doi:10.1038/nature13992

This PDF file includes:

Materials and Methods
Supplementary Figures 1 to 22
Supplementary Tables 1, 2, 3, 9, 10, 14, 15, 20, 27, 29, 33.
References

The following Supplementary Material for this manuscript is available at the mouse ENCODE portal website: http://mouse.encodedcc.org)

Supplementary Tables 4-8, 11-13, 16-19, 21-26, 28, 30-32.

Table of contents

Supplementary tables:

Supplementary Table 11: List of chromatin states identified in the15 mouse tissues and cell types and six human cell lines.
Supplementary Table 12: List of replication timing profiles for 11 mouse and 9 human cell types.
Supplementary Table 13: List of human and mouse orthologous gene pairs (MIT and Emsemble)
Supplementary Table 14: Summary of the 1:1 and many to 1 human/mouse orthologous genes.
Supplementary Table 15: GO categories for 1:many human/mouse orthologous genes.
Supplementary Table 16: List of the 151,257 orthologous exon in the mouse genome.
Supplementary Table 17: List of the 204,887 orthologous splice junction pairs in the mouse genome.
Supplementary Table 18: List of the 2,717 (3,446) novel human(mouse) exons.
Supplementary Table 19: List of the 2,327 human transcripts (corresponding to 1,679 genes) homologous to 5,067 putative mouse transcripts
Supplementary Table 20: RNA-seq datasets used for PCA analysis.
Supplementary Table 21: Complete list of all GO categories for analysis of conservation of gene expression programs.
Supplementary Table 22: List of the 164,428 candidate mouse enhancers with homologous sequences in the human genome.
Supplementary Table 23: List of the 51,661 candidate mouse promoters with homologous sequences in the human genome.
Supplementary Table 24: List of 918,002 DHS sites with homologous sequences in the human genome.
Supplementary Table 25: List of 149,521 TF binding sites with homologous sequences in the human genome.
Supplementary Table 26: List of 12,387 predicted mouse-specific promoters.
Supplementary Table 27: List of the predicted mouse-specific predicted promoters tested in reporter assay.
Supplementary Table 28: List of 48,245 predicted mouse-specific enhancers.
Supplementary Table 29: List of the predicted mouse-specific enhancers tested in reporter assays performed with the mouse and human cells.
Supplementary Table 30. List of mouse-specific TF binding sites.
Supplementary Table 31: Predicted orthologous promoters in the mouse and human genomes.
Supplementary Table 32: Predicted orthologous enhancers in the mouse and human genomes.
Supplementary Table 33: List of suspicious 1:1 human/mouse orthologous genes with manual curation.

Materials and Methods

RNA-Seq assays

RNA-Seq assays performed in the Gingeras group:

Mouse Tissue Acquisition: C57Bl/6 mice were used for all tissue resections. The following tissues were taken from 8-week old littermates: Adrenal, Duodenum, Stomach, Genital Fat Pad, Subcutaneous Fat Pad, Large Intestine, Small Intestine, Ovary, Testis, Spleen, Colon, Lung, Heart, Kidney, Liver, Thymus, Mammary Gland, Placenta (from pregnant mice), Cortex, Frontal Lobe, Cerebellum, Bladder, Liver. Central nervous system (CNS) taken from stage E11.5 littermates. Central nervous system (CNS) and liver were taken from E14 littermates. Liver, Limb and Whole Brain were taken from E14.5 littermates. Central nervous system (CNS) and liver were taken from E18 littermates.

Human Cell Culture and RNA Isolation: Cells were grown according the ENCODE cell culture standards. RNA was prepared as was described.[1]

Library Construction: We generated directional (stranded) libraries for Paired End (PE) sequencing on the Illumina platform as described in [2] and as was also used in [1], to generate the human ENCODE data. Briefly, 100 ng of Ribominus, (Invitrogen Inc.) treated PolyA(+) RNA with length > 200 nt were mixed with 2 ng of exogenous RNA spike-in, pool 14 [3]. A mixture of random hexamers and oligo-dT_{21} were used to prime the reverse transcriptase reaction. Entry sites for second strand synthesis catalyzed by E. coli DNA Polymerase are generated through RNAse H nicks of the DNA:RNA duplex. dTTP is replaced with dUTP during the second strand synthesis. The (ds)cDNA is then sheared using sonification (Covaris). Staggered ends generated during shearing, are repaired and adenylated to prime them for adapter ligation with Illumina Y-adapters. The second strand containing dUTP is eliminated using UNG digestion. The resulting (ss)cDNA is run on an agarose gel and bands with the desired insert sizes of ~ 200 nt are cut out. Cluster compatible sequences are appended in an 18-cycle PCR reaction and the final library is gel purified. All libraries from biological replicates (littermates) were prepared in parallel to minimize day-to-day variation in the experimental procedure. Finally, the libraries were sequenced on the Illumina GAIIx or Hi-Seq platform to an average depth of ~100 million mate pairs per sample. The human ENCODE RNA-Seq data is at GEO: GSE30567. The mouse ENCODE RNA-Seq data is at GEO: GSE36025.

RNA-Seq assays performed in the Snyder group

Total RNA was extracted from cells grown under the same conditions as described for ChIP-seq (below) after indicated time stimulation with TNF-alpha or without stimulation, respectively, using RNeasy kit (Qiagen) according to the manufacturer's instructions.

Strand-specific RNA-seq: Total RNA was extracted using RNeasy kit (Qiagen) according to the manufacturer's instructions. Ribo-Zero™ Magnetic Gold Kit (Epicentre Cat # MRZG126 , Madison, WI) was used to deplete ribosomal RNA. The isolated RNA was then sheared using the RNA Fragmentation Reagents (Life Technologies, Grand Island, NY, USA) at 80 °C for 1 min. The RNA fragment size range was mainly 100~500 bp on 2% E-gel (Invitrogen Cat # G4020-02, Carlsbad, CA). Doublestranded cDNA synthesis was primed with random hexamers and ds cDNA Kit (Invitrogen, Cat # 11917-010, Carlsbad, CA) following the manufacturer's instruction with some modification. Briefly, after First-Strand cDNA synthesis, NucAway Spin Column (Life Technologies, Grand Island, NY, USA) was used to remove dNTP. Then during instruction with some modification. Briefly, after First-Strand cDNA synthesis, NucAway Spin Column (Life Technologies, Grand Island, NY, USA) was used to remove dNTP. Then during the second strand reaction, dUTP was incorporated instead of dTTP. The reaction then was purified with QIAquick PCR spin column (Qiagen, #28106, Valencia, CA). The cDNA libraries were constructed and sequenced as described previously . Briefly, cDNA was end-repaired with the End-It kit from Epicentre (Cat# ER0720, Madison, WI). After treatment with Klenow fragment (NEB, Cat # M0212s, Ipswich, MA) and dATP Illumina adapters were ligated to the protruding 3_-'A' base (LigaFast, Promega Cat # M8221, Madison, WI). Ligated product was size selected on 2% E-gel (Invitrogen) (250-300 bp). The dUTP-containing second strands were removed by incubating with Uracil-DNA Glycosylase (New England Biolabs, Ipswich, MA, USA), and the libraries were amplified with Illumina genomic DNA primers 1.1 and 2.1 at the following conditions: 98°C 30 sec, 15 cycles of (98°C 10 sec, 65°C 30 sec, 72°C 30 sec), 72°C 5 min. The final amplified libraries were further purified by agarose gel excision and extraction. The library was subjected to 101b paired-end sequencing using Illumina's HiSeq 2000 Sequencer.

RNA-Seq assays performed in the Ren group

RNA samples from mouse tissues and primary cells were extracted from Trizol® according to manufacturer's protocol (Invitrogen). polyA+RNA was purified with the Dynabeads mRNA purification kit (Invitrogen). The mRNA libraries were prepared for strand-specific sequencing as described previously [2].

RNA-Seq assays performed in the Hardison group

Cells were grown according to the approved ENCODE cell culture protocols. Total RNA was extracted from 5-10 million cells using TRIzol reagent, followed by mRNA selection, fragmentation and cDNA synthesis, which were performed as described previously [4]. Double-stranded cDNA samples were processed for library construction for Illumina sequencing, using the Illumina ChIP-seq Sample Preparation Kit. Strand-specific libraries were generated in a similar manner, except for a couple of modifications described previously [2]. Briefly, instead of dTTP, dUTP was used during second-strand cDNA synthesis to label the second-strand cDNA. During library preparation, the dUTP-labeled cDNA was treated with Uracil N Glycosylase, prior to the PCR amplification step. This was done to remove uracil from the second-strand, following which the DNA was subjected to high heat to facilitate abasic scission of the second strand. Library quality and size were evaluated using the Agilent Bioanalzyer 2100 (Aglient Technologies, Santa Clara, CA), and libraries were quantified with Quant-iT dsDNA HS Assay Kit (Molecular Probes, Eugene OR), and qPCR. DNA libraries were sequenced on either the Illumina Genome Analyzer II sequencing system or the HiSeq 2000. Cluster generation and sequencing chemistry were performed using Illumina-supplied kits as appropriate. All samples were determined as biological replicates.

ChIP-seq assays

ChIP-Seq assays of transcription factors (Snyder group)

Biological replicates were grown in separate batches and at separate times. Briefly, 5 x 10^7 cells were grown to a density of 0.6-0.8 x 10^6/mL, then treated with 25 ng/mL human recombinant TNF-alpha (eBioscience #14-8329, San Diego, CA) for 30 minutes, 60 minutes or

120 minutes at 37°C, 5% CO_2. After stimulation, cells were cross-linked in 1% formaldehyde for 10 minutes at room temperature. Nuclear lysates were sonicated using a Branson 250 Sonifier (power setting 7, 100% duty cycle for 12 × 20-s intervals), such that the chromatin fragments ranged from 50-2000kb. Clarified lysates were divided in half and treated overnight at 4oC with 12µg of either anti-NFκB p65 (C-20) rabbit polyclonal antibody (Santa Cruz Biotechnology sc-372, Santa Cruz, CA), anti-pol II (Covance MMS-126R, Princeton, NJ), Anti-pol II CTD repeat YSPTSPS (phospho S2) (Abcam ab5095, Cambridge, MA) (Pol II-s2) antibody or normal rabbit / mouse IgG (Santa Cruz Biotechnology Sc-2027, Santa Cruz, CA; Millipore #12-371, Billerica, MA). Protein-DNA complexes were captured on Protein A/G agarose beads (Millipore #16-156/16-266, Billerica, MA) and eluted in 1% SDS TE buffer at 65°C. Following cross-link reversal and purification, the ChIP DNA sequencing libraries were prepared according to Illumina DNA Sample Kit Instructions (Illumina Part # 0801– 0303, San Diego, CA). Libraries were sequenced on an Illumina Genome Analyzer II.

ChIP-Seq assays of transcription factors and chromatin modification marks (Ren group)

Mouse Tissues and Cell Culture: Adult bone marrow, cerebellum, cortex, heart, intestine, kidney, liver, lung, olfactory bulb, spleen, testis, brown adipocytes, and thymus were dissected from 8-week old male C57Bl/6 mice (Charles River). Placenta was extracted from C57Bl/6 pregnant mice. E14.5 brain, heart, limb and liver, and mouse embryonic fibroblast (MEF) cells were derived from E14.5 C57Bl/6 mouse embryos. Tissues were minced to fine pieces in PBS and fixed with 1% formaldehyde at room temperature for 20 minutes. For bone marrow derived macrophage, Murine BMDMs were generated by flushing bone marrow cells from the femurs and tibias of mice using a 27 1/2G syringe. These cells were cultured for 7 d in DMEM (Life Technologies, Rockville, MD) containing 10% FBS (Omega Scientific, Tarzana, CA), 1% penicillin/streptomycin (Life Technologies), and 10% conditioned media (CM) from L929 cells overexpressing M-CSF at a cell density of 250, 000 cells/mL on 10cm tissue cultured treated dishes. CM was replaced on day 4 of differentiation and every 2 day thereafter. MEF cells were genotyped to select male MEF cells used for this study. Placenta was dissected from pregnant C57Bl/6 mice at E14.5. mESC lines E14 and Bruce4 were maintained on mitomycin C-inactivated MEF feeder layers in DME containing 15% fetal calf serum, leukemia inhibiting factor, penicillin/streptomycin, L-glutamine and non-essential amino acids. mESCs were passaged on 0.2% gelatin twice to deplete feeder cells before harvest for experiments. CH12 and MEL cells were obtained and grown according to standard ENCODE cell culture protocols [5]. All cultured cells were fixed with 1% formaldehyde at room temperature for 15 minutes.

ChIP-Seq was carried out as previously described [6] with 500 µg chromatin and 5 µg antibody with the following antibodies, H3K4me3 (Millipore 05-745), H3K4me2 (Abcam, ab32356), H3K4me1 (Abcam, ab8895), H3K9ac (Active Motif, AM39137), H3K9me3 (Abcam, ab8898), H3K27ac (Active Motif, 39133), H3K27me3 (Active Motif, AM39155), H3K36me3 (Abcam, ab9050), H3K79me2 (Active Motif, AM39143), polII (Covance, MMS-126R), CTCF (ref), and P300 (Santa Cruz, sc585). ChIP and input library preparation and sequencing procedures were carried out as described previously according to Illumina protocols with minor modifications (Illumina, San Diego, CA).

ChIP-seq assays of transcription factors and chromatin modifications (Hardison Lab)

Cells were grown according to the approved ENCODE cell culture protocols. The chromatin immunoprecipitation followed published methods [7]. Information on antibodies used is available via the hyperlinks in the "Select subtracks" menu. Samples passing initial quality thresholds (enrichment and depletion for positive and negative controls - if available - by quantitative PCR of ChIP material) are processed for library construction for Illumina sequencing, using the ChIP-seq Sample Preparation Kit, or more recently the TruSeq ChIP Sample Prep Kit, purchased from Illumina. Starting with a 5-10 ng sample of ChIP DNA, DNA fragments were repaired to generate blunt ends and a single A nucleotide was added to each end. Double-stranded Illumina adaptors were ligated to the fragments, and the ligation products were amplified by 18 cycles of PCR. Amplified DNA products between 250-350 bp were gel purified. Library quality and size were evaluated using the Agilent Bioanalzyer 2100 (Aglient Technologies, Santa Clara, CA), and libraries were quantified with Quant-iT dsDNA HS Assay Kit (Molecular Probes, Eugene OR), and qPCR. DNA libraries were sequenced on either the Illumina Genome Analyzer II sequencing system or the HiSeq 2000. Cluster generation and sequencing chemistry were performed using Illumina-supplied kits as appropriate. All samples were determined as biological replicates.

DNase-Seq assays (Stamatoyannopoulos group)

DNaseI hypersensitivity mapping including nuclei isolation, DNaseI treatment, DNaseI double-hit fragment purification, library construction, and sequencing was performed essentially as previously described [8-12]. Different handling protocols were specifically developed and utilized for the following classes of mouse cell/tissue samples; the detailed protocols can be found at the associated URLs:

(i) Cultured mouse cell lines http://www.mouseencode.org/media/protocols/data_assay/dnasei-mouse/Culturedcells_SOP_nuclei_DNase_crosslink_RNA_V1_mouse_dnase.pdf
(ii) Primary mouse immune cells (T cells, B cells)
http://www.mouseencode.org/media/protocols/data_assay/dnasei/03042011_mouse_Rudensky_V5.pdf
(iii) Fresh mouse tissues
http://www.mouseencode.org/media/protocols/data_assay/dnasei/01122011_nuclei_isolation_mouse__tissue_V2.pdf
(iv) Frozen mouse tissues
http://www.mouseencode.org/media/protocols/data_assay/dnasei/06242011_nuclei_isolation_cryomouse_tissue-Dounce_V1.pdf

Conventional DHS data sets were sequenced to an average target depth of 30 million uniquely mapping 36bp reads. Data sets for genomic footprinting were sequenced to an average depth of 270 million uniquely mapping 36bp reads.

Replication Timing analysis (Gilbert group)

Protocols for generating and quality control for Repli-chip profiles were performed as previously described [13,14]. CH12 and MEL cells were obtained and grown according to standard ENCODE cell culture protocols [5]. Cells were labeled with 100 uM BrdU for 2 h, and BrdU-substituted DNA was extracted using anti-BrdU antibody and hybridized onto NimbleGen Mouse CGH

3x720K Whole-Genome Tiling Arrays (NimbleGen, 100718_MM9_WG_CGH) according to the manufacturer's protocol. Human and mouse Repli-chip profiles are available at GEO (GSE45716; GSE18019; GSE20027), www.ReplicationDomain.org, and the UCSC genome browser.

Luciferase reporter assays for functional validation of candidate promoters

Predicted promoter sequences randomly selected for validation are listed in Supplementary Table 9. A total of 75 previously unknown promoters were tested, including 25 mESC-specific promoters, 25 MEF-specific promoters and 25 found to be active in both mESCs and MEF. Half of these tested promoter candidates were associated with H3K4me3 signal only, while the other half were associated with both H3K4me3 and polII signals. All candidate promoters were tested in both orientations. Cloning and reporter assays were carried out as previously described [15]. For novel promoter sequences, we tested both orientations of the candidate sequences. Fragments were designated as active if their relative luciferase value was significantly higher than random genomic fragments (P value < 0.01).

A high throughput reporter assay for functional validation of the candidate enhancers

The chromosome coordinates and primers of the tested genomic fragments are listed in Supplementary Table 10. To test the accuracy of enhancer predictions, we performed quantitative high throughput reporter assay [16] with 183 candidate mESC-specific enhancers and 30 negative control regions. Briefly, candidate enhancer regions (1kb) were first amplified from C57BL/6 mouse genomic DNA by PCR with high fidelity 2X KAPA PCR mix. The amplified enhancer sequences were pooled and randomly sonicated into fragments of 250-500bp and size selected on a 2% agarose gel. Tru-Seq adaptors were then ligated to purified enhancer fragments according standard Illumina library preparation procedure. The ligation products were then amplified by PCR with primers (Forward: GCC GTG TAA TTC TAG AAA TGA TAC GGC GAC CAC CGA GAT CTA CAC, and Reverse: CCG CCC CGA CTC TAG ACA AGC AGA AGA CGG CAT ACG AGC T), and purified by size selection on a 2% agarose gel. The purified PCR products were then cloned (Clontech In-fusion HD) into pGL3 promoter (Promega) vector that were linearized at XbaI site. The In-fusion cloning product was ethanol precipitated, dissolved in 10 ul H2O, and transformed with DH10B electrophorate competent cells (Life Technology) to form reporter library. The reporter library was prepared with Qiagen.

The reporter library was transfected into mESCs according to Amaxa nucleofactor protocol (program A-13). RNA and DNA were harvested 24 hours post transfection with Qiagen Allprep kit. The input DNA library was generated with Tru-Seq PCR primers. For RNA library, 1ug of total RNA was treated with DNaseI and subject to first strand synthesize with Invitrogen SuperscriptIII using a reporter-RNA specific primer (TGTATCTTATCATGTCTGCTCGAA). RNA-Seq library was made with Tru-Seq PCR primers with 1st strand DNA. Each PCR reaction was monitored so that minimal cycle number was used for library construction.

Both Input DNA and RNA-Seq library were sequenced on an Illumina Hi-Seq 2000 instrument, and the sequence reads were mapped to mm9 with bowtie. The relative enhancer activity was calculated as the log ratio of (RNA-Seq reads)/(DNA-Seq reads) for each candidate enhancers or negative region.

For each putative enhancer region, RPKM was calculated as the number of reads per kilobase of the enhancer region per million reads mapped, with a pseudocount of 0.05 RPKM. Negative controls sets were chosen based on the following criteria: 1) they are in the gene desert area, and 2) they don't exhibit any chromatin modification signals, TF binding sites or DHS sites. The median of the negative control is -0.95 and standard deviation is 0.51. For the tested enhancers, we decided to use median + 1.63SD (-0.119 RPKM) to determine whether an enhancer is tested positive or negative.

Data processing methods

Mapping of RNA-seq reads:

Gingeras group: STAR 1.9 [17] was used to map the RNA-seq data to the reference mouse genome mm9. Up to 10 mismatches per paired alignment were allowed. Only alignments for reads mapping to 10 or fewer loci were reported. Annotations were not utilized for mapping the data.

Snyder group: RNA-seq reads were mapped by Tophat (version 1.3) to mouse mm9 reference genome and FPKM were calculated with cufflink by using ensembl genes as guided reference. Orthologous genes between human and mouse were normalized by quantile normalization and differentially expressed genes were calculated by using limma package on log transformed expression value .

Hardison group: FastQ files for the resulting sequence reads (single read and paired-end, directional and non-directional) were moved to a data library in Galaxy, and tools implemented in Galaxy were used for further processing via workflows [18-20]. Data processing was also performed on the CyberSTAR high-performance computing system at Penn State. The reads were mapped to the mouse genome (mm9 assembly) using the program TopHat [21,22]. Signal tracks were created using BEDtools [23] and SAMtools [24].

Ren group: Raw reads were mapped in FASTQ format to the mouse genome (mm9) with TopHat software version 1.20 35 [22]. The wig files for RNA-Seq data were generated by TopHat. We assigned expression value (FPKM) for each RefSeq gene with Cufflinks [25]. To normalize the gene expression levels between different tissues, we used the quantile normalization function in R.

To investigate whether different mapping tools might affect the results, we systematically performed comparison between two different mappers (STAR and Tophat) and found their results are consistent, with an average correlation 0.94. Please refer to Supplementary Fig. 18 for more details.

Contig assembly from RNA-Seq reads (Gingeras group)

Transcript contigs were generated as described in [1]. Contigs are defined by collapsing the RNA-seq coverage signal from merged biological replicates. Each contig is quantified against individual replicates to enable non-parametric irreproducible discovery rate (npIDR) analysis [17]. Contigs are required to have non-zero signal in both replicates. Only uniquely mapping reads are used for building and quantifying contigs. Neighboring contigs are merged if the gap between them is smaller than 25 bases. Contigs are called in a strand-specific way. To eliminate possible artifacts of stranded library construction, contigs are required to have > 1/9 signal of the contigs on the opposite strand. Each contig is quantified with: (1) BPKM (Bases per Kilobase per Million mapped bases) averaged between the replicates. (2) npIDR between the replicates. (3) The total number of mapped bases in the contig in both replicates (sum of wiggle track signal).

ChIP-Seq data processing

Snyder Lab

Uniform pipeline: We used a uniform processing pipeline to identify high confidence binding events (peaks) in human and mouse.

Read Mapping: For human ChIP-seq, mapped reads in the form of BAM files were downloaded from ENCODE UCSC DCC (http://encodeproject.org/ENCODE/downloads.html. These BAM files were generated by the individual data production labs using different mappers and mapping parameters. For mouse ChIP-seq, reads were mapped by BWA. In order to standardize the mapping protocol, we used custom mappability tracks to filter out multi-mapping reads and only retain unique mapping reads i.e. reads that map to exactly one location in the genome. We also filtered all positional and PCR duplicates.

Quality control: A number of quality metrics for all replicate experiments of each dataset were computed. In brief, these metrics measure ChIP enrichment and signal-to-noise ratios, sequencing depth and library complexity and reproducibility of peak calling. These measures will be reported at the ENCODE portal at http://encodeproject.org/ENCODE/qualityMetrics.html. Datasets that did not pass the minimum quality control thresholds were discarded and not used in any analyses. Datasets that passed most but not all quality metrics were flagged.

Peak Calling: All ChIP-seq experiments were scored against an appropriate control designated by the production groups (either input DNA or DNA obtained from a control immunoprecipitation). We used the SPP [26] peak caller to identify and score (rank) potential binding sites/peaks. For obtaining optimal thresholds, we used the Irreproducible Discovery Rate (IDR [27]) framework to determine high confidence binding events by leveraging the reproducibility and rank consistency of peak identifications across replicate experiments of a dataset. Code and detailed step-by-step instructions to call peaks using the IDR framework are available at https://sites.google.com/site/anshulkundaje/projects/idr.

Black list: All peak sets were then screened against specially curated empirical blacklists for each species. Briefly, these blacklist regions typically show the following characteristics:
- Unstructured and extreme high signal in sequenced input-DNA and control datasets as well as open chromatin datasets irrespective of cell type identity.
- An extreme ratio of multi-mapping to unique mapping reads from sequencing experiments.
- Overlap with specific types of repeat regions such as centromeric, telomeric and satellite repeats that often have few unique mappable locations interspersed in repeats. The human blacklist can be downloaded from: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz). The Mouse blacklist can be downloaded from: (http://www.broadinstitute.org/~anshul/projects/mouse/blacklist/mm9-blacklist.bed.gz)

Comparison with DHS: To systematically investigate to what degree TF binding sites overlap with DNase I HSS, we performed the following analysis in MEL and CH12 cell lines. First we merged all the binding sites from different transcription factors in the same cell line (33 TFs in CH12 and 43 TFs in MEL), and compared the union of the binding sites with DHS in the same cell type. We found 61% and 72% of TFBS in MEL and CH12 overlap with DHS, and 81% and 67% of DHS in MEL and CH12 overlap with TFBS in the same cell type (Supplementary Fig. 19).

Ren lab

ChIP-Seq reads were aligned to the mouse genome build mm9 with Bowtie version 0.12 [21]. We used the first 25 bp for the alignment and only kept the reads with less than two mismatches. Duplicated reads from the same biological library were removed. For each chromatin mark, we have at least two good biological replicates. After we verified that biological replicates are highly correlated, they were pooled for all subsequent data analysis.

Hardison Lab

The ChIP-seq reads were moved to a data library in Galaxy, and the tools implemented in Galaxy were used for further processing via workflows [19]. There are two biological replicates for each ChIP target in each condition. The reads from two replicates were pooled and then mapped to the mouse genome (mm9 assembly) using the program Bowtie [21], and the mapped reads for the ChIP sample and from the "input" control (no antibody) were processed by MACS [28] to calculate signals (tag counts for every 1 bp segment). Peaks or enrichment domains were identified by MACS for TF occupancy, or by SICER [29] for histone modifications.

DNase-seq data analysis (Stamatoyannopoulos group)

Sequence alignment to the mouse genome was performed using Bowtie [21]. Only uniquely mapping reads were utilized in the analysis. DHS peaks were identified as described in [8] using a false discovery rate threshold of 1%. DNaseI footprints were identified within DNaseI hotspot regions as described in [30] using a false discovery rate threshold of 1%. In cases where more than one biological replicate was available for analysis, the replicate with the highest proportion of uniquely mapped reads falling within DNaseI hotspots [8] was selected for analysis.

Replication Timing analysis (Gilbert group)

Replication timing profiles were smoothed from raw early/late CGH timing values ($\log_2$ Cy3/Cy5 enrichment) for individual probes using loess with a span of 700kb, and timing values were normalized to an equivalent interquartile range between experiments. Replication boundaries were identified as the borders of transitions between relatively earlier and later replicating regions (timing transition regions; TTRs) with a slope above ±9e-7 RT units / bp. The density of replication boundary positions was determined for each cell type, and significantly aligned boundary positions were defined as those with a density above the 95th percentile of control sets with equivalent numbers of random positions. To determine the level of boundary usage/preservation between cell types, we used methods from the MALDIquant R package to align and consolidate peaks between multiple replicates into a unified set of positions. For comparisons of replication boundary conservation between species, windows ±100 kb from boundary positions were converted between mouse (mm9) and human (hg19) assemblies using UCSC liftOver, using standard parameters. Fractions of overlapping boundary positions were determined using the GenomicRanges package in R. Spearman correlations between RT and epigenetic profiles (Fig. 7i) were calculated from the number of significantly enriched sites (peaks of UW DNase I, Stanford/Yale TFBS, and BROAD histone marks relative to input control profiles) within nonoverlapping 200kb windows.

Prediction of candidate promoters and enhancer from chromatin modification profiles (Ren Lab)

To predict promoters, we relied exclusively on the presence of H3K4me3 [15]. To predict enhancers, we used the following three chromatin marks, H3K4me1, H3K4me3 and H3K27ac. We divided the mouse or human genome into 100 bp bins, and counted the number of reads that fell within each bin. We normalized the tag counts in each bin according to the total number of reads. Input reads were processed in the same way and their normalized signal intensity values were subtracted from the IP tracks. Therefore, the height of each 100 bp bin in genome browser is computed as: *Delta* normalized signal intensity = normalized signal $\text{intensity}_{\text{IP}}$ - normalized signal $\text{intensity}_{\text{input}}$. The normalized signals were used as the input file for enhancer prediction. Then we predicted enhancers with the RFECS method as described in [31]. We only kept the predicted enhancers with probability greater than 0.7 and at least 3Kb away from a known TSS. We also compared the predicted enhancers with DNase I sites and found the median distance is only between 300-400bp (Supplementary Fig. 20).

Annotation of protein-coding and non-coding orthologs in the human and mouse genomes

We first compiled a list of orthologous protein coding genes by combining the results from the Ensembl GeneTrees v65 (http://e65.ensembl.org, [32]) and a novel pipeline (modENCODE) developed specifically for comparisons among the human, mouse and modENCODE projects. The two methods are both based on phylogenetic reconstruction, but they differ in the software implementation and the reference species used. While Ensembl GeneTrees focus on vertebrate genomes, the modENCODE set is built on 12 Drosophila, 5 Caenorhabditis, the human, mouse, and yeast genomes (*Wu et al, in prep*.). Each method provides a list of 1:1, 1:many, many:1 and many:many orthologs. While there are some differences among the list, the large majority of the 1:1 orthologs are identical in both sets. To merge the results, we have decomposed the orthologues into pairs of orthologous genes. The pairs present in both sets are used in the final set. These pairs are supplemented with pair from either set that is further supported by orthology to either rat or dog (*Wu et al., in prep.).* The final set of orthologues contain a larger number of 1:1 and a lower set of many:many relationships, which suggests an increase in the quality of the results (Supplementary Table 13). We have assessed the final set using 3 different criteria: the BLAST e-values for the pairs of orthologous proteins, the bootstrap score for the speciation nodes relating the orthologues and a simple count of human and mouse genes sharing the same gene symbol. All the criteria show an improvement in the final set with respect to either of the individual results (*Wu et al., in prep*.).

Next, to infer orthologies among short ncRNA genes, we used a similar phylogenetic approach as for the protein coding genes (*Pignatelli et al, submitted*.). In short, the genes are classified according to their RFam family [33] and aligned with Infernal [34] and different trees are built with RAxML [35] using the 16 different secondary structure models. In addition, the gene loci including their flanking regions are aligned with PRANK [36] and two further trees are built with TreeBeST [37], namely a Neighbour Joining and a Maximum Likelihood tree. All trees are merged with TreeBeST merging algorithm, producing one final tree per family. The human-mouse short ncRNA orthologues correspond to the release 67 of Ensembl (http://e67.ensembl.org, [32]).

Additionally, to identify the orthologous long non-protein coding RNA genes, we used the PipeR framework developed in [38] and adapted it so that it would report multiple homologues between the human and mouse. The original pipeline was also completed with a synteny analysis module, relying on the ENSEMBL comparative orthology dataset (ENSEMBL 71). Altogether, this allowed us to map the 17,547 human long non-coding RNA (lncRNA) transcripts annotated in Gencode v10 onto the mouse genome. We found 2,327 human transcripts (corresponding to 1,679 genes) homologous to 5,067 putative mouse transcripts (corresponding to 3,887 putative genes)(Supplementary Fig. 3). This result suggests that unlike protein coding genes, only a fraction of lncRNA is constrained enough at the primary sequence level to retain a similarity statistically significant between human and mouse. As a consequence, most lncRNA appear to be species specific, at least from a primary sequence point of view. This observation suggests they may be very distinct from proteins, or other highly structured lncRNA (like ribosomal subunits) that remain highly similar, even at high evolutionary distances, owing to the constraint imposed on the primary sequence by the secondary structure maintenance.

Following the efforts to build an automated set of orthologs by merging two initial sets (Wu et al, in prep.) and the analysis of the co-expression of 1-to-1 orthologs, we have built a list of 80 suspicious pairs of 1-to-1 orthologs (Supplementary table S33). This list includes the cases where the orthologous gene is not the one with a matching gene symbol. Although a few gene symbols are well-known to disagree between these two species, the expectation is that most of the 1-to-1 orthologs between human and mouse should have matching gene symbols. The majority of the suspicious cases involves genes from the same family and reflects the difficulty to differentiate orthologs and close paralogs.

We have curated the list manually by checking the quality of the pairwise alignments. We have also compared this list to a more recent Ensembl release (Ensembl 75, http://e75.ensembl.org) that features the latest Mouse assembly (GRCm38). Although the information from pairwise alignments might be misleading, we have relied on cases where the pairwise alignments are clearly showing one best reciprocal hit in terms of alignment score (BLAST e-value), coverage (% of the protein that is aligned) and identity (% identity between both proteins in the alignment). In some cases, this information is not sufficient and we have looked at the synteny data to elucidate the right pairing.

Based on this analysis, 31 (39%) of the cases were clearly misannotated as 1-to-1 orthologs. These involved difficult corner cases like 5 histone proteins that are too similar to be correctly classified using their protein sequence only; an additional 8 erroneous pairs that are related by partial alignments only; another 5 cases involve genes that appear to be retrotransposed copies in mouse; and 4 annotated pairs are close paralogues instead of orthologues. The rest of the cases are more challenging to classify, as they involve either complex families like the protocadherin beta genes or additional gene copies that could be either recent duplications or dubious annotations in the mouse genome.

There are 10 genes (13%) for which it is very difficult to confirm or reject the 1-to-1 relationship. These typically involve large families of very similar genes. 14 genes (17%) have been deprecated or re-annotated as pseudogenes since the phylogenetic analysis was performed. The remaining 25 cases (31%) appear to be correct 1-to-1 orthologs or at least the closest ortholog for complex relationships. This includes well-known cases like human PGM1 and mouse Pgm2, and human PGM2 and mouse Pgm1. The list also includes 3 cases that are more controversial as one of the they involve a read-through transcript that spans two genes in the other species. We consider these to be correct annotations as the loci are indeed orthologous.
We note that the full list of 1-to-1 orthologs contain nearly 14,000 pairs with matching gene symbols. The list of suspicious 1-to-1 orthologs represents a very small fraction of the predicted pairs and only a 40% (less than 0.25% of the total) could be confirmed as false positives.

Determination of transcribed sequences in the human and mouse genomes

The mouse and human genomes (mm9 and hg19 respectively) were partitioned into genomic domains at the nucleotide level, in an unstranded way. They were first partitioned into genic, intergenic and gap regions based on annotated long transcripts (ensembl v65 for mouse and gencode v10 for human) and assembly gaps of the respective assemblies available at the UCSC genome browser. Genic nucleotides were then further partitioned into exonic and intronic regions, giving priority to exons over introns.

For each species, the RNA-seq genome and genomic domain (exonic, intronic, intergenic) coverage was computed using all RNA-seq contigs with an IDR value lower or equal to 0.1 (hereafter simply called contigs). More precisely, the RNA-seq genome coverage was computed as the fraction of the genomic nucleotides not included in gaps, that overlapped with RNA-seq contigs, and the genomic domain (exonic, intronic, intergenic) coverage as the fraction of the genomic domain nucleotides that overlapped with RNA-seq contigs.

Analysis of splicing patterns

Exon boundaries of human and mouse genes were extracted from the genomic annotations (Gencode v.10 and Ensembl NCBIM37.65). Additionally, novel splice sites predicted from the RNAseq data were included in the analysis if (i) they were supported by non-zero SJ counts in at least 15% of samples and (ii) the other boundary of the supporting SJ was an annotated splice site (i.e., the SJ with both unannotated boundaries were not considered). One-to-one correspondence between human and mouse splice sites was established as follows. Pairwise whole-genome (chain) alignments from the UCSC genome browser database [39] were used to construct (non-unique) projections of human splice sites to the mouse genome and, vice versa, mouse splice sites to human genome. Unique projections were obtained from multiple projections by filtering, where the objective function was to maximize the length of continuously aligned region per each gene [40]. Only splice sites that were mapped uniquely and bijectively (i.e., the human-to-mouse and mouse-to-human projections were mutually inverse as functions) were retained for the comparative analysis. One-to-one correspondence between segments (exons or introns) was such that a human segment was said to be orthologous to a mouse segment if the corresponding boundaries were orthologous (as defined above). The orthology relationship on protein-coding genes that was induced by one-to-one correspondence of splice sites was identical to that in the human-mouse ortholog list in more than 94% gene pairs (Wu *et al.*, in prep.). At that, 2,673 novel human and 4,087 novel mouse exons were predicted when one of the boundaries was unannotated in one of the species but supported by both SJ in RNAseq and bijective mappings. Quantitative splicing estimates were computed by *bam2ssj* software using only annotated SJs and requiring minimum 10 counts in the denominator of the fraction defining ψ_5, ψ_3, θ_5 and θ_3 [40]. Pooled values of ψ_i and θ_i (i=5,3), denoted by ψ and θ, respectively, were used to compute sample statistics. At that, ψ and θ samples with more than 25% missing values were excluded from the analysis. The integrated analysis pipeline, as well as the lists of orthologous exons *etc*. are available at http://genome.crg.eu/~dmitri/splicing_pipelines/

Homology based prediction of the mouse lncRNA complement

The lncRNAs were searched using the homology based strategy reported in [38,41]. It involved using spliced transcripts of Human Long Non Coding RNAs (Gencode 10) as queries, blasting them (17151342) against the mm9 assembly of the mouse genome previously masked for interspersed repeats and mouse specific low complexity regions using repeat masker (http://www.repeatmasker.org). The hits thus obtained were used as anchor points, the surrounding genomic regions were extracted and realigned to the original query using exonerate in split mapping alignment mode, which allows for intron modeling between the query and the extracted genomic sequence. Hits aligned to the query but covered with more than 20% of known ancestral repeats were removed from the analysis in order to avoid the inclusion of non-related sequences.

Gene expression analysis

We performed Principal Component Analysis (PCA) of gene expression on several human and mouse samples. Given the heterogeneity of the data, which include cell lines for human and tissues for mouse, we complemented the human data set with tissue samples from the Illumina Human Body Map project (HBM, www.illumina.com; ArrayExpress ID: E-MTAB-513). To obtain more construable results we selected only samples from related organs between human and mouse. The dataset used to generate Fig. 2a,b and Supplementary Fig. 4a,b consisted of 10 human tissue samples from HBM and 10 matching mouse tissue samples from Cold Spring Harbor Laboratory(CSHL). The final dataset for Fig. 2c,d and Supplementary Fig. 4c,d consisted of 8 and 3 human cell samples from CSHL and Caltech, respectively, 12 human tissue samples from HBM and 15 mouse tissues from CSHL.

The analysis was performed within R with the function prcomp(). The measure used for gene expression is log10-transformed RPKM. A pseudocount of 0.01 was added to genes which were not detected or were detected at inconsistent levels between replicates, when replicates were available (IDR<=0.1). Genes with a maximum RPKM greater than 0.1 in each dataset, namely human cell lines (CSHL and Caltech), human tissues (HBM) and mouse tissues (CSHL), were kept. The PCA in Fig. 2a,c and Supplementary Fig.4a,c was performed on expression values that were normalized across all samples, by centering and scaling the log10-transformed RPKM across samples for a given gene. The PCA in Fig. 2b,d and Supplementary Fig. 4b,d was performed on centered and scaled expression values across each dataset.

Variance decomposition

To assess the contribution of tissue and species to gene expression variation, we used a linear mixed model (LMM). Gene expression was modeled as a function of tissue and the species (considered as random factors). The LMM was implemented in the R package lme4 (*ref)*. To restrict our analysis to matching samples, we used a dataset consisting of 10 human tissue samples from HBM and 10 matching mouse tissue samples from Cold Spring Harbor Laboratory (CSHL). This dataset is the same used to generate Fig. 2a and Extended Data 1c. In the same way, only genes with a maximum RPKM greater than 0.1 in each species were kept. We used log10 (RPKM) to normalize the data and a pseudocount of 0.01 to deal with zero expression values. The restricted maximum likelihood (REML) estimators for the random effects of tissue, species and residual variance were normalized by their sum to give the variance components (Fig. 2b). We selected genes whose fraction of variance explained by tissues or species was in their respective top quartile and higher than the other fraction to cluster the samples in Extended Data 1a,b. Euclidean distance and complete linkage were used for clustering both samples and genes. Extended Data 1a and b show the gene expression values, as described above, for genes with a higher fraction of variance explained by tissues or species, respectively.

Neighborhood analysis of conserved co-expression (NACC) (Mike Beer)

Since comparisons of transcript abundance between mouse and human revealed both species specific differences and similarities, we next developed methods to quantify the conservation of coexpression across the mouse and human samples, focusing on the variation in orthologous transcript levels in each species in a way that did not require precisely matched cell lines, tissues, or developmental stages. For the set of human genes with mouse orthologs, we calculated the correlation between every pair of human genes separately across all human samples (C_{ij}^{Hs}) and the correlation of the orthologs of the pair across all mouse tissues (C_{ij}^{Mm}) and we plot the joint distribution of these quantities in Supplementary Fig. 22a. The tilt of this distribution quantifies the conserved coexpression of all orthologous gene pairs. For a negative control, for each pair of human genes we chose a random pair of mouse genes and show this distribution in Supplementary Fig. 22b. We can quantify the differences between these two joint distributions by calculating the Z-score for the change between the actual and control distributions for each bin, and this Z-score is shown in Supplementary Fig. 22d. The actual distribution (a) has more pairs (positive Z-score) in the upper right and lower left (both pairs positively correlated or both pairs anti-correlated) and fewer pairs (negative Z-score) in the upper left and lower right (where the human pair is correlated, but the mouse pair is anti-correlated or vice versa), demonstrating a statistically significant tendency for both correlated and anti-correlated expression patterns to be preserved between species.

A measure of the conservation of expression for each gene, Δ D, is computed by comparing each gene's non-normalized Euclidean distance to the orthologs of a set of simliarly expressed genes, which we refer to as Neighborhood Analysis of Conserved Co-expression (NACC). For each human gene (the test gene), we define the most similarly expressed set of genes (N=20) across all the human samples as that gene's expression neighborhood. We then quantify the average distance between the mouse ortholog of the test gene and each mouse ortholog of the neighborhood genes across the mouse samples. We can also flip species, and choose a mouse test gene and define a similarly expressed neighborhood in the mouse samples, and calculate the average distance between the orthologs of the test gene and neighborhood genes across the human samples. The average change of the human to mouse and mouse to human calculations is a symmetic measure of the degree of conservation of the co-expression for each gene, ΔD. The distribution of this quantity for each gene is shown in Fig. 2e in the main paper, the blue distribution is the change in the distance relative to the orthologs of the gene's expression neighborhood, the red distribution is the change relative to a random gene set, showing that genes in one species show a strong tendency to be co-expressed with orthologs of similarly expressed genes in the other species. This quantity is a global measure of the significance of the conservation of patterns of co-expression between orthologous genes in human and mouse that does not rely on precise matching of samples by biological origin. Since this is a gene-by-gene measure of expression conservation, we can now quantify which biological processes tend to be more or less conserved between human and mouse. We calculated the average of ΔD over each functional GO category. Since this varies with the size of the set, we calculated the significance of large or small ΔD using the Z-score for finding a set of size N_{GO} with greater than observed ΔD via random sampling. Fig 2f shows this Z-score for each functional GO category, highlighting those biological processes which tend to be more conserved between human and mouse (intracellular and nuclear processes) and those processes that have been less conserved (membrane, transmembrane and signaling receptors, and extracellular, see Supplementary Table 21).

In addition to quantifying the conservation of gene expression patterns, this approach (NACC) can also be applied to quantify the correlation of regulatory element activity across cell-types and tissues in the two species. First, we collected regulatory elements, separately using either H3K27ac chromatin marks or Dnase I signal, and then restricted these sets to clearly orthologous regulatory elements by stringent sequence alignment (>70% nucleotide conservation), and classified them as promoters if the distance to the nearest TSS was <3000bp. This process yielded independent sets of 71513 distal enhancers and 6282 promoters as defined from H3K27Ac signal, and 74713 distal enhancers and 7335 promoters as defined from DNase I signal. For these sets of orthologous regulatory elements then we calculated the correlation in H3K27ac activity or DNase signal activity across all available human and mouse tissues. The distribution of ΔD for chromatin promoters, chromatin enhancers, DNase promoters, and DNase enhancers are shown in Fig. 4a-d. Both the chromatin and DNase signal at promoters show significant conservation, and the distribution of ΔD for promoters is very similar to the distribution calculated based on gene expression levels (Fig. 2e). The distribution of ΔD at distal enhancers also shows conservation in patterns of regulatory element activity. From these distributions we can observe that most of the sequence alignable regulatory elements are conserved in their activity.

To demonstrate the sensitivity and specificity of NACC in detecting both conserved and diverged gene expression programs between mismatched human and mouse samples, we carried the following analysis. We simulated two related species by starting with the mouse expression data analyzed above (30 tissues), and added variation to each gene by adding a fraction of the expression pattern of a randomly chosen gene to generate two classes of genes, a more conserved

NEG class (90%) and a less conserved POS class (10%), with euclidean distance distribution between the two species as shown in Supp. Fig. 6(a). We varied the mean distance between genes in the less conserved class (D_{POS}=10 in (a)) and scaled the entire distribution to this value, generating pairs of species at 1x, .5x, .25x and .1x the observed difference between human and mouse expression data sets, while preserving two distinct classes. We approximately matched the actual human-mouse divergence using two distinct methods. First, we measured the NACC ΔD on the full 30 conditions of the simulated data, as shown in Supp Fig 2(b), which compares quite favorably with the actual human-mouse NACC ΔD in Fig. 2e in the main text. D_{POS}=10 was chosen as 1xHM to match the mean of this distribution to the human-mouse distribution in Fig. 2e. The green distribution (.5xHM) uses D_{POS}=5 and is approximately ½ of the human-mouse divergence. We also evaluated the difference in expression between the two simulated datasets using the joint distribution of pairwise correlation coefficients as in Supp. Fig. 22, as shown in Supp Fig. 6(c). The distribution for D_{POS}=10 (1xHM) compares quite favorably with the actual human-mouse distribution shown in Supp. Fig. 22(a), whereas using .5x the human mouse divergence shows much more conserved coexpression (.5xHM). We then asked whether using NACC on a sampled subset of the full 30 conditions could still distriminate the less conserved genes from the more conserved genes, with the class determined using the euclidean distance across all conditions, but the NACC ΔD using the partially matched samples. We removed up to 20 conditions for each sample independently. We used area under the ROC curve (AUC) (Supp. Fig. 6d) as a measure of the accuracy of NACC to discriminate the NEG and POS class, and we plot AUC vs. the number of overlapping (matched) samples in (Supp. Fig. 6e). If the distance between the simulated species is comparable or even half of the observed human-mouse divergence, NACC can accurately detect the rapidly evolving genes. Only if the distance between the species is quite small (.1xHM) and the number of matched samples is very small, does the accuracy of NACC drop significantly. In the former case the gene's evolution is not sufficient to change the membership of its neighborhood. For larger divergence, as long as most genes are expressed in several tissues/cell-types in the selected panel of mouse/human samples, it is only when there is no overlap of any of the expressed conditions that the NACC neighborhood expression pattern is an unreliable reference. This is quantified in (Supp. Fig. 6f), which shows that the majority of mouse and human gene expression patterns have similar entropy, consistent with expression across many tissues ($S \sim \log(N_C)$) and very few genes are expressed in only one sample ($S \sim \log(1)=0$).

Sensitivity of evolving/conserved GO Categories to condition sampling

To assess the sensitivity of NACC in detecting evolving/conserved GO categories, we recalculated the most rapidly evolving and most conserved GO categories by sampling 100 subsets of the mouse and human conditions with 50% probability of inclusion and performing the NACC analysis in Fig 2d using these subsets. The GO categories with large positive and negative z-scores where very robust to this sampling of conditions, indicating that our sample of human and mouse cell-lines/tissues are diverse enough that our detection of evolving and species specific genes is not strongly biased by our sample selection. We selected all GO categories that were either in the top 10 or bottom 10 of the ranked z-score list, and plot the z-scores of each of these GO categories in Supplemental Fig. 7a. We also list the number of times each of these GO categories was in the top or bottom 10, and their average z-score, in Supplemental Fig. 7(b) and (c).

We developed NACC because of the difficulty or impossibility of analyzing perfectly matched biological samples. But to confirm the conserved and species specific expression detected in our broader ENCODE dataset, we repeated NACC on a limited collection of more closely matched tissues and primary cell types. We mapped CAGE tags from 14 matched primary cells from the Riken Fantom 5 dataset (Nature 507, 462–470 2014) to gencode TSSs (aortic smooth muscle cells, CD4+ T cells, CD8+ T cells, marrow derived mesenchymal stem cells, meningeal cells, cerebral neurons, schwann cells, cerebellar astrocytes, hepatic sinusoidal endothelial cells, hepatocytes, hepatic lipocyte stellate cells, cardiac myocytes, CD19+ B cells, and common myeloid progenitor cells) and used log(raw counts) for expression level, after removing any genes with fewer than 5 tags in mouse or human (13219 orthologous genes detected). For the matched tissues we used the 13 tissues in (Lin et al) (small intestine, sigmoid, spleen, lung, liver, brain ovary, testis, pancreas, heart, kidney, fat, adrenal). As shown in Supplementary Fig. 8a and b, the most significantly conserved or species specific GO categories in the ENCODE samples (highlighted) are also the most significant in the matched datasets, with slightly less significance for vasculature, immune system, and motility, presumably due to the specific tissues/cells selected or the smaller size of the matched sets.

Chromatin state analysis (Hardison group)

Chromatin states from segmentations

ChromHMM [42] was applied on the ChIP-seq data of four histone modifications in 15 mouse cell types/lines and 6 human cell lines to learn a multivariate HMM model for segmentation of mapped genome in each cell type. Specifically, the ChIP-seq mapped reads were first pooled from replicates for each of the four histone modifications (H3K4me3, H3K4me1, H3K36me3, and H3K27me3) in 15 mouse cell types/lines (G1E, G1E-ER4+E2, Erythroblasts, Megakaryocytes, CH12, Mel, ESbruce4, Thymus, Kidney, Small intestine, E14 Whole brain, Liver, Spleen, Heart, and Testis) and 6 human cell lines (GM12878, H1-hESC, HepG2, NT2-D1, NH-A, and K562) separately. These mapped reads were first processed by ChromHMM into binarized data in every 200 bp window over the entire mapped genome, with ChIP input reads as the background control. To learn the model jointly from mouse and human, a pseudo genome table was first constructed by concatenating mouse mm9 table and human hg19 table, then the model was learned from the binarized data in all 21 cell types, giving a single model with a common set of emission parameters and transition parameters, which was then used to produce segmentations in all cell types based on the most likely state assignment of the model. We tried models with up to 15 states and selected a seven-state model as it appeared most parsimonious in the sense that all seven states had clearly distinct emission properties, while the interpretability of distinction between states in models with additional states was less clear (Supplementary Fig. 2b). In addition, we found significant overlap between predicted enhancers and H3k4me1 state from chromHMM segmentation (Supplementary Fig. 21).

The majority of the genome in all human and mouse cell types was in the quiescent state, which is a segmentation category characterized by very low signal for any dynamic histone modification. This was observed using the 25-state models described in [43] and it is true for the 7-state models described here. One would expect some regions that are quiescent in one cell type would be active in another, corresponding to differential gene activation, but some regions could be quiescent in all cell types. We find that about 41% of the mouse genome was in a quiescent state in all the 15 cell types examined (Supplementary Fig. 10). This value appears to be quite different from the fraction of quiescent state in human genome (21%) [43], but the difference could be largely explained by the different stringency for calling that state between the two models. The human model separated a “low signal” state from the quiescent state, but this separation was not resolved in the mouse model. The total coverage of these two states in human genome reaches 44%, which is quite close to 41% in mouse. When the same calculation is done on the three human cell types segmented by the 8-state model, we find 47% is quiescent in all three cell types.

Chromatin state maps

State-maps were plotted to display the distribution of chromatin states on the neighborhoods of human-mouse one-to-one orthologous genes. Each gene (from the TSS to the polyA addition signal) was divided into 50 equal sized intervals, and 10kb regions 5’ and 3’ to each gene were also divided into 50 intervals of 200bp. Each interval was assigned to a chromatin state (based on the state assignment of the boundary nucleotide), and the state map displayed as a series of 150 dots colored by the state assignment. Then these strings of dots were stacked to show the state-map over all human-mouse one-to-one orthologous gene neighborhoods. In each cell type or line, the gene neighborhoods were sorted by the gene transcription levels which were average RPKM values of replicates (data from RNAseq Dashboard). The transcription levels of the genes were shown by white dots in each panel.

Based on the state-maps, first, in each gene neighborhood state-map which is a string of 150 points of states, the number of points in activating states (state 1 to 5) was count to represent the chromatin activeness of this gene neighborhood. We applied two approaches to normalize these activeness values for clustering and PCA. In the first approach, the mean of the chromatin activeness of this gene was calculated in the cell types within each species separately. This mean was then subtracted from the chromatin activeness of the corresponding gene in individual cell types or lines, respectively for mouse and human, which resulted in species-normalized chromatin activeness for each gene neighborhood. In the second approach, the mean subtraction was computed across all cell types/tissues including both species. Hierarchical clustering and PCA were performed using the activeness values normalized by the two approaches respectively.

State maps were also plotted in differentially expressed genes between different lineages or between different stages in the same lineage. The differential expressed genes were identified in mouse cell lines G1E (erythroid progenitor) versus erythroblasts which is a comparison within lineage, and in erythroblasts versus megakaryocytes which is a comparison between lineages. For each pair of comparison, the differentially expressed genes were separated into two groups based on whether they were higher expressed in one cell line or the other, and their chromatin state-maps were displayed. The plot conventions are the same as in the state-maps of all ortholog gene neighborhoods.

The quiescent regions are depleted of transcripts in the major biotypes: protein-coding, non-coding and small non-coding RNAs in all eight tissues examined (Supplementary Fig. 12). As expected, chromatin states with active marks and modifications associated with transcriptional elongation are enriched for these RNAs. These enrichment/depletion calculations are done on a base-wise manner, with each nucleotide assigned as either being represented in RNA or not. We also examined the level of expression in each segmentation class for protein-coding genes. While the genes in the segmentation classes with active marks have higher expression scores, and those with repressive marks have lower scores, the genes in quiescent segments are expressed at low levels (Supplementary Fig. 12). Thus the quiescent zones are depleted of transcripts, and the transcripts that are covered are produced at low levels.

Identification of predicted orthologous cis regulatory elements (Ren and Synder groups)

First we built a one-to-one mapping of human and mouse cis bases derived from reciprocal chained blastz alignments [44] available from the UCSC. Next we created a merged list of predicted promoters, enhancers and TF binding sites (Supplementary Table 5-8). We used program bnMapper to find the orthologous sequences between human and mouse (available at https://bitbucket.org/james_taylor/bx-python/wiki/bnMapper, MCP06).

We used the following command to located the human orthologous sequences for the mouse cis elements:

bnMapper.py all.mouse.enhancers all.mouse.promoters Mouse_TFs_merged multi-tissue.master.mm9.bed.c4 mm9.hg19.rBest.chain -f BED12 -t 0.1 -o mm9_to_hg19_0.5

For Supplementary Fig. 9, we set the parameter –t as 0.5. For the random control, we generated same amount of reads with the same genomic coverage with the program shuffleBed from bedtools software package [23]. We run the simulations 100 times and took the average.

Identification of lineage specific elements (Ren and Synder groups)

We first used a very similar approach as described above to find the orthologous sequences for the mouse cis regulatory elements.

bnMapper.py all.mouse.promoters mm9ToHg19.over.chain -fBED12 -o all.mouse.promoters.hg19

The main difference is that we used a different chain file that allows 1-to-many mapping (mm9ToHg19.over.chain, available from UCSC genome browser). In this case, the cis elements that cannot be mapped to the human genome even with the 1-to-many chains were used as the mouse-specific elements.

Self-Organizing Map Methods:

The SOM was trained as described before (ENCODE Consortium, 2012) with the following modifications. The twelve ChromHMM segmentations in mouse tissues were combined into a single segmentation, using the ERANGE 3.3 partitioning script with a minimum segment size of 200bp, and all histone mark datasets read densities in RPKM were calculated to build a training matrix of 1.56 million segments x 60 ChIP-seq datasets was built using the ERANGE v3.3 buildMatrix.sh script, with a maximum threshold of 100 RPKM and the rescale option. Self-Organizing Maps were trained and analyzed using ERANGE v3.3. For every SOM instance, randomly split the matrix into a training set and scoring set, randomly initialized the toroid map from the training set, and incrementally trained the SOM with map size 30x45 using 156 million timesteps, which is equivalent to about 200 epochs (i.e. the number of times that we have gone through the training dataset), starting with an update radius of 15, and a learning rate of 0.2, both of which decreased exponentially over the course of training. Each segment was assigned to its best matching unit. We selected for analysis the best of 10 trials based on the lowest quantization error on the scoring set, which is defined as the average Euclidean distance of all segments to the prototype vector of their assigned unit.

GWAS SNP entries from the NHGRI Catalog were mapped onto mouse using the UCSC Genome Browser liftOver utility and were then mapped to their corresponding segments. To get the probability values for this test, each unit was analyzed to see the random chance of the GWAS SNPs appearing in that unit. The SNPs fit a binomial distribution because they are assumed to be independent as part of the significance test (SNPs on the same segment were collapsed to insure independence). Because several significance tests were performed simultaneously, a method needed to be selected to control the Familywise error rate using The Šidák version of the Holm-Bonferroni method correction. This method was applied to the probability values calculated for each unit, using the standard significance value of 5%.

Supplementary Figures Legends

Supplementary Figure 1: Functional validation of the candidate promoters and enhancers by a high throughout reporter assay in the mouse ES cells and embryonic fibroblast cells. (**A**) Flowchart of quantitative high throughput reporter assay (for details, please refer to the supplementary method section). (**B**) Percentage of candidate promoters or enhancers that show promoter/enhancer activities by luciferase reporter assay.

Supplementary Figure 2: The genomic distribution of 7 chromatin states in each of the 15 mouse and 6 human tissues or cell types. a, The chromHMM model was learned based on four histone modifications (H3K4me1, H3K4me3, H3K36me3, H3K27me3) from all of the 15 mouse cell lines/tissues plus six human cell lines, and then applied in each cell type to segment the genome into seven states. The fraction of the genome in each state is shown as a pie chart for each cell type. **b**, The chromHMM model was learned with 1 to 15 states. The log(likelihood) of the model output by the program increased as number of states increased, while the extent of increment declined, especially after 5-state model.

Supplementary Figure 3: Homology based identification of mouse lncRNA transcripts. a, The small green circle represents the subset of the human Gencode lncRNA transcripts that could be mapped to the mouse genome. Within this circle, the figure in black represents the number of mouse putative transcripts found, and the figure in white the corresponding number of human Gencode transcripts. **b**, Overlap between the ENSEMBL set of mouse lncRNAs (in green) and the homology based subset (in red). In the intersection, the black number corresponds to ENSEMBL transcripts, while the white number corresponds to homology predicted transcripts.

Supplementary figure 4: Principal Component Analysis (PCA) was performed for RNA-seq data for 10 human and mouse matching tissues (a) and for a set of 20 human cell lines and tissues and 15 mouse primary tissues (b). The analysis in a and b is the same as Fig. 2 a and Extended Data Fig. 1C, and the only difference is that PC2 and PC3 is shown here.

Supplementary Figure 5: NACC analysis for biological replicates in human and mouse.

Supplemental Figure 6: Simulations of ability of NACC to detect rapidly evolving genes in unmatched conditions. We simulated two related species by starting with 30 conditions of mouse expression data, and added variation to each gene by adding a fraction of the expression pattern of a randomly chosen gene to generate two classes of genes, a more conserved NEG class (90%) and a less conserved POS class (10%), with euclidean distance distribution between the two species as shown in (**a**). We varied the mean distance between genes in the less conserved class (D_{POS}=10 in (a)) and scaled the entire distribution to this value, generating pairs of species at 1x, .5x, .25x and .1x the observed difference between human and mouse expression data sets, as measured by NACC ΔD as shown in (**b**), which compares quite favorably with the actual human-mouse NACC ΔD in Fig. 2e. The green distribution (.5xHM) uses D_{POS}=5 and is approximately ½ of the human-mouse divergence. We can also evaluate the difference in expression between the two simulated datasets using the joint distribution of pairwise correlation coefficients as in Supplementary Fig. 22, as shown in (**c**). D_{POS}=10 (1xHM) compares quite favorably with the actual human-mouse distribution shown in Supplemental Fig. 22(a), whereas using .5x the human mouse divergence shows much more conserved coexpression (.5xHM). We then asked whether using NACC on a sampled subset of the full 30 conditions could still discriminate the less conserved genes from the more conserved genes. We removed up to 20 conditions for each sample independently. We used area under the ROC curve (AUC) (**d**) as a measure of the accuracy of NACC to discriminate the NEG and POS class, and we plot AUC vs. the number of overlapping (matched) samples in (**e**). If the distance between the simulated species is comparable or even half of the observed human-mouse divergence, NACC can accurately detect the rapidly evolving genes. Only if the distance between the species is quite small (.1xHM) and the number of matched samples is very small, does the accuracy of NACC drop significantly. This is due to the fact that most genes are expressed in several tissues/cell-types in the selected panel of mouse/human samples, and as quantified in (**f**), the majority of mouse and human gene expression patterns have similar entropy.

Supplemental Figure 7: Species specific and conserved GO categories are robust to random condition sampling. We sampled 100 random subsets of the mouse and human conditions and repeated the NACC analysis of Figure 2d. The z-score of any GO category which was ever in the top or bottom 10 z-scores in any random sample is plotted for each sample in (a). The average z-score and number of times each of these GO categories was in the top or bottom 10 of the ranked z-score list are shown in (b) and (c).

Supplemental Figure 8: Species specific and conserved GO categories detected in matched samples. We repeated the NACC analysis on (a) matched primary cell types (RIKEN) and (b) matched mouse and human tissues. The most significantly conserved or species specific GO categories in the ENCODE samples (circles) are also detected as significant in the matched datasets, with slightly less significance for vasculature, immune system, and cell motility.

Supplementary Figure 9: Conservation of predicted mouse cis regulatory elements in the human genome. To define the orthologous sequences for the mouse cis elements in the human genome, we used program bnMapper and the reciprocal chained blastz alignments. For the random control, we generate same amount of reads with the same genomic coverage and run bnMapper with the same parameter. We run the simulations 100 times and took the average. For more details, please refer to the supplementary method section. * indicates $P < 2.2E\text{-}16$ by chi-squared test.

Supplementary Figure 10: Distribution of chromatin states around the orthologous genes in the human and mouse tissue and cell types. a, Heat maps displaying the distribution of the chromatin states over the neighborhoods of human-mouse one-to-one orthologous genes in 15 mouse cell lines/tissues and four human cell lines. The gene neighborhood intervals were sorted by the transcription levels of the genes in the corresponding cell type, which were shown by white dots. **b**, Distribution of chromatin states in differentially expressed genes. From all the human-mouse one-to-one orthologs, the differentially expressed genes were identified between G1E and ER4.

Supplementary Figure 11: Correlation matrix for the fraction of gene neighborhoods in the quiescent state among seven mouse cell lines and six human cell lines. For the members of orthologous gene pairs in human and mouse, the fraction of their gene neighborhood (TSS - 10kb to polyA site + 10kb) residing in state 6 (quiescent) in each cell type was computed. The matrix of pairwise correlations is shown (Pearson correlation coefficient).

Supplementary Figure 12: Distribution of various classes of transcript in the segmentations. **a**, Enrichment (red) or depletion (blue) of RNA-seq transcript categories ("RNA biotypes") in each state for seven segmentations. **b**, Distribution of expression levels in segmentation states. The level of expression of each protein-coding RNA-seq contig intersecting a protein-coding gene in each state for liver and spleen segmentions was extracted, and the distribution of those values for each state is shown in the box plot.

Supplementary Figure 13: Alignment of replication boundary positions across cell types. For each cell type, the density of replication boundary positions (early and late borders of timing transition regions) are calculated. Boundary positions with significant alignment in each cell type are then identified from the local maxima of the boundary densities, and aligned as follows: 1) all positions are put into a sorted vector, 2) differences in position are calculated between each neighboring boundary, 3) the position vector is divided at the largest gap, 4) repeat step 3 until i) differences in peak positions with the bin mean are below 50kb tolerance, or ii) each bin contains a single peak (adapted from MALDIquant methods).

Supplementary Figure 14: Definitions and distributions of boundary classes. After alignment between cell types (Supplementary Fig. 12), boundaries were further classified into nine categories **a**, according to the orientation of TTRs to either side of each boundary. Consolidation occurs preferentially in classes 3 and 6, which harbor especially small early and late domains with a single boundary. Analysis was focused on class 1 through 6 boundaries since 7-9 are most likely to be false boundary calls. Also, since consensus boundary positions do not precisely align with boundary positions in each individual cell type, the absolute timing transition between early and late boundaries occasionally dropped below the threshold for defining a TTR, resulting in a more stringent filtering of a small fraction of boundaries for preservation and conservation analysis. **b** and **c**, Number of RT boundaries present among the set of all aligned boundary positions for each boundary class and cell type shown.

Supplementary Figure 15: Examples of conserved boundaries of each class. Replication timing profiles for all mouse cell types are plotted for the first eight examples of conserved boundaries (those found with the same orientation and position in 10/11 cell types) for classes 1:6.

Supplementary Figure 16: Mouse and Human Boundary Comparisons and Effect of Domain Consolidation During Development. Pairwise percentage of boundaries conserved between cell types in mouse (**a**) and human (**b**) as a fraction of the number of boundaries in each type, with comparisons from rows to columns. Fewer boundaries are shared between widely divergent cell types (e.g., hESCs to HeLa; 42%) than those in similar states or differentiation intermediates (mESC->EBM3->EBM6->NPC; 77-81%). Note that the fact that only 13% of domain boundaries are constitutive does not contradict constitutive replication timing of approximately 50% of the genome, since constitutively replicated domains can be flanked by developmentally regulated boundaries depending on the RT of their neighboring domains (e.g. boundaries 2 and 5 in Supplementary Figure 14a). (**c**, **d**) Significance of boundary number reduction from hESCs, mESCs. Consolidation was predominantly found for small domains detected as a single boundary with timing transitions to both sides (classes 3 and 6 in Supplementary Figure 14a).

Supplementary Figure 17: Comparison of chromatin features to RT. Spearman correlations between chromatin features mapped in specific cell types and RT. The bottom right panel is an expanded version of Fig. 7I, to show the remaining changes in RT vs. changes in other chromatin features mapped in common between the indicated cell types.

Supplementary Figure 18: The joint distribution of the expression (computed in RPKM by the flux capacitor program) of human genes based on long polyA+ RNAseq data, when mapping is done with the STAR mapper (x axis) versus with the TopHat mapper (y axis). The different plots represent different RNAseq library protocols (dUTP, TruSeq), different cell lines (K562, GM12878) and different bioreplicates.

Supplementary Figure 19: Comparison of DHS sites with TFBS in CH12 and MEL. First we merged all the binding sites from different transcription factors in the same cell line (33 TFs in CH12 and 43 TFs in MEL), and compared the union of the binding sites with DHS in the same cell type. We found 61% and 72% of TFBS in MEL and CH12 overlap with DHS, and 81% and 67% of DHS in MEL and CH12 overlap with TFBS in the same cell type.

Supplementary Figure 20: Median distance between candidate enhancers and nearest DNase I Hypersensative sites.

Supplementary Figure 21: Comparison between chromHMM states with candidate enhancers. **a**. Significant overlap between predicted enhancers and H3k4me1 state from chromHMM segmentation. The closest or overlapping chromHMM state was identified for each of the predicted enhancers. The proportion of the states for each tissue/cell type is shown as a piechart. **b**, Relationship between predicted enhancers and chromHMM states in locations. For each of the predicted enhancers, the closest distance to each of the seven chromHMM states was computed and displayed by empirical cumulative distribution plot.

Supplementary Figure 22: Analysis of conserved of co-expression of gene pairs in the human and mouse genomes. For the set of human genes with mouse orthologs, we calculated the correlation between every pair of human genes separately across all human samples (C_{human}) and the correlation of the orthologs of the pair across all mouse tissues (C_{mouse}) and plot the joint distribution of these quantities in (**a**). The tilt of this distribution quantifies the conserved coexpression of all orthologous gene pairs. For a negative control, for each pair of human genes we chose a random pair of mouse genes (**b**), the Z-score for the differences between the actual and control distributions in (**d**) shows a statistically significant tendency for both correlated and anticorrelated expression patterns to be preserved between species.

Supplementary References

1 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108, doi:10.1038/nature11233 (2012).

2 Parkhomchuk, D. *et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic acids research* **37**, e123, doi:10.1093/nar/gkp596 (2009).

3 Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome research* **21**, 1543-1551, doi:10.1101/gr.121095.111 (2011).

4 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).

5 Stamatoyannopoulos, J. A. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome biology* **13**, 418, doi:10.1186/gb-2012-13-8-418 (2012).

6 Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell* **6**, 479-491, doi:10.1016/j.stem.2010.03.018 (2010).
7 Welch, J. J. *et al.* Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136-3147, doi:10.1182/blood-2004-04-1603 (2004).
8 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).
9 John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature genetics* **43**, 264-268, doi:10.1038/ng.759 (2011).
10 Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature methods* **6**, 283-289, doi:10.1038/nmeth.1313 (2009).
11 Samstein, R. M. *et al.* Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* **151**, 153-166, doi:10.1016/j.cell.2012.06.053 (2012).
12 Sabo, P. J. *et al.* Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nature methods* **3**, 511-518, doi:10.1038/nmeth890 (2006).
13 Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS biology* **6**, e245, doi:10.1371/journal.pbio.0060245 (2008).
14 Ryba, T., Battaglia, D., Pope, B. D., Hiratani, I. & Gilbert, D. M. Genome-scale analysis of replication timing: from bench to bioinformatics. *Nature protocols* **6**, 870-895, doi:10.1038/nprot.2011.328 (2011).
15 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).
16 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
17 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
18 Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome research* **15**, 1451-1455, doi:10.1101/gr.4086505 (2005).
19 Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 19**, Unit 19 10 11-21, doi:10.1002/0471142727.mb1910s89 (2010).
20 Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* **11**, R86, doi:10.1186/gb-2010-11-8-r86 (2010).
21 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
22 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

23 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

24 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

25 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

26 Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology* **26**, 1351-1359, doi:10.1038/nbt.1508 (2008).

27 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813-1831, doi:10.1101/gr.136184.111 (2012).

28 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

29 Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958, doi:10.1093/bioinformatics/btp340 (2009).

30 Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90, doi:10.1038/nature11212 (2012).

31 Rajagopal, N. *et al.* RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS computational biology* **9**, e1002968, doi:10.1371/journal.pcbi.1002968 (2013).

32 Flicek, P. *et al.* Ensembl 2013. *Nucleic acids research* **41**, D48-55, doi:10.1093/nar/gks1236 (2013).

33 Gardner, P. P. *et al.* Rfam: Wikipedia, clans and the "decimal" release. *Nucleic acids research* **39**, D141-145, doi:10.1093/nar/gkq1129 (2011).

34 Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337, doi:10.1093/bioinformatics/btp157 (2009).

35 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690, doi:10.1093/bioinformatics/btl446 (2006).

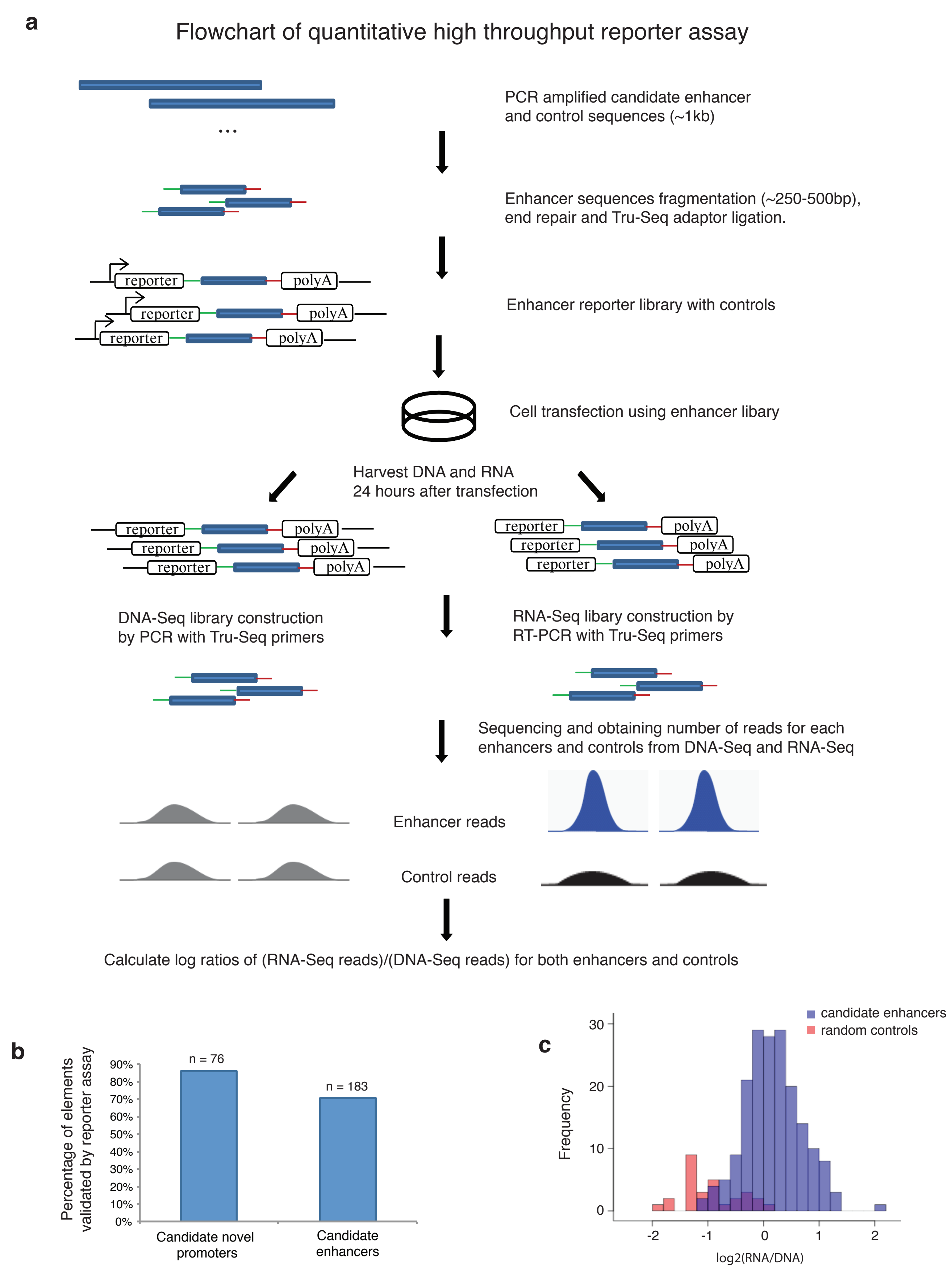
36 Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635, doi:10.1126/science.1158395 (2008).

37 Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic acids research* **34**, D572-580, doi:10.1093/nar/gkj118 (2006).

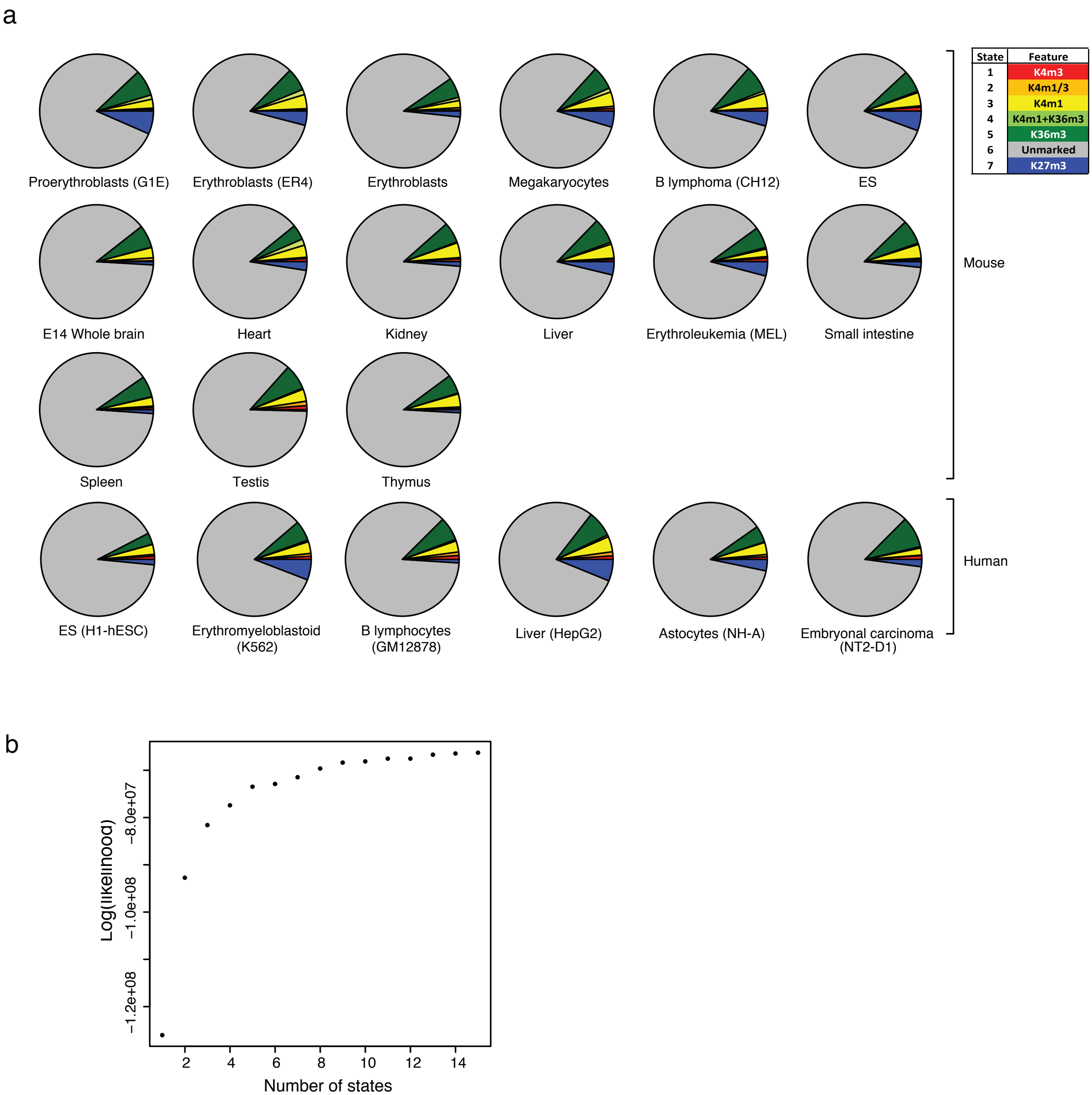
38 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775-1789, doi:10.1101/gr.132159.111 (2012).

39 Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic acids research* **41**, D64-69, doi:10.1093/nar/gks1048 (2013).

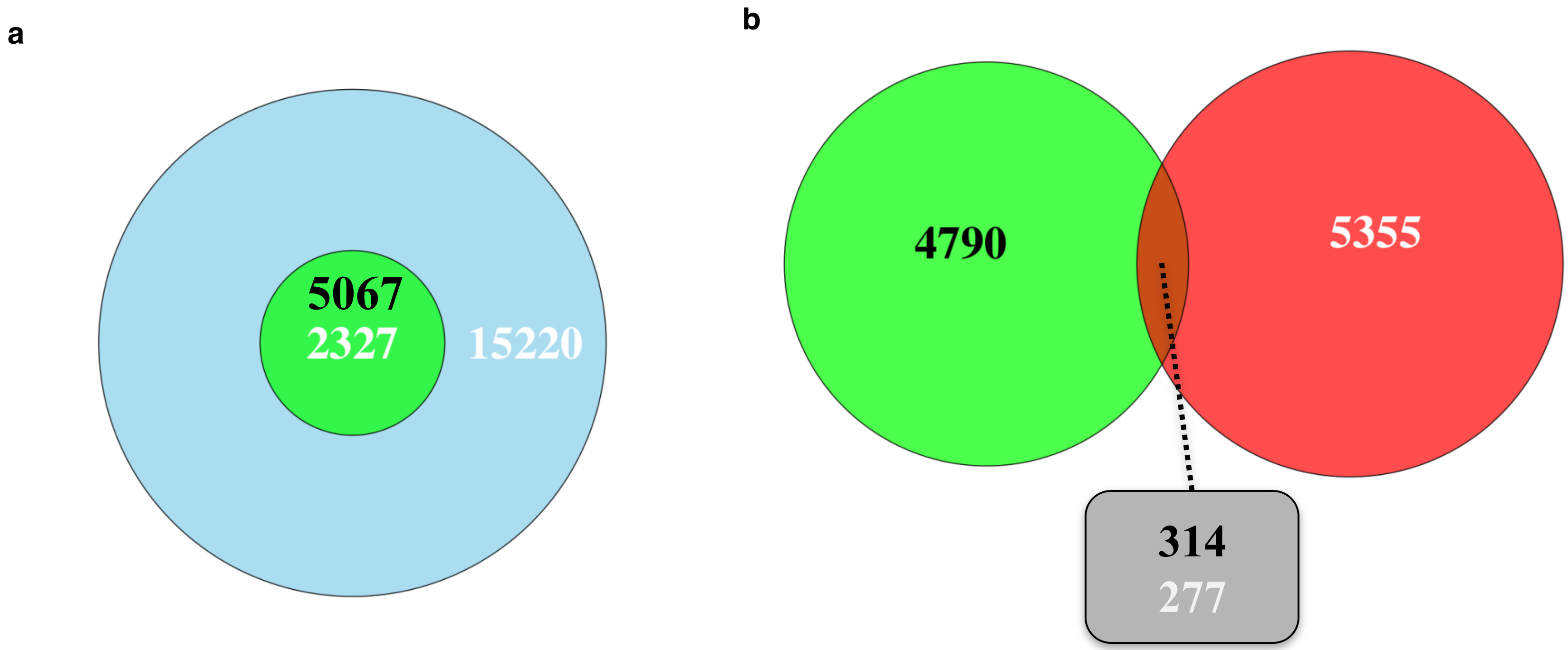
40 Pervouchine, D. D., Knowles, D. G. & Guigo, R. Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* **29**, 273-274, doi:10.1093/bioinformatics/bts678 (2013).
41 Esteve-Codina, A. *et al.* Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. *BMC genomics* **12**, 552, doi:10.1186/1471-2164-12-552 (2011).
42 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
43 Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic acids research* **41**, 827-841, doi:10.1093/nar/gks1284 (2013).
44 Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome research* **13**, 103-107, doi:10.1101/gr.809403 (2003).



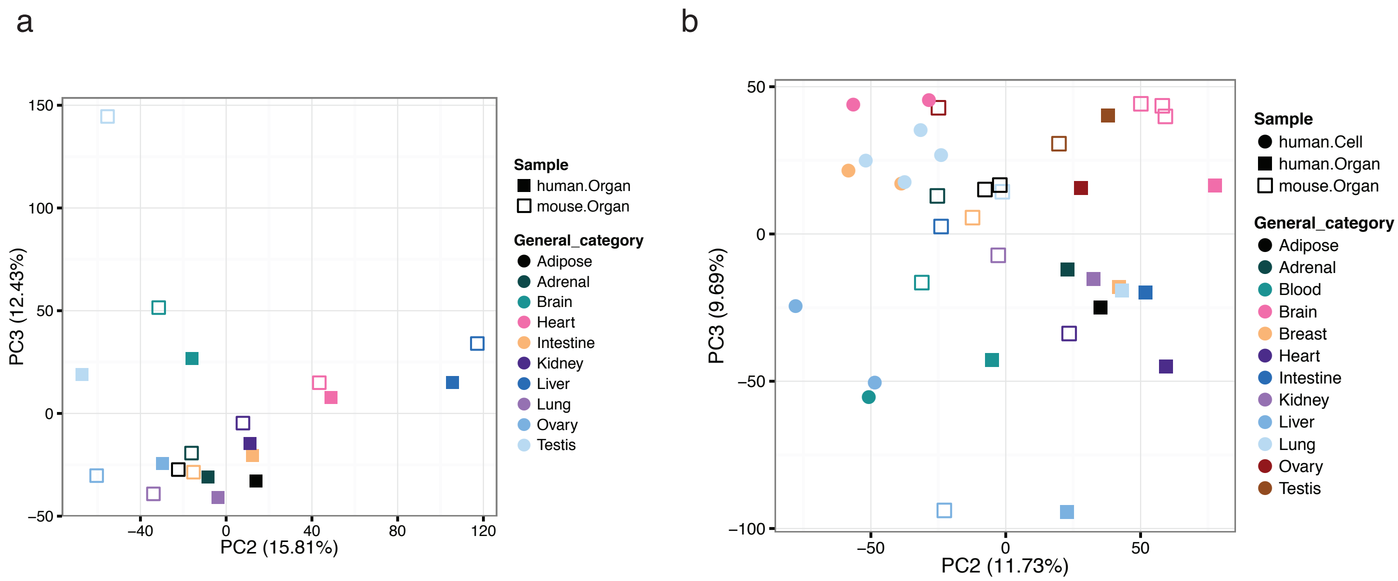
Supplementary Figure 1: Functional validation of the candidate promoters and enhancers by a high throughout reporter assay in the mouse ES cells and embryonic fibroblast cells. **a**, Flowchart of quantitative high throughput reporter assay (for details, please refer to the supplementary method section). **b**, Percentage of candidate promoters or enhancers that show promoter/enhancer activities by luciferase reporter assay. (C) Histogram showing the distribution of ratio of RNA/DNA for the enhancer assay.



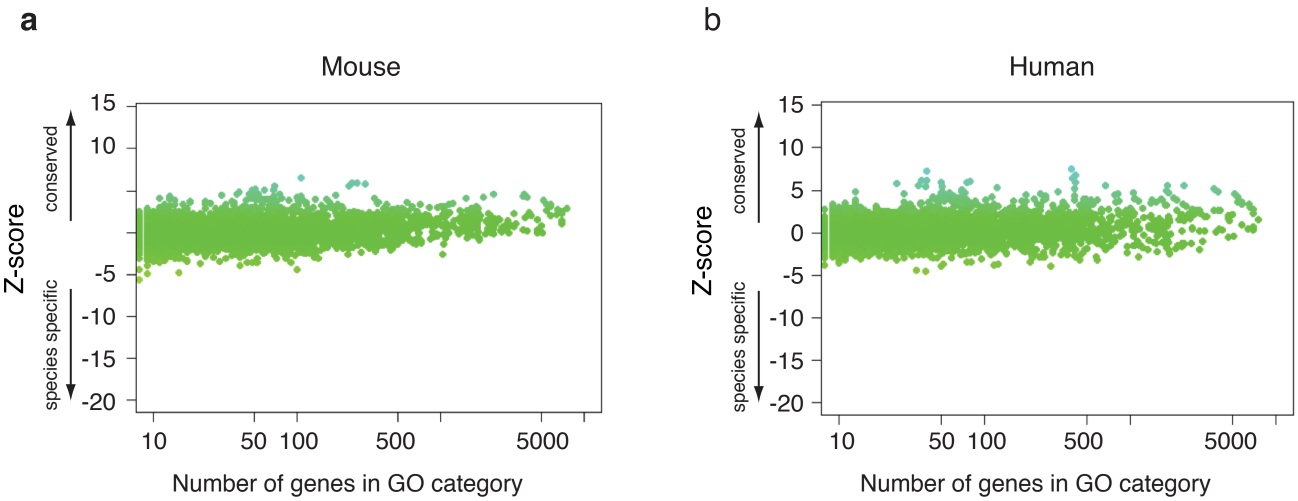
Supplementary Figure 2: The genomic distribution of 7 chromatin states in each of the 15 mouse and 6 human tissues or cell types. The chromHMM model was learned based on four histone modifications (H3K4me1, H3K4me3, H3K36me3, H3K27me3) from all of the 15 mouse cell lines/tissues plus six human cell lines, and then applied in each cell type to segment the genome into eight states. **B.** The chromHMM model was learnt with one to 15 states. The log(likelihood) of the model output by the program increased as number of states increased, while the extent of increment declined, especially after 5-state model.



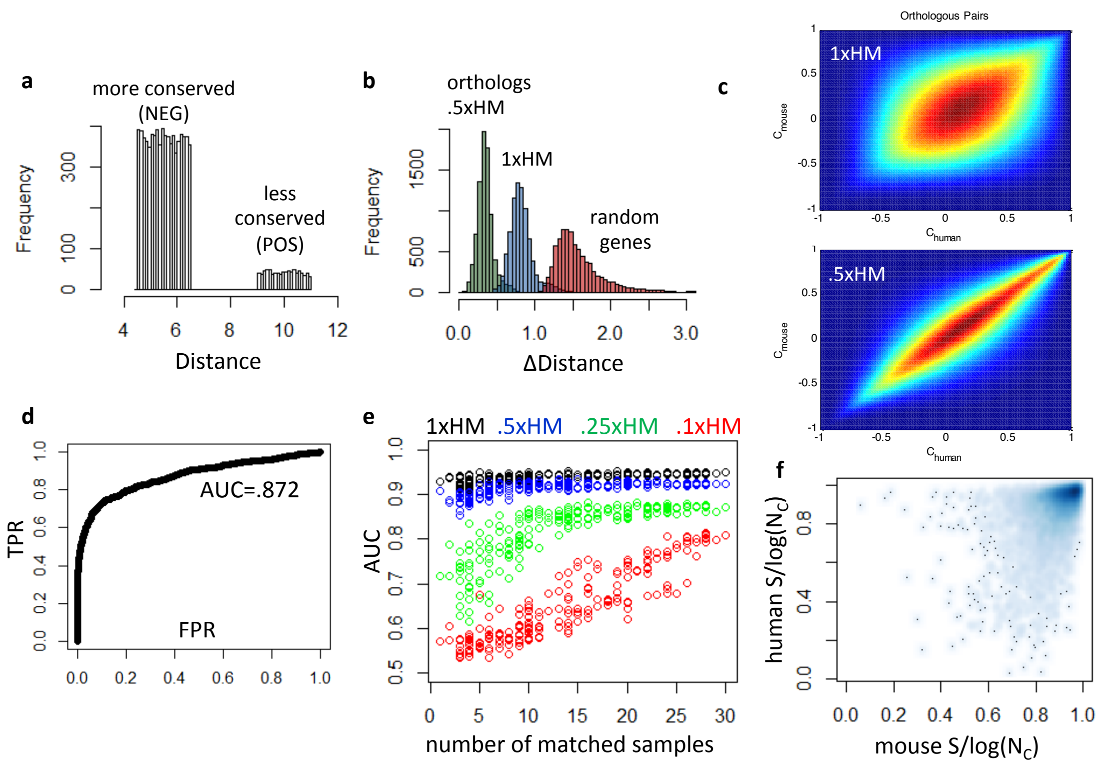
Supplementary Figure 3: Homology based identification of mouse lncRNA transcripts. **a**, The small green circle represents the subset of the human GencodelncRNA transcripts that could be mapped to the mouse genome. Within this circle, the figure in black represents the number of mouse putative transcripts found, and the figure in white the corresponding number of human Gencode transcripts. **b**, Overlap between the ENSEMBL set of mouse lncRNAs (in green) and the homology based subset (in red). In the intersection, the black number correspond to ENSEMBL transcripts, while the white number corresponds to homology predicted transcripts.



Supplementary figure 4: Principal Component Analysis (PCA) was performed for RNA-seq data for 10 human and mouse matching tissues (a) and for a set of 20 human cell lines and tissues and 15 mouse primary tissues (b). The analysis in a and b is the same as Fig. 2 a and Extended Data Fig. 1C, and the only difference is that PC2 and PC3 is shown here.

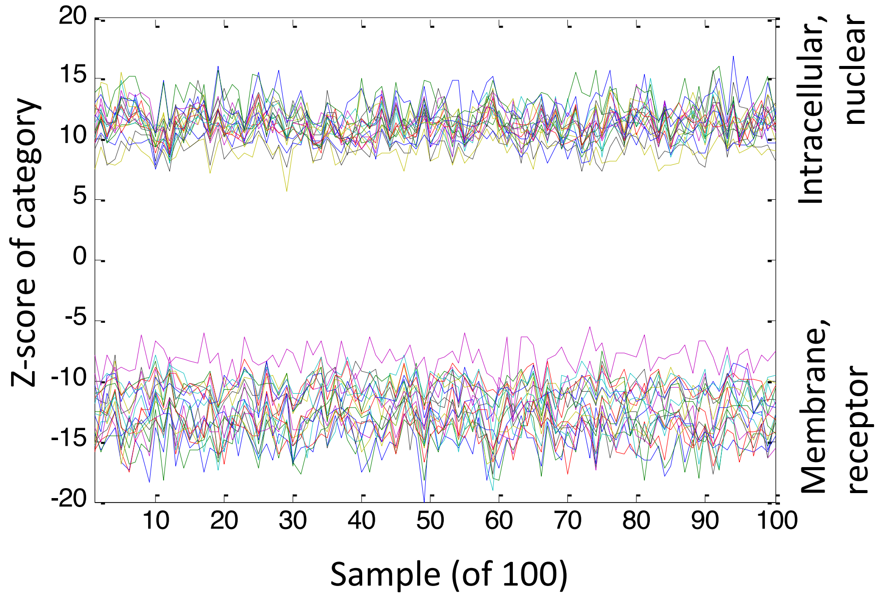


Supplementary Figure 5: NACC analysis between biological replicates in human and in mouse.



Supplemental Figure 6: Simulations of ability of NACC to detect rapidly evolving genes in unmatched conditions. We simulated two related species by starting with 30 conditions of mouse expression data, and added variation to each gene by adding a fraction of the expression pattern of a randomly chosen gene to generate two classes of genes, a more conserved NEG class (90%) and a less conserved POS class (10%), with euclidean distance distribution between the two species as shown in (**a**). We varied the mean distance between genes in the less conserved class (D_{POS}=10 in (a)) and scaled the entire distribution to this value, generating pairs of species at 1x, .5x, .25x and .1x the observed difference between human and mouse expression data sets, as measured by NACC ΔD as shown in (**b**), which compares quite favorably with the actual human-mouse NACC ΔD in Fig. 2e. The green distribution (.5xHM) uses D_{POS}=5 and is approximately ½ of the human-mouse divergence. We can also evaluate the difference in expression between the two simulated datasets using the joint distribution of pairwise correlation coefficients as in Supplementary Fig. 22, as shown in (**c**). D_{POS}=10 (1xHM) compares quite favorably with the actual human-mouse distribution shown in Supplemental Fig. 22(a), whereas using .5x the human mouse divergence shows much more conserved coexpression (.5xHM). We then asked whether using NACC on a sampled subset of the full 30 conditions could still discriminate the less conserved genes from the more conserved genes. We removed up to 20 conditions for each sample independently. We used area under the ROC curve (AUC) (**d**) as a measure of the accuracy of NACC to discriminate the NEG and POS class, and we plot AUC vs. the number of overlapping (matched) samples in (**e**). If the distance between the simulated species is comparable or even half of the observed human-mouse divergence, NACC can accurately detect the rapidly evolving genes. Only if the distance between the species is quite small (.1xHM) and the number of matched samples is very small, does the accuracy of NACC drop significantly. This is due to the fact that most genes are expressed in several tissues/cell-types in the selected panel of mouse/human samples, and as quantified in (**f**), the majority of mouse and human gene expression patterns have similar entropy.

a



b

Species specific:

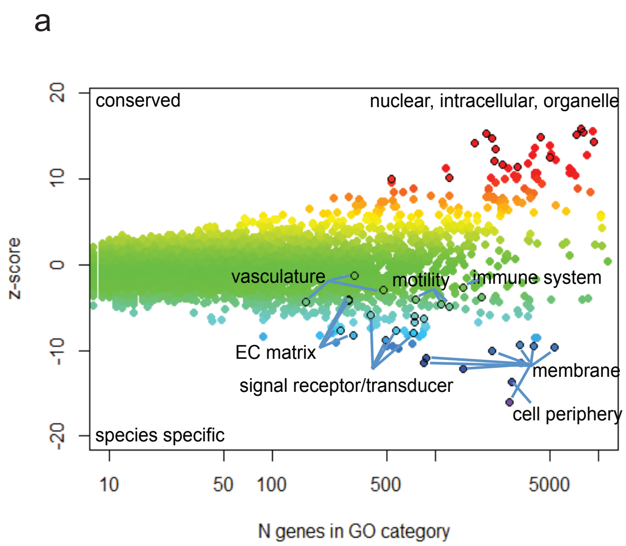
go term	N	<z-score>	
GO:0005886	99	-15.19479	plasma membrane
GO:0071944	100	-15.05556	cell periphery
GO:0005576	100	-14.51038	extracellular region
GO:0004888	94	-13.98638	transmembrane signaling receptor
GO:0044459	94	-13.75342	plasma membrane part
GO:0004872	94	-13.6127	receptor activity
GO:0038023	91	-13.58393	signaling receptor activity
GO:0005887	81	-13.13734	integral to plasma membrane
GO:0031226	85	-13.09087	intrinsic to plasma membrane
GO:0044421	74	-12.9195	extracellular region part
GO:0031012	37	-11.79722	extracellular matrix
GO:0005578	15	-11.06345	proteinaceous extracellular matrix
GO:0031224	10	-10.83791	intrinsic to membrane
GO:0005615	7	-10.80225	extracellular space
GO:0004930	10	-10.78896	G-protein coupled receptor activity
GO:0016021	6	-10.65094	integral to membrane
GO:0060089	1	-10.29409	molecular transducer activity
GO:0004871	1	-10.14543	signal transducer activity
GO:0002455	1	-8.16477	humoral immune response

c

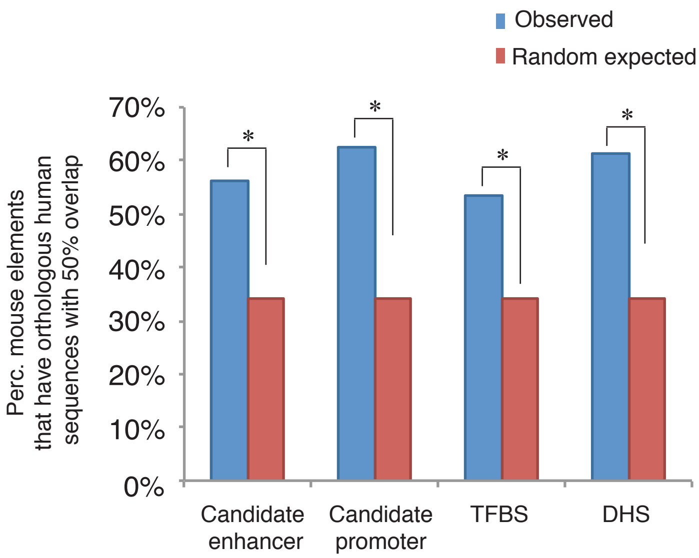
Conserved:

go term	N	<z-score>	
GO:0043170	2	9.09661	macromolecule metabolic process
GO:0005654	4	9.42091	nucleoplasm
GO:0044237	11	9.95219	cellular metabolic process
GO:0043233	33	10.78154	organelle lumen
GO:0031974	48	10.94642	membrane-enclosed lumen
GO:0070013	47	11.04836	intracellular organelle lumen
GO:0031981	44	11.05255	nuclear lumen
GO:0044260	57	11.12734	cellular macromolecule metabolism
GO:0044422	56	11.1508	organelle part
GO:0044424	56	11.16491	intracellular part
GO:0044446	56	11.26187	intracellular organelle part
GO:0005622	63	11.36701	intracellular
GO:0005634	76	11.63825	nucleus
GO:0044428	77	11.78031	nuclear part
GO:0043229	85	11.95167	intracellular organelle
GO:0043226	89	11.96123	organelle
GO:0043227	96	12.87334	membrane-bounded organelle
GO:0043231	100	13.1441	Intracellular membrane organelle

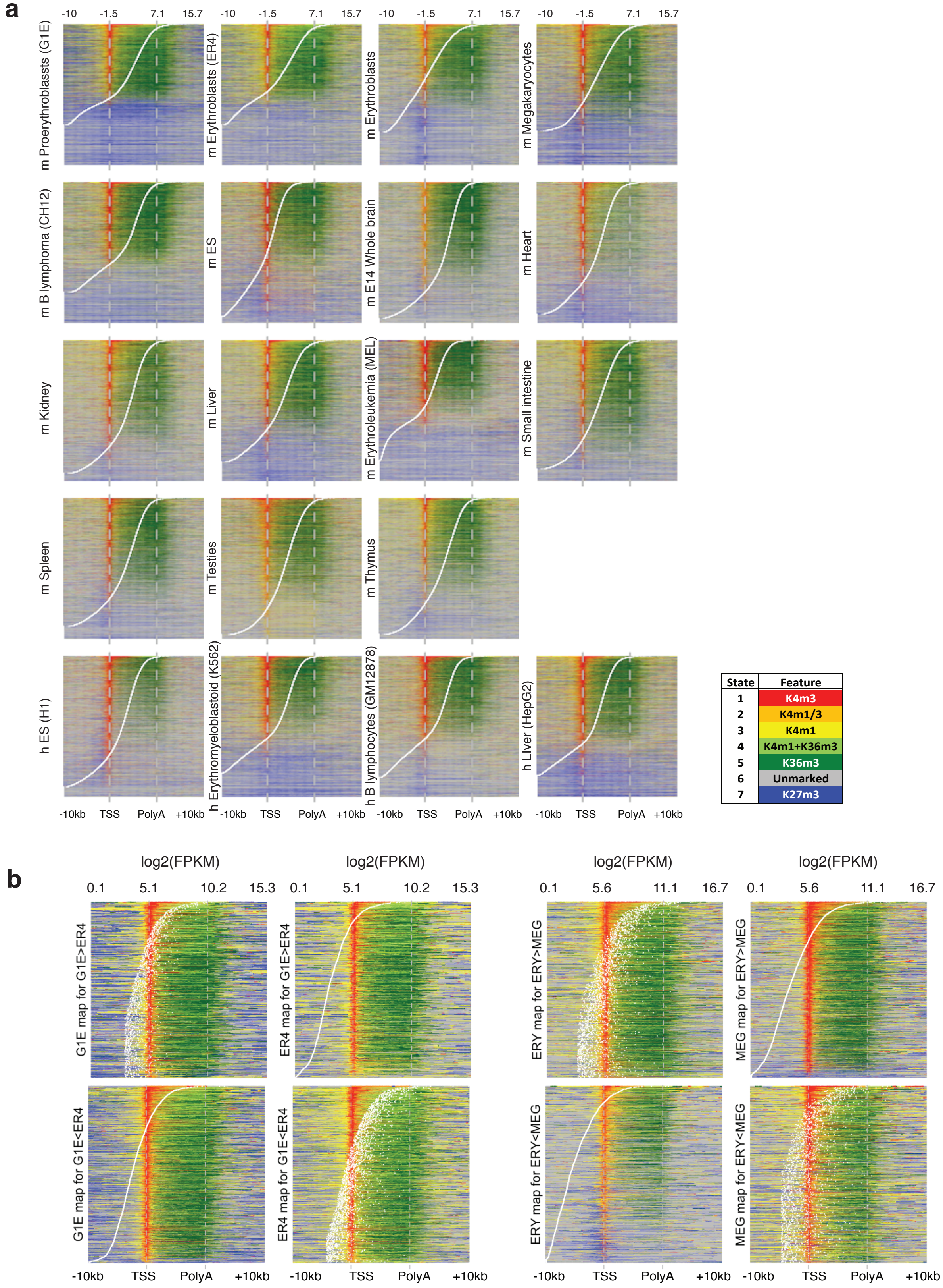
Supplementary Figure 7: Species specific and conserved GO categories are robust to random condition sampling. We sampled 100 random subsets of the mouse and human conditions and repeated the NACC analysis of Figure 2d. The z-score of any GO category which was ever in the top or bottom 10 z-scores in any random sample is plotted for each sample in (a). The average z-score and number of times each of these GO categories was in the top or bottom 10 of the ranked z-score list are shown in (b) and (c).



Supplemental Figure 8: Species specific and conserved GO categories detected in matched samples. We repeated the NACC analysis on (a) matched primary cell types (RIKEN) and (b) matched mouse and human tissues. The most significantly conserved or species specific GO categories in the ENCODE samples (circles) are also detected as significant in the matched datasets, with slightly less significance for vasculature, immune system, and cell motility.



Supplementary Figure 9: Conservation of predicted mouse cis regulatory elements in the human genome. To define the orthologous sequences for the mouse cis regulatory elements in the human genome, we used program bnMapper and the reciprocal chained blastz alignments. For the random control, we generate same amount of reads with the same genomic coverage and run bnMapper with the same parameter. We run the simulations 100 times and took the average. For more details, please refer to the supplementary method section. * indicates $P < 2.2E-16$ by chi-squared test.



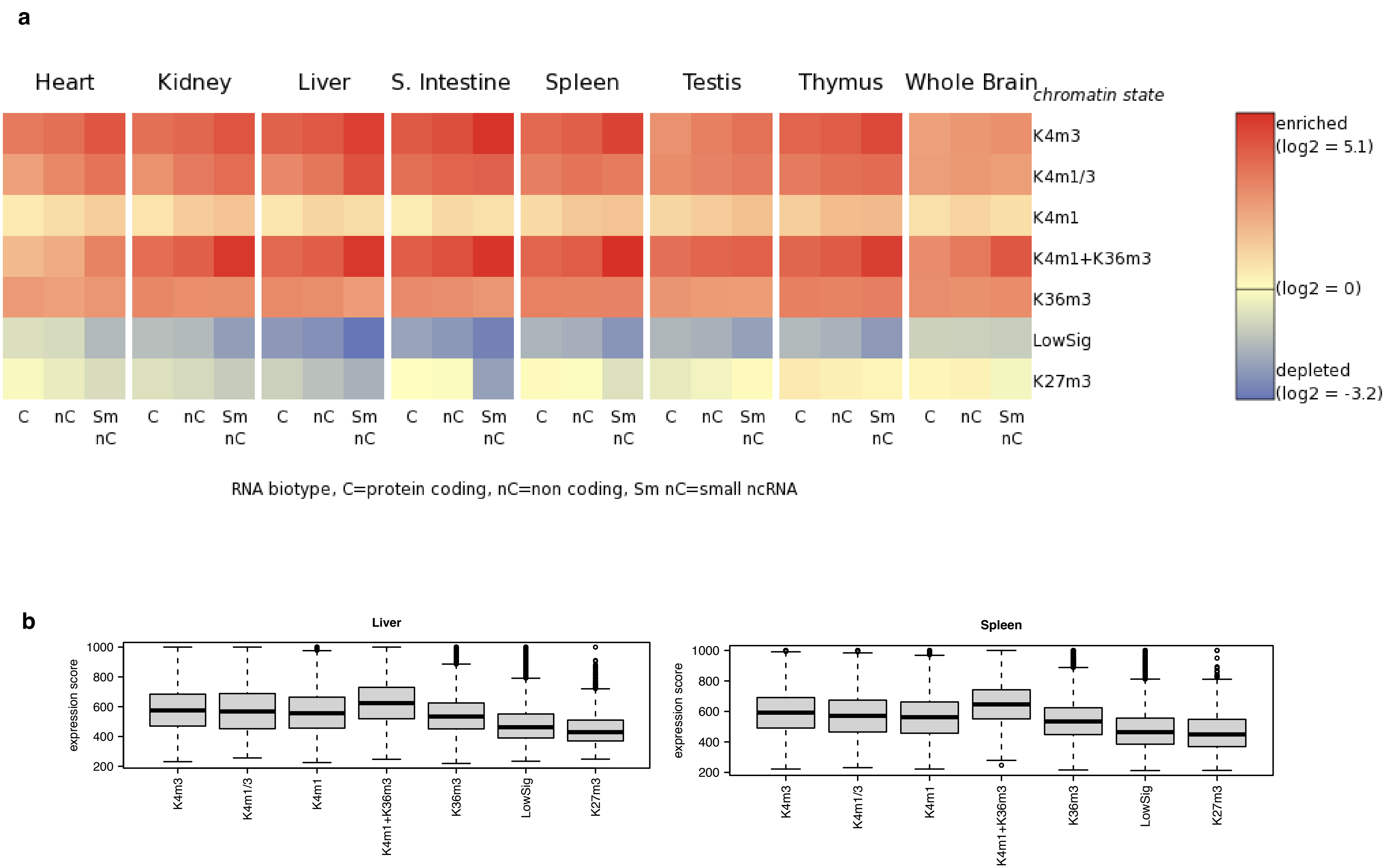
Supplementary Figure 10: Distribution of chromatin states around the orthologous genes in the human and mouse tissue and cell types. a, Heat maps displaying the distribution of the chromatin states over the neighborhoods of human-mouse one-to-one orthologous genes in 15 mouse cell lines/tissues and three human cell lines. The gene neighborhood intervals were sorted by the transcription levels of the genes in the corresponding cell type, which were shown by white dots. **b**, Distribution of chromatin states in differentially expressed genes. From all the human-mouse one-to-one orthologs, the differentially expressed genes were identified between G1E and ER4 (left panel) and between ERY and MEG (right panel).

Raw Coefficient

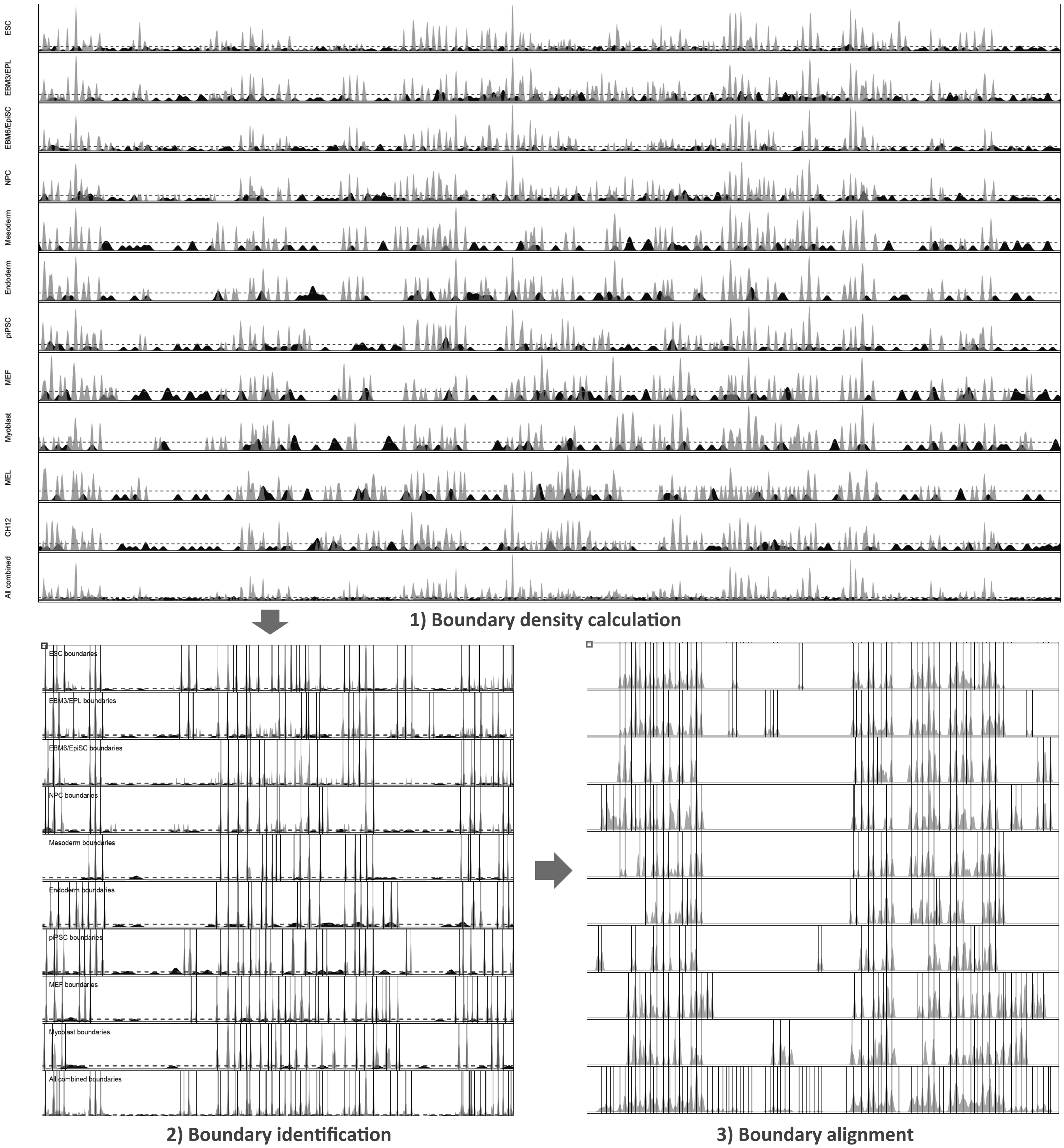
	ER4	MEL	K562	CH12	GM12878	mES	H1-hESC	W. brain	NH-A	Liver	HepG2	Testis	NT2-D1
ER4		0.76	0.59	0.70	0.62	0.64	0.51	0.65	0.50	0.63	0.56	0.70	0.62
MEL	0.26		0.57	0.66	0.61	0.60	0.47	0.61	0.49	0.59	0.49	0.63	0.54
K562	0.17	0.21		0.58	0.68	0.58	0.66	0.57	0.67	0.54	0.68	0.57	0.69
CH12	0.03	0.05	0.10		0.64	0.64	0.52	0.66	0.51	0.64	0.57	0.67	0.63
GM12878	0.09	0.15	-0.11	0.11		0.62	0.68	0.65	0.69	0.60	0.64	0.63	0.72
mES	-0.30	-0.16	-0.08	-0.17	-0.17		0.69	0.80	0.63	0.70	0.58	0.74	0.71
H1-hESC	-0.23	-0.17	-0.34	-0.20	-0.28	0.23		0.71	0.79	0.59	0.69	0.62	0.82
W. brain	-0.38	-0.25	-0.19	-0.18	-0.15	0.23	0.26		0.68	0.74	0.59	0.79	0.74
NH-A	-0.21	-0.04	-0.27	-0.16	-0.23	0.08	0.06	0.17		0.60	0.69	0.59	0.77
Liver	-0.34	-0.18	-0.07	-0.14	-0.09	-0.02	0.03	0.01	0.11		0.59	0.70	0.63
HepG2	0.07	-0.04	-0.12	0.08	-0.29	-0.06	-0.28	-0.12	-0.22	0.10		0.57	0.73
Testis	-0.11	-0.13	-0.03	-0.11	-0.05	0.06	0.07	0.14	0.00	-0.11	-0.04		0.68
NT2-D1	0.03	-0.18	-0.32	0.01	-0.21	0.07	0.06	0.14	-0.18	-0.07	-0.20	0.06	

Normalized Coefficient

Supplementary Figure 11: Correlation of unmarked state proportion in neighborhood of human-mouse 1:1 orthlog genes across tissue/cell types. In each of the ortholog genes, chromatin state was observed on 150 positions over the neighborhood region, and the number of positions that fell in unmarked state was count in each tissue/cell type. The coefficients of Person correlation of these counts between tissue/types were computed and displayed on the right part of the table. The correlation strength is colored in grey scale. In addition, a normalization approach was applied between species. In this approach, the average of each gene's unmarked position counts was calculated within mouse group and human group separately, and then was subtracted from the count in each of the tissue/cell types in each group respectively. The correlation coefficients from these normalized counts are displayed in the left part of the table, in which yellow-light green-green coloring represents increasing correlation coefficient.

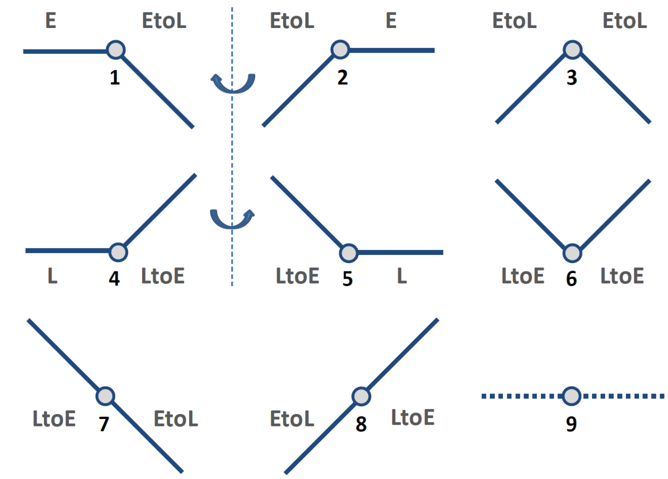


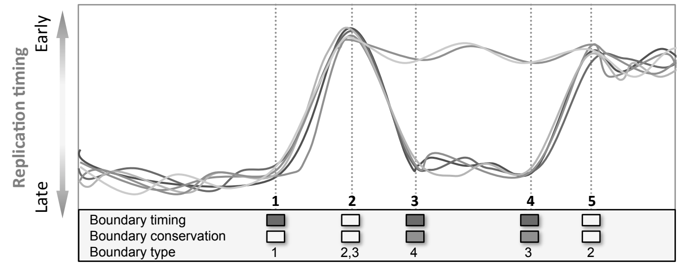
Supplementary Figure 12: Distribution of various classes of transcript in the segmentations. **a**, Enrichment (red) or depletion (blue) of RNA-seq transcript categories ("RNA biotypes") in each state for eight segmentations. **b**, Distribution of expression levels in segmentation states. The level of expression of each protein-coding RNA-seq contig intersecting a protein-coding gene in each state for liver and spleen segmentions was extracted, and the distribution of those values for each state is shown in the box plot.



Supplementary Figure 13. Alignment of replication boundary positions across cell types. For each cell type, the density of replication boundary positions (early and late borders of timing transition regions) are calculated. Boundary positions with significant alignment in each cell type are then identified as shown in Figure 1B and Figure S1, and their positions identified from the local maxima of the boundary densities, and aligned as follows: 1) all positions are put into a sorted vector, 2) differences in position are calculated between each neighboring boundary, 3) the position vector is divided at the largest gap, 4) repeat step 3 until i) differences in peak positions with the bin mean are below 50kb tolerance, or ii) each bin contains a single peak (adapted from MALDIquant methods).

A





Boundary Classes:

(1) Relatively early left, early-to-late TTR right
(2) Early-to-late TTR left, relatively early right
(3) Early-to-late TTR to left and right
(4) Relatively late left, late-to-early TTR right
(5) Late-to-early TTR left, relatively late right
(6) Late-to-early TTR to left and right
(7) Early-to-late TTR to left and right
(8) Late-to-early TTR to left and right
(9) No clear TTR to either side

B

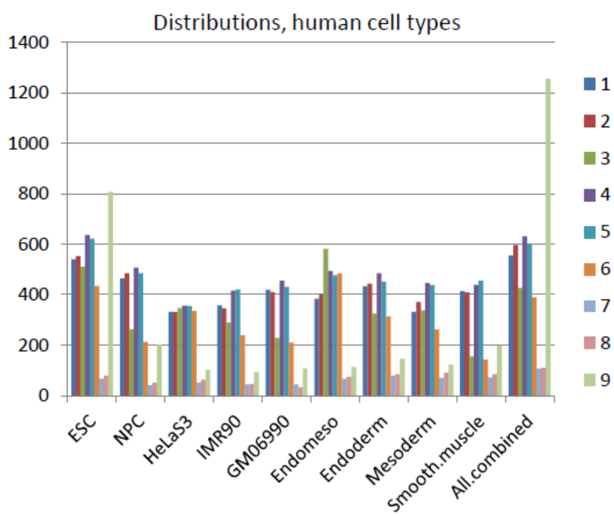
Human cell types

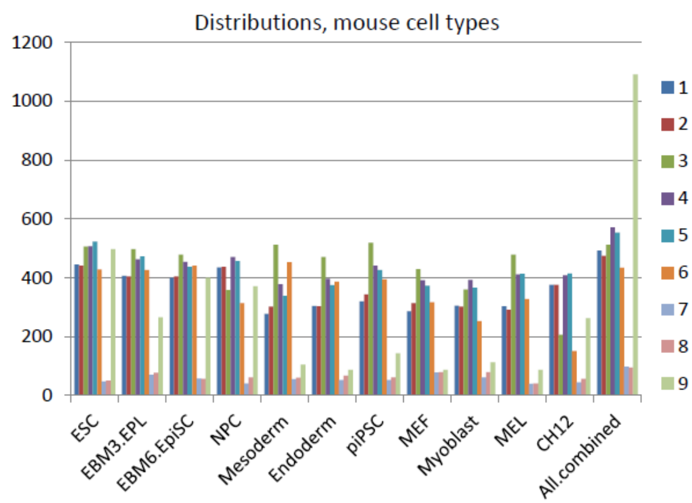
Class	ESC	NPC	HeLa-S3	IMR90	GM06990	Endomesoderm	Endoderm	Mesoderm	Smooth muscle	All combined
1	540	465	332	357	419	384	432	332	414	555
2	552	485	330	346	410	401	443	371	409	596
3	510	263	347	289	228	581	324	337	154	427
4	637	506	356	416	455	494	485	446	439	631
5	621	485	355	421	431	476	451	437	455	602
6	434	213	335	239	210	484	314	261	143	390
7	66	40	51	44	44	67	79	70	72	107
8	79	51	63	45	33	74	85	90	84	110
9	807	203	102	94	108	113	145	122	197	1255
NA's	1620	3155	3595	3615	3528	2792	3108	3400	3499	1193

Mouse cell types

Class	ESC	EBM3/EPL	EBM6/EpiSC	NPC	Mesoderm	Endoderm	piPSC	MEF	Myoblast	MEL	CH12	All combined
1	445	406	400	435	277	304	320	286	305	303	376	492
2	441	405	405	437	302	303	344	314	302	291	376	475
3	506	497	478	358	512	471	519	429	360	478	206	512
4	507	463	454	471	378	396	441	391	392	411	409	571
5	523	472	437	456	339	375	426	373	366	413	415	553
6	428	426	441	314	453	386	395	317	253	328	151	434
7	48	70	58	41	55	53	53	78	61	40	44	98
8	51	77	57	62	60	67	62	79	79	41	56	95
9	497	265	400	371	105	87	144	87	113	87	263	1090
NA's	1924	2289	2240	2425	2889	2928	2666	3016	3139	2978	3074	1050

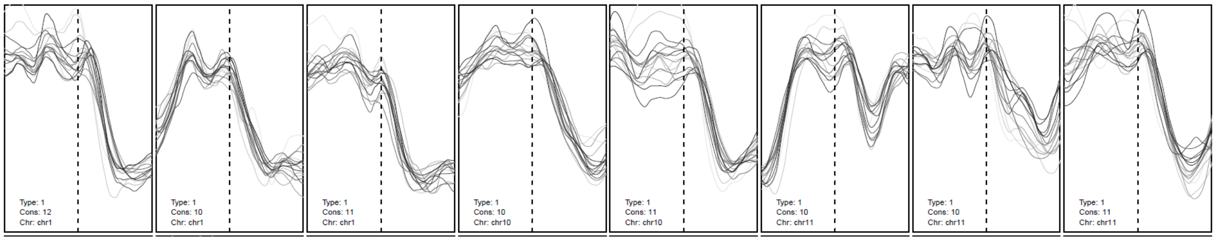
C



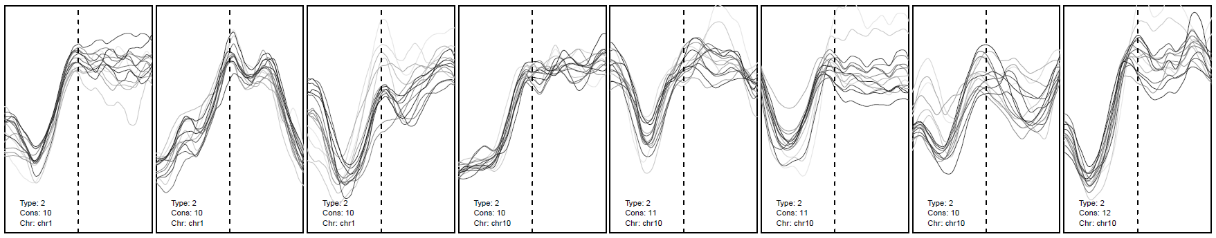


Supplementary Figure 14: Definitions and distributions of boundary classes. After alignment between cell types (Supplementary Fig. 13), boundaries were further classified into nine categories **a**, according to the orientation of TTRs to either side of each boundary. Consolidation occurs preferentially in classes 3 and 6, which harbor especially small early and late domains with a single boundary. Analysis was focused on class 1 through 6 boundaries since 7-9 are most likely to be false boundary calls. Also, since consensus boundary positions do not precisely align with boundary positions in each individual cell type, the absolute timing transition between early and late boundaries occasionally dropped below the threshold for defining a TTR, resulting in a more stringent filtering of a small fraction of boundaries for preservation and conservation analysis. **b** and **c**, Number of RT boundaries present among the set of all aligned boundary positions for each boundary class and cell type shown.

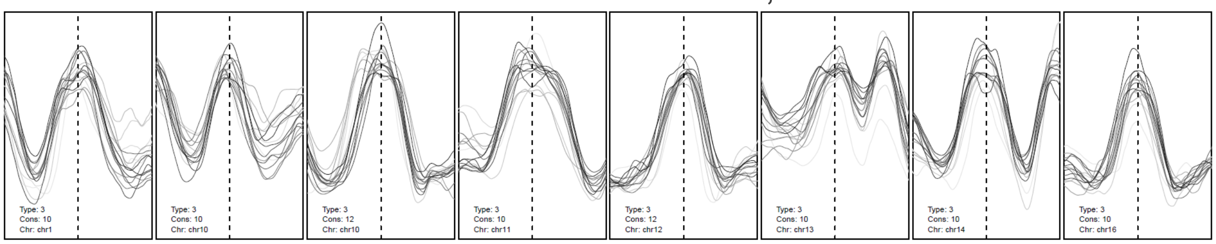
Mouse conserved 10/11 boundaries – **class 1**, n=76/396



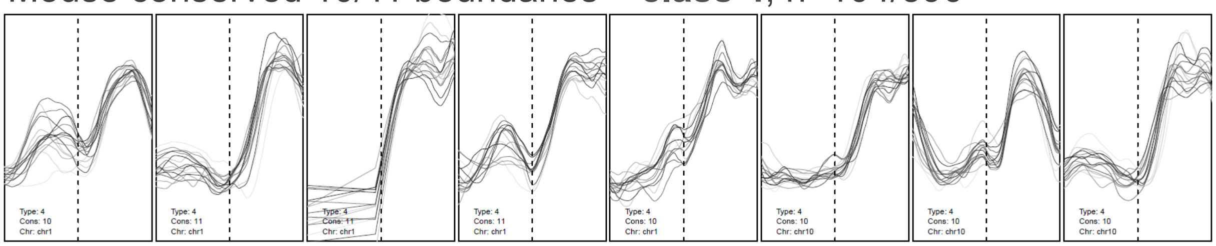
Mouse conserved 10/11 boundaries – **class 2**, n=73/396



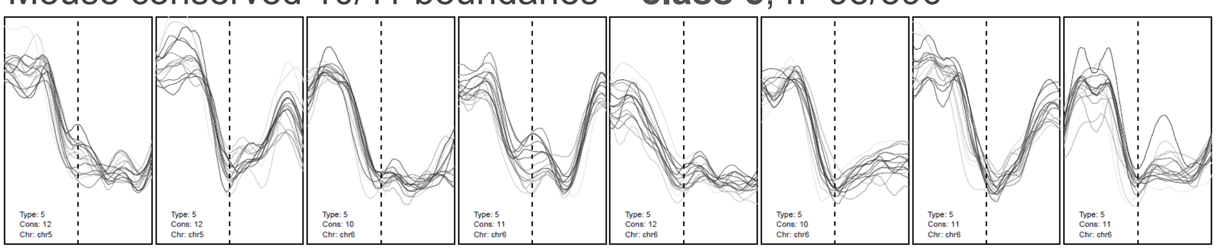
Mouse conserved 10/11 boundaries – **class 3**, n=80/396



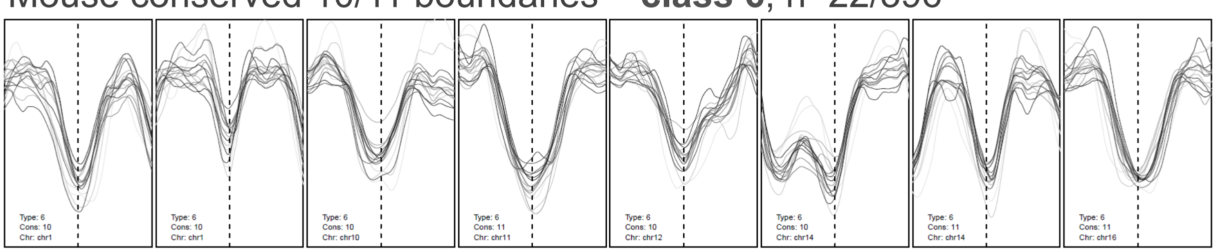
Mouse conserved 10/11 boundaries – **class 4**, n=104/396



Mouse conserved 10/11 boundaries – **class 5**, n=95/396



Mouse conserved 10/11 boundaries – **class 6**, n=22/396



Supplementary Figure 15: Examples of conserved boundaries of each class. Replication timing profiles for all mouse cell types are plotted for the first eight examples of conserved boundaries (those found with the same orientation and position in 10/11 cell types) for classes 1:6.

A

Percentage of overlapping RT boundaries (+/- 50kb):	n	ESC	EBM3/EPL	EBM6/EpiSC	NPC	Mesoderm	Endoderm	piPSC	MEF	Myoblast	MEL	CH12	All combined
ESC	3446	100%	77%	74%	70%	60%	59%	64%	53%	52%	55%	51%	95%
EBM3/EPL	3082	86%	100%	81%	76%	64%	64%	68%	56%	55%	59%	54%	95%
EBM6/EpiSC	3131	81%	80%	100%	78%	67%	65%	69%	57%	55%	59%	54%	96%
NPC	2945	82%	79%	83%	100%	67%	66%	69%	59%	57%	61%	55%	94%
Mesoderm	2481	84%	80%	84%	80%	100%	77%	74%	65%	63%	65%	58%	94%
Endoderm	2443	83%	81%	84%	80%	78%	100%	74%	65%	63%	66%	60%	94%
piPSC	2704	81%	77%	79%	75%	68%	66%	100%	63%	60%	63%	58%	93%
MEF	2354	77%	74%	76%	74%	69%	67%	72%	100%	68%	64%	57%	91%
Myoblast	2231	80%	75%	78%	76%	70%	69%	73%	72%	100%	64%	59%	91%
MEL	2392	79%	76%	77%	75%	68%	68%	72%	63%	60%	100%	61%	91%
CH12	2296	76%	73%	73%	71%	62%	63%	68%	59%	58%	64%	100%	89%
All combined	4322	75%	68%	69%	64%	54%	53%	58%	49%	47%	50%	47%	100%

Mouse

B

Percentage of overlapping RT boundaries (+/- 50kb):	n	ESC	NPC	HeLa	IMR90	Lymphoblast	Endomeso	Endoderm	Mesoderm	Smooth muscle	All combined
ESC	4248	100%	51%	42%	43%	43%	61%	54%	48%	45%	95%
NPC	2712	80%	100%	49%	52%	52%	65%	62%	58%	54%	88%
HeLa	2271	79%	59%	100%	57%	57%	67%	63%	62%	57%	87%
IMR90	2251	81%	62%	58%	100%	57%	70%	65%	63%	61%	89%
GM06990	2339	79%	60%	55%	55%	100%	67%	64%	61%	56%	87%
Endomeso	3075	85%	58%	50%	51%	51%	100%	69%	61%	54%	92%
Endoderm	2759	82%	60%	52%	53%	54%	77%	100%	65%	59%	91%
Mesoderm	2466	83%	63%	57%	57%	58%	76%	73%	100%	65%	92%
Smooth muscle	2368	81%	62%	55%	58%	55%	70%	68%	67%	100%	89%
All combined	4675	86%	51%	43%	43%	44%	61%	54%	48%	45%	100%

Human

C

Mouse

Cell type	Boundaries	ESC Diff.	p-value
ESC	3446	0	NA
EBM3/EPL	3082	364	9.20E-20
EBM6/EpiSC	3131	315	2.20E-15
NPC	2945	501	1.03E-34
piPSC	2704	742	4.83E-70
Mesoderm	2481	965	2.76E-112
Endoderm	2443	1003	2.27E-120
MEF	2354	1092	2.64E-140
Myoblast	2231	1215	3.60E-170
MEL	2392	1054	1.27E-131
CH12	2296	1150	4.72E-154

D

Human

Cell type	Boundaries	ESC Diff.	p-value
ESC	4248	0	NA
Endomeso	3075	1173	1.77E-316
Endoderm	2759	1489	<10E-324
NPC	2712	1536	<10E-324
Mesoderm	2466	1782	<10E-324
Smooth muscle	2368	1880	<10E-324
GM06990	2339	1909	<10E-324
HeLa	2271	1977	<10E-324
IMR90	2251	1997	<10E-324

Supplementary Figure 16: Mouse and Human Boundary Comparisons and Effect of Domain Consolidation During Development. Pairwise percentage of boundaries conserved between cell types in mouse (**a**) and human (**b**) as a fraction of the number of boundaries in each type, with comparisons from rows to columns. Fewer boundaries are shared between widely divergent cell types (e.g., hESCs to HeLa; 42%) than those in similar states or differentiation intermediates (mESC->EBM3->EBM6->NPC; 77-81%). Note that the fact that only 13% of domain boundaries are constitutive does not contradict constitutive replication timing of approximately 50% of the genome, since constitutively replicated domains can be flanked by developmentally regulated boundaries depending on the RT of their neighboring domains (e.g. boundaries 2 and 5 in Supplementary Figure 14a). (**c**, **d**) Significance of boundary number reduction from hESCs, mESCs. Consolidation was predominantly found for small domains detected as a single boundary with timing transitions to both sides (classes 3 and 6 in Supplementary Figure 14a).

H1 hESC	RT.H1ESC
RT.H1ESC	1.00
Rbbp5	0.62
Chd1	0.60
DNaseRep1	0.60
H3k4me1	0.57
H3k9ac	0.55
Hdac2	0.55
H3k4me2	0.55
Sin3	0.54
Ezh2	0.54
Gtf2f1	0.53
H3k27ac	0.53
Chd7	0.53
Tbp	0.51
Znf143	0.50
Mxi1	0.50
Chd2	0.50
Jund	0.50
Ctcf	0.48
Bach1	0.48
Sap3	0.47
Jmjd2a	0.47
H3k4me3	0.47
P300	0.47
Phf8	0.46
H4k20me1	0.45
Cmyc	0.45
Rad21	0.45
Usf2	0.45
Jarid1a	0.43
Sirt6	0.40
H3k36me3	0.39
Cebpb	0.38
Brca1	0.36
Chd1	0.36
Suz12	0.35
Nrf1	0.34
Hdac6	0.34
Plu1	0.30
Rfx5	0.29
H3k27me3	0.28
H3k79me2	0.27
Cjun	0.24
Mafk	0.14
H2az	-0.04
H3k9me3	-0.06

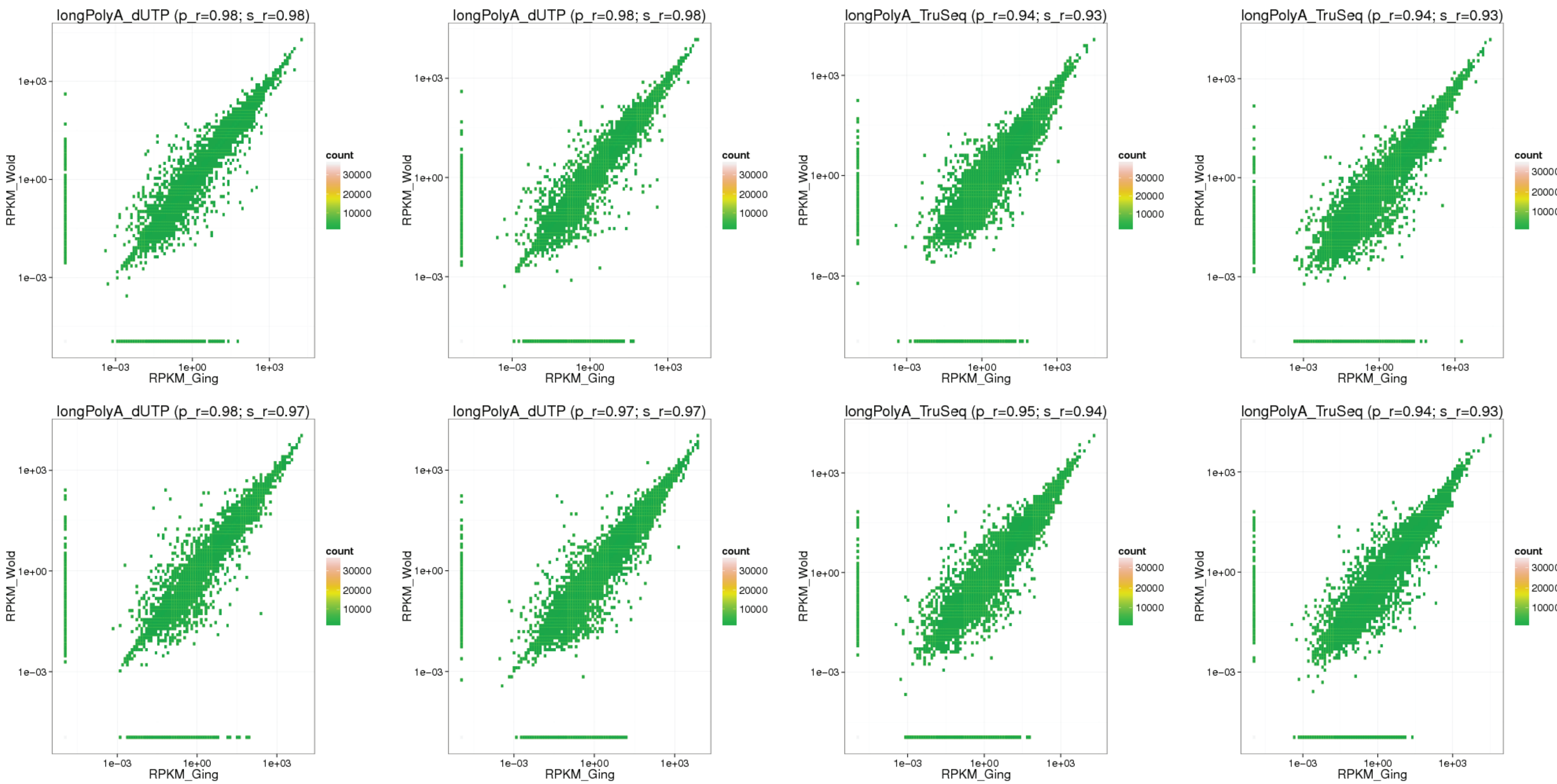
GM12878	RT.GM06990
RT.GM06990	1.00
DNaseRep1	0.71
DNaseRep2	0.69
H3k4me1	0.67
Znf14	0.67
H3k4me3	0.65
H3k4me2	0.64
Stat1	0.64
Smc3	0.64
H2az	0.64
Ctcf	0.63
Ctcf	0.63
Mazab	0.63
Bhlhe40	0.62
H3k27ac	0.61
Whip	0.61
Max	0.61
Ebf1	0.61
Tblr1	0.61
Corest	0.60
Mxi1	0.58
Chd2	0.58
Ezh2	0.57
H3k9ac	0.57
Usf2	0.56
Znf3	0.56
Elk1	0.55
Nrf1	0.55
P300b	0.53
Rfx5200401194	0.53
P300	0.52
Chd1a301218a	0.52
P300	0.50
Tbp	0.50
Cdp	0.49
Pol2s2	0.48
Rad21	0.48
Pol2	0.48
Nfyb	0.47
Sin3	0.46
Stat3	0.43
H4k20me1	0.42
Brca1	0.41
H3k9me3	0.40
Nfe2	0.40
Ikzf1	0.39
E2f4	0.33
Irf3	0.32
H3k79me2	0.32
Erra	0.28
H3k36me3	0.25
Nfkb	0.25
Srebp1	0.21
Gcn5	0.19
Jund	0.18
H3k27me3	0.16
Nfya	0.16
Srebp2	0.12
Tr4	0.12
Yy1	0.08
Spt20	0.06
Mafk	-0.01
Zzz3	-0.07

HeLa	RT.HeLa_S3
RT.HeLa_S3	1.00
Cebpb	0.66
Smc3	0.66
Rad21	0.63
DNaseRep2	0.63
DNaseRep1	0.62
Corest	0.61
Rfx5	0.61
Usf2	0.60
H3k4me1	0.60
Max	0.59
H3k4me2	0.58
Ctcf	0.57
Mafk	0.56
Maz	0.55
H3k4me3	0.55
Chd2	0.54
Mxi1	0.53
Gtf2f1	0.52
Brca1	0.52
H3k27ac	0.52
Znf143	0.52
Zkscan1	0.51
H3k9ac	0.51
Pol2b	0.45
Elk1	0.45
H2az	0.44
H3k79me2	0.41
Tbp	0.41
H4k20me1	0.41
Pol2s2	0.38
Jund	0.35
Nfya	0.34
P300	0.33
Hcfc1	0.29
Nfyb	0.29
Stat3	0.28
Ini1	0.27
Cjun	0.24
Prdm1	0.23
Baf155	0.22
H3k36me3	0.20
Nrf1	0.11
Tf3c110	0.08
Rpc155	0.07
Baf170	0.06
Irf3	0.05
H3k9me3	-0.06
Gcn5	-0.06
H3k27me3	-0.06
Ezh2	-0.07
Brf2	-0.07
Zzz3	-0.14
Bdp1	-0.20
Spt20	-0.21
Brf1	-0.26

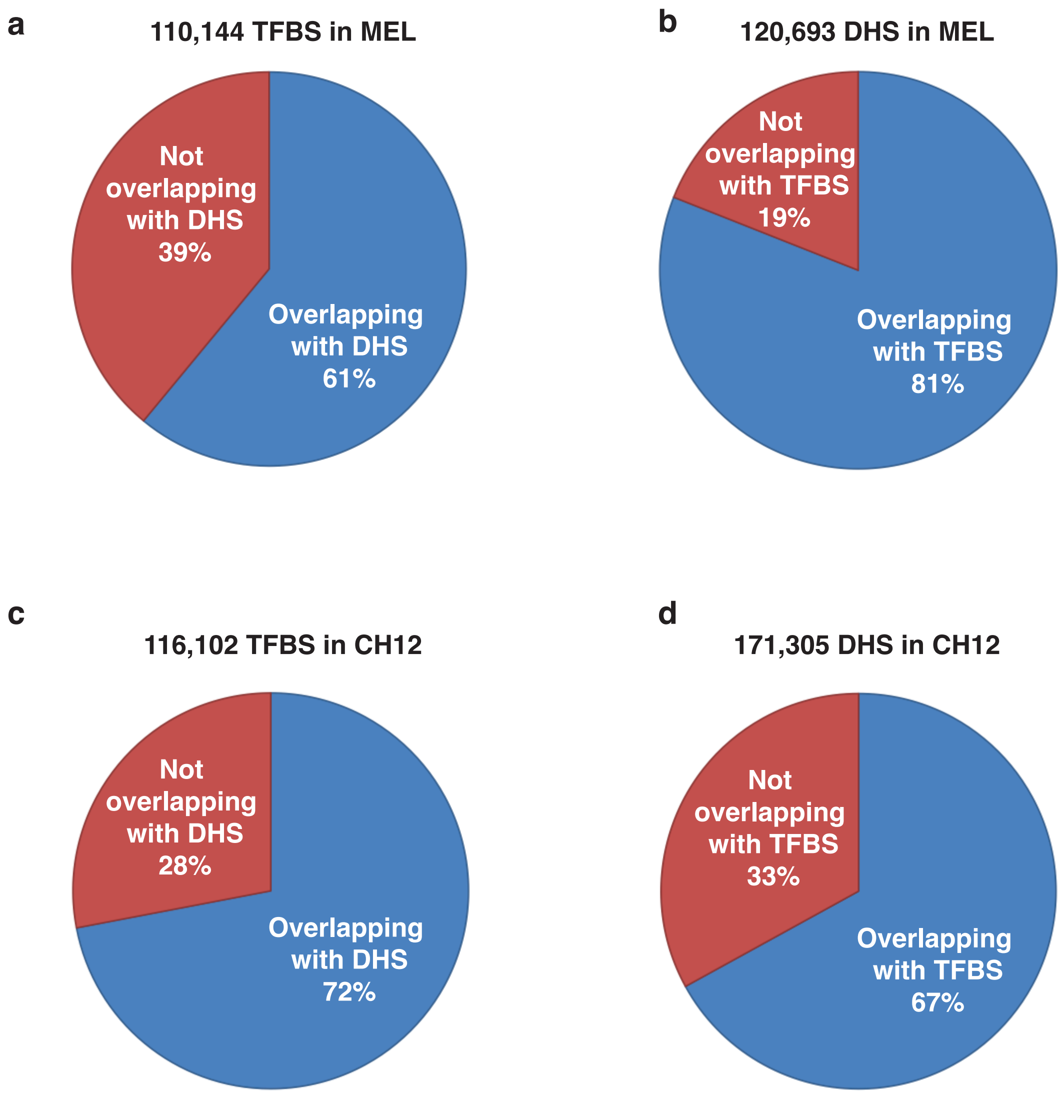
RT vs. Mark Spearman correlation, 200kb windows	Mean R	H1hESC	HeLa	GM06990
DNase	0.64	0.60	0.63	0.69
H3k4me1	0.61	0.57	0.67	0.60
H3k4me2	0.59	0.55	0.64	0.58
Znf143	0.56	0.50	0.52	0.67
Ctcf	0.56	0.48	0.57	0.63
H3k4me3	0.56	0.47	0.65	0.55
H3k27ac	0.55	0.53	0.61	0.52
Chd2	0.54	0.50	0.54	0.58
Usf2	0.54	0.45	0.60	0.56
Mxi1	0.54	0.50	0.53	0.58
Rad21	0.52	0.45	0.63	0.48
H3k9ac	0.51	0.55	0.57	0.41
Rfx5	0.48	0.29	0.61	0.53
Tbp	0.47	0.51	0.41	0.50
P300	0.44	0.47	0.33	0.52
Brca1	0.43	0.36	0.52	0.41
H4k20me1	0.43	0.45	0.42	0.41
Ezh2	0.35	0.54	-0.07	0.57
H2az	0.35	-0.04	0.44	0.64
Jund	0.34	0.50	0.35	0.18
H3k79me2	0.33	0.27	0.32	0.41
Nrf1	0.33	0.34	0.11	0.55
H3k36me3	0.28	0.39	0.25	0.20
Mafk	0.23	0.14	0.56	-0.01
H3k27me3	0.12	0.28	0.16	-0.06
H3k9me3	0.09	-0.06	0.40	-0.06

ΔRT vs. ΔMark Spearman correlation, 200kb windows	Mean R	H1hESC vs. HeLa	H1hESC vs. GM06990	HeLa vs. GM06990
DNase	0.33	0.31	0.26	0.42
H3k4me1	0.31	0.29	0.26	0.38
H3k27ac	0.31	0.31	0.27	0.35
H3k4me2	0.30	0.27	0.27	0.36
P300	0.29	0.32	0.20	0.36
H3k9ac	0.26	0.25	0.25	0.29
H3k4me3	0.26	0.23	0.23	0.32
Tbp	0.23	0.24	0.21	0.25
Chd2	0.22	0.22	0.20	0.26
Znf143	0.19	0.16	0.19	0.23
Rad21	0.19	0.25	0.13	0.20
Mxi1	0.19	0.19	0.17	0.22
Jund	0.19	0.26	0.11	0.20
Usf2	0.18	0.18	0.15	0.21
Ctcf	0.17	0.16	0.17	0.19
H3k36me3	0.17	0.16	0.16	0.18
Rfx5	0.15	0.12	0.10	0.22
Brca1	0.15	0.14	0.11	0.19
H2az	0.14	0.00	0.11	0.30
Nrf1	0.13	0.11	0.10	0.18
Ezh2	0.12	0.09	0.22	0.06
H3k79me2	0.10	0.06	0.18	0.07
Mafk	0.10	0.11	0.08	0.10
H3k9me3	0.10	0.01	0.15	0.13
H4k20me1	-0.02	-0.07	0.11	-0.11
H3k27me3	-0.03	-0.05	0.04	-0.08

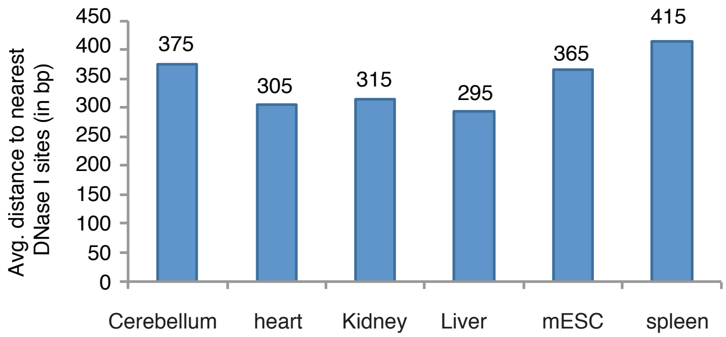
Supplementary Figure 17: Comparison of chromatin features to RT. Spearman correlations between chromatin features mapped in specific cell types and RT. The bottom right panel is an expanded version of Fig. 7I, to show the remaining changes in RT vs. changes in other chromatin features mapped in common between the indicated cell types.



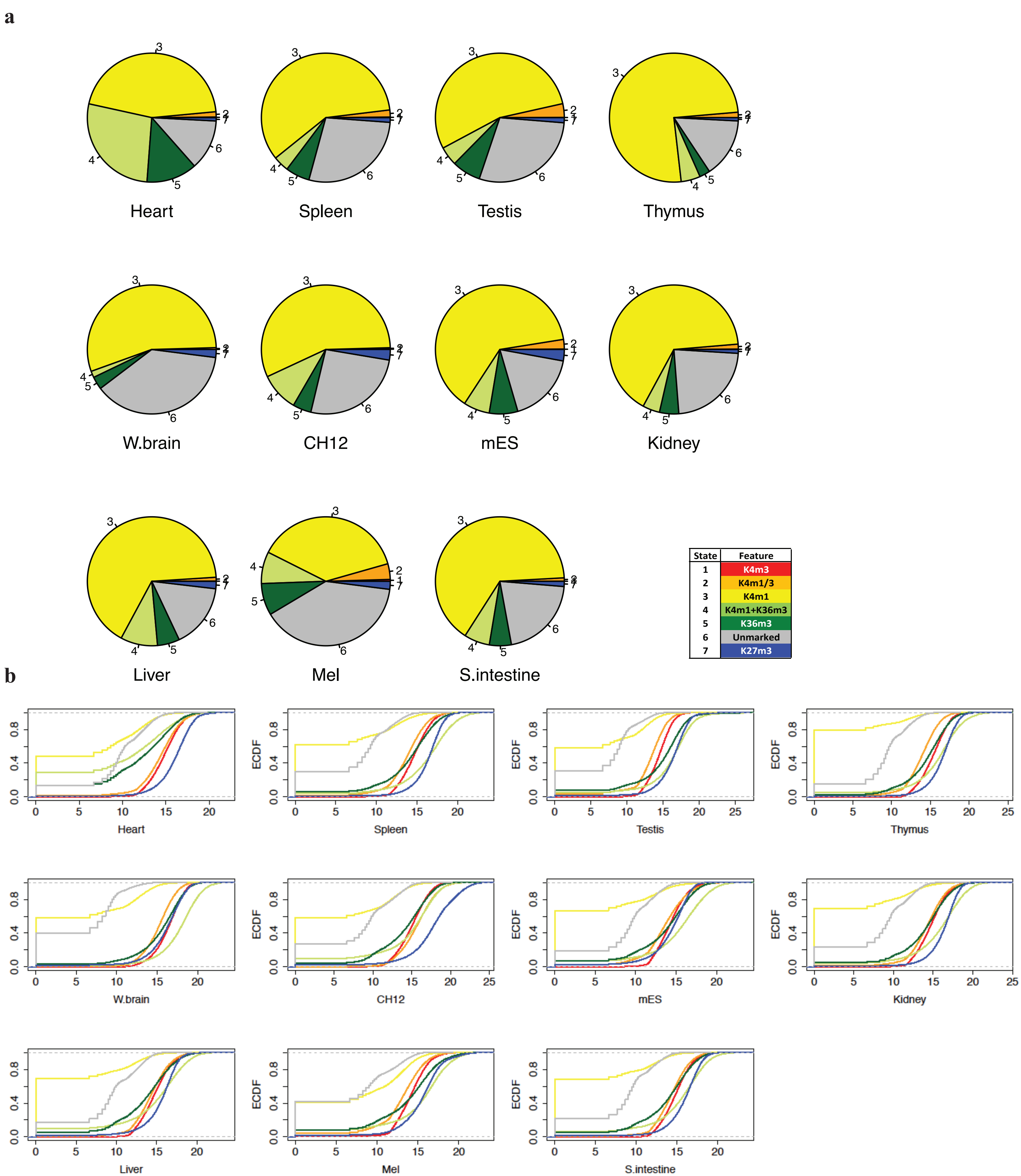
Supplemenary Figure 18: The joint distribution of the expression (computed in RPKM by the flux capacitor program) of human genes based on long polyA+ RNAseq data, when mapping is done with the STAR mapper (x axis) versus with the TopHat mapper (y axis). The different plots represent different RNAseq library protocols (dUTP, TruSeq), different cell lines (K562, GM12878) and different bioreplicates.



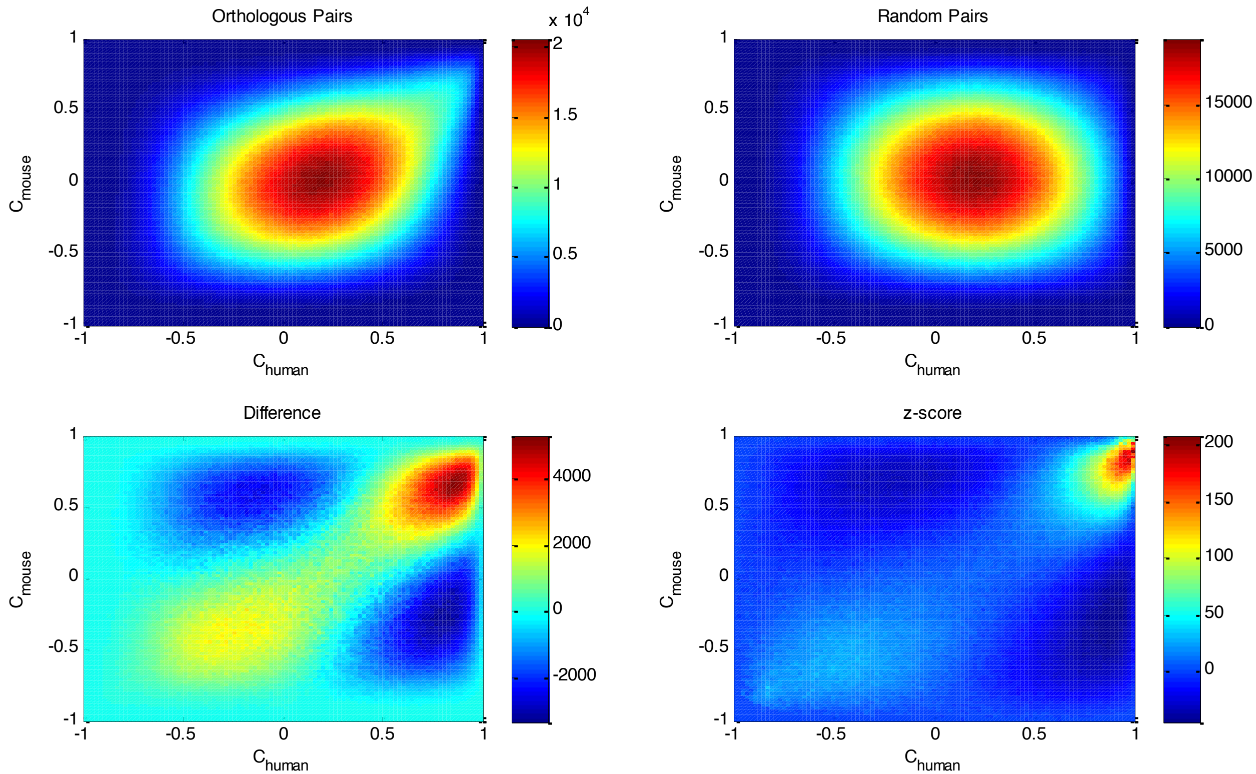
Supplementary Figure 19: Comparison of DHS sites with TFBS in CH12 and MEL. First we merged all the binding sites from different transcription factors in the same cell line (33 TFs in CH12 and 43 TFs in MEL), and compared the union of the binding sites with DHS in the same cell type. We found 61% and 72% of TFBS in MEL and CH12 overlap with DHS, and 81% and 67% of DHS in MEL and CH12 overlap with TFBS in the same cell type.



Supplementary Figure 20: Meidan distance between candidate enhancers based hisone modifications to nearest DHS sites is around 300 - 400 bp.



Supplemenatary Figure 21: Comparison between chromHMM states with candidate enhancers. a. Significant overlap between predicted enhancers and H3k4me1 state from chromHMM segmentation. The closest or overlapping chromHMM state was identified for each of the predicted enhancers. The proportion of the states for each tissue/cell type is shown as a piechart. **b,** Relationship between predicted enhancers and chromHMM states in locations. For each of the predicted enhancers, the closest distance to each of the seven chromHMM states was computed and displayed by empirical cumulative distribution



Supplementary Figure 22: Analysis of conserved of co-expression of gene pairs in the human and mouse genomes. For the set of human genes with mouse orthologs, we calculated the correlation between every pair of human genes separately across all human samples (C_{human}) and the correlation of the orthologs of the pair across all mouse tissues (C_{mouse}) and plot the joint distribution of these quantities in (**a**). The tilt of this distribution quantifies the conserved coexpression of all orthologous gene pairs. For a negative control, for each pair of human genes we chose a random pair of mouse genes (**b**), the Z-score for the differences between the actual and control distributions in (**d**) shows a statistically significant tendency for both correlated and anti-correlated expression patterns to be preserved between species.

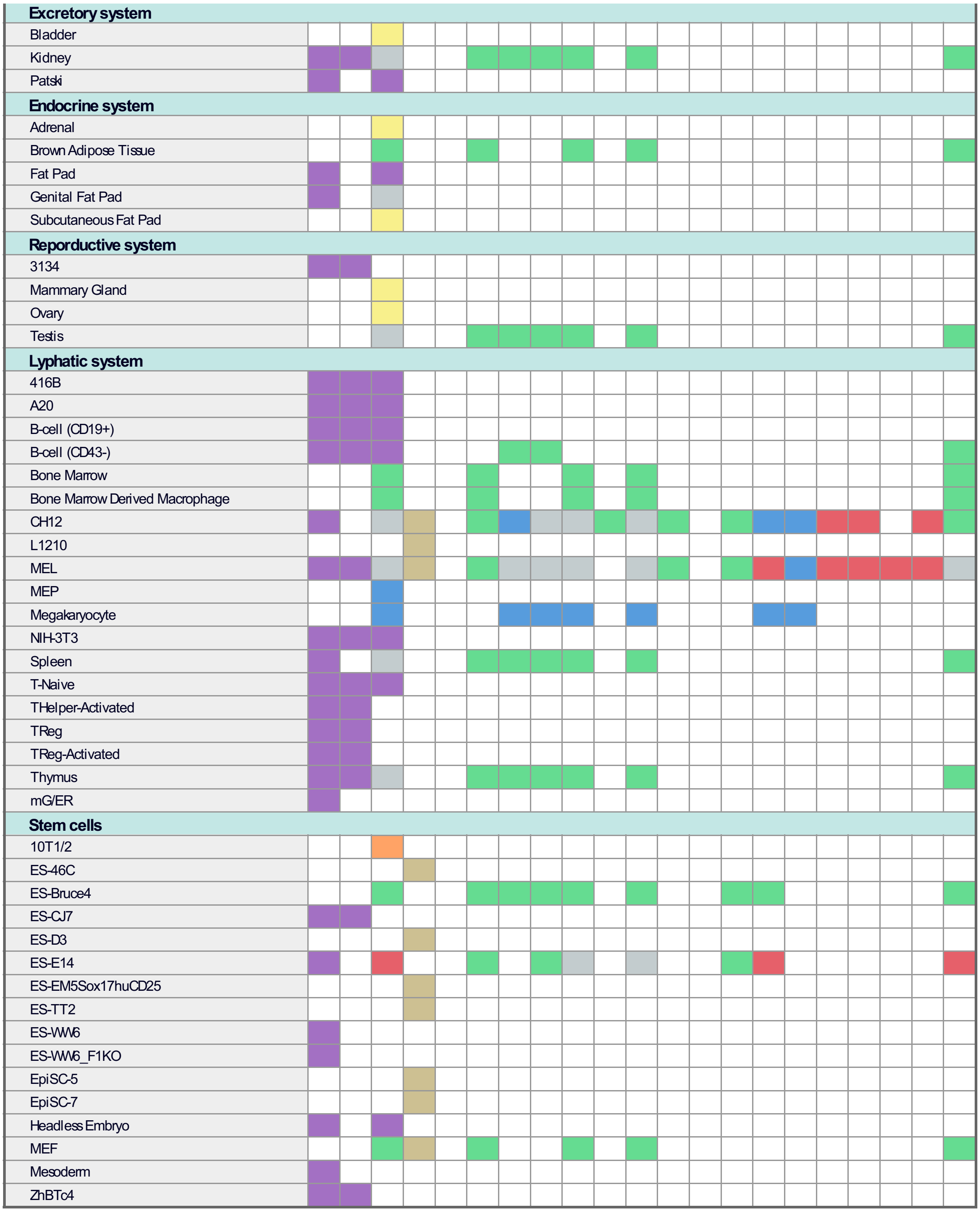
Supplementary Table 1: Experimental datasets used in the current study.

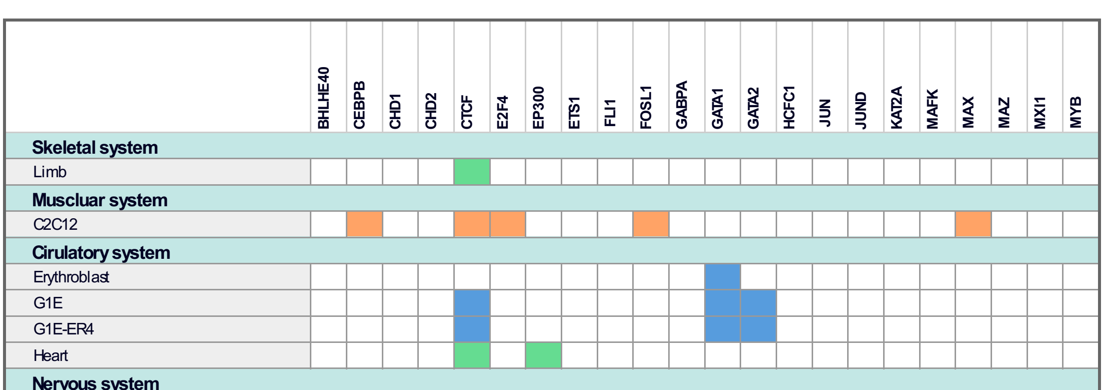
	# of tissue or cell lines	# of experiments	# of data sets
Histone modifications	33	157	310
Transcription factor	29	109	299
RNA-Seq	69	104	193
Dnase-Seq	53	53	123
Replication timing	18	18	33
ChIP controls	34	36	108
Total	123	477	1066

Cells and Tissues

- University of Washington
- Ludwig Institute for Cancer Research
- Pennsylvania State University
- Stanford/Yale
- California Institute of Technology
- Cold Spring Harbor Laboratory
- Florida State University
- Multiple centers

	DNase-seq	DNase-DGF	RNA-seq	Repli-chip	H3ac	H3K27ac	H3K27me3	H3K36me3	H3K4me1	H3K4me2	H3K4me3	H3K79me2	H3K79me3	H3K9ac	H3K9me3	Input (genomic)	Input (IgG-mus)	Input (IgG-rab)	Input (IgG-rat)	Input (IgG-Yale)	Input (Standard)
Skeletal system																					
Fore Limb Bud	University of Washington																				
Hind Limb Bud	University of Washington																				
Limb			Multiple centers			Ludwig Institute for Cancer Research			Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Muscluar system																					
C2C12			California Institute of Technology		California Institute of Technology		California Institute of Technology	California Institute of Technology		California Institute of Technology	California Institute of Technology	California Institute of Technology	California Institute of Technology			California Institute of Technology					
J185a				Florida State University																	
Skeletal Muscle	University of Washington		University of Washington																		
Cirulatory system																					
EPC (CD117+ CD71+ TER119+)	University of Washington																				
EPC (CD117+ CD71+ TER119-)	University of Washington																				
EPC (CD117+ CD71- TER119-)	University of Washington																				
EPC (CD117- CD71+ TER119+)	University of Washington																				
Erythroblast			Pennsylvania State University				Pennsylvania State University	Pennsylvania State University	Pennsylvania State University		Pennsylvania State University				Pennsylvania State University	Pennsylvania State University					
FV-progenitor			Pennsylvania State University																		
FVL-stem			Pennsylvania State University																		
G1E	Pennsylvania State University		Pennsylvania State University				Pennsylvania State University	Pennsylvania State University	Pennsylvania State University		Pennsylvania State University				Pennsylvania State University	Pennsylvania State University					
G1E-ER4	Pennsylvania State University		Pennsylvania State University				Pennsylvania State University	Pennsylvania State University	Pennsylvania State University		Pennsylvania State University				Pennsylvania State University	Pennsylvania State University					
Heart	University of Washington	University of Washington	Multiple centers			Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research							Ludwig Institute for Cancer Research
Placenta			Multiple centers			Ludwig Institute for Cancer Research			Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Nervous system																					
CNS			Cold Spring Harbor Laboratory																		
Cerebellum	University of Washington		Multiple centers			Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Cerebrum	University of Washington		University of Washington																		
Cortex			Multiple centers			Ludwig Institute for Cancer Research			Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Frontal Lobe			Cold Spring Harbor Laboratory																		
Olfactory Bulb			Ludwig Institute for Cancer Research			Ludwig Institute for Cancer Research			Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Retina	University of Washington	University of Washington																			
Whole Brain	University of Washington	University of Washington	Multiple centers			Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research				Ludwig Institute for Cancer Research						Ludwig Institute for Cancer Research
Respiratory system																					
Fibroblast	University of Washington																				
Lung	University of Washington	University of Washington	Multiple centers						Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Digestive system																					
Colon			Cold Spring Harbor Laboratory																		
Duodenum			Cold Spring Harbor Laboratory																		
Large Intestine	University of Washington		Multiple centers																		
Liver	University of Washington	University of Washington	Multiple centers			Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research							Ludwig Institute for Cancer Research
Small Intestine			Multiple centers			Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research	Ludwig Institute for Cancer Research		Ludwig Institute for Cancer Research										Ludwig Institute for Cancer Research
Stomach			Cold Spring Harbor Laboratory																		


Excretory system
Bladder
Kidney
Patski
Endocrine system
Adrenal
Brown Adipose Tissue
Fat Pad
Genital Fat Pad
Subcutaneous Fat Pad
Reporductive system
3134
Mammary Gland
Ovary
Testis
Lyphatic system
416B
A20
B-cell (CD19+)
B-cell (CD43-)
Bone Marrow
Bone Marrow Derived Macrophage
CH12
L1210
MEL
MEP
Megakaryocyte
NIH-3T3
Spleen
T-Naive
THelper-Activated
TReg
TReg-Activated
Thymus
mG/ER
Stem cells
10T1/2
ES-46C
ES-Bruce4
ES-CJ7
ES-D3
ES-E14
ES-EM5Sox17huCD25
ES-TT2
ES-WW6
ES-WW6_F1KO
EpiSC-5
EpiSC-7
Headless Embryo
MEF
Mesoderm
ZhBTc4


BHLHE40
CEBPB
CHD1
CHD2
CTCF
E2F4
EP300
ETS1
FLI1
FOSL1
GABPA
GATA1
GATA2
HCFC1
JUN
JUND
KAT2A
MAFK
MAX
MAZ
MXI1
MYB
Skeletal system
Limb
Musclear system
C2C12
Cirulatory system
Erythroblast
G1E
G1E-ER4
Heart
Nervous system

Cerebellum																						
Cortex																						
Olfactory Bulb																						
Whole Brain																						
Respiratory system																						
Lung																						
Digestive system																						
Liver																						
Small Intestine																						
Excretory system																						
Kidney																						
Reporductive system																						
Testis																						
Lyphatic system																						
Bone Marrow																						
Bone Marrow Derived Macrophage																						
CH12																						
MEL																						
Megakaryocyte																						
Spleen																						
Thymus																						
Stem cells																						
ES-Bruce4																						
ES-E14																						
MEF																						

	MYC	MYOD1	MYOG	PAX5	POLR2A	RAD21	RCOR1	RDBP	REST	SIN3A	SMC3	SRF	TAL1	TBP	TCF12	TCF3	UBTF	USF1	USF2	ZC3H11A	ZKSCAN1	ZMIZ1	ZNF384
Skeletal system																							
Limb																							
Muscluar system																							
C2C12																							
Cirulatory system																							
Erythroblast																							
G1E																							
G1E-ER4																							
Heart																							
Nervous system																							
Cerebellum																							
Cortex																							
Olfactory Bulb																							
Whole Brain																							
Respiratory system																							
Lung																							
Digestive system																							
Liver																							
Small Intestine																							
Excretory system																							
Kidney																							
Reporductive system																							
Testis																							
Lyphatic system																							
Bone Marrow																							
Bone Marrow Derived Macrophage																							
CH12																							
MEL																							
Megakaryocyte																							
Spleen																							
Thymus																							
Stem cells																							
ES-Bruce4																							
ES-E14																							
MEF																							

Supplementary Table 3: List of genomci experiments conducted in this study.

	Tisse/cell type	Replicate	Note1	Note2	DCC number	GEO number
RNA-Seq, CSHL	Adrenal	1	1AWC	LID20728	wgEncodeEM002520	GSM900188
	Adrenal	2	1BWC	LID20729	wgEncodeEM002520	GSM900188
	Bladder	1	30AWC	LID47030	wgEncodeEM003062	GSM1000564
	Bladder	2	30BWC	LID47031	wgEncodeEM003062	GSM1000564
	CNS	1	23AWC	LID46948	wgEncodeEM003056	GSM1000569
	CNS	1	24AWC	LID46950	wgEncodeEM003057	GSM1000570
	CNS	2	23BWC	LID46949	wgEncodeEM003056	GSM1000569
	CNS	2	24BWC	LID46951	wgEncodeEM003057	GSM1000570
	CNS	1	21AWC	LID46946	wgEncodeEM003052	GSM1000573
	CNS	2	21BWC	LID46947	wgEncodeEM003052	GSM1000573
	Cerebellum	1	20AWC	LID47036	wgEncodeEM003058	GSM1000567
	Cerebellum	2	20BWC	LID47037	wgEncodeEM003058	GSM1000567
	Colon	1	14AWC	LID21040	wgEncodeEM002512	GSM900198
	Colon	2	14BWC	LID21041	wgEncodeEM002512	GSM900198
	Cortex	1	6AWC	LID47032	wgEncodeEM003063	GSM1000563
	Cortex	2	6BWC	LID47033	wgEncodeEM003063	GSM1000563
	Duodenum	1	2AWC	LID20730	wgEncodeEM002521	GSM900187
	Duodenum	2	2BWC	LID20731	wgEncodeEM002521	GSM900187
	Frontal Lobe	1	7AWC	LID47081	wgEncodeEM003064	GSM1000562
	Frontal Lobe	2	7BWC	LID47082	wgEncodeEM003064	GSM1000562
	Genital Fat Pad	1	4AWC	LID21179	wgEncodeEM002522	GSM900190
	Genital Fat Pad	2	4BWC	LID21180	wgEncodeEM002522	GSM900190
	Heart	1	16AWC	LID20870	wgEncodeEM002513	GSM900199
	Heart	2	16BWC	LID20871	wgEncodeEM002513	GSM900199
	Kidney	1	17AWC	LID20872	wgEncodeEM002514	GSM900194
	Kidney	2	17BWC	LID20873	wgEncodeEM002514	GSM900194
	Large Intestine	1	8AWC	LID21183	wgEncodeEM002523	GSM900189
	Large Intestine	2	8BWC	LID21184	wgEncodeEM002523	GSM900189
	Limb	1	29AWC	LID46985	wgEncodeEM003059	GSM1000568
	Limb	2	29BWC	LID46986	wgEncodeEM003059	GSM1000568
	Liver	1	26AWC	LID47144	wgEncodeEM003053	GSM1000574
	Liver	1	28AWC	LID47146	wgEncodeEM003054	GSM1000571
	Liver	2	26BWC	LID47145	wgEncodeEM003053	GSM1000574
	Liver	2	28BWC	LID47147	wgEncodeEM003054	GSM1000571
	Liver	1	27AWC	LID47148	wgEncodeEM003060	GSM1000566
	Liver	2	27BWC	LID47149	wgEncodeEM003060	GSM1000566
	Liver	1	18AWC	LID21042	wgEncodeEM002515	GSM900195
	Liver	2	18BWC	LID21043	wgEncodeEM002515	GSM900195
	Lung	1	15AWC	LID20920	wgEncodeEM002516	GSM900196
	Lung	2	15BWC	LID20921	wgEncodeEM002516	GSM900196
	Mammary Gland	1	22AWC	LID20924	wgEncodeEM002524	GSM900184
	Mammary Gland	2	22BWC	LID20925	wgEncodeEM002524	GSM900184
	Ovary	1	11AWC	LID20821	wgEncodeEM002525	GSM900183
	Ovary	2	11BWC	LID20822	wgEncodeEM002525	GSM900183
	Placenta	1	10AWC	LID46983	wgEncodeEM003061	GSM1000565
	Placenta	2	10BWC	LID46984	wgEncodeEM003061	GSM1000565
	Small Intestine	1	9AWC	LID20819	wgEncodeEM002526	GSM900186
	Small Intestine	2	9BWC	LID20820	wgEncodeEM002526	GSM900186
	Spleen	1	13AWC	LID21038	wgEncodeEM002517	GSM900197
	Spleen	2	13BWC	LID21039	wgEncodeEM002517	GSM900197
	Stomach	1	3AWC	LID20732	wgEncodeEM002527	GSM900185
	Stomach	2	3BWC	LID20733	wgEncodeEM002527	GSM900185
	Subcutaneous Fat Pad	1	5AWC	LID21181	wgEncodeEM002528	GSM900191
	Subcutaneous Fat Pad	2	5BWC	LID21182	wgEncodeEM002528	GSM900191
	Testis	1	12AWC	LID20868	wgEncodeEM002519	GSM900193
	Testis	2	12BWC	LID20869	wgEncodeEM002519	GSM900193
	Thymus	1	19AWC	LID20922	wgEncodeEM002518	GSM900192
	Thymus	2	19BWC	LID20923	wgEncodeEM002518	GSM900192
	Whole Brain	1	25AWC	LID46987	wgEncodeEM003055	GSM1000572
	Whole Brain	2	25BWC	LID46988	wgEncodeEM003055	GSM1000572
RNA-Seq, CalTech	10T1/2	1	EqS_2.0pct_60hr		wgEncodeEM002734	GSM929772
	10T1/2	1	None		wgEncodeEM002735	GSM929773
	C2C12	1	EqS_2.0pct_60hr		wgEncodeEM002733	GSM929775
	C2C12	1	None		wgEncodeEM002732	GSM929774
RNA-Seq, LICR	Brown Adipose Tissue	1	Adult 24 weeks		wgEncodeEM002635	GSM929703
	Brown Adipose Tissue	2	Adult 24 weeks		wgEncodeEM002635	GSM929703
	Bone Marrow Derived Macroph	1	Adult 8 weeks		wgEncodeEM002636	GSM929705
	Bone Marrow Derived Macroph	2	Adult 8 weeks		wgEncodeEM002636	GSM929705
	Bone Marrow	1	Adult 8 weeks		wgEncodeEM001706	GSM929717
	Bone Marrow	2	Adult 8 weeks		wgEncodeEM001706	GSM929717
	Cerebellum	1	Adult 8 weeks		wgEncodeEM001710	GSM929709
	Cerebellum	2	Adult 8 weeks		wgEncodeEM001710	GSM929709
	Cortex	1	Adult 8 weeks		wgEncodeEM001711	GSM929708
	Cortex	2	Adult 8 weeks		wgEncodeEM001711	GSM929708
	ES-Bruce4	1	Embryonic day 0 (stem		wgEncodeEM001707	GSM929718
	ES-Bruce4	2	Embryonic day 0 (stem		wgEncodeEM001707	GSM929718
	Heart	1	Embryonic day 14.5		wgEncodeEM002391	GSM929724
	Heart	2	Embryonic day 14.5		wgEncodeEM002391	GSM929724
	Heart	1	Adult 8 weeks		wgEncodeEM001712	GSM929707
	Heart	2	Adult 8 weeks		wgEncodeEM001712	GSM929707
	Kidney	1	Adult 8 weeks		wgEncodeEM001713	GSM929706
	Kidney	2	Adult 8 weeks		wgEncodeEM001713	GSM929706
	Limb	1	Embryonic day 14.5		wgEncodeEM002387	GSM929713
	Limb	2	Embryonic day 14.5		wgEncodeEM002387	GSM929713
	Liver	1	Embryonic day 14.5		wgEncodeEM002392	GSM929721
	Liver	2	Embryonic day 14.5		wgEncodeEM002392	GSM929721
	Liver	1	Adult 8 weeks		wgEncodeEM001714	GSM929711

	Liver	2	Adult 8 weeks		wgEncodeEM001714	GSM929711
	Lung	1	Adult 8 weeks		wgEncodeEM001715	GSM929710
	Lung	2	Adult 8 weeks		wgEncodeEM001715	GSM929710
	MEF	1	Adult 8 weeks		wgEncodeEM001708	GSM929719
	MEF	2	Adult 8 weeks		wgEncodeEM001708	GSM929719
	MEL	1	Immortal cells		wgEncodeEM002634	GSM929704
	MEL	2	Immortal cells		wgEncodeEM002634	GSM929704
	Olfactory Bulb	1	Adult 8 weeks		wgEncodeEM002385	GSM929712
	Olfactory Bulb	2	Adult 8 weeks		wgEncodeEM002385	GSM929712
	Placenta	1	Adult 8 weeks		wgEncodeEM002393	GSM929722
	Placenta	2	Adult 8 weeks		wgEncodeEM002393	GSM929722
	Small Intestine	1	Adult 8 weeks		wgEncodeEM002388	GSM929716
	Small Intestine	2	Adult 8 weeks		wgEncodeEM002388	GSM929716
	Spleen	1	Adult 8 weeks		wgEncodeEM001709	GSM929720
	Spleen	2	Adult 8 weeks		wgEncodeEM001709	GSM929720
	Testis	1	Adult 8 weeks		wgEncodeEM002389	GSM929715
	Testis	2	Adult 8 weeks		wgEncodeEM002389	GSM929715
	Thymus	1	Adult 8 weeks		wgEncodeEM002386	GSM929714
	Thymus	2	Adult 8 weeks		wgEncodeEM002386	GSM929714
	Whole Brain	1	Embryonic day 14.5		wgEncodeEM002390	GSM929723
	Whole Brain	2	Embryonic day 14.5		wgEncodeEM002390	GSM929723
RNA-Seq, UW	416B	1	B6D2F1/J	Immortal cells	wgEncodeEM002947	GSM970855
	416B	2	B6D2F1/J	Immortal cells	wgEncodeEM002947	GSM970855
	A20	1	BALB/cAnN	Immortal cells	wgEncodeEM002948	GSM970854
	A20	2	BALB/cAnN	Immortal cells	wgEncodeEM002948	GSM970854
	B-cell (CD19+)	1	C57BL/6	Adult 8 weeks	wgEncodeEM002951	GSM970867
	B-cell (CD43-)	1	C57BL/6	Adult 8 weeks	wgEncodeEM002952	GSM970868
	B-cell (CD43-)	2	C57BL/6	Adult 8 weeks	wgEncodeEM002952	GSM970868
	Cerebellum	1	C57BL/6	Adult 8 weeks	wgEncodeEM001986	GSM970856
	Cerebellum	2	C57BL/6	Adult 8 weeks	wgEncodeEM001986	GSM970856
	Cerebrum	1	C57BL/6	Adult 8 weeks	wgEncodeEM001987	GSM970857
	Cerebrum	2	C57BL/6	Adult 8 weeks	wgEncodeEM001987	GSM970857
	Fat Pad	1	C57BL/6	Adult 8 weeks	wgEncodeEM001988	GSM970858
	Genital Fat Pad	1	C57BL/6	Adult 8 weeks	wgEncodeEM002953	GSM970869
	Genital Fat Pad	2	C57BL/6	Adult 8 weeks	wgEncodeEM002953	GSM970869
	Headless Embryo	1	CD-1	Embryonic day 11.5	wgEncodeEM001989	GSM970859
	Heart	1	C57BL/6	Adult 8 weeks	wgEncodeEM002954	GSM970862
	Heart	2	C57BL/6	Adult 8 weeks	wgEncodeEM002954	GSM970862
	Kidney	1	C57BL/6	Adult 8 weeks	wgEncodeEM001990	GSM970871
	Large Intestine	1	C57BL/6	Adult 8 weeks	wgEncodeEM002955	GSM970863
	Large Intestine	2	C57BL/6	Adult 8 weeks	wgEncodeEM002955	GSM970863
	Liver	1	129	Embryonic day 14.5	wgEncodeEM002961	GSM970851
	Liver	1	C57BL/6	Embryonic day 14.5	wgEncodeEM001991	GSM970870
	Liver	1	C57BL/6	Adult 8 weeks	wgEncodeEM001992	GSM970873
	Liver	2	C57BL/6	Adult 8 weeks	wgEncodeEM001992	GSM970873
	Liver	3	C57BL/6	Adult 8 weeks	wgEncodeEM001992	GSM970873
	Liver	4	C57BL/6	Adult 8 weeks	wgEncodeEM001992	GSM970873
	Lung	1	C57BL/6	Adult 8 weeks	wgEncodeEM002956	GSM970864
	MEL	1	Unknown	Adult 8 weeks	wgEncodeEM001993	GSM970872
	MEL	2	Unknown	Adult 8 weeks	wgEncodeEM001993	GSM970872
	NIH-3T3	1	NIH/Swiss	Immortal cells	wgEncodeEM002949	GSM970853
	NIH-3T3	2	NIH/Swiss	Immortal cells	wgEncodeEM002949	GSM970853
	Patski	1	Spretus.BL6-Xist	Immortal cells	wgEncodeEM002950	GSM970866
	Patski	2	Spretus.BL6-Xist	Immortal cells	wgEncodeEM002950	GSM970866
	Skeletal Muscle	1	C57BL/6	Adult 8 weeks	wgEncodeEM002957	GSM970865
	Skeletal Muscle	2	C57BL/6	Adult 8 weeks	wgEncodeEM002957	GSM970865
	Spleen	1	C57BL/6	Adult 8 weeks	wgEncodeEM002958	GSM970860
	T-Naive	1	C57BL/6	Adult 8 weeks	wgEncodeEM002959	GSM970861
	Thymus	1	C57BL/6	Adult 8 weeks	wgEncodeEM002960	GSM970852
	Thymus	2	C57BL/6	Adult 8 weeks	wgEncodeEM002960	GSM970852
	Whole Brain	1	C57BL/6	Embryonic day 14.5	wgEncodeEM001995	GSM970874
	Whole Brain	1	C57BL/6	Embryonic day 18.5	wgEncodeEM001997	GSM970875
	Whole Brain	2	C57BL/6	Embryonic day 14.5	wgEncodeEM001995	GSM970874
	Whole Brain	2	C57BL/6	Embryonic day 18.5	wgEncodeEM001997	GSM970875
RNA-Seq, PSU	CH12	1	None	1x41	wgEncodeEM003177	GSM995540
	CH12	2	None	1x41	wgEncodeEM003177	GSM995540
	Erythroblast	1	None	2x99D	wgEncodeEM003180	GSM995533
	Erythroblast	2	None	2x99D	wgEncodeEM003180	GSM995533
	FVL-stem	1	None	2x99D	wgEncodeEM003182	GSM995535
	FVL-stem	2	None	2x99D	wgEncodeEM003182	GSM995535
	FV-progenitor	1	None	2x99D	wgEncodeEM003181	GSM995534
	FV-progenitor	2	None	2x99D	wgEncodeEM003181	GSM995534
	G1E	1	None	1x36	wgEncodeEM003178	GSM995530
	G1E	2	None	1x55	wgEncodeEM003178	GSM995530
	G1E	1	None	2x99D	wgEncodeEM003183	GSM995536
	G1E	2	None	2x99D	wgEncodeEM003183	GSM995536
	G1E-ER4	1	None	2x99D	wgEncodeEM003187	GSM995532
	G1E-ER4	2	None	2x99D	wgEncodeEM003187	GSM995532
	G1E-ER4	1	Estradiol 14 hour diff p	2x99D	wgEncodeEM003185	GSM995527
	G1E-ER4	2	Estradiol 14 hour diff p	2x99D	wgEncodeEM003185	GSM995527
	G1E-ER4	1	Estradiol 24 hour diff p	1x36	wgEncodeEM003179	GSM995529
	G1E-ER4	2	Estradiol 24 hour diff p	1x36	wgEncodeEM003179	GSM995529
	G1E-ER4	1	Estradiol 24 hour diff p	2x99D	wgEncodeEM003189	GSM995539
	G1E-ER4	2	Estradiol 24 hour diff p	2x99D	wgEncodeEM003189	GSM995539
	G1E-ER4	1	Estradiol 30 hour diff p	2x99D	wgEncodeEM003192	GSM995541
	G1E-ER4	2	Estradiol 30 hour diff p	2x99D	wgEncodeEM003192	GSM995541
	G1E-ER4	1	Estradiol 3 hour diff pr	2x99D	wgEncodeEM003188	GSM995538
	G1E-ER4	2	Estradiol 3 hour diff pr	2x99D	wgEncodeEM003188	GSM995538
	G1E-ER4	1	Estradiol 7 hour diff pr	2x99D	wgEncodeEM003186	GSM995531

	G1E-ER4	2	Estradiol 7 hour diff pr	2x99D	wgEncodeEM003186	GSM995531
	MEL	1	DMSO_2.0pct	1x45	wgEncodeEM003191	GSM995526
	MEL	2	DMSO_2.0pct	1x45	wgEncodeEM003191	GSM995526
	MEL	1	None	1x45	wgEncodeEM003190	GSM995528
	MEL	2	None	1x45	wgEncodeEM003190	GSM995528
	MEP	1	None	2x99D	wgEncodeEM003184	GSM995525
	MEP	2	None	2x99D	wgEncodeEM003184	GSM995525
	Megakaryocyte	1	None	2x99D	wgEncodeEM003193	GSM995537
	Megakaryocyte	2	None	2x99D	wgEncodeEM003193	GSM995537
RNA-Seq, Yale/Standford	CH12	1	None	FastqRd1	wgEncodeEM002000	GSM973234
	CH12	2	None	FastqRd1	wgEncodeEM002000	GSM973234
	ES-E14	1	None	FastqRd1	wgEncodeEM002899	GSM973235
	ES-E14	2	None	FastqRd1	wgEncodeEM002899	GSM973235
	MEL	1	DMSO_2.0pct	FastqRd1	wgEncodeEM001998	GSM973233
	MEL	2	DMSO_2.0pct	FastqRd1	wgEncodeEM001998	GSM973233
	MEL	1	None	FastqRd1	wgEncodeEM001999	GSM973232
	MEL	2	None	FastqRd1	wgEncodeEM001999	GSM973232

	Tisse/cell type	**Replicate**	**Note1**	**Note2**	**Notes3**	**Gender**	**DCC number**	**GEO number**
DHS, UW	3134	2	RIII	Immortal cells		M	wgEncodeEM001721	GSM1014196
	3134	1	RIII	Immortal cells		M	wgEncodeEM001721	GSM1014196
	416B	2	B6D2F1/J	Immortal cells		M	wgEncodeEM001717	GSM1014163
	416B	1	B6D2F1/J	Immortal cells		M	wgEncodeEM001717	GSM1014163
	A20	2	BALB/cAnN	Immortal cells		M	wgEncodeEM001733	GSM1014167
	A20	1	BALB/cAnN	Immortal cells		M	wgEncodeEM001733	GSM1014167
	B-cell (CD19+)	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001727	GSM1014190
	B-cell (CD19+)	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001727	GSM1014190
	B-cell (CD43-)	4	C57BL/6	Adult 8 weeks		M	wgEncodeEM001734	GSM1014170
	B-cell (CD43-)	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001734	GSM1014170
	B-cell (CD43-)	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001734	GSM1014170
	B-cell (CD43-)	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001734	GSM1014170
	CH12	2	B10.H-2aH-4bp/Wts	Immortal cells		F	wgEncodeEM003416	GSM1014153
	CH12	1	B10.H-2aH-4bp/Wts	Immortal cells		F	wgEncodeEM003416	GSM1014153
	Cerebellum	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001716	GSM1014164
	Cerebellum	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001716	GSM1014164
	Cerebellum	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001716	GSM1014164
	Cerebrum	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001718	GSM1014168
	Cerebrum	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001718	GSM1014168
	Cerebrum	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001718	GSM1014168
	EPC (CD117+ CD71+ TER119+)	1	CD-1	Embryonic day 14.5		M	wgEncodeEM003412	GSM1014157
	EPC (CD117+ CD71+ TER119-)	1	CD-1	Embryonic day 14.5		M	wgEncodeEM003413	GSM1014158
	EPC (CD117+ CD71- TER119-)	1	CD-1	Embryonic day 14.5		M	wgEncodeEM003414	GSM1014155
	EPC (CD117- CD71+ TER119+)	1	CD-1	Embryonic day 14.5		M	wgEncodeEM003415	GSM1014156
	ES-CJ7	2	129S1/SVImJ	Embryonic day 0 (stem cell)		M	wgEncodeEM001728	GSM1014187
	ES-CJ7	1	129S1/SVImJ	Embryonic day 0 (stem cell)		M	wgEncodeEM001728	GSM1014187
	ES-E14	2	129/Ola	Embryonic day 0 (stem cell)		M	wgEncodeEM003417	GSM1014154
	ES-E14	1	129/Ola	Embryonic day 0 (stem cell)		M	wgEncodeEM003417	GSM1014154
	ES-WW6	2	Unknown	Embryonic day 0 (stem cell)		M	wgEncodeEM003410	GSM1014159
	ES-WW6	1	Unknown	Embryonic day 0 (stem cell)		M	wgEncodeEM003410	GSM1014159
	ES-WW6_F1KO	2	Unknown	Embryonic day 0 (stem cell)		M	wgEncodeEM003411	GSM1014160
	ES-WW6_F1KO	1	Unknown	Embryonic day 0 (stem cell)		M	wgEncodeEM003411	GSM1014160
	Fat Pad	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001731	GSM1014165
	Fat Pad	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001731	GSM1014165
	Fibroblast	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001719	GSM1014199
	Fibroblast	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001719	GSM1014199
	Fore Limb Bud	2	CD-1	Embryonic day 11.5		M	wgEncodeEM001931	GSM1014174
	Fore Limb Bud	1	CD-1	Embryonic day 11.5		M	wgEncodeEM001931	GSM1014174
	Genital Fat Pad	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001932	GSM1014173
	Genital Fat Pad	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001932	GSM1014173
	Headless Embryo	2	CD-1	Embryonic day 11.5		M	wgEncodeEM001933	GSM1014172
	Headless Embryo	1	CD-1	Embryonic day 11.5		M	wgEncodeEM001933	GSM1014172
	Heart	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001730	GSM1014166
	Heart	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001730	GSM1014166
	Hind Limb Bud	2	CD-1	Embryonic day 11.5		M	wgEncodeEM001934	GSM1014179
	Hind Limb Bud	1	CD-1	Embryonic day 11.5		M	wgEncodeEM001934	GSM1014179
	Kidney	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001722	GSM1014193
	Kidney	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001722	GSM1014193
	Large Intestine	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003397	GSM1014186
	Large Intestine	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003397	GSM1014186
	Liver	1	129	Embryonic day 14.5		M	wgEncodeEM003419	GSM1014162
	Liver	2	129.ŒîLCR/ŒîLCR	Embryonic day 14.5		M	wgEncodeEM003418	GSM1014161
	Liver	1	129.ŒîLCR/ŒîLCR	Embryonic day 14.5		M	wgEncodeEM003418	GSM1014161
	Liver	1	C57BL/6	Embryonic day 14.5		M	wgEncodeEM003401	GSM1014183
	Liver	9	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	8	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	7	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	6	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	5	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	4	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	14	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	13	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	12	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	11	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	10	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Liver	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001720	GSM1014195
	Lung	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001723	GSM1014194
	Lung	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001723	GSM1014194

	Lung	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001723	GSM1014194
	MEL	3	Unknown	Immortal cells		M	wgEncodeEM001724	GSM1014191
	MEL	2	Unknown	Immortal cells		M	wgEncodeEM001724	GSM1014191
	MEL	1	Unknown	Immortal cells		M	wgEncodeEM001724	GSM1014191
	Mesoderm	2	CD-1	Embryonic day 11.5		M	wgEncodeEM001935	GSM1014178
	Mesoderm	1	CD-1	Embryonic day 11.5		M	wgEncodeEM001935	GSM1014178
	NIH-3T3	2	NIH/Swiss	Immortal cells		M	wgEncodeEM001936	GSM1014177
	NIH-3T3	1	NIH/Swiss	Immortal cells		M	wgEncodeEM001936	GSM1014177
	Patski	2	Spretus.BL6-Xist	Immortal cells		M	wgEncodeEM001736	GSM1014171
	Patski	1	Spretus.BL6-Xist	Immortal cells		M	wgEncodeEM001736	GSM1014171
	Retina	1	C57BL/6	Adult 1 week		M	wgEncodeEM003398	GSM1014198
	Retina	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003402	GSM1014175
	Retina	1	C57BL/6	Newborn 1 day		M	wgEncodeEM003400	GSM1014188
	Skeletal Muscle	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003399	GSM1014189
	Skeletal Muscle	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003399	GSM1014189
	Spleen	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003394	GSM1014182
	Spleen	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003394	GSM1014182
	T-Naive	4	C57BL/6	Adult 8 weeks		M	wgEncodeEM001725	GSM1014192
	T-Naive	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001725	GSM1014192
	T-Naive	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001725	GSM1014192
	T-Naive	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001725	GSM1014192
	THelper-Activated	2	C57BL/6	Adult 8 weeks	no peak files	M	wgEncodeEM003403	GSM1014149
	THelper-Activated	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003403	GSM1014149
	TReg	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001732	GSM1014148
	TReg	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001732	GSM1014148
	TReg	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001732	GSM1014148
	TReg-Activated	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003404	GSM1014200
	TReg-Activated	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003404	GSM1014200
	Thymus	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003395	GSM1014185
	Thymus	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003395	GSM1014185
	Whole Brain	2	C57BL/6	Embryonic day 14.5		M	wgEncodeEM001726	GSM1014197
	Whole Brain	1	C57BL/6	Embryonic day 14.5		M	wgEncodeEM001726	GSM1014197
	Whole Brain	2	C57BL/6	Embryonic day 18.5	no peak files	M	wgEncodeEM003396	GSM1014184
	Whole Brain	1	C57BL/6	Embryonic day 18.5		M	wgEncodeEM003396	GSM1014184
	Whole Brain	7	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	6	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	5	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	4	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	3	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	Whole Brain	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM001729	GSM1014151
	ZhBTc4	2	129/Ola	Embryonic day 0 (stem cell)		M	wgEncodeEM001735	GSM1014169
	ZhBTc4	1	129/Ola	Embryonic day 0 (stem cell)		M	wgEncodeEM001735	GSM1014169
	ZhBTc4	1	129/Ola	Embryonic day 0 (stem cell)	24 hour diff protocol B	M	wgEncodeEM003406	GSM1014152
	ZhBTc4	2	129/Ola	Embryonic day 0 (stem cell)	24 hour diff protocol B	M	wgEncodeEM003406	GSM1014152
	ZhBTc4	1	129/Ola	Embryonic day 0 (stem cell)	6 hour diff protocol B	M	wgEncodeEM003407	GSM1014150
	ZhBTc4	2	129/Ola	Embryonic day 0 (stem cell)	6 hour diff protocol B	M	wgEncodeEM003407	GSM1014150
	mG/ER	2	C57BL/6	Adult 8 weeks		M	wgEncodeEM003405	GSM1014176
	mG/ER	1	C57BL/6	Adult 8 weeks		M	wgEncodeEM003405	GSM1014176
	mG/ER	1	C57BL/6	Adult 8 weeks	24 hour diff protocol C	M	wgEncodeEM003408	GSM1014181
	mG/ER	2	C57BL/6	Adult 8 weeks	24 hour diff protocol C	M	wgEncodeEM003408	GSM1014181
	mG/ER	1	C57BL/6	Adult 8 weeks	48 hour diff protocol C	M	wgEncodeEM003409	GSM1014180
	mG/ER	2	C57BL/6	Adult 8 weeks	48 hour diff protocol C	M	wgEncodeEM003409	GSM1014180
TF, CalTech	Tissue/cell type	Replicate	TF	notes1	Note2	Notes3	DCC number	GEO number
	C2C12	1	CEBPB (sc-150)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002122	GSM915180
	C2C12	2	CEBPB (sc-150)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002122	GSM915180
	C2C12	1	CEBPB (sc-150)	Control 50bp	None	PCR 1-round	wgEncodeEM002123	GSM915179
	C2C12	1	CTCF (07-729)	32bp	None	PCR 2-round	wgEncodeEM002108	GSM915188
	C2C12	1	E2F4 (sc-866)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002109	GSM915187
	C2C12	1	FOSL1 (sc-605)	Control 36bp	None	PCR 1-round	wgEncodeEM002124	GSM915182
	C2C12	1	Input	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002110	GSM915169
	C2C12	1	Input	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002112	GSM915171
	C2C12	1	Input	32bp	None	PCR 2-round	wgEncodeEM002111	GSM915170
	C2C12	1	Input	Control 36bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002125	GSM915181
	C2C12	1	Input	Control 36bp	None	PCR 1-round	wgEncodeEM002137	GSM915160
	C2C12	1	Input	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002126	GSM915184
	C2C12	2	Input	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002126	GSM915184
	C2C12	1	Input	Control 50bp	None	PCR 1-round	wgEncodeEM002113	GSM915172
	C2C12	2	Input	Control 50bp	None	PCR 1-round	wgEncodeEM002113	GSM915172
	C2C12	3	Input	Control 50bp	None	PCR 1-round	wgEncodeEM002113	GSM915172
	C2C12	1	MAX (sc-197)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002114	GSM915173
	C2C12	1	MAX (sc-197)	Control 50bp	None	PCR 1-round	wgEncodeEM002115	GSM915174
	C2C12	1	MYOD (sc-32758)	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002127	GSM915183
	C2C12	1	MYOD (sc-32758)	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002129	GSM915185
	C2C12	1	MYOD (sc-32758)	32bp	None	PCR 2-round	wgEncodeEM002128	GSM915186
	C2C12	1	MYOD (sc-32758)	Control 50bp	EqS_2.0pct_7d	PCR 1-round	wgEncodeEM002130	GSM915165
	C2C12	1	Myogenin (sc-12732)	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002136	GSM915159
	C2C12	1	Myogenin (sc-12732)	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002132	GSM915163
	C2C12	1	Myogenin (sc-12732)	32bp	None	PCR 2-round	wgEncodeEM002131	GSM915166
	C2C12	1	Myogenin (sc-12732)	Control 50bp	EqS_2.0pct_7d	PCR 1-round	wgEncodeEM002133	GSM915164
	C2C12	1	NRSF	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002116	GSM915175
	C2C12	1	POL2 (MMS-126R)	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002117	GSM915176
	C2C12	1	L2 (phospho S2) (ab50!	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002118	GSM915167
	C2C12	1	SRF (sc-335)	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002119	GSM915168
	C2C12	1	TCF12 (sc-357)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002120	GSM915178
	C2C12	1	TCF3 (sc-349)	32bp	EqS_2.0pct_5d	PCR 2-round	wgEncodeEM002121	GSM915177
	C2C12	1	USF1 (sc-229)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002134	GSM915161
	C2C12	1	USF1 (sc-229)	Control 50bp	None	PCR 1-round	wgEncodeEM002135	GSM915162
TF, LICR	Bone Marrow Derived Macroph	1	CTCF (07-729)	F	Adult 8 weeks		wgEncodeEM002663	GSM918726

Bone Marrow Derived Macroph	2	CTCF (07-729)	F	Adult 8 weeks	wgEncodeEM002663	GSM918726
Bone Marrow Derived Macroph	1	Input	F	Adult 8 weeks	wgEncodeEM002660	GSM918737
Bone Marrow Derived Macroph	2	Input	F	Adult 8 weeks	wgEncodeEM002660	GSM918737
Bone Marrow Derived Macroph	1	POL2 (MMS-126R)	F	Adult 8 weeks	wgEncodeEM002664	GSM918720
Bone Marrow Derived Macroph	2	POL2 (MMS-126R)	F	Adult 8 weeks	wgEncodeEM002664	GSM918720
Bone Marrow	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001687	GSM918757
Bone Marrow	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001687	GSM918757
Bone Marrow	1	Input	M	Adult 8 weeks	wgEncodeEM001447	GSM918721
Bone Marrow	2	Input	M	Adult 8 weeks	wgEncodeEM001447	GSM918721
Bone Marrow	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001688	GSM918760
Bone Marrow	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001688	GSM918760
Cerebellum	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001689	GSM918759
Cerebellum	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001689	GSM918759
Cerebellum	1	Input	M	Adult 8 weeks	wgEncodeEM001448	GSM918733
Cerebellum	2	Input	M	Adult 8 weeks	wgEncodeEM001448	GSM918733
Cerebellum	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001692	GSM918725
Cerebellum	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001692	GSM918725
Cortex	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001690	GSM918727
Cortex	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001690	GSM918727
Cortex	1	Input	M	Adult 8 weeks	wgEncodeEM001449	GSM918732
Cortex	2	Input	M	Adult 8 weeks	wgEncodeEM001449	GSM918732
Cortex	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001691	GSM918728
Cortex	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001691	GSM918728
ES-Bruce4	1	CTCF (07-729)	M	Embryonic day 0 (stem c	wgEncodeEM001703	GSM918748
ES-Bruce4	2	CTCF (07-729)	M	Embryonic day 0 (stem c	wgEncodeEM001703	GSM918748
ES-Bruce4	1	Input	M	Embryonic day 0 (stem c	wgEncodeEM001683	GSM918754
ES-Bruce4	2	Input	M	Embryonic day 0 (stem c	wgEncodeEM001683	GSM918754
ES-Bruce4	1	POL2 (MMS-126R)	M	Embryonic day 0 (stem c	wgEncodeEM001704	GSM918749
ES-Bruce4	2	POL2 (MMS-126R)	M	Embryonic day 0 (stem c	wgEncodeEM001704	GSM918749
ES-Bruce4	1	P300/EP300 (sc-585)	M	Embryonic day 0 (stem c	wgEncodeEM001705	GSM918750
ES-Bruce4	2	P300/EP300 (sc-585)	M	Embryonic day 0 (stem c	wgEncodeEM001705	GSM918750
Heart	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001684	GSM918756
Heart	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001684	GSM918756
Heart	1	Input	M	Adult 8 weeks	wgEncodeEM001450	GSM918755
Heart	2	Input	M	Adult 8 weeks	wgEncodeEM001450	GSM918755
Heart	1	Input	U	Embryonic day 14.5	wgEncodeEM002506	GSM918729
Heart	2	Input	U	Embryonic day 14.5	wgEncodeEM002506	GSM918729
Heart	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001694	GSM918723
Heart	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001694	GSM918723
Heart	1	P300/EP300 (sc-585)	M	Adult 8 weeks	wgEncodeEM001702	GSM918747
Heart	2	P300/EP300 (sc-585)	M	Adult 8 weeks	wgEncodeEM001702	GSM918747
Kidney	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001685	GSM918731
Kidney	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001685	GSM918731
Kidney	1	Input	M	Adult 8 weeks	wgEncodeEM001451	GSM918716
Kidney	2	Input	M	Adult 8 weeks	wgEncodeEM001451	GSM918716
Kidney	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001686	GSM918758
Kidney	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001686	GSM918758
Limb	1	CTCF (07-729)	U	Embryonic day 14.5	wgEncodeEM002589	GSM918741
Limb	2	CTCF (07-729)	U	Embryonic day 14.5	wgEncodeEM002589	GSM918741
Limb	2	Input	U	Embryonic day 14.5	wgEncodeEM002482	GSM918719
Limb	2	Input	U	Embryonic day 14.5	wgEncodeEM002482	GSM918719
Limb	1	POL2 (MMS-126R)	U	Embryonic day 14.5	wgEncodeEM002590	GSM918708
Limb	2	POL2 (MMS-126R)	U	Embryonic day 14.5	wgEncodeEM002590	GSM918708
Liver	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001696	GSM918715
Liver	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001696	GSM918715
Liver	1	Input	M	Adult 8 weeks	wgEncodeEM001452	GSM918718
Liver	2	Input	M	Adult 8 weeks	wgEncodeEM001452	GSM918718
Liver	1	Input	U	Embryonic day 14.5	wgEncodeEM002570	GSM918753
Liver	2	Input	U	Embryonic day 14.5	wgEncodeEM002570	GSM918753
Liver	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001693	GSM918738
Liver	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001693	GSM918738
Lung	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001697	GSM918722
Lung	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001697	GSM918722
Lung	1	Input	M	Adult 8 weeks	wgEncodeEM001453	GSM918739
Lung	2	Input	M	Adult 8 weeks	wgEncodeEM001453	GSM918739
Lung	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001695	GSM918724
Lung	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001695	GSM918724
MEF	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001698	GSM918743
MEF	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001698	GSM918743
MEF	1	Input	M	Adult 8 weeks	wgEncodeEM001456	GSM918740
MEF	2	Input	M	Adult 8 weeks	wgEncodeEM001456	GSM918740
MEF	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001699	GSM918761
MEF	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001699	GSM918761
MEL	1	CTCF (07-729)	M	Immortal cells	wgEncodeEM002661	GSM918744
MEL	2	CTCF (07-729)	M	Immortal cells	wgEncodeEM002661	GSM918744
MEL	1	Input	M	Immortal cells	wgEncodeEM002652	GSM918712
MEL	2	Input	M	Immortal cells	wgEncodeEM002652	GSM918712
MEL	1	POL2 (MMS-126R)	M	Immortal cells	wgEncodeEM002662	GSM918707
MEL	2	POL2 (MMS-126R)	M	Immortal cells	wgEncodeEM002662	GSM918707
Olfactory Bulb	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002585	GSM918736
Olfactory Bulb	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002585	GSM918736
Olfactory Bulb	1	Input	M	Adult 8 weeks	wgEncodeEM002473	GSM918714
Olfactory Bulb	2	Input	M	Adult 8 weeks	wgEncodeEM002473	GSM918714
Olfactory Bulb	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002586	GSM918735
Olfactory Bulb	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002586	GSM918735
Small Intestine	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002591	GSM918709
Small Intestine	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002591	GSM918709
Small Intestine	1	Input	M	Adult 8 weeks	wgEncodeEM002486	GSM918717
Small Intestine	2	Input	M	Adult 8 weeks	wgEncodeEM002486	GSM918717

	Small Intestine	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002592	GSM918710
	Small Intestine	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002592	GSM918710
	Spleen	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001700	GSM918745
	Spleen	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM001700	GSM918745
	Spleen	1	Input	M	Adult 8 weeks	wgEncodeEM001459	GSM918763
	Spleen	2	Input	M	Adult 8 weeks	wgEncodeEM001459	GSM918763
	Spleen	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001701	GSM918746
	Spleen	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM001701	GSM918746
	Testis	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002593	GSM918711
	Testis	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002593	GSM918711
	Testis	1	Input	M	Adult 8 weeks	wgEncodeEM002490	GSM918751
	Testis	2	Input	M	Adult 8 weeks	wgEncodeEM002490	GSM918751
	Testis	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002594	GSM918704
	Testis	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002594	GSM918704
	Thymus	1	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002587	GSM918734
	Thymus	2	CTCF (07-729)	M	Adult 8 weeks	wgEncodeEM002587	GSM918734
	Thymus	1	Input	M	Adult 8 weeks	wgEncodeEM002477	GSM918705
	Thymus	2	Input	M	Adult 8 weeks	wgEncodeEM002477	GSM918705
	Thymus	1	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002588	GSM918742
	Thymus	2	POL2 (MMS-126R)	M	Adult 8 weeks	wgEncodeEM002588	GSM918742
	Whole Brain	1	CTCF (07-729)	U	Embryonic day 14.5	wgEncodeEM002595	GSM918730
	Whole Brain	2	CTCF (07-729)	U	Embryonic day 14.5	wgEncodeEM002595	GSM918730
	Whole Brain	2	Input	U	Embryonic day 14.5	wgEncodeEM002494	GSM918752
	Whole Brain	2	Input	U	Embryonic day 14.5	wgEncodeEM002494	GSM918752
	Whole Brain	1	POL2 (MMS-126R)	U	Embryonic day 14.5	wgEncodeEM002596	GSM918706
	Whole Brain	2	POL2 (MMS-126R)	U	Embryonic day 14.5	wgEncodeEM002596	GSM918706
TF, PSU	CH12	2		PAX5 (ab12000)		wgEncodeEM002353	GSM923584
	CH12	1		PAX5 (ab12000)		wgEncodeEM002353	GSM923584
	CH12	2		Input		wgEncodeEM001923	GSM923569
	CH12	1		Input		wgEncodeEM001923	GSM923569
	CH12	2		CTCF (07-729)		wgEncodeEM001922	GSM923568
	CH12	1		CTCF (07-729)		wgEncodeEM001922	GSM923568
	Erythroblast	3		TAL1 (sc-12984)		wgEncodeEM002359	GSM923582
	Erythroblast	2		TAL1 (sc-12984)		wgEncodeEM002359	GSM923582
	Erythroblast	1		TAL1 (sc-12984)		wgEncodeEM002359	GSM923582
	Erythroblast	2		Input		wgEncodeEM002350	GSM923585
	Erythroblast	1		Input		wgEncodeEM002350	GSM923585
	Erythroblast	2		GATA1 (sc-265)		wgEncodeEM002349	GSM923575
	Erythroblast	1		GATA1 (sc-265)		wgEncodeEM002349	GSM923575
	G1E	2		TAL1 (sc-12984)		wgEncodeEM001930	GSM923579
	G1E	1		TAL1 (sc-12984)		wgEncodeEM001930	GSM923579
	G1E	2		Pol2-4H8 (ab5408)		wgEncodeEM002354	GSM923589
	G1E	1		Pol2-4H8 (ab5408)		wgEncodeEM002354	GSM923589
	G1E	2		Input		wgEncodeEM001916	GSM923580
	G1E	1		Input		wgEncodeEM001916	GSM923580
	G1E	2		GATA2 (sc-9008)		wgEncodeEM002356	GSM923587
	G1E	1		GATA2 (sc-9008)		wgEncodeEM002356	GSM923587
	G1E	2		GATA1 (sc-265)		wgEncodeEM002358	GSM923581
	G1E	1		GATA1 (sc-265)		wgEncodeEM002358	GSM923581
	G1E	2		CTCF (07-729)		wgEncodeEM001925	GSM923570
	G1E	1		CTCF (07-729)		wgEncodeEM001925	GSM923570
	G1E-ER4	1	None	GATA1 (sc-265)	Timecourse	wgEncodeEM003196	GSM995445
	G1E-ER4	1	None	Input	Timecourse	wgEncodeEM003202	GSM995441
	G1E-ER4	1	adiol 14 hour diff prot	GATA1 (sc-265)	Timecourse	wgEncodeEM003197	GSM995444
	G1E-ER4	1	adiol 14 hour diff prot	Input	Timecourse	wgEncodeEM003203	GSM995440
	G1E-ER4	2	adiol 24 hour diff prot	CTCF (07-729)		wgEncodeEM001926	GSM923571
	G1E-ER4	1	adiol 24 hour diff prot	CTCF (07-729)		wgEncodeEM001926	GSM923571
	G1E-ER4	2	adiol 24 hour diff prot	GATA1 (sc-265)		wgEncodeEM001927	GSM923572
	G1E-ER4	1	adiol 24 hour diff prot	GATA1 (sc-265)		wgEncodeEM001927	GSM923572
	G1E-ER4	1	adiol 24 hour diff prot	GATA1 (sc-265)	Timecourse	wgEncodeEM003198	GSM995449
	G1E-ER4	2	adiol 24 hour diff prot	GATA2 (sc-9008)		wgEncodeEM002357	GSM923588
	G1E-ER4	1	adiol 24 hour diff prot	GATA2 (sc-9008)		wgEncodeEM002357	GSM923588
	G1E-ER4	2	adiol 24 hour diff prot	Input		wgEncodeEM001921	GSM923567
	G1E-ER4	1	adiol 24 hour diff prot	Input		wgEncodeEM001921	GSM923567
	G1E-ER4	1	adiol 24 hour diff prot	Input	Timecourse	wgEncodeEM003204	GSM995439
	G1E-ER4	3	adiol 24 hour diff prot	Pol2-4H8 (ab5408)		wgEncodeEM002355	GSM923590
	G1E-ER4	2	adiol 24 hour diff prot	Pol2-4H8 (ab5408)		wgEncodeEM002355	GSM923590
	G1E-ER4	1	adiol 24 hour diff prot	Pol2-4H8 (ab5408)		wgEncodeEM002355	GSM923590
	G1E-ER4	2	adiol 24 hour diff prot	TAL1 (sc-12984)		wgEncodeEM002348	GSM923576
	G1E-ER4	1	adiol 24 hour diff prot	TAL1 (sc-12984)		wgEncodeEM002348	GSM923576
	G1E-ER4	1	adiol 30 hour diff prot	GATA1 (sc-265)	Timecourse	wgEncodeEM003199	GSM995448
	G1E-ER4	1	adiol 30 hour diff prot	Input	Timecourse	wgEncodeEM003205	GSM995438
	G1E-ER4	1	radiol 3 hour diff proto	GATA1 (sc-265)	Timecourse	wgEncodeEM003200	GSM995443
	G1E-ER4	1	radiol 3 hour diff proto	Input	Timecourse	wgEncodeEM003206	GSM995437
	G1E-ER4	1	radiol 7 hour diff proto	GATA1 (sc-265)	Timecourse	wgEncodeEM003201	GSM995442
	G1E-ER4	1	radiol 7 hour diff proto	Input	Timecourse	wgEncodeEM003207	GSM995436
	MEL	2		TAL1 (sc-12984)		wgEncodeEM002360	GSM923578
	MEL	1		TAL1 (sc-12984)		wgEncodeEM002360	GSM923578
	MEL	2		Pol2-4H8 (ab5408)		wgEncodeEM002361	GSM923577
	MEL	1		Pol2-4H8 (ab5408)		wgEncodeEM002361	GSM923577
	MEL	2		Input		wgEncodeEM001929	GSM923574
	MEL	1		Input		wgEncodeEM001929	GSM923574
	MEL	2		CTCF (07-729)		wgEncodeEM001928	GSM923573
	MEL	1		CTCF (07-729)		wgEncodeEM001928	GSM923573
	Megakaryocyte	4		TAL1 (sc-12984)		wgEncodeEM003194	GSM995447
	Megakaryocyte	3		TAL1 (sc-12984)		wgEncodeEM003194	GSM995447
	Megakaryocyte	2		TAL1 (sc-12984)		wgEncodeEM003194	GSM995447
	Megakaryocyte	1		TAL1 (sc-12984)		wgEncodeEM003194	GSM995447
	Megakaryocyte	2		Input		wgEncodeEM002352	GSM923583

	Megakaryocyte	1		Input		wgEncodeEM002352	GSM923583
	Megakaryocyte	3		GATA1 (sc-265)		wgEncodeEM002351	GSM923586
	Megakaryocyte	2		GATA1 (sc-265)		wgEncodeEM002351	GSM923586
	Megakaryocyte	1		GATA1 (sc-265)		wgEncodeEM002351	GSM923586
	Megakaryocyte	2		FLI1 (sc-356)		wgEncodeEM003195	GSM995446
	Megakaryocyte	1		FLI1 (sc-356)		wgEncodeEM003195	GSM995446
TF, Yale/Stanford	CH12	1	BHLHE40 (NB100-1800	IgG-rab	None	wgEncodeEM001979	GSM912922
	CH12	2	BHLHE40 (NB100-1800	IgG-rab	None	wgEncodeEM001979	GSM912922
	CH12	1	CHD1 (NB100-60411)	IgG-rab	None	wgEncodeEM003334	GSM1003803
	CH12	2	CHD1 (NB100-60411)	IgG-rab	None	wgEncodeEM003334	GSM1003803
	CH12	1	CHD2 (ab68301)	IgG-rab	None	wgEncodeEM001975	GSM912926
	CH12	2	CHD2 (ab68301)	IgG-rab	None	wgEncodeEM001975	GSM912926
	CH12	1	COR1/COREST (sc3018	IgG-rab	None	wgEncodeEM002783	GSM1003786
	CH12	2	COR1/COREST (sc3018	IgG-rab	None	wgEncodeEM002783	GSM1003786
	CH12	1	CTCF (sc-15914)	IgG-rab	None	wgEncodeEM001954	GSM912909
	CH12	2	CTCF (sc-15914)	IgG-rab	None	wgEncodeEM001954	GSM912909
	CH12	1	E2F4 (sc-866)	IgG-rab	None	wgEncodeEM001937	GSM912912
	CH12	2	E2F4 (sc-866)	IgG-rab	None	wgEncodeEM001937	GSM912912
	CH12	1	ETS1 (sc-350)	IgG-rab	None	wgEncodeEM002778	GSM1003774
	CH12	2	ETS1 (sc-350)	IgG-rab	None	wgEncodeEM002778	GSM1003774
	CH12	1	GCN5	IgG-rab	None	wgEncodeEM003335	GSM1003802
	CH12	2	GCN5	IgG-rab	None	wgEncodeEM003335	GSM1003802
	CH12	1	HCFC1_(NB100-68209)	IgG-rab	None	wgEncodeEM003345	GSM1003795
	CH12	2	HCFC1_(NB100-68209)	IgG-rab	None	wgEncodeEM003345	GSM1003795
	CH12	1	Input	IgG-mus	None	wgEncodeEM001938	GSM912918
	CH12	1	Input	IgG-rab	None	wgEncodeEM001939	GSM912917
	CH12	1	JUND (sc-74)	IgG-rab	None	wgEncodeEM001940	GSM912902
	CH12	2	JUND (sc-74)	IgG-rab	None	wgEncodeEM001940	GSM912902
	CH12	1	MAZ (ab85725)	IgG-rab	None	wgEncodeEM003336	GSM1003801
	CH12	2	MAZ (ab85725)	IgG-rab	None	wgEncodeEM003336	GSM1003801
	CH12	1	MAFK (ab50322)	IgG-rab	None	wgEncodeEM001980	GSM912898
	CH12	2	MAFK (ab50322)	IgG-rab	None	wgEncodeEM001980	GSM912898
	CH12	1	MAX (sc-197)	IgG-rab	None	wgEncodeEM001955	GSM912908
	CH12	2	MAX (sc-197)	IgG-rab	None	wgEncodeEM001955	GSM912908
	CH12	1	MXI1 (AF4185)	IgG-rab	None	wgEncodeEM001976	GSM912925
	CH12	2	MXI1 (AF4185)	IgG-rab	None	wgEncodeEM001976	GSM912925
	CH12	1	NELF-E (sc-32912)	IgG-rab	None	wgEncodeEM001977	GSM912924
	CH12	2	NELF-E (sc-32912)	IgG-rab	None	wgEncodeEM001977	GSM912924
	CH12	1	Nrf2 (sc-22810)	IgG-rab	None	wgEncodeEM003337	GSM1003800
	CH12	2	Nrf2 (sc-22810)	IgG-rab	None	wgEncodeEM003337	GSM1003800
	CH12	1	POL2 (MMS-126R)	IgG-mus	None	wgEncodeEM001941	GSM912891
	CH12	2	POL2 (MMS-126R)	IgG-mus	None	wgEncodeEM001941	GSM912891
	CH12	1	L2 (phospho S2) (ab50	IgG-rab	None	wgEncodeEM001970	GSM912931
	CH12	2	L2 (phospho S2) (ab50	IgG-rab	None	wgEncodeEM001970	GSM912931
	CH12	1	RAD21 (ab992)	IgG-rab	None	wgEncodeEM001956	GSM912911
	CH12	2	RAD21 (ab992)	IgG-rab	None	wgEncodeEM001956	GSM912911
	CH12	1	SIN3A (NB600-1263)	IgG-rab	None	wgEncodeEM002784	GSM1003781
	CH12	2	SIN3A (NB600-1263)	IgG-rab	None	wgEncodeEM002784	GSM1003781
	CH12	1	SMC3 (ab9263)	IgG-rab	None	wgEncodeEM001971	GSM912930
	CH12	2	SMC3 (ab9263)	IgG-rab	None	wgEncodeEM001971	GSM912930
	CH12	1	TBP (ab62126)	IgG-mus	None	wgEncodeEM001942	GSM912900
	CH12	2	TBP (ab62126)	IgG-mus	None	wgEncodeEM001942	GSM912900
	CH12	1	UBF (sc-13125)	IgG-rab	None	wgEncodeEM002786	GSM1003783
	CH12	2	UBF (sc-13125)	IgG-rab	None	wgEncodeEM002786	GSM1003783
	CH12	1	USF2 (ab60931)	IgG-mus	None	wgEncodeEM001957	GSM912910
	CH12	2	USF2 (ab60931)	IgG-mus	None	wgEncodeEM001957	GSM912910
	CH12	1	C3H11A_(NB100-7465(	IgG-rab	None	wgEncodeEM003346	GSM1003792
	CH12	2	C3H11A_(NB100-7465(	IgG-rab	None	wgEncodeEM003346	GSM1003792
	CH12	1	ZKSCAN1 (HPA006672)	IgG-rab	None	wgEncodeEM002787	GSM1003782
	CH12	2	ZKSCAN1 (HPA006672)	IgG-rab	None	wgEncodeEM002787	GSM1003782
	CH12	1	ZMIZ1 (ab65767)	IgG-rab	None	wgEncodeEM003347	GSM1003793
	CH12	2	ZMIZ1 (ab65767)	IgG-rab	None	wgEncodeEM003347	GSM1003793
	CH12	1	ZNF384 (HPA004051)	IgG-rab	None	wgEncodeEM003348	GSM1003796
	CH12	2	ZNF384 (HPA004051)	IgG-rab	None	wgEncodeEM003348	GSM1003796
	CH12	1	c-Jun (sc-1694)	IgG-rab	None	wgEncodeEM001943	GSM912901
	CH12	2	c-Jun (sc-1694)	IgG-rab	None	wgEncodeEM001943	GSM912901
	CH12	1	c-Myc (sc-764)	IgG-rab	None	wgEncodeEM001944	GSM912906
	CH12	2	c-Myc (sc-764)	IgG-rab	None	wgEncodeEM001944	GSM912906
	CH12	1	P300/EP300 (sc-584)	IgG-rab	None	wgEncodeEM001958	GSM912920
	CH12	2	P300/EP300 (sc-584)	IgG-rab	None	wgEncodeEM001958	GSM912920
	ES-E14	1	HCFC1_(NB100-68209)	Standard Control	None	wgEncodeEM003349	GSM1003799
	ES-E14	2	HCFC1_(NB100-68209)	Standard Control	None	wgEncodeEM003349	GSM1003799
	ES-E14	1	Input	Standard Control	None	wgEncodeEM002788	GSM1003811
	ES-E14	1	MAFK (ab50322)	Standard Control	None	wgEncodeEM002794	GSM1003809
	ES-E14	2	MAFK (ab50322)	Standard Control	None	wgEncodeEM002794	GSM1003809
	ES-E14	1	C3H11A_(NB100-7465(	Standard Control	None	wgEncodeEM003350	GSM1003810
	ES-E14	2	C3H11A_(NB100-7465(	Standard Control	None	wgEncodeEM003350	GSM1003810
	ES-E14	1	ZNF384 (HPA004051)	Standard Control	None	wgEncodeEM003351	GSM1003807
	ES-E14	2	ZNF384 (HPA004051)	Standard Control	None	wgEncodeEM003351	GSM1003807
	MEL	1	BHLHE40 (NB100-1800	IgG-rab	None	wgEncodeEM003344	GSM1003794
	MEL	2	BHLHE40 (NB100-1800	IgG-rab	None	wgEncodeEM003344	GSM1003794
	MEL	1	CHD1 (NB100-60411)	IgG-rab	None	wgEncodeEM003338	GSM1003805
	MEL	2	CHD1 (NB100-60411)	IgG-rab	None	wgEncodeEM003338	GSM1003805
	MEL	1	CHD2 (ab68301)	IgG-rab	None	wgEncodeEM001972	GSM912929
	MEL	2	CHD2 (ab68301)	IgG-rab	None	wgEncodeEM001972	GSM912929
	MEL	1	COR1/COREST (sc3018	IgG-rab	None	wgEncodeEM003343	GSM1003789
	MEL	2	COR1/COREST (sc3018	IgG-rab	None	wgEncodeEM003343	GSM1003789
	MEL	1	CTCF (sc-15914)	IgG-rab	None	wgEncodeEM001968	GSM912896
	MEL	2	CTCF (sc-15914)	IgG-rab	None	wgEncodeEM001968	GSM912896

	MEL	1	CTCF (sc-15914)	Standard Control	DMSO_2.0pct		wgEncodeEM002781	GSM1003784
	MEL	2	CTCF (sc-15914)	Standard Control	DMSO_2.0pct		wgEncodeEM002781	GSM1003784
	MEL	1	E2F4 (sc-866)	IgG-rab	None		wgEncodeEM001953	GSM912914
	MEL	2	E2F4 (sc-866)	IgG-rab	None		wgEncodeEM001953	GSM912914
	MEL	1	ETS1 (sc-350)	IgG-rab	None		wgEncodeEM002789	GSM1003777
	MEL	2	ETS1 (sc-350)	IgG-rab	None		wgEncodeEM002789	GSM1003777
	MEL	1	GATA1	IgG-rat	None		wgEncodeEM001945	GSM912907
	MEL	2	GATA1	IgG-rat	None		wgEncodeEM001945	GSM912907
	MEL	1	GATA1	Standard Control	DMSO_2.0pct		wgEncodeEM002790	GSM1003808
	MEL	2	GATA1	Standard Control	DMSO_2.0pct		wgEncodeEM002790	GSM1003808
	MEL	1	GCN5	IgG-rab	None		wgEncodeEM003339	GSM1003804
	MEL	2	GCN5	IgG-rab	None		wgEncodeEM003339	GSM1003804
	MEL	1	HCFC1_(NB100-68209)	IgG-rab	None		wgEncodeEM003352	GSM1003778
	MEL	2	HCFC1_(NB100-68209)	IgG-rab	None		wgEncodeEM003352	GSM1003778
	MEL		Input	IgG-Yale	DMSO_2.0pct		wgEncodeEM001963	GSM913032
	MEL		Input	IgG-mus	None		wgEncodeEM001946	GSM912904
	MEL		Input	IgG-rab	None		wgEncodeEM001947	GSM912905
	MEL	1	Input	IgG-rat	DMSO_2.0pct		wgEncodeEM002782	GSM1003787
	MEL		Input	IgG-rat	None		wgEncodeEM001948	GSM912894
	MEL	1	Input	Standard Control	DMSO_2.0pct		wgEncodeEM001983	GSM912897
	MEL	1	Input	Standard Control	DMSO_2.0pct		wgEncodeEM002791	GSM1003812
	MEL	1	Input	Standard Control	None		wgEncodeEM001982	GSM912916
	MEL	1	JUND (sc-74)	IgG-rab	None		wgEncodeEM001952	GSM912915
	MEL	2	JUND (sc-74)	IgG-rab	None		wgEncodeEM001952	GSM912915
	MEL	1	MAZ (ab85725)	IgG-rab	None		wgEncodeEM003340	GSM1003790
	MEL	2	MAZ (ab85725)	IgG-rab	None		wgEncodeEM003340	GSM1003790
	MEL	1	MAFK (ab50322)	IgG-rab	None		wgEncodeEM001981	GSM912899
	MEL	2	MAFK (ab50322)	IgG-rab	None		wgEncodeEM001981	GSM912899
	MEL	1	MAFK (ab50322)	Standard Control	DMSO_2.0pct		wgEncodeEM002792	GSM1003806
	MEL	2	MAFK (ab50322)	Standard Control	DMSO_2.0pct		wgEncodeEM002792	GSM1003806
	MEL	1	MAX (sc-197)	IgG-rab	None		wgEncodeEM001959	GSM912919
	MEL	2	MAX (sc-197)	IgG-rab	None		wgEncodeEM001959	GSM912919
	MEL	1	MXI1 (AF4185)	IgG-rab	None		wgEncodeEM001973	GSM912928
	MEL	2	MXI1 (AF4185)	IgG-rab	None		wgEncodeEM001973	GSM912928
	MEL	1	NELF-E (sc-32912)	IgG-rab	None		wgEncodeEM001964	GSM912932
	MEL	2	NELF-E (sc-32912)	IgG-rab	None		wgEncodeEM001964	GSM912932
	MEL	1	Nrf2 (sc-22810)	IgG-rab	None		wgEncodeEM003341	GSM1003791
	MEL	2	Nrf2 (sc-22810)	IgG-rab	None		wgEncodeEM003341	GSM1003791
	MEL	1	POL2 (MMS-126R)	IgG-mus	None		wgEncodeEM001949	GSM912895
	MEL	2	POL2 (MMS-126R)	IgG-mus	None		wgEncodeEM001949	GSM912895
	MEL	1	POL2 (MMS-126R)	IgG-rab	DMSO_2.0pct		wgEncodeEM002779	GSM1003775
	MEL	2	POL2 (MMS-126R)	IgG-rab	DMSO_2.0pct		wgEncodeEM002779	GSM1003775
	MEL	1	L2 (phospho S2) (ab50	IgG-rab	None		wgEncodeEM001974	GSM912927
	MEL	2	L2 (phospho S2) (ab50	IgG-rab	None		wgEncodeEM001974	GSM912927
	MEL	1	RAD21 (ab992)	IgG-rab	DMSO_2.0pct		wgEncodeEM001965	GSM912933
	MEL	2	RAD21 (ab992)	IgG-rab	DMSO_2.0pct		wgEncodeEM001965	GSM912933
	MEL	1	RAD21 (ab992)	IgG-rab	None		wgEncodeEM001966	GSM912935
	MEL	2	RAD21 (ab992)	IgG-rab	None		wgEncodeEM001966	GSM912935
	MEL	1	SIN3A (NB600-1263)	IgG-rab	None		wgEncodeEM002785	GSM1003780
	MEL	2	SIN3A (NB600-1263)	IgG-rab	None		wgEncodeEM002785	GSM1003780
	MEL	1	SMC3 (ab9263)	IgG-rab	None		wgEncodeEM001978	GSM912923
	MEL	2	SMC3 (ab9263)	IgG-rab	None		wgEncodeEM001978	GSM912923
	MEL	1	TBP (ab62126)	IgG-mus	None		wgEncodeEM001950	GSM912913
	MEL	2	TBP (ab62126)	IgG-mus	None		wgEncodeEM001950	GSM912913
	MEL	1	UBF (sc-13125)	IgG-mus	None		wgEncodeEM003342	GSM1003788
	MEL	2	UBF (sc-13125)	IgG-mus	None		wgEncodeEM003342	GSM1003788
	MEL	1	USF2 (ab60931)	IgG-mus	None		wgEncodeEM001960	GSM912892
	MEL	2	USF2 (ab60931)	IgG-mus	None		wgEncodeEM001960	GSM912892
	MEL	1	USF2 (ab60931)	IgG-rab	None		wgEncodeEM002780	GSM1003785
	MEL	2	USF2 (ab60931)	IgG-rab	None		wgEncodeEM002780	GSM1003785
	MEL	1	C3H11A_(NB100-7465	IgG-rab	None		wgEncodeEM003353	GSM1003776
	MEL	2	C3H11A_(NB100-7465	IgG-rab	None		wgEncodeEM003353	GSM1003776
	MEL	1	ZKSCAN1 (HPA006672)	IgG-rab	None		wgEncodeEM002793	GSM1003779
	MEL	2	ZKSCAN1 (HPA006672)	IgG-rab	None		wgEncodeEM002793	GSM1003779
	MEL	1	ZMIZ1 (ab65767)	IgG-rab	None		wgEncodeEM003354	GSM1003798
	MEL	2	ZMIZ1 (ab65767)	IgG-rab	None		wgEncodeEM003354	GSM1003798
	MEL	1	ZNF384 (HPA004051)	IgG-rab	None		wgEncodeEM003355	GSM1003797
	MEL	2	ZNF384 (HPA004051)	IgG-rab	None		wgEncodeEM003355	GSM1003797
	MEL	1	c-Myb (sc-7874)	IgG-rab	None		wgEncodeEM001967	GSM912903
	MEL	2	c-Myb (sc-7874)	IgG-rab	None		wgEncodeEM001967	GSM912903
	MEL	1	c-Myc (sc-764)	IgG-rab	None		wgEncodeEM001951	GSM912934
	MEL	2	c-Myc (sc-764)	IgG-rab	None		wgEncodeEM001951	GSM912934
	MEL	1	P300/EP300 (sc-585)	IgG-rab	None		wgEncodeEM001969	GSM912921
	MEL	2	P300/EP300 (sc-585)	IgG-rab	None		wgEncodeEM001969	GSM912921
	MEL	1	P300/EP300 (sc-584)	IgG-rab	None		wgEncodeEM001961	GSM912893
	MEL	2	P300/EP300 (sc-584)	IgG-rab	None		wgEncodeEM001961	GSM912893
Histone, CalTech	C2C12	1	H3K27me3 (07-449)	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002143	GSM918414
	C2C12	1	H3K27me3 (07-449)	32bp	None	PCR 2-round	wgEncodeEM002138	GSM918408
	C2C12	1	H3K36me3 (ab9050)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002139	GSM918409
	C2C12	1	H3K36me3 (ab9050)	Control 50bp	None	PCR 1-round	wgEncodeEM002140	GSM918417
	C2C12	1	H3K4me2 (ab32356)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002144	GSM918413
	C2C12	1	H3K4me3 (07-473)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002141	GSM918416
	C2C12	1	H3K4me3 (07-473)	Control 50bp	None	PCR 1-round	wgEncodeEM002142	GSM918415
	C2C12	1	H3K79me2 (ab3594)	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002145	GSM918412
	C2C12	1	H3K79me2 (ab3594)	Control 50bp	None	PCR 1-round	wgEncodeEM002146	GSM918411
	C2C12	1	H3K79me3 (ab2621)	Control 50bp	None	PCR 1-round	wgEncodeEM002147	GSM918410
	C2C12	1	H3ac (06-599)	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002148	GSM918423
	C2C12	1	H3ac (06-599)	32bp	None	PCR 2-round	wgEncodeEM002149	GSM918422
	C2C12	1	Input	32bp	EqS_2.0pct_24hr	PCR 2-round	wgEncodeEM002110	GSM918418

	C2C12	1	Input	32bp	EqS_2.0pct_60hr	PCR 2-round	wgEncodeEM002112	GSM918420
	C2C12	1	Input	32bp	None	PCR 2-round	wgEncodeEM002111	GSM918419
	C2C12	1	Input	Control 50bp	EqS_2.0pct_60hr	PCR 1-round	wgEncodeEM002126	GSM918407
	C2C12	1	Input	Control 50bp	None	PCR 1-round	wgEncodeEM002113	GSM918421
Histone, LICR	B-cell (CD43-)	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002719	GSM1000088
	B-cell (CD43-)	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002719	GSM1000088
	B-cell (CD43-)	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002720	GSM1000148
	B-cell (CD43-)	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002720	GSM1000148
	B-cell (CD43-)	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM003163	GSM1000120
	B-cell (CD43-)	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM003163	GSM1000120
	Brown Adipose Tissue	1	H3K27ac (ab4729)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002653	GSM1000071
	Brown Adipose Tissue	2	H3K27ac (ab4729)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002653	GSM1000071
	Brown Adipose Tissue	1	H3K4me1 (ab8895)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002654	GSM1000076
	Brown Adipose Tissue	2	H3K4me1 (ab8895)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002654	GSM1000076
	Brown Adipose Tissue	1	H3K4me3 (07-473)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002655	GSM1000075
	Brown Adipose Tissue	2	H3K4me3 (07-473)	M	Adult 24 weeks	C57BL/6	wgEncodeEM002655	GSM1000075
	Brown Adipose Tissue	1	Input	M	Adult 24 weeks	C57BL/6	wgEncodeEM002656	GSM1000201
	Brown Adipose Tissue	2	Input	M	Adult 24 weeks	C57BL/6	wgEncodeEM002656	GSM1000201
	Bone Marrow Derived Macroph	1	H3K27ac (ab4729)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002657	GSM1000074
	Bone Marrow Derived Macroph	2	H3K27ac (ab4729)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002657	GSM1000074
	Bone Marrow Derived Macroph	1	H3K4me1 (ab8895)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002658	GSM1000066
	Bone Marrow Derived Macroph	2	H3K4me1 (ab8895)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002658	GSM1000066
	Bone Marrow Derived Macroph	1	H3K4me3 (07-473)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002659	GSM1000065
	Bone Marrow Derived Macroph	2	H3K4me3 (07-473)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002659	GSM1000065
	Bone Marrow Derived Macroph	1	Input	F	Adult 8 weeks	C57BL/6	wgEncodeEM002660	GSM1000206
	Bone Marrow Derived Macroph	2	Input	F	Adult 8 weeks	C57BL/6	wgEncodeEM002660	GSM1000206
	Bone Marrow	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002478	GSM1000108
	Bone Marrow	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002478	GSM1000108
	Bone Marrow	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001435	GSM769024
	Bone Marrow	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001435	GSM769024
	Bone Marrow	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001436	GSM769021
	Bone Marrow	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001436	GSM769021
	Bone Marrow	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001447	GSM769011
	Bone Marrow	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001447	GSM769011
	CH12	1	H3K27ac (ab4729)	F	Immortal cells	C57BL/6	wgEncodeEM003167	GSM1000117
	CH12	2	H3K27ac (ab4729)	F	Immortal cells	C57BL/6	wgEncodeEM003167	GSM1000117
	CH12	1	H3K36me3 (ab9050)	F	Immortal cells	C57BL/6	wgEncodeEM002707	GSM1000091
	CH12	2	H3K36me3 (ab9050)	F	Immortal cells	C57BL/6	wgEncodeEM002707	GSM1000091
	CH12	1	H3K4me1 (ab8895)	F	Immortal cells	C57BL/6	wgEncodeEM003165	GSM1000119
	CH12	2	H3K4me1 (ab8895)	F	Immortal cells	C57BL/6	wgEncodeEM003165	GSM1000119
	CH12	1	H3K4me2 (ab7766)	F	Immortal cells	C57BL/6	wgEncodeEM003166	GSM1000116
	CH12	2	H3K4me2 (ab7766)	F	Immortal cells	C57BL/6	wgEncodeEM003166	GSM1000116
	CH12	1	H3K4me3 (07-473)	F	Immortal cells	C57BL/6	wgEncodeEM003168	GSM1000114
	CH12	2	H3K4me3 (07-473)	F	Immortal cells	C57BL/6	wgEncodeEM003168	GSM1000114
	CH12	1	H3K79me2 (39143)	F	Immortal cells	C57BL/6	wgEncodeEM003170	GSM1000122
	CH12	2	H3K79me2 (39143)	F	Immortal cells	C57BL/6	wgEncodeEM003170	GSM1000122
	CH12	1	H3K9ac (ab4441)	F	Immortal cells	C57BL/6	wgEncodeEM003169	GSM1000115
	CH12	2	H3K9ac (ab4441)	F	Immortal cells	C57BL/6	wgEncodeEM003169	GSM1000115
	CH12	1	Input	F	Immortal cells	C57BL/6	wgEncodeEM003164	GSM1000118
	CH12	2	Input	F	Immortal cells	C57BL/6	wgEncodeEM003164	GSM1000118
	Cerebellum	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002495	GSM1000097
	Cerebellum	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002495	GSM1000097
	Cerebellum	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002708	GSM1000090
	Cerebellum	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002708	GSM1000090
	Cerebellum	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001440	GSM769018
	Cerebellum	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001440	GSM769018
	Cerebellum	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001439	GSM769027
	Cerebellum	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001439	GSM769027
	Cerebellum	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001448	GSM769020
	Cerebellum	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001448	GSM769020
	Cortex	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002496	GSM1000100
	Cortex	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002496	GSM1000100
	Cortex	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001437	GSM769022
	Cortex	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001437	GSM769022
	Cortex	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001438	GSM769026
	Cortex	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001438	GSM769026
	Cortex	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001449	GSM769019
	Cortex	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001449	GSM769019
	ES-Bruce4	1	H3K27ac (ab4729)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002497	GSM1000099
	ES-Bruce4	2	H3K27ac (ab4729)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002497	GSM1000099
	ES-Bruce4	1	H3K27me3 (07-449)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002709	GSM1000089
	ES-Bruce4	2	H3K27me3 (07-449)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002709	GSM1000089
	ES-Bruce4	1	H3K36me3 (ab9050)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002710	GSM1000109
	ES-Bruce4	2	H3K36me3 (ab9050)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002710	GSM1000109
	ES-Bruce4	1	H3K4me1 (ab8895)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001681	GSM769009
	ES-Bruce4	2	H3K4me1 (ab8895)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001681	GSM769009
	ES-Bruce4	1	H3K4me3 (07-473)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001682	GSM769008
	ES-Bruce4	2	H3K4me3 (07-473)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001682	GSM769008
	ES-Bruce4	1	H3K9ac (ab4441)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM003176	GSM1000127
	ES-Bruce4	2	H3K9ac (ab4441)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM003176	GSM1000127
	ES-Bruce4	1	H3K9me3 (ab8898)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002721	GSM1000147
	ES-Bruce4	2	H3K9me3 (ab8898)	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM002721	GSM1000147
	ES-Bruce4	1	Input	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001683	GSM769010
	ES-Bruce4	2	Input	M	Embryonic day 0 (stem c	C57BL/6	wgEncodeEM001683	GSM769010
	ES-E14	1	H3K27ac (ab4729)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003174	GSM1000126
	ES-E14	2	H3K27ac (ab4729)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003174	GSM1000126
	ES-E14	1	H3K36me3 (ab9050)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003175	GSM1000125
	ES-E14	2	H3K36me3 (ab9050)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003175	GSM1000125
	ES-E14	1	H3K4me1 (ab8895)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003171	GSM1000121

ES-E14	2	H3K4me1 (ab8895)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003171	GSM1000121
ES-E14	1	H3K4me3 (07-473)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003172	GSM1000124
ES-E14	2	H3K4me3 (07-473)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003172	GSM1000124
ES-E14	1	H3K9ac (ab4441)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003173	GSM1000123
ES-E14	2	H3K9ac (ab4441)	M	Embryonic day 0 (stem c	129/Ola	wgEncodeEM003173	GSM1000123
Heart	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002498	GSM1000093
Heart	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002498	GSM1000093
Heart	1	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002503	GSM1000137
Heart	2	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002503	GSM1000137
Heart	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002637	GSM1000131
Heart	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002637	GSM1000131
Heart	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002638	GSM1000130
Heart	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002638	GSM1000130
Heart	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001433	GSM769025
Heart	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001433	GSM769025
Heart	1	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002504	GSM1000136
Heart	2	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002504	GSM1000136
Heart	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001441	GSM769017
Heart	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001441	GSM769017
Heart	1	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002505	GSM1000135
Heart	2	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002505	GSM1000135
Heart	1	H3K79me2 (39143)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002639	GSM1000129
Heart	2	H3K79me2 (39143)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002639	GSM1000129
Heart	1	H3K9ac (ab4441)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002640	GSM1000149
Heart	2	H3K9ac (ab4441)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002640	GSM1000149
Heart	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001450	GSM769032
Heart	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001450	GSM769032
Heart	1	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002506	GSM1000208
Heart	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002506	GSM1000208
Kidney	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002499	GSM1000092
Kidney	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002499	GSM1000092
Kidney	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002711	GSM1000077
Kidney	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002711	GSM1000077
Kidney	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002712	GSM1000063
Kidney	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002712	GSM1000063
Kidney	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001434	GSM769023
Kidney	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001434	GSM769023
Kidney	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001442	GSM769016
Kidney	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001442	GSM769016
Kidney	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001451	GSM769033
Kidney	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001451	GSM769033
Limb	1	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002479	GSM1000107
Limb	2	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002479	GSM1000107
Limb	1	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002480	GSM1000085
Limb	2	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002480	GSM1000085
Limb	1	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002481	GSM1000086
Limb	2	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002481	GSM1000086
Limb	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002482	GSM1000202
Limb	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002482	GSM1000202
Liver	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002500	GSM1000140
Liver	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002500	GSM1000140
Liver	1	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002571	GSM1000113
Liver	2	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002571	GSM1000113
Liver	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002641	GSM1000150
Liver	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002641	GSM1000150
Liver	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002642	GSM1000151
Liver	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002642	GSM1000151
Liver	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001443	GSM769015
Liver	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001443	GSM769015
Liver	1	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002573	GSM1000111
Liver	2	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002573	GSM1000111
Liver	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001444	GSM769014
Liver	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001444	GSM769014
Liver	1	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002572	GSM1000110
Liver	2	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002572	GSM1000110
Liver	1	H3K79me2 (39143)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002643	GSM1000152
Liver	2	H3K79me2 (39143)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002643	GSM1000152
Liver	1	H3K9ac (ab4441)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002644	GSM1000153
Liver	2	H3K9ac (ab4441)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002644	GSM1000153
Liver	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001452	GSM769034
Liver	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001452	GSM769034
Liver	1	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002570	GSM1000112
Liver	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002570	GSM1000112
Lung	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001445	GSM769013
Lung	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001445	GSM769013
Lung	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001446	GSM769012
Lung	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001446	GSM769012
Lung	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001453	GSM769035
Lung	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001453	GSM769035
MEF	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002501	GSM1000139
MEF	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002501	GSM1000139
MEF	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001454	GSM769028
MEF	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001454	GSM769028
MEF	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001455	GSM769029
MEF	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001455	GSM769029
MEF	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001456	GSM769030
MEF	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001456	GSM769030
MEL	1	H3K27ac (ab4729)	M	Immortal cells	Unknown	wgEncodeEM002649	GSM1000142
MEL	2	H3K27ac (ab4729)	M	Immortal cells	Unknown	wgEncodeEM002649	GSM1000142

MEL	1	H3K27me3 (07-449)	M	Immortal cells	Unknown	wgEncodeEM002645	GSM1000154
MEL	2	H3K27me3 (07-449)	M	Immortal cells	Unknown	wgEncodeEM002645	GSM1000154
MEL	1	H3K36me3 (ab9050)	M	Immortal cells	Unknown	wgEncodeEM002646	GSM1000155
MEL	2	H3K36me3 (ab9050)	M	Immortal cells	Unknown	wgEncodeEM002646	GSM1000155
MEL	1	H3K4me1 (ab8895)	M	Immortal cells	Unknown	wgEncodeEM002650	GSM1000073
MEL	2	H3K4me1 (ab8895)	M	Immortal cells	Unknown	wgEncodeEM002650	GSM1000073
MEL	1	H3K4me3 (07-473)	M	Immortal cells	Unknown	wgEncodeEM002651	GSM1000087
MEL	2	H3K4me3 (07-473)	M	Immortal cells	Unknown	wgEncodeEM002651	GSM1000087
MEL	1	H3K79me2 (39143)	M	Immortal cells	Unknown	wgEncodeEM002647	GSM1000156
MEL	2	H3K79me2 (39143)	M	Immortal cells	Unknown	wgEncodeEM002647	GSM1000156
MEL	1	H3K9ac (ab4441)	M	Immortal cells	Unknown	wgEncodeEM002648	GSM1000141
MEL	2	H3K9ac (ab4441)	M	Immortal cells	Unknown	wgEncodeEM002648	GSM1000141
MEL	1	Input	M	Immortal cells	Unknown	wgEncodeEM002652	GSM1000200
MEL	2	Input	M	Immortal cells	Unknown	wgEncodeEM002652	GSM1000200
Olfactory Bulb	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002471	GSM1000105
Olfactory Bulb	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002471	GSM1000105
Olfactory Bulb	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002472	GSM1000104
Olfactory Bulb	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002472	GSM1000104
Olfactory Bulb	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002511	GSM1000128
Olfactory Bulb	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002511	GSM1000128
Olfactory Bulb	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002473	GSM1000205
Olfactory Bulb	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002473	GSM1000205
Placenta	1	H3K27ac (ab4729)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002507	GSM1000134
Placenta	2	H3K27ac (ab4729)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002507	GSM1000134
Placenta	1	H3K4me1 (ab8895)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002508	GSM1000133
Placenta	2	H3K4me1 (ab8895)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002508	GSM1000133
Placenta	1	H3K4me3 (07-473)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002509	GSM1000132
Placenta	2	H3K4me3 (07-473)	F	Adult 8 weeks	C57BL/6	wgEncodeEM002509	GSM1000132
Placenta	1	Input	F	Adult 8 weeks	C57BL/6	wgEncodeEM002510	GSM1000207
Placenta	2	Input	F	Adult 8 weeks	C57BL/6	wgEncodeEM002510	GSM1000207
Small Intestine	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002483	GSM1000084
Small Intestine	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002483	GSM1000084
Small Intestine	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002713	GSM1000064
Small Intestine	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002713	GSM1000064
Small Intestine	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002714	GSM1000069
Small Intestine	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002714	GSM1000069
Small Intestine	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002484	GSM1000082
Small Intestine	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002484	GSM1000082
Small Intestine	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002485	GSM1000083
Small Intestine	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002485	GSM1000083
Small Intestine	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002486	GSM1000080
Small Intestine	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002486	GSM1000080
Spleen	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002502	GSM1000138
Spleen	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002502	GSM1000138
Spleen	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002722	GSM1000146
Spleen	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002722	GSM1000146
Spleen	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002715	GSM1000070
Spleen	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002715	GSM1000070
Spleen	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001457	GSM769031
Spleen	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001457	GSM769031
Spleen	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001458	GSM769036
Spleen	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM001458	GSM769036
Spleen	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001459	GSM769037
Spleen	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM001459	GSM769037
Testis	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002487	GSM1000081
Testis	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002487	GSM1000081
Testis	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002723	GSM1000145
Testis	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002723	GSM1000145
Testis	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002716	GSM1000067
Testis	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002716	GSM1000067
Testis	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002488	GSM1000078
Testis	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002488	GSM1000078
Testis	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002489	GSM1000079
Testis	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002489	GSM1000079
Testis	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002490	GSM1000203
Testis	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002490	GSM1000203
Thymus	1	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002474	GSM1000103
Thymus	2	H3K27ac (ab4729)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002474	GSM1000103
Thymus	1	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002724	GSM1000144
Thymus	2	H3K27me3 (07-449)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002724	GSM1000144
Thymus	1	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002717	GSM1000068
Thymus	2	H3K36me3 (ab9050)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002717	GSM1000068
Thymus	1	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002475	GSM1000102
Thymus	2	H3K4me1 (ab8895)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002475	GSM1000102
Thymus	1	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002476	GSM1000101
Thymus	2	H3K4me3 (07-473)	M	Adult 8 weeks	C57BL/6	wgEncodeEM002476	GSM1000101
Thymus	1	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002477	GSM1000204
Thymus	2	Input	M	Adult 8 weeks	C57BL/6	wgEncodeEM002477	GSM1000204
Whole Brain	1	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002491	GSM1000094
Whole Brain	2	H3K27ac (ab4729)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002491	GSM1000094
Whole Brain	1	H3K27me3 (07-449)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002725	GSM1000143
Whole Brain	2	H3K27me3 (07-449)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002725	GSM1000143
Whole Brain	1	H3K36me3 (ab9050)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002718	GSM1000072
Whole Brain	2	H3K36me3 (ab9050)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002718	GSM1000072
Whole Brain	1	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002492	GSM1000096
Whole Brain	2	H3K4me1 (ab8895)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002492	GSM1000096
Whole Brain	1	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002493	GSM1000095
Whole Brain	2	H3K4me3 (07-473)	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002493	GSM1000095
Whole Brain	1	H3K9me3 (ab8898)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002726	GSM1000106

	Whole Brain	2	H3K9me3 (ab8898)	M	Embryonic day 14.5	C57BL/6	wgEncodeEM002726	GSM1000106
	Whole Brain	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002494	GSM1000098
	Whole Brain	2	Input	U	Embryonic day 14.5	C57BL/6	wgEncodeEM002494	GSM1000098
Histone, PSU	CH12	2	Input				wgEncodeEM001923	GSM946521
	CH12	1	Input				wgEncodeEM001923	GSM946521
	CH12	2	H3K36me3 (ab9050)				wgEncodeEM002364	GSM946530
	CH12	1	H3K36me3 (ab9050)				wgEncodeEM002364	GSM946530
	CH12	2	H3K27me3 (07-449)				wgEncodeEM002362	GSM946532
	CH12	1	H3K27me3 (07-449)				wgEncodeEM002362	GSM946532
	CH12	2	H3K9me3 (ab8898)				wgEncodeEM002372	GSM946548
	CH12	1	H3K9me3 (ab8898)				wgEncodeEM002372	GSM946548
	CH12	2	H3K4me3 (07-473)				wgEncodeEM002366	GSM946528
	CH12	1	H3K4me3 (07-473)				wgEncodeEM002366	GSM946528
	CH12	2	H3K4me1 (ab8895)				wgEncodeEM002370	GSM946546
	CH12	1	H3K4me1 (ab8895)				wgEncodeEM002370	GSM946546
	Erythroblast	2	Input				wgEncodeEM002350	GSM946544
	Erythroblast	1	Input				wgEncodeEM002350	GSM946544
	Erythroblast	2	H3K36me3 (ab9050)				wgEncodeEM002375	GSM946543
	Erythroblast	1	H3K36me3 (ab9050)				wgEncodeEM002375	GSM946543
	Erythroblast	2	H3K27me3 (07-449)				wgEncodeEM002371	GSM946547
	Erythroblast	1	H3K27me3 (07-449)				wgEncodeEM002371	GSM946547
	Erythroblast	2	H3K9me3 (ab8898)				wgEncodeEM002373	GSM946549
	Erythroblast	1	H3K9me3 (ab8898)				wgEncodeEM002373	GSM946549
	Erythroblast	2	H3K4me3 (07-473)				wgEncodeEM002380	GSM946524
	Erythroblast	1	H3K4me3 (07-473)				wgEncodeEM002380	GSM946524
	Erythroblast	2	H3K4me1 (ab8895)				wgEncodeEM002376	GSM946536
	Erythroblast	1	H3K4me1 (ab8895)				wgEncodeEM002376	GSM946536
	G1E	2	Input				wgEncodeEM001916	GSM946538
	G1E	1	Input				wgEncodeEM001916	GSM946538
	G1E	2	H3K36me3 (ab9050)				wgEncodeEM002365	GSM946529
	G1E	1	H3K36me3 (ab9050)				wgEncodeEM002365	GSM946529
	G1E	2	H3K27me3 (07-449)				wgEncodeEM002363	GSM946531
	G1E	1	H3K27me3 (07-449)				wgEncodeEM002363	GSM946531
	G1E	2	H3K9me3 (ab8898)				wgEncodeEM002374	GSM946542
	G1E	1	H3K9me3 (ab8898)				wgEncodeEM002374	GSM946542
	G1E	2	H3K4me3 (07-473)				wgEncodeEM001919	GSM946533
	G1E	1	H3K4me3 (07-473)				wgEncodeEM001919	GSM946533
	G1E	2	H3K4me1 (ab8895)				wgEncodeEM001915	GSM946535
	G1E	1	H3K4me1 (ab8895)				wgEncodeEM001915	GSM946535
	G1E-ER4	1	H3K27me3 (07-449)	Estradiol 24 hour diff protocol			wgEncodeEM001917	GSM946537
	G1E-ER4	2	H3K27me3 (07-449)	Estradiol 24 hour diff protocol			wgEncodeEM001917	GSM946537
	G1E-ER4	1	H3K36me3 (ab9050)	Estradiol 24 hour diff protocol			wgEncodeEM002368	GSM946526
	G1E-ER4	2	H3K36me3 (ab9050)	Estradiol 24 hour diff protocol			wgEncodeEM002368	GSM946526
	G1E-ER4	1	H3K4me1 (ab8895)	Estradiol 24 hour diff protocol			wgEncodeEM001918	GSM946534
	G1E-ER4	2	H3K4me1 (ab8895)	Estradiol 24 hour diff protocol			wgEncodeEM001918	GSM946534
	G1E-ER4	1	H3K4me3 (07-473)	Estradiol 24 hour diff protocol			wgEncodeEM001920	GSM946519
	G1E-ER4	2	H3K4me3 (07-473)	Estradiol 24 hour diff protocol			wgEncodeEM001920	GSM946519
	G1E-ER4	1	H3K9me3 (ab8898)	Estradiol 24 hour diff protocol			wgEncodeEM002377	GSM946545
	G1E-ER4	2	H3K9me3 (ab8898)	Estradiol 24 hour diff protocol			wgEncodeEM002377	GSM946545
	G1E-ER4	1	Input	Estradiol 24 hour diff protocol			wgEncodeEM001921	GSM946520
	G1E-ER4	2	Input	Estradiol 24 hour diff protocol			wgEncodeEM001921	GSM946520
	Megakaryocyte	2	Input				wgEncodeEM002352	GSM946539
	Megakaryocyte	1	Input				wgEncodeEM002352	GSM946539
	Megakaryocyte	2	H3K36me3 (ab9050)				wgEncodeEM002378	GSM946540
	Megakaryocyte	1	H3K36me3 (ab9050)				wgEncodeEM002378	GSM946540
	Megakaryocyte	2	H3K27me3 (07-449)				wgEncodeEM002381	GSM946523
	Megakaryocyte	1	H3K27me3 (07-449)				wgEncodeEM002381	GSM946523
	Megakaryocyte	2	H3K9me3 (ab8898)				wgEncodeEM002379	GSM946541
	Megakaryocyte	1	H3K9me3 (ab8898)				wgEncodeEM002379	GSM946541
	Megakaryocyte	2	H3K4me3 (07-473)				wgEncodeEM002367	GSM946527
	Megakaryocyte	1	H3K4me3 (07-473)				wgEncodeEM002367	GSM946527
	Megakaryocyte	2	H3K4me1 (ab8895)				wgEncodeEM002369	GSM946525
	Megakaryocyte	1	H3K4me1 (ab8895)				wgEncodeEM002369	GSM946525
Histone, Yale/Stanford	CH12	1	H3K4me3 (07-473)	IgG-Yale	None		wgEncodeEM002004	GSM798327
	CH12	2	H3K4me3 (07-473)	IgG-Yale	None		wgEncodeEM002004	GSM798327
	CH12		Input	IgG-Yale	None		wgEncodeEM002001	GSM798326
	ES-E14	1	H3K4me1 (ab8895)	Standard Control	None		wgEncodeEM003322	GSM1003750
	ES-E14	2	H3K4me1 (ab8895)	Standard Control	None		wgEncodeEM003322	GSM1003750
	ES-E14	1	H3K4me3 (07-473)	Standard Control	None		wgEncodeEM003328	GSM1003756
	ES-E14	2	H3K4me3 (07-473)	Standard Control	None		wgEncodeEM003328	GSM1003756
	ES-E14	1	H3K9me3 (ab8898)	Standard Control	None		wgEncodeEM003323	GSM1003751
	ES-E14	2	H3K9me3 (ab8898)	Standard Control	None		wgEncodeEM003323	GSM1003751
	ES-E14	1	Input	Standard Control	None		wgEncodeEM002788	GSM1003746
	MEL	1	H3K27ac (ab4729)	IgG-rab	None		wgEncodeEM003324	GSM1003752
	MEL	2	H3K27ac (ab4729)	IgG-rab	None		wgEncodeEM003324	GSM1003752
	MEL	1	H3K27me3 (9733S)	Standard Control	DMSO_2.0pct		wgEncodeEM003331	GSM1003745
	MEL	2	H3K27me3 (9733S)	Standard Control	DMSO_2.0pct		wgEncodeEM003331	GSM1003745
	MEL	1	H3K36me3 (9763S)	Standard Control	DMSO_2.0pct		wgEncodeEM003332	GSM1003749
	MEL	2	H3K36me3 (9763S)	Standard Control	DMSO_2.0pct		wgEncodeEM003332	GSM1003749
	MEL	1	H3K4me1 (ab8895)	IgG-rab	None		wgEncodeEM003330	GSM1003744
	MEL	2	H3K4me1 (ab8895)	IgG-rab	None		wgEncodeEM003330	GSM1003744
	MEL	1	H3K4me1 (ab8895)	Standard Control	DMSO_2.0pct		wgEncodeEM003329	GSM1003757
	MEL	2	H3K4me1 (ab8895)	Standard Control	DMSO_2.0pct		wgEncodeEM003329	GSM1003757
	MEL	1	H3K4me3 (07-473)	IgG-Yale	DMSO_2.0pct		wgEncodeEM002005	GSM798328
	MEL	2	H3K4me3 (07-473)	IgG-Yale	DMSO_2.0pct		wgEncodeEM002005	GSM798328
	MEL	1	H3K4me3 (07-473)	IgG-Yale	None		wgEncodeEM002002	GSM798324
	MEL	2	H3K4me3 (07-473)	IgG-Yale	None		wgEncodeEM002002	GSM798324
	MEL	1	H3K4me3 (07-473)	IgG-rab	None		wgEncodeEM003325	GSM1003753
	MEL	2	H3K4me3 (07-473)	IgG-rab	None		wgEncodeEM003325	GSM1003753

	MEL	1	H3K4me3 (07-473)	Standard Control	DMSO_2.0pct		wgEncodeEM003326	GSM1003754
	MEL	2	H3K4me3 (07-473)	Standard Control	DMSO_2.0pct		wgEncodeEM003326	GSM1003754
	MEL	1	H3K9me3 (ab8898)	IgG-rab	None		wgEncodeEM003327	GSM1003755
	MEL	2	H3K9me3 (ab8898)	IgG-rab	None		wgEncodeEM003327	GSM1003755
	MEL	1	H3K9me3 (ab8898)	Standard Control	DMSO_2.0pct		wgEncodeEM003333	GSM1003748
	MEL	2	H3K9me3 (ab8898)	Standard Control	DMSO_2.0pct		wgEncodeEM003333	GSM1003748
	MEL	1	Input	IgG-Yale	DMSO_2.0pct		wgEncodeEM001963	GSM798323
	MEL		Input	IgG-Yale	None		wgEncodeEM002003	GSM798325
	MEL	1	Input	IgG-rab	None		wgEncodeEM001947	GSM1003747
	MEL	1	Input	Standard Control	DMSO_2.0pct		wgEncodeEM002791	GSM1003758
Replication Timing	CH12	1	F	None	GPL10989	CH12_R1	wgEncodeEM002962	GSM1113245
	CH12	2	F	None	GPL10989	CH12_R2	wgEncodeEM002962	GSM1113246
	ES-46C	1	M	None	GPL9156	ES-46C_R1	wgEncodeEM002964	GSM450272
	ES-46C	1	M	diffProtF_6d	GPL9156	NP-46C_R1	wgEncodeEM002963	GSM450283
	ES-D3	1	M	None	GPL9156	ES-D3_R1	wgEncodeEM002968	GSM450273
	ES-D3	2	M	None	GPL9156	ES-D3_R2	wgEncodeEM002968	GSM450273
	ES-D3	1	M	diffProtE_3d	GPL9156	EBM3_R1	wgEncodeEM002965	GSM450279
	ES-D3	2	M	diffProtE_3d	GPL9156	EBM3_R2	wgEncodeEM002965	GSM450279
	ES-D3	1	M	diffProtE_6d	GPL9156	EBM6_R1	wgEncodeEM002966	GSM450282
	ES-D3	2	M	diffProtE_6d	GPL9156	EBM6_R2	wgEncodeEM002966	GSM450282
	ES-D3	1	M	diffProtE_9d	GPL9156	NP-D3_R1	wgEncodeEM002967	GSM450285
	ES-D3	2	M	diffProtE_9d	GPL9156	NP-D3_R2	wgEncodeEM002967	GSM450285
	ES-D3	1	M	diffProtG_3d	GPL9156		wgEncodeEM002976	GSM450278
	ES-D3	2	M	diffProtG_3d	GPL9156		wgEncodeEM002976	GSM450278
	ES-EM5Sox17huCD25	1	U	diffProtH_Sox17+	GPL9156		wgEncodeEM002978	GSM450287
	ES-EM5Sox17huCD25	2	U	diffProtH_Sox17+	GPL9156		wgEncodeEM002978	GSM450287
	ES-EM5Sox17huCD25	1	U	diffProtH_Sox17-	GPL9156		wgEncodeEM002977	GSM450286
	ES-EM5Sox17huCD25	2	U	diffProtH_Sox17-	GPL9156		wgEncodeEM002977	GSM450286
	ES-TT2	1	M	None	GPL9156	ES-TT2_R1	wgEncodeEM002970	GSM450274
	ES-TT2	2	M	None	GPL9156	ES-TT2_R2	wgEncodeEM002970	GSM450274
	ES-TT2	1	M	diffProtF_9d	GPL9156	NP-TT2_R1	wgEncodeEM002969	GSM450284
	ES-TT2	2	M	diffProtF_9d	GPL9156	NP-TT2_R2	wgEncodeEM002969	GSM450284
	EpiSC-5	1	M	None	GPL9156	EpiSC5_R1	wgEncodeEM002971	GSM450280
	EpiSC-5	2	M	None	GPL9156	EpiSC5_R2	wgEncodeEM002971	GSM450280
	EpiSC-7	1	F	None	GPL9156	EpiSC7_R1	wgEncodeEM002972	GSM450281
	EpiSC-7	2	F	None	GPL9156	EpiSC7_R2	wgEncodeEM002972	GSM450281
	J185a	1	U	None	GPL9156		wgEncodeEM002979	GSM450293
	J185a	2	U	None	GPL9156		wgEncodeEM002979	GSM450293
	L1210	1	F	None	GPL9156	L1210_R1	wgEncodeEM002973	GSM1112425
	L1210	2	F	None	GPL9156	L1210_R2	wgEncodeEM002973	GSM1112426
	MEF	1	M	None	GPL9156	MEF_R1	wgEncodeEM002974	GSM450292
	MEL	1	M	None	GPL10989	MEL_R1	wgEncodeEM002975	GSM1113247
	MEL	2	M	None	GPL10989	MEL_R2	wgEncodeEM002975	GSM1113248

Supplementary Table 9: List of 76 predicted promoters tested by reporter assay.

				Forward	Reservse	tested positive mESCs in one direstion	tested positive in mESCs in both direstions	tested negative in mESCs	tested positive in MEF in one direstion	tested positive in MEF in both direstions	tested negative in MEF
mESC and MEF	me3+polII	chr1:72,282,608-72,283,862	1	CCCCATGGCTACTAGCAAGA	TCAACAAGACCCGGGTAACT	1				1	
		chr11:120,176,972-120,178,515	2	GGCTCGCGCATATTAACACT	CCCTGCTAATTGGCTCTCTG		1		1		
		chr1:88,383,424-88,384,740	3	CCACTGAGCCATCTTTCTCC	ACGACGGCTTTTGTTGTACC	1			1		
		chr2:32,821,808-32,823,140	4	TAGCACATTCCCCTGTCCTC	CCCTGAAGCACTCTGCTACC		1			1	
		chr8:46,019,901-46,021,052	5	GCATTTGGAAAGGATTTGGA	TGAAACACCCCCACGTTATT	1				1	
		chr16:94,534,005-94,535,452	6	AAACCCCTCTCTCCGATGAC	GAGCTGGGGATCCAGATTG	1			1		
		chr4:133,466,066-133,467,413	7	AGGTTTTGGATGCTTGTTCG	GGCGGTGTCTGGAGAGTAGA	1			1		
		chr10:81,640,535-81,641,906	8	TTCTGCTTAAGTTCCTGAAGTTTT	TGCGCGGAGTTAACTGTAGA	1			1		
		chr12:113,931,008-113,932,428	9	AGCCTTCACCTTTGCACTGT	AGCTGTCTGCAACATCATGG		1			1	
		chr16:21,332,693-21,334,078	11	GCTGCTGAACAGAACCTTCC	AAGAACCTGTTCCGCACTGT			1		1	
		chr6:117,829,050-117,830,421	12	GTACCGTTCGGTCCCTACAC	GGGTGCTTTGAGATTTTCGT		1			1	
		chr7:88,667,757-88,668,832	13	CAAGTCAAAGCACACACAGGA	CAACAGCTCTGTGCATGTGA			1	1		
	me3 ONLY	chr1:4,561,098-4,562,521	1	CAAGTGCCAGACCAGTTTGA	TTGATCCCATTTTCCCAGAG			1			1
		chr13:23,487,516-23,488,786	2	GGTGATGCTTTCCTGGGTTA	CAGATCCCGCCTCTCTACTG			1	1		
		chr1:172,963,082-172,964,359	3	CCAGGGTAGTAAAATGTCTTCTG	AGATGGTGCCCTTTTTGTTG			1	1		
		chr18:24,625,130-24,626,521	4	CCTTGGGGTCCTGATGTCTA	GAGAAGCAACCGAGAGCAAC			1			1
		chr19:61,275,284-61,276,321	5	ACCCAAACACGACACCATCT	CTCCTTGCACACCCTGTTTT	1			1		
		chr13:21,366,741-21,367,942	6	GCTGCTCGTTGGAGTAGACC	CCACAAGACAAAATGCTCCA	1			1		
		chr11:37,049,105-37,050,498	7	GTAGCTTCGCTCCCTGACAC	ATCGGGTTTAGCAGAGCAGA	1			1		
		chr14:51,702,299-51,703,553	8	AGAGGCAGGTGCTCAGAAAA	CTGCCTGGTTGTGGAGATTT			1		1	
		chr4:88,513,577-88,514,670	9	GGACTTCAGATTCCCCCAAG	GTGGTGTCAGGTGCTGTGAC			1		1	
		chr8:52,000,477-52,001,938	10	GCTACTGCTGCTTCCAAACC	CACTTAGTGGGGAGGAGAGG			1	1		
		chr13:21,947,190-21,948,224	11	CAGCGTTGGCTTTCCTTACT	TCTCATGATCGCAAAACCAA			1			1
		chr19:38,555,283-38,556,504	12	CCCTCCTGGATCTGAGACAA	ACCAAGTGTGCCAAGGCTTA	1			1		
		chr6:70,962,454-70,963,748	13	GGAGGGAGGATGAAGTAGGG	AGCCTGGAGGAAGCTTTAGG			1		1	
mESC only	polII+me3	chr2:75,471,401-75,472,561	1	CATTCACTTTGGTGGGCTCT	TCATTGGGCTAATGTCAAAGG			1			
		chr12:18,393,657-18,394,810	2	AGTCGAAGGTCATGGGTTTG	AACACCACCGCTCACCTC	1					
		chr15:100,922,977-100,924,412	3	CTGAAACCCACACTCCCATC	TGGAACTGAAGGAACCCAAG		1				
		chr2:166,903,330-166,904,794	4	CACTGCCGGAAGGTTAGAAG	GCTGGCCTTAGGAGTTCAGA		1				
		chr2:168,805,839-168,807,165	5	CTGGGGCTATGAACCGAATA	TGACCATCTCGACACCCATA			1			
		chr4:138,262,409-138,263,533	6	CTCAGGCGGTCCTAAGAATG	TAGCACTGTGCGTTTGCTCT		1				
		chr17:30,118,145-30,119,424	7	AGCTCAGACCACACCGTTCT	GCTATGCCTGGCTATCTTAGTTC		1				
		chr8:109,632,424-109,633,774	8	AATCTGACCGCCAATAGCTG	GTTGACTCTGGCAGGGACTG		1				
		chr4:154,028,643-154,030,048	9	GCTGGGACTCTGAGAACTGG	TGAGTGCAGAGAGGTCATGG		1				
		chr8:73,214,953-73,216,209	10	CGTCTGCATCTGTTTTGTGG	GGACACTGATCCGTCCAGTC	1					
		chr9:114,572,308-114,573,743	11	TGGGCAGGACTTATTCAACC	AGAGGGCCACAGCCTAAAAT	1					
		chr7:50,508,728-50,509,835	12	CTGCAAGTTGAATCCTCAGC	AAAAGTTGGGATGGGAGGTC		1				
		chr5:110,514,708-110,515,903	13	CCCTGGTTGGCACATTACTT	CTGGGCACCCTTCCTCTTAT	1					
	me3 only	chr8:91,576,369-91,577,522	1	GCTACAGCCATAGAATCCAATTTT	AAATGCTCGGAGCTGAAAGA	1					
		chr14:76,915,379-76,916,761	2	AACAGGAGCAGCATGGAGAT	ACAATGGAGCAGAGGTGTCC		1				
		chr4:133,512,988-133,514,350	3	TGGGCAATAAGAGCTGGACT	TGGTTGGTTGGGTTTTGG			1			
		chr9:113,954,352-113,955,827	4	CCCACGGTATGGAATAATCG	CAGACCCTGCCATACTGGAC	1					
		chr2:167,088,587-167,089,586	5	CATCTCCTCTGCCTTTGACA	TCTCATTCCGCTTTTAAACCA	1					
		chr19:55,803,480-55,804,780	6	TATGGACCGAAGCACAAACA	AATTGTTCCTGATGGCGAAG	1					
		chr5:33,876,910-33,878,336	7	CAGGGACAGGGTTAAGAGCA	ACTGCGGTTTGAGGTGAGTT			1			
		chr8:28,332,972-28,334,480	8	AGGAGGCGCCTAACACTTCT	CGTCATTCCTCAAACCTGCT	1					
		chr4:140,840,789-140,842,238	9	GCAGGATCTGACTTGGGGTA	GGGCTATTGGGAGGTTTAGG	1					
		chr4:44,965,616-44,967,013	10	TGAACTTGGGAGGGGTACAG	GAGGCATTGAAAGCATCTGG	1					
		chr17:37,107,047-37,108,640	11	CTCCAAGTGCCTCAGTAGCC	TCCCTCCAGACTTTCCACAC			1			
		chr3:96,373,157-96,374,474	12	GAGCCCTCACTCCAGTCCTA	ATATCTAGGCGGCCGTGTC	1					
		chr17:29,655,802-29,657,008	13	GGCTGCATATAACTCAACACCTC	GGTTCCAGCCACTCAGGTTA			1			
MEF only	polII+me3	chr7:106,503,912-106,505,303	1	TCCATTTGCAGTCAGTGGAG	TCCAGTTCTGCGTCTTCCTT				1		
		chr12:70,602,990-70,604,165	2	CGTTAGAGCCAGAAGCCAGT	CGCCCTACACCATAACCAAT				1		
		chr8:131,177,663-131,179,109	3	GGCAAGGCAAACACTACCAT	CCCTTTTCTAGCCTGCCTTC					1	
		chr6:31,037,647-31,038,728	4	CAGTCCTGAACAGCGACAAT	CAGGCGGTCTCCTAAAAATG					1	
		chr6:31,037,357-31,038,733	6	ACACACGACAACCAGCAAAG	ACCTAGCCCTGTGTGTCGAG				1		
		chr10:5,056,630-5,057,713	7	TCCATCCACATAAGGGTGAG	GGCATTTTCCAAGCTGAATG					1	
		chr6:4,406,665-4,407,901	8	ATCTCCGGAAGCCCTAACTC	GGAAGGAAGGCAAAGGAAAC					1	
		chr2:30,319,171-30,320,263	9	TTTCCAGTTGGTGGATGACA	AAGCAGCAACAGCACATCAC					1	
		chr1:184,449,607-184,450,750	10	TCAACTCCCAGCACTTAGCC	AGGCTTAGTCCAGTCCACCA					1	
		chr11:94,793,896-94,795,214	11	GCTTCAGTCTCCATGTTCCTG	AGCAAAAGCCAGAATCTCCA						1
		chr1:173,192,458-173,193,486	12	AGTTCTCAATGCTGGGCAAC	CCCCCAGCAACAGTCAAT					1	
		chr6:4,438,664-4,440,057	13	TCCCCCAATTTTTCTCTGTG	GCATGGAATTACGCTGTGTG				1		
	me3 only	chr19:54,189,128-54,190,170	1	GACAAGACACCATGGCAGAA	TCACACTGTCTGCTGGGAAG				1		
		chr1:78,199,770-78,201,041	2	TGGTAGTGGGGATCCATGTT	GGGGAGATAGAAGGGATTGG				1		
		chr4:8,954,712-8,956,015	3	TGGAGATGGGGACAGAGTTC	TACGTGGGGAAAGACGTAGG						1
		chr11:96,176,070-96,177,089	4	CGGTTTTCCGGATACACAAT	GCCCTCCTATAGGCCAGACT				1		
		chr18:13,107,342-13,108,338	5	TGCACGTGACACTGTGAGTC	GCTTGCCTTGGTGAACTCTG				1		
		chr3:69,907,851-69,908,955	6	AATGAACCAAACCAGAGAACG	TGGAAAAGGATTGTGGGAAG				1		
		chr14:118,400,191-118,401,401	7	CAGGGTCAGTGAGCTTGACA	AATCCCACGTACAGGCTTTG					1	
		chr5:13,628,493-13,629,571	8	AGTCGAAAAACTGCCACCAT	GGGAACGACAACAACAACAA				1		
		chr1:156,730,738-156,732,002	9	ACCCCACCAAGGAACATACA	GATTCTAGCGGGGTCTAGGG				1		
		chr14:68,230,847-68,232,082	10	CTACCCCAGCTTCCACAAAA	CGATCAGACTGGGATTTGCT						1
		chr10:126,246,319-126,247,564	11	CTGGTCAGCACGCATAACAT	ACCTGTCTCTCCCCAGTGTG				1		
		chr2:65,656,036-65,657,085	12	TTTTTGCATGGAAAGCCAGT	CCACATGAAAAACAGAGTTTGC					1	
		chr14:87,967,514-87,968,917	13	ATCTCGGGTTCTGGTGACTG	GGTTTATGGCGTGCTGACTT				1		

Supplementary Table 10: List of 183 predicted enhancers tested in a high throughput reporter assay.

ID	chr	start	end	# reads RNA	RPKM	# reads DNA	RPKM	log2(RNA/DNA)	Validation
ESC-E1	chr1	182854696	182855674	99971	4910.627	114185	3986.0389	0.301	Positive
ESC-E10	chr2	171811722	171812747	119000	5577.5654	183281	6104.9676	-0.1303	Negative
ESC-E100	chr15	61826135	61827093	38497	1930.4596	41533	1480.1274	0.3832	Positive
ESC-E101	chr1	55165093	55166107	196788	9323.4378	181721	6118.604	0.6077	Positive
ESC-E102	chr1	89625428	89626417	96502	4687.561	171590	5923.3878	-0.3376	Negative
ESC-E103	chr1	120586239	120587188	87195	4413.8145	61221	2202.401	1.0029	Positive
ESC-E104	chr1	121295046	121296025	86568	4247.9324	93656	3266.0731	0.3792	Positive
ESC-E105	chr1	133045795	133046795	44945	2159.2282	62554	2135.7032	0.0158	Positive
ESC-E106	chr1	137753335	137754335	75408	3622.6845	101514	3465.8347	0.0639	Positive
ESC-E107	chr1	145402355	145403355	104899	5039.4456	143837	4910.7823	0.0373	Positive
ESC-E108	chr1	166447135	166448135	86535	4157.2307	119261	4071.7342	0.03	Positive
ESC-E109	chr10	60534775	60535775	88918	4271.7111	110747	3781.0582	0.176	Positive
ESC-E11	chr17	15816990	15818001	136245	6474.1774	190643	6438.0367	0.0081	Positive
ESC-E110	chr10	75111855	75112855	103280	4961.6681	63960	2183.7054	1.184	Positive
ESC-E111	chr10	99626895	99627895	174194	8368.4088	310776	10610.239	-0.3424	Negative
ESC-E112	chr11	5872835	5873835	124998	6005.0107	89621	3059.7964	0.9727	Positive
ESC-E113	chr11	9021995	9022995	88593	4256.0979	150058	5123.1732	-0.2675	Negative
ESC-E114	chr1	21245800	21246800	165597	7955.405	123709	4223.5932	0.9135	Positive
ESC-E115	chr1	36901550	36902550	87005	4179.8097	88044	3005.9561	0.4756	Positive
ESC-E116	chr1	37956700	37957700	126586	6081.2989	144061	4918.4299	0.3062	Positive
ESC-E117	chr1	38115028	38115996	146995	7294.9605	232935	8215.2983	-0.1714	Negative
ESC-E118	chr1	53923075	53924075	81676	3923.8021	178256	6085.8798	-0.6332	Negative
ESC-E12	chr10	92826779	92827779	120126	5770.9576	134607	4595.6613	0.3285	Positive
ESC-E120	chr2	20689850	20690850	77534	3724.8186	138298	4721.6756	-0.3421	Negative
ESC-E121	chr2	20701043	20702018	301239	14842.396	292618	10246.206	0.5346	Positive
ESC-E124	chr2	28929984	28930939	123541	6214.3813	133756	4781.5604	0.3781	Positive
ESC-E125	chr2	29584734	29585699	54615	2718.842	57738	2042.7018	0.4125	Positive
ESC-E126	chr2	30336557	30337536	131794	6467.1674	115913	4042.2313	0.678	Positive
ESC-E127	chr2	32794674	32795659	79008	3853.3723	145350	5037.9303	-0.3867	Negative
ESC-E128	chr2	50862975	50863975	137561	6608.5429	181269	6188.7465	0.0947	Positive
ESC-E129	chr2	109748850	109749850	85117	4089.1093	201730	6887.3045	-0.7522	Negative
ESC-E13	chr8	74857499	74858512	164899	7820.311	249553	8410.7969	-0.105	Positive
ESC-E130	chr2	120339667	120340635	191375	9497.4036	166049	5856.3351	0.6975	Positive
ESC-E131	chr3	33905123	33906100	108730	5346.3302	236214	8254.2788	-0.6266	Negative
ESC-E132	chr11	20136900	20137872	13752	679.7139	7583	266.3909	1.3514	Positive
ESC-E133	chr7	51757050	51758050	141885	6816.2698	168198	5742.4901	0.2473	Positive
ESC-E134	chr7	52654300	52655300	27015	1297.8629	53714	1833.8972	-0.4988	Negative
ESC-E135	chr3	34544700	34545700	61508	2954.9222	86048	2937.8108	0.0084	Positive
ESC-E137	chr3	94875800	94876800	116657	5604.3052	176733	6033.8831	-0.1066	Positive
ESC-E138	chr3	107767000	107768000	105290	5058.2294	103497	3533.5363	0.5175	Positive
ESC-E139	chr3	129276350	129277350	137918	6625.6933	190948	6519.1968	0.0234	Positive
ESC-E14	chr12	25854076	25855103	199325	9324.2126	182374	6062.9378	0.621	Positive
ESC-E140	chr4	10970754	10971716	105276	5257.1262	175656	6233.7576	-0.2458	Negative
ESC-E142	chr4	48634200	48635200	141674	6806.1333	191425	6535.482	0.0585	Positive
ESC-E143	chr4	55003500	55004500	79981	3842.3735	195862	6686.9654	-0.7994	Negative
ESC-E144	chr4	108306800	108307800	115099	5529.4582	148064	5055.0961	0.1294	Positive
ESC-E145	chr4	116842950	116843950	294217	14134.368	333854	11398.144	0.3104	Positive
ESC-E147	chr5	38911677	38912647	4036	199.9317	1403	49.4297	2.0161	Positive
ESC-E148	chr5	67111900	67112900	36989	1777.0183	47945	1636.938	0.1185	Positive
ESC-E149	chr5	76022475	76023475	142416	6841.7793	128531	4388.2209	0.6407	Positive
ESC-E15	chr4	115583944	115584945	104069	4994.5825	106089	3618.4149	0.465	Positive
ESC-E150	chr5	128129750	128130750	60046	2884.6871	73107	2495.9927	0.2088	Positive
ESC-E151	chr6	30103450	30104450	182968	8789.9158	120901	4127.7254	1.0905	Positive
ESC-E152	chr6	30107125	30108125	92425	4440.189	151301	5165.6104	-0.2183	Negative
ESC-E153	chr6	30147450	30148450	137014	6582.2648	149179	5093.1633	0.37	Positive
ESC-E154	chr6	35185250	35186250	78070	3750.5682	134610	4595.7637	-0.2932	Negative
ESC-E155	chr6	49082350	49083350	110865	5326.0549	124127	4237.8641	0.3297	Positive
ESC-E156	chr6	54622600	54623600	127322	6116.6567	155960	5324.6731	0.2001	Positive

ESC-E158	chr6	100402200	100403200	100071	4807.5063	151370	5167.9661	-0.1043	Positive
ESC-E16	chr16	13732499	13733518	193782	9136.0153	176201	5903.6635	0.63	Positive
ESC-E160	chr6	125391550	125392550	175850	8447.9638	186422	6364.6748	0.4085	Positive
ESC-E161	chr6	136761475	136762475	64536	3100.3887	160257	5471.3768	-0.8195	Negative
ESC-E162	chr7	4773250	4774250	201263	9668.8159	224619	7668.7568	0.3343	Positive
ESC-E163	chr7	36247400	36248400	118584	5696.8792	169491	5786.6343	-0.0226	Positive
ESC-E164	chr7	50802475	50803475	131815	6332.5024	122851	4194.3003	0.5943	Positive
ESC-E165	chr7	52654300	52655300	27015	1297.8629	53714	1833.8972	-0.4988	Negative
ESC-E166	chr7	71974650	71975650	120744	5800.6466	193127	6593.5899	-0.1849	Negative
ESC-E167	chr7	86284050	86285050	218858	10514.088	251297	8579.5692	0.2933	Positive
ESC-E168	chr7	87383100	87384100	135604	6514.5277	78630	2684.5532	1.279	Positive
ESC-E169	chr7	134000050	134001050	91807	4410.5	88010	3004.7953	0.5537	Positive
ESC-E17	chr2	5981071	5982079	1833	87.41	5218	176.785	-1.0161	Negative
ESC-E171	chr7	152074175	152075175	123845	5949.62	98740	3371.1277	0.8196	Positive
ESC-E172	chr8	24509700	24510700	147245	7073.7666	122988	4198.9776	0.7524	Positive
ESC-E173	chr8	42173350	42174350	147532	7087.5542	198502	6777.0975	0.0646	Positive
ESC-E174	chr8	72536950	72537950	116816	5611.9437	173263	5915.414	-0.076	Positive
ESC-E175	chr8	74851250	74852250	91732	4406.8969	138280	4721.061	-0.0993	Positive
ESC-E177	chr8	89298950	89299950	142272	6834.8615	217445	7423.8296	-0.1193	Negative
ESC-E178	chr8	91497750	91498750	94639	4546.5505	139775	4772.1018	-0.0699	Positive
ESC-E179	chr8	119458400	119459400	125547	6031.3849	152383	5202.5509	0.2133	Positive
ESC-E18	chr7	4773684	4774673	244295	11866.493	224904	7763.803	0.6121	Positive
ESC-E180	chr8	124404400	124405400	54451	2615.9008	82966	2832.5883	-0.1148	Positive
ESC-E181	chr9	60669000	60670000	219020	10521.87	195455	6673.0701	0.657	Positive
ESC-E182	chr9	99053600	99054600	122931	5905.711	131838	4501.125	0.3918	Positive
ESC-E183	chr9	118991100	118992100	188801	9070.1357	128401	4383.7826	1.0489	Positive
ESC-E184	chr9	121523275	121524275	178960	8597.3696	129001	4404.2671	0.965	Positive
ESC-E185	chr10	42366000	42367000	98274	4721.1776	173275	5915.8237	-0.3254	Negative
ESC-E187	chr11	49619300	49620300	112432	5401.3343	139534	4763.8738	0.1812	Positive
ESC-E188	chr12	77310200	77311200	191581	9203.6882	175722	5999.3666	0.6174	Positive
ESC-E189	chr13	112509750	112510750	128728	6184.2016	159107	5432.1147	0.1871	Positive
ESC-E19	chr7	36252539	36253574	118044	5479.3533	149415	4928.8837	0.1527	Positive
ESC-E190	chr14	27860050	27861050	99358	4773.2534	113443	3873.1022	0.3015	Positive
ESC-E191	chr14	30763950	30764950	302653	14539.638	216059	7376.5103	0.979	Positive
ESC-E192	chr14	56169700	56170700	119567	5744.103	223161	7618.9793	-0.4075	Negative
ESC-E193	chr15	99380100	99381100	167678	8055.3772	138335	4722.9388	0.7703	Positive
ESC-E194	chr16	65648850	65649850	37381	1795.8502	73106	2495.9585	-0.4749	Negative
ESC-E195	chr17	35635750	35636750	114367	5494.2926	168392	5749.1134	-0.0654	Positive
ESC-E196	chrX	73843900	73844900	27761	1333.701	87931	3002.0982	-1.1705	Negative
ESC-E197	chr4	55527350	55528397	115409	5295.7037	74457	2428.0796	1.125	Positive
ESC-E198	chr16	48303488	48304467	24221	1188.5719	20876	728.0493	0.7071	Positive
ESC-E2	chr1	120544330	120545378	214978	9855.1199	124276	4048.805	1.2834	Positive
ESC-E20	chr15	84487613	84488671	131931	5990.95	128351	4142.0779	0.5324	Positive
ESC-E200	chr8	74857330	74858302	149773	7402.268	203643	7152.69	0.0495	Positive
ESC-E201	chr5	120162398	120163368	119338	5910.23	121477	4275.5268	0.4671	Positive
ESC-E202	chr2	168594397	168595368	96819	4790.0499	98344	3457.782	0.4702	Positive
ESC-E21	chr12	55211145	55212173	156089	7294.5932	169511	5629.8405	0.3737	Positive
ESC-E22	chr13	64081221	64082236	181170	8575.0426	171837	5780.1147	0.569	Positive
ESC-E23	chr17	41036531	41037557	128473	6015.7006	222079	7390.0897	-0.2969	Negative
ESC-E26	chr7	86284032	86285007	218858	10783.402	251297	8799.3314	0.2933	Positive
ESC-E27	chr8	94445910	94446903	26871	1300.0358	29449	1012.5473	0.3606	Positive
ESC-E28	chr10	27731682	27732646	69031	3440.0467	95025	3365.3229	0.0317	Positive
ESC-E29	chr2	171704503	171705469	46255	2300.2913	68339	2415.2431	-0.0704	Positive
ESC-E3	chr2	118824013	118825028	179728	8506.7911	244471	8223.2975	0.0489	Positive
ESC-E30	chr4	108052394	108053383	101765	4943.2072	75978	2622.8332	0.9143	Positive
ESC-E31	chr8	42173639	42174632	156135	7553.6688	249913	8592.4045	-0.1859	Negative
ESC-E33	chr14	76480936	76481885	6903	349.4762	12438	447.4919	-0.3567	Negative
ESC-E34	chr18	4403814	4404743	6181	319.6575	10005	367.7079	-0.202	Negative
ESC-E36	chr2	58480149	58481060	65584	3458.2033	88305	3309.076	0.0636	Positive
ESC-E37	chrX	99182025	99182925	51401	2743.4426	76145	2888.2437	-0.0742	Positive
ESC-E38	chr6	17571043	17571991	23794	1205.7588	24090	867.5716	0.4749	Positive

ESC-E39	chr12	120004596	120005499	73161	3891.8667	91732	3467.9148	0.1664	Positive
ESC-E4	chr8	97710983	97711993	121611	5784.5108	188911	6385.8574	-0.1427	Negative
ESC-E41	chr2	147118424	147119401	27810	1367.4744	44666	1560.8524	-0.1908	Negative
ESC-E42	chr12	71038410	71039332	112040	5837.3576	120517	4462.325	0.3875	Positive
ESC-E44	chr14	20680136	20681089	60048	3026.9032	58522	2096.4807	0.5299	Positive
ESC-E45	chr7	4653663	4654607	51640	2627.8699	61909	2238.9346	0.2311	Positive
ESC-E46	chr5	53301450	53302370	80708	4214.085	104564	3880.055	0.1191	Positive
ESC-E47	chr17	7444695	7445609	48431	2545.3766	38776	1448.327	0.8135	Positive
ESC-E48	chr11	100136383	100137293	59738	3153.4095	120007	4501.9716	-0.5136	Negative
ESC-E49	chr18	25547503	25548453	39178	1981.1342	55633	1999.2751	-0.0132	Positive
ESC-E5	chr6	55741713	55742726	167420	7939.8683	207256	6985.2506	0.1848	Positive
ESC-E50	chr1	3472699	3473648	57321	2901.6085	104845	3771.7213	-0.3784	Negative
ESC-E51	chr2	31995916	31996896	94182	4616.8396	90364	3148.0604	0.5524	Positive
ESC-E52	chr10	33949762	33950731	91008	4511.8414	180903	6373.6346	-0.4984	Negative
ESC-E53	chr9	58418474	58419426	101713	5132.5002	142529	5111.2169	0.006	Positive
ESC-E54	chr1	91772860	91773820	84939	4250.4024	104880	3729.7939	0.1885	Positive
ESC-E55	chr17	71121493	71122443	38527	1948.2155	64781	2328.0172	-0.2569	Negative
ESC-E56	chr13	86237126	86238080	37618	1894.2835	44116	1578.7604	0.2629	Positive
ESC-E57	chr3	137496052	137497014	142053	7093.6272	177134	6286.2091	0.1743	Positive
ESC-E58	chr2	141092234	141093199	111037	5527.5885	159794	5653.2333	-0.0324	Positive
ESC-E6	chr2	28462521	28463544	134206	6302.5544	142589	4758.8318	0.4053	Positive
ESC-E60	chr14	79509623	79510586	92812	4629.9145	74243	2632.064	0.8148	Positive
ESC-E61	chr18	44525448	44526440	27494	1331.5153	43865	1509.7081	-0.1812	Negative
ESC-E62	chr3	107798742	107799702	36286	1815.8035	42403	1507.986	0.268	Positive
ESC-E63	chr3	104618095	104619084	22042	1070.7233	38845	1340.991	-0.3247	Negative
ESC-E64	chr8	58730595	58731581	61913	3016.5677	61327	2123.511	0.5065	Positive
ESC-E65	chr13	16921782	16922764	63554	3109.1204	88584	3079.7718	0.0137	Positive
ESC-E66	chr16	56901932	56902893	47611	2380.0306	39730	1411.4601	0.7538	Positive
ESC-E67	chr14	30763757	30764720	298268	14878.949	206702	7327.9134	1.0218	Positive
ESC-E68	chr1	17921912	17922867	66205	3330.2788	79997	2859.7833	0.2197	Positive
ESC-E69	chr6	68802399	68803383	20196	986.035	24504	850.2292	0.2138	Positive
ESC-E7	chr2	129410365	129411379	52590	2491.6499	80591	2713.5525	-0.1231	Negative
ESC-E70	chr4	10970134	10971087	47627	2400.795	69878	2503.2861	-0.0603	Positive
ESC-E71	chr2	36270509	36271461	47863	2415.2227	89698	3216.6687	-0.4134	Negative
ESC-E72	chr7	97687774	97688743	97734	4845.2881	160626	5659.2342	-0.224	Negative
ESC-E73	chr3	83985406	83986372	33310	1656.542	72452	2560.6022	-0.6283	Negative
ESC-E74	chr6	4982813	4983806	68853	3331.0732	76794	2640.3339	0.3353	Positive
ESC-E75	chr14	99747416	99748389	109553	5408.9194	133109	4670.4942	0.2118	Positive
ESC-E76	chr7	65923805	65924781	71225	3505.7849	108904	3809.4712	-0.1199	Negative
ESC-E77	chr15	9382274	9383260	100065	4875.4048	97133	3363.3017	0.5356	Positive
ESC-E78	chr9	9268943	9269894	65060	3286.4339	45016	1616.0437	1.0241	Positive
ESC-E79	chr1	142084819	142085770	40548	2048.2562	62367	2238.9136	-0.1284	Negative
ESC-E8	chr8	91525792	91526825	262634	12214.435	177411	5863.7228	1.0587	Positive
ESC-E80	chr11	97215732	97216701	19638	973.6189	33522	1181.099	-0.2787	Negative
ESC-E81	chr19	55493055	55494004	70199	3553.486	114724	4127.1063	-0.2159	Negative
ESC-E82	chr5	18070759	18071737	54917	2697.5738	43437	1516.356	0.8311	Positive
ESC-E83	chr5	18070851	18071816	54917	2733.8759	42510	1503.9664	0.8622	Positive
ESC-E84	chr19	41021229	41022183	43544	2192.6839	78284	2801.4773	-0.3535	Negative
ESC-E85	chr12	37806323	37807280	98389	4938.8588	142100	5069.2368	-0.0376	Positive
ESC-E86	chr18	31728208	31729158	72561	3669.187	130713	4697.3466	-0.3564	Negative
ESC-E87	chr19	20998236	20999229	25212	1219.7755	30380	1044.5564	0.2237	Positive
ESC-E88	chr6	64951947	64952897	109009	5512.2253	169175	6079.5154	-0.1413	Negative
ESC-E89	chr12	4358406	4359368	90212	4504.8883	88356	3135.6415	0.5227	Positive
ESC-E9	chr4	10859231	10860280	46182	2115.1194	80203	2610.4739	-0.3036	Negative
ESC-E90	chr11	9951289	9952244	91930	4624.2919	144651	5171.0349	-0.1612	Negative
ESC-E91	chr8	109785728	109786686	75394	3780.6381	93882	3345.6456	0.1763	Positive
ESC-E92	chr1	84886510	84887461	89974	4544.9178	83513	2998.0171	0.6002	Positive
ESC-E93	chr10	68497183	68498132	55188	2793.6372	136761	4919.8606	-0.8165	Negative
ESC-E94	chr4	84307834	84308786	14097	711.3863	38189	1369.5282	-0.945	Negative
ESC-E95	chr11	8571508	8572457	29177	1476.9742	60246	2167.3266	-0.5533	Negative
ESC-E96	chr17	35635745	35636736	114367	5544.1396	168392	5801.2723	-0.0654	Positive

ESC-E97	chr3	34566928	34567886	47019	2357.7903	92131	3283.2467	-0.4777	Negative
ESC-E98	chr4	55554567	55555518	62006	3132.1668	164767	5914.8901	-0.9172	Negative
ESC-E99	chr6	122665619	122666581	48264	2410.1674	49284	1749.0483	0.4626	Positive
Neg1	chr8	24322802	24323828	70856	3317.8283	146748	4883.3282	-0.5576	
Neg10	chr6	10435078	10436122	81772	3763.0092	423197	13840.048	-1.8789	
Neg11	chr18	60511562	60512551	102322	4970.2631	227051	7837.918	-0.6571	
Neg12	chr1	6285016	6285968	83123	4194.4466	265527	9521.9915	-1.1828	
Neg13	chr1	7844595	7845548	12782	644.3552	45752	1639.0216	-1.3469	
Neg14	chr2	7728265	7729228	40284	2009.5904	132569	4699.7991	-1.2257	
Neg15	chr3	18648999	18649997	63136	3039.2043	99583	3406.7151	-0.1647	
Neg16	chr4	12792115	12793107	50011	2421.957	183775	6324.8516	-1.3849	
Neg17	chr5	10115318	10116288	25573	1266.5454	115566	4067.4846	-1.6832	
Neg18	chr6	6463178	6464160	76126	3724.1443	188592	6556.6618	-0.8161	
Neg19	chr7	10925589	10926543	19641	989.0616	50089	1792.5069	-0.8578	
Neg2	chr1	4026272	4027308	87789	4071.0633	153002	5042.343	-0.3087	
Neg20	chr8	15664223	15665213	41955	2035.9257	62828	2166.7027	-0.0898	
Neg21	chr9	5813564	5814538	66919	3300.5976	236944	8305.2639	-1.3313	
Neg22	chr11	10534440	10535390	56092	2836.4115	171224	6153.1482	-1.1173	
Neg23	chr12	10534097	10535080	48392	2364.9874	165723	5755.7402	-1.2832	
Neg24	chr13	10676652	10677603	60054	3033.5652	205425	7374.4395	-1.2815	
Neg25	chr14	9870734	9871711	80596	3962.9746	142159	4967.6341	-0.326	
Neg26	chr15	6076833	6077804	56762	2808.2795	209268	7357.8212	-1.3896	
Neg27	chr16	12407909	12408885	49282	2425.738	88054	3080.146	-0.3446	
Neg28	chr17	17036087	17037040	54694	2757.0229	173508	6215.6184	-1.1728	
Neg29	chr18	8424984	8425980	12480	602.0003	43857	1503.3771	-1.3204	
Neg3	chr2	4004184	4005157	130127	6424.7025	178340	6257.5308	0.038	
Neg30	chr19	8519452	8520441	31357	1523.1923	79921	2758.9467	-0.857	
Neg4	chr3	6326753	6327840	91772	4056.2791	233033	7319.8338	-0.8517	
Neg5	chr4	7456053	7456979	117701	6105.8363	292732	10792.007	-0.8217	
Neg6	chr6	7362387	7363366	106281	5215.2472	225610	7867.6449	-0.5932	
Neg7	chr9	6027395	6028301	163816	8685.4577	546883	20606.2	-1.2464	
Neg9	chr3	10996635	10997622	91772	4466.8286	436409	15095.514	-1.7568	

Supplementary Table S14: Summary of the 1:1 and many to 1 human/mouse orthologous genes.

	1:1	1:many (hsap:mmus)	many:1 (hsap:mmus)	many:many
MIT	15566	2293 (625:2293)	977 (977:389)	6177 (771:1099)
Ensembl	15782	1492 (437:1492)	703 (703:294)	2580 (458:552)
Merged set	15736	1527 (484:1527)	758 (758:321)	1835 (396:503)

Supplementary Table 15: Enriched GO term for 1:many human/mouse orthologous genes.

GO:0007608	sensory perception of smell	309	81	8.93	< 1e-30	8.63e-27
GO:0007606	sensory perception of chemical stimulus	358	84	10.35	< 1e-30	8.63e-27
GO:0007600	sensory perception	635	88	18.36	< 1e-30	8.63e-27
GO:0007186	G-protein coupled receptor signaling pat...	938	95	27.12	1.1e-27	9.49e-24
GO:0050877	neurological system process	1146	98	33.14	5.9e-23	5.09e-19
GO:0003008	system process	1459	105	42.19	6.2e-19	5.35e-15
GO:0007166	cell surface receptor signaling pathway	2037	112	58.9	6.6e-12	5.7e-08
GO:0032501	multicellular organismal process	4645	189	134.31	3.0e-08	0.000259
GO:0044707	single-multicellular organism process	4534	184	131.1	6.5e-08	0.000561
GO:0006952	defense response	637	43	18.42	2.2e-07	0.0019
GO:0001906	cell killing	65	11	1.88	2.3e-06	0.0198
GO:0006956	complement activation	33	8	0.95	3.4e-06	0.0293
GO:0045087	innate immune response	231	21	6.68	3.8e-06	0.0328
GO:0030449	regulation of complement activation	10	5	0.29	4.4e-06	0.038
GO:2000257	regulation of protein activation cascade	10	5	0.29	4.4e-06	0.038

Supplementary Table 20: RNA-Seq datasets used for PCA analysis

organism	lab	cell	category	labExpId	localization	readType	rnaExtract	seqPlatform
human	CSHL	SK-N-SH	Brain	LID46598-LID46599	cell	2x101D	longPolyA	Illumina_HiSeq_2000
human	Caltech	HepG2	Liver	Hepg2R2x75Il200Rep1-Hepg2R2x75Il200Rep2	cell	2x75	longPolyA	.
mouse	CSHL	Kidney	Kidney	LID20872-LID20873	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Adipose	Adipose	hbmadiposePolyAP50	cell	2x50	polyA	NA
human	Caltech	MCF-7	Breast	Mcf7R2x75Il200Rep1-Mcf7R2x75Il200Rep2	cell	2x75	longPolyA	.
human	illumina	Whiteblood	Blood	hbmwhitebloodPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Lung	Lung	LID20920-LID20921	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	SubcFatPad	Adipose	LID21181-LID21182	cell	2x76D	longPolyA	Illumina_GA2x
human	CSHL	NHLF	Lung	LID8692-LID8701	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Colon	Intestine	hbmcolonPolyAP50	cell	2x50	polyA	NA
human	illumina	Brain	Brain	hbmbrainPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Testis	Testis	LID20868-LID20869	cell	2x76D	longPolyA	Illumina_GA2x
human	CSHL	MCF-7	Breast	LID8686-LID8687	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	GenitalFatPad	Adipose	LID21179-LID21180	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Breast	Breast	hbmbreastPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Cerebellum	Brain	LID47036-LID47037	cell	2x101D	longPolyA	Illumina_HiSeq_2000
human	illumina	Kidney	Kidney	hbmkidneyPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Heart	Heart	LID20870-LID20871	cell	2x76D	longPolyA	Illumina_GA2x
human	CSHL	A549	Lung	LID8963-LID8964	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	Colon	Intestine	LID21040-LID21041	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	Cortex	Brain	LID47032-LID47033	cell	2x101D	longPolyA	Illumina_HiSeq_2000
human	illumina	Lung	Lung	hbmlungPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Liver	Liver	LID21042-LID21043	cell	2x76D	longPolyA	Illumina_GA2x

mouse	CSHL	Spleen	Blood	LID21038-LID21039	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Liver	Liver	hbmliverPolyAP50	cell	2x50	polyA	NA
human	illumina	Testis	Testis	hbmtestisPolyAP50	cell	2x50	polyA	NA
mouse	CSHL	Adrenal	Adrenal	LID20728-LID20729	cell	2x76D	longPolyA	Illumina_GA2x
human	Caltech	NHLF	Lung	NhlfR2x75Il200Rep1-NhlfR2x75Il200Rep2	cell	2x75	longPolyA	.
human	CSHL	HepG2	Liver	LID16635-LID16636	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Ovary	Ovary	hbmovaryPolyAP50	cell	2x50	polyA	NA
human	CSHL	SK-N-SH_RA	Brain	LID8967-LID8968	cell	2x76D	longPolyA	Illumina_GA2x
human	illumina	Adrenal	Adrenal	hbmadrenalPolyAP50	cell	2x50	polyA	NA
human	illumina	Heart	Heart	hbmheartPolyAP50	cell	2x50	polyA	NA
human	CSHL	CD20+	Blood	LID44498-LID44499	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	MammaryGland	Breast	LID20924-LID20925	cell	2x76D	longPolyA	Illumina_GA2x
mouse	CSHL	FrontalLobe	Brain	LID47081-LID47082	cell	2x101D	longPolyA	Illumina_HiSeq_2000
mouse	CSHL	Ovary	Ovary	LID20821-LID20822	cell	2x76D	longPolyA	Illumina_GA2x
human	CSHL	IMR90	Lung	LID45016-LID45017	cell	2x101D	longPolyA	Illumina_HiSeq_2000

Supplementary Table S27: List of candidte mouse-specific promoters tested in reporter assay.

			Tested in mouse	Tested in human
chr1	72282608	72283862	Positive	Positive
chr16	94534005	94535452	Positive	Positive
chr10	81640535	81641906	Positive	Positive
chr12	113931008	113932428	Positive	Positive
chr6	117829050	117830421	Positive	Positive
chr13	23487516	23488786	Positive	Positive
chr19	61275284	61276321	Positive	Positive
chr13	21366741	21367942	Positive	Positive
chr14	51702299	51703553	Positive	Positive
chr4	88513577	88514670	Positive	Positive
chr2	168805839	168807165	Negative	N/A
chr4	154028643	154030048	Positive	Positive
chr7	50508728	50509835	Positive	Positive
chr5	110514708	110515903	Positive	Positive
chr8	91576369	91577522	Positive	Positive
chr4	133512988	133514350	Negative	N/A
chr9	113954352	113955827	Positive	Positive
chr8	28332972	28334480	Positive	Positive
chr18	13107342	13108338	Positive	Positive
chr10	126246319	126247564	Positive	Positive

Supplementary Table 29: List of candidte mouse-specific enhancers tested in reporter assays.

	start	end	Tested in mESC	Tested in hESC
chr1	121295046	121296025	Positive	Positive
chr11	9021995	9022995	Negative	Negative
chr1	21245800	21246800	Positive	Positive
chr2	109748850	109749850	Negative	Negative
chr3	33905123	33906100	Negative	Negative
chr3	107767000	107768000	Positive	Positive
chr5	38911677	38912647	Positive	Negative
chr5	128129750	128130750	Positive	Positive
chr6	30147450	30148450	Positive	Positive
chr6	49082350	49083350	Positive	Positive
chr6	54622600	54623600	Positive	Negative
chr7	50802475	50803475	Positive	Positive
chr8	72536950	72537950	Positive	Negative
chr9	99053600	99054600	Positive	Positive
chr10	42366000	42367000	Negative	Negative
chr12	77310200	77311200	Positive	Negative
chrX	73843900	73844900	Negative	Negative
chr16	48303488	48304467	Positive	Negative
chr17	41036531	41037557	Negative	Negative
chr8	97710983	97711993	Negative	Negative
chr2	147118424	147119401	Negative	Positive
chr5	53301450	53302370	Positive	Negative
chr18	25547503	25548453	Positive	Positive
chr1	3472699	3473648	Negative	Negative
chr2	31995916	31996896	Positive	Positive
chr13	86237126	86238080	Positive	Negative
chr3	107798742	107799702	Positive	Positive
chr8	58730595	58731581	Positive	Negative
chr13	16921782	16922764	Positive	Negative
chr16	56901932	56902893	Positive	Negative
chr1	17921912	17922867	Positive	Positive
chr6	68802399	68803383	Positive	Positive
chr2	36270509	36271461	Negative	Negative
chr9	9268943	9269894	Positive	Positive
chr1	142084819	142085770	Negative	Negative
chr12	4358406	4359368	Positive	Positive
chr1	84886510	84887461	Positive	Positive