then promoter activity is very often present as well (heatmap enrichment concentrated on the diagonal of the plot). In contrast, most enhancer regions (orange) are rapidly evolving within older DNA sequences, reflected in increased heatmap enrichment toward the lower x axis. Color scales and dashed contour lines indicate absolute numbers of active promoter or enhancer regions (logarithmic scale).

See also Figure S3.

existence of over 2,000 highly conserved regions is greater than expected by chance (p value $< 1 \times 10^{-4}$, random permutation test, Experimental Procedures).

Highly conserved regions were classified as promoters or enhancers based on their consensus histone mark enrichment across all 20 mammals (Experimental Procedures). Of these

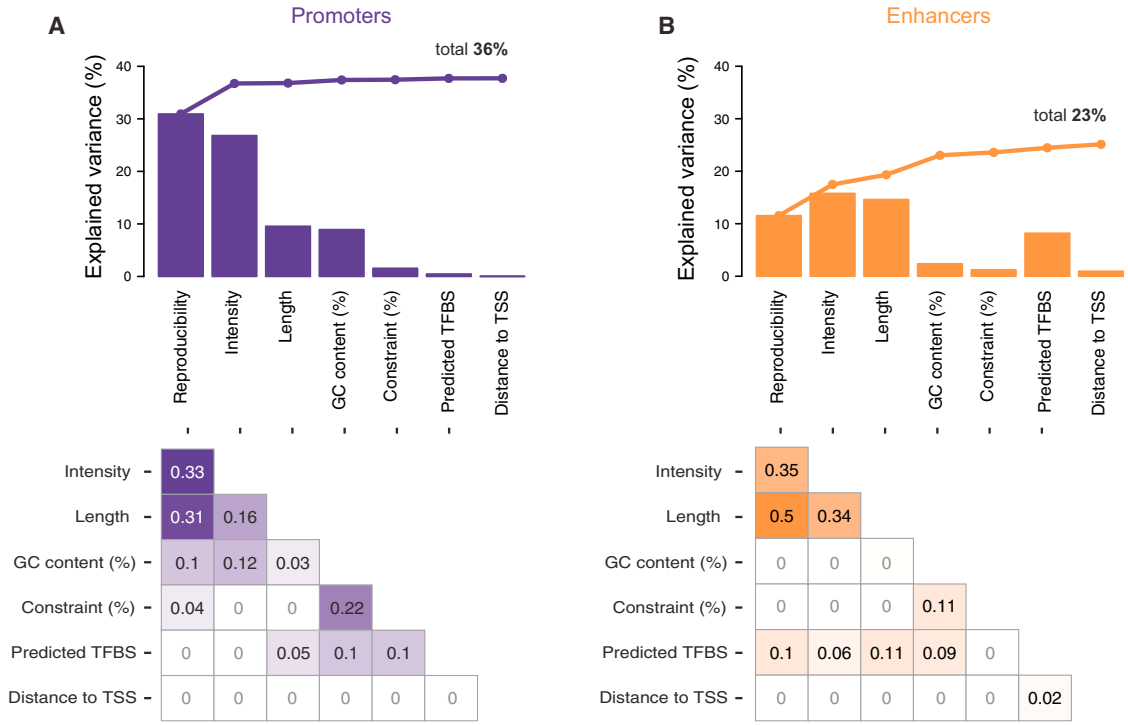


Figure 3. Features Contributing to Conservation of Promoter and Enhancer Activity Identified in Human Liver

(A) For all human proximal promoters active in liver, the depth of conservation was correlated with experimental features (reproducibility, peak intensity, peak length, distance to nearest transcription start site) as well as underlying genomic features (GC content, sequence constraint, TF binding sites). Each feature in isolation explained a significant fraction of the variance in conservation of promoter activity (e.g., peak length explained 10%). The fraction explained by the features in combination, when added left to right using multiple regression analysis, are plotted as a line above, in sum totaling 36%. The increases in explained variance with the addition of each feature are attenuated due to strong inter-correlation of features, quantified in the bottom panel as R^2 values between features (Experimental Procedures).

(B) The same analysis was performed for human liver enhancers, where experimental and genomic features together explained a more modest fraction (23%) of the conservation of enhancer activity in other species.

2,151 highly conserved regulatory regions, 1,871 elements (87%) were enriched for both H3K27ac and H3K4me3, consistent with acting as promoters (Santos-Rosa et al., 2002).The vast majority of highly conserved promoters occupied the transcription start sites of genes (Figure 5B). On the other hand, a subset of 279 regions showed enrichment only for H3K27ac occupancy, consistent with acting as enhancers (Creyghton et al., 2010). Most highly conserved enhancers were tens to hundreds of kilobases away from the nearest gene (Figure 5B). The single region uniformly enriched across placentals for only H3K4me3 is not shown.

In human liver, there are 11,838 promoter regions enriched for both H3K27ac and H3K4me3, and 28,963 enhancer regions containing only H3K27ac. Although nearly three times as common as promoters, the activity of only 1% of these enhancers is highly conserved. In contrast, the activity of 16% of promoters is highly conserved (Figure 5A).

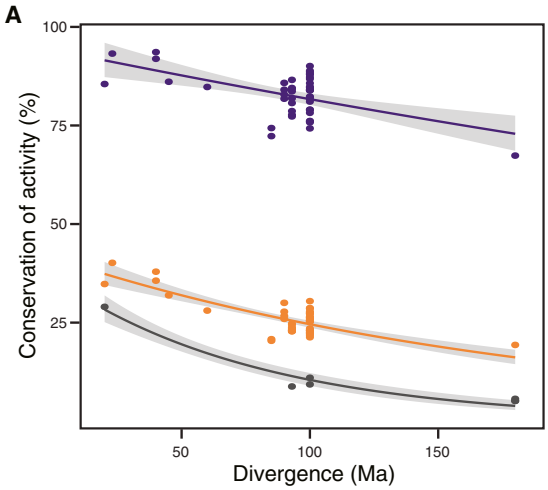
Three independent lines of evidence support the functionality of the sequences we identify as highly conserved regulatory regions in liver. First, all show enhanced sequence constraint (Figure 5C). Second, genes near highly conserved enhancers are strongly enriched for liver-specific functions, and genes near conserved proximal promoters are enriched for housekeeping functions (Figure S5, Tables S3 and S6) (Forrest et al., 2014). Third, highly conserved enhancers are enriched for TF binding motifs for liver-specific regulators such as CEBPA and PBX1, whereas highly conserved proximal promoters appear dominated by transcriptional initiation regulatory sequences (Figure S5, Table S7).

In sum, in adult mammals comparatively few enhancers are evolutionarily stable. In contrast, a substantial fraction of the proximal promoters found in human liver appear to be highly conserved across mammals.

Recently Evolved Regulatory Activity Is Pervasive in Mammals

Even for proximal promoters, the number of highly conserved regulatory elements active in liver is a small fraction of the total number experimentally identified in any single species (Figure 5 and Table S4). We sought to identify and analyze the molecular features of more recently evolved regulatory regions.

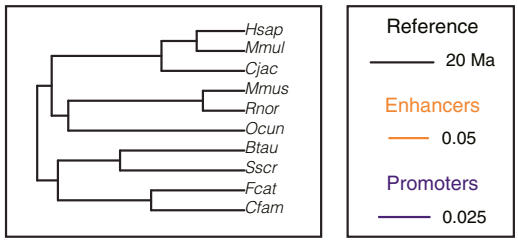
From each placental order, we selected a representative species (human, mouse, cow, dog) and then identified a set of newly evolved or, more formally, apomorphic active promoters and enhancers in liver (Figure 6 and Figure S7). For each of these four species, we started with all active regions and then removed

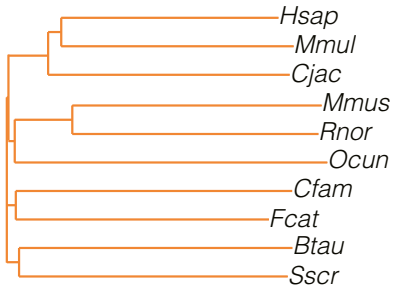


B

	Half life (Ma)	**Mean lifetime** (Ma)
Active promoters	939	1355
Active enhancers	296	427
CEBPA	144	207

C





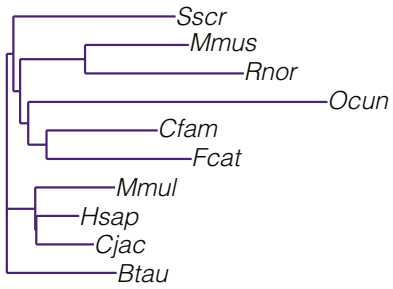


Figure 4. Empirically Determined Rates of Promoter, Enhancer, and TF Binding Divergence in Liver across 180 Million Years of Mammalian Evolution

(A) For promoters (purple), enhancers (orange), and TF binding sites (CEBPA, black), the fraction of ChIP-seq peaks present at the orthologous location between pairs of mammals are shown as a function of evolutionary distance. Solid lines represent an exponential decay fit, surrounded by gray shading of a 95% confidence interval (Experimental Procedures). For liver promoters and enhancers, we used data from the ten highest-quality placental genomes, while CEBPA data have been previously reported (Schmidt et al., 2010).

(B) Comparative half-lives and mean-lifetimes (in million years) for active promoters, enhancers and CEBPA transcription factor binding locations, as calculated from the exponential decay fits in (A).

(C) Neighbor-joining phylogenetic trees based on pairwise conservation levels of enhancer and promoter activity, as measured in (A). Enhancer evolution (orange) recapitulates the known relationships among the studied mammals (black). The low divergence of promoter activity is insufficient to resolve the phylogenetic groups (purple).

See also Figure S4.

those that showed any activity within alignable regions in any other study species (see Experimental Procedures). We found that a typical mammalian liver deploys between 1,000 to 2,000 promoters and 10,000 enhancers not found in any other study species; we henceforth refer to these enhancers and promoters as *recently evolved*.

These numbers are comparable to the extent of enhancer gains previously reported in inter-primate comparisons (Cotney et al., 2013; Shibata et al., 2012) and the extent of promoter evolution estimated from mouse-human comparisons (Forrest et al., 2014; Frith et al., 2006). Especially for enhancers, recently evolved regions are 10–20 times more abundant than those conserved across placentals or shared across multiple species in a particular lineage (Table S4). Both highly conserved and recently evolved regulatory regions active in liver are associated with increased expression of neighboring genes (Figure S6).

Exaptation Drives Recently Evolved Enhancer, but Not Promoter, Activity

Using these tens of thousands of apomorphic regulatory regions, we tested whether functional exaptation of ancestral DNA, recently reported for human-specific enhancers active in embryonic limb (Cotney et al., 2013), is a prevalent mechanism in mammalian genome evolution.

We first asked whether recently evolved proximal promoters are primarily found in ancestral DNA sequences older than 100 Ma (Figure 6A, Figure S7). To our surprise, we discovered that across four orders of mammals, the recent evolution of promoters occurred within evolutionarily younger DNA segments (i.e., not shared with other study species) about three to four times as often as occurred by exaptation of ancestral DNA. For instance in mouse, 1,400 recently evolved promoters occurred in DNA sequences present only in this species (i.e., not shared even with rat); in contrast, only 260 recently evolved promoters were found in ancestral DNA.

Within the ancestral DNA commandeered into new promoters, and regardless of species interrogated, diverse ERV repeat elements are over-represented, consistent with previous reports that ERVs are pre-primed to transcriptional initiation (Fort et al., 2014).

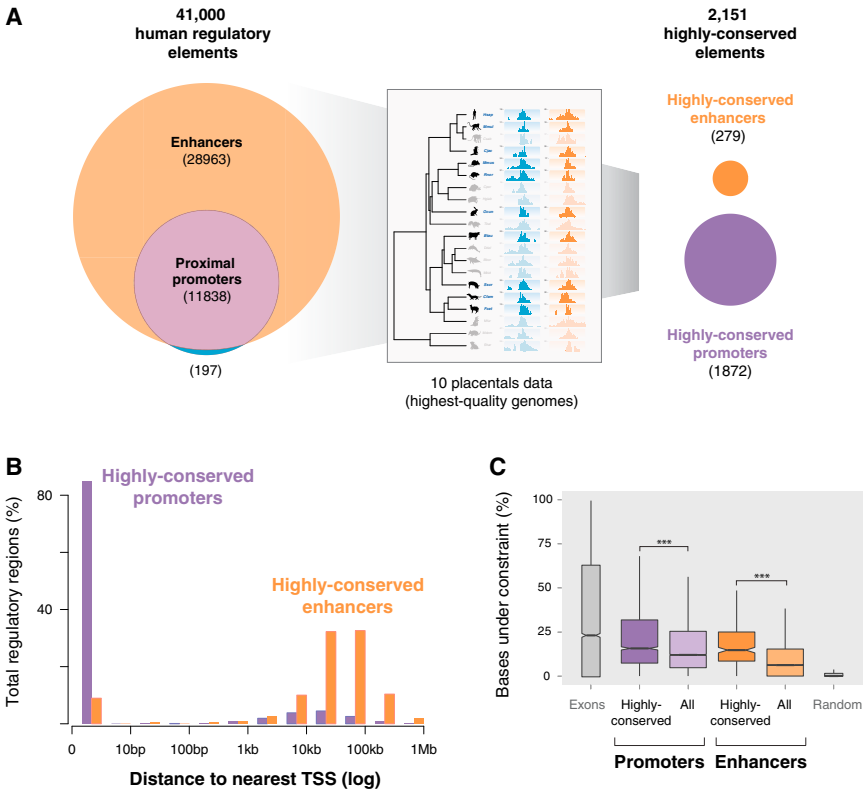


Figure 5. Most Highly Conserved Liver Regulatory Regions Are Proximal Promoters

(A) The ~41,000 regulatorily active regions in human liver are shown on the left panel (enhancers: orange; promoters: purple). The regulatory elements with conserved activity in the ten placental species with highest quality genomes (boxed inset) were determined by cross-species comparison (Experimental Procedures), identifying approximately 300 enhancers and 1,800 promoters (labeled as highly conserved, right panel).

(B) Almost all highly conserved promoter regions (purple) are located at transcription start sites as expected, whereas conserved enhancer regions (orange) are typically tens to hundreds of kilobases from the nearest gene.

(C) Regions of highly conserved enhancer and promoter activity show a corresponding, but modest, increase in selective constraint in their underlying DNA sequence. The distribution of the fraction of bases under constraint in each region within each category is shown as a boxplot, with human exons and randomly selected regions shown for comparison (Experimental Procedures).*** indicates p value $< 2 \times 10^{-16}$, Wilcoxon test.

See also Figures S5 and S6 and Tables S3, S6, and S7.

In contrast, the vast majority of enhancers in liver are recently evolved (Table S4)—as well as far more likely to exapt ancestral DNA (Figure 6B). Of the typically 10,000 recently evolved enhancers in a given species, 52%–77% contained sequences of ancestral DNA over 100 Ma old. The remaining recently evolved enhancers were found in younger DNA, and enriched for mobile repetitive element families, including LTRs in all lineages and lineage-specific SINEs and DNA transposons exclusive to primates, carnivores, or ungulates (Figure 6B).

In a typical mammalian species, the 1,000 to 2,000 recently evolved liver promoters occur predominantly in younger DNA typically less than 40 Ma old, whereas the 10,000 recently evolved enhancers are formed predominantly by exaptation of ancestral DNA. Only a minority of recently evolved enhancers and promoters appear driven by repeat element expansions (Figure 6, Figure S7). Across our study's 20 mammals, exaptation of ancestral DNA generates more of the recently evolved regulatory genome than do repeat-driven expansions.

Functional Annotation of Genes under Positive Selection

Comparing genome sequences can suggest which genes drive phenotypic adaptations by using inference of regions under positive selection and by analyzing amino acid substitution patterns in proteins (Nielsen et al., 2007). Both approaches primarily employ coding-sequence alignments and thus provide limited insight into regulatory adaptations. We therefore asked whether genes under positive selection are associated with apomorphic enhancers, perhaps evolving synergistically (Shibata et al., 2012).

We compared recently evolved enhancers and positively selected genes in two newly sequenced species: (1) naked mole rat, a cancer-resistant rodent (Kim et al., 2011); and (2) dolphin, a marine mammal metabolically adapted to an aquatic environment (Sun et al., 2013). In both species, we found that recently evolved enhancers are over-represented near positively selected genes (Experimental Procedures) (p values = 0.022 [naked mole rat] and 0.023 [dolphin], hypergeometric test. See Table S5).

Illustrative examples are shown in Figure 7. First, a recently evolved enhancer in naked mole rat is shown upstream of the *thymopoietin* gene (*TMPO*), identified previously as positively selected (Kim et al., 2011). The orthologous *TMPO* regions in human, mouse, cow, and dog show no enhancer activity, though a number of partially conserved enhancers are present nearby (Figure 7A). Second, the genomic region around the *TRIP12* gene, under positive selection in dolphin (Sun et al., 2013), contains a recently evolved dolphin enhancer not active in human, mouse, dog, and cow. Moreover, this regulatory element appears to be the main enhancer in this region (Figure 7B).

In sum, recently evolved active regions identified in this study, and in particular rapidly evolving enhancers, can functionally annotate lineage-specific adaptations.

DISCUSSION

We experimentally dissected the evolution of regulatory regions in mammalian liver by mapping the genome-wide landscape of

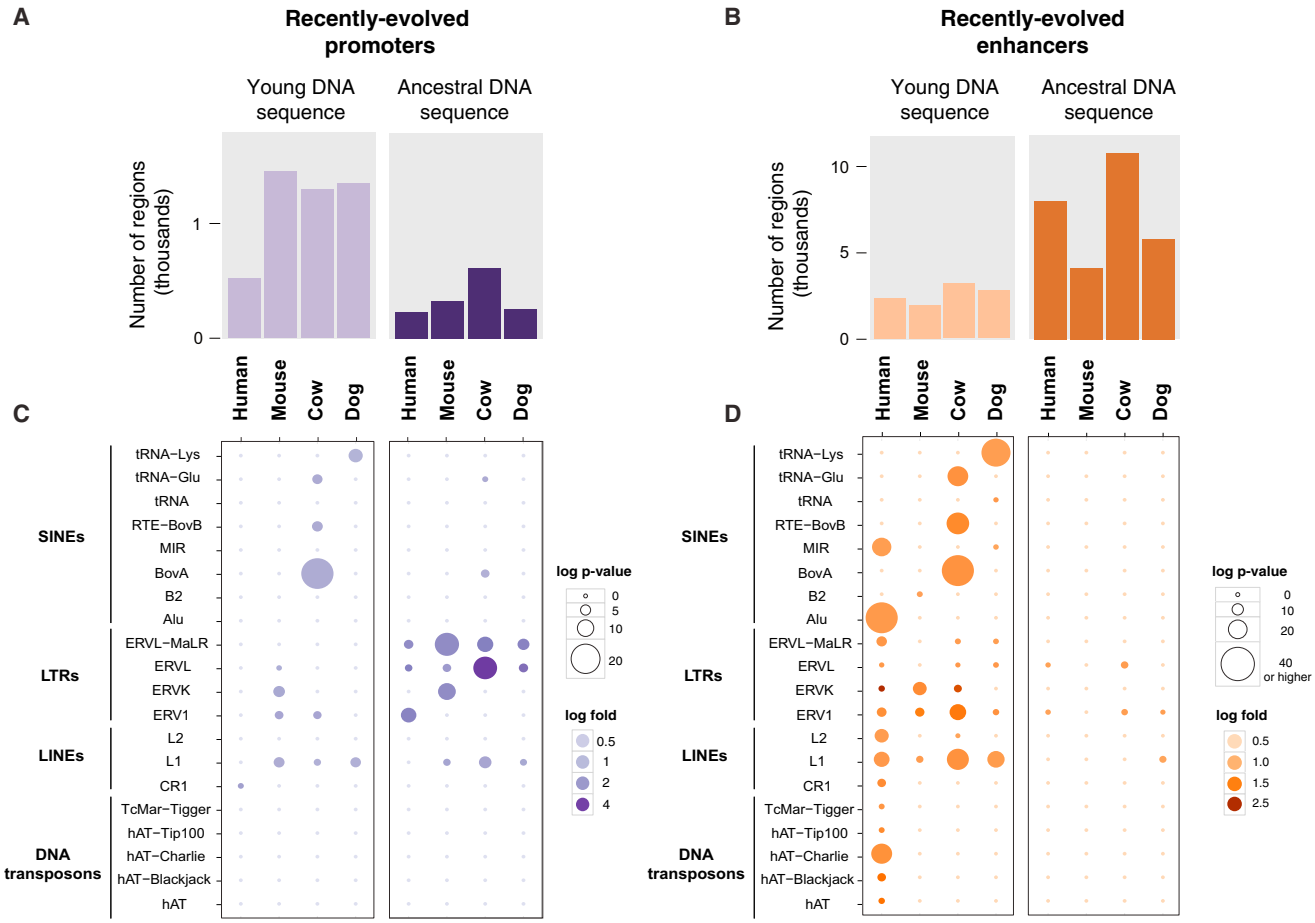


Figure 6. Recently Evolved Promoters Are Largely Derived from Young DNA, While Recently Evolved Enhancers Are Mostly Exapted from Ancestral DNA Sequences

Regions with recently evolved promoter and enhancer activity in liver were identified in a representative species for each placental order (primate:human, rodent:mouse, ungulate:cow, and carnivore:dog). These regions were categorised into those falling in (1) young DNA sequences (0–40 Ma) or (2) ancestral DNA sequences (>100 Ma).

(A) Typically three times as many recently evolved active promoters reside in young DNA as are found in ancestral DNA sequences present across placental mammals.

(B) Conversely, typically twice as many recently evolved enhancers are exapted from evolutionarily ancestral DNA as are found in young DNA.

(C and D) Repeat classes and families enriched in recently evolved promoters and enhancers were identified using a binomial test (see Experimental Procedures). Plots show enrichments for each repeat family (y axis) and each species (x axis). Circle sizes represent the statistical significance of enrichment, and color shades denote the fold change of the enrichment (both in logarithmic scale).

See also Figures S6 and S7 and Tables S3, S4, S6, and S7.

active promoters and enhancers from 20 diverse species. The evolutionary distances spanning four distinct orders within class *Mammalia* enabled rigorous analysis of the mechanisms underlying regulatory evolution. The combination of rapid enhancer and slower promoter evolution appears to be a fundamental property of the mammalian regulatory genome, shared by species separated by up to 180 million years. A sizable number of the 10,000–15,000 active promoters are functionally shared across most mammals, and are associated with ubiquitous cellular functions; highly conserved enhancers are much less common, and are found near liver-specific genes. Remarkably, almost half of 20,000–25,000 active enhancers in each species have rapidly evolved in a lineage- or species-specific manner. Our genome-wide mapping of enhancers in previously uncharacterized species has enabled us to identify regulatory regions near genes under positive selection that may help drive phenotypic adaptations.

A Global Overview of Enhancer and Promoter Evolution in Mammals

We used a powerful and unbiased strategy to confirm, extend, and explicitly quantify previous results showing higher conservation of active promoter regions compared to distal enhancers in selected representatives of mammals (Xiao et al., 2012) or within primates (Cotney et al., 2013).

Our study has a number of limitations. First, the relationship between different histone marks and the activity of enhancers is not perfectly understood. Most active enhancers are marked

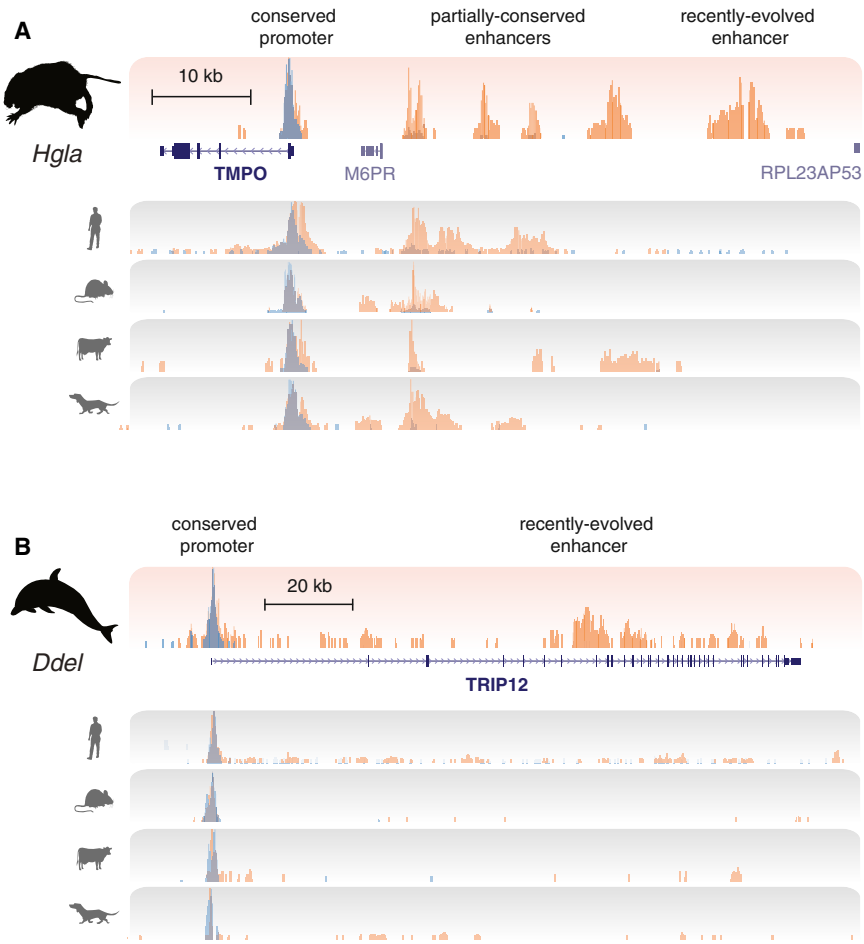


Figure 7. Recently Evolved Enhancers Associate with Genes under Positive Selection during Naked Mole Rat and Dolphin Evolution

(A) The liver enhancer and promoter landscape surrounding the *TMPO* locus, which is under positive selection in naked mole rat (Kim et al., 2011), is shown (upper track). The bottom four tracks display overlaid H3K4me3 (blue) and H3K27ac (orange) levels in the orthologous regions of human, mouse, dog, and cow. Shown (left to right) are a promoter present in all species, four enhancer regions shared in a subset of species, and a naked mole rat-specific enhancer whose recently evolved activity is not present in other study species.

(B) The enhancer and promoter landscape surrounding the *TRIP12* locus, which is under positive selection in dolphins (Sun et al., 2013), is shown. In this case, no mammals other than dolphin show liver enhancer activity near this gene; this enhancer is thus a good candidate to contain the regulatory regions associated with positive selection in dolphin.

See also Table S5.

by H3K27ac (Andersson et al., 2014; Creyghton et al., 2010; Zhu et al., 2013), and typically over two-thirds of regions enriched for H3K27ac show independent evidence in transgenesis assays for regulatory activity (Nord et al., 2013). Global mapping of H3K4me1 and p300 can also detect poised enhancer activity genome-wide, which can partly differ from that identified by H3K27ac (Heintzman et al., 2007; Krebs et al., 2011; Visel et al., 2009). Second, other approaches to map regulatory sequences, such as DNase-seq (Shibata et al., 2012) or ATAC-seq (Buenrostro et al., 2013), can reveal all regions of open chromatin genome-wide, but cannot distinguish promoters and enhancers. Third, our approach does not directly reveal which transcription factors control these regulatory regions, as would a more direct comparison (Kunarso et al., 2010; Paris et al., 2013; Schmidt et al., 2010), which in turn can only capture a modest subset of active regions. Fourth, our results generalize to other mammalian somatic tissues to the extent that adult liver is a representative tissue. However, other studies have suggested rapid enhancer evolution in mammals, using embryonic limb buds (Cotney et al., 2013), adipocytes (Mikkelsen et al., 2010), and embryonic stem cells (Xiao et al., 2012). These studies and others (Barbosa-Morais et al., 2012; Brawand et al., 2011) suggest that regulation in other somatic tissues evolves similarly, though embryonic tissues and their enhancers may be under stronger evolutionary constraint (Faure et al., 2012; He et al., 2011; Nord et al., 2013). Fifth, we cannot directly evaluate how often regions with regulatory activity are fully tissue-specific, particularly among those we assign as enhancers (Zhu et al., 2013).

One powerful strategy to dissect the regulatory genome has been to identify regions under high sequence constraint (Lindblad-Toh et al., 2011). Testing for activity has revealed that thousands of constrained noncoding regulatory sequences can act as enhancers in embryonic tissues (Pennacchio et al., 2006). The complementary approach we used additionally captures rapidly evolving regulatory regions. The enhancer regions we mapped likely range in function from essential to dispensible, which is reflected both in the modest sequence constraint and rapid evolution between species. Most of these regions would likely be missed by any sequence-conservation based approach. On the other hand, many DNA sequences we do not identify as enhancers may be active in other tissues or embryonic states, which we anticipate to be an area of active investigation.

Rapid enhancer and slow promoter evolution is a fundamental property of the mammalian regulatory genome. Active enhancer elements have a mean lifetime three times shorter than active promoters do, despite similar alignability of their underlying DNA sequences. Comparative sequence-based approaches have limited power to detect regulatory regions, in part because of their rapid evolution (Alföldi and Lindblad-Toh, 2013; Lindblad-Toh et al., 2011); indeed, our data indicate that sequence-based features such as sequence constraint or TF binding site density are poor predictors of enhancer

conservation. Nevertheless, previous work across *Drosophila* species has indicated that specific TF motifs may be preferentially preserved in functionally conserved enhancers (Arnold et al., 2014). In agreement, we found motifs for the liver-specific transcription factor CEBPA enriched in highly conserved liver enhancers.

Active Mammalian Enhancers Are Predominantly Apomorphic

Our results also newly reveal thousands of functionally active regulatory regions conserved across placental mammals, the vast majority of which are proximal promoter sequences. Placental-conserved proximal promoters in mammalian liver are commonly associated with ubiquitously expressed genes. In contrast, only 12% of highly conserved regulatory regions are active enhancers and these are near genes associated with liver-specific activities.

Perhaps our most surprising finding is that representative mammals typically deploy over 10,000 enhancers in a lineage- and probably most often species-specific manner. In total, almost half of all enhancers in each species appear to be recently evolved. Our results confirm and extend the concept that exaptation is a widespread phenomenon across placental mammals (Cotney et al., 2013), and redeployment of ancestral DNA is the dominant mechanism to generate active enhancers across a diverse cross-section of mammals. Interestingly, a recent study comparing enhancer activity across the much smaller genomes of five *Drosophila* species (Arnold et al., 2014) found a similar proportion of gained enhancers, especially for more distant species.

Another mechanism to create regulatory sequences is repeat-carried expansion of regulatory elements. Recent studies have indicated the involvement of specific repeat element expansions in the de novo creation of TF binding sites for CTCF (Bourque et al., 2008; Schmidt et al., 2012), Oct4/Nanog (Kunarso et al., 2010), and NRSF (Mortazavi et al., 2006). Our results show that repeat-carriage of newborn enhancers is not the dominant evolutionary process in mammals: repeat element enrichment is only significant among the recently evolved enhancers found in DNA less than 40 Ma old. Two technical limitations may have caused us to underestimate the repeat-driven creation of recently evolved enhancers (also, see Jacques et al., 2013): the difficulty of mapping reads to recently duplicated regions, and the incomplete representation of repeat regions in genome assemblies.

Recently Evolved Promoters, Though Less Common Than Enhancers, Are Mostly Found in Young DNA

Promoters are far more evolutionarily stable than are enhancers. Nevertheless, the absolute number of promoters deeply conserved across all 20 study species is similar to the number of recently evolved promoters in any one species. Compared to the tens of thousands of newborn enhancers arising from exaptation of ancestral DNA, there are few newborn promoters—and these often arise from DNA sequences that are themselves evolutionarily young. We were not able to identify sequence features that account for the birth of promoters in young DNA. In contrast, the recently evolved promoters arising in ancestral sequences overlap LTR repeats, which enrich for latent non-coding RNA activity (Fort et al., 2014).

A Strategy for Identifying the Enhancer Repertoire of Unannotated Genomes

Finally, extending an approach pioneered in well-annotated primate genomes (Cotney et al., 2013; Shibata et al., 2012), we provide examples of how experimental mapping of enhancers and promoters in newly sequenced mammals can annotate the regulatory network of genes, which have been identified computationally as under positive selection. Across representative species, we discovered that recently evolved enhancers are significantly over-represented in the vicinity of positively-selected genes and can often suggest candidate regulatory elements that could mediate species-specific adaptations. This result was obtained using only a single somatic tissue. Similarly, significant associations likely also exist in between the newly evolved enhancers specific to other somatic tissues and positively selected genes, which would uncover an extensive repertoire of highly evolvable, potentially synergistic regulatory connections.

Future Directions

Our quantitation and analysis of the evolution of promoters and enhancers across a wide cross-section of mammals has revealed how dynamic and rapid enhancer evolution is. Within this regulatory diversity are the instructions by which a small number of founder species have radiated into surprising new niches, including marine (cetaceans) and aerial environments (bats). By combining detailed investigations of carefully selected subclades with new tools for modifying any sequenced genome, future studies will identify, formalize, and explore the functional instructions directing the diversity of mammalian forms.

EXPERIMENTAL PROCEDURES

We performed ChIP-seq using liver tissue isolated from 20 mammalian species (Table S1). At least two independent biological replicates from different animals, generally young adult males, were performed for each species and antibody. The only exception was *Balaenoptera borealis*, for which a single individual was profiled, and dolphin, for which we profiled a single individual from two closely-related species. ChIP-seq experiments were performed as recently described (Aldridge et al., 2013) with antibodies against H3K4me3 (Millipore 05-1339) and H3K27ac (Abcam ab4729). To match inter-individual variability for the two histone marks, the same tissue samples were used for both antibodies and control input DNA in each species.

Sequencing reads were aligned to the appropriate reference genome with BWA v.0.5.9 (Table S2) and regions of enrichment determined with MACS v1.4.2. Regions enriched in two to four biological replicates and overlapping by a minimum 50% of their length were merged and categorized into active promoters (H3K4me3-enriched regions, with or without overlapping H3K27ac enrichment) or enhancers (regions enriched only for H3K27ac). Cross-species comparisons were performed through the Ensembl API. Human, macaque, vervet, marmoset, mouse, rat, rabbit, cow, pig, dog, and cat were directly cross-compared using the 13 eutherian mammals EPO alignment available from Ensembl (Flicek et al., 2014). Species not included in the EPO alignment were compared to the reference species of their respective clade (human, mouse, cow, dog, or opossum) using Lastz aligments. Promoters or enhancers were considered as having conserved activity between species when their orthologous location in the second species overlapped a marked region by a minimum of 50% in length. All pairwise comparisons correspond to average values of reciprocal comparisons between species. Genome annotations (including gene ontology and repetitive and constrained elements) were downloaded from Ensembl v73. See also Extended Experimental Procedures.

ACCESSION NUMBERS

Data have been deposited under ArrayExpress accession number E-MTAB-2633.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.01.006.

AUTHOR CONTRIBUTIONS

D.V., C.B., P.F., and D.T.O. designed experiments; D.V. and S.A. performed experiments; C.B., D.V., T.F.R.,and M.L. analyzed the data; T.J.P., R.D., J.T.E., A.J.J., J.M.A.T., M.F.B., and E.P.M. provided tissue samples; M.P. generated LastZ whole-genome alignments; D.V., C.B., P.F., and D.T.O. wrote the manuscript; P.F. and D.T.O. oversaw the work. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Stephen Watt, Frances Connor, the CRUK-CI Genomics and Bioinformatics cores, Biological Resources Unit (Matthew Clayton), Margaret Brown (West Yorkshire bat hospital), Julie E. Horvath (North Carolina Central University), and Chris Dillingham (University of Cardiff) for technical assistance; Matthieu Muffato for assistance with whole-genome alignments; Claudia Kutter, Gordon Brown, Christine Feig, and Christina Ernst for useful comments and discussions, and the EBI systems team for management of computational resources. This research was supported by Cancer Research UK (D.V., D.T.O.), the European Molecular Biology Laboratory (C.B., P.F.), the Wellcome Trust (WT095908) (P.F.) and (WT098051) (P.F., D.T.O.), the European Research Council, EMBO Young Investigator Programme (D.T.O.), the National Science Foundation (0744979) (T.J.P.), NIH (P40 OD010965, R01 OD010980, R37 MH060233) (A.J.J.) and MRC (U117588498) (J.M.A.T.). Cetacean samples were collected by the UK Cetacean Strandings Investigation Programme, funded by Defra and the Governments of Scotland and Wales.

Received: August 29, 2014
Revised: October 31, 2014
Accepted: December 15, 2014
Published: January 29, 2015

REFERENCES

Aldridge, S., Watt, S., Quail, M.A., Rayner, T., Lukk, M., Bimson, M.F., Gaffney, D., and Odom, D.T. (2013). AHT-ChIP-seq: a completely automated robotic protocol for high-throughput chromatin immunoprecipitation. Genome Biol. *14*, R124.

Alföldi, J., and Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. Genome Res. *23*, 1063–1068.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. Nat. Genet. *46*, 685–692.

Ballester, B., Medina-Rivera, A., Schmidt, D., Gonzàlez-Porta, M., Carlucci, M., Chen, X., Chessman, K., Faure, A.J., Funnell, A.P., Goncalves, A., et al. (2014). Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. eLife *3*, e02626.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al. (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343–348.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cain, C.E., Blekhman, R., Marioni, J.C., and Gilad, Y. (2011). Gene expression differences among primates are associated with changes in a histone epigenetic modification. Genetics *187*, 1225–1234.

Chan, E.T., Quon, G.T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R.A., Aubin, J., Ratcliffe, M.J., Wilde, A., Brudno, M., et al. (2009). Conservation of core gene expression in vertebrate tissues. J. Biol. *8*, 33.

Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell *154*, 185–196.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. Nature *482*, 390–394.

Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res. *22*, 2163–2175.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res. *42*, D749–D755.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassman, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., et al.; FANTOM Consortium (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat. Genet. *46*, 558–566.

Frith, M.C., Ponjavic, J., Fredman, D., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). Evolutionary turnover of mammalian transcription start sites. Genome Res. *16*, 713–722.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008). Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet. *4*, e1000106.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat. Genet. *43*, 414–420.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone

modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Jacques, P.E., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. *9*, e1003504.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome Res. *23*, 800–811.

Kim, J., He, X., and Sinha, S. (2009). Evolution of regulatory sequences in 12 Drosophila species. PLoS Genet. *5*, e1000330.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223–227.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol. Cell *44*, 410–423.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. PLoS Biol. *3*, e93.

Lynch, M., Bobay, L.M., Catania, F., Gout, J.F., and Rho, M. (2011). The repatterning of eukaryotic genomes by random genetic drift. Annu. Rev. Genomics Hum. Genet. *12*, 347–366.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216–219.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593–1599.

Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S., and Rosen, E.D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. Cell *143*, 156–169.

Mortazavi, A., Leeper Thompson, E.C., Garcia, S.T., Myers, R.M., and Wold, B. (2006). Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. Genome Res. *16*, 1208–1221.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A.G. (2007). Recent and ongoing selection in the human genome. Nat. Rev. Genet. *8*, 857–868.

Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell *155*, 1521–1531.

Paris, M., Kaplan, T., Li, X.Y., Villalta, J.E., Lott, S.E., and Eisen, M.B. (2013). Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. PLoS Genet. *9*, e1003748.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499–502.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E.M., Couronne, O., and Pennacchio, L.A. (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res. *16*, 855–863.

Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014). 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet. *10*, e1004525.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

Schwaiger, M., Schönauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res. *24*, 639–650.

Shibata, Y., Sheffield, N.C., Fedrigo, O., Babbitt, C.C., Wortham, M., Tewari, A.K., London, D., Song, L., Lee, B.K., Iyer, V.R., et al. (2012). Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. PLoS Genet. *8*, e1002789.

Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

Sun, Y.B., Zhou, W.P., Liu, H.Q., Irwin, D.M., Shen, Y.Y., and Zhang, Y.P. (2013). Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. Genome Biol. Evol. *5*, 130–139.

Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. Nat. Rev. Genet. *15*, 221–233.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C.C., Musselman, M., Xie, M., West, F.D., Lewin, H.A., et al. (2012). Comparative epigenomic annotation of regulatory DNA. Cell *149*, 1381–1392.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.