Supplementary Materials for

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

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

Correspondence to: jstam@uw.edu

This PDF file includes:

Materials and Methods

Mouse strains

The mouse cell lines and tissues used in this study are outlined in **table S1**.

Nuclei isolation from solid mouse tissues

Solid mouse tissues were minced in 2 mm square pieces and resuspended in 3 mL of homogenization buffer (20 mM tricine, 25 mM D-sucrose, 15 mM NaCl, 60 mM KCl, 2 mM $MgCl_2$, 0.5 mM spermidine, pH 7.8) per gram of tissue. The nuclei were released by 5-10 strokes in a Dounce homogenizer with a loose-fitting type-A pestle and the resulting homogenate was filtered through a 100μm filter. For some samples, the homogenate was cryopreserved before DNaseI treatment (10% DMSO was added for a controled freeze to –80°C; stored in liquid nitrogen). Prior to DNase I treatment an additional buffer exchange was performed (for both fresh and frozen samples) by first adding 15 mL of sucrose buffer (10 mM Tris-HCl, 250 mM D-sucrose, 1 mM $MgCl_2$, pH 7.5), collecting nuclei by centrifugation (600*g* for 10 minutes at 4°C), and then resuspending nuclei pellet in 10 mL of fresh sucrose buffer. The nuclei were filtered through a 20μm filter and collected by centrifugation (600*g* for 10 minutes at 4°C). The nuclei pellet was washed once in 10 mL of buffer A (15 mM Tris-HCl, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine) and resuspended at concentration of 2 x 10^6 per mL.

Nuclei isolation of mouse cultured and primary cells

Cells (primary or cultured) were washed once with Dulbecco's PBS (without $MgCl_2$ or $CaCl_2$). Nuclei were extracted by resuspending cells in buffer A supplemented with detergent (IGEPAL-CA630) (Sigma) and incubating for 10 minutes on ice. Following incubation, the nuclei were collected by centrifugation (600*g*) and resuspended in buffer A at a concentration of 2 x 10^6 nuclei per mL. Optimal detergent concentrations for nuclei extraction were empirically derived for each cell type (commonly ranging from 0.010–0.10%).

DNase I digestion of mouse nuclei

Fresh nuclei were incubated for 3 minutes at 37°C with limiting concentrations of the DNA endonuclease deoxyribonuclease I (DNase I) (Sigma) in buffer A supplemented with Ca^{2+}. The digestion was stopped with stop buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1% SDS, 100 mM EDTA, 1 mM spermidine, 0.5 spermine, pH 8.0) and the samples were treated with proteinase K and RNase A. The small 'double-hit' fragments (<750 bp) were recovered by sucrose ultracentrifugation, end-

repaired and ligated with adapters compatible with the Illumina sequencing platform. A detailed protocol describing genome-wide mapping of DNase I hypersensitivity can found in (*27*).

DNase I fragment sequence alignment and normalization

Mouse (NCBI37) and human (GRCh37) 36 bp sequence reads were mapped using bowtie (*28*), version 0.12.7 with parameters: "`--mm -v 3 -k 2`", with "`--phred33-quals`" for Illumina HiSeq sequencer runs or "`--phred64-quals`" for Illumina GAII sequencer runs. Only uniquely mapping reads with up to 2 mismatches were retained; this was accomplished by additionally filtering the "`-k 2`" results, when present, for actual uniqueness within the potential 2-3 mismatch alignments (any remaining 3-mismatch-only alignments were discarded). Signal tracks were generated using BEDOPS (*29*), summing reads within a window size of ±75bp in 20 bp steps and subsequently normalizing to the total number of reads per dataset and then scaling to one million reads. The NCBI and Mouse Genome Sequencing Consortium build 37 (MGSCv37/mm9) including chromosome Y was used as the reference assembly for all sequence alignments.

DHS identification and master list creation

We identified DNase I hypersensitive regions of chromatin accessibility (hotspots) and more highly accessible DNaseI hypersensitive sites (DHSs, or peaks) within the hotspots, using the hotspot algorithm (*30*). Briefly, the hotspot algorithm is a scan statistic that uses the binomial distribution to gauge enrichment of tags based on a local background model estimated around every tag. General-sized regions of enrichment are identified as hotspots, and then 150 bp peaks within hotspots are called by looking for local maxima in the tag density profile (sliding window tag count in 150 bp windows, stepping every 20 bp). Further stringencies are applied to the local maximum detections to prevent overcalling of spurious peaks. Hotspot also includes a FDR (false discovery rate) estimation procedure for thresholding hotspots and peaks, based on a simulation approach. Random reads are generated at the same sequencing depth as the target sample, hotspots are called on the simulated data, and the random and observed hotspots are compared via their *Z*-scores (based on the binomial model) to estimate the FDR.

The DHSs called on individual cell-types were consolidated into a master list of unique, non-overlapping DHS positions by first merging the FDR 1% peaks across all cell-types. Then, for each resulting interval of merged sites, the DHS with the highest Z-score was selected for the master list. Any DHSs overlapping the peaks selected for the master list were then discarded. The remaining DHSs were then merged and the process repeated until each original DHS was either in the master

list or discarded.

Due to the variability in sequencing depth within the mouse DNase I experiments, we down-sampled each mouse dataset to 25 million sequencing tags (random sampling without replacement) and used the sampled tags to detect hotspots and DHSs.

Data availability

All mouse sequence data generated for this study can be accessed with GEO accession numbers found within **table S1**. Processed data such as hotspots and peaks are released as part of the Mouse ENCODE Consortium and available for download at the data portal website (`http://www.mouse-encode.org`).

Human DNase I data

Human DNase I data was generated as part of the ENCODE Project (*31*) or the Roadmap Epigenomics Mapping Consortium (*10*). Raw sequence tag data can be accessed through the GEO accession numbers in **table S2**. Processed data such as hotspots and DHS peaks can be accessed at `http://www.encode-roadmap.org/`.

Genomic annotations

Genome annotations used for all analysis correspond to the GENCODE version 10 (*32*) (human) or ENSEMBL 65 (*33*) (mouse). Promoters are defined as 1 kb upstream of a TSS.

Alignment of DHSs between mouse and human

Pair-wise genome alignments ("over" chain files) between mouse (mm9) and human (hg19) were downloaded from the UCSC Genome Browser (*34*). Using these chain files, DHSs were mapped between species using the software "bnMapper" (bx-python software package; available at http://bitbucket.org/james_taylor/bx-python) using the following parameters: "`-fBED12 --gap 20 --threshold 0.1`". The mapped blocks were then intersected with a DHS peak list from the query species using the software BEDOPS (*29*), requiring only 1 base-pair of overlap. We applied this mapping strategy for each target-query pair (i.e., human → mouse and mouse → human) and then retained DHSs in each list that were in strict reciprocal relationships.

Alignment of mouse DHS sequence to the vertebrate phylogeny

Pairwise genome alignments ("over" chain files) were downloaded from the UCSC Genome Browser

between mouse (mm9) or human (hg19) to the following genomes: panda (ailMel1), lizard (anoCar2), cow (bosTau7), lancelet (braFlo1), marmoset (calJac3), dog (canFam2), guinea pig (cavPor3), zebrafish (danRer7), tenrec (echTel1), horse (equCab2), cat (felCat4), fugu (fr2), chicken (galGal3), stickleback (gasAcu1), human (hg19), elephant (loxAfr3), turkey (melGal1), opossum (monDom5), platypus (ornAna1), rabbit (oryCun2), medaka (oryLat2), sheep (oviAri1), chimp (panTro3), lamprey (petMar1), orangutan (ponAbe2), macaque (rheMac2), rat (rn5), pig (susScr2), tetradon (tetNig2), and frog (xenTro3). Using the sequence alignment strategy described above (see *Cross-alignment of DHSs between mouse and human*), each mouse DHS was to each of these genomes.

To estimate the proportion of regulatory elements conserved throughout vertebrate evolution, we overlayed the sequence alignments on a recent proposed vertebrate phylogenies obtained from (*35*-*37*). We considered a DHS to be conserved if it aligned successfully to one or more species within a branch of the phylogeny that share a MRCA with mouse.

Evolutionary sequence constraint

The phastCons (*38*) element track corresponding to a 46-way multiple alignment of vertebrate species was downloaded from the UCSC Genome Browser (*34*). To assess conservation, we computed the fraction of phastCons elements that overlapped mouse DHSs grouped by conservation status with human.

Nucleotide diversity

Human nucleotide diversity measurements (π) were calculated using whole genome sequences from 53 unrelated, publicly available human genomes released by Complete Genomics (*39*) (version 1.1034) as previously described (*8*, *40*). π for a heterozygous site is $2pq$, where p is the major allele frequency and q the minor allele frequency. π was calculated for each human tissue by summing $2pq$ all variants and dividing by the total number of nucleotides considered. Repetitive elements identified by RepeatMasker (*41*, *42*) (see *Analysis of repeat content within DHS*) were removed from all π calculations.

Functional variation within human DHSs

A catalog of single nucleotide polymorphisms (SNPs) identified by genome-wide association studies was downloaded from NHGRI GWAS Catalog on December 3, 2013 from: `http://www.genome.gov/admin/gwascatalog.txt`. SNPs within coding regions were removed

yielding a total of 6,571 total loci. We randomly sampled the same number of SNPs from the 1000 Genomes Project (*43*) matched by distance to TSS, intergenic or intronic, and allele frequency as in (*44*). We repeated this sampling process 1,000 times to estimate the parameters of a normal distribution (μ, σ). These parameters were used to calculate the upper-tail *p*-value of the observed overlap of GWAS SNPs.

Analysis of repeat content within DHSs

We scanned both hg19 and mm9 for repeats using the RepeatMasker program (*41*) (`http://repeatmasker.org`) using the default parameters except for the specification of the species (e.g., human or mouse). The RepeatMasker database version 2012-04-18 was obtained from RepBase (*42*).

Tissue-selective mouse DHSs conserved in human

We calculated the maximum density (tags per 150 bp, tiled at 20 bp genome-wide and normalized for sequencing depth) of DNase I cleavage within each DHS master peak across all 45 mouse tissues/experiments. To limit the effects of outliers, we normalized each value to 90% of the maximum (across all experiments) at each DHS peak. We then performed *k*-means clustering on a matrix with the columns comprising mouse DNase I experiments and the rows corresponding to individual DHSs contained within the master DHS peak list. The clustering was performed using an efficient implementation of the k-means++ algorithm (*45*) (GraphLab API, `http://graphlab.org/toolkits/clustering/`) setting $k = 45$.

TF recognition sequence predictions

Transcription factor binding sites identified by scanning the entire genome for consensus sequences using the FIMO (*46*) tool from the MEME Suite (*47*) (version 4.6). A 5th order Markov model was generated from 36 bp mappable genome sequence and used as the background model. Motif models were curated from TRANSFAC (*48*) (version 11), JASPAR (*49*), and a SELEX-derived set from (*50*). Putative binding sites with a FIMO $P < 10^{-4}$ were retained.

Grouping TF recognition sequence models by similarity

Motif models used for the genome-wide scans were compared pairwise using the software TOMTOM (*51*) tool from the MEME Suite (*47*) (version 4.6) with the following parameters: "`-dist kullback -query-pseudo 0.1 -target-pseudo 0.1 -text -min-overlap 0 -thresh 1`".

The same 5th order Markov model background model as the FIMO genome-wide scans was used. The resulting pairwise comparisons were hierarchically clustered using Pearson correlation as the distance metric and complete linkage. Clusters were selected by cutting the tree at a height of 0.1.

TF recognition sequence enrichment within tissue-selective DHSs

We computed the number of DHSs containing an instance of each motif model. Using the cumulative hypergeometric distribution, a *p*-value for the number of observed DHSs containing a TF recognition sequence within a particular cluster of DHSs with respect the overall prevalence of the recognition sequence within all mouse DHSs conserved in human. The *p*-values were thresholded using the Bonferroni correction method.

Gene ontology analysis of tissue-selective DHSs

DHSs within lineage-specific clusters were supplied as input to GREAT (*52*). The analysis was run in the "basal plus extension" configuration, such that proximal regions were defined as 5 kb upstream and 1 kb downstream and distal regions were limited to 1 Mb.

Analysis of conserved DHS landscape usage mouse and human tissues

To compare the usage of the conserved DHS landscape in between two tissues we computed the Jaccard index. The Jaccard index is defined as $\frac{A \cap B}{A \cup B}$, where A and B are the number DHSs active in a mouse or human tissue. To compare multiple independent samples of similar tissues (as in **fig. S12A**) we used the median Jaccard index of all pairwise combinations.

Comparison of tissue activity of mouse DHSs conserved in human

For each shared DHS peak, tissue activity was defined as whether the DHS peak was identified via DNase I assay within that tissue (binarization of DHS signal into "accessible" or "inaccessible"). We associated tissues/cell types/experiments into tissue categories and took the union of all accessible DHSs within the datasets comprising the group. As a control, we compared the conservation of tissue activity against human DHSs with shuffled tissue activity profiles by sampling DHS from all shared human DHS without replacement keeping the number sampled constant with the number of active DHSs in each tissue.

TF recognition sequence conservation

To identify positionally conserved transcription factor recognition sequences, predicted

transcription factor binding sites identified with FIMO (see *TF recognition sequence predictions*) within all mouse DHSs were aligned to the human genome, using the "over chain" pairwise alignment downloaded from UCSC Genome Browser (*34*). We then obtained the human coordinates for the aligned mouse binding sites and overlapped them with predicted transcription factor binding sites in human. Importantly, both mouse and human genomes were scanned using the same motif models. A relaxed threshold (FIMO $P < 10^{-4}$) was used for human genome scans. A motif was labeled as conserved if the same motif was identified in mouse and human and the human motif matched the exact coordinates as the aligned mouse motif.

Positional and operational conservation of TF recognition sequences

Functionally conserved TF recognition sequences was identified by first filtering all DHSs with a one or more positionally conserved TF recognition sequence (see above). We then searched for shared DHSs that contained independent instances in mouse and human of a motif that corresponded to the recognition sequence for the same transcription factor.

TF recognition sequence turnover at repurposed DHSs

To assess the dynamics of TF recognition sequence evolution within shared DHSs with respect to conservation of tissue-specific DNase I accessibility we partitioned DHSs active in each mouse tissue, for both mouse and human, into two groups: conserved and repurposed. For each of these groups, we then computed the proportion of DHSs containing an instance of a motif (FIMO $P < 10^{-5}$). The assumption of this analysis is that the TF recognition sequence density should either be maintained (neutral loss or gain) or increase (net gain) when comparing mouse DHS with conserved accessibility in human vs. DHSs that have been repurposed human. For human, however, we expect to see a relative reduction in TF recognition sequences in a comparison of DHSs with conserved vs. repurposed accessibility. Tissue-selective TF recognition sequences were defined by the motifs enriched ($P < 10^{-5}$) within the tissue-specific clusters (see **fig. 2C**).

Comparison of *cis*-regulatory sequence content

We examined the *cis*-regulatory sequence content between mouse and human tissues by computing the proportion of all DHSs in a tissue (regardless of DHSs sequence or functional conservation) containing at least one instance of each TF recognition sequence (see *TF recognition sequence predictions*; FIMO $P < 10^{-5}$). Brain- and T cell-selective TF recognition sequences were defined by the motifs enriched ($P < 10^{-4}$) within the tissue-specific clusters (see **fig. 2C**).

To assess the similarity of each mouse tissue to all human tissues we computed the Euclidean distance pairwise between the proportions of DHSs contain each TF recognition sequences. When comparing a single tissue to multiple tissues, as in **fig. 5E**, we used the median Euclidean distance between all possible pairwise combinations.

References and Notes

27. S. John *et al.*, Genome-scale mapping of DNase I hypersensitivity. *Curr Protoc Mol Biol* **Chapter 27**, Unit 21.27 (2013).
28. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genom. Biol.* **10**, R25 (2009).
29. S. Neph *et al.*, BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
30. S. John *et al.*, Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
31. ENCODE Project Consortium *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
32. J. Harrow *et al.*, GENCODE: the reference human genome annotation for The ENCODE Project. *Genom. Res.* **22**, 1760–1774 (2012).
33. P. Flicek *et al.*, Ensembl 2013. *Nucleic Acids Res.* **41**, D48–55 (2013).
34. L. R. Meyer *et al.*, The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–9 (2013).
35. S. B. Hedges, The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838–849 (2002).
36. W. J. Murphy, T. H. Pringle, T. A. Crider, M. S. Springer, W. Miller, Using genomic data to unravel the root of the placental mammal phylogeny. *Genom. Res.* **17**, 413–421 (2007).
37. O. R. P. Bininda-Emonds *et al.*, The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
38. A. Siepel *et al.*, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genom. Res.* **15**, 1034–1050 (2005).
39. R. Drmanac *et al.*, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
40. B. Vernot *et al.*, Personal and population genomics of human regulatory variation. *Genom. Res.* **22**, 1689–1697 (2012).
41. A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0 Software Package (available at http://www.repeatmasker.org).
42. J. Jurka *et al.*, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
43. 1000 Genomes Project Consortium *et al.*, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

44. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

45. D. Arthur, S. Vassilvitskii, *k-means++: the advantages of careful seeding* (Society for Industrial and Applied Mathematics, 2007), pp. 1027–1035.

46. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

47. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8 (2009).

48. V. Matys *et al.*, TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–10 (2006).

49. J. C. Bryne *et al.*, JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36**, D102–6 (2008).

50. A. Jolma *et al.*, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).

51. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs. *Genom. Biol.* **8**, R24 (2007).

52. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

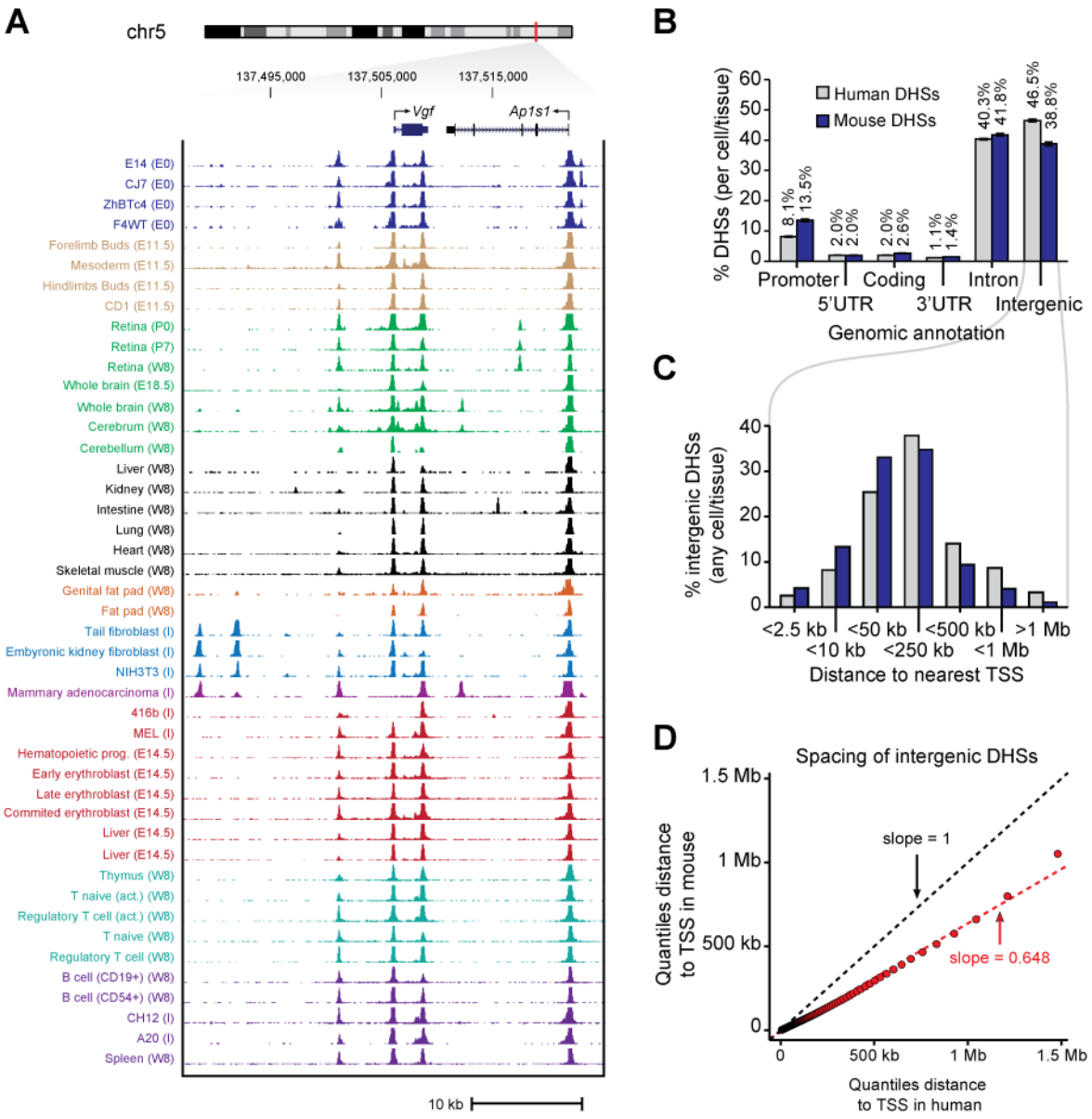

Figure S1. The accessible landscape of the mouse genome

(A) Exemplar DNase I cleavage profile within the ~35 kb surrounding the *Vgf* locus in 45 diverse mouse cell and tissue types. (B) Distribution (mean of tissues ± SEM) of mouse and human DHSs with respect to genome annotations (per cell- or tissue-type). Promoters are defined as 1 kb upstream of an annotated TSS.. (C) Distribution of distances between all intergenic DHSs identified to the nearest annotated TSS inboth human and mouse. (D) Q-Q plot of distances from intergenic DHSs to the nearest annotated TSS reveals a linear compression in regulatory DNA spacing. For clarity, the bottom and top 1% of the data points are not shown. Dashed black line, expected relationship between two identical distributions. Red line, linear regression of mouse vs. human distance quantiles.

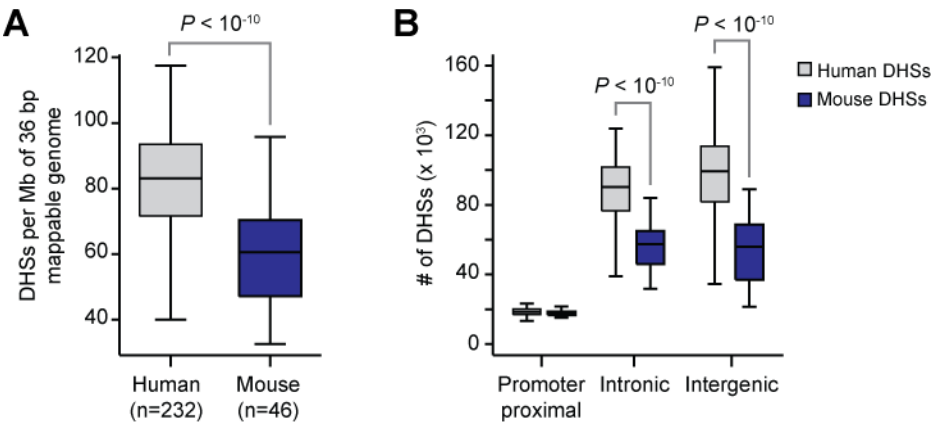

Figure S2. Expansion of the human *cis*-regulatory landscape

(A) Comparison of densities of the human and mouse DHS landscapes reveals a relative increase in human regulatory DNA ($P < 10^{-10}$, Wilcoxon rank-sum test). Densities were normalized by the size of the 36 bp alignable genome. (B) Genomic distribution of DHSs in mouse and human by annotation. Mouse and human have nearly equivalent amounts of promoter proximal DHSs, while human tissues have significant increase in the quantity of distal regulatory elements ($P < 10^{-10}$, Wilcoxon rank-sum test).

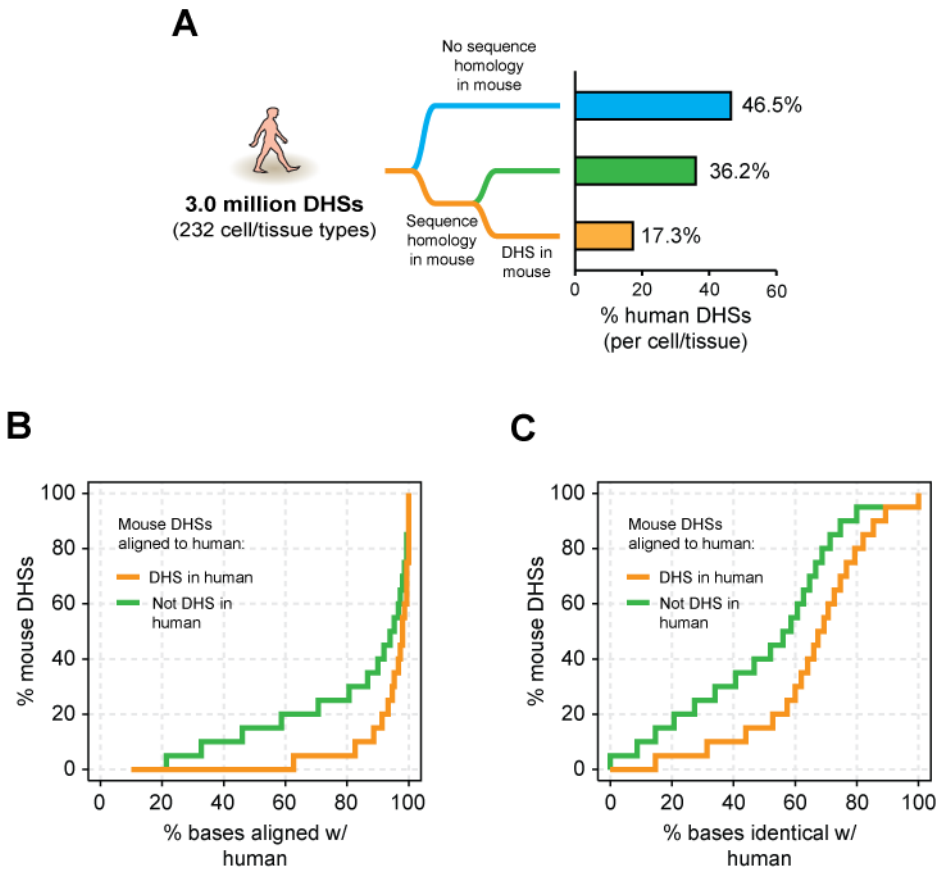

Figure S3. Sequence and functional alignment of mouse and human DHSs

(A) Proportions of the human regulatory DNA landscape with sequence homology and functional conservation in mouse. (B) Cumulative fraction of mouse DHSs as a function of the proportion of bases that align to the human using the mm9 to hg19 "over" chain downloaded from the UCSC Genome Browser. (C) Cumulative fraction of mouse DHSs as a function of the proportion of aligned nucleotides identical to human.

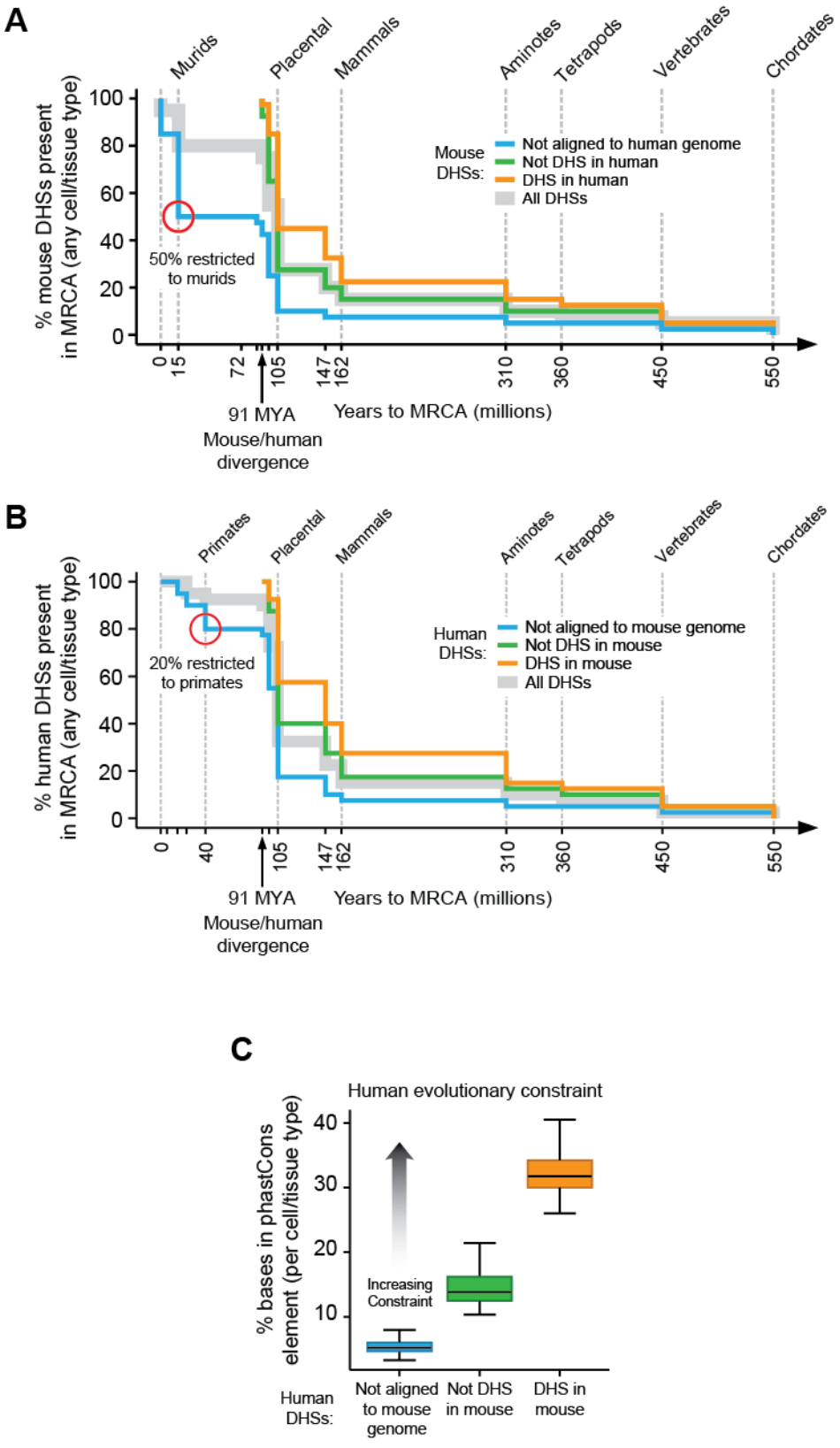
A
Murids
Placental
Mammals
Aminotes
Tetrapods
Vertebrates
Chordates
% mouse DHSs present in MRCA (any cell/tissue type)
100
80
60
40
20
0
Mouse DHSs:
Not aligned to human genome
Not DHS in human
DHS in human
All DHSs
50% restricted to murids
0
15
72
91
105
147
162
310
360
450
550
91 MYA
Mouse/human divergence
Years to MRCA (millions)
B
Primates
Placental
Mammals
Aminotes
Tetrapods
Vertebrates
Chordates
% human DHSs present in MRCA (any cell/tissue type)
100
80
60
40
20
0
Human DHSs:
Not aligned to mouse genome
Not DHS in mouse
DHS in mouse
All DHSs
20% restricted to primates
0
40
91
105
147
162
310
360
450
550
91 MYA
Mouse/human divergence
Years to MRCA (millions)
C
Human evolutionary constraint
% bases in phastCons element (per cell/tissue type)
40
30
20
10
Increasing Constraint
Human DHSs:
Not aligned to mouse genome
Not DHS in mouse
DHS in mouse

Figure S4. Evolutionary sequence constraint on mouse and human regulatory DNA

(A) Proportion of mouse DHSs with conserved sequence across the chordate phylogeny with respect to DHSs conservation with human. (B) Same as A for human DHSs. (C) Percentage of human DHSs containing a conserved phastCons element. Boxes indicate 25%- and 75%-iles for all human tissues. Whiskers denote 1.5× the interquartile range.

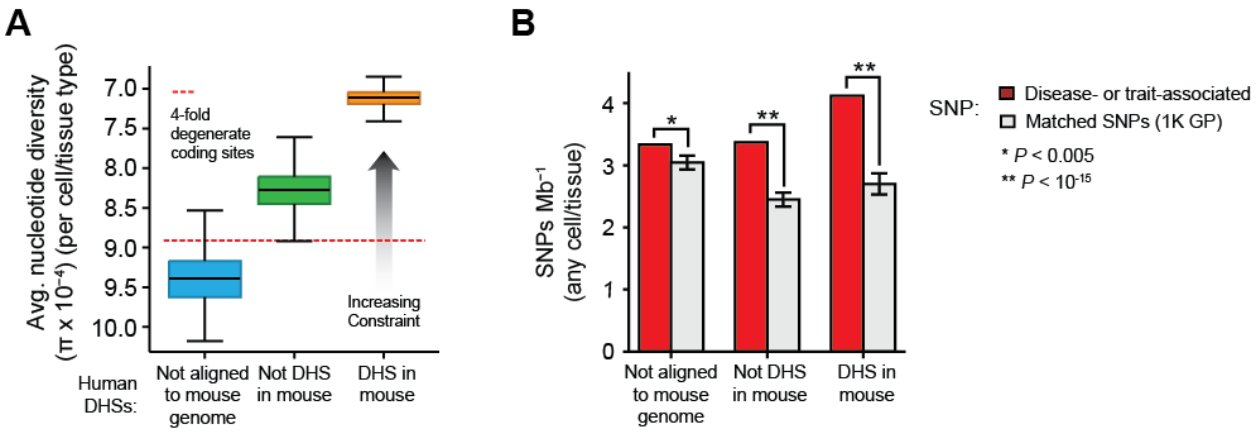

Figure S5. Recent sequence constraint on regulatory DNA

(A) Nucleotide diversity (π) with human DHSs estimated from the complete genomes of 53 unrelated individuals. Boxes indicate the 25%- and 75%-iles for all human tissues. Whiskers denote 1.5× the interquartile range. Red dashed line indicates π at four-fold degenerate coding sites. (B) Disease- and trait-associated variants are enriched within both conserved and species-specific regulatory elements. The observed variant density is compared against 1,000 sets of randomly matched SNPs.

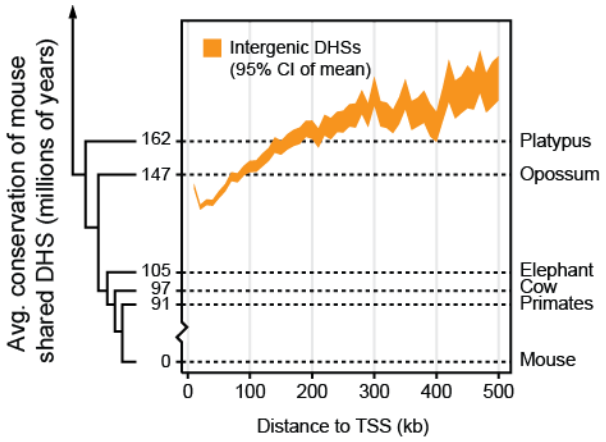

Figure S6. Increased conservation of far distal elements

The average conservation of intergenic shared DHSs increases with distance to the nearest TSS. The brown tree shows part of the mammalian phylogeny. Orange shading indicates the 95% percent confidence interval on the mean estimated by bootstrap analysis (500 replicates).

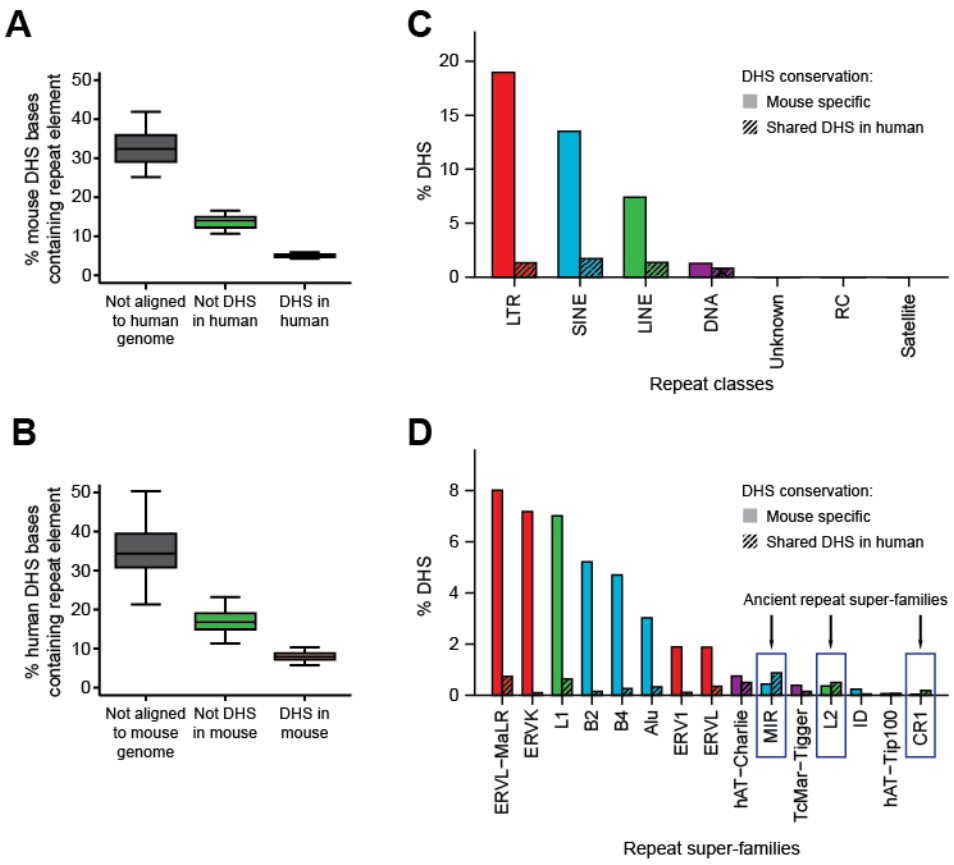

Figure S7. Repeat-associated innovation of species-specific regulatory DNA

(A-B) Proportion of the mouse (A) and human (B) DHS landscapes that overlap a repetitive element identified by RepeatMasker. (C) Mouse specific regulatory DNA is enriched for all classes repetitive elements. (D) LTR endogenous retroelements, L1 and SINE elements constitute the majority of mouse specific regulatory DNA that has arisen via repetitive elements. Enrichment of ancient repeat-families (blue boxes; SINE/MIR, LINE/L2 and LINE/CR1) within conserved DHSs indicate that repetitive elements contributed to innovations in mammalian regulatory DNA long before the mouse and human divergence ~91 MYA.

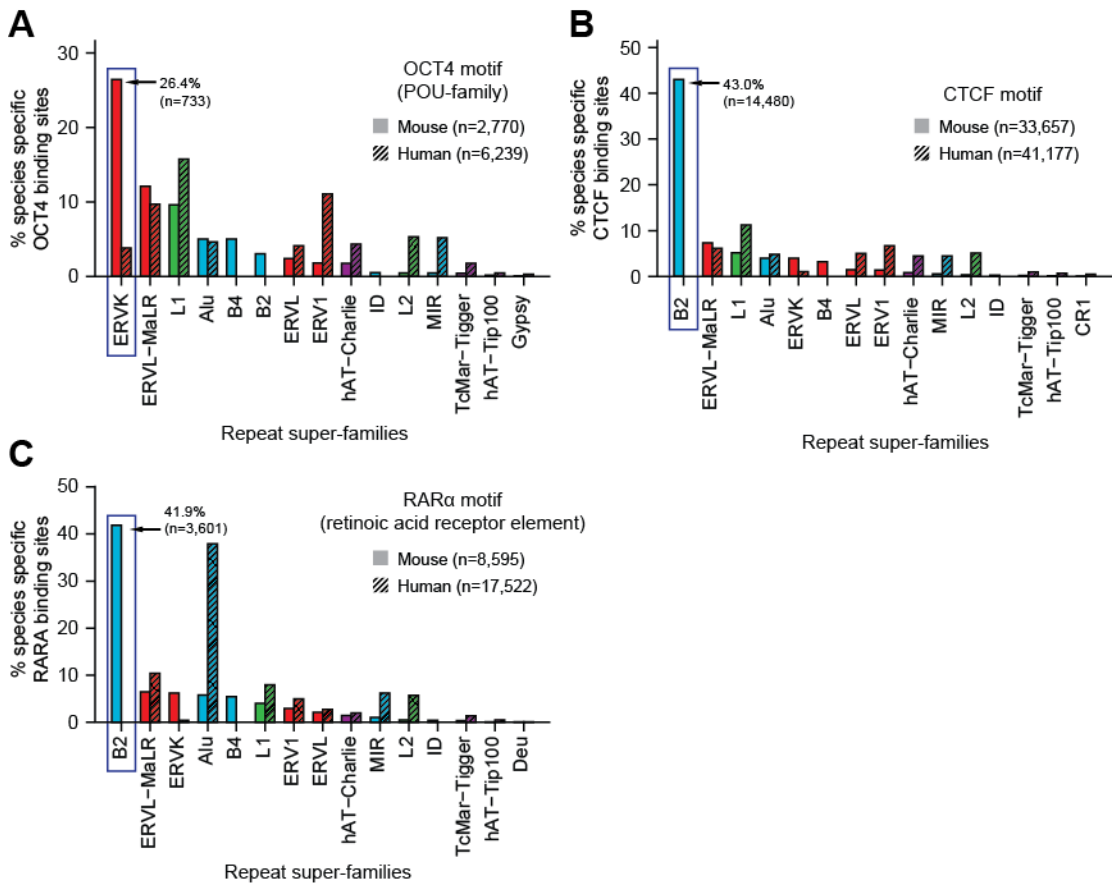

Figure S8. Stereotypical expansion of TF binding sites on specific repetitive elements

Examples of repeat mediated expansions of TF recognition sequences. (A) Many of the sequence elements recognized by the POU family of TFs that includes the canonical pluripotency master-regulator OCT4, has arisen on LTR/ERVL retrotransposable elements. (B) Enrichment of mouse specific CTCF recognition sequences within rodent specific SINE/B2 elements. (C) Retinoic acid receptor elements sequences (represented by RARα) have also expanded on SINE/B2 elements.

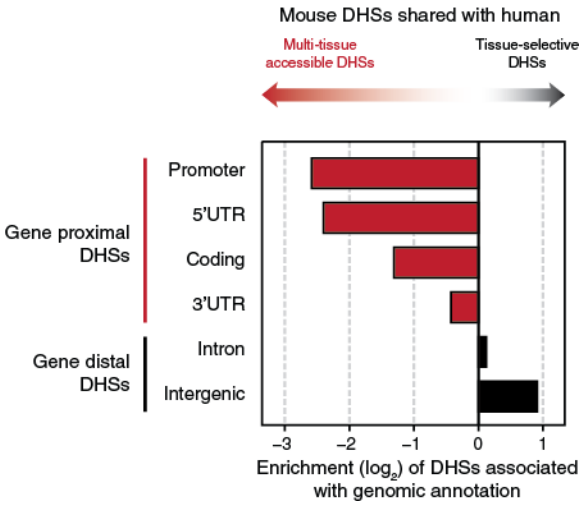

Figure S9. Tissue-selective DHSs are predominantly distal to genes

Enrichment (log_2) of tissue-selective shared DHSs categorized by genomic annotation.

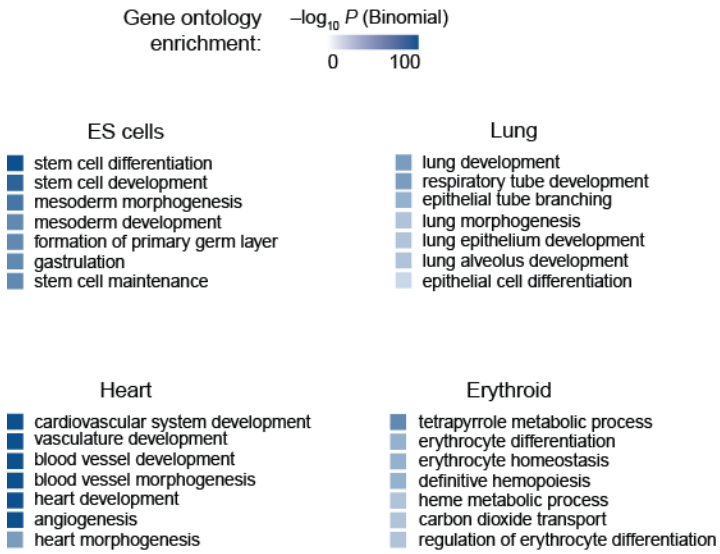

Figure S10. Shared DHSs localize to genes important in development and differentiation

Gene ontology terms associated with tissue-selective DHSs in mouse discovered using GREAT (*52*).

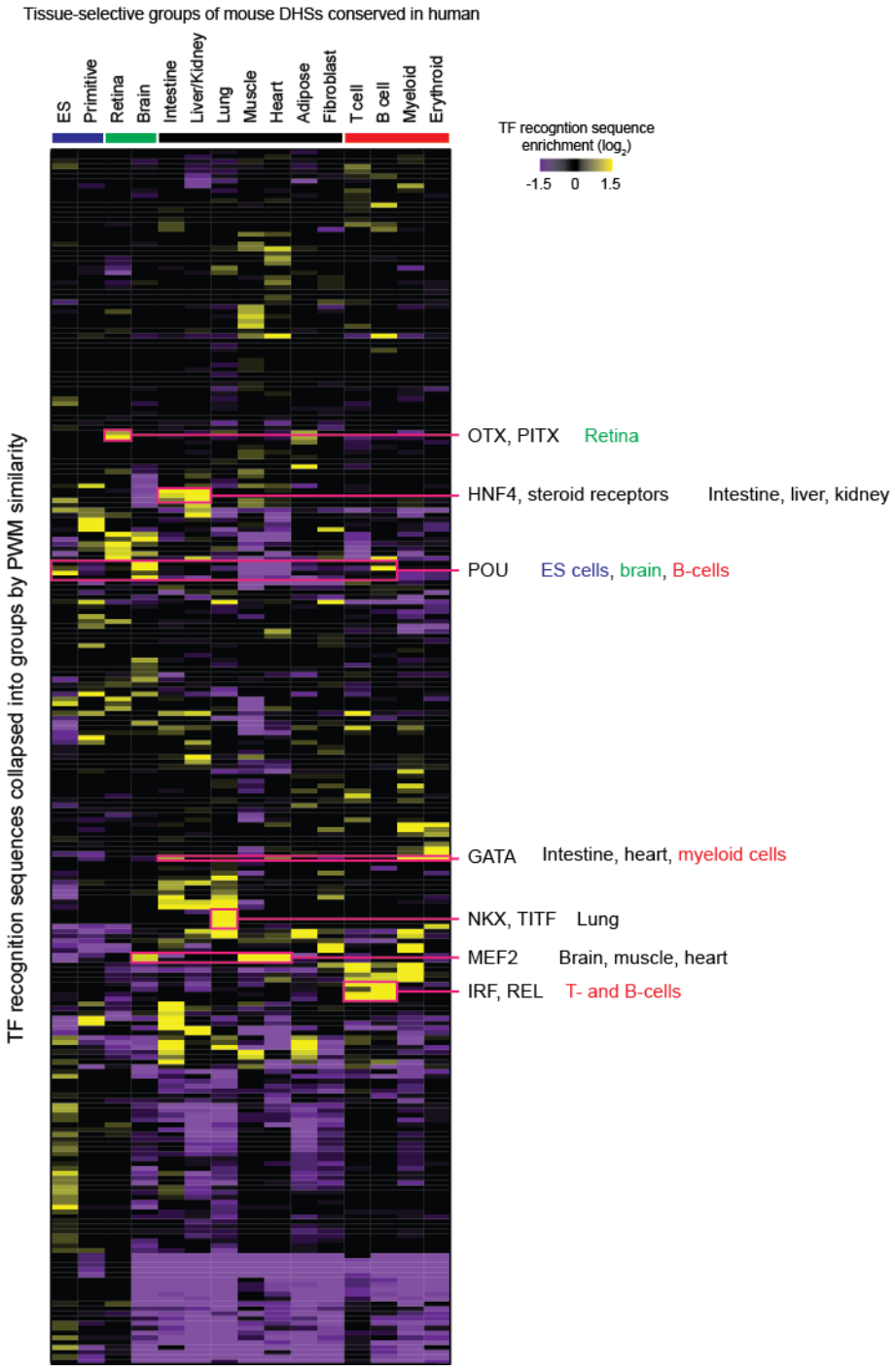

Figure S11. Content and localization of tissue-selective shared DHSs in mouse

Individual position weight matrices were clustered to form groups of transcription factors with similar recognition sequences. Each cell indicates the enrichment of individual PWM instances comprising the group (rows) in the tissue-selective mouse DHSs conserved in human (columns).

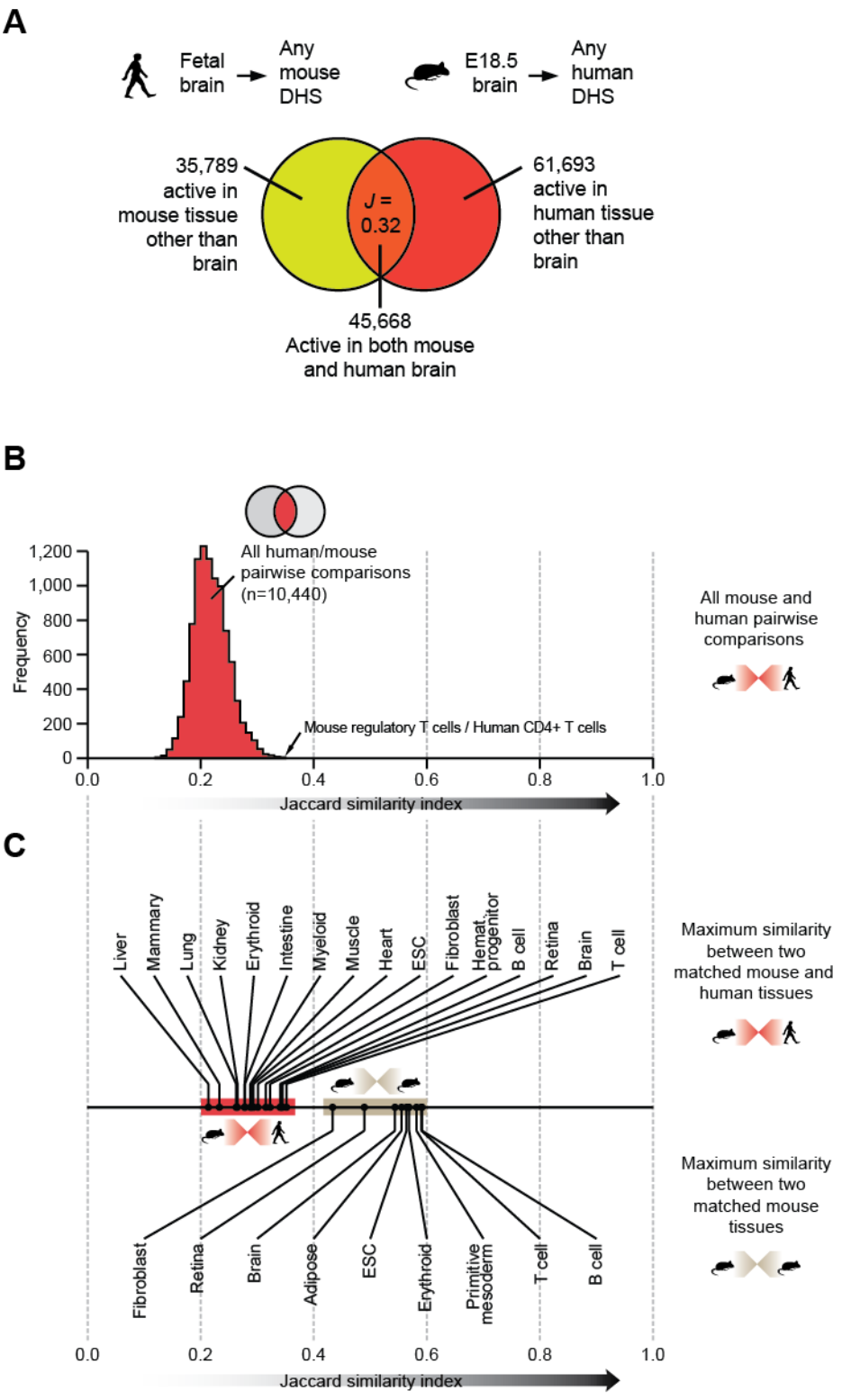

Figure S12. Conservation of accessibility within shared DHSs

(A) Usage of the shared DHS landscape in human fetal brain and mouse embryonic brain indicate

limited conservation in tissue-level accessibility patterns. (B) Histogram of all pairwise Jaccard similarity indices comparing the shared accessibility of DHSs common to both mouse and human tissues. (C) The maximum shared DHS landscape usage between matched mouse and human tissues (top, red) is markedly lower than a similar comparison between comparing mouse tissues with similar embryological origins (bottom, brown).

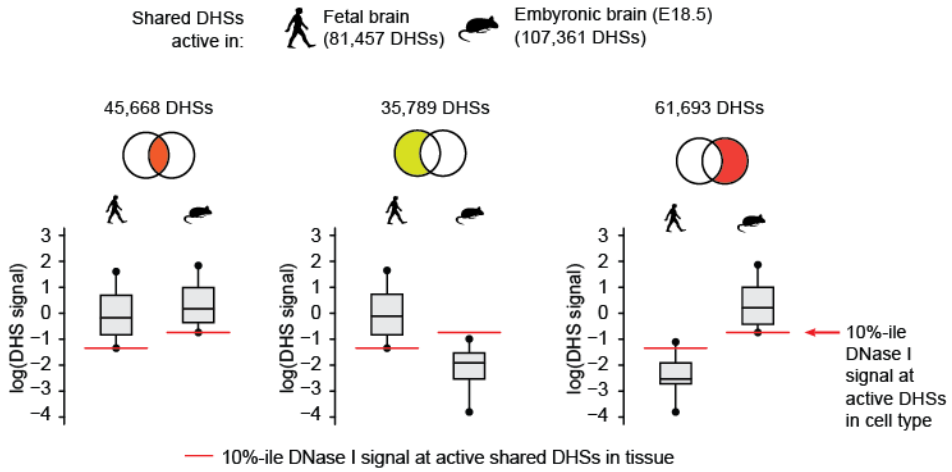

Figure S13. Identification of conserved and repurposed DHSs is robust to DNase I cleavage intensity

Box plots show the normalized DNase I cleavage intensity within mouse and human shared DHSs active in fetal and embryonic brain. Left, shared DHSs active in both mouse and human brain. Middle, shared DHSs active in human brain, but not active in mouse embryonic brain. Right, shared DHS active in mouse brain, but not active in human fetal brain. Boxes indicate the 25%- and 75%-iles and the whiskers denote the 10%- and 90%-iles. Red line indicates the DNase I cleavage intensity at the weakest 10% shared DHSs in each tissues.

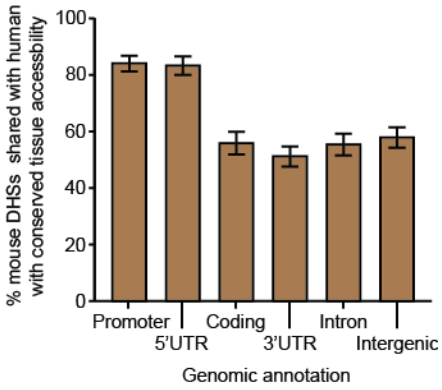

Figure S14. Genomic distribution of shared DHSs with conserved tissue accessibility

The proportion (mean of tissues ± SEM) of DHSs with conserved tissue accessibility within each genomic compartment. Promoters are defined as 1 kb upstream of an annotated TSS.

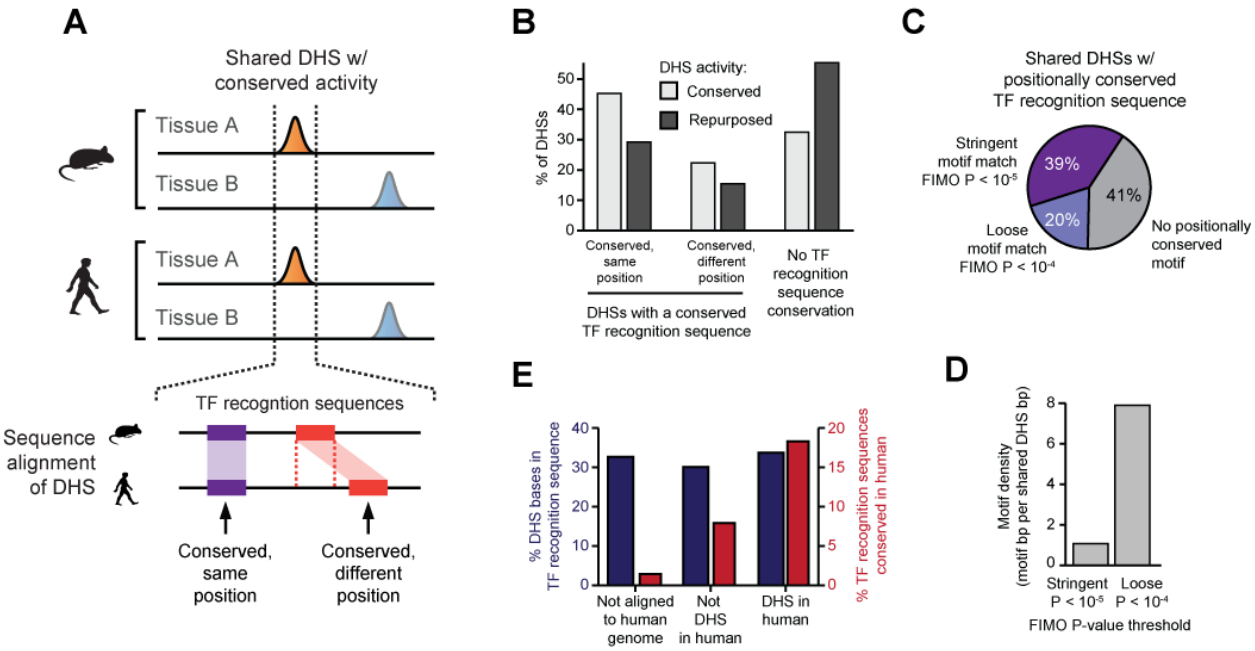

Figure S15. Conservation of transcription factor recognition sequences within DHS

(A) Model depicting the relationship of TF recognition sequences and conservation of accessibility at shared DHSs. DHSs with similar tissues activity patterns contain positionally and/or operationally conserved TF recognition sequences. (B) DHSs with conserved activity patterns are more likely to contain at least one conserved TF recognition sequences when compared to DHSs with repurposed activity (χ^2 test, *P*-value < 10^{-15}). Light grey, percentages of shared DHSs with conserved activity patterns containing conserved recognition elements. Dark gray, same for DHSs with repurposed activity patterns. (C–D) Effect of motif thresholds on the identification of conserved motifs. (C) Proportion of mouse DHSs shared with human that contain a positionally conserved transcription factor recognition sequence at stringent (purple) and relaxed (lavender) thresholds. (D) The number of putative TF recognition sequences increases drastically with reduced *P*-value thresholds. (E) The density (blue, % of DHS covered by TF recognition sequences) of TF recognition sequences within mouse DHSs is uniform across with respect to functional conservation, while the proportion of TF recognition sequence base-pairs conserved (red, % of TF recognition sequence bases conserved) is increased in mouse DHSs conserved in human.

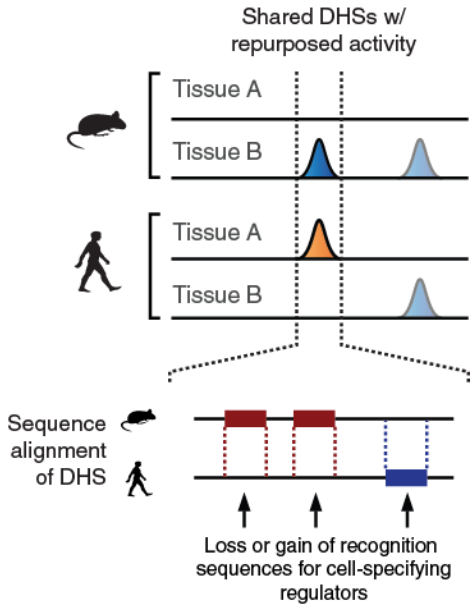

Figure S16. TF recognition sequence dynamics at repurposed DHSs

Model depicting the relationship of TF recognition sequence turnover and the repurposing of DHSs. The loss or gain of recognition sequences for cell-specifying TFs within individual DHSs underpin the cell and tissue activity patterns of regulatory DNA.

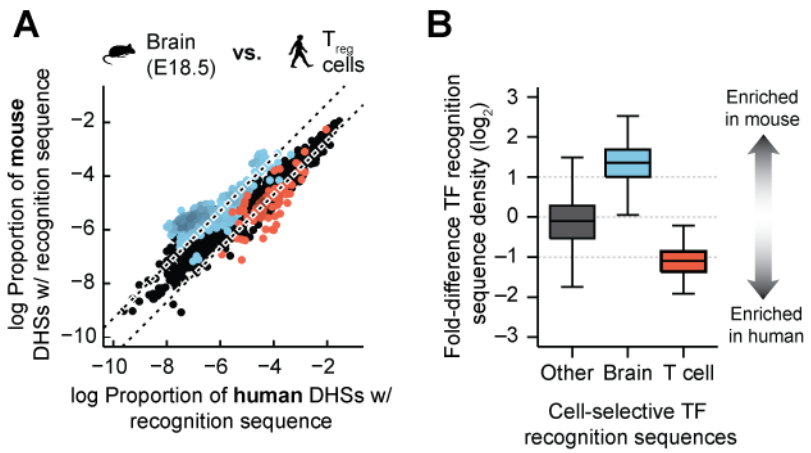

Figure S17. Comparison of TF recognition sequence densities between different mouse and human reveals cell-selective regulators

(A) Scatterplot of TF recognition sequence density in human fetal brain to mouse regulatory T cells. The recognition sequences enriched in brain or T cells (identified in **fig. 2C**) are colored blue and red, respectively. Dotted black lines demarcate a 2-fold difference in density between mouse and human. (B) The relative fold-difference (log_2) in density of TF recognition sequences. Boxes indicate the 25%- and 75%-ile and the whiskers denote 1.5× the interquartile range.

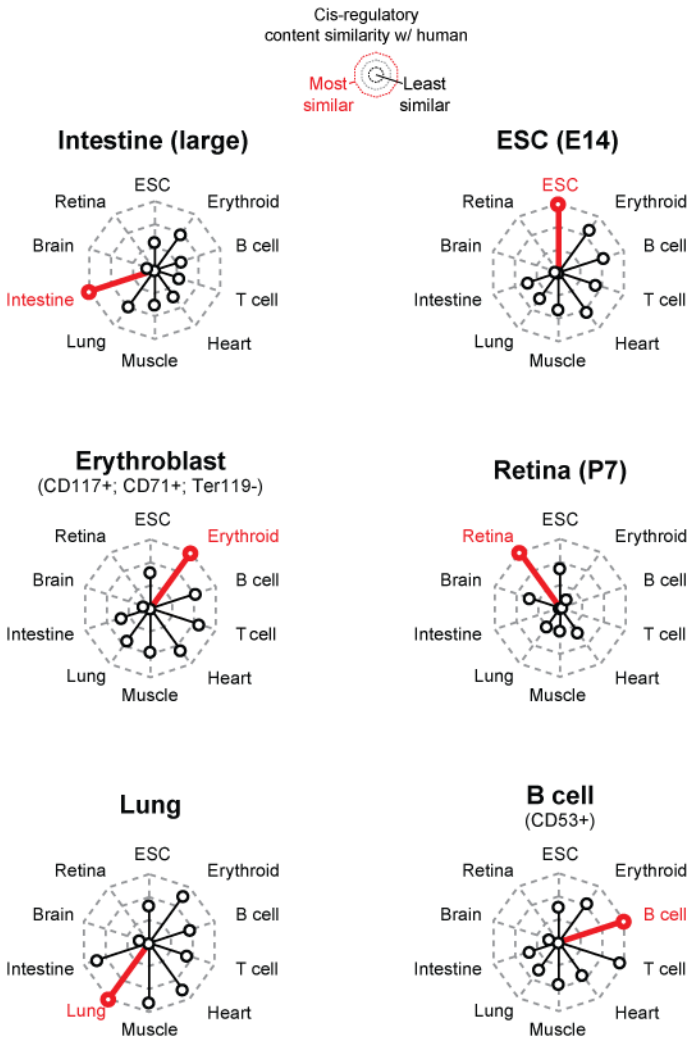

Figure S18. Conservation of *cis*-regulatory content between mouse and human tissues

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.

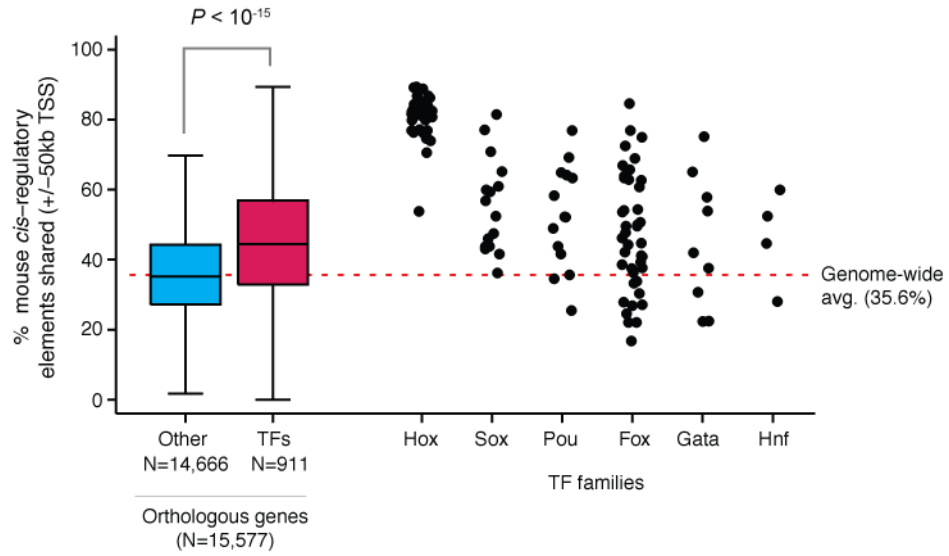

Figure S19. Preferential conservation of *cis*-regulatory elements surrounding transcription factor genes

Conservation of *cis*-regulatory elements ±50kb surrounding the transcription start sites of mouse genes with human orthologs. Genes that are not annotated as transcription factors (blue) have less cis-regulatory conservation than genes annotated as transcriptions factors (pink) (Wilcoxon rank-sum test, $P < 10^{-15}$). Stripcharts on the right show the *cis*-regulatory conservation surrounding TF genes grouped by canonical TF families. Dash red line indicates the genome-wide average (35.6%) DHS conservation between mouse and human.

Table S1. Mouse cell- and tissue-types used in this study

Overview of strain, developmental age and DHS mapping statistics for each mouse tissue used within this study.

Table S2. Overview of human cell- and tissue-types used in this study

Overview of developmental age and DHS mapping statistics for each human tissue used within this study.

Table S3. Conservation of mouse DHSs in human by cell- and tissue-type

Conservation statistics for DHSs identified within individual mouse tissues.

Table S4. Conservation of human DHSs in mouse by cell- and tissue-type

Conservation statistics for DHSs identified within individual human tissues.

Table S1

Cell- or tissue-type	Strain*	Age#	Tissue	GEO Accession	SPOT	Tags	Tags (-chrM)	DHS peaks (FDR 1%)	Downsampled DHS peaks (FDR 1%)
E14	129/Ola	E0	ES	GSM1014154	0.457	33,828,611	32,646,012	175,237	165,014
CJ7	129S1/SVimJ	E0	ES	GSM1014187	0.427	35,432,673	35,123,245	161,628	146,807
ZhBTc4	129/Ola	E0	ES	GSM1014169	0.558	30,861,073	29,216,912	164,855	158,561
ESC (F4WT)	WW6	E0	ES	GSM1014159	0.427	24,221,226	23,283,969	167,073	167,055
416b	B6D2F1/J	I	Myeloid	GSM1014163	0.538	42,933,235	40,389,545	141,483	129,350
MEL	Unknown	I	Erythroid	GSM1014191	0.502	28,778,970	28,632,367	129,572	125,720
Erythroblast (CD117+;CD71-;Ter119-)	CD-1	E14.5	Hem. Prog.	GSM1014155	0.435	26,102,618	26,038,902	147,180	146,269
Erythroblast (CD117+;CD71+;Ter119-)	CD-1	E14.5	Erythroid	GSM1014158	0.491	26,514,730	26,250,932	116,843	115,265
Erythroblast (CD117-;CD71+;Ter119+)	CD-1	E14.5	Erythroid	GSM1014156	0.324	39,265,301	39,135,425	74,521	68,271
Erythroblast (CD117+;CD71+;Ter119+)	CD-1	E14.5	Erythroid	GSM1014157	0.406	51,563,832	51,115,669	109,017	95,539
Naïve T cell (activated)	C57BL/6*	W8	T cell	GSM1014149	0.557	28,386,529	28,157,831	85,736	83,850
Regualtory T cell (activated)	C57BL/6*	W8	T cell	GSM1014200	0.543	32,305,988	32,264,326	100,369	97,162
Naïve T cell (resting)	C57BL/6*	W8	T cell	GSM1014192	0.519	33,637,621	33,535,765	100,535	93,602
Regulatory T cell (resting)	C57BL/6*	W8	T cell	GSM1014148	0.560	28,564,791	28,422,561	93,807	91,491
B cell (CD19+)	C57BL/6*	W8	B cell	GSM1014190	0.425	28,983,865	28,854,216	94,291	92,120
B cell (CD53+)	C57BL/6J	W8	B cell	GSM1014170	0.491	29,461,779	29,172,586	92,539	88,889
Liver	129*	E14.5	Erythroid	GSM1014162	0.367	29,791,189	29,512,936	102,545	97,849
Liver	C57BL/6J	E14.5	Erythroid	GSM1014183	0.351	44,019,764	43,575,379	112,444	97,047
Liver	C57BL/6J	W8	Liver	GSM1014195	0.423	19,812,754	19,748,170	104,995	104,979
A20 (lymphoma)	BALB/cAnN	I	B cell	GSM1014167	0.446	29,879,112	29,050,553	123,283	120,322
Mammary adenocarcinoma	RIII	I	Mammary	GSM1014196	0.686	32,345,915	32,111,316	91,073	84,282
Whole brain	C57BL/6J	E18.5	Brain	GSM1014197	0.483	25,257,085	25,249,691	167,715	166,503
Whole brain	C57BL/6J	W8	Brain	GSM1014151	0.689	36,089,477	36,079,464	224,978	204,809
Cerebrum	C57BL/6J	W8	Brain	GSM1014168	0.437	43,877,302	43,834,852	242,095	201,693
Cerebellum	C57BL/6J	W8	Brain	GSM1014164	0.397	21,727,916	21,690,965	106,514	106,499
Lung	C57BL/6J	W8	Lung	GSM1014194	0.423	21,041,166	20,978,206	163,431	163,463
Kidney	C57BL/6J	W8	Kidney	GSM1014193	0.451	46,472,928	46,357,946	184,503	160,029
Fat pad (mammary)	C57BL/6J	W8	Adipose	GSM1014165	0.355	19,628,835	19,566,613	139,661	139,658
Fibroblast (tail)	C57BL/6J	I	Fibroblast	GSM1014199	0.580	36,767,603	30,042,276	163,505	156,571
Fibroblast (NIH3T3)	NIH/Swiss	I	Fibroblast	GSM1014177	0.516	30,124,872	28,204,510	131,639	129,066
Fibroblast (embryonic kidney)	Spretus.BL6-Xist	I	Fibroblast	GSM1014171	0.463	25,485,737	22,217,276	137,147	137,138
Heart	C57BL/6J	W8	Heart	GSM1014166	0.433	28,947,564	28,077,411	146,214	141,807
Embyro (headless)	CD-1	E11.5	Prim. mesoderm	GSM1014172	0.402	29,943,418	29,909,171	149,434	138,370
Embryo (forelimb buds)	CD-1	E11.5	Prim. mesoderm	GSM1014174	0.458	26,816,011	26,734,717	150,619	147,330
Fat pad (genital)	C57BL/6J	W8	Adipose	GSM1014173	0.439	25,555,735	24,690,742	169,155	169,143
Embryo (hindlimb buds)	CD-1	E11.5	Prim. mesoderm	GSM1014179	0.389	30,266,505	30,132,401	157,008	144,567
Embryo (mesoderm)	CD-1	E11.5	Prim. mesoderm	GSM1014178	0.456	103,121,762	102,805,749	228,571	148,193
Intestine (large)	C57BL/6J	W8	Intestine	GSM1014186	0.398	30,356,956	29,995,670	131,435	129,237
Muscle (skeletal)	C57BL/6J	W8	Muscle	GSM1014189	0.389	102,881,281	97,647,043	193,949	126,944
Retina	C57BL/6J	W8	Retina	GSM1014175	0.323	28,638,758	28,252,986	100,570	97,088
Retina	C57BL/6J	D7	Retina	GSM1014198	0.418	27,024,790	26,973,855	109,073	108,605
Retina	C57BL/6J	D0	Retina	GSM1014188	0.484	36,128,496	35,959,145	134,712	123,526
Spleen	C57BL/6J	W8	B cell	GSM1014182	0.610	25,037,085	24,916,157	86,664	86,661
Thymus	C57BL/6J	W8	T cell	GSM1014185	0.495	26,399,759	26,286,891	105,449	103,977
CH12 (lymphoma)	B10.H-2aH-4bp/Wts	I	B cell	GSM1014153	0.536	68,983,868	68,760,538	171,305	141,262
							Total distinct peaks		1,334,703

* indicates that the substrain is not known
# E = days post-conception, D = days post-natal, W = weeks post-natal, I = immortalized or malignant

Table S2

Cell- or tissue-type	Source	Age#	Tissue	GEO Accession	SPOT	Tags	Tags (-chrM)	DHS peaks (FDR 1%)
A549	ENCODE		Lung	GSM736580	0.438	33,328,713	20,075,666	118,965
AG04449	ENCODE		Fibroblast	GSM736562	0.462	29,035,304	22,616,899	166,158
AG04450	ENCODE		Fibroblast	GSM736514	0.464	27,024,153	22,671,606	148,086
AG09309	ENCODE		Fibroblast	GSM736551	0.695	38,396,845	26,118,731	201,320
AG09319	ENCODE		Fibroblast	GSM736531	0.670	28,332,355	19,707,391	141,216
AG10803	ENCODE		Fibroblast	GSM736598	0.747	33,534,037	25,715,256	171,180
AoAF	ENCODE		Blood vessel	GSM736583	0.716	38,226,561	31,210,444	173,907
BE2_C	ENCODE		Brain	GSM736508	0.614	44,063,888	42,391,862	175,969
BJ	ENCODE		Fibroblast	GSM736518	0.749	42,763,886	24,605,012	162,086
CACO2	ENCODE		Intestine	GSM736500	0.707	27,117,636	25,576,071	122,479
CD14	ENCODE		Myeloid	GSM736513	0.429	67,698,560	67,035,406	175,178
CD19	REMC	y34	B cell	GSM701493	0.483	24,575,305	23,887,915	84,515
CD20	ENCODE		B cell	GSM1024765	0.572	36,983,818	36,413,583	104,139
CD34	REMC	y33	Hem. prog.	GSM530657	0.769	22,001,770	21,463,361	139,457
CD34	ENCODE		Hem. prog.	GSM1024770	0.691	49,756,223	48,561,124	164,050
CD4	REMC	y37	T cell	GSM701539	0.625	28,031,950	27,199,471	93,360
CD4pos_N	ENCODE		T cell	GSM1024789	0.356	24,083,134	22,482,769	82,384
CMK	ENCODE		Myeloid	GSM736607	0.572	31,319,078	21,490,670	133,032
fAdrenal	REMC	d96	Adrenal	GSM530653	0.440	27,428,606	27,094,195	188,072
fAdrenal	REMC	d85	Adrenal	GSM665799	0.352	33,450,853	32,872,463	181,935
fAdrenal	REMC	d108	Adrenal	GSM817167	0.313	28,128,420	27,626,714	136,447
fAdrenal	REMC	d101	Adrenal	GSM1027311	0.291	44,006,828	43,589,652	180,997
fAdrenal	REMC	d113	Adrenal	GSM878658	0.358	36,200,139	34,885,120	199,720
fBrain	REMC	d122	Brain	GSM530651	0.720	25,521,213	25,472,137	182,501
fBrain	REMC	d117	Brain	GSM595920	0.590	25,078,460	25,016,845	195,888
fBrain	REMC	d85	Brain	GSM595923	0.405	23,665,122	23,608,044	184,688
fBrain	REMC	d96	Brain	GSM595928	0.555	20,987,395	20,887,542	177,090
fBrain	REMC	d112	Brain	GSM665804	0.386	34,750,897	34,671,585	191,014
fBrain	REMC	d142	Brain	GSM665819	0.434	34,909,221	34,462,349	167,734
fBrain	REMC	d101	Brain	GSM878650	0.416	30,829,149	30,786,898	216,130
fBrain	REMC	d104	Brain	GSM878651	0.584	35,036,073	34,982,802	191,232
fBrain	REMC	d109	Brain	GSM878652	0.436	24,953,459	24,905,787	178,475
fBrain	REMC	d105	Brain	GSM1027328	0.490	35,021,221	34,978,828	204,455
fHeart	REMC	d96	Heart	GSM530654	0.580	25,623,855	25,253,759	173,135
fHeart	REMC	d101	Heart	GSM530661	0.540	32,812,462	26,756,542	209,039
fHeart	REMC	d117	Heart	GSM665809	0.574	20,497,225	20,237,506	157,105
fHeart	REMC	d103	Heart	GSM665814	0.560	24,283,747	23,937,451	172,946
fHeart	REMC	d147	Heart	GSM665824	0.412	34,249,454	33,857,207	187,545
fHeart	REMC	d110	Heart	GSM665830	0.649	36,948,524	36,362,253	189,064
fHeart	REMC	d105	Heart	GSM774203	0.600	57,235,297	56,289,045	220,074
fHeart	REMC	d120	Heart	GSM878630	0.596	39,316,934	38,082,008	217,926
fHeart	REMC	d91	Heart	GSM817220	0.585	34,557,454	33,668,945	202,827
fIntestine_Lg	REMC	d103	Intestine	GSM665815	0.343	25,940,579	25,753,257	148,014
fIntestine_Lg	REMC	d105	Intestine	GSM665818	0.367	26,525,135	26,385,840	165,904
fIntestine_Lg	REMC	d110	Intestine	GSM665826	0.391	36,983,856	36,509,010	174,469
fIntestine_Lg	REMC	d107	Intestine	GSM701490	0.357	21,461,793	21,194,682	165,770
fIntestine_Lg	REMC	d108	Intestine	GSM701495	0.439	27,932,860	27,773,196	84,894
fIntestine_Lg	REMC	d115	Intestine	GSM774213	0.391	22,437,785	22,188,540	147,422
fIntestine_Lg	REMC	d113	Intestine	GSM774214	0.296	53,885,997	53,430,571	174,813
fIntestine_Lg	REMC	d91	Intestine	GSM774220	0.439	21,201,047	21,106,664	154,797
fIntestine_Lg	REMC	d120	Intestine	GSM701531	0.349	29,832,797	29,469,112	166,626
fIntestine_Lg	REMC	d98	Intestine	GSM774228	0.453	81,556,735	80,494,968	286,072
fIntestine_Sm	REMC	d110	Intestine	GSM665825	0.412	36,648,756	36,283,133	182,941
fIntestine_Sm	REMC	d115	Intestine	GSM701487	0.406	19,998,953	19,734,999	136,731
fIntestine_Sm	REMC	d105	Intestine	GSM665835	0.406	40,799,365	40,095,822	193,223
fIntestine_Sm	REMC	d87	Intestine	GSM817161	0.428	29,166,499	28,923,298	172,118
fIntestine_Sm	REMC	d91	Intestine	GSM774205	0.476	22,479,708	22,244,548	173,146
fIntestine_Sm	REMC	d107	Intestine	GSM774210	0.401	23,494,853	23,287,680	162,640
fIntestine_Sm	REMC	d108	Intestine	GSM701496	0.402	24,154,098	23,959,814	149,240
fIntestine_Sm	REMC	d120	Intestine	GSM774225	0.343	63,431,414	63,044,896	189,437
fIntestine_Sm	REMC	d98	Intestine	GSM774229	0.417	35,105,452	34,694,378	183,531
fKidney	REMC	d122	Kidney	GSM530655	0.450	26,390,341	26,243,980	168,058
fKidney	REMC	d121	Kidney	GSM878666	0.380	24,583,989	24,482,089	185,281
fKidney	REMC	d105	Kidney	GSM1027329	0.452	23,680,869	23,598,291	207,896
fKidney	REMC	d85	Kidney	GSM1027342	0.585	35,548,444	35,467,952	240,229
fKidney_L	REMC	d147	Kidney	GSM665822	0.398	29,828,784	29,783,443	174,558
fKidney_L	REMC	d110	Kidney	GSM665829	0.356	35,168,628	35,102,532	192,878
fKidney_L	REMC	d115	Kidney	GSM665834	0.620	26,541,650	26,356,072	218,057
fKidney_L	REMC	d98	Kidney	GSM1027336	0.431	49,114,616	48,912,663	238,901
fKidney_R	REMC	d117	Kidney	GSM665810	0.360	35,268,222	35,173,250	194,782
fKidney_R	REMC	d107	Kidney	GSM817163	0.437	24,749,718	24,724,172	177,763
fKidney_R	REMC	d87	Kidney	GSM1027346	0.519	27,437,943	27,358,757	204,630
fKidney_renal_cortex	REMC	d108	Kidney	GSM701502	0.540	25,114,272	25,032,574	202,065
fKidney_renal_cortex	REMC	d113	Kidney	GSM701511	0.583	29,566,179	29,481,922	241,258
fKidney_renal_cortex	REMC	d120	Kidney	GSM701532	0.440	29,888,464	29,710,664	204,864
fKidney_renal_cortex	REMC	d97	Kidney	GSM878629	0.536	19,863,251	19,749,761	180,904
fKidney_renal_cortex	REMC	d96	Kidney	GSM1027316	0.519	27,413,186	27,173,032	199,449
fKidney_renal_cortex	REMC	d89	Kidney	GSM878667	0.491	34,431,625	34,140,860	227,044
fKidney_renal_pelvis	REMC	d91	Kidney	GSM774222	0.426	38,430,223	38,227,807	198,240
fKidney_renal_pelvis	REMC	d127	Kidney	GSM817177	0.302	31,071,466	30,965,110	175,309
fKidney_renal_pelvis	REMC	d103	Kidney	GSM878662	0.438	51,210,201	50,856,330	247,500
fLung	REMC	d122	Lung	GSM530656	0.470	25,021,977	24,742,635	157,966
fLung	REMC	d101	Lung	GSM530662	0.670	24,455,569	24,229,966	220,166
fLung	REMC	d103	Lung	GSM595916	0.400	41,845,529	41,604,307	196,107

Cell- or tissue-type	Source	Age#	Tissue	GEO Accession	SPOT	Tags	Tags (-chrM)	DHS peaks (FDR 1%)
fLung	REMC	d67	Lung	GSM595921	0.416	20,778,962	20,718,549	166,407
fLung	REMC	d85	Lung	GSM595924	0.662	24,489,273	24,450,572	205,880
fLung	REMC	d96	Lung	GSM595927	0.554	17,950,380	17,899,596	158,920
fLung	REMC	d112	Lung	GSM665805	0.471	36,058,035	35,938,309	199,936
fLung	REMC	d82	Lung	GSM665806	0.401	37,705,535	37,364,344	175,444
fLung_L	REMC	d110	Lung	GSM665828	0.496	38,447,839	38,319,871	186,870
fLung_L	REMC	d113	Lung	GSM701512	0.505	39,104,651	38,962,497	187,442
fLung_L	REMC	d108	Lung	GSM701524	0.581	21,555,838	21,499,339	195,089
fLung_L	REMC	d115	Lung	GSM774237	0.663	89,759,962	89,203,598	233,911
fLung_L	REMC	d87	Lung	GSM1027345	0.668	40,500,507	40,386,730	189,797
fLung_R	REMC	d117	Lung	GSM665807	0.410	28,258,863	28,164,047	166,666
fLung_R	REMC	d91	Lung	GSM774206	0.563	38,209,949	38,084,356	194,569
fLung_R	REMC	d107	Lung	GSM817164	0.552	30,722,966	30,611,703	181,390
fLung_R	REMC	d98	Lung	GSM774227	0.485	71,593,015	71,347,894	218,885
fLung_R	REMC	d105	Lung	GSM774231	0.546	53,568,139	53,408,427	203,664
fMuscle_arm	REMC	d115	Muscle	GSM701506	0.437	29,500,649	29,372,970	196,225
fMuscle_arm	REMC	d91	Muscle	GSM774223	0.478	90,984,367	90,581,737	324,206
fMuscle_back	REMC	d98	Muscle	GSM701536	0.517	29,018,101	28,803,299	216,313
fMuscle_back	REMC	d105	Muscle	GSM774235	0.478	89,523,543	89,154,405	243,666
fMuscle_back	REMC	d85	Muscle	GSM817217	0.500	31,601,635	31,512,300	207,352
fMuscle_back	REMC	d104	Muscle	GSM878639	0.472	71,300,684	66,777,257	248,101
fMuscle_leg	REMC	d127	Muscle	GSM774242	0.556	26,120,290	25,999,443	191,850
fMuscle_leg	REMC	d96	Muscle	GSM878626	0.509	41,739,601	41,562,780	213,588
fMuscle_leg	REMC	d97	Muscle	GSM817213	0.471	25,731,993	25,477,370	206,751
fMuscle_leg	REMC	d101	Muscle	GSM878631	0.488	35,011,994	34,889,731	207,372
fMuscle_leg	REMC	d113	Muscle	GSM878653	0.588	37,504,372	37,317,264	222,306
fMuscle_trunk	REMC	d120	Muscle	GSM701533	0.521	30,627,558	30,400,821	204,765
fMuscle_trunk	REMC	d121	Muscle	GSM878664	0.368	40,333,900	40,149,841	185,668
fOvary	REMC	pool	Ovary	GSM1027306	0.303	28,055,073	27,948,213	135,870
fPlacenta	REMC	d108	Placenta	GSM774215	0.320	74,216,981	73,680,118	233,243
fPlacenta	REMC	d91	Placenta	GSM774219	0.380	28,484,130	28,356,431	183,711
fPlacenta	REMC	d85	Placenta	GSM817219	0.523	32,115,085	31,947,808	198,726
fPlacenta	REMC	d113	Placenta	GSM878659	0.430	28,287,203	28,136,595	184,044
fPlacenta	REMC	d105	Placenta	GSM1027332	0.566	41,258,148	40,911,915	213,832
fRetina	REMC	d125	Retina	submitted	0.700	43,144,177	43,011,675	184,852
fRetina	REMC	d87	Retina	submitted	0.490	35,442,562	34,942,536	209,105
fRetina	REMC	d103	Retina	submitted	0.410	46,845,714	46,713,047	343,803
fSpinal_cord	REMC	d105	Spinal cord	GSM817189	0.324	41,246,058	41,088,815	227,684
fSpinal_cord	REMC	d96	Spinal cord	GSM1027308	0.307	32,060,944	31,943,861	170,412
fSpinal_cord	REMC	d113	Spinal cord	GSM878661	0.405	39,555,509	39,344,297	205,017
fSpinal_cord	REMC	d89	Spinal cord	GSM878663	0.371	35,237,430	34,879,741	191,578
fSpinal_cord	REMC	d87	Spinal cord	GSM1027339	0.367	22,834,430	22,683,314	167,471
fSpleen	REMC	d112	B cell	GSM701509	0.324	28,603,001	28,216,178	172,219
fStomach	REMC	d147	Stomach	GSM774202	0.325	29,750,284	29,366,538	157,567
fStomach	REMC	d107	Stomach	GSM774212	0.357	25,534,879	24,955,355	164,391
fStomach	REMC	d91	Stomach	GSM701528	0.269	21,448,212	21,054,457	131,796
fStomach	REMC	d98	Stomach	GSM701538	0.427	30,676,457	30,139,257	202,942
fStomach	REMC	d105	Stomach	GSM774232	0.346	28,486,554	28,351,708	168,103
fStomach	REMC	d127	Stomach	GSM817173	0.277	30,624,637	30,374,867	146,317
fStomach	REMC	d101	Stomach	GSM817199	0.286	35,346,480	34,995,534	163,730
fStomach	REMC	d96	Stomach	GSM1027318	0.301	57,000,342	56,662,485	185,823
fStomach	REMC	d108	Stomach	GSM878660	0.368	30,410,923	30,126,473	177,326
fStomach	REMC	d121	Stomach	GSM878665	0.387	30,430,524	30,060,424	180,573
fThymus	REMC	d147	T cell	GSM665823	0.315	34,790,401	34,559,542	130,756
fThymus	REMC	d105	T cell	GSM774204	0.343	28,903,878	28,696,303	116,826
fThymus	REMC	d98	T cell	GSM774230	0.341	77,445,578	76,852,917	150,360
fThymus	REMC	d127	T cell	GSM817172	0.293	84,341,630	83,920,582	161,208
fThymus	REMC	d104	T cell	GSM1027313	0.336	41,442,960	41,201,419	129,323
fThymus	REMC	d108	T cell	GSM878656	0.381	30,036,394	29,784,612	113,568
fThymus	REMC	d113	T cell	GSM878657	0.396	28,428,579	28,166,498	120,968
GM04503D	ENCODE		Fibroblast	GSM1024777	0.646	42,011,708	34,097,581	200,043
GM04504A	ENCODE		Fibroblast	GSM1024775	0.751	35,761,055	31,272,576	190,772
GM06990	ENCODE		B cell	GSM736558	0.546	22,440,189	19,729,256	92,709
GM12864	ENCODE		B cell	GSM736525	0.475	29,163,780	22,094,200	137,656
GM12865	ENCODE		B cell	GSM736512	0.525	37,660,121	34,216,112	143,716
GM12878	ENCODE		B cell	GSM736496	0.500	26,277,477	22,759,410	117,684
H7_hESC_T14	ENCODE		ES	GSM736638	0.372	33,507,208	26,763,822	140,102
H7_hESC_T2	ENCODE		ES	GSM736638	0.286	24,392,212	24,354,266	156,697
H7_hESC_T5	ENCODE		ES	GSM736638	0.343	44,134,819	42,890,856	211,653
HAc	ENCODE		Brain	GSM736586	0.422	45,892,152	40,763,545	180,083
HAEpiC	ENCODE		Placenta	GSM736631	0.764	31,211,767	29,842,070	205,033
HAh	ENCODE		Brain	GSM736594	0.485	32,927,492	27,040,382	200,014
HAsp	ENCODE		Spinal cord	GSM736537	0.425	39,414,103	34,517,745	194,537
HBMEC	ENCODE		Blood vessel	GSM736509	0.543	45,157,036	35,967,071	199,815
HBVP	ENCODE		Blood vessel	GSM1024750	0.359	47,747,677	45,100,866	209,369
HBVSMC	ENCODE		Blood vessel	GSM1024768	0.355	26,910,879	25,087,761	160,148
HCFaa	ENCODE		Heart	GSM736494	0.518	32,427,458	29,403,643	184,440
HCF	ENCODE		Heart	GSM736568	0.688	35,153,888	27,752,878	174,667
HCM	ENCODE		Heart	GSM736516	0.721	38,668,400	30,918,209	193,375
HConF	ENCODE		Fibroblast	GSM736547	0.506	36,022,255	32,701,475	153,668
HCPEpiC	ENCODE		Brain	GSM736569	0.742	26,199,536	24,538,293	210,380
HCT116	ENCODE		Intestine	GSM736600	0.454	39,022,377	26,710,185	114,060
HEEpiC	ENCODE		Esophagael	GSM736585	0.569	39,750,750	31,034,571	209,838
Hela	ENCODE		Cervical	GSM736564	0.579	36,519,330	26,208,222	123,470
HEPG2	ENCODE		Liver	GSM736637	0.570	22,126,494	18,183,774	90,775
HESC	ENCODE		ES	GSM736582	0.358	24,431,583	20,890,242	150,729
hESCT0	ENCODE		ES	GSM736638	0.635	33,752,751	27,784,900	266,618
HFF	ENCODE		Fibroblast	GSM736602	0.545	46,526,167	37,514,733	192,282
HFF_MyC	ENCODE		Fibroblast	GSM736524	0.484	39,972,176	29,812,887	209,807

Cell- or tissue-type	Source	Age#	Tissue	GEO Accession	SPOT	Tags	Tags (-chrM)	DHS peaks (FDR 1%)
HGF	ENCODE		Fibroblast	GSM736579	0.483	30,751,179	30,148,924	145,887
HIPEpiC	ENCODE		Iris	GSM736589	0.560	32,033,391	26,191,760	225,744
HL60	ENCODE		Myeloid	GSM736626	0.589	33,148,093	32,066,521	161,716
HMEC	ENCODE		Mammary	GSM736634	0.425	43,782,139	31,914,218	140,574
HMF	ENCODE		Fibroblast	GSM736628	0.798	33,118,548	27,188,138	179,452
HMVEC_dAd	ENCODE		Blood vessel	GSM1024745	0.377	28,492,574	25,367,629	125,234
HMVEC_dBlAd	ENCODE		Blood vessel	GSM736609	0.726	46,050,351	42,643,211	162,593
HMVEC_dBlNeo	ENCODE		Blood vessel	GSM736571	0.529	54,866,074	49,657,528	168,436
HMVEC_dLyAd	ENCODE		Blood vessel	GSM736599	0.575	48,098,604	33,370,795	127,713
HMVEC_dLyNeo	ENCODE		Blood vessel	GSM736577	0.578	44,427,289	41,368,217	153,107
HMVEC_dNeo	ENCODE		Blood vessel	GSM736611	0.586	41,076,964	29,813,741	141,037
HMVEC_LBl	ENCODE		Blood vessel	GSM736542	0.485	50,131,824	47,911,926	169,983
HMVEC_LLy	ENCODE		Blood vessel	GSM736507	0.605	52,381,779	41,515,731	144,886
HNPCEpiC	ENCODE		Blood vessel	GSM736621	0.605	30,449,370	20,884,189	212,433
HPAEC	ENCODE		Blood vessel	GSM1024763	0.298	33,475,651	32,804,748	123,918
HPAF	ENCODE		Blood vessel	GSM736555	0.716	54,924,185	47,644,186	188,071
HPdLF	ENCODE		Fibroblast	GSM736632	0.686	25,814,740	23,004,769	171,349
HPF	ENCODE		Lung	GSM736574	0.672	30,693,222	27,902,228	154,397
HRCE	ENCODE		Kidney	GSM736549	0.657	26,420,059	24,358,945	192,147
HRE	ENCODE		Kidney	GSM736527	0.534	25,850,984	22,900,558	187,131
HRGEC	ENCODE		Kidney	GSM736499	0.424	20,987,760	19,623,147	141,865
HRPEpiC	ENCODE		Retina	GSM736630	0.741	38,479,119	27,376,806	227,086
HSMM_D	ENCODE		Muscle	GSM736530	0.498	34,007,549	28,658,027	221,749
HSMM	ENCODE		Muscle	GSM736560	0.637	44,844,225	37,479,300	228,282
hTH17	ENCODE		T cell	GSM1024790	0.268	18,274,006	17,865,182	77,514
hTH1	ENCODE		T cell	GSM1024753	0.333	20,742,528	20,459,985	82,245
hTH2	ENCODE		T cell	GSM1024739	0.689	43,122,486	42,711,501	128,977
hTR	ENCODE		T cell	GSM1024744	0.504	40,015,882	37,326,516	125,848
HUVEC	ENCODE		Blood vessel	GSM736575	0.401	23,090,713	20,957,203	119,094
HVMF	ENCODE		Connective	GSM736534	0.591	22,052,455	19,811,379	170,340
Jurkat	ENCODE		T cell	GSM736501	0.497	67,808,791	62,773,068	159,613
K562	ENCODE		Erythroid	GSM736629	0.542	35,811,565	22,416,086	142,986
LHCN_M2_D4	ENCODE		Muscle	GSM1024787	0.727	64,634,390	37,759,025	218,246
LHCN_M2	ENCODE		Muscle	GSM1024787	0.709	37,903,192	30,643,938	192,347
LNCap	ENCODE		Prostate	GSM736565	0.620	49,624,105	28,747,933	183,224
M059J	ENCODE		Brain	GSM1024773	0.690	61,673,188	56,087,660	220,835
MCF7	ENCODE		Mammary	GSM736581	0.683	44,891,456	32,444,710	207,878
MCF7	ENCODE		Mammary	GSM736581	0.437	29,211,804	22,787,939	126,717
MCF7_ER	ENCODE		Mammary	GSM736581	0.652	63,497,188	45,615,312	220,065
NB4	ENCODE		Myeloid	GSM736604	0.531	36,021,761	32,684,492	143,838
NHA	ENCODE		Brain	GSM736544	0.562	40,038,155	34,637,438	191,510
NHBE_RA	ENCODE		Broncheal	GSM1024781	0.344	41,688,982	29,967,781	149,972
NHDF_Ad	ENCODE		Fibroblast	GSM736567	0.805	44,829,222	44,170,209	230,696
NHDF_Neo	ENCODE		Fibroblast	GSM736498	0.698	38,935,890	33,799,898	187,962
NHEK	ENCODE		Skin	GSM736545	0.357	27,995,767	24,307,712	145,203
NHLF	ENCODE		Lung	GSM736612	0.706	41,241,537	33,520,502	206,254
NT2_D1	ENCODE		ES	GSM1024751	0.351	36,885,055	33,804,957	184,238
PANC1	ENCODE		Pancreas	GSM736517	0.418	26,460,252	22,683,675	116,642
PrEC	ENCODE		Prostate	GSM1024742	0.323	35,291,903	32,675,776	167,623
RPMI_7951	ENCODE		Skin	GSM1024779	0.682	38,648,543	34,251,134	167,310
RPTEC	ENCODE		Kidney	GSM736543	0.487	35,166,095	22,996,745	169,261
SAEC	ENCODE		Broncheal	GSM736608	0.620	24,221,968	22,257,395	198,442
SKMC	ENCODE		Muscle	GSM736593	0.801	28,517,820	28,117,588	205,493
SK_N_MC	ENCODE		Brain	GSM736522	0.353	26,552,446	24,428,899	146,328
SKNSH	ENCODE		Brain	GSM736559	0.622	19,593,615	18,614,810	89,968
T_47D	ENCODE		Mammary	GSM1024762	0.584	55,382,458	41,173,122	153,537
WERI_Rb1	ENCODE		Retina	GSM736495	0.546	35,473,957	33,340,703	191,374
WI_38	ENCODE		Lung	GSM736613	0.700	26,765,352	20,405,387	166,381
WI_38_TAM	ENCODE		Lung	GSM736613	0.622	28,554,163	24,833,322	205,334
						Total distinct peaks		3,022,372

y = years post-natal, d = days post-conception, blank = primary, cultured or malignant cells

Table S3

Cell- or tissue-type	Age*	Not aligned # peaks	Not aligned % peaks	Not DHS in human # peaks	Not DHS in human % peaks	DHS in human # peaks	DHS in human % peaks	Total#
E14	E0	80,583	45.5%	26,273	14.8%	70,120	39.6%	176,976
CJ7	E0	75,270	47.4%	22,106	13.9%	61,331	38.6%	158,707
ZhBTc4	E0	77,919	45.8%	25,009	14.7%	67,164	39.5%	170,092
ESC (F4WT)	E0	82,365	46.1%	26,526	14.8%	69,807	39.1%	178,698
416b	I	58,933	42.7%	22,653	16.4%	56,363	40.9%	137,949
MEL	I	60,654	45.3%	21,290	15.9%	52,055	38.8%	133,999
Erythroblast (CD117+;CD71-;Ter119-)	E14.5	61,617	38.3%	24,796	15.4%	74,311	46.2%	160,724
Erythroblast (CD117+;CD71+;Ter119-)	E14.5	52,228	41.9%	18,410	14.8%	53,914	43.3%	124,552
Erythroblast (CD117-;CD71+;Ter119+)	E14.5	29,920	40.2%	10,006	13.5%	34,460	46.3%	74,386
Erythroblast (CD117+;CD71+;Ter119+)	E14.5	42,430	40.7%	14,691	14.1%	47,232	45.3%	104,353
Naïve T cell (activated)	W8	34,496	37.2%	12,749	13.7%	45,558	49.1%	92,803
Regualtory T cell (activated)	W8	41,758	39.4%	15,106	14.2%	49,249	46.4%	106,113
Naïve T cell (resting)	W8	40,697	39.7%	15,604	15.2%	46,262	45.1%	102,563
Regulatory T cell (resting)	W8	38,518	38.2%	13,688	13.6%	48,598	48.2%	100,804
B cell (CD19+)	W8	38,256	37.9%	14,083	13.9%	48,643	48.2%	100,982
B cell (CD53+)	W8	38,280	39.2%	13,590	13.9%	45,769	46.9%	97,639
Liver	E14.5	41,251	38.3%	15,036	13.9%	51,537	47.8%	107,824
Liver	E14.5	41,073	38.2%	14,842	13.8%	51,734	48.1%	107,649
Liver	W8	42,346	37.9%	20,548	18.4%	48,890	43.7%	111,784
A20 (lymphoma)	I	54,396	42.6%	20,896	16.4%	52,336	41.0%	127,628
Mammary adenocarcinoma	I	29,491	32.4%	11,840	13.0%	49,635	54.6%	90,966
Whole brain	E18.5	39,091	21.9%	31,667	17.8%	107,361	60.3%	178,119
Whole brain	W8	62,899	28.8%	43,684	20.0%	112,014	51.2%	218,597
Cerebrum	W8	60,346	28.2%	48,480	22.7%	105,164	49.1%	213,990
Cerebellum	W8	32,608	28.5%	18,803	16.4%	62,965	55.1%	114,376
Lung	W8	59,285	33.4%	32,120	18.1%	85,947	48.5%	177,352
Kidney	W8	59,542	34.8%	30,698	17.9%	80,980	47.3%	171,220
Fat pad (mammary)	W8	55,297	36.5%	28,524	18.8%	67,563	44.6%	151,384
Fibroblast (tail)	I	52,603	31.2%	29,090	17.3%	86,887	51.5%	168,580
Fibroblast (NIH3T3)	I	47,016	33.8%	21,600	15.5%	70,671	50.7%	139,287
Fibroblast (embryonic kidney)	I	47,897	32.4%	26,985	18.2%	73,010	49.4%	147,892
Heart	W8	45,619	29.5%	24,139	15.6%	85,013	54.9%	154,771
Embyro (headless)	E11.5	40,219	26.4%	21,259	13.9%	90,981	59.7%	152,459
Embryo (forelimb buds)	E11.5	45,168	28.1%	24,222	15.1%	91,401	56.8%	160,791
Fat pad (genital)	W8	67,057	37.1%	35,318	19.6%	78,197	43.3%	180,572
Embryo (hindlimb buds)	E11.5	44,267	28.1%	24,695	15.7%	88,798	56.3%	157,760
Embryo (mesoderm)	E11.5	43,181	26.6%	24,325	15.0%	94,953	58.4%	162,459
Intestine (large)	W8	53,391	38.6%	23,019	16.6%	62,061	44.8%	138,471
Muscle (skeletal)	W8	41,230	29.5%	20,894	15.0%	77,526	55.5%	139,650
Retina	W8	30,844	29.0%	14,902	14.0%	60,469	56.9%	106,215
Retina	D7	33,666	28.3%	15,969	13.4%	69,356	58.3%	118,991
Retina	D0	36,126	26.9%	19,810	14.7%	78,419	58.4%	134,355
Spleen	W8	36,278	38.1%	13,352	14.0%	45,499	47.8%	95,129
Thymus	W8	44,140	38.9%	16,902	14.9%	52,478	46.2%	113,520
CH12 (lymphoma)	I	63,397	42.1%	26,139	17.3%	61,138	40.6%	150,674

* E = days post-conception, D = days post-natal, W = weeks post-natal, I = immortalized or malignant
# Total peaks from master peaks list (see Supporting Methods)

Table S4

Cell- or tissue-type	Age*	Not aligned # peaks	Not aligned % peaks	Not DHS in mouse # peaks	Not DHS in mouse % peaks	DHS in mouse # peaks	DHS in mouse % peaks	Total#
A549		56,327	38.8%	36,445	25.1%	52,537	36.2%	145,309
AG04449		73,878	36.6%	58,713	29.1%	69,302	34.3%	201,893
AG04450		65,008	35.8%	49,633	27.3%	67,180	36.9%	181,821
AG09309		89,264	36.9%	70,819	29.3%	81,928	33.9%	242,011
AG09319		58,682	33.7%	47,490	27.3%	67,822	39.0%	173,994
AG10803		73,912	35.6%	58,341	28.1%	75,129	36.2%	207,382
AoAF		75,624	35.7%	58,858	27.8%	77,085	36.4%	211,567
BE2_C		76,347	36.9%	58,379	28.2%	72,261	34.9%	206,987
BJ		70,919	35.8%	58,337	29.4%	69,043	34.8%	198,299
CACO2		52,579	35.6%	41,067	27.8%	54,211	36.7%	147,857
CD14		86,352	43.0%	53,618	26.7%	61,053	30.4%	201,023
CD19	y34	37,781	35.9%	23,176	22.0%	44,230	42.0%	105,187
CD20		44,868	36.0%	27,592	22.1%	52,157	41.9%	124,617
CD34	y33	64,748	38.5%	41,169	24.5%	62,410	37.1%	168,327
CD34		75,036	38.6%	47,615	24.5%	71,881	37.0%	194,532
CD4	y37	40,534	35.3%	24,588	21.4%	49,685	43.3%	114,807
CD4pos_N		35,699	36.3%	22,677	23.0%	40,083	40.7%	98,459
CMK		66,174	41.8%	38,850	24.5%	53,261	33.6%	158,285
fAdrenal	d96	81,661	35.9%	64,442	28.3%	81,421	35.8%	227,524
fAdrenal	d85	77,024	34.2%	63,716	28.3%	84,713	37.6%	225,453
fAdrenal	d108	58,885	35.9%	43,321	26.4%	61,722	37.7%	163,928
fAdrenal	d101	76,831	35.4%	60,523	27.9%	79,806	36.7%	217,160
fAdrenal	d113	86,410	36.5%	69,264	29.2%	81,146	34.3%	236,820
fBrain	d122	65,968	30.6%	64,895	30.1%	84,531	39.2%	215,394
fBrain	d117	70,308	30.1%	71,318	30.6%	91,777	39.3%	233,403
fBrain	d85	63,723	28.1%	67,364	29.7%	95,625	42.2%	226,712
fBrain	d96	64,562	30.4%	62,892	29.6%	84,916	40.0%	212,370
fBrain	d112	68,107	29.1%	75,146	32.1%	90,532	38.7%	233,785
fBrain	d142	57,884	28.4%	63,803	31.3%	82,264	40.3%	203,951
fBrain	d101	74,876	29.0%	79,855	30.9%	103,390	40.1%	258,121
fBrain	d104	68,753	30.4%	66,989	29.6%	90,316	40.0%	226,058
fBrain	d109	61,976	29.2%	63,351	29.8%	87,251	41.0%	212,578
fBrain	d105	73,427	30.1%	73,343	30.1%	97,025	39.8%	243,795
fHeart	d96	70,589	33.0%	59,591	27.9%	83,492	39.1%	213,672
fHeart	d101	87,103	33.9%	75,635	29.4%	94,348	36.7%	257,086
fHeart	d117	63,750	32.2%	56,484	28.6%	77,564	39.2%	197,798
fHeart	d103	71,673	33.1%	65,427	30.2%	79,361	36.7%	216,461
fHeart	d147	77,034	33.2%	73,569	31.7%	81,337	35.1%	231,940
fHeart	d110	78,676	33.5%	67,543	28.8%	88,482	37.7%	234,701
fHeart	d105	90,590	34.2%	75,571	28.5%	98,821	37.3%	264,982
fHeart	d120	88,246	33.4%	77,863	29.5%	98,246	37.2%	264,355
fHeart	d91	83,176	33.7%	70,734	28.7%	92,972	37.7%	246,882
fIntestine_Lg	d103	61,072	32.2%	48,760	25.7%	79,643	42.0%	189,475
fIntestine_Lg	d105	75,869	36.6%	54,478	26.3%	76,969	37.1%	207,316
fIntestine_Lg	d110	79,221	36.2%	59,820	27.3%	79,759	36.5%	218,800
fIntestine_Lg	d107	74,910	36.3%	56,451	27.3%	75,225	36.4%	206,586
fIntestine_Lg	d108	37,563	34.6%	27,122	25.0%	43,890	40.4%	108,575
fIntestine_Lg	d115	62,644	34.0%	46,380	25.2%	75,358	40.9%	184,382
fIntestine_Lg	d113	79,272	37.3%	57,025	26.8%	76,378	35.9%	212,675
fIntestine_Lg	d91	67,186	35.1%	48,357	25.3%	75,831	39.6%	191,374
fIntestine_Lg	d120	76,130	37.0%	55,063	26.8%	74,358	36.2%	205,551
fIntestine_Lg	d98	134,881	39.7%	98,046	28.8%	107,213	31.5%	340,140
fIntestine_Sm	d110	83,624	36.4%	63,576	27.6%	82,777	36.0%	229,977
fIntestine_Sm	d115	60,130	35.3%	41,929	24.6%	68,262	40.1%	170,321
fIntestine_Sm	d105	88,488	37.0%	66,555	27.9%	83,810	35.1%	238,853
fIntestine_Sm	d87	74,051	34.9%	54,681	25.8%	83,166	39.2%	211,898
fIntestine_Sm	d91	76,506	35.9%	55,677	26.1%	81,038	38.0%	213,221
fIntestine_Sm	d107	69,329	34.4%	51,749	25.7%	80,171	39.8%	201,249
fIntestine_Sm	d108	69,186	36.9%	49,896	26.6%	68,172	36.4%	187,254
fIntestine_Sm	d120	84,929	36.7%	61,618	26.6%	84,730	36.6%	231,277
fIntestine_Sm	d98	82,490	36.4%	59,834	26.4%	84,068	37.1%	226,392
fKidney	d122	67,815	31.8%	55,748	26.1%	89,780	42.1%	213,343
fKidney	d121	72,854	31.3%	65,120	28.0%	94,674	40.7%	232,648
fKidney	d105	84,763	32.6%	74,115	28.5%	100,989	38.9%	259,867
fKidney	d85	103,037	34.9%	83,422	28.3%	108,643	36.8%	295,102
fKidney_L	d147	69,030	30.8%	65,513	29.2%	89,463	39.9%	224,006
fKidney_L	d110	78,574	31.9%	74,341	30.2%	93,040	37.8%	245,955
fKidney_L	d115	91,843	33.5%	75,945	27.7%	106,496	38.8%	274,284
fKidney_L	d98	99,602	33.8%	86,395	29.3%	108,716	36.9%	294,713

Cell- or tissue-type	Age*	Not aligned # peaks	Not aligned % peaks	Not DHS in mouse # peaks	Not DHS in mouse % peaks	DHS in mouse # peaks	DHS in mouse % peaks	Total#
fKidney_R	d117	80,948	32.3%	73,424	29.3%	96,164	38.4%	250,536
fKidney_R	d107	69,926	31.4%	60,148	27.0%	92,751	41.6%	222,825
fKidney_R	d87	84,389	33.0%	70,543	27.6%	100,749	39.4%	255,681
fKidney_renal_cortex	d108	82,569	32.3%	72,489	28.3%	100,743	39.4%	255,801
fKidney_renal_cortex	d113	102,935	34.1%	87,225	28.9%	111,276	36.9%	301,436
fKidney_renal_cortex	d120	84,310	32.5%	73,978	28.5%	100,858	38.9%	259,146
fKidney_renal_cortex	d97	71,790	31.6%	60,590	26.7%	94,514	41.7%	226,894
fKidney_renal_cortex	d96	80,546	32.3%	68,368	27.4%	100,344	40.3%	249,258
fKidney_renal_cortex	d89	94,585	33.5%	81,256	28.8%	106,134	37.6%	281,975
fKidney_renal_pelvis	d91	82,257	33.3%	66,083	26.8%	98,685	39.9%	247,025
fKidney_renal_pelvis	d127	69,085	31.2%	61,380	27.7%	91,090	41.1%	221,555
fKidney_renal_pelvis	d103	103,589	34.0%	89,644	29.5%	111,127	36.5%	304,360
fLung	d122	63,460	32.0%	53,397	26.9%	81,302	41.0%	198,159
fLung	d101	94,196	34.5%	81,299	29.8%	97,612	35.7%	273,107
fLung	d103	82,352	33.5%	75,419	30.7%	88,087	35.8%	245,858
fLung	d67	68,319	32.5%	57,978	27.6%	83,720	39.9%	210,017
fLung	d85	85,272	33.2%	76,009	29.6%	95,765	37.3%	257,046
fLung	d96	62,810	31.4%	56,107	28.1%	80,912	40.5%	199,829
fLung	d112	83,034	33.2%	76,799	30.7%	90,608	36.2%	250,441
fLung	d82	71,530	32.3%	67,830	30.6%	82,387	37.2%	221,747
fLung_L	d110	77,121	32.6%	70,644	29.9%	88,642	37.5%	236,407
fLung_L	d113	77,822	32.7%	68,552	28.8%	91,683	38.5%	238,057
fLung_L	d108	79,252	32.6%	71,016	29.2%	92,935	38.2%	243,203
fLung_L	d115	97,920	34.5%	81,141	28.6%	104,725	36.9%	283,786
fLung_L	d87	76,916	33.1%	67,133	28.9%	88,630	38.1%	232,679
fLung_R	d117	66,759	31.5%	61,657	29.1%	83,796	39.5%	212,212
fLung_R	d91	78,225	32.6%	68,454	28.5%	93,602	39.0%	240,281
fLung_R	d107	71,058	31.6%	64,334	28.6%	89,454	39.8%	224,846
fLung_R	d98	91,338	33.8%	79,063	29.2%	100,088	37.0%	270,489
fLung_R	d105	82,470	32.8%	72,146	28.7%	96,787	38.5%	251,403
fMuscle_arm	d115	76,837	31.4%	75,098	30.6%	93,131	38.0%	245,066
fMuscle_arm	d91	140,585	36.8%	120,026	31.4%	121,621	31.8%	382,232
fMuscle_back	d98	85,784	31.9%	79,664	29.6%	103,328	38.4%	268,776
fMuscle_back	d105	97,212	32.8%	88,002	29.7%	110,872	37.4%	296,086
fMuscle_back	d85	80,229	31.6%	75,674	29.8%	97,798	38.5%	253,701
fMuscle_back	d104	103,912	34.6%	90,807	30.3%	105,186	35.1%	299,905
fMuscle_leg	d127	73,330	31.0%	67,296	28.4%	95,923	40.6%	236,549
fMuscle_leg	d96	82,734	31.8%	78,408	30.1%	99,001	38.1%	260,143
fMuscle_leg	d97	79,719	31.4%	76,368	30.1%	97,614	38.5%	253,701
fMuscle_leg	d101	80,365	31.6%	74,766	29.4%	98,995	39.0%	254,126
fMuscle_leg	d113	87,071	32.2%	81,261	30.0%	102,386	37.8%	270,718
fMuscle_trunk	d120	81,120	31.9%	74,126	29.2%	98,762	38.9%	254,008
fMuscle_trunk	d121	70,422	30.8%	65,285	28.6%	92,650	40.6%	228,357
fOvary	pool	54,725	32.8%	40,692	24.4%	71,443	42.8%	166,860
fPlacenta	d108	116,138	42.4%	73,834	27.0%	83,665	30.6%	273,637
fPlacenta	d91	90,994	41.8%	55,110	25.3%	71,338	32.8%	217,442
fPlacenta	d85	101,630	44.2%	60,218	26.2%	68,195	29.6%	230,043
fPlacenta	d113	91,694	42.6%	55,904	26.0%	67,825	31.5%	215,423
fPlacenta	d105	109,086	44.4%	65,882	26.8%	70,684	28.8%	245,652
fRetina	d125	72,230	32.2%	62,616	27.9%	89,552	39.9%	224,398
fRetina	d87	78,004	30.6%	73,024	28.6%	104,094	40.8%	255,122
fRetina	d103	127,649	32.1%	140,437	35.4%	129,095	32.5%	397,181
fSpinal_cord	d105	77,694	27.9%	91,064	32.7%	109,827	39.4%	278,585
fSpinal_cord	d96	56,804	26.8%	60,443	28.6%	94,399	44.6%	211,646
fSpinal_cord	d113	72,777	29.0%	71,650	28.5%	106,887	42.5%	251,314
fSpinal_cord	d89	63,565	27.1%	69,648	29.7%	101,447	43.2%	234,660
fSpinal_cord	d87	53,882	26.2%	60,496	29.4%	91,642	44.5%	206,020
fSpleen	d112	75,576	35.2%	57,468	26.7%	81,904	38.1%	214,948
fStomach	d147	61,626	31.3%	49,890	25.3%	85,604	43.4%	197,120
fStomach	d107	65,459	31.6%	52,623	25.4%	89,089	43.0%	207,171
fStomach	d91	52,436	31.0%	41,716	24.7%	74,922	44.3%	169,074
fStomach	d98	83,927	32.7%	68,202	26.6%	104,575	40.7%	256,704
fStomach	d105	65,861	31.4%	54,465	26.0%	89,306	42.6%	209,632
fStomach	d127	59,019	32.1%	44,425	24.2%	80,389	43.7%	183,833
fStomach	d101	64,719	31.5%	53,793	26.2%	87,116	42.4%	205,628
fStomach	d96	74,012	31.9%	61,046	26.3%	96,665	41.7%	231,723
fStomach	d108	69,215	31.2%	58,683	26.4%	94,240	42.4%	222,138
fStomach	d121	73,677	32.7%	58,625	26.0%	93,026	41.3%	225,328
fThymus	d147	60,164	37.3%	40,987	25.4%	60,059	37.3%	161,210
fThymus	d105	48,894	34.2%	33,197	23.2%	60,748	42.5%	142,839
fThymus	d98	68,133	37.9%	45,099	25.1%	66,619	37.0%	179,851
fThymus	d127	73,567	38.2%	49,173	25.5%	69,990	36.3%	192,730
fThymus	d104	57,055	36.7%	37,830	24.3%	60,581	39.0%	155,466

Cell- or tissue-type	Age*	Not aligned		Not DHS in mouse		DHS in mouse		Total#
		# peaks	% peaks	# peaks	% peaks	# peaks	% peaks	
fThymus	d108	48,765	35.6%	31,881	23.3%	56,443	41.2%	137,089
fThymus	d113	51,506	35.3%	34,246	23.5%	60,161	41.2%	145,913
GM04503D		89,876	37.4%	70,603	29.4%	79,709	33.2%	240,188
GM04504A		84,495	37.1%	65,604	28.8%	77,926	34.2%	228,025
GM06990		43,530	38.0%	27,508	24.0%	43,604	38.0%	114,642
GM12864		67,850	41.4%	40,417	24.6%	55,753	34.0%	164,020
GM12865		69,664	41.0%	42,297	24.9%	57,913	34.1%	169,874
GM12878		55,743	39.3%	34,566	24.3%	51,686	36.4%	141,995
H7_hESC_T14		59,597	34.5%	44,000	25.4%	69,296	40.1%	172,893
H7_hESC_T2		60,813	30.8%	52,290	26.5%	84,526	42.8%	197,629
H7_hESC_T5		101,067	40.4%	70,850	28.3%	78,116	31.2%	250,033
HAc		77,240	35.4%	61,257	28.0%	79,958	36.6%	218,455
HAEpiC		97,121	39.8%	69,829	28.6%	76,895	31.5%	243,845
HAh		86,423	35.2%	70,941	28.9%	88,082	35.9%	245,446
HAsp		88,285	37.4%	75,303	31.9%	72,765	30.8%	236,353
HBMEC		92,189	38.3%	69,819	29.0%	78,658	32.7%	240,666
HBVP		96,603	38.5%	74,378	29.6%	80,001	31.9%	250,982
HBVSMC		67,366	34.3%	54,369	27.7%	74,805	38.1%	196,540
HCFaa		86,530	39.2%	64,883	29.4%	69,583	31.5%	220,996
HCF		76,528	35.7%	57,601	26.9%	80,296	37.4%	214,425
HCM		87,385	37.2%	64,247	27.4%	83,160	35.4%	234,792
HConF		64,326	34.1%	49,361	26.2%	74,723	39.7%	188,410
HCPEpiC		97,182	38.4%	72,293	28.5%	83,826	33.1%	253,301
HCT116		53,125	38.8%	31,591	23.1%	52,187	38.1%	136,903
HEEpiC		101,853	41.6%	68,683	28.1%	74,232	30.3%	244,768
Hela		59,123	39.5%	39,411	26.3%	51,264	34.2%	149,798
HEPG2		39,242	35.0%	26,199	23.4%	46,656	41.6%	112,097
HESC		71,678	39.3%	45,463	24.9%	65,350	35.8%	182,491
hESCT0		136,414	44.1%	85,660	27.7%	86,969	28.1%	309,043
HFF		85,055	36.4%	65,221	27.9%	83,648	35.8%	233,924
HFF_MyC		98,289	38.6%	70,575	27.7%	85,532	33.6%	254,396
HGF		62,573	35.1%	48,782	27.4%	66,760	37.5%	178,115
HIPEpiC		107,400	39.9%	78,797	29.3%	82,905	30.8%	269,102
HL60		81,343	43.2%	47,679	25.3%	59,430	31.5%	188,452
HMEC		63,060	37.1%	46,803	27.5%	60,317	35.4%	170,180
HMF		80,142	37.3%	59,590	27.7%	75,340	35.0%	215,072
HMVEC_dAd		51,193	33.7%	37,593	24.7%	63,319	41.6%	152,105
HMVEC_dBlAd		71,178	36.5%	51,785	26.6%	71,870	36.9%	194,833
HMVEC_dBlNeo		74,224	36.6%	54,039	26.6%	74,553	36.8%	202,816
HMVEC_dLyAd		53,129	34.4%	39,295	25.5%	61,810	40.1%	154,234
HMVEC_dLyNeo		65,563	35.4%	48,152	26.0%	71,372	38.6%	185,087
HMVEC_dNeo		59,515	34.7%	43,238	25.2%	68,778	40.1%	171,531
HMVEC_LBl		75,549	37.0%	55,622	27.2%	73,266	35.8%	204,437
HMVEC_LLy		61,126	35.1%	44,974	25.8%	67,968	39.0%	174,068
HNPCEpiC		98,777	38.8%	74,578	29.3%	80,926	31.8%	254,281
HPAEC		52,042	34.6%	38,140	25.4%	60,126	40.0%	150,308
HPAF		84,471	37.1%	61,520	27.1%	81,433	35.8%	227,424
HPdLF		74,324	35.8%	59,192	28.5%	73,981	35.7%	207,497
HPF		65,833	34.7%	50,672	26.7%	73,258	38.6%	189,763
HRCE		88,432	38.3%	65,038	28.1%	77,596	33.6%	231,066
HRE		85,205	37.6%	63,436	28.0%	78,219	34.5%	226,860
HRGEC		63,575	37.0%	47,061	27.4%	61,400	35.7%	172,036
HRPEpiC		100,268	37.4%	80,359	30.0%	87,440	32.6%	268,067
HSMM_D		101,244	37.5%	80,135	29.7%	88,246	32.7%	269,625
HSMM		104,311	38.4%	83,219	30.7%	83,901	30.9%	271,431
hTH17		35,035	37.7%	22,567	24.3%	35,419	38.1%	93,021
hTH1		35,919	36.6%	22,049	22.5%	40,238	41.0%	98,206
hTH2		59,737	40.0%	36,764	24.6%	52,780	35.4%	149,281
hTR		57,337	38.3%	34,943	23.3%	57,424	38.4%	149,704
HUVEC		50,074	33.7%	39,793	26.8%	58,633	39.5%	148,500
HVMF		78,271	37.4%	58,697	28.1%	72,042	34.5%	209,010
Jurkat		80,694	43.8%	47,148	25.6%	56,320	30.6%	184,162
K562		75,406	44.5%	41,067	24.3%	52,826	31.2%	169,299
LHCN_M2_D4		96,994	38.0%	72,396	28.3%	86,072	33.7%	255,462
LHCN_M2		87,242	38.2%	64,367	28.2%	76,491	33.5%	228,100
LNCap		86,111	41.2%	61,243	29.3%	61,736	29.5%	209,090
M059J		104,568	40.4%	73,394	28.3%	81,008	31.3%	258,970
MCF7		103,258	42.7%	65,227	27.0%	73,364	30.3%	241,849
MCF7		62,412	41.2%	40,681	26.8%	48,493	32.0%	151,586
MCF7_ER		110,407	43.3%	69,194	27.2%	75,090	29.5%	254,691
NB4		70,437	41.2%	41,264	24.1%	59,308	34.7%	171,009
NHA		85,719	36.9%	68,675	29.5%	78,051	33.6%	232,445
NHBE_RA		68,025	37.4%	46,201	25.4%	67,562	37.2%	181,788

Cell- or tissue-type	Age*	Not aligned		Not DHS in mouse		DHS in mouse		Total#
		# peaks	% peaks	# peaks	% peaks	# peaks	% peaks	
NHDF_Ad		102,754	37.7%	80,642	29.6%	88,807	32.6%	272,203
NHDF_Neo		80,333	35.4%	66,749	29.4%	80,125	35.3%	227,207
NHEK		67,911	39.1%	47,560	27.4%	58,321	33.6%	173,792
NHLF		92,846	37.4%	74,196	29.9%	81,323	32.7%	248,365
NT2_D1		88,796	41.0%	55,529	25.6%	72,255	33.4%	216,580
PANC1		55,251	38.7%	36,159	25.3%	51,479	36.0%	142,889
PrEC		81,066	40.9%	55,146	27.9%	61,781	31.2%	197,993
RPMI_7951		74,892	37.3%	53,533	26.7%	72,143	36.0%	200,568
RPTEC		79,909	39.2%	56,747	27.8%	67,167	33.0%	203,823
SAEC		96,014	41.2%	66,279	28.5%	70,540	30.3%	232,833
SKMC		91,154	36.8%	73,780	29.8%	82,738	33.4%	247,672
SK_N_MC		63,753	37.0%	52,338	30.3%	56,359	32.7%	172,450
SKNSH		31,456	28.0%	28,969	25.8%	52,062	46.3%	112,487
T_47D		71,740	40.0%	45,070	25.1%	62,636	34.9%	179,446
WERI_Rb1		88,488	40.7%	63,636	29.3%	65,350	30.0%	217,474
WI_38		75,170	37.4%	56,069	27.9%	69,727	34.7%	200,966
WI_38_TAM		97,277	39.9%	71,270	29.2%	75,185	30.8%	243,732

* y = years post-natal, d = days post-conception, blank = primary, cultured or malignant cells
# Total peaks from master peaks list (see Supporting Methods)