



Published in final edited form as:

Nat Genet. 2013 July ; 45(7): 836–841. doi:10.1038/ng.2649.

DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape

Mingchao Xie^{1,*}, Chibo Hong^{2,*}, Bo Zhang^{1,*}, Rebecca Lowdon¹, Xiaoyun Xing¹, Daofeng Li¹, Xin Zhou¹, Hyung Joo Lee¹, Cecile L. Maire³, Keith L. Ligon^{3,4}, Philippe Gascard⁵, Mahvash Sigaroudinia⁵, Thea D. Tlsty⁵, Theresa Kadlecsek⁶, Arthur Weiss^{6,7}, Henriette O'Geen⁸, Peggy J. Farnham⁹, Pamela A.F. Madden¹⁰, Andrew J. Mungall¹¹, Angela Tam¹¹, Baljit Kamoh¹¹, Stephanie Cho¹¹, Richard Moore¹¹, Martin Hirst^{11,12}, Marco A. Marra¹¹, Joseph F. Costello², and Ting Wang¹

¹Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108

²Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA 94143.

³Department of Medical Oncology, Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215.

⁴Departments of Pathology, Harvard Medical School, Boston, MA 02115

⁵Department of Pathology, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA 94143.

⁶Howard Hughes Medical Institute, University of California San Francisco, CA 94143

⁷Division of Rheumatology, University of California San Francisco, CA 94143.

⁸Genome Center, University of California Davis, Davis, CA 95616

⁹Department of Biochemistry & Molecular Biology, Norris Comprehensive Cancer Center, University of Southern California, CA 90089

¹⁰Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110

¹¹BC Cancer Agency, Canada's Michael Smith Genome Sciences Centre Vancouver, British Columbia, Canada.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to J.F.C. (jcostello@cc.ucsf.edu) or T.W. (twang@genetics.wustl.edu).

*These authors contributed equally.

Author contributions J.F.C and T.W. designed the study. C.L.M, K.L.L., P.G., M.S., T.D.T., T.K, and A.W. collected samples. C.H., H.O., P.J.F., A.J.M., A.T., B.K., S.C., R.M., M.H., and M.A.M. performed sequencing assays. M.X., B.Z., R.L., D.L., X.Z., H.J.L., P.A.F.M, and T.W. performed data analysis. C.H., X.X., and M.X. performed bisulfite validation and reporter gene assays. M.X., J.F.C. and T.W. wrote the manuscript. All authors discussed the results and contributed to writing the manuscript.

Competing financial interests The authors declare no competing financial interests.

Accession codes Complete datasets used in this study:

¹²Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

Introduction

Transposable element (TE) derived sequences comprise half of our genome and DNA methylome, and are presumed densely methylated and inactive. Examination of the genome-wide DNA methylation status within 928 TE subfamilies in human embryonic and adult tissues revealed unexpected tissue-specific and subfamily-specific hypomethylation signatures. Genes proximal to tissue-specific hypomethylated TE sequences were enriched for functions important for the tissue type and their expression correlated strongly with hypomethylation of the TEs. When hypomethylated, these TE sequences gained tissue-specific enhancer marks including H3K4me1 and occupancy by *p300*, and a majority exhibited enhancer activity in reporter gene assays. Many such TEs also harbored binding sites for transcription factors that are important for tissue-specific functions and exhibited evidence for evolutionary selection. These data suggest that sequences derived from TEs may be responsible for wiring tissue type-specific regulatory networks, and have acquired tissue-specific epigenetic regulation.

A large portion of eukaryotic genomes is derived from transposable elements (TEs)¹. TEs have been described as parasitic or junk DNA. However, there is mounting evidence for their evolutionary contribution to the wiring of gene regulatory networks²⁻⁷, a theory rooted in Barbara McClintock's discovery that TEs can control gene expression^{3,8,9}. TEs contain functional binding sites for transcription factors^{6,10,11}; TE DNAs are presumed to be methylated in somatic cells to suppress transposition and TE-mediated changes in gene expression¹²⁻¹⁴. However, the extent to which DNA methylation silences TEs and how DNA methylation-mediated silencing of TEs is reconciled with the known regulatory function of TE sequences remain unexplored.

To construct TE DNA methylation profiles we assayed 29 human samples representing 11 cell types using two complementary DNA methylomics methods: MeDIP-seq and MRE-seq^{15,16}. Tissue and cell types included embryonic stem cells (ESC H1); fetal brain tissue and primary neural progenitor cells (derived from cortex or ganglionic eminence regions); primary adult breast epithelial cells (luminal epithelial cells, myoepithelial cells, and a progenitor cell-enriched population); unfractionated peripheral blood mononuclear cells (PBMC), and adult immune cells including CD4+ naïve, CD4+ memory, and CD8+ naïve cells.

Mapping short-read data to TEs is difficult due to the high copy number of these elements. Standard mapping often discard or mis-align high quality reads derived from TEs (Supplementary Note). We developed a computational strategy termed Repeat Analysis Pipeline (RAP) that allows mapping of reads derived from repetitive elements to one of 1,395 specific families of human repeats including 928 TE families (Supplementary Fig. 1-5, Note). RAP includes features of three previously published methods¹⁷⁻²⁰ combined with novel technical modifications (**Methods**).

As expected, sequences of the majority of TE families were methylated in all samples examined. The total MeDIP-seq signal, which represents the proportion of individual TE families that are methylated, correlated tightly with the total number of CpGs in that TE family, consistent with the high level of DNA methylation in TEs ($R^2=0.95$, Supplementary Fig. 6-9). In contrast to TE families, total MeDIP-seq signal was 4.9% in promoter CpG islands after normalizing for CpG content, consistent with the unmethylated status of promoter CpG islands. Conversely, MRE-seq signal, which measures unmethylated DNA, was 6.7-fold more enriched over promoter CpG islands than in TEs (Supplementary Fig. 6-9).

Strikingly, we found sequences of numerous TE families that were differentially methylated in specific cell-types. Unsupervised clustering of samples based on TE methylation revealed a clear relationship among tissue-types, indicating that TE methylation is a signature that can distinguish tissue- or possibly cell-types (Fig. 1a, b). We identified 14 TE families with significant ($p<0.05$, ANOVA) hypomethylation patterns in brain samples, 55 in breast samples, 13 in blood samples, and 13 in ESC (total 95 TE families, $p<0.05$, ANOVA). More than 800 other families were consistently methylated across cell types from these 29 samples (Supplementary Note). Most tissue-specific hypomethylated TEs belonged to the ERV/LTR class (69/95), whereas 12 were DNA transposon families (Supplementary Table 1). These findings are consistent with previous studies that have shown that LTR-elements participate in regulation of mammalian genes^{3,21-24}, and support the hypothesis that LTRs might play a role in the epigenetic regulation of cell-type specific gene expression. For each TE family, we identified individual copies that were uniquely mappable and were tissue-specifically hypomethylated. The complete list of TE families and coordinates of individual elements are provided at our website (Supplementary Note).

We next investigated the genomic distribution of members of TE families showing tissue-specific hypomethylation. Their proximities to “known genes” were not different from being expected by chance (Supplementary Fig. 10). However, genes near members of these TE families were significantly enriched for functions specific to the tissue type in which they were hypomethylated (Table 1 and Supplementary Table 2). For example, hypomethylation of the UCON29 DNA transposon was restricted to fetal brain, and 11 of the 60 genes with a nearby UCON29 element are involved in neuron development ($p<6.6\times 10^{-23}$, binomial test). Another brain-specific hypomethylated retroelement, LFSINE, was located near 19 out of 87 genes involved in telencephalon development ($p<1.5\times 10^{-5}$, binomial test). Similarly, genes associated with LTR12 and LTR77, two ERVs hypomethylated in immune cells, were enriched for immune-related functions, including ‘antigen processing and presentation of peptide or polysaccharide antigen via MHC class II’ ($p<7.4\times 10^{-6}$, binomial test), and ‘oxidation reduction’ ($p<3.7\times 10^{-6}$, binomial test). While antigen processing and presentation is a known function of lymphocytes and other antigen-presenting hematopoietic cells, the enrichment of genes in the oxidation-reduction process was interesting because T-cell activation, differentiation and proliferation are sensitive to the redox potential^{25,26}.

DNA hypomethylation has been associated with distal regulatory regions²⁷. We next asked if TE sequences with tissue-specific DNA hypomethylation possessed other tissue-specific epigenetic signatures. We generated histone modification data (H3K4me1, H3K4me3,

H3K27me3, H3K36me3 and H3K9me3) from these same tissues, and collected *p300* genome-wide locations from related tissues²⁸ (Fig. 2). Sequences within hypomethylated TE families displayed remarkably strong tissue-specific H3K4me1 signals. For example, LTR77, a TE of the ERV class, had the lowest methylated (MeDIP-seq) signal and the highest unmethylated (MRE-seq) signal in blood (Fig. 2a). When we applied RAP to H3K4me3 and H3K4me1 ChIP-seq data from the same samples, we found much stronger signals within the LTR77 family in T cells compared to the three other cell and tissue types (Supplementary Fig. 11). Using data from CD8+ naïve cells, we identified a “histone signature” for all 148 LTR77 copies along with a 3kb region flanking the LTR (Fig. 2b,c). We observed a strong H3K4me1 peak over the LTR element itself, suggesting that at least some LTR77 elements had this enhancer mark. The H3K4me3 peak detected 3kb downstream suggested nearby promoter activities, potentially from genes regulated by enhancers embedded in LTR77. LFSINE and UCON29 displayed H3K4me1 enrichment specifically in fetal brain (Fig. 2f,g, and Supplementary Fig. 12). Moreover, LFSINE and UCON29 both accumulate *p300* binding signals in the neuroblastoma cell-line SK-N-SH, but not in any non-neural cell lines including ESC, HepG2, or GM12878 (Fig. 2h, Supplementary Fig. 12). Similarly, the T cell-specific hypomethylated TE LTR77 accumulated *p300* binding signal in GM12878 (a lymphoblastoid cell-line), but not in any other cell type (Fig. 2d). These results suggested that hypomethylated DNA sequences derived from TEs might serve as tissue-specific enhancers.

We next asked if any of these hypomethylated, enhancer-like sequences within TE might contribute to tissue-specific gene expression. We selected candidate TEs that could be uniquely mapped using our data. As a proof of principle, we focused on two putative target genes: *ERAPI*, a gene in the generation of most HLA class I-binding peptides, and the glial cell line-derived neurotrophic factor (*GDNF*) family receptor alpha-1 *GFRA1*, a neurotrophic factor involved in the control of neuron survival and differentiation²⁹ (Fig. 3a,d). A LTR77 element was detected 2kb upstream of an *ERAPI* alternative transcription start site. Our genome-wide data suggested that this element was hypomethylated in T-cells, a prediction confirmed by locus-specific bisulfite-sequencing (Fig. 3b). In addition to enhancer-like signature, *NF-kB* and *Pol2* ChIP-seq peaks were observed in a lymphoblastoid cell-line (GM12878), but not in a non-lymphoblastoid cell-line (HepG2). Consistently, *ERAPI* exhibited the highest expression in T-cells (Fig. 3c). This LTR77 element exhibited modest enhancer activity in 293T, SK-N-SH, and GM12878 cells based on reporter assay (Supplementary Fig. 13, LTR77-1). In the brain samples, *GFRA1* appeared as a putative target of an LFSINE element (Fig. 3d). We observed tissue-specific H3K4me1 marks and a H3K4me3 mark in the promoter region in fetal brain, but not in T-cells (Fig. 3d). Transcription factor binding motifs, such as that for *SOX10*, a regulator of neural crest and glial cell development^{30,31}, were identified in the hypomethylated LFSINE element upstream of *GFRA1*. Consistent with the hypothesis that LFSINE is a tissue-specific enhancer, *GFRA1* was highly and specifically expressed in neuronal cells (Fig. 3f). This element exhibited enhancer activity in 293T and SK-N-SH cells but not in GM12878 (Supplementary Fig. 13, LFSINE-1). Hypomethylation of these TEs did not appear to be a result of increased expression of nearby genes, since the hypomethylation was not observed for other TE families in the same genomic neighborhood (Fig 3a, d). Additional members of

the LTR77, LTR12, UCON29 and LFSINE subfamilies were validated and shown to exhibit tissue-specific hypomethylation and associate with nearby tissue-specific gene expression (Supplementary Fig. 14, 15). Of the 36 TE derived candidates for which we performed reporter gene assay, 26 showed enhancer activities ranging from 5- to 1000-fold increase in at least one of the three cell-lines tested (Supplementary Fig. 13). These hypomethylated TE sequences have not been previously annotated as functional elements, but our results suggest that they may influence tissue-specific gene expression.

We next examined the relationship between sequences of TEs, their epigenetic status, and transcription factor binding. We analyzed histone modification and binding data of transcription factors of two cell-lines (GM12878 and SK-N-SH) published by ENCODE^{32,33}. We focused on individual copies of two TE families that exhibited tissue-specific hypomethylation in either blood (LTR77) or fetal brain (LFSINE). Consistent with our previous findings, members of these two TE families enriched for enhancer marks in a cell type-specific manner (Fig. 4) – LTR77 exhibited H3K4me1 mark and *p300* binding in GM12878, but not in SK-N-SH; LFSINE exhibited *p300* binding in SK-N-SH, but they did not enrich for H3K4me1 or *p300* signal in GM12878. Binding sites of several transcription factors were enriched in LTR77 and LFSINE and showed cell type specificity (Fig. 4). For example, *NF- κ B* binding overlapped specifically with LTR77 in GM12878; *Rad21* bound within LFSINE more than within LTR77; and *Rad21* bound within LFSINE more in SK-N-SH than in GM12878 (Fig. 4). Not surprisingly, many TEs were predicted to contain a sequence motif when scanned using position specific weight matrices of transcription factors (Fig. 4). Having a motif was neither necessary nor sufficient for the actual binding, which correlated strongly with cell type-specific enhancer mark. Taken together, ENCODE data confirmed that sequences of specific TE families exhibited cell type-specific enhancer signatures and cell type-specific transcription factor binding. Whether there is a causal relationship between the TEs' epigenetic mark and transcription factor binding awaits further investigation.

For decades, TEs have been deemed as parasitic DNA as a result of the impact of their transposition in the genome^{34,35}. Transposition of TEs may be deleterious when they disrupt coding sequences or normal gene expression, resulting in human diseases³⁶⁻³⁸. Thus, it is believed that cells have acquired epigenetic mechanisms to cope with TEs so that transposon-derived sequences are completely methylated and transcriptionally silent in somatic tissues^{14,39}.

However, TE transpositions might provide diverse genetic material for natural selection, which would contribute to the evolution of species-specific traits and population biodiversity^{40,41}. Many functional elements were born by “exaptation”, a process in which DNAs of a transposon are co-opted to benefit the host⁴²⁻⁴⁴. TE insertions with regulatory functions have been described in mammals^{4,5,7,45}. A substantial proportion of constrained non-coding sequences arose from TEs^{46,47}, pointing to transposons as a driving force in the evolution of regulation network. Some hypomethylated TE subfamilies identified here were conserved based on their PhastCons and PhyloP scores, suggesting that this conservation might be a consequence of selection (Supplementary Fig. 16, 17). While we do not know how many TEs could have regulatory functions, previous reports indicate that 5% of TEs are

under evolutionary constraint^{46,47}. TE sequences were incorporated in gene networks under the control of transcription factors including *TP53*⁶, *OCT4*^{4,7}, *CTCF*⁴⁸, and *MER20* was reported to have contributed to the origin of pregnancy in placental mammals⁵. TE-derived sequences can directly regulate expression. For example, *ISL1* is regulated by a SINE element⁴⁹, and so is *FGF8* in the forebrain⁵⁰. In both cases, TEs provide distal enhancers that help control expression of host genes, and their hypomethylation status in brain cells was confirmed by our genome-wide data (Supplementary Fig. 14).

Our findings help to resolve the conflicting observations that TE sequences are globally suppressed by epigenetic mechanisms, including DNA methylation, but that they can mediate gene regulation in some instances. In this study, we challenge the general notion that TEs are constitutively methylated by examining the extent to which TE methylation differs between cell-types and the relationship between epigenetic silencing and TE sequences' potential to impact gene regulation. Epigenetic control of TEs may contribute to developmental stage-specific, cell type-specific, and perhaps health condition-specific gene regulation. Distal regulatory regions are methylated at low levels, display enhancer chromatin marks, and are occupied by cell type-specific transcription factors²⁷. Our results suggest that some TE sequences match this profile of distal enhancers. With a few exceptions^{51,52}, majority of human TEs were fixed and no longer active. Sequences within these TEs, however, could be adapted to serve as enhancers, and these sequences might be the reason for their epigenetic regulation. The mechanisms through which DNA within TEs is demethylated and obtains enhancer chromatin marks, and the relationship between TE-derived enhancers and other regulatory elements remain to be elucidated. A recent report demonstrated transposons on a human chromosome acquired activating histone modifications and changed DNA methylation status in mouse cells⁵³. In rodents, some endogenous retroviruses function as species-specific enhancers in the placenta⁵⁴. Therefore, as a source of new regulatory elements, TEs' regulatory potential could be controlled by tissue- or cell type-specific epigenetic regulation. In our study, examination of DNA methylation in four distinct tissue types showed that while sequences of many TE families are globally hypermethylated, about 10% of TE families are hypomethylated in a tissue-specific manner and gain distal enhancer signatures. Analysis of a more extensive panel of tissues may reveal that a much larger portion of sequences derived from TEs may harbor gene regulatory function.

Online Methods

Further details for computational analyses are provided in the Supplementary Note.

1. Sample preparation

Blood—Buffy coats were obtained from the Stanford Blood Center (Palo Alto, CA). Blood was drawn and processed on the same day. Peripheral Blood Mononuclear cells (PBMC) were isolated by Histopaque 1077 (Sigma-Aldrich, Saint-Louis, MO) density gradient centrifugation according to the manufacturer's protocol. Further purification of CD4 memory, CD4 naïve, and CD8 naïve T lymphocytes was performed using a Robosep instrument and isolation kits for each subpopulation as listed below (STEMCELL

Technologies, Vancouver, BC, Canada). Total PBMC were karyotyped (Molecular Diagnostic Services Inc. San Diego, CA) and analyzed for cell cycle. PBMC and T cell subpopulations were stained with antibodies and analyzed by FACS for purity. Cells were aliquoted for DNA and RNA samples, and were washed in PBS. Cell pellets for RNA samples were resuspended in 1 ml TRIzol reagent (Invitrogen, Carlsbad, CA), and frozen at -80°C . Cell pellets for DNA samples were flash frozen in liquid nitrogen and stored at -80°C . Reagents and Antibodies:

- Anti-CD3 TRI-COLOR, Invitrogen
- Anti-CD4 PE, BD Biosciences
- Anti-CD8 FITC, BD Biosciences
- Anti-CD4 TRI-COLOR, Invitrogen
- Anti-CD45RO PE, Invitrogen
- Anti-CD45RA FITC, BD Biosciences
- Anti-CD8 TRI-COLOR, Invitrogen
- EasySep[®] Human Memory CD4 T Cell Enrichment Kit,
- EasySep[®] Human Naive CD4+ T Cell Enrichment Kit,
- Custom Human Naïve CD8 T cell Enrichment Kit, STEMCELL Technologies

Breast—Breast tissues were obtained from disease-free pre-menopausal women undergoing reduction mammoplasty in accordance with institutionally approved IRB protocol # 10-01563 (previously CHR # 8759-34462-01). All tissues were obtained as de-identified samples and linked only with minimal dataset (age, ethnicity and in some cases parity/gravidity). Tissue was dissociated mechanically and enzymatically, as previously described⁵⁶. Briefly, tissue was minced and dissociated in RPMI 1640 with L-glutamine and 25mm HEPES (Fisher, cat # MT10041CV) supplemented with 10% fetal bovine serum (JR Scientific, Inc, cat # 43603), 100 units/ml penicillin, 100 $\mu\text{g}/\text{ml}$ streptomycin sulfate, 0.25 $\mu\text{g}/\text{ml}$ fungizone, gentamycin (Lonza, Cat # CC4081G), 200U/ml collagenase 2 (Worthington, cat # CLS-2) and 100U/ml hyaluronidase (Sigma-Aldrich, cat # H3506-SG) at 37°C for 16h. The cell suspension was centrifuged at 1,400rpm for 10min followed by a wash with RPMI 1640/10% FBS. Clusters enriched in epithelial cells (referred to as organoids) were recovered after serial filtration through a 150- μm nylon mesh (Fisher, cat # NC9445658), and a 40- μm nylon mesh (Fisher, cat # NC9860187). The final filtrate contained primarily mammary stromal cells (fibroblasts, immune cells and endothelial cells) and some single epithelial cells. Following centrifugation at 1,200rpm for 5min, the epithelial organoids and filtrate were frozen for long-term storage. The day of cell sorting, epithelial organoids were thawed out and further digested with 0.5g/L 0.05% trypsin-EDTA and dispase-DNAse I (STEMCELL Technologies, cats # 7913 and # 7900, respectively). Generation of single cell suspensions was monitored visually. Single cell suspensions were filtered through a 40- μm cell strainer (Fisher, cat # 087711), spun down and allowed to “regenerate” in MEGM medium (Lonza) supplemented with 2% fetal calf serum for 60-90min at 37°C . This “regeneration” step enables quenching of trypsin and re-expression

of the cell surface markers prior to staining as their extra cellular domain had been cleaved by trypsin.

The single cell suspension obtained as described above was stained for cell sorting with three human-specific primary antibodies, anti-CD10 labeled with PE-Cy7 (BD Biosciences, cat # 341092) to isolate myoepithelial cells, anti-CD227/MUC1 labeled with FITC (BD Biosciences cat # 559774) to isolate luminal epithelial cells or anti-CD73 labeled with PE (BD Biosciences, cat # 550257) to isolate a stem cell-enriched cell population, and with biotinylated antibodies for lineage markers, anti-CD2, CD3, CD16, CD64 (BD Biosciences, cat # 555325, 555338, 555405 and 555526), CD31 (Invitrogen, cat # MHCD3115), CD45, CD140b (BioLegend, cat #s 304003 and 323604) to specifically remove hematopoietic, endothelial and leukocyte lineage cells, respectively, by negative selection. Sequential incubation with primary antibodies was performed for 20min at room temperature in PBS with 1% bovine serum albumin (BSA), followed by washing in PBS with 1% BSA. Biotinylated primary antibodies were revealed with an anti-human secondary antibody labeled with streptavidin-Pacific Blue conjugate (Invitrogen, cat # S11222). After incubation, cells were washed once in PBS with 1% BSA and cell sorting was performed using a FACSAria II cell sorter (BD Biosciences).

Fetal Brain—Post-mortem human fetal neural tissues were obtained from a case of twin non-syndrome fetuses whose death was attributed to environmental/placental etiology. Tissues were obtained with appropriate patient consent according to Partner's Healthcare/Brigham and Women's Hospital IRB guidelines (Protocol #2010P001144). All samples and tissues were de-identified and linked only with minimal dataset (age, gender, brain location). Fetal brain tissue and fetal neural progenitor cells were derived from manually dissected regions of the brain (telencephalon), specifically the neocortex (pallium; GSM666914, GSM669615, GSM669610, GSM669612) and ganglionic eminences (subpallium; GSM669611, GSM669613). The tissues were minced and dissociated by combination of mechanical agitation (gentleMACS device) during enzymatic treatment with papain according to manufacturer's protocol (Miltenyi Biotec, Neural tissue dissociation kit #130-092-628). Cell suspensions were then washed twice in DMEM and plated at low density in human NeuroCult NS-A media (Stem cell technology # 05751) supplemented with heparin, EGF (20ng/ml) and FGF (10ng/ml) in ultra low attachment cell culture flasks (Corning #3814).

ESC H1—Data were obtained from a previous publication¹⁵.

2. High-throughput sequencing assays

All assays were performed as part of the NIH Roadmap Epigenomics Mapping Centers' repository for human reference epigenome atlas⁵⁷. Experiments were performed under the guidelines of Roadmap Epigenomics project (<http://www.roadmapepigenomics.org/protocols>). Specifically, MeDIP-seq and MRE-seq were performed as previously described¹⁶. ChIP-seq was performed as described in⁵⁸. All data have been submitted to NCBI (Supplementary Table 3).

3. Bisulfite validation

Total genomic DNA underwent bisulfite conversion following an established protocol⁵⁹ with modification of: 95 °C for 1 min, 50 °C for 59 min for a total of 16 cycles. Regions of interest were amplified with PCR primers (see below) and were subsequently cloned using pCR2.1/TOPO (Invitrogen). Individual bacterial colonies were subjected to PCR using vector-specific primers and sequenced using an ABI 3700 automated DNA sequencer. The data were analyzed with online software BISMA⁶⁰. Result is summarized in Supplementary Fig. 13. Genomic locations of candidates and primer information are summarized in Supplementary Table 4.

4. Reporter gene assay

TE candidates were amplified from genomic DNA using Pfu-polymerase (Agilent) and primers containing KpnI- or BglIII- restriction sites. PCR products were gel-purified using Qiagen Gel purification kit, and then digested by the corresponding restriction enzymes (NEB). The digested PCR products were cloned into the pGL4.23[luc2/minP]-vector (Promega, E8411) using T4-ligase(NEB) and transformed into chemical competent DH5 α -cells. The positive clones were verified by enzyme digestion and sequencing. 800 ng of reporter plasmid (or empty pGL4.23[luc2/minP]-vector control) were transfected into 3 different cell lines, 293T, GM12878, and SK-N-SH_RA which were differentiated with 6 μ M of retinoic acid for 48 hours from SK-N-SH cells, using X-tremeGENE (Roche) in triplicate. In order to normalize the transfection, 200 ng of renilla luciferase plasmid driven by a TK promoter were co-transfected. The luciferase activity was measured after 48 hours, and normalized by the relative renilla control. Genomic locations of candidates and primer information are summarized in Supplementary Table 5.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the many collaborators in Reference Epigenome Mapping Centers (REMCs), Epigenome Data Analysis and Coordination Center and NCBI who have generated and processed data which were used in this project. We acknowledge the dedicated system administrators at Washington University Center for Genome Sciences and Systems Biology who have provided an excellent computing environment. We thank UCSC Genome Browser bioinformatics team for providing processed ENCODE data. We acknowledge support from NIH Roadmap Epigenomics Program, sponsored by the National Institute on Drug Abuse (NIDA) and the National Institute of Environmental Health Sciences (NIEHS). J.F.C., T.W., P.F. and M.H. are supported by NIH grant 5U01ES017154. B.Z and X.Z. are supported by NIDA's R25 program DA027995. K.L.L. and C.M. are supported by NIH grant P01CA095616 and P01CA142536. T.W. is supported in part by the March of Dimes Foundation, the Edward Jr. Mallinckrodt Foundation, P50CA134254 and a generous start up package from Department of Genetics, Washington University School of Medicine.

References

1. Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]
2. Bourque G, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res*. 2008; 18:1752–62. Epub 2008 Aug 5. [PubMed: 18682548]

3. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 2008; 9:397–405. [PubMed: 18368054]
4. Kunarso G, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics.* 2010; 42:631–4. Epub 2010 Jun 6. [PubMed: 20526341]
5. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet.* 2011; 43:1154–9. [PubMed: 21946353]
6. Wang T, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A.* 2007; 104:18613–8. Epub 2007 Nov 14. [PubMed: 18003932]
7. Xie D, et al. Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Research.* 2010
8. McClintock B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol.* 1956; 21:197–216. [PubMed: 13433592]
9. Mc CB. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A.* 1950; 36:344–55. [PubMed: 15430309]
10. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 2003; 19:68–72. [PubMed: 12547512]
11. Polavarapu N, Marino-Ramirez L, Landsman D, McDonald JF, Jordan IK. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics.* 2008; 9:226. [PubMed: 18485226]
12. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet.* 1999; 23:314–8. [PubMed: 10545949]
13. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007; 8:272–85. [PubMed: 17363976]
14. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002; 16:6–21. [PubMed: 11782440]
15. Harris RA, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol.* 2010; 28:1097–105. [PubMed: 20852635]
16. Maunakea AK, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010; 466:253–7. [PubMed: 20613842]
17. Day DS, Luquette LJ, Park PJ, Kharchenko PV. Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* 2011; 11:R69. [PubMed: 20584328]
18. Chung D, et al. Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS Comput Biol.* 2011; 7:e1002111. [PubMed: 21779159]
19. Wang J, Huda A, Lunyak VV, Jordan IK. A Gibbs sampling strategy applied to the mapping of ambiguous short-sequence tags. *Bioinformatics.* 2010; 26:2501–8. [PubMed: 20871106]
20. Schmid CD, Bucher P. MER41 repeat sequences contain inducible STAT1 binding sites. *PLoS One.* 2010; 5:e11425. [PubMed: 20625510]
21. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol.* 1990; 10:2513–20. [PubMed: 1692956]
22. Medstrand P, Landry JR, Mager DL. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J Biol Chem.* 2001; 276:1896–903. [PubMed: 11054415]
23. Dunn CA, Medstrand P, Mager DL. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. *Proc Natl Acad Sci U S A.* 2003; 100:12841–6. [PubMed: 14534330]
24. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene.* 2009; 448:105–14. [PubMed: 19577618]

25. Yan Z, Banerjee R. Redox remodeling as an immunoregulatory strategy. *Biochemistry*. 2010; 49:1059–66. [PubMed: 20070126]
26. Angelini G, et al. Antigen-presenting dendritic cells provide the reducing extracellular microenvironment required for T lymphocyte activation. *Proc Natl Acad Sci U S A*. 2002; 99:1491–6. [PubMed: 11792859]
27. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2012; 480:490–5. [PubMed: 22170606]
28. Birney E, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. [PubMed: 17571346]
29. Roussa E, von Bohlen und Halbach O, Kriegelstein K. TGF-beta in dopamine neuron development, maintenance and neuroprotection. *Adv Exp Med Biol*. 2009; 651:81–90. [PubMed: 19731553]
30. Britsch S, et al. The transcription factor Sox10 is a key regulator of peripheral glial development. *Genes Dev*. 2001; 15:66–78. [PubMed: 11156606]
31. Wegner M, Stolt CC. From stem cells to neurons and glia: a Soxist's view of neural development. *Trends Neurosci*. 2005; 28:583–8. [PubMed: 16139372]
32. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
33. Rosenbloom KR, et al. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res*. 2012; 40:D912–7. [PubMed: 22075998]
34. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*. 1980; 284:601–3. [PubMed: 6245369]
35. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature*. 1980; 284:604–7. [PubMed: 7366731]
36. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet*. 2001; 35:501–38. [PubMed: 11700292]
37. Martinez-Garay I, et al. Intronic L1 insertion and F268S, novel mutations in RPS6KA3 (RSK2) causing Coffin-Lowry syndrome. *Clin Genet*. 2003; 64:491–6. [PubMed: 14986828]
38. Claverie-Martin F, Gonzalez-Acosta H, Flores C, Anton-Gamero M, Garcia-Nieto V. De novo insertion of an Alu sequence in the coding region of the CLCN5 gene results in Dent's disease. *Hum Genet*. 2003; 113:480–5. [PubMed: 14569459]
39. Fazzari MJ, Grealley JM. Epigenomics: beyond CpG islands. *Nat Rev Genet*. 2004; 5:446–55. [PubMed: 15153997]
40. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci U S A*. 1997; 94:7704–11. [PubMed: 9223252]
41. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*. 2002; 3:370–9. [PubMed: 11988762]
42. Brosius J. Retroposons--seeds of evolution. *Science*. 1991; 251:753. [PubMed: 1990437]
43. Britten RJ. Cases of ancient mobile element DNA insertions that now affect gene regulation. *Mol Phylogenet Evol*. 1996; 5:13–7. [PubMed: 8673282]
44. Miller WJ, McDonald JF, Nouaud D, Anxolabehere D. Molecular domestication--more than a sporadic episode in evolution. *Genetica*. 1999; 107:197–207. [PubMed: 10952213]
45. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet*. 2003; 19:530–6. [PubMed: 14550626]
46. Lowe CB, Bejerano G, Haussler D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A*. 2007 in press.
47. Lindblad-Toh K, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011; 478:476–82. [PubMed: 21993624]
48. Schmidt D, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*. 2012; 148:335–48. [PubMed: 22244452]
49. Bejerano G, et al. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*. 2006; 441:87–90. [PubMed: 16625209]

50. Sasaki T, et al. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci U S A*. 2008; 105:4220–5. [PubMed: 18334644]
51. Beck CR, et al. LINE-1 retrotransposition activity in human genomes. *Cell*. 2010; 141:1159–70. [PubMed: 20602998]
52. Iskow RC, et al. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*. 2010; 141:1253–61. [PubMed: 20603005]
53. Ward MC, et al. Latent Regulatory Potential of Human-Specific Repetitive Elements. *Mol Cell*. 2012
54. Chuong EB, Rumi MA, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet*. 2013; 45:325–9. [PubMed: 23396136]
55. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28:495–501. [PubMed: 20436461]
56. Romanov SR, et al. Normal human mammary epithelial cells spontaneously escape senescence and acquire genomic changes. *Nature*. 2001; 409:633–7. [PubMed: 11214324]
57. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010; 28:1045–8. [PubMed: 20944595]
58. O’Geen H, Echipare L, Farnham PJ. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol Biol*. 2011; 791:265–86. [PubMed: 21913086]
59. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res*. 2001; 29:E65–5. [PubMed: 11433041]
60. Rohde C, Zhang Y, Reinhardt R, Jeltsch A. BISMAs--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*. 2010; 11:230. [PubMed: 20459626]

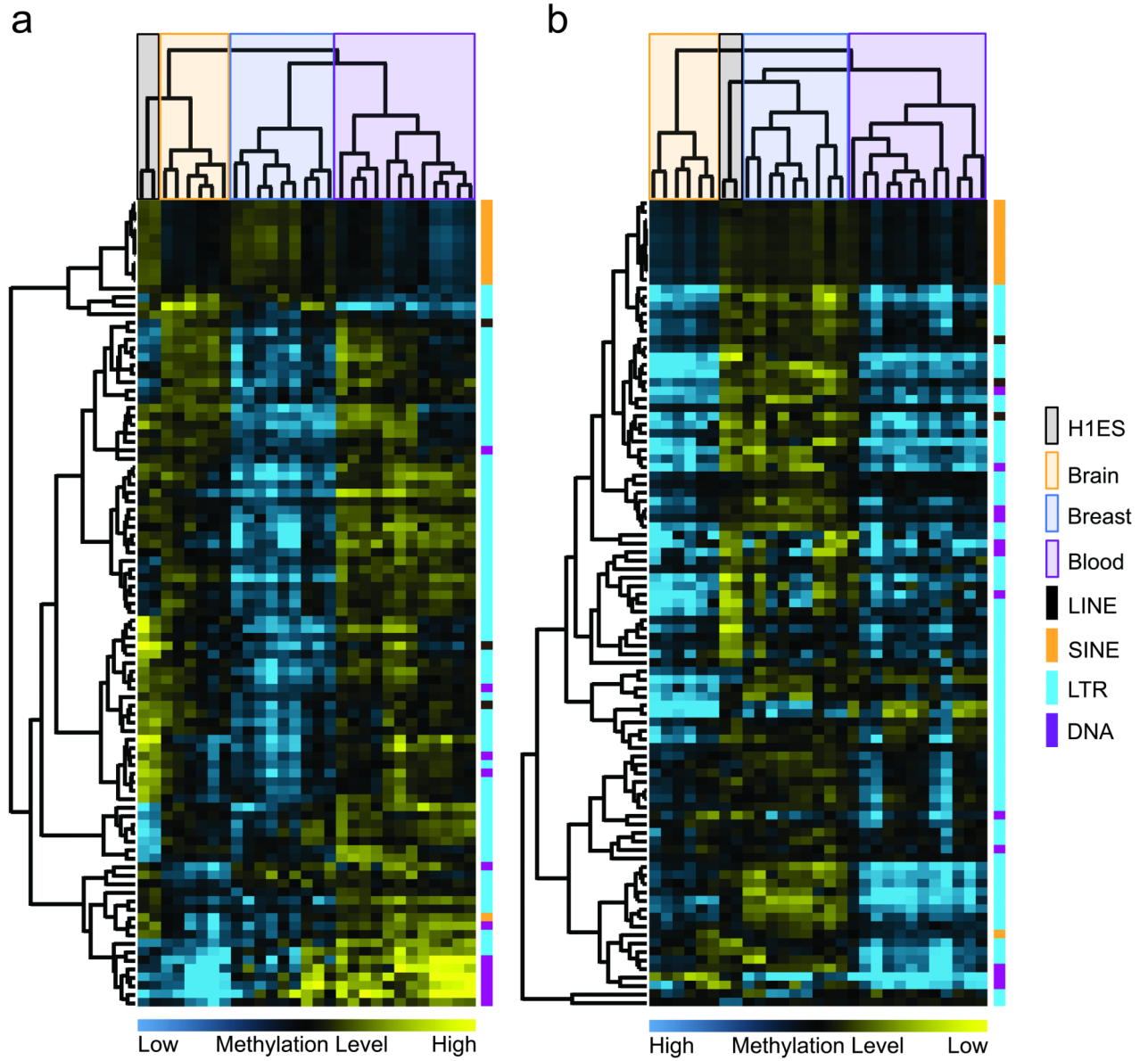


Figure 1. Clustering of TE families based on their DNA methylation profile reveals tissue specificity

TE families (rows) were clustered based on their MeDIP-seq (a) or MRE-seq (b) enrichment values across 29 samples (see Online Methods). The samples (columns) were clustered into four major groups, which were consistent with their tissue types: ESC H1 (gray), Brain (orange), Breast (blue), and Blood (purple). The vertical bar on the right side of the heatmap represents TE classes: LTR (blue), DNA transposon (purple), SINE (orange), and LINE (black). The corresponding methylation enrichment values are represented as horizontal bar with varying color gradients at the bottom of each panel.

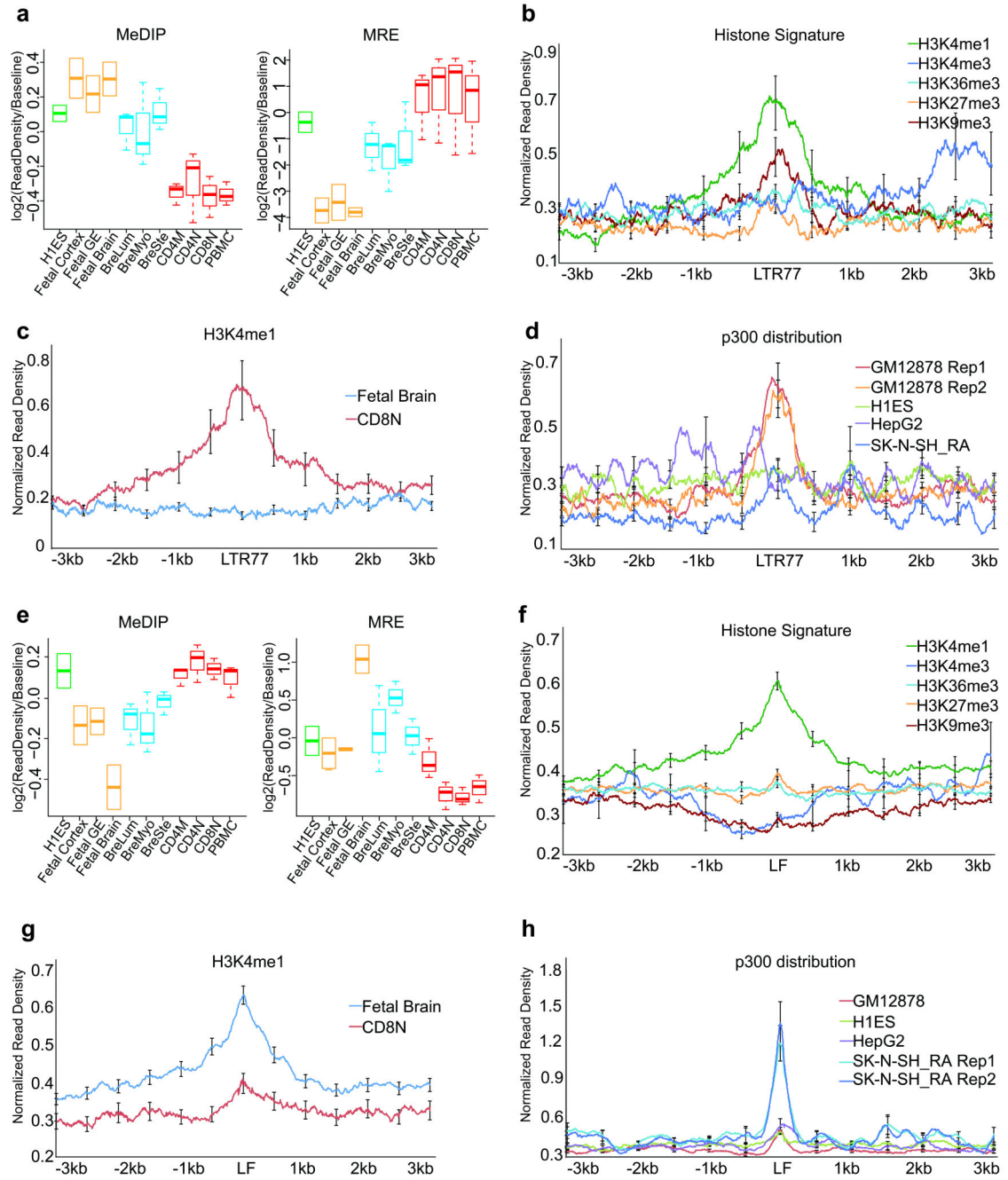


Figure 2. Tissue-specific enhancer signatures of LTR77 and LFSINE

LTR77 (a-d) and LFSINE (e-h) are specifically hypomethylated in blood samples and brain samples, respectively. (a) Boxplots of MeDIP-seq and MRE-seq enrichment scores of LTR77 in multiple cell/tissue types. (b) Histone modification signatures of LTR77 in CD8+ Naïve cells. (c) Comparison of H3K4me1 signal of LTR77 between fetal brain sample and CD8+ Naïve cells. (d) p300 binding signal on LTR77 in four cell lines. (e) Boxplots of MeDIP-seq and MRE-seq enrichment scores of LFSINE in multiple cell/tissue types. (f) Histone modification signatures of LFSINE in fetal brain sample. (g) Comparison of

H3K4me1 signal of LFSINE between fetal brain sample and CD8+ Naïve cells. **(h)** p300 binding signal on LFSINE in four cell lines. Signals of different histone modification or p300 binding for each genomic copy of the TE family including 3kb upstream and downstream flanking regions were averaged in 5bp tiling windows. Error bar represents 1 standard deviation.

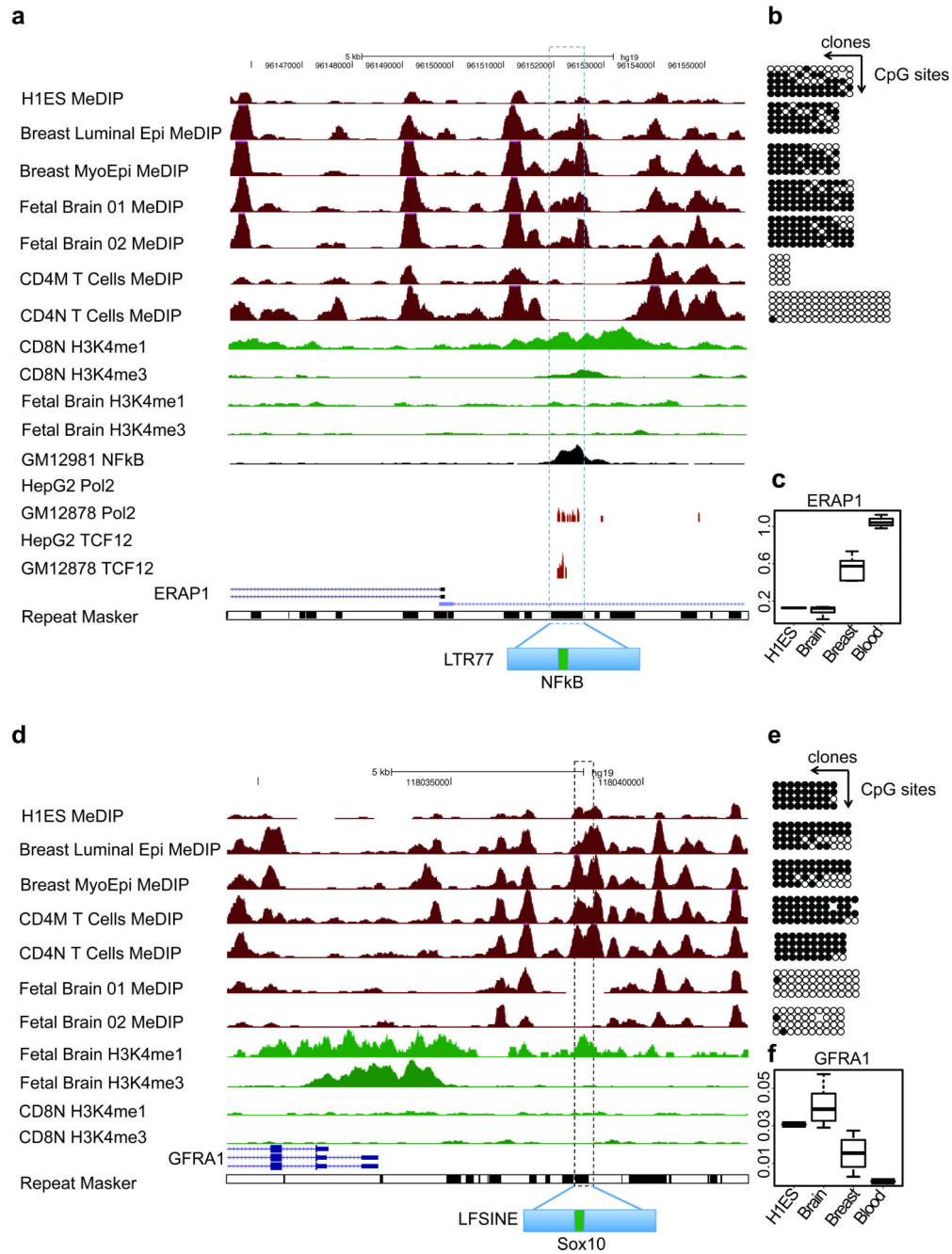


Figure 3. Tissue-specific hypomethylated TEs correlate with gene expression

(a) Genome Browser view of an LTR77 element upstream of the *ERAP1* gene. Displayed tracks include: DNA methylation (MeDIP-seq) for human ESC H1, breast, brain and blood samples; histone modification (H3K4me1 and H3K4me3) tracks for a CD8 naïve sample and a fetal brain cell sample; transcription factor binding tracks (ENCODE) for *NFkB*, *Pol2*, and *TCF12* in three cell lines; gene annotation and RepeatMasker. (b) Bisulfite sequencing validation of DNA methylation status of the LTR77 element (5 CpG sites) in human ESC H1, breast, brain and blood samples. Black circle represents methylated CpG sites and white

circle represents unmethylated CpG sites. (e) Boxplots of expression levels of *ERAPI* in 4 different tissues. (d) Genome Browser view of an LFSINE element upstream of the *GFRA1* gene. Displayed tracks include: DNA methylation (MeDIP-seq) for human ESC H1, breast, brain, and blood samples; histone modification (H3K4me3 and H3K4me1) tracks for a fetal brain sample and a CD8+ naïve cell sample; gene annotation and RepeatMasker. (e) Bisulfite sequencing validation of DNA methylation status of the LFSINE element (4 CpG sites) in human ESC H1, breast, brain, and blood samples. (f) Boxplots of expression levels of *GFRA1* in 4 different tissues.

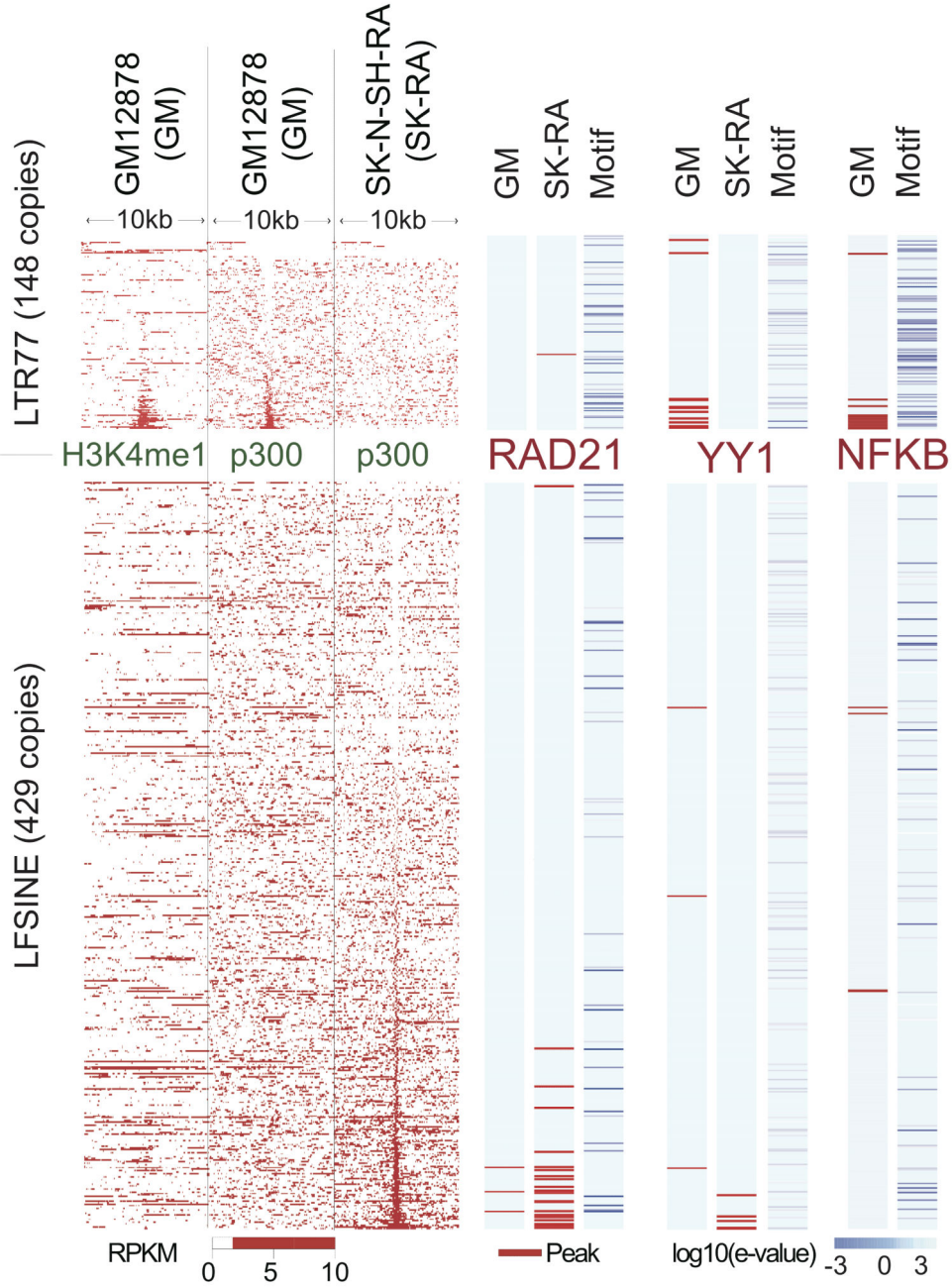


Figure 4. Correlation between cell type-specific enhancer marks, binding of transcription factors, and sequence motifs
 Histone modification, transcription factor binding, and sequence motif prediction data were displayed for individual genomic copies of LTR77 and LFSINE. Each row represents one element. Data were obtained from UCSC ENCODE portal³³. For H3K4me1 histone modification and *p300* ChIP-seq data, RPKM values at 50bp resolution were plotted for a 10kb region centered on the TE copy. For transcription factor binding data, a red tick indicates that the TE copy overlaps with a peak predicted using ChIP-seq data of the given transcription factor in the given cell type. For sequence motif data, each TE copy was scored

using position specific weight matrix of the given transcription factor. A blue tick indicates log-transformed e-value of observing a sequence motif by chance.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

GO enrichment of genes associated with hypomethylated TEs.

TE	GO Biological Process	P-Value	FDR	Gene Hits	Fold Enrichment
LFSINE	Telencephalon development	1.49E-05	2.74E-03	19/87	3.55
	Pallium development	9.35E-05	1.24E-02	12/56	3.48
	Neuron migration	1.50E-04	1.79E-02	16/69	3.77
UCON29	Generation of neurons	6.6031E-23	3.6419E-20	11/656	4.9126
	Neuron differentiation	3.3780E-22	1.4247E-19	10/500	5.8593
	Neuron recognition	5.01E-5	4.49E-2	5/23	11.04
LTR12	Oxidation reduction	3.73E-06	2.67E-02	17/647	2.24
	Antigen processing and presentation of peptide via MHC class II	7.40E-06	2.65E-02	2/20	8.53
LTR77	Homophilic cell adhesion	7.0555E-7	5.0588E-3	10/105	11.70
	Cell-cell adhesion	4.5389E-6	1.6272E-2	12/266	5.55

Genomic coordinates of individual TE copies of the TE families were used as input for GREAT analysis⁵⁵. Each gene was assigned a basal regulatory domain of 5kb upstream and 1kb downstream of the TSS (regardless of other nearby genes). The gene regulatory domain was extended in both directions to the nearest gene's basal domain but no more than a maximum of 1Mb extension in one direction. GO enrichment, p-values and FDR values were computed by GREAT.

Experiment	Sample	GEO ID
MeDIP-seq	H1Es Batch1	GSM543016
	H1Es Batch2	GSM456941
	Breast Luminal Epithelial Cells RM066	GSM613856
	Breast Luminal Epithelial Cells RM070	GSM613843
	Breast Luminal Epithelial Cells RM071	GSM613852
	Breast MyoEpithelial Cells RM066	GSM613857
	Breast MyoEpithelial Cells RM070	GSM613846
	Breast MyoEpithelial Cells RM071	GSM613850
	Breast Stem Cells RM066	GSM613859
	Breast Stem Cells RM070	GSM613847
	Breast Stem Cells RM071	GSM613853
	CD4 Memory Primary Cells TC003	GSM613862
	CD4 Memory Primary Cells TC007	GSM613914
	CD4 Memory Primary Cells TC009	GSM669608
	CD4 Naive Primary Cells TC003	GSM543025
	CD4 Naive Primary Cells TC007	GSM613913
	CD4 Naive Primary Cells TC009	GSM669607
	CD8 Naive Primary Cells TC003	GSM543027
	CD8 Naive Primary Cells TC007	GSM613917
	CD8 Naive Primary Cells TC009	GSM669609
	Fetal Brain HuFNSC01	GSM669614
	Fetal Brain HuFNSC02	GSM669615
	Neurosphere Cultured Cells, Cortex Derived HuFNSC01	GSM669610
	Neurosphere Cultured Cells, Cortex Derived HuFNSC02	GSM669612
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC01	GSM669611
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC02	GSM669613
	Peripheral Blood Mononuclear Primary Cells TC03	GSM543023
	Peripheral Blood Mononuclear Primary Cells TC007	GSM613911
	Peripheral Blood Mononuclear Primary Cells TC009	GSM669606
	MRE-seq	H1Es Batch1
H1Es Batch2		GSM450236
Breast Luminal Epithelial Cells RM066		GSM613833
Breast Luminal Epithelial Cells RM070		GSM613818
Breast Luminal Epithelial Cells RM071		GSM613826
Breast MyoEpithelial Cells RM066		GSM613834
Breast MyoEpithelial Cells RM070		GSM613821
Breast MyoEpithelial Cells RM071		GSM613908
Breast Stem Cells RM066		GSM613837

Experiment	Sample	GEO ID
	Breast Stem Cells RM070	GSM613907
	Breast Stem Cells RM071	GSM613829
	CD4 Memory Primary Cells TC003	GSM613842
	CD4 Memory Primary Cells TC007	GSM613903
	CD4 Memory Primary Cells TC009	GSM669599
	CD4 Naive Primary Cells TC003	GSM543011
	CD4 Naive Primary Cells TC007	GSM613901
	CD4 Naive Primary Cells TC009	GSM613920
	CD8 Naive Primary Cells TC003	GSM543013
	CD8 Naive Primary Cells TC007	GSM613905
	CD8 Naive Primary Cells TC009	GSM613923
	Fetal Brain HuFNSC01	GSM669604
	Fetal Brain HuFNSC02	GSM669605
	Neurosphere Cultured Cells, Cortex Derived HuFNSC01	GSM669600
	Neurosphere Cultured Cells, Cortex Derived HuFNSC02	GSM669602
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC01	GSM669601
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC02	GSM669603
	Peripheral Blood Mononuclear Primary Cells TC03	GSM543009
	Peripheral Blood Mononuclear Primary Cells TC007	GSM613898
	Peripheral Blood Mononuclear Primary Cells TC009	GSM613919
Histone ChIP-seq	CD8 Naive Primary Cells TC001 H3K4me1	GSM613814
	CD8 Naive Primary Cells TC001 H3K4me3	GSM613811
	CD8 Naive Primary Cells TC001 H3K36me3	GSM669593
	CD8 Naive Primary Cells TC001 H3K27me3	GSM613815
	CD8 Naive Primary Cells TC001 H3K9me3	GSM613812
	Fetal Brain HuFNSC01 H3K4me1	GSM806942
	Fetal Brain HuFNSC01 H3K4me3	GSM806943
	Fetal Brain HuFNSC01 H3K36me3	GSM806946
	Fetal Brain HuFNSC01 H3K27me3	GSM806945
	Fetal Brain HuFNSC01 H3K9me3	GSM806944
p300 (ENCODE/HAIB)	GM12878 rep1	GSM803387
	GM12878 rep2	GSM803387
	H1	GSM803542
	HepG2	GSM803499
	SK-N-SH RA rep1	GSM803495
	SK-N-SH RA rep2	GSM803495
mRNA-seq	Breast Luminal Epithelial Cells RM035	GSM543029
	Breast Luminal Epithelial Cells RM080	GSM669620
	Breast MyoEpithelial Cells RM035	GSM543031

Experiment	Sample	GEO ID
	Breast MyoEpithelial Cells RM080	GSM669621
	CD4 Memory TC014	GSM669618
	CD4 Naïve TC014	GSM669617
	CD8 Naïve TC014	GSM669619
	Fetal Brain HuFNSC01	GSM751274
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC01	GSM751271
	Neurosphere Cultured Cells, Ganglionic Eminence Derived HuFNSC02	GSM751273
	H1ES	GSM484408
TF ChIP-seq (ENCODE)	RAD21 GM12878 Rep1	GSM803416
	RAD21 SK-N-SH RA Rep1	GSM803497
	YY1 GM12878 Rep	GSM803406
	YY1 SK-N-SH RA Rep	GSM803498
	NFKB GM12878 Rep1	GSM935478