MOUSE GENOMICS

Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution

Jeff Vierstra,[1] Eric Rynes,[1] Richard Sandstrom,[1] Miaohua Zhang,[2] Theresa Canfield,[1] R. Scott Hansen,[3] Sandra Stehling-Sun,[1] Peter J. Sabo,[1] Rachel Byron,[2] Richard Humbert,[1] Robert E. Thurman,[1] Audra K. Johnson,[1] Shinny Vong,[1] Kristen Lee,[1] Daniel Bates,[1] Fidencio Neri,[1] Morgan Diegel,[1] Erika Giste,[1] Eric Haugen,[1] Douglas Dunn,[1] Matthew S. Wilken,[4] Steven Josefowicz,[5,6] Robert Samstein,[5,6] Kai-Hsin Chang,[7] Evan E. Eichler,[1,6] Marella De Bruijn,[8] Thomas A. Reh,[4] Arthur Skoultchi,[9] Alexander Rudensky,[5,6] Stuart H. Orkin,[6,10] Thalia Papayannopoulou,[7] Piper M. Treuting,[11] Licia Selleri,[12] Rajinder Kaul,[1,3] Mark Groudine,[2,13] M. A. Bender,[2,14] John A. Stamatoyannopoulos[1,15]*

To study the evolutionary dynamics of regulatory DNA, we mapped >1.3 million deoxyribonuclease I–hypersensitive sites (DHSs) in 45 mouse cell and tissue types, and systematically compared these with human DHS maps from orthologous compartments. We found that the mouse and human genomes have undergone extensive cis-regulatory rewiring that combines branch-specific evolutionary innovation and loss with widespread repurposing of conserved DHSs to alternative cell fates, and that this process is mediated by turnover of transcription factor (TF) recognition elements. Despite pervasive evolutionary remodeling of the location and content of individual cis-regulatory regions, within orthologous mouse and human cell types the global fraction of regulatory DNA bases encoding recognition sites for each TF has been strictly conserved. Our findings provide new insights into the evolutionary forces shaping mammalian regulatory DNA landscapes.

The laboratory mouse *Mus musculus* is the major model organism for mammalian biology and has provided extensive insights into human developmental and disease processes (*1*). At 2.7 Gb, the mouse genome is comparable to the 3.3-Gb human genome in size, structure, and sequence composition (*2*, *3*), and >80% of mouse genes have human orthologs (*1*, *4*).

Human-to-mouse transgenic experiments have collectively demonstrated that the mouse is capable of recapitulating salient features of human gene regulation, often with striking precision and even in the case of human genes that lack mouse orthologs (*5*). By contrast, comparative analyses of regulatory regions governing individual gene systems (*6*), as well as the occupancy patterns of several TFs (*7*), have highlighted the potential for cis-regulatory divergence. However, broader efforts to identify and quantify the major forces shaping the evolution of the mammalian cis-regulatory landscape have been hampered by the lack of expansive and highly detailed regulatory DNA maps from diverse cell fates that can be directly compared between mouse and human.

Deoxyribonuclease I (DNase I)–hypersensitive sites (DHSs) mark all major classes of cis-regulatory elements in their cognate cellular context, and systematic delineation of DHSs across human cell types and states has provided fundamental insights into many aspects of genome control (*8*). In conjunction with the Mouse ENCODE Project (*9*), we undertook comprehensive mapping of DHSs in diverse mouse cell and tissue types and systematically compared the resulting maps to those from orthologous and nonorthologous human cells and tissues.

We mapped DHSs in 45 mouse cell and tissue types including adult primary tissues (*n* = 19), purified adult and primitive primary cells (*n* = 10), primary embryonic tissues (*n* = 4), embryonic

stem cell lines (*n* = 4), and model immortalized primary (*n* = 3) and malignant (*n* = 5) cell lines (Fig. 1A, fig. S1A, and table S1). We identified between 74,386 and 218,597 DHSs per cell type at a false discovery rate threshold of 1%, and collectively delineated 1,334,703 distinct ~150–base pair DHSs, each of which was detected in one or more mouse cell or tissue types. The genomic distribution of DHSs relative to annotated genes and transcripts was similar to that observed in human (*8*) (fig. S1B). On average, 13.5% of DHSs marked promoters, with the remaining 86.5% distributed across the intronic and intergenic compartments in roughly equal proportions; the vast majority were located within 250 kb of the nearest annotated transcriptional start site (TSS) (fig. S1C). However, average intergenic DHS-to-TSS distances in the mouse genome were markedly compressed (median 48.7 kb versus 91.6 kb for human) relative to genome size (2.7 Gb versus 3.3 Gb), indicating differential rates of genome remodeling within DHS-rich regions (fig. S1D), with a pronounced difference in both size and density of distal elements (fig. S2, A and B).

To gain insight into the evolution of mammalian regulatory DNA, we comprehensively integrated the mouse DHS maps with human maps generated using the same methods from 232 cell or tissue types from the ENCODE Project (*n* = 103) (*8*) and the Roadmap Epigenomics Project (*n* = 126) (*10*). These human maps collectively encompass ~3 million distinct DHSs from primary cells, adult and fetal tissues, immortalized and malignant lines, and embryonic stem cells (table S2). We used high-quality pairwise alignments and a conservative reciprocal mapping and filtering strategy to project the genomic sequence underlying all mouse and human DHSs to the other species (Fig. 1, B and C, and fig. S3A). Collectively, 59.5% of mouse DHSs (52.5 to 78.8% per cell type) could be aligned with high confidence to the human genome, of which 35.6% (38.6 to 60% per cell type) coincided with a human DHS (Fig. 1B and table S3). The remaining 23.9% (13 to 22.7% per cell type) may correspond to human DHSs not yet defined, or to human lineage–specific extinction of an ancestral element. In support of the latter, mouse DHSs aligning outside of human DHSs showed excess sequence divergence, as evidenced by fewer alignable or identical nucleotides relative to mouse DHSs that aligned with human DHSs (fig. S3, B and C). A smaller proportion of human DHSs aligned with a mouse DHS (17.3%; fig. S3A and table S4); however, this was largely because there are more than twice as many DHSs identified in human. Given the breadth of mouse and human tissues analyzed, these values suggest upper and lower limits of regulatory DNA conservation between mouse and human.

To trace the evolutionary origins and dynamics of individual regulatory regions, we aligned all mouse and human DHS sequences to >30 vertebrate genomes spanning ~550 million years of evolutionary distance (fig. S4, A and B). Despite the deep sequence conservation of many DHSs, turnover of individual regulatory regions

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [2]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [3]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA. [4]Department of Biological Structure, University of Washington, Seattle, WA 98195, USA. [5]Immunology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA. [6]Howard Hughes Medical Institute. [7]Division of Hematology, Department of Medicine, University of Washington, Seattle, WA 98195, USA. [8]Medical Research Council (MRC) Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK. [9]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [10]Division of Hematology/Oncology, Children's Hospital Boston and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115, USA. [11]Department of Comparative Medicine, University of Washington, Seattle, WA 98195, USA. [12]Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, NY 10065, USA. [13]Department of Radiation Oncology, University of Washington, Seattle, WA 98109, USA. [14]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. [15]Division of Oncology, Department of Medicine, University of Washington, Seattle, WA 98195, USA.
*Corresponding author. E-mail: jstam@uw.edu

within different branches of the evolutionary tree appeared frequently. Of the 80% of mouse DHS sequences that predate the divergence of humans from a common ancestor, only 58.5% were detectable in human, and comparison of mouse DHSs aligning to a human DHS or to a non-DHS region yielded nearly identical evolutionary profiles (fig. S4, A and B). Overall, the proportion of DHSs that encompassed evolutionarily conserved sequence elements increased with alignability and conservation of DNase I hypersensitivity (fig. S4B). Unexpectedly, however, ~40% of mouse-human shared DHSs lacked conserved sequence elements.

The aforementioned trends are also reflected in patterns of human variation. Analysis of nucleotide diversity (π) within DHSs indicated graded constraint depending on the extent of sequence and DHS conservation (fig. S5A). Notably, mean π within human-specific DHSs approximated that of fourfold synonymous sites within coding regions, compatible with relaxed (but not absent)

nucleotide-level constraint. Despite decreased constraint (both evolutionary and recent), human-specific DHSs are significantly enriched (versus all DHSs) in disease- and trait-associated variants identified by genome-wide association studies (fig. S5B; permutation test, $P_{null} < 0.005$). The above results indicate that although mouse-human shared DHSs are collectively under selection over evolutionary time scales and within human populations, the sequence information with the cis-regulatory compartment is continuing to evolve rapidly in both mice and humans.

Whereas the overall density of mouse-human shared DHSs was higher in gene-proximal regions such as promoters, exons, and UTRs (Fig. 1D), the relative proportion of shared DHSs (to all DHSs) increased markedly with distance from the TSS (Fig. 1E and fig. S6). From 10 to 50 kb upstream of the TSS, the proportion of DHSs that are shared with human (average 27%) was lower than the average for intergenic regions (average 31%; Fig. 1E), whereas in far distal regions this proportion

increased substantially to a plateau of ~38%. These data suggest that regulatory elements functioning over long range (>100 kb) (*11*) constitute a genomic compartment that may be operationally distinct from a more rapidly evolving gene-proximal region, and hence less buffered against evolutionary alteration.

The genesis of novel regulatory DNA sequences appears to have played a substantial role in shaping the DHS landscape of both mouse and human (Fig. 1B and fig. S2A). More than 50% of the mouse and human genomes consist of repetitive DNA (*2*, *3*), which is proportionately reflected in their respective DHS compartments (fig. S7, A and B). Species-specific DHSs were enriched (relative to all DHSs) for nearly all classes of repetitive elements (fig. S7C), and 5 to 10% of shared DHSs overlapped ancient repeats that predate mouse/human divergence (fig. S7D)—a finding compatible with an important role for transposons in the evolution of mammalian regulatory genomes.
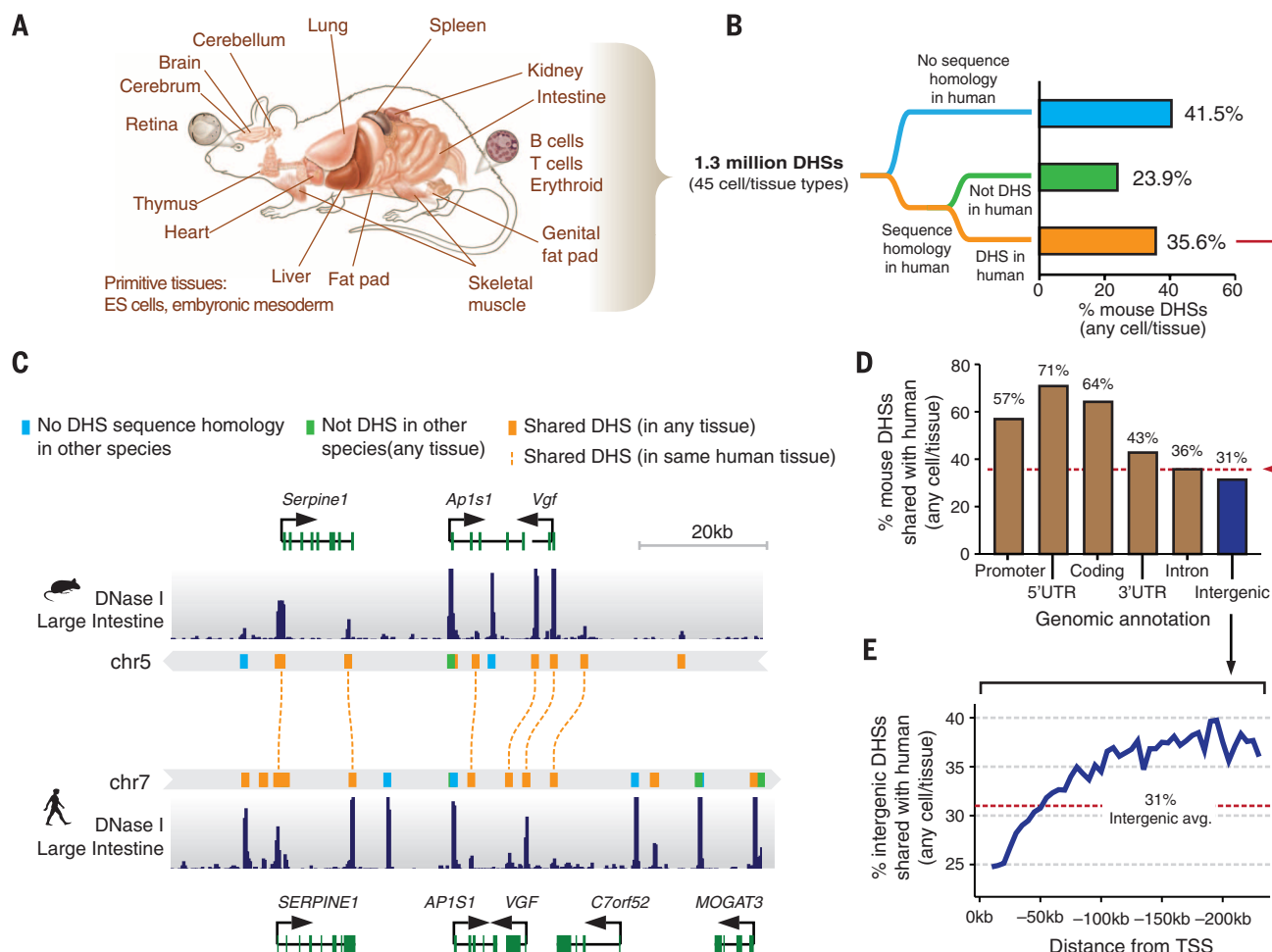
Fig. 1. Conservation of mouse regulatory DNA in humans. (**A**) The accessible landscape of the mouse was derived from 45 tissues and cell types. (**B**) Proportions of the mouse regulatory DNA landscape with sequence homology and functional conservation with human. (**C**) Example of the conservation of the cis-regulatory elements surrounding within the *Vgf/VGF* locus in mouse and human intestine. (**D**) Gene-proximal DHSs are more likely to be conserved than distal DHSs. Dashed red line indicates the average conservation of DHSs. (**E**) The rate of intergenic DHS conservation versus distance to nearest TSS indicates a rapidly evolving cis-regulatory domain.

Transposable elements have recently been implicated in the rapid expansion of TF recognition elements (*12*, *13*). To test the generality of this phenomenon, we estimated the total proportion of TF recognition sequences residing within species-specific DHSs that arose from transposon expansion during mouse and human evolution, which revealed substantial asymmetries (fig. S8, A to C). For example, the recognition motif for the pluripotency factor OCT4 (and other POU family TFs) has been greatly expanded in the murine lineage on a LTR/ERVL element (*12*), accounting for >25% of mouse-specific sites versus <5% in humans with a similar class of retroelement (fig. S8A). By contrast, expansions of CTCF (*13*) and retinoic acid receptor recognition elements (*14*) have been driven chiefly by short interspersed elements (SINEs) in both mouse and human (fig. S8, B and C). These results indicate that expansion of TF recognition sequences by repetitive elements is a general feature shaping mammalian cis-regulatory landscapes.

DHS patterns encode cellular fate and identity in a manner that reflects both current and future regulatory potential and informs an organism's developmental trajectory (*15*). To visualize cell- and tissue-selective activity patterns, we clustered shared DHSs by normalized DNase I cleavage measured in each of the 45 mouse cell and tissue types (Fig. 2A). The vast majority of shared DHSs (78.8%) displayed tissue-selective accessibility and were readily organized into distinct cohorts. A minority (21.2%) exhibited high accessibility across multiple tissue types, whereas <5% were constitutive (Fig. 2B). Tissue-selective shared DHSs were enriched in distal regions (fig. S9) and reflected both tissue organization and anatomic or functional compartments within tissues. For example, the 91,951 shared brain-selective DHSs in turn comprised four subclusters corresponding to distinct anatomical and developmental par-

titions (Fig. 2A, green box). Similarly, shared blood-selective DHSs were subcompartmentalized into major hematopoietic lineages, including T, B, myeloid, and erythroid cell cohorts (Fig. 2A, red boxes). Across all compartments, cell- or tissue-selective shared DHSs were preferentially localized around genes critical for the development and maintenance of their respective cell or tissue type (fig. S10).

We hypothesized that tissue-selective shared DHSs should encode information critical for basic mammalian regulatory processes such as development and differentiation, and that this would be reflected in their TF recognition sequence content. We thus computed, for each TF, the number of DHSs within each cluster that contained its recognition sequence, and compared this value to the overall distribution of recognition sequences within all shared DHSs. Tissue-selective DHSs showed pronounced enrichment for nearly all known lineage-specifying or cell
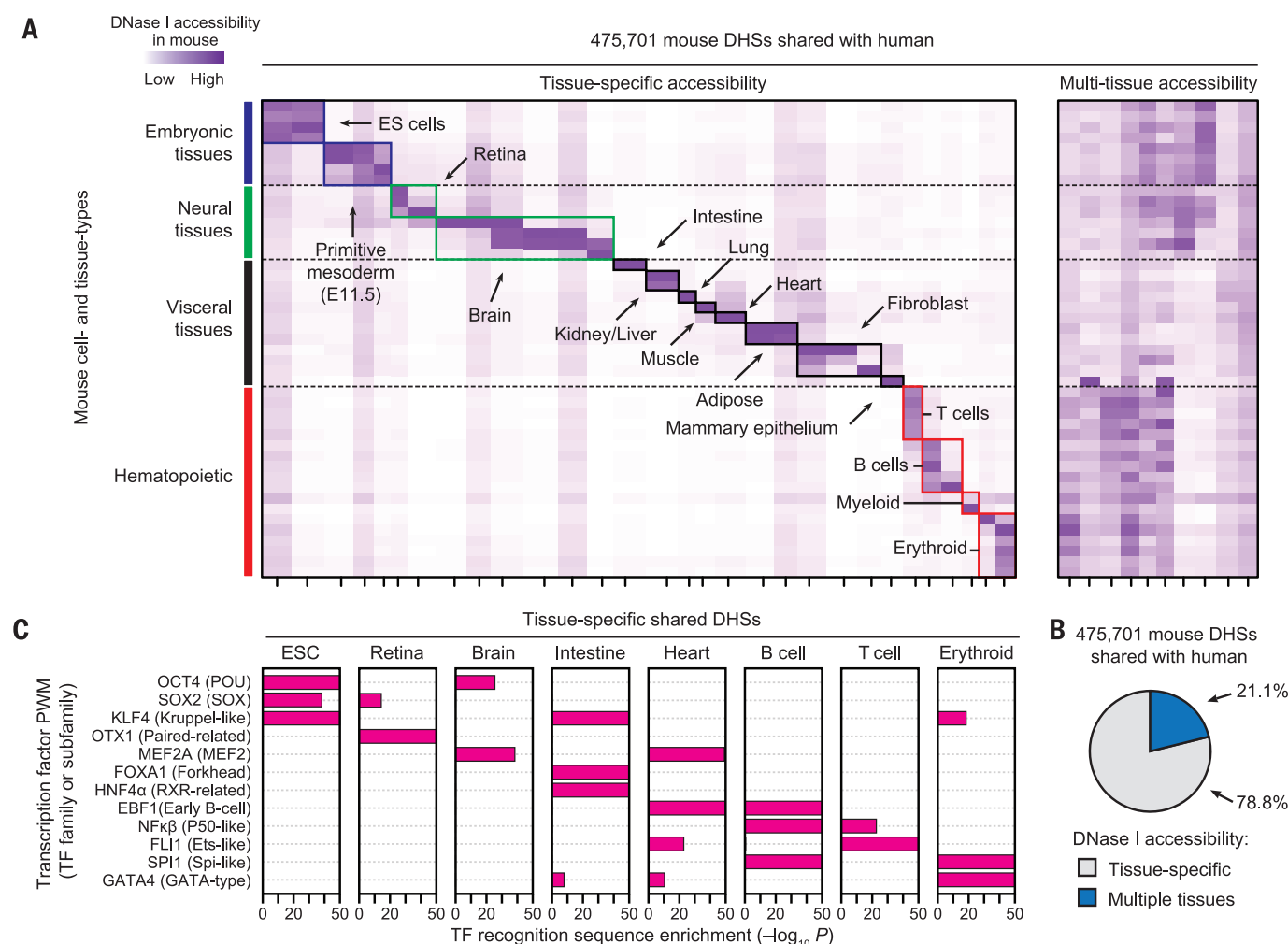
Fig. 2. Cell and tissue lineage encoding within shared regulatory elements. (**A**) *k*-means clustering of DHSs by accessibility at each of the 475,701 mouse DHSs shared with human. Columns correspond clusters of mouse DHSs that are also accessible in human; rows correspond to the 45 mouse cell or tissue types. Colors (axes and boxes) distinguish tissue groupings. Left, tissue-selective clusters; right, clusters containing DHSs active in multiple tissues. (**B**) Proportion of shared DHSs that are tissue-selective or active in multiple tissues. (**C**) Enrichment of TF recognition sequences within tissue-selective DHSs computed using the cumulative hypergeometric distribution.

identity–specifying regulators, which were further organized combinatorially into their respective functional compartments (Fig. 2C and fig. S11). For example, OCT4, SOX2, and KLF4 recognition sites were collectively concentrated within embryonic stem cell–selective shared DHS landscapes, consistent with coordinated expression of their cognate factors in embryonic stem cells. KFL4 recognition sites were also enriched within intestine- and erythroid-specific DHSs, consistent with the known role of Krüppel-like TFs (many of which share the KLF4 recognition sequence) in intestinal epitheliogenesis (16) and in erythropoiesis (17). Analogously, sequence elements recognized by the cardiac regulators MEF2A, EBF1, FLI1, and GATA4 (18–20) were enriched within heart-selective shared DHSs, compatible with important functions for these TFs or their cognates in defining their respective cell fates (18, 21, 22). Nonetheless, the tissue-selective enrichments we observed are consistent with the known cell-selective activity of TFs even after recognition sequences are systematically grouped by similarity (fig. S11). Together, these results indicate that mouse-human shared DHSs densely encode regulatory information fundamental to diverse cell and tissue specification programs, and thus collectively define a core mammalian regulon.

Because most shared DHSs showed strong cell or tissue selectivity in mouse, we next asked to what degree these patterns were preserved in human. Computing the Jaccard similarity index over all possible combinations of mouse and human cell types revealed surprisingly limited similarity in the tissue-selective usage of shared DHSs (fig. S12, A to C), even when accounting for variability in DNase I cleavage density and peak identification parameters (fig. S13). Unsupervised hierarchical clustering resulted in loose groupings of shared DHSs by cells or tissues derived from the same progenitor or developmental lineage (Fig. 3A).

Weak correspondence between orthologous tissues suggested that a substantial fraction of shared DHSs had undergone functional "repurposing" via alteration of tissue activity patterns from one tissue type in mouse to a different one in human (Fig. 3, B and C). Indeed, analysis of well-matched mouse and human tissue pairs confirmed substantial repurposing ranging from 22.9 to 69% of shared DHSs, depending on the tissue (Fig. 3B). For example, of the 77,060 shared DHSs active in mouse muscle, 59,658 (77.4%) were also DHSs in human muscle; the remaining 17,402 (22.6%) were DHSs in a different human tissue (Fig. 3B, 7th from top). Overall, we found

that at least 35.7% of shared DHSs (12.7% of mouse DHSs overall) have undergone repurposing (Fig. 3D), chiefly affecting distal elements (fig. S14). Facile repurposing of regulatory DNA from one tissue context to another thus emerges as an important evolutionary mechanism shaping the mammalian cis-regulatory landscape.

To examine the conservation of individual TF recognition elements within the shared DHS compartment, we distinguished between elements that were positionally conserved versus those that were operationally conserved (i.e., have arisen independently at a different position within the DHS) (fig. S15A). In shared DHSs, 39.1% of TF recognition sequences were positionally conserved and 19.6% were operationally conserved (Fig. 4A). Both positional and operational conservation were significantly concentrated (χ^2 test, $P < 10^{-15}$) within shared DHSs that maintained their tissue activity profile (Fig. 4B and fig. S15B). Surprisingly, 41.3% of shared DHSs (chiefly repurposed DHSs) lacked any positionally or operationally conserved TF recognition elements (Fig. 4, A and B, and fig. S15, C and D). Additionally, the overall density of TF recognition elements did not differ substantially between shared DHSs with positionally, operationally, or nonconserved TFs (fig. S15E). This indicates that new regulatory features are

Fig. 3. Conservation and repurposing of regulatory DNA activity. (**A**) Pairwise comparison (median Jaccard distance) of shared DHS landscape usage between all mouse (rows) and human (columns) tissues largely mirrors their conserved morphological and embryological origins. (**B**) Conservation of mouse cis-regulatory DNA accessibility in human for individual tissue types. Orange ticks indicate the expected overlap of randomly selected DHSs. (**C**) The activity patterns of individual shared DHSs during mouse and human evolution may have been conserved (activity in at least one similar tissue) or repurposed to another tissue. (**D**) Overall conservation of tissue-level accessibility patterns of mouse DHSs shared with human.
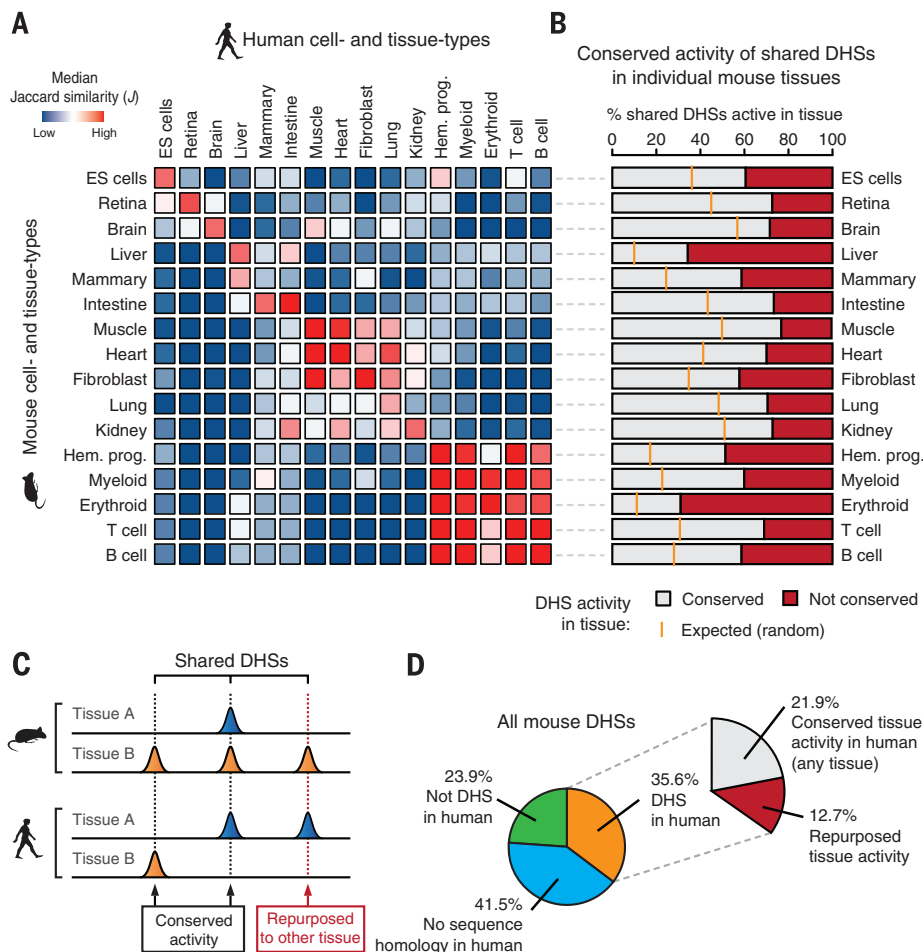
continuously evolving within the same ancestral DNA segment.

We next elaborated the relationship between conservation of TF recognition sites and the maintenance of tissue accessibility patterns. Reasoning that known regulators of cell fate would play an outsized role in repurposing, we hypothesized that recognition sequences for such TFs would be preferentially maintained (or gained) in DHSs with conserved tissue activity spectra but would be preferentially lost at repurposed DHSs (fig. S16). We found this to be the case across a spectrum of lineage-regulating TFs. For example, recognition sites for the retinal master regulator OTX1 (and other paired-related homeodomain family TFs) within mouse retinal DHSs that had undergone repurposing in human were depleted

by a factor of >4 relative to orthologous DHSs that had conserved retinal activity (Fig. 4C). Analogously, sequence elements recognized by the intestinal master regulator HNF1β (and by other POU-homeobox TFs) were selectively depleted in repurposed intestinal DHSs, and those recognized by the major erythroid regulator GATA1 (and by other GATA-type factors) were selectively depleted in repurposed erythroid DHSs (Fig. 4C). Overall, we found that recognition sites for cell fate–modifying TFs were consistently depleted within repurposed DHSs (Fig. 4D), linking the conservation and repurposing of DHSs to preservation versus turnover of specific TF recognition sequences.

The above results also suggest an incremental process whereby the composition of TFs within a given DHS is remodeled over evolutionary time

via sequential small mutations (23) that could ultimately affect function and phenotype (24). The presence of a substantial population of shared DHSs without conserved TF recognition sites but with preserved tissue selectivity patterns highlights the plasticity of individual cis-regulatory templates. Such a finding indicates that the same higher-level regulatory outcome may be encoded by many different combinations of instructive TF recognition events.

To investigate how the marked plasticity of TF recognition elements within the evolving cis-regulatory landscape is reflected in global patterns of the types and quantities of such elements, we computed the global density of recognition sequences for each of 744 TFs within all mouse and human DHSs (separately, and irrespective of conservation status) from each cell or tissue type. This analysis revealed striking conservation of the proportion of the regulatory DNA landscape of each cell type devoted to recognition sites of each TF. Shown in Fig. 5, A and B, are examples for mouse versus human regulatory T cell DHSs and for mouse brain versus human fetal brain. In each case a linear relationship is observed, indicating that the proportion of the DHS compartment devoted to recognition sequences of each of the 744 TFs has been strictly conserved (Fig. 5A). It is noteworthy that this finding obtains across a wide spectrum of TFs that encompass diverse functional roles and biophysical mechanisms of DNA recognition. These findings are in marked contrast to the weak conservation (~25%) of individual mouse regulatory T cell and brain DHSs (Fig. 5, C and D). TF recognition sequence content varied between cell types and between tissue types, with effector TFs selectively enriched within their cognate cell type (fig. S17), and TF recognition sequence density was consistently more similar between orthologous cell or tissue pairs than between non-orthologous cells or tissues (Fig. 5E and fig. S18).

It has been proposed that in large genomes such as mouse and human, maximization of the occupancy of any given TF requires an excess of its recognition sites, so as to ensure high occupancy of sites with critical regulatory roles across a range of TF concentrations (25). Consistent with this model, the majority of DHSs in both the mouse and human genomes show relaxed sequence constraint over evolutionary distances (fig. S4C) and within human populations (fig. S5A). This model also predicts that the cis-regulatory programs of TF genes themselves should be more highly conserved than other gene classes. Comparing DHSs within 50 kb of the TSSs of TF genes ($n = 911$) relative to those of all orthologous genes ($n = 14,666$ with at least 10 identified DHSs in mouse) revealed an overall increase in the conservation of TF-linked DHSs (Wilcoxon rank sum test, $P < 10^{-15}$) (fig. S19), particularly for DHSs surrounding the TSSs of genes within canonical TF families, such as Hox and Sox factors. As such, TFs are distinguished from other trans-acting regulators in that their activity appears to directly shape their cis-regulatory landscape.
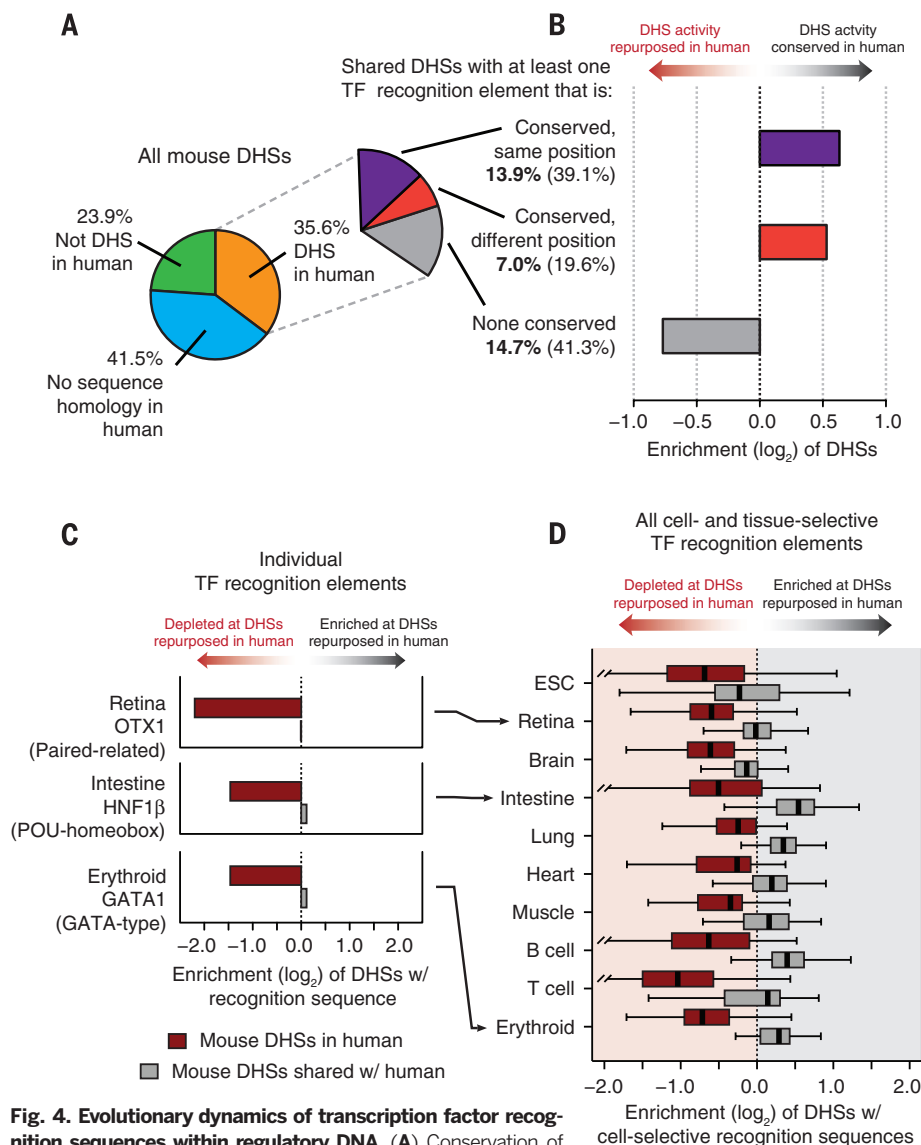
Fig. 4. Evolutionary dynamics of transcription factor recognition sequences within regulatory DNA. (**A**) Conservation of TF recognition sequences within shared DHSs. (**B**) Conserved TF recognition sequences (both positional and operational conservation) are enriched within DHSs that have conserved tissue activity patterns. (**C**) Recognition sequences for cell-selective TFs are preferentially lost at mouse DHSs that are repurposed in human but are maintained or gained in human. Representative examples of individual TF regulators in retina, intestine, and erythroid tissues are shown. (**D**) Same as (C) for recognition sequences of all cell-selective TF regulators (identified in Fig. 2C) within mouse DHSs repurposed in human.

Fig. 5. Conservation of global cis-regulatory content predominates that of individual regulatory elements. (**A**) Density of individual TF recognition sequences in human (*x* axis) and mouse (*y* axis) regulatory T cells. Dotted black lines demarcate a factor of 2 difference in density between mouse and human. (**B**) Same as (A) for human and mouse brain. (**C** and **D**) Proportion of mouse DHSs that are conserved in a matched human tissue. Top, mouse regulatory T cell DHSs that are conserved in human regulatory T cells; bottom, mouse embryonic brain DHSs that are conserved in human fetal brain. (**E**) Radar plots showing the median similarity (Euclidean distance between the distributions of TF recognition sequence densities) of the cis-regulatory content between mouse and human tissues.
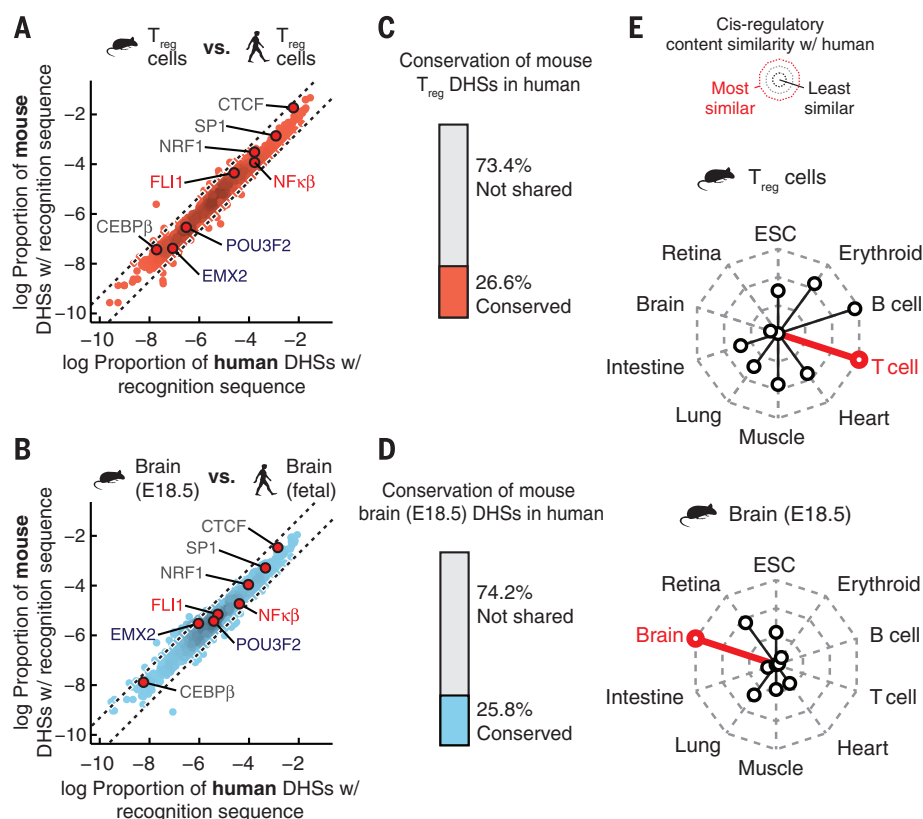
Taken together, our results have important implications for understanding the major mechanisms and forces governing the evolution of mammalian regulatory DNA. Performing genomic footprinting on 25 of the cell and tissue samples analyzed herein reveals that the effective in vivo recognition repertoires of human and mouse TFs are highly similar, and that the high turnover of individual TF occupancy sites within regulatory DNA is accompanied by striking evolutionary stability at the level of regulatory networks (*26*). As such, the combination of a highly conserved trans-regulatory environment with a large genome (under weakened selection) may function to potentiate both the de novo creation and the cis-migration of operational TF binding elements. We speculate that high cis-regulatory plasticity may be a key facilitator of mammalian evolution by increasing the potential for innovation of novel functions in the context of an evolutionarily inflexible trans-regulatory environment.

REFERENCES AND NOTES

1. S. N. Hardouin, A. Nagy, *Clin. Genet.* **57**, 237–244 (2000).
2. A. T. Chinwalla *et al.*, *Nature* **420**, 520–562 (2002).
3. E. S. Lander *et al.*, *Nature* **409**, 860–921 (2001).
4. J. L. Guénet, *Genome Res.* **15**, 1729–1740 (2005).
5. K. R. Peterson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11207–11211 (1993).
6. E. T. Dermitzakis, A. G. Clark, *Mol. Biol. Evol.* **19**, 1114–1121 (2002).
7. D. Schmidt *et al.*, *Science* **328**, 1036–1040 (2010).
8. R. E. Thurman *et al.*, *Nature* **489**, 75–82 (2012).
9. Mouse ENCODE Project Consortium, An integrated and comparative encyclopedia of DNA elements in the mouse genome. *Nature* 10.1038/nature13992 (2014).
10. B. E. Bernstein *et al.*, *Nat. Biotechnol.* **28**, 1045–1048 (2010).
11. L. A. Lettice *et al.*, *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
12. G. Bourque *et al.*, *Genome Res.* **18**, 1752–1762 (2008).
13. D. Schmidt *et al.*, *Cell* **148**, 335–348 (2012).
14. D. Laperriere, T.-T. Wang, J. H. White, S. Mader, *BMC Genomics* **8**, 23 (2007).
15. A. B. Stergachis *et al.*, *Cell* **154**, 888–903 (2013).
16. J. M. Shields, R. J. Christy, V. W. Yang, *J. Biol. Chem.* **271**, 20009–20017 (1996).
17. Q. Xiong *et al.*, *BMC Genomics* **14**, 587 (2013).
18. S. Garel, F. Marín, R. Grosschedl, P. Charnay, *Development* **126**, 5285–5294 (1999).
19. D. G. Edmondson, G. E. Lyons, J. F. Martin, E. N. Olson, *Development* **120**, 1251–1263 (1994).
20. W. Schachterle, A. Rojas, S.-M. Xu, B. L. Black, *Dev. Biol.* **361**, 439–449 (2012).
21. M. K. Anderson, G. Hernandez-Hoyos, R. A. Diamond, E. V. Rothenberg, *Development* **126**, 3131–3148 (1999).
22. M. Merika, S. H. Orkin, *Mol. Cell. Biol.* **13**, 3999–4010 (1993).
23. J. L. Payne, A. Wagner, *Science* **343**, 875–877 (2014).
24. B. Prud'homme *et al.*, *Nature* **440**, 1050–1053 (2006).
25. S. Lin, A. D. Riggs, *Cell* **4**, 107–111 (1975).
26. A. B. Stergachis *et al.*, *Nature* 10.1038/nature13972 (2014).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6212/1007/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S4
References (*27–52*)

25 September 2013; accepted 27 October 2014
10.1126/science.1246426