Gene 475 (2011) 39–48

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Epigenetic regulation of transposable element derived human gene promoters

Ahsan Huda [a], Nathan J. Bowen [b], Andrew B. Conley [a], I. King Jordan [a,*]

[a] School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA 30332, USA
[b] Integrated Cancer Research Center, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Article history:
Accepted 22 December 2010
Available online 6 January 2011

Received by A.J. van Wijnen

Keywords:
Transposable elements
Gene regulation
Gene expression
Chromatin
Histone modifications

ABSTRACT

It was previously thought that epigenetic histone modifications of mammalian transposable elements (TEs) serve primarily to defend the genome against deleterious effects associated with their activity. However, we recently showed that, genome-wide, human TEs can also be epigenetically modified in a manner consistent with their ability to regulate host genes. Here, we explore the ability of TE sequences to epigenetically regulate individual human genes by focusing on the histone modifications of promoter sequences derived from TEs. We found 1520 human genes that initiate transcription from within TE-derived promoter sequences. We evaluated the distributions of eight histone modifications across these TE-promoters, within and between the GM12878 and K562 cell lines, and related their modification status with the cell-type specific expression patterns of the genes that they regulate. TE-derived promoters are significantly enriched for active histone modifications, and depleted for repressive modifications, relative to the genomic background. Active histone modifications of TE-promoters peak at transcription start sites and are positively correlated with increasing expression within cell lines. Furthermore, differential modification of TE-derived promoters between cell lines is significantly correlated with differential gene expression. LTR-retrotransposon derived promoters in particular play a prominent role in mediating cell-type specific gene regulation, and a number of these LTR-promoter genes are implicated in lineage-specific cellular functions. The regulation of human genes mediated by histone modifications targeted to TE-derived promoters is consistent with the ability of TEs to contribute to the epigenomic landscape in a way that provides functional utility to the host genome.

1. Introduction

Transposable elements (TEs) form a substantial fraction of eukaryotic genomes. TEs do not proliferate by increasing the fitness of their hosts, rather their abundance can be attributed to their ability to out-replicate their host genomes (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). In fact, transposition is primarily deleterious and may cause major disruptions for host genomes. Accordingly, eukaryotic genomes have evolved a variety of mechanisms to control the proliferation of the TEs. For instance, it is thought that epigenetic regulatory systems originally evolved to mitigate the deleterious effects of TEs by suppressing element transcription and/or preventing ectopic recombination between dispersed TE sequences (Henikoff and Matzke, 1997; McDonald, 1998; Matzke et al., 2000; Lippman et al., 2004; McDonald et al., 2005). The phenomena by which host genomes epigenetically control TEs are collectively referred to as 'genome defense' mechanisms (Yoder et al., 1997).

Epigenetic modifications can be classified into two processes: direct modification of the DNA molecule (Fazzari and Greally, 2004) or modification of the histone proteins that form the core of eukaryotic nucleosomes (Li, 2002). A number of eukaryotic species use these epigenetic mechanisms to shut down the activity of TE sequences. This process has been well studied in plants and yeast where TEs are targeted by DNA and histone methylation. This targeting causes *de novo* formation of heterochromatin and is thought to be mediated by the RNA interference pathway (Gendrel et al., 2002; Grewal and Elgin, 2007; Grewal and Jia, 2007; Slotkin and Martienssen, 2007; Suzuki and Bird, 2008; Weil and Martienssen, 2008). For instance, TE insertions have been shown to recruit repressive histone modifications and initiate the formation of local heterochromatin in *Arabidopsis thaliana* (Gendrel et al., 2002; Lippman et al., 2004). In

Abbreviations: TE, Transposable elements; ENCODE, Encyclopedia of DNA Elements; LTR, Long Terminal Repeats; LINE, Long Interspersed Nuclear Elements; SINE, Short Interspersed Nuclear Elements; MIR, Mammalian Inverted Repeats; H3K4me1, histone H3 Lysine 4 mono-methylation; H3K4me2, histone H3 Lysine 4 di-methylation; H3K4me3, histone H3 Lysine 4 tri-methylation; H3K9ac, histone H3 Lysine 4 acetylation; H3K27ac, histone H3 Lysine 27 acetylation; H3K27me3, histone H3 Lysine 27 tri-methylation; H3K36me3, histone H3 Lysine 36 tri-methylation; H4K20me1, histone H4 Lysine 20 mono-methylation; TSS, Transcription Start Site; UCSC, University of California, Santa Cruz; ChIP-seq, Chromatin Immunoprecipitation followed by high throughput sequencing; CAGE, Cap Analysis of Gene Expression; GEO, Gene Expression Omnibus; NCBI, National Center for Biotechnology Information; ANOVA, Analysis of Variance.

* Corresponding author.
E-mail addresses: ahsan.huda@gatech.edu (A. Huda), bowen@gatech.edu (N.J. Bowen), aconley@gatech.edu (A.B. Conley), king.jordan@biology.gatech.edu (I.K. Jordan).

0378-1119/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2010.12.010

another well studied system, *Schizosaccharomyces pombe*, TE insertions can attract repressive modifications which in turn lead to gene silencing and heterochromatin formation (Volpe et al., 2002).

To date, a handful of studies have explicitly evaluated the relationship between TEs and epigenetic modifications in mammals. In 2003, Kondo and Issa investigated the distribution of the histone modification histone H3 Lysine 9 dimethylation (H3K9me2) in different regions of the human genome and found Alu elements to be highly enriched for that repressive modification (Kondo and Issa, 2003). Similarly in the mouse genome, Martens et al. found TEs to be enriched for repressive marks albeit at varying levels among different TE families and cell lines (Martens et al., 2005). In another mouse study by the Bernstein and Lander groups, young long terminal repeat (LTR)-retrotransposons families (IAP and ETn) were found to be enriched for several repressive histone marks (Mikkelsen et al., 2007). Following up on this study, Matsui et al. recently reported that H3K9 methylation is the primary mechanism by which endogenous retroviruses are repressed in mouse embryonic stem cells (Matsui et al., 2010). The Jenuwein group also reported enrichment of H3K27me3, a classic repressive modification, in human SINEs (Pauler et al., 2009). All of these studies focus on repressive modifications of mammalian TEs and thus point directly to genome defense as a critical role for the histone modifications of these elements.

On the other hand, TEs are not solely deleterious; there are numerous documented cases where formerly selfish TE sequences now provide some functional utility for their host genomes (Kidwell and Lisch, 2000). This occurs most often when TEs donate regulatory sequences that help to control the expression of host genes (Feschotte, 2008). For instance, TEs are known to provide transcription factor binding sites (Jordan et al., 2003; Polavarapu et al., 2008), transcription start sites (Conley et al., 2008; Cohen et al., 2009) and enhancer elements (Bejerano et al., 2006; Santangelo et al., 2007) to their host genomes. Cases such as these, where TE sequences provide functional elements to their host genomes, can be considered as genomic 'exaptations' (Brosius and Gould, 1992). Exaptation refers to the phenomenon whereby an organismic feature plays a role for which it was not originally evolved (Gould and Vrba, 1982). TE regulatory sequences originally evolved to ensure that the TEs could replicate within genomes, thus ensuring their long term survival. Only later were these sequences exapted to serve the needs of their hosts.

Thus TEs provide sequences that help to regulate their host genomes, and TEs are frequently targeted by epigenetic modifications. These two facts led us to hypothesize that many of the regulatory effects of TEs may be mediated by epigenetic modifications. We recently tested this hypothesis for human TEs by evaluating the distributions of active and repressive histone modifications targeted to TEs genome-wide (Huda et al., 2010). This study revealed that numerous human TEs are enriched for active histone modifications, TEs closer to genes are more epigenetically modified than TEs further away from genes and more conserved TEs bear proportionally more histone modifications than younger TEs. All of these features are consistent with the notion that many TE histone modifications represent exaptations that are now used to regulate the human genome.

While the global landscape of histone modifications of human TEs indicates their overall epigenetic regulatory potential, exaptation is a phenomenon that occurs for individual organismic features on a case-by-case basis. With respect to TEs, this means that TE exaptation probably operates at the local scale, on individual TE sequence insertions, as opposed to genome-wide on all TEs. In fact, it is almost certainly the case that the majority of human TE sequences do not play a role in epigenetically regulating host genes. Therefore, our aim in this study was to identify the individual TE sequences that help to regulate human genes and to evaluate the extent to which their regulatory effects may be epigenetically mediated. To that end, we focused on histone modifications of TEs that provide promoter sequences, specifically transcription start sites (TSS), to human genes.

We used histone modification data provided by the ENCODE project to study the epigenetic regulation of TE-derived promoters in two human cell lines: GM12878 and K562 (Celniker et al., 2009). GM12878 is a lymphoblastoid cell line derived from a female donor of northern and western European descent. K562 is an immortalized cancer cell line obtained from a female donor with Chronic Myelogenous Leukemia. The data consist of genome-wide maps of the locations of eight histone modifications: H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H3K27me3, H3K36me3 and H4K20me1. All but one (H3K27me3) of these modifications are activating, *i.e.* their effect leads to euchromatin formation and promotion of gene expression. H3K27me3 is a repressive histone mark that is associated with heterochromatin and gene silencing. We also analyzed gene expression data provided by the ENCODE pilot project using exon array experiments in both the cell lines.

We co-located UCSC gene annotation data with RepeatMasker TE annotation data to obtain a set of human gene TSS derived from TE sequences. To explore the epigenetic regulatory potential of TE-derived TSS across the two cell lines, we mapped histone modification data to the TSS in both cell types and ranked them according to differential histone modifications between cell types. The cell type specific expression levels of genes with TE-derived TSS were then evaluated to determine if gene expression divergence corresponds to TE modification differences between the cell lines. These analyses demonstrated that TE-derived promoters are in fact epigenetically modified in such a way as to facilitate differential gene regulation between cell lineages. Thus, we present evidence for the epigenetically mediated exaptation of TE sequences in driving cell-type specific gene expression.

2. Results and discussion

2.1. TE-derived human gene promoters

We focused our analysis on human genes with promoter sequences derived from TEs in order to evaluate the epigenetic modifications of TEs that are most likely to have regulatory consequences for the human genome. To do this, we surveyed the human genome sequence (NCBI Build 36.1; UCSC hg18) for genes that have TE-derived promoters by comparing the locations of gene models with TE annotations. We found 1520 human genes whose TSS lie within annotated TE boundaries. These TE-derived promoters were classified according to their constituent families: Alu, L1, LTR, DNA, L2 and MIR. Using the genomic abundances of these families as background, we determined over- and under-represented families of TEs that donate TSS. Alu, DNA, LTR L2 and MIR elements are over-represented, whereas L1 is the only family that is under-represented in donating TSS (Fig. 1 and Supplementary Table 1; χ^2 test, $P \leq 5.24E-101$).

The age of TEs can be ascertained by comparison of individual element sequence insertions with sub-family consensus sequences (Kapitonov and Jurka, 1996). In the human genome, Alus and L1s are the youngest TE families, LTRs and DNA elements are of intermediate age, and the L2 and MIR families are the most ancient (Lander et al., 2001). In general, older families of TEs donate proportionally more TSS than younger ones do (Fig. 1). MIR, which is the oldest TE family in the human genome, is the most over-represented family in this set followed by L2. Both MIR and L2 sequences have previously been implicated as having regulatory function based on anomalously low levels of between species sequence divergence (Silva et al., 2003). Their over-representation among TE-derived promoters is consistent with this result and with the principle of phylogenetic footprinting, which holds that functionally important sequences are more likely to be evolutionarily conserved (Marino-Ramirez et al., 2005).

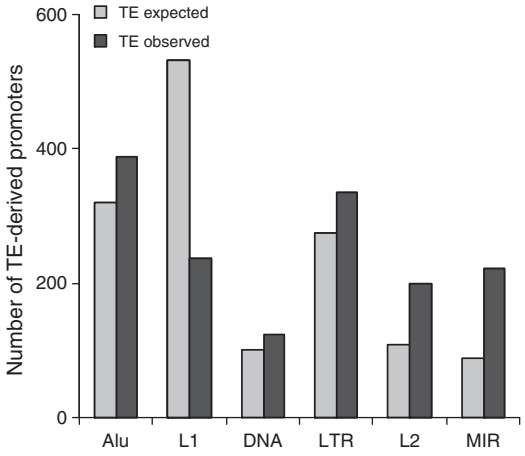

Fig. 1. Contribution of different TE families to human gene TSS. The number of promoters (TSS) derived from different TE families. Expected values were calculated based on the relative genomic abundance of the TE families. Statistical significance was calculated using the χ^2 test (Supplementary Table 1).

On the other hand, TEs are among the most rapidly evolving and lineage-specific components of eukaryotic genomes. Accordingly, if TEs donate regulatory sequences, they may help to drive regulatory divergence between evolutionary lineages (Marino-Ramirez et al., 2005; Marino-Ramirez and Jordan, 2006). Indeed, primate-specific Alu elements contribute more TSS overall than any other TE family (Fig. 1), and promoters derived from Alus may be expected to yield primate-specific patterns of gene expression. However, the abundance of Alu-derived TSS may simply reflect the overall high numbers of Alus in the genome along with the fact that they are known to be enriched in gene regions (Soriano et al., 1983; Lander et al., 2001).

LTR elements are of intermediate age and they have previously been noted for their exceptional role in providing promoter sequences to human genes (Samuelson et al., 1990; Medstrand et al., 2001; Dunn et al., 2003; van de Lagemaat et al., 2003; Bannert and Kurth, 2004; Medstrand et al., 2005; Dunn et al., 2006; Romanish et al., 2007; Conley et al., 2008; Cohen et al., 2009). Recently, Cohen et al. compiled a list of known LTR derived promoters in the human genome, which included 24 cases supported by experimental data (Cohen et al., 2009). The computational approach to identifying TE-derived promoters employed here identified 16 out of these 24 cases. The 8 missed instances may be attributed to the fact that the LTR-derived promoters evaluated by Cohen et al. correspond to alternative promoters that are not represented by gene models, which tend to identify dominant or canonical promoters. Overall, our analysis also indicates that LTR elements are over-represented in TE-derived promoters albeit marginally. LTR elements, however, turn out to be more prominent among differentially regulated genes, as will be shown in later sections.

2.2. Genome-wide maps of epigenetic histone modifications

Histone modifications were analyzed in order to evaluate how the epigenetic modifications of TE-derived promoters relate to cell-type specific human gene expression. In order to do this, we needed to first characterize the regulatory effects of individual histone modifications genome-wide. To this end, we analyzed genome-wide histone modification data, characterized as part of the ENCODE project, for the related human cell lines GM12878 and K562 (Celniker et al., 2009). Genome-wide maps of histone modifications for these cell lines were generated by the Broad Institute using chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq). There are genomic location data available for eight histone modifications – H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27me3, H3K27ac, H3K36me3 and H4K20me1 – in both cell lines.

Histone modifications serve to either activate or repress the transcription of genes. We defined the effect for each of the eight individual histone modifications analyzed here as active or repressive based on their associations with genes expressed at different levels, regardless of whether these genes contained a TE-derived promoter. To do this, we established presence/absence calls for each modification over all human genes based on its enrichment at a gene locus as described in the Materials and methods. Then for each modification, the log normalized ratio of the average expression level for genes that are marked present for the modification over the average expression levels of genes that are absent for the modification was calculated. These ratios classify the eight histone modifications into seven active modifications – H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H3K36me3, H4K20me1 – and one repressive modification – H3K27me3 – in each of the two cell lines (Fig. 2). These results are statistically significant (Supplementary Table 2; Student's t test, $0 \leq P \leq 2.1E-148$), qualitatively identical for each cell line and consistent with previous results (Barski et al., 2007; Wang et al., 2008; Huda et al., 2010).

2.3. Epigenetic modifications at TE-derived promoters

Having established demonstrable regulatory effects for the eight histone modifications over all human genes in the GM12878 and K562 cell lines, we wished to evaluate the enrichment of TE-derived promoters with respect to these histone marks in the two cell lines. To do this, we mapped ChIP-seq tags corresponding to eight epigenetic histone modifications in both cell lines to our dataset of 1520 TE-derived promoters. The total numbers of tags for each modification in these promoters were converted into log enrichment ratios by comparing the gene tag counts against the genomic background tag counts as described in the Materials and methods (Fig. 3). Across both cell lines, TE-derived promoters were found to be significantly enriched or depleted for 15 out of 16 comparisons (8 modifications × 2 cell lines) relative to the genomic background (Supplementary Table 3; χ^2 test, $0.39 \leq P \leq 0$). The GM12878 cell line shows significant enrichments for all 7 activating modifications (H3K4me2, H3K9ac,

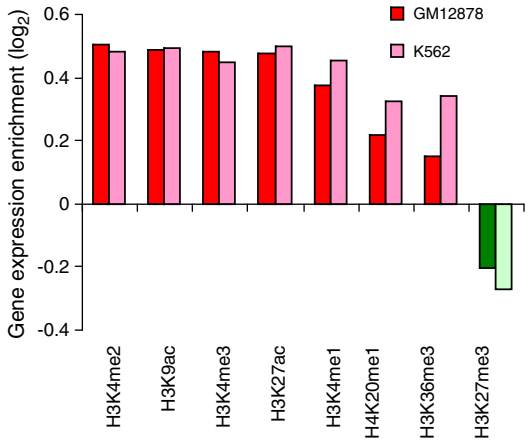

Fig. 2. Characterization of individual histone modifications as active or repressive. For each histone modification, in both GM12878 and K562 cells, the average expression level of all genes marked present for the modification were divided by the average expression level of all genes marked absent for the modification and this ratio was log normalized. Presence and absence calls for each modification at human gene promoters were determined using Poisson distributions parameterized by the genomic background tag count of each modification (Materials and methods). Activating histone modifications show positive gene expression enrichment ratios, and repressive histone modifications show negative enrichment ratios (Wang et al., 2008). Statistical significance values for the enrichment ratios of each modification were calculated using the Student's t test (Supplementary Table 2).

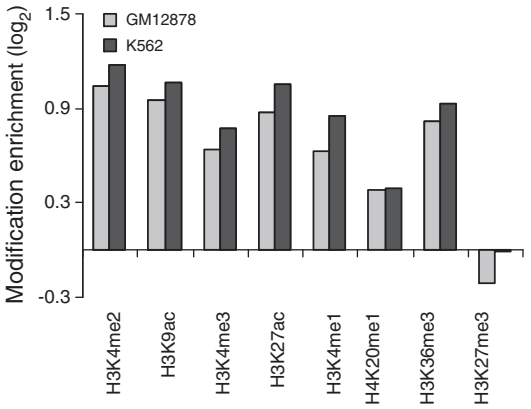

Fig. 3. Histone modification enrichment in TE-derived promoters. Enrichment values for the eight individual histone modifications were calculated over the 1520 TE-derived promoters in both GM12878 and K562 cells. Log_2 enrichment values are computed by comparing the average ChIP-seq tag counts in TE-derived promoters against the genomic background tag counts (Materials and methods). Statistical significance values for each modification were calculated using the χ^2 test (Supplementary Table 3).

H3K4me3, H3K27ac, H3K4me1, H3K36me3 and H4K20me1) and a significant depletion for the only repressive modification (H3K27me3). The same promoters in the K562 cell line are also significantly enriched for all activating modifications, and they are depleted, albeit not significantly so, for the repressive modification HeK27me3. There are a number of other well characterized repressive histone modifications that are not available for genome-wide analysis in these cell lines. This introduces some bias into the analysis, and it is a formal possibility that the TE sequences that are found to provide promoters here are enriched for other repressive modifications. However, since active and repressive modifications tend to be co-located along the genome, this is not likely to be the case.

As a control analysis, we compared the enrichment of these same histone modifications across TE families genome-wide. The patterns of histone modifications at TE-derived promoters are entirely distinct from those seen for TEs genome-wide. In almost all cases, TE sequences that provide promoters are more enriched for active modifications and more depleted for the repressive modification than any of the TE families genome-wide. For instance, TE-promoter sequences show the highest enrichment for 6 out of 7 active marks in GM12878 (Supplementary Fig. 1 and Supplementary Table 4; χ^2 test $0 \leq P = 7E{-}4$) and all 7 active marks in K562 (Supplementary Fig. 2 and Supplementary Table 4; χ^2 test $0 \leq P = 2E{-}3$). In addition, on average, TE promoter sequences are more depleted for the repressive mark H3K27me3 than are TE sequences genome-wide. This is consistent with previous results that show TE sequences to be targeted by H3K27me3 (Pauler et al., 2009).

Taken together, these data indicate that TE-derived promoters are enriched for activating histone modifications, suggesting that these formerly selfish sequences are no longer epigenetically repressed and may instead help to mediate the epigenetic activation of human genes.

2.4. TE-promoter epigenetic modifications and gene expression

Given the observation that TE-derived promoters are epigenetically modified, we next tried to relate their epigenetic modifications to the levels of cell-type specific expression of the genes they regulate. To do this, microarray gene expression data for the GM12878 (20 samples) and K562 (21 samples) cell lines were taken from the NCBI Gene Expression Omnibus (GSE12760), and processed as described in the Materials and methods section to yield cell-type specific expression levels for human genes with TE-derived promoters. To evaluate the functional effect of TE histone modifications on gene expression, we examined the modification landscape of the promoter regions of genes expressed at varying levels. 10 kb genomic regions were centered on the TE-derived TSS, and histone modification tag counts of the seven active modifications were summed across promoters in both cell lines. Genes with TE-derived promoters were grouped into equal sized bins of high, medium and low expression for each cell type. For each bin, we computed and plotted the total number of tags-per-position along the TE-TSS centered genomic regions (Fig. 4a and b). Overall, histone modification tag counts peak around the TE-derived TSS as can be expected for epigenetically regulated genes. Furthermore, genes with high expression levels have more active modifications, whereas genes with medium and lower expression have successively fewer active modifications. These plots demonstrate that gene expression is proportional to the enrichment of active modifications for both cell types indicating that TE-derived promoters are involved in epigenetic gene regulation.

We attempted to explore the relationship between histone modifications at TE-derived promoters and gene expression in a more quantitative way by devising a 'modification index', which represents the nature and extent to which a gene is epigenetically modified. The modification index takes into account the level at which a modification is deemed to be active or repressive, as well as the number of tags for that modification present at the gene locus (see Materials and methods). In other words, genes that are enriched for active modifications will have higher modification indices and *vice versa*. We computed the modification index of TE-derived promoters and plotted these values against their expression levels in GM12878 and K562 cell lines. The modification indices of genes in our dataset are significantly correlated with their expression levels in each cell line (Fig. 4c and d GM12878: $\rho = 0.51$, $P = 9E{-}52$, K562: $\rho = 0.46$, $P = 9E{-}44$). Thus, genes with TE-derived promoters that bear more activating modifications tend to have a higher expression values whereas genes that are enriched for repressive modifications have lower expression. Taken together, the two approaches described in this section support the notion that TE-derived promoters are epigenetically regulated to drive the expression of human genes. In other words, the set of human genes with TE-derived promoters analyzed here represents a collection of TE exaptations whose regulatory effects are, at least in part, epigenetically mediated.

2.5. Cell-type specific epigenetic regulation of TE-derived promoter genes

Next, we wanted to evaluate whether histone modifications of TE-derived promoters could underlie cell-type specific gene expression. To address this question, we compared gene expression divergence with cell-type specific TE-promoter histone modifications for the GM12878 and K562 cell lines. To uncover differentially expressed TE-promoter genes, we performed ANOVA on the 20 and 21 samples of GM12878 and K562 microarray expression data. Using a *P*-value cutoff of 1E−4, we found 522 out of 1520 genes with significantly divergent expression between the two cell lines (Fig. 5). 296 genes are up-regulated in K562 and down-regulated in GM12878, whereas another 226 genes are up-regulated in GM12878 and down-regulated in K562. For differentially regulated TE-promoter genes, histone modification divergence values were calculated as the differences between the modification indices of each gene in each cell line (GM12878 mod. index − K562 mod. index). TE-promoter histone modification divergence values can be seen to be largely concordant with gene expression divergence between cell lines (Fig. 5). In addition, TE-promoter histone modification divergence values are significantly positively correlated with gene expression divergence values (Fig. 6a; $\rho = 0.61$, $P = 3E{-}54$). Thus, TE-promoters with greater modification divergence tend to regulate genes with higher expression divergence, *i.e.* genes with divergently modified TE-promoters are also divergently expressed between the GM12878 and K562 cell lines. This significant positive correlation holds when

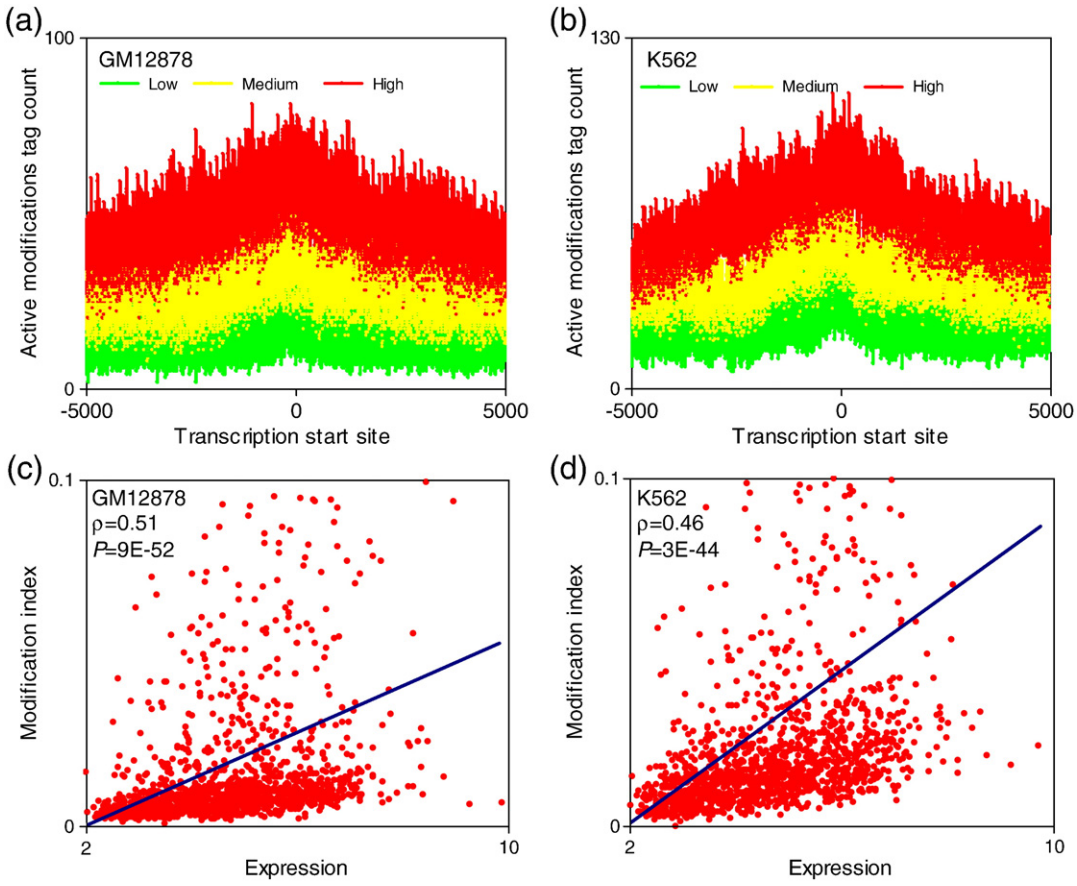

Fig. 4. Relationship between TE-promoter histone modifications and gene expression. (a and b) 10 kb regions surrounding TE-derived TSS were analyzed for all 1520 TE-promoters. The numbers of ChIP-seq tags-per-position are plotted for active modifications in genes with low, medium and high expression in (a) GM12878 and (b) K562 cell lines. (c and d) Scatter-plots of the TE-promoter histone modification indices against gene expression levels are shown for (c) for GM12878 and (d) for K562. Linear trend lines along with Spearman's rank correlations and statistical significance values are shown.

different ANOVA cut-off *P*-values are used or when all 1520 TE-promoter genes are considered (Supplementary Fig. 3). Taken together, these data underscore the ability of TE-derived promoters to participate in the epigenetic regulation of cell-type specific gene expression.

It can also be seen that individual TE-families have characteristic values of both promoter modification divergence and gene expression divergence between cell lines (Fig. 6b). As is the case for individual TE-promoter genes, the TE-family-specific modification and expression divergence values are significantly positively correlated ($\rho = 0.94$, $P = 2E{-}9$). The TE-promoters derived from the LTR class of elements have the highest levels of both modification and expression divergence between cell lines (Fig. 6b). In addition, LTR-derived promoters are significantly over-represented among both the top 100 most divergently modified promoters (χ^2 test, $P = 0.02$) and the top 100 most divergently expressed TE-promoters (Fig. 6c; χ^2 test $P = 0.003$). These observations suggest that LTRs may play a special role in the epigenetically mediated regulation of cell-type specific human gene expression.

2.6. Epigenetic regulation of LTR-derived promoters

The implication of LTR-elements as having a prominent role in the epigenetic regulation of cell-type specific gene expression is noteworthy in light of numerous previous studies showing that LTR-elements, primarily endogenous retroviruses, participate in the regulation of mammalian genes (Samuelson et al., 1990; Medstrand et al., 2001; Dunn et al., 2003; van de Lagemaat et al., 2003; Bannert and Kurth, 2004; Medstrand et al., 2005; Dunn et al., 2006; Romanish et al., 2007; Conley et al., 2008; Cohen et al., 2009). There is also evidence that LTR-elements are involved in the epigenetic regulation of mammalian genes. An LTR retrotransposon, intra-cisternal A particle (IAP), inserted upstream of the agouti locus becomes activated as a cryptic promoter of the gene upon local hypomethylation (Morgan et al., 1999; Whitelaw and Martin, 2001). Expression of agouti driven from the hypomethylated LTR promoter of the IAP results in a syndrome of phenotypes including yellow fur, obesity and diabetes as well an increased tumor-genesis. The discovery of the LTR-driven epigenetic regulation of the agouti locus, when considered together with the abundance of mammalian LTR elements, was taken to indicate that this kind of mechanism may be widespread. Consistent with this notion, we have uncovered evidence for LTR-mediated epigenetic regulation of numerous human genes (Additional File: Locations of LTR-derived promoters). Below, we describe a few individual cases of human genes that show evidence for epigenetic regulation of cell-type specific expression mediated by LTR-derived promoters.

GM12878 and K562 cells are derived from hematopoietic stem cells that differentiate into a variety of blood cell types (Alberts et al., 2002). GM12878 cells are lymphoblast precursors derived from lymphoid stem cells, whereas K562 cells are cancerous myeloid stem cells derived from Chronic Myelogenous Leukemia (Supplementary Fig. 4). SAGE1 (sarcoma antigen 1) encodes a cell surface antigen that is known to be expressed in tumor tissues relative to normal tissues and has been investigated as a potential target for cancer immunotherapy (Atanackovic et al., 2006). SAGE1 expression is driven by an LTR-promoter derived from the ERVL-MaLR LTR-retrotransposon subfamily, and it is upregulated in K562 cells relative

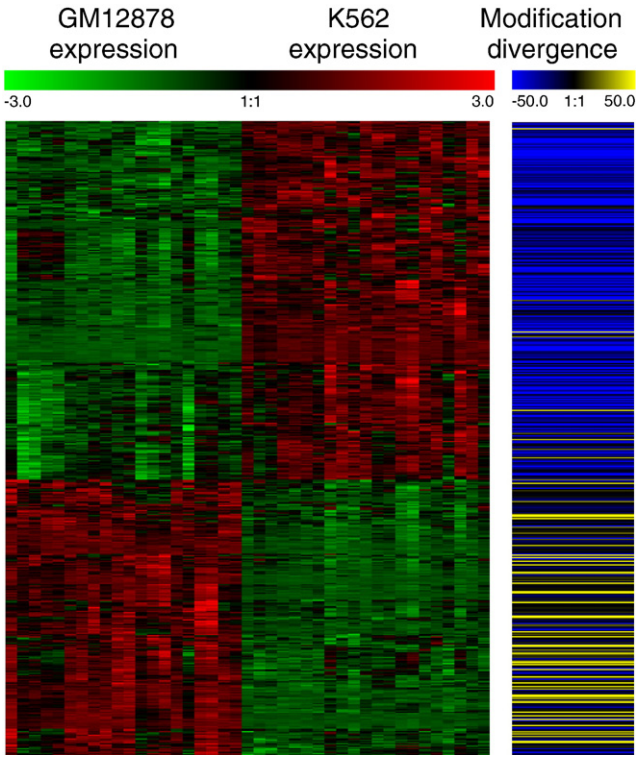

Fig. 5. Gene expression and TE-promoter histone modification divergence for differentially expressed genes. Cell-type specific gene expression levels, along with their corresponding TE-promoter histone modification divergence values, are shown for 522 differentially expressed genes. (Left) Normalized exon array expression data in 20 GM12878 and 21 K562 samples are presented as a heat map, and (right) corresponding histone modification divergence values are represented as horizontal bars with varying color intensity. Genes were clustered based on their expression levels using hierarchical clustering. Histone modification divergence values are calculated as the difference between promoter histone modification indices for GM12878-K562.

to GM12878 (ANOVA F = 398.1, $P \sim 0$). Accordingly, the ERVL promoter is enriched for active modifications and depleted for repressive modifications in K562 as compared to GM12878.

Like SAGE1, expression of CT45 (cancer/testes antigen 45, CT45-1, CT45-4, CT45-6) genes is also characteristic of hematological malignancies and their upregulation is indicative of cancer progression (Chen et al., 2010). SAGE1 and these three CT genes are paralogs that appear to use the same kind of ERVL LTR as a promoter. Presumably, these genes arose via duplication after the promoter was derived from an ERVL insertion. In all cases, the CT45 genes are upregulated in K562 relative to G12878 (ANOVA F = 865.3, $P \sim 0$) and also relatively enriched for active histone modifications at the LTR promoter.

IL23R (Interleukin 23 Receptor) is a receptor for the cytokine IL23 and is expressed on the surface of a number of hematopoietic cell types derived from the lymphoid and myeloid lineages. It has been implicated in a number of functions related to immune responses to diseases including cancer (Langowski et al., 2006), and has previously been shown to be upregulated in Chronic Myelogenous Leukemia (CML) cells (Zhang et al., 2006). Consistent with these findings, our

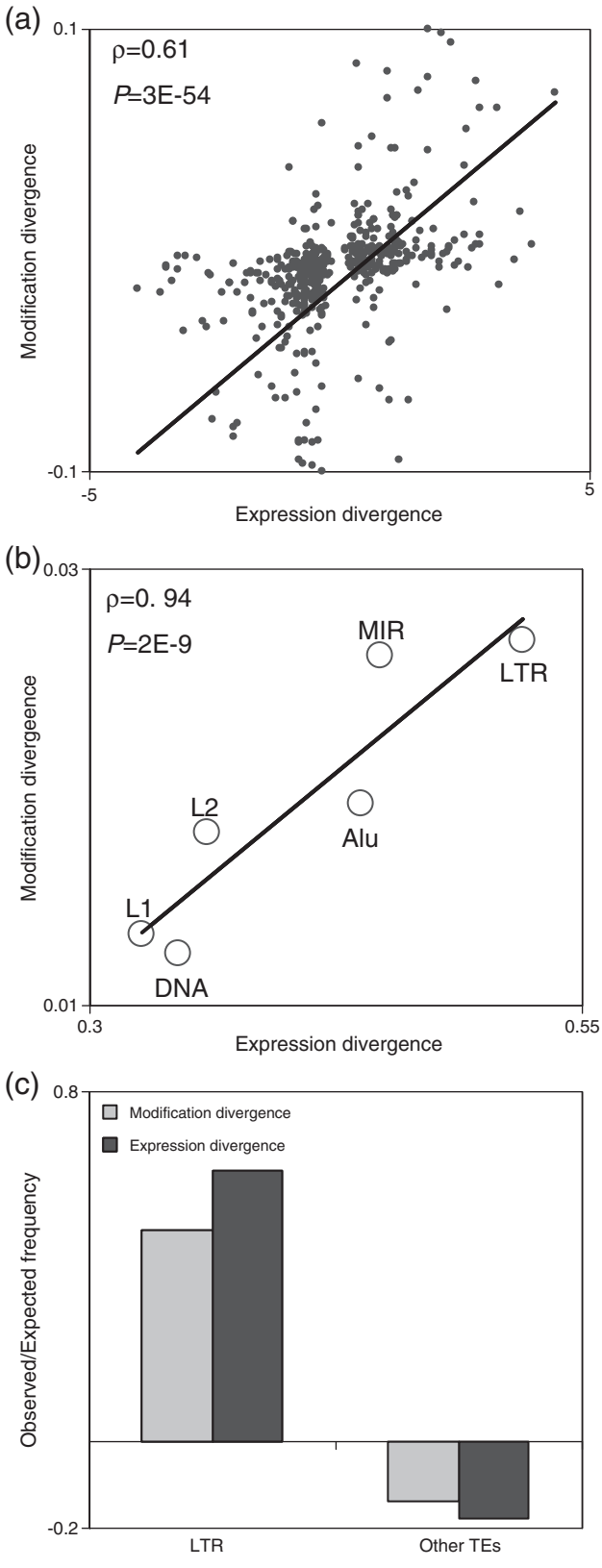

Fig. 6. Comparison of TE-promoter histone modification divergence and gene expression divergence. (a) Scatter-plot of TE-promoter histone modification divergence against gene expression divergence, between GM12878 and K562 cells, for differentially expressed genes. The linear trend line along with the Spearman's rank correlation and statistical significance value are shown. (b) Scatter-plot of average TE-promoter histone modification divergence against average gene expression divergence for individual TE families. The linear trend line along with the Spearman's rank correlation and statistical significance value are shown. (c) The relative frequencies of LTR-derived TE-promoters are compared against the relative frequencies of all other TE-derived promoters for the 100 most divergently expressed and the most divergently modified TE-promoter genes. Relative frequencies were calculated as (observed-expected)/expected TE-promoter counts for the different TE-families, where expected counts were based on the TE-family counts over all 1520 TE-promoter genes.

analysis indicates that the LTR derived promoter for the IL23R gene is epigenetically modified with active modifications in the K562 cell line and also expressed at higher levels in K562 compared with the GM12878 cell line (ANOVA F = 283.3, $P \sim 0$; Fig. 7).

There are also examples of epigenetically regulated LTR-promoter genes that may play a role in the developmental specification of cell-type specific function along the lymphoid lineage (GM12878). IL1R2 (Interleukin-1 Receptor 2) is another gene that initiates transcription from within an ERVL LTR-element. The IL1R2 LTR-promoter is enriched for active modifications in GM12878 relative to K562, and accordingly it drives upregulated expression in GM12878. The IL1R2 gene encodes a receptor for the cytokine IL1 which is responsible for activating B and T lymphocytes (Kuno and Matsushima, 1994). IL1R2 upregulation in GM12878 suggests a role for its ERVL promoter in driving lymphoid specific expression via differential epigenetic histone modifications.

2.7. Alternative TE-derived promoters

In most cases, human genes have multiple transcript variants including alternative promoters, and TE sequences often serve as minor promoters that drive a specific subset of the overall gene expression (Cohen et al., 2009). Thus, it may be the case that many of the TE-derived promoters analyzed here are alternative promoters that complement dominant non-TE derived promoters. To evaluate this possibility, we searched for alternative non-TE-derived promoters for those genes that have TE-derived promoters (see Materials and methods). We identified 2481 additional promoters that do not initiate transcription within TEs, and correspond to 1211 out of 1520 genes with TE-derived promoters. For each of these genes, we used CAGE (Cap Analysis of Gene Expression) data to compare the levels of transcriptional initiation generated from TE-derived promoters versus non-TE-derived promoters. For 42% (1044/2481) of cases, the non-TE-derived promoter serves as the dominant promoter, in the sense that it initiates higher levels of transcription, whereas 31% (476/1520) of the TE-derived promoters are dominant. Thus, there is a slight tendency for non-TE derived promoters to drive higher levels of expression than TE-derived promoters (Sign test $Z = 14.5$ $P = 6.8E-48$), consistent with previous results (Birney et al., 2007; Rosenbloom et al., 2010).

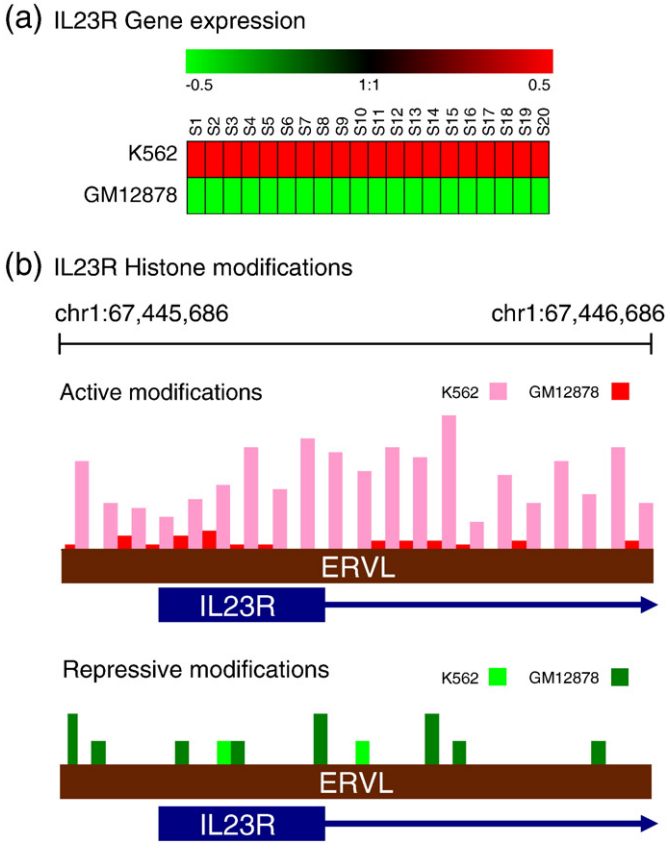

Fig. 7. Cell-type specific gene expression and TE-promoter histone modifications for IL23R. (a) Relative IL23R gene expression levels are shown across replicate samples for the K562 and GM12878 cell lines. (b) Cell-type specific histone modifications of the IL23R TE-derived promoter. Locations of the TSS and first exon of the gene are shown along with the location of the ERVL sequence from which they are derived. Relative ChIP-seq tag counts for active and repressive histone modifications, binned in 50 bp windows, are shown for the ERVL-derived promoter in both cell types.

We also compared the epigenetic regulation of TE-derived promoters with that of non-TE-derived promoters. To do this, we analyzed the histone modification landscape of non-TE-derived promoters, together with gene expression data, and found their epigenetic regulation to be similar to TE-derived promoters. We computed the enrichment of histone modification in our set of non-TE-derived promoters and found them to be significantly enriched for all modifications in GM12878 cell line and significantly enriched for all but one (H3K27me3) histone mark, which was found to be significantly depleted (Supplementary Fig. 5 and Supplementary Table 5; χ^2 test, $0 \leq P \leq 6.9E-33$). We also computed the modification index of non-TE-derived promoters and related that to their expression in each cell line. As is the case with TE-derived promoters, the modification index is significantly positively correlated with gene expression in non-TE-derived promoters (Supplementary Fig. 6 and 7; GM12878: $\rho = 0.49$, $P = 8E-150$, K562: $\rho = 0.55$, $P = 6E-196$). Similarly, modification divergence between the GM12878 and K562 cell line is also significantly correlated with expression divergence in the two cell types (Supplementary Fig. 8; $\rho = 0.48$, $P = 8E-143$). Thus, TE-derived promoters resemble non-TE-derived promoters in our measures of epigenetic modifications and are similarly related to cell type specific gene expression.

We also evaluated the sequence conservation of TE-derived TSS to assess the extent to which TE-derived promoter activity may be evolutionarily conserved. Comparative analysis of nucleotide conservation scores, based on whole genome sequence alignment of 18 mammalian species (UCSC PhyloP) (Siepel et al., 2006), reveals that TE-derived TSS are significantly less conserved than the human genomic background (TE-TSS conservation $= 0.31 \pm 0.15$, genome conservation $= 0.43$; $z = 30.3$, $P \approx 0$). The lack of apparent conservation of TE-derived promoter activity is not surprising when you consider that TE sequences are the most lineage-specific and rapidly evolving sequences in eukaryotic genomes (Marino-Ramirez et al., 2005). In fact, we have previously shown that numerous functionally active regulatory sequences are derived from TE sequences that are not evolutionarily conserved (Marino-Ramirez et al., 2005; Marino-Ramirez and Jordan, 2006; Piriyapongsa and Jordan, 2007; Polavarapu et al., 2008). This result is consistent with our finding that the majority of TE-derived TSS serve as alternative promoters in the sense that TE-derived sequences may also serve as alternative promoters over evolutionary time-scales. Evolutionarily emergent TE-derived promoters may be expected to provide lineage-specific regulatory functions that could encode phenotypic differences between species.

3. Conclusions

Previous studies on the epigenetic modifications of mammalian TEs have focused on repressive histone modifications that presumably serve to mitigate the deleterious effects of TEs (Kondo and Issa, 2003; Martens et al., 2005; Mikkelsen et al., 2007; Pauler et al., 2009; Matsui et al., 2010). However, we recently showed that, genome-wide, human TEs are epigenetically modified in a way that suggests some elements have been exapted to regulate their host genome (Huda

et al., 2010). In this report, we demonstrate one specific way that histone modifications of human TEs can facilitate the regulation of host genes. We show that TE-derived promoters are epigenetically modified to regulate gene expression in a cell-type specific manner, and the TE-mediated regulation of these genes may be related to lineage-specific cellular functions. These data underscore the potential for epigenetically mediated TE exaptations to influence the regulation of hundreds of human genes.

The results reported here, taken together with previous studies on TE epigenetic modifications, indicate that histone modifications of TE-derived sequences serve both repressive and activating roles. Initially, TE sequences are selfish genetic elements that do not play any adaptive role for the host genomes in which they reside. Indeed, active TEs can be considerably deleterious for the host and are consequently repressed by a number of mechanisms including histone modifications and DNA methylation. However, over the course of evolution TE-derived sequences may become exapted to play some functional role for their host genome, and this has occurred a number of times via the donation of regulatory sequences including novel promoters. Our analysis of TE-derived promoters indicates that the regulatory exaptation of TE sequences is mediated in part by epigenetic histone modifications, which can be cell-type specific and help to facilitate differential expression.

4. Materials and methods

4.1. Identification of TE-derived promoters

We downloaded the annotations for UCSC genes (Hsu et al., 2006) and TE RepeatMasker annotations from the March 2006 build (NCBI Build 36.1; UCSC hg18) of the human genome using the UCSC table browser (Karolchik et al., 2003; Karolchik et al., 2004). The start coordinates of genes were intersected with the TE annotation coordinates to identify TE-derived promoters, which are defined as TSS of UCSC genes that are located within TE sequences. This analysis yielded 1533 genes that initiate transcription in TE sequences.

4.2. Identification of alternative promoters

UCSC gene models were also used to identify alternative promoters for genes with TE-derived promoters. To do this, genes with TE-derived promoters were evaluated for the presence of overlapping gene models (*i.e.* mapped transcripts) that shared part or most of the exon/intron structure but differed with respect to the TSS. The locations of these TSS were taken as non-TE-derived alternative promoters that could be paired with congenic TE-derived promoters.

UCSC gene models that share part or most of the exon/intron structure with Overlapping gene models with distinct TSS that have the same basic exon/intron structure were taken as non-TE-derived alternative promoters.

CAGE tag data from GM12878 and K562 cell lines, provided as part of the ENCODE data sets (Birney et al., 2007; Rosenbloom et al., 2010), were used to quantify the different levels of transcription initiation from TE-derived promoters versus non-TE-derived promoters at the same locus. Transcription initiation levels for individual promoters were measured as the number of CAGE clusters that map within 200 bp of the TSS.

4.3. Gene expression analysis

We downloaded Affymetrix exon array signal intensity data from the GEO database under accession number GSE12760. This dataset contains 20 samples of GM12878 and 21 samples of K562 cell line analyzed as part of the ENCODE project (Birney et al., 2007). We normalized the dataset using the MAS5 algorithm provided by the Bioconductor package Exonmap (Miller et al., 2007). The normalized data was mapped to a genomic locus by averaging the expression values of all probes whose genomic coordinates lay within that the boundaries of that locus for all replicates. The genomic locus of a TE-derived promoter gene was defined as bounded by the transcription start site to the transcription end site. In our dataset of 1533 genes, we were able to obtain expression probes for all but 13 genes. We eliminated these genes from consideration to obtain a final dataset of 1520 TE-derived promoter genes.

4.4. Gene expression and histone modification enrichment analysis

We downloaded the ENCODE (Birney et al., 2007) histone modification data in GM12878 and K562 cell lines from the USCS genome browser for 8 histone modifications: H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27me3, H3K27ac, H3K36me3, H4K20me1. These data are taken from the 2009 ENCODE release based on the scale-up phase of the project that covers the entire human genome sequence (Rosenbloom et al., 2010). We defined proximal promoters of TE-derived genes as 1 KB upstream and downstream of the TSS. In this region, we found the number of ChIP-seq tags of each histone modification and used it to calculate a binary presence/absence call of that modification in each promoter using the Poisson distribution as background. We associated the promoters with their respective genes and computed gene expression as described in the preceding section. We combined the gene expression data from both cell lines to obtain gene expression fold change in genes that bear different histone modifications as follows:

$$\textit{Expression fold change}(fc) = log_2\left(\frac{\textit{average expesion of genes with modification}}{\textit{average expession of genes without modification}}\right)$$

4.5. Histone modification enrichment of TE-derived promoters

We computed the number of tags of each of the eight histone modifications that mapped within 1 KB upstream and downstream of the TE-derived TSS. The enrichment of a particular modification was calculated using the following formula:

$$\textit{Modification enrichment}_{\substack{\text{H3K4me1, H3K4me2,}\\ \text{H3K4me3, H3K9me1,}\\ \text{H3K27me3, H3K36me1,}\\ \text{H3K36me2, H4K20me2}}} = \left(\frac{\textit{Normalized tag count in TE derived promoter}}{\textit{Normalized tag count in genomic background}}\right)$$

where

$$\textit{Normalized tag count}_{m=\text{H3K4me1, H3K4me2, H3K4me3, H3K9me1, H3K27me3, H3K36me1, H3K36me2, H4K20me2}} = \frac{\textit{number of tags}_m \textit{ in genomic locus}}{\textit{length of genomic locus}}$$

4.6. Modification index

We calculated the modification index as a measure of the combined effect of all histone modifications at a genomic locus. Modification index for all modifications in a particular cell line is computed as follows:

$$\textit{Modification Index}\left(m_{\substack{\text{H3K4me1, H3K4me2, H3K4me3, H3K9me1,}\\ \text{H3K27me3, H3K36me1, H3K36me2, H4K20me2}}\right) = \frac{\sum_{m=1}^{8}(\textit{number of tags of } m \times fc_m)}{\textit{length of genomic locus}}$$

fc = Expression fold change defined earlier.

4.7. Statistical analyses

We used a two tailed χ^2 test with $d.f. = 5$ to determine the statistical significance of the over- and under-represented TE-families that donate human transcription start sites (Fig. 1). The genomic abundance of TE families was used to compute the expected number of promoters derived from each family. The χ^2 test was also used to ascertain the statistical significance of individual histone modification enrichments in TE-derived promoters (Fig. 3). In this case, the total number of mapped tags of each histone modification in each cell line was normalized by the length of the genome and taken as background.

To determine the regulatory effect of individual histone modifications, we modeled the genomic background tag distributions of each histone modification using the Poisson distribution parameterized with the genomic average tag count per position (Wang et al., 2008). From each histone modification-specific genomic background tag count Poisson distribution, we determined the threshold for the number of tags present at a genomic locus to be considered modified using a significance cutoff of $P = 0.001$. The presence or absence calls were used to calculate expression fold change as discussed in an earlier section (Materials and methods — Gene expression and histone modification enrichment analysis, Fig. 2). Statistical significance was calculated using the two tailed Student's t test with ($n = 32{,}621$ -2) degrees of freedom, where n is the total number of genes considered for expression fold change analysis.

Differentially expressed genes were identified using one way ANOVA (Analysis of variance) on two samples from GM12878 and K562 cell lines with 20 and 21 replicates ($d.f. = 39$) in each sample respectively. A Bonferroni adjusted significance cutoff of $P = 1E{-}4$ (0.05/1500 tests) was used to calculate ANOVA using the Genesis program (Sturn et al., 2002).

Spearman's rank correlation coefficients ρ were calculated for all correlation analyses using the R program. The distribution of Spearman's rank correlation coefficients ρ was determined using the formula $t = r\sqrt{(n-2)/(1-r)^2}$ with $d.f. = n-2$ to determine statistical significance (Sokal and Rohlf, 1981).

The tendency for non-TE derived promoters to serve as dominant promoters relative to alternative TE-derived promoters was evaluated using the Sign test. For this test, $Z = (|x-y|-1)/\sqrt{N}$ where $x = \#$ of genes with a dominant non-TE promoter, $y = \#$ of genes with a dominant TE-promoter and $N = $ total # of genes. The corresponding P-value was approximated with a Standard Normal distribution.

Supplementary materials related to this article can be found online at doi:10.1016/j.gene.2010.12.010.

Acknowledgements

IKJ and AH were supported by an Alfred P. Sloan Research Fellowship in Computational and Evolutionary Molecular Biology (BR-4839). AH and AC were supported by the School of Biology at the Georgia Institute of Technology. NJB was supported by the Integrated Cancer Research Center at the Georgia Institute of Technology. The authors would like to thank Lee S. Katz and Leonardo Mariño-Ramírez for helpful discussions and technical advice.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. Molecular biology of the cell. Garland Science.
Atanackovic, D., et al., 2006. Expression of cancer—testis antigens as possible targets for antigen-specific immunotherapy in head and neck squamous cell carcinoma. Cancer Biol Ther 5, 1218–1225.
Bannert, N., Kurth, R., 2004. Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci USA 101 (Suppl 2), 14572–14579.
Barski, A., et al., 2007. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.
Bejerano, G., et al., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90.
Birney, E., et al., 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.
Brosius, J., Gould, S.J., 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci USA 89, 10706–10710.
Celniker, S.E., et al., 2009. Unlocking the secrets of the genome. Nature 459, 927–930.
Chen, Y.T., Chadburn, A., Lee, P., Hsu, M., Ritter, E., Chiu, A., Gnjatic, S., Pfreundschuh, M., Knowles, D.M. and Old, L.J., 2010. Expression of cancer testis antigen CT45 in classical Hodgkin lymphoma and other B-cell lymphomas. *Proc Natl Acad Sci U S A* 107, pp. 3093-8.
Cohen, C.J., Lock, W.M., Mager, D.L., 2009. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene 448, 105–114.
Conley, A.B., Piriyapongsa, J., Jordan, I.K., 2008. Retroviral promoters in the human genome. Bioinformatics 24, 1563–1567.
Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601–603.
Dunn, C.A., Medstrand, P., Mager, D.L., 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1, 3-galactosyltransferase 5 in the colon. Proc Natl Acad Sci USA 100, 12841–12846.
Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., Mager, D.L., 2006. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene 366, 335–342.
Fazzari, M.J., Greally, J.M., 2004. Epigenomics: beyond CpG islands. Nat Rev Genet 5, 446–455.
Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9, 397–405.
Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., Martienssen, R.A., 2002. Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science 297, 1871–1873.
Gould, S.J., Vrba, E.S., 1982. Exaptation; a missing term in the science of form. Paleobiology 8, 4–15.
Grewal, S.I., Elgin, S.C., 2007. Transcription and RNA interference in the formation of heterochromatin. Nature 447, 399–406.
Grewal, S.I., Jia, S., 2007. Heterochromatin revisited. Nat Rev Genet 8, 35–46.
Henikoff, S., Matzke, M.A., 1997. Exploring and explaining epigenetic effects. Trends Genet 13, 293–295.
Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., Haussler, D., 2006. The UCSC known genes. Bioinformatics 22, 1036–1046.
Huda, A., Marino-Ramirez, L., Jordan, I.K., 2010. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. Mob DNA 1, 2.
Jordan, I.K., Rogozin, I.B., Glazko, G.V., Koonin, E.V., 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19, 68–72.
Kapitonov, V., Jurka, J., 1996. The age of Alu subfamilies. J Mol Evol 42, 59–65.
Karolchik, D., et al., 2003. The UCSC Genome Browser Database. Nucleic Acids Res 31, 51–54.
Karolchik, D., et al., 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32, D493–D496.
Kidwell, M.G., Lisch, D.R., 2000. Transposable elements and host genome evolution. Trends Ecol Evol 15, 95–99.
Kondo, Y., Issa, J.P., 2003. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. J Biol Chem 278, 27658–27662.
Kuno, K., Matsushima, K., 1994. The IL-1 receptor signaling pathway. J Leukoc Biol 56, 542–547.
Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.
Langowski, J.L., et al., 2006. IL-23 promotes tumour incidence and growth. Nature 442, 461–465.
Li, E., 2002. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3, 662–673.
Lippman, Z., et al., 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471–476.
Marino-Ramirez, L., Jordan, I.K., 2006. Transposable element derived DNaseI-hypersensitive sites in the human genome. Biol Direct 1, 20.
Marino-Ramirez, L., Lewis, K.C., Landsman, D., Jordan, I.K., 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. Cytogenet Genome Res 110, 333–341.
Martens, J.H., et al., 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 24, 800–812.
Matsui, T., et al., 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464, 927–931.
Matzke, M.A., Mette, M.F., Matzke, A.J., 2000. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol Biol 43, 401–415.
McDonald, J.F., 1998. Transposable elements, gene silencing and macroevolution. Trends Ecol Evol 13, 94–95.
McDonald, J.F., Matzke, M.A., Matzke, A.J., 2005. Host defenses to transposable elements and the evolution of genomic imprinting. Cytogenet Genome Res 110, 242–249.
Medstrand, P., Landry, J.R., Mager, D.L., 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J Biol Chem 276, 1896–1903.
Medstrand, P., van de Lagemaat, L.N., Dunn, C.A., Landry, J.R., Svenback, D., Mager, D.L., 2005. Impact of transposable elements on the evolution of mammalian gene regulation. Cytogenet Genome Res 110, 342–352.
Mikkelsen, T.S., et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.

Miller, C.J., Okoniewski, M.J., Yates, T., 2007. Description of exonmap: simple analysis and annotation tools for Affymetrix exon arrays.
Morgan, H.D., Sutherland, H.G., Martin, D.I., Whitelaw, E., 1999. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 23, 314–318.
Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. Nature 284, 604–607.
Pauler, F.M., et al., 2009. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res 19 (2), 221–233.
Piriyapongsa, J., Jordan, I.K., 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS ONE 2, e203.
Polavarapu, N., Marino-Ramirez, L., Landsman, D., McDonald, J.F., Jordan, I.K., 2008. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. BMC Genomics 9, 226.
Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A., Mager, D.L., 2007. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet 3, e10.
Rosenbloom, K.R., et al., 2010. ENCODE whole-genome data in the UCSC Genome Browser. Nucleic Acids Res 38, D620–D625.
Samuelson, L.C., Wiebauer, K., Snow, C.M., Meisler, M.H., 1990. 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 10, 2513–2520.
Santangelo, A.M., de Souza, F.S., Franchini, L.F., Bumaschny, V.F., Low, M.J., Rubinstein, M., 2007. Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet 3, 1813–1826.
Siepel, A., Pollard, K.S., Haussler, D., 2006. New methods for detecting lineage-specific selection., 10*th Int'l Conf. on Research in Computational Molecular Biology (RECOMB '06)*.
Silva, J.C., Shabalina, S.A., Harris, D.G., Spouge, J.L., Kondrashovi, A.S., 2003. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. Genet Res 82, 1–18.
Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8, 272–285.
Sokal, R.R., Rohlf, J.F., 1981. Biometry: The Principles and Practice of Statistics in Biological Research. W. H. Freeman, San Francisco.
Soriano, P., Meunier-Rotival, M., Bernardi, G., 1983. The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. Proc Natl Acad Sci USA 80, 1816–1820.
Sturn, A., Quackenbush, J., Trajanoski, Z., 2002. Genesis: cluster analysis of microarray data. Bioinformatics 18, 207–208.
Suzuki, M.M., Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9, 465–476.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., Medstrand, P., 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19, 530–536.
Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., Martienssen, R.A., 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297, 1833–1837.
Wang, Z., et al., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40, 897–903.
Weil, C., Martienssen, R., 2008. Epigenetic interactions between transposons and genes: lessons from plants. Curr Opin Genet Dev 18, 188–192.
Whitelaw, E., Martin, D.I., 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. Nat Genet 27, 361–365.
Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13, 335–340.
Zhang, X.Y., et al., 2006. Identification and expression analysis of alternatively spliced isoforms of human interleukin-23 receptor gene in normal lymphoid cells and selected tumor cells. Immunogenetics 57, 934–943.