A novel mode of enhancer evolution: The *Tal1* stem cell enhancer recruited a MIR element to specifically boost its activity

Aileen M. Smith, Maria-Jose Sanchez, George A. Follows, et al.

Genome Res. 2008 18: 1422-1432 originally published online August 7, 2008
Access the most recent version at doi:10.1101/gr.077008.108

Letter

A novel mode of enhancer evolution: The *Tal1* stem cell enhancer recruited a MIR element to specifically boost its activity

Aileen M. Smith,[1] Maria-Jose Sanchez,[1,3] George A. Follows,[1] Sarah Kinston,[1] Ian J. Donaldson,[2] Anthony R. Green,[1,4] and Berthold Göttgens[1,4,5]

[1]*University of Cambridge Department of Haematology, Cambridge Institute for Medical Research, Cambridge CB2 2XY, United Kingdom;* [2]*University of Manchester Faculty of Life Sciences, Manchester M13 9PT, United Kingdom*

Altered *cis*-regulation is thought to underpin much of metazoan evolution, yet the underlying mechanisms remain largely obscure. The stem cell leukemia TAL1 (also known as SCL) transcription factor is essential for the normal development of blood stem cells and we have previously shown that the *Tal1* +19 enhancer directs expression to hematopoietic stem cells, hematopoietic progenitors, and to endothelium. Here we demonstrate that an adjacent region 1 kb upstream (+18 element) is in an open chromatin configuration and carries active histone marks but does not function as an enhancer in transgenic mice. Instead, it boosts activity of the +19 enhancer both in stable transfection assays and during differentiation of embryonic stem (ES) cells carrying single-copy reporter constructs targeted to the *Hprt* locus. The +18 element contains a mammalian interspersed repeat (MIR) which is essential for the +18 function and which was transposed to the *Tal1* locus ~160 million years ago at the time of the mammalian/marsupial branchpoint. Our data demonstrate a previously unrecognized mechanism whereby enhancer activity is modulated by a transposon exerting a "booster" function which would go undetected by conventional transgenic approaches.

[Supplemental material is available online at www.genome.org.]

Increased biological complexity during evolution does not correlate with gene number but rather with increasing complexity of gene regulation (Levine and Tjian 2003). However, the evolution of *cis*-regulatory mechanisms remains poorly understood. This is largely due to the fact that, compared with coding sequences, very little is known about the way in which regulatory information is encoded in the genome. Developmental programs are thought to be executed through transcriptional networks which provide the capability to propagate transcriptional changes through feed forward as well as positive and negative feedback loops (Davidson 2006). Individual *cis*-regulatory elements are key components of these transcription factor networks since they function as information-processing units by integrating the inputs of their respective upstream regulators into tightly controlled spatiotemporal expression patterns.

Repetitive DNA, largely derived from transposable elements, is known to make up ~50% of the human genome (Lander et al. 2001) and has historically been thought of as "junk" DNA. Following the recent completion of the opossum genome project, it has become increasingly clear that transposable element-derived DNA may play a larger role in gene regulation than previously thought (Gentles et al. 2007). Several topical reports have highlighted that thousands of these elements are under strong purifying selection and are therefore likely to have gained a beneficial function that promotes host fitness (Kazazian 2004; Bejerano et al. 2006; Xie et al. 2006). Indeed, it has been suggested that at least 16% of eutherian-specific conserved noncoding elements are derived from transposons and that these are commonly found in the gene loci of key developmental genes, suggesting that they play a role in controlling expression of these genes (Mikkelsen et al. 2007). To date, very few of these elements have been functionally validated. However, there have been specific instances where repetitive elements have been shown to function as silencers (Donnelly et al. 1999) and polyadenylation or alternative splice sites (Medstrand et al. 2005), although experimental validation has so far been largely restricted to those elements derived from the more recent, primate-specific *Alu* family of repeats. However, a role for more ancient repeats in gene regulation is also now emerging with the discovery of two neuronal enhancers that are derived from ancient short interspersed nucleotide elements (SINE) retroposons (Bejerano et al. 2006; Santangelo et al. 2007).

Hematopoietic stem cell (HSC) specification and subsequent differentiation into the many mature blood lineages is one of the best-understood vertebrate developmental systems. The stem cell leukemia (SCL) transcription factor, also known as TAL1, is essential for the specification of HSC and the development of hematopoiesis (Porcher et al. 1996; Robb et al. 1996). As part of a systematic analysis of the transcriptional mechanisms regulating the *Tal1* locus, we have identified multiple cell type-specific enhancers, each of which directs expression to a subdomain of the overall *Tal1* expression pattern (Gottgens et al. 1997; Sanchez et al. 1999; Sinclair et al. 1999; Gottgens et al. 2000, 2001; Sanchez et al. 2001; Gottgens et al. 2002a,b, 2004; Delabesse et al. 2005; Silberstein et al. 2005; Ogilvy et al. 2007). The *Tal1* +19 enhancer lies 19 kb downstream from the first exon of *Tal1* and directs

[3]Present address: Centro Andaluz de Biologia del Desarrollo, Universidad Pablo de Olavide, Carretera de Utrera Km1, Sevilla 41013, Spain.
[4]Joint senior authors.
[5]Corresponding author.
E-mail bg200@cam.ac.uk; fax +44-1223-762670.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.077008.108. Freely available online through the *Genome Research* Open Access option.

18:1422–1432 ©2008 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/08; www.genome.org

expression to hemangioblasts, HSCs, hematopoietic progenitors, and endothelium (Sanchez et al. 1999; Gottgens et al. 2002b). Activity of this element is dependent on two conserved Ets and one conserved GATA binding site. We have also developed bio informatic tools to identify functionally related regulatory enhancers situated in the *Fli1*, *Hhex* (also known as *Hex*), and *Smad6* gene loci (Donaldson et al. 2005a,b; Donaldson and Gottgens 2006; Pimanda et al. 2007a), and are thus beginning to define key components of the transcriptional networks controlling HSCs (Pimanda et al. 2007a,b).

In this paper we demonstrate that activity of the +19 enhancer is modulated by a novel "booster" element created by the insertion of a mammalian interspersed repeat (MIR) transposon ~160 million years ago.

Results

Chromatin structure of the mouse *Tal1* stem cell enhancer

Our previous analysis of the regulation of the murine *Tal1* gene identified a 5.5-kb fragment 3′ to the *Tal1* coding region in which we mapped the approximate positions of two regions of DNase I hypersensitivity (+18 and +19 regions; nomenclature reflects the approximate distance in kilobases from the beginning of exon 1; Gottgens et al. 1997). Transgenic analysis showed that this fragment directed expression of a *lacZ* reporter gene to endothelium and blood progenitors throughout ontogeny (Sanchez et al. 1999) as well as to the vast majority of long term repopulating HSCs from fetal liver and adult bone marrow (Sanchez et al. 2001). Subsequent transgenic studies demonstrated that a 640-bp fragment containing the +19 hypersensitive site was sufficient to drive reporter gene expression in embryonic tissues (Gottgens et al. 2002a) but was prone to being silenced in adult tissues (Silberstein et al. 2005). The function of the +18 remained obscure.

To assess the potential significance of the +18 region we first performed comparative sequence analysis of the intergenic segment between *Tal1* and *Pdzk1ip1*. This analysis demonstrated that, in addition to the previously described +19 and +23 enhancers (Gottgens et al. 2000), the region at +18 was also conserved between mouse and human (Fig. 1A). Real-time PCR mapping (Follows et al. 2007) was used to confirm the presence of a DNase I-hypersensitive site in the +18 region of the myeloid progenitor cell line 416B (Fig. 1B; Dexter et al. 1979). The location of the hypersensitive site was further refined using a restriction endonuclease accessibility assay with a resolution of ~100 bp (Gottgens et al. 2001) and was found to correspond precisely with the +18 peak of sequence homology (Fig. 1C). Chromatin immunoprecipitation (ChIP) analysis of 416B cells demonstrated clear binding of GATA2 to the +19 region, confirming our previous results (Gottgens et al. 2002b), but little binding to the +18 region (Fig. 1B). In contrast, histone H3K9 acetylation was more prominent over the +18 region compared to the +19 region (Fig. 1B). Taken together these results demonstrate that the +18 region is conserved and contains a DNase I-hypersensitive site together with acetylation of H3K9, all features of a *cis*-regulatory element.

The +18 region does not function as a classical enhancer

While analysis of regulatory elements in cell lines is relatively high throughput and cost effective, only in vivo analysis using transgenic animals can identify true in vivo enhancer activity across different tissues. We generated *lacZ* reporter constructs containing the +18 region (460-bp fragment), the +19 region

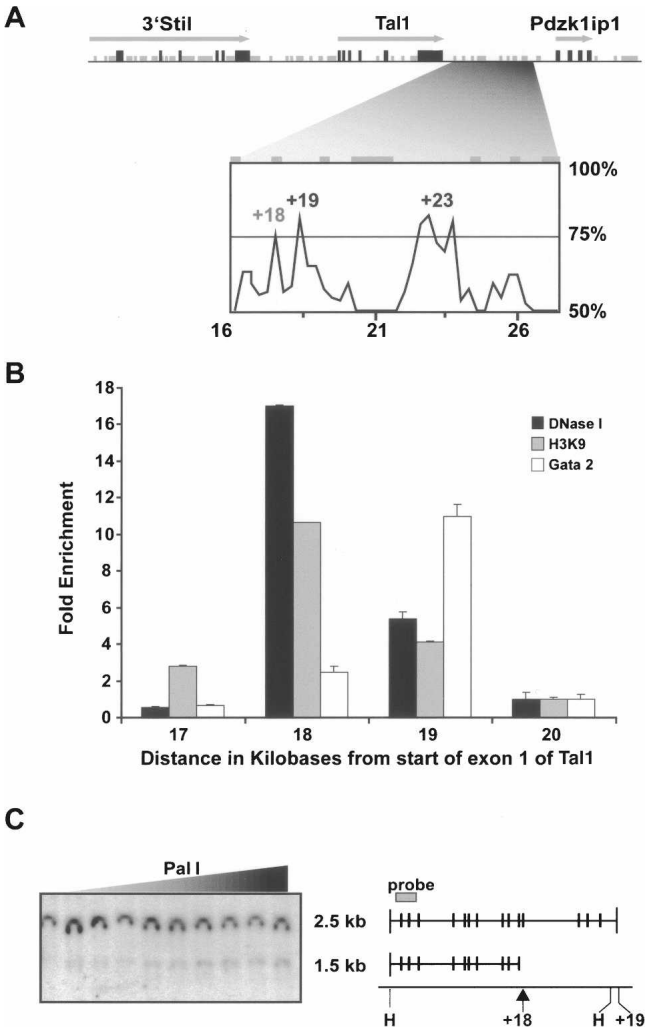

Figure 1. Identification of the + 18 region as a region of open chromatin. (*A*) Sequence homology plot between mouse and human highlighting the +19 and +23 elements with known regulatory functions and the conserved peak at +18. Numbering corresponds to distances in kilobases from the start of exon 1a with the exons shown as black boxes and the repeat regions marked in gray. (*B*) Chromatin status and transcription factor occupancy in the *Tal1* 3′ flanking region. DNase I/ChIP assays were performed in the 416B cell line with either DNase I (black bars), AcK9 (gray bars), or GATA2 (white bars) and analyzed by real-time PCR using primers across a 4-kb segment of the *Tal1* locus from +17 to +20. The levels of enrichment were normalized to untreated DNase I control or to IgG and plotted as fold increase over the +20 control region. (*C*) Restriction endonuclease accessibility assay showing the mapping of the +18 hypersensitive site (arrowhead). Nuclei were incubated with increasing concentrations of PalI and the extracted DNA digested with HindIII. Southern blot analysis was performed with the indicated probe.

(640-bp fragment), or both (1.1-kb fragment) together with the SV40 minimal promoter (Fig. 2A). These constructs were used to create transgenic embryos which were analyzed at mid-gestation by staining for β-galactosidase activity. The +19 and the +18,19 constructs both gave rise to strong hematopoietic and endothelial staining in 5/7 and 2/2 transgenic embryos, respectively (Fig. 2B; Table 1). In contrast, this pattern was absent in all transgenic embryos carrying the +18 construct ($n = 21$) and in all those with the construct containing the SV40 minimal promoter alone ($n = 9$) (Fig. 2B; Table 1). Although all four *Tal1* enhancers

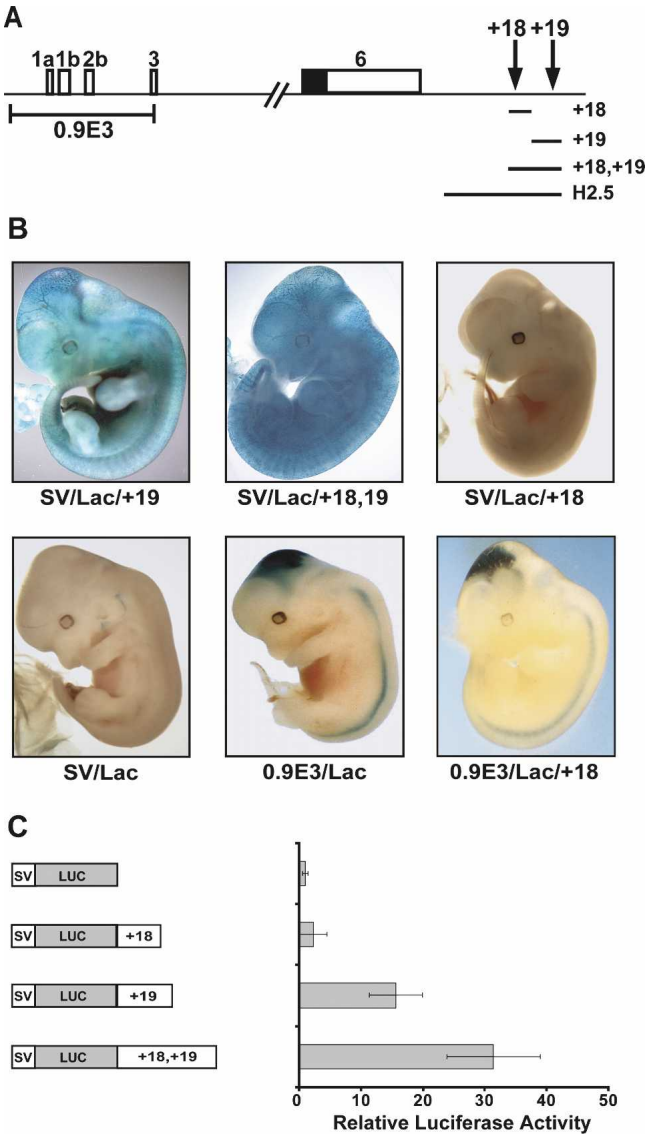

Figure 2. The +18 region does not behave as a classical enhancer. (*A*) Schematic of the *Tal1* locus indicating the positions of the 0.9E3 *Tal1* promoter, +18 and +19 regions. (*B*) Whole-mount staining of mid-gestation transgenic embryos expressing *lacZ* under the control of either the SV40 minimal promoter or the 0.9E3 *Tal1* promoter with and without the +18, the +19, or the +18/+19 elements combined. (*C*) The +18 region boosts the activity of the *Tal1* +19 enhancer in 416B stable transfection assays. Shown on the *left* are the reporter constructs of the +18 element, the 640-bp +19 element, and the 1.1-kb +18/+19 fragment inserted downstream from the SV40 luciferase reporter cassette. The results are shown on the *right* with the luciferase values presented as fold increase over the SV40 luciferase control, which was assigned a value of 1.

described so far are able to function with the SV40 promoter (Sanchez et al. 1999; Sinclair et al. 1999; Gottgens et al. 2000, 2002b, 2004; Delabesse et al. 2005), we considered the possibility that the +18 element may be unable to interact with this promoter and may require the endogenous *Tal1* promoter. We therefore generated constructs carrying the *Tal1* promoter region with and without the +18 element. As previously described, a 3.8-kb *Tal1* promoter fragment directs expression to the midbrain and spinal cord (Sinclair et al. 1999). However, inclusion of the +18 element did not result in any additional activity (Fig. 2B; Table 1). Taken together these data demonstrate that the +18 region does not show any enhancer activity in transgenic mouse embryos at mid-gestation, a time point when all organ systems are already developing.

The +18 region can boost the activity of the +19 enhancer in stable transfection assays

Having established that the +18 region did not function as an independent enhancer, we next asked whether it might be able to modulate activity of the +19 enhancer. Consistent with our transgenic results a luciferase reporter construct containing the +18 element on its own was inactive in stable transfection experiments using 416B cells (Fig. 2C). In contrast, the + 19 enhancer generated a 15-fold increase in luciferase activity compared to a construct containing the SV40 promoter alone. Inclusion of the +18 element as well as the +19 enhancer reliably produced a further twofold increase in luciferase activity. These results therefore suggested that the +18 region may have a role in boosting the activity of the +19 enhancer, which is consistent with our observation that deletion of the +18 region from a 2.5-kb *Tal1* 3′ enhancer fragment results in a 50% reduction of enhancer activity in stable transfection assays (Supplemental Fig. 1). The activity of the +18 region was further assessed using a beta-geo reporter cassette in stable transfection assays using cotransfection of a puromycin marker for stable selection. Pools of stably transfected clones were analyzed after 3-wk selection in puromycin by flow cytometry. Cells containing the combined +18 and +19 elements showed a fourfold higher mean fluorescence than cells containing the +19 alone (Supplemental Fig. 2), demonstrating that the +18 region increases the transcriptional activity of the +19 enhancer when stably integrated into chromatin.

Single-copy analysis of +18 function during embryonic stem (ES) cell differentiation

The interpretation of stable transfection experiments is complicated by the fact that the reporter constructs are usually integrated as multiple concatenated copies and are also subject to position effects. Furthermore, transformed cell lines are unlikely to provide a normal transcriptional environment. To circumvent these issues, we used a gene targeting method that allows single-copy transgenes to be inserted upstream of the X-linked *Hprt1* gene (Bronson et al. 1996; Misra et al. 2001) in HM-1 ES cells. This male ES cell line carries an inactivating deletion of the promoter and exons 1 and 2 of the *Hprt1* gene. Upon successful

Table 1. Activity of *Tal1* regulatory elements in transgenic embryos

Construct	No. of transgenic embryos	Expression		
		Blood/endothelium	Midbrain	Ectopic
SV/Lac	9	0	0	4
SV/Lac/+18	21	0	0	4
SV/Lac/+19	7	5	0	0
SV/Lac/+18,19	2	2	0	0
0.9E3/Lac	8	0	6	4
0.9E3/Lac/+18	5	0	5	1

F_0 transgenic embryos were generated using the constructs shown on the left and collected at embryonic days 11–12.5 followed by overnight staining with X-gal to determine *lacZ* activity. *lacZ* expression patterns for each of the constructs are shown on the right.

homologous recombination with an exogenous *Hprt1* targeting vector, the functional activity of *Hprt1* is restored, permitting selection of correctly targeted clones in HAT media. Targeting vectors were generated using three different *lacZ* reporter constructs: SV40 minimal promoter alone; the promoter with the +19 enhancer; the promoter with the +19 enhancer and the +18 element (Fig. 3A). Each targeting vector was used to generate ES cell lines carrying a single copy of the appropriate reporter construct targeted to the *Hprt* locus. Activity of the reporter constructs was then assessed during ES cell differentiation.

Several groups have shown that *Tal1* switches on around day 2.5 of ES cell differentiation and that this expression coincides with the onset of hemangioblast specification (Elefanty et al. 1997; Fehling et al. 2003). Prior to plating in differentiation assays, X-gal was used to stain each of the ES cell lines targeted with SV/Lac, SV/Lac/+19, and SV/Lac/+18,19. As would be expected no staining was observed in any of the targeted cell lines (Fig. 3B).

Embryoid bodies (EBs) were then generated from each of the targeted ES cell lines and *lacZ* activity assessed at day 3 and day 8 of differentiation (Fig. 3B). No staining was observed in EBs with the SV40 *lacZ* control vector at either day of differentiation. X-gal staining of EBs carrying the +19 enhancer alone (SV/Lac/+19) showed high levels of *lacZ* expression in day 3 EBs, which markedly decreased by day 8 of differentiation. In striking contrast, EBs containing both +18 and +19 regions (SV/Lac/+18,19) showed high levels of *lacZ* activity observed by X-gal staining at both days 3 and 8 of differentiation. *lacZ* activity in EBs was also quantified using a colorimetric assay. In cells carrying SV/Lac/+19 alone, the mean *lacZ* activity per 1×10^6 cells showed a twofold reduction in *lacZ* activity between days 3 and 8 of differentiation whereas cells containing SV/Lac/+18,19 showed a fourfold increase (Supplemental Fig. 3). These data demonstrate that, during hematopoietic differentiation of ES cells, a single copy of the +18 element boosts activity of an adjacent +19 stem cell enhancer by almost an order of magnitude.

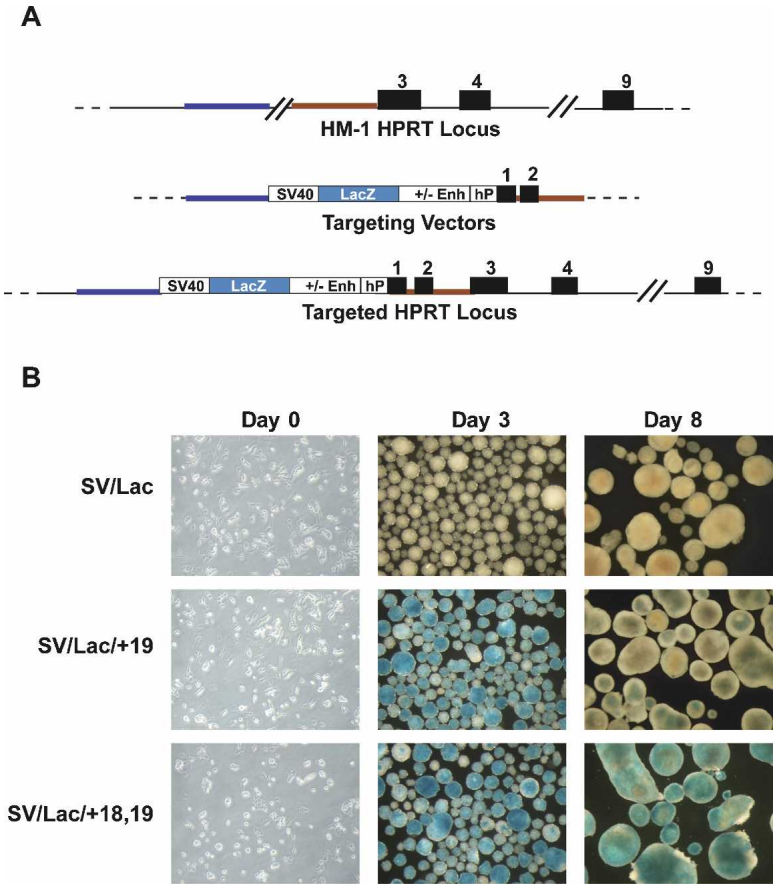

Figure 3. The +18 region can boost the activity of the *Tal1* +19 stem cell enhancer. (*A*) Schematic representation of the *Hprt* locus in the HM-1 ES cell line, which carries a deletion of exons 1 and 2. SV40 *lacZ* reporter constructs with or without the 1.1-kb +18/+19 fragment or the +19 element were cloned into an *Hprt* targeting vector containing the promoter and exon 1 of the human *HPRT* locus and 2.9 kb of the mouse genomic locus including exon 2. The targeted *Hprt* locus contains the SV40 *lacZ* reporter cassettes integrated upstream of the promoter of the fully functional *Hprt* locus. (*B*) ES cells expressing either the SVLac control, SVLac/+19, or SVLac/+18,+19 were plated into differentiation assays and analyzed at days 0, 3, and 8 of differentiation. *lacZ* activity was assessed by staining with X-gal for 2 h.

Functional activity of the +18 region resides in a mammalian interspersed repeat

To investigate the +18 region more closely, we focused on the sequence conservation plot between mouse and human (Fig. 1A), which showed several blocks of sequence identity. To our surprise, these coincided with a stretch of 211 bp masked as repetitive sequence in the UCSC and Ensembl genome browsers. To determine the nature of the repeat, we screened the DNA sequence of the +18 region with the RepeatMasker program (A.F.A. Smit, R. Hubley, and P. Green, 1996–2004. RepeatMasker Open-3.0; http://www.repeatmasker.org), which identified the +18 repeat as a member of the mammalian interspersed repeat (MIR) family (Fig. 4A; Smit and Riggs 1995). In stable transfection assays using 416B cells (Fig. 4B), a 1.1-kb fragment containing both +18 and +19 elements produced a 90-fold increase in luciferase activity relative to a construct containing the SV40 promoter alone. Deletion of 460 bp containing the +18 element halved activity of the construct. A similar effect was observed following deletion of the 200 bp containing the MIR repeat.

In vivo DMS footprinting was performed to search for DNA sequence motifs that might mediate the boosting effect of the +18 region. No footprints were observed in the 400-bp +18 region apart from the protection of three bases (Fig. 4C, open circles) around the first block of sequence conservation (M1, Fig. 4A) situated between the MIR A and B box sequences. We therefore generated a construct whereby we mutated the M1 region to a BglII restriction site and assessed its activity in our 416B assay. Analysis of the mutant construct, SV/Luc/+18,19M1, also showed a decrease in luciferase activity comparable to the SV/

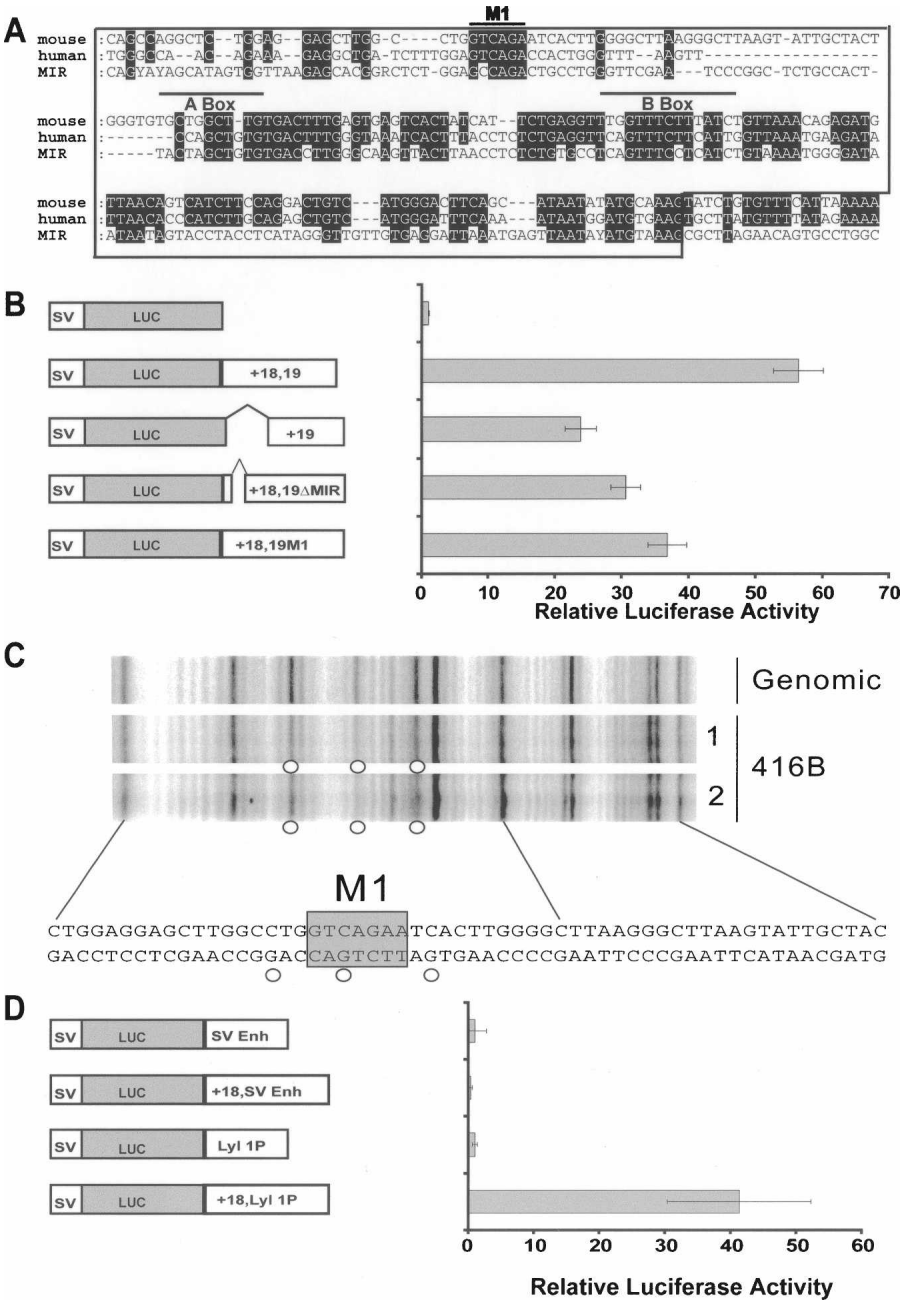

Figure 4. A MIR repeat within the +18 region boosts the activity of the *Tal1* +19 enhancer. (*A*) Mouse/human nucleotide sequence of the +18 region aligned to the MIR consensus sequence, in which the A and B boxes of the MIR tRNA region are underlined. The nucleotide sequence contained within the boxed region denotes the 200 bp of the +18 region that was deleted in the SV/Luc/+18,19ΔMIR construct, and the conserved sequence that was mutated to a BglII restriction site for SV/Luc/+18,19M1 is marked. (*B*) Mutation analysis of the +18 region. Shown on the *left* are the luciferase reporter constructs depicting the deletion in the +18 region and the mutation in the first block of sequence conservation highlighted in *A*. The mutation constructs were analyzed in 416B stable transfection assay and the luciferase values presented as fold increase over the SV40 luciferase control, which was assigned a value of 1. Data are presented as the mean luciferase activity of three biological replicates performed in quadruplicate plus or minus the standard error of the mean. (*C*) In vivo DMS footprinting of the +18 MIR repeat region (antisense strand). Protections (○) were observed over the M1 region of the MIR repeat in two independent 416B samples compared to the naked genomic control. (*D*) Shown on the *left* are the luciferase reporter with constructs, either the heterologous SV40 enhancer or the tissue-specific *Lyl1* promoter proximal enhancer, inserted downstream from the SV40 luciferase reporter cassette with and without the +18 region. The luciferase activities are shown as fold increase over SV40 luciferase with either the SV40 enhancer or the *Lyl1* enhancer, each of which was assigned a value of 1.

Luc/+18,19ΔMIR construct (Fig. 4C). Bioinformatic analysis of the conserved motif M1 using the TRES web server (TRES: comparative promoter sequence analysis; Katti et al. 2000) showed that this conserved motif did not match any of the 222 TRANSFAC consensus binding site positional weight matrices or any of the 5919 experimentally characterized sites present in the TRES database. We also used the recently described software tool STAMP, which is specifically designed to predict the identity or structural class of proteins binding to a given sequence motif (Mahony et al. 2007). Using the JASPAR and TRANSFAC data sets, STAMP also failed to identify any match to the motif M1. Taken together these results indicate that the boosting effect of the +18 region is mediated by sequence derived from a MIR repeat and that much of this activity depends on the conserved sequence that lies within the tRNA promoter-like domain of this repeat element.

To determine whether the boosting activity of the +18 region is nonspecific or restricted to a subset of enhancers, we inserted the 460-bp +18 element upstream of an SV40 enhancer and also a *Lyl1* promoter/enhancer element (*Lyl1*) that we have previously shown to con tain an Ets/Ets/GATA motif related to the *Tal1* +19 enhancer (Chan et al. 2007). Both the SV/luc/SVEnh and SV/luc/Lyl1P constructs are highly active in stable transfection assays in 416B cells compared to the SV40 promoter alone (data not shown). When inserted into these two constructs, the +18 region failed to increase activity of the SV40 enhancer any further but dramatically boosted the activity of the *Lyl1* element (Fig. 4D). These results demonstrate that the +18 element is able to increase the activity of two functionally related hematopoietic/endothelial enhancers (*Tal1* +19 and *Lyl1*), but that it does not act nonspecifically on all enhancers. The question as to whether the activity of other enhancers may be modulated in a similar way to the *Tal1* +18/+19 interaction will require comprehensive analysis using the multiple assays carried out here. However, we have noted that another hematopoietic enhancer (*Fli+12*; Donaldson et al. 2005b) is situated <1 kb downstream from an exapted MER repeat. Moreover, stable transfection assays suggest that this repeat region is also capable of functioning as a positive modulator enhancer activity (see Supplemental Fig. 4).

This led us to examine if the presence of MIR repeats within close proximity to *cis*-regulatory regions might be a more widespread phenomenon. To this end, we considered 2768 lymphoid DNase I-hypersensitive sites that have been mapped as part of the ENCODE project (Sabo et al. 2004). As a control, we performed the same analysis on 1000 sets of 2768 randomly chosen genomic coordinates to determine a normal distribution, allowing us to assign z-scores to any subsequent results. On average, we found 20% of randomly chosen coordinates have MIR repeats within 500 bp. In contrast, when we analyzed the 2768 lymphoid DNase I-hypersensitive sites, we found that 26% have a MIR repeat within 500 bp (*Z*-score 7.65 when compared to random genomic coordinates). MIR repeats had previously been reported to be evenly distributed with respect to the location of coding genes (Medstrand et al. 2002). Given the recent availability of data sets containing genome-wide patterns of histone modifications (Barski et al. 2007), we chose to further investigate the distribution pattern of MIR repeats (see Methods). We found that 36% of H3K4me1 sites, 23% of H3K4me3 sites, and 28% of H3K27me3 sites contain a MIR within 500 bp, suggesting that MIR repeat distribution throughout the genome is not random with the prevalence of MIR repeats next to H3K4me1 sites being particularly striking. Finally, we interrogated a recent genome-wide data set of DNase I-hypersensitive sites in human CD4 T-cells which reported fairly precise regions of hypersensitivity (Boyle et al. 2008). Out of the 95,723 DNase I-hypersensitive sites reported in the paper, 19% overlap with sequence annotated as MIR repeat compared to 12% of 95,723 randomly chosen regions of the same length as the 95,723 DNase I-hypersensitive regions. Taken together, the above analyses therefore indicated that MIR repeats may be enriched near regulatory regions consistent with a more widespread role in *cis*-regulation.

Sequential evolution of the *Tal1* stem cell enhancer

Our observation that the +18 region was derived from a MIR repeat suggested that it evolved from a transposition event sometime during early mammalian evolution. To investigate the timing of this transposition, we obtained genomic sequences of the *Tal1* locus for 10 mammals and also chicken as an outgroup for comparative analysis. Mouse, human, dog, platypus, and chicken sequences were obtained from BAC clones isolated by us previously and sequenced by the Wellcome Trust Sanger Institute (Gottgens et al. 2000, 2002a), while the remaining sequences were downloaded from the Ensembl genome browser. The 11 sequences were aligned using seven different alignment programs (ClustalW, MLagan, MAVID, Dialign, DialignT, TCoffee, MUSCLE) to minimize the likelihood of missing significant alignment blocks.

Our comparative sequence analysis demonstrated that both the +18 and +19 elements were present in all eight eutherian mammals that were analyzed (Fig. 5A; data not shown), but only the +19 could be detected in the more distantly related opossum and platypus. To confirm that our failure to detect homology with the +18 region in platypus and opossum was not a shortcoming of all seven alignment programs, we made use of the fact that the eutherian +18 sequence is recognized as MIR-derived by the RepeatMasker program. No MIR repeats were identified within 10 kb of the platypus +19 sequence, and a MIR-derived sequence 600 bp upstream of the opossum +19 element was inserted in the opposite orientation to the MIR-derived eutherian +18 region. Our results are therefore consistent with a model where the +18 MIR element integrated next to the +19 enhancer at some point between the marsupial branchpoint and the base of extant eutherians (~160 to 120 million years ago), and was subsequently conserved throughout the radiation of placental mammals (Fig. 5B).

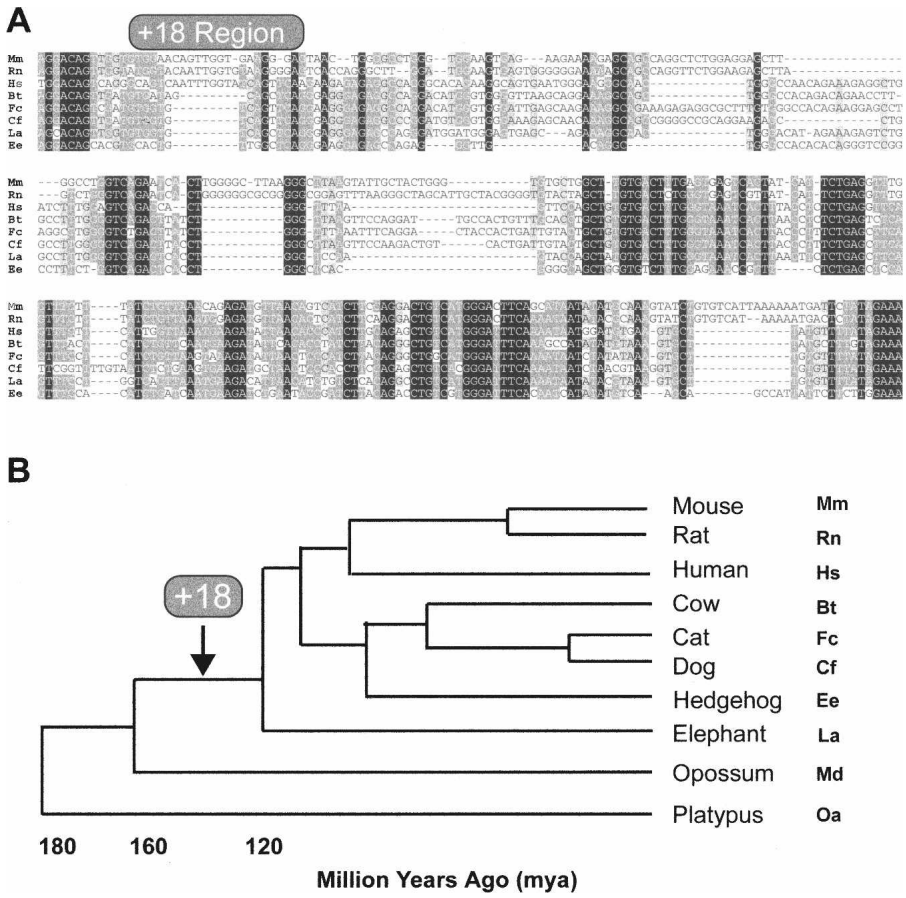

Figure 5. Evolution of the *Tal1* stem cell enhancer region. (*A*) Multiple species sequence alignment of the +18 region detected in eight of the mammals analyzed. Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Hs, *Homo sapiens*; Bt, *Bos taurus*; Fc, *Felis catus*; Cf, *Canis familiaris*; La, *Loxodonta africana*; Ee, *Erinaceus europaeus*. (*B*) Phylogentic tree depicting the divergence of the species listed on the *right* (adapted from Murphy et al. 2007). Time window for the insertion of the *Tal1* +18 region is indicated. Md, *Monodelphis domestica*; Oa, *Ornitorhynchus anatinus*.

Discussion

Transcriptional control is a key mechanism controlling the formation and subsequent behavior of blood stem cells. Many transcription factors such as TAL1, GATA2, ETV6 (also known as TEL), FLI1, or RUNX1 are critical for the specifica-

tion and/or subsequent development of HSCs. It is the combinatorial activity of these (and additional as yet unknown) transcription factors that characterizes the transcriptional environment of HSCs, also termed the "regulatory state" (Davidson 2006). The *Tal1* +19 enhancer represents the best-characterized regulatory element that specifically responds to the HSC regulatory state. Using reporter genes, we have shown that the +19 enhancer directs expression to HSCs in transgenic mice. The *Tal1* enhancer has also been used to generate mouse models for inducible transgene expression, retroviral gene transfer, and Cre-mediated recombination as well as the development of acute myeloid leukemia mouse models (Murphy et al. 2003; Eguchi et al. 2005; Gothert et al. 2005; Koschmieder et al. 2005). At the molecular level, activity of the +19 element critically depends on binding sites for the Ets and GATA families of transcription factors and this combination of transcription factors constitutes a regulatory code employed by several other enhancers (Landry et al. 2005; Pimanda et al. 2006, 2007a,b; Chan et al. 2007; Landry et al. 2008) that display similar expression patterns in transgenic assays.

Much less well understood are the mechanisms that control quantitative aspects of gene expression, even though expression levels of key regulators are critical both to the initial development and maintenance of HSCs as well as their transformation into malignant leukemia cells. For example, reducing the level of key regulators can cause leukemia development, as illustrated by the deletion of an upstream regulatory element of the mouse *PU.1* (also known as *SFPI1*) transcription factor, which results in an 80% reduction of *PU.1* expression without any gross alterations to the overall expression pattern (Rosenbauer et al. 2004). Interestingly, a single nucleotide polymorphism in the corresponding human element reduces enhancer activity and is associated with specific subtypes of human acute myeloid leukemia (Steidl et al. 2007). Several possible mechanisms for this so-called "dosing effect" of gene expression can be proposed: Firstly, within a given regulatory element there may be binding sites that mediate the quantitative effect. One such example seems to be the *PU.1* enhancer where the human SNP appears to affect binding of SATB1, which apparently plays no part in setting expression patterns but instead boosts the overall activity of the enhancer (Steidl et al. 2007). Alternatively, multiple tissue-specific enhancers with overlapping expression patterns may be required to act in an additive fashion, thus ensuring appropriate levels of gene expression as seen for example in the β-globin LCR (for review, see Grosveld 1999). Here, we propose a third and novel mechanism whereby domestication of transposable element-derived DNA sequence functions in a quantitative manner to boost the activity of a tissue-specific enhancer.

The new element (+18 region) is located 1 kb upstream of the +19 *Tal1* enhancer and can boost its activity, suggesting that the *Tal1* stem cell enhancer may encompass a larger gene regulatory domain. From our ES cell differentiation assays it appears that the +18 region may contribute to the function of the *Tal1* stem cell enhancer by maintaining expression of the +19 enhancer, and this may explain the lack of transgene expression in adult mice when the +19 enhancer is used alone (Silberstein et al. 2005). Given that the +18 and +19 regions show the same tissue-specific pattern of DNase I hypersensitivity, their physical proximity, and specificity of enhancement, it is likely that they constitute a regulatory unit within the *Tal1* gene locus. The DNA sequence in the +18 region is derived from an ancient MIR repeat. This family of repeats can be found in all mammalian orders and is thought to have arisen more than 130 million years ago (Jurka et al. 1995; Smit and Riggs 1995). MIR repeats constitute ~1%–2% of our DNA, initially characterized by a 70-bp core sequence (Donehower et al. 1989) and more recently were shown to be part of a longer 260-bp tRNA-derived SINE (Jurka et al. 1995; Smit and Riggs 1995; Gilbert and Labuda 1999).

When the MIR-derived +18 region was assayed in standard classical enhancer assays, it failed to show any activity and therefore may easily have been disregarded when screening for regulatory elements. However, when tested upstream of hematopoietic enhancers, it behaved in a quantitative manner to boost the activity of these elements. We considered the possibility that the function of the +18 region was related to boundary elements or insulators. These latter elements are thought to play key roles in organizing higher order chromatin structure and can serve to protect a given gene locus from repressing influences of neighboring regions of heterochromatin. At the molecular level, insulator/boundary elements are characterized by the presence of ubiquitous DNase I-hypersensitive sites and binding by the multi-zinc finger protein CTCF (Xi et al. 2007). In contrast, the *Tal1* +18 region is characterized by a tissue-specific hypersensitive site and not bound by CTCF (see Supplemental Fig. 5). Moreover, within our single-copy *Hprt* targeting constructs, the +18 region is located between the SV promoter and the *Tal1* +19 enhancer and therefore not in the right position to protect the +19 enhancer from encroaching inhibitory chromatin. We also demonstrate that MIR elements are commonly found in the vicinity of DNase I-hypersensitive sites, raising the possibility of a more widespread regulatory role for this family of repeats. The precise mechanism whereby the +18 element boosts the activity of an associated enhancer remains unclear. Nevertheless, it is clear from the data presented here that we have defined a novel role for transposon-derived DNA sequence and thus added to the growing list of possible functions of repetitive DNA, which until fairly recently was considered to be "junk" DNA.

Regulatory variation is fundamental to evolution, and transposable repeat sequences (which make up at least 50% of the human genome; Kazazian 2004) may make a significant contribution (Britten and Davidson 1971). With the advent of whole genome sequencing projects and the availability of comprehensive EST data sets, it has indeed been possible to verify that up to 25% of genes incorporate transposable elements into their promoters and/or untranslated regions (Jordan et al. 2003; van de Lagemaat et al. 2003). Moreover, in the few cases studied in detail, transposable element-derived DNA was found to play a functional role in promoter activity (Landry et al. 2002; Dunn et al. 2003). Promoters only represent a relatively small proportion of active gene regulatory elements as evidenced by a recent genome-wide survey of DNase I-hypersensitive sites, which demonstrate that only 16%–21% of DNase I-hypersensitive sites were found in promoters or first exons of known genes (Boyle et al. 2008). The vast majority of gene regulatory activity therefore appears to reside in distal regulatory elements. However, very little is known about a possible role for transposable elements in the evolution of distal regulatory elements, a situation most likely explained by the relative ease of identifying promoters compared to distal regulatory elements. Despite this, a few examples of repetitive DNA gaining specific functions as distal regulatory elements have emerged where transposable elements have been shown to act as enhancers (Hambor et al. 1993; Bejerano et al. 2006; Santangelo et al. 2007) and silencers (Donnelly et al. 1999). Indeed, a recent report suggests that many exapted nonexonic

elements may be preferentially involved in distal *cis*-regulation (Lowe et al. 2007)

Given the recent reports that 16% of eutherian-specific conserved noncoding sequence is derived from transposable elements (Mikkelsen et al. 2007) and that a proportion of these exapted noncoding elements are under strong purifying selection (Lowe et al. 2007), it is becoming increasingly clear that repetitive elements have been a driving force in carving the landscape of the human genome. Here we describe a novel function, whereby a repeat-derived sequence that would generally have been discounted in many of the techniques used to delineate gene regulatory pathways has become domesticated and acts to boost the activity of an adjacent tissue-specific enhancer.

Methods

Preparation of DNase I-treated template, chromatin immunoprecipitation, and real-time PCR analysis

DNase I-digested material was prepared from the 416B cell line as previously described (Follows et al. 2006, 2007). ChIP assays were performed as previously described (Landry et al. 2005). Briefly 416B cells were treated with 0.4% formaldehyde and the cross-linked chromatin was retrieved by nuclei isolation and lysis. The chromatin was sonicated to yield an average fragment size of ~300 bp and immunoprecipitated with anti-GATA2 (sc-9008x, Santa Cruz) to recover DNA-bound transcription factors and an anti-acetyl H3K9 antibody (06-599, Upstate Biotechnology) to recover acetylated histones. Enrichment was measured by real-time PCR using Sybr green (Stratagene). The levels of enrichment were normalized to those obtained with a control rabbit antibody and were calculated as a fold increase over that measured at a control region. The *Tal1* +20 region was used as a control for both the enrichment of GATA2 and acetyl H3K9. The following pairs of primers were used: 17PW1f: CCGAG-GATTCTGATCACTCC; 17PW1r: GGCACTTTTCTTCCCTTGTG; 18PW1f: AGGAAGG ACAGT-TGGTGTGG; 18PW1R: CTCAAAGTCACAAGCCAGCA; 19PW1f: GCAAGGTTGGGAGGTAT-CTG; 19PW1R: GAGCTG GAAGGCAAAGTGAG; 20PW1f: AAGAATGGGATGTGCTTTGG; 20PW1R: GCAGCTGTTCTCCACTACGG.

Restriction endonuclease accessibility assay and transgenic mouse analysis

416B cells were maintained as described previously (Bockamp et al. 1997). The restriction endonuclease assay was performed as described (Gottgens et al. 2001) using PalI digestion and a 210-bp HindIII/StuI fragment ~2 kb 3′ of exon 6 as a probe. Transgenic *lacZ* reporter constructs were driven by the 3.8-kb 0.9E3 fragment of *Tal1* (Sanchez et al. 1999; Sinclair et al. 1999) or the SV40 minimal promoter. DNA was linearized using BamHI and SalI, and F_0 transgenic embryos were generated by pronuclear injection as described (Sanchez et al. 1999). Embryos were harvested at mid-gestation (embryonic day 11–12.5) and analyzed as described (Sanchez et al. 1999).

Reporter constructs and transfection assays

Reporter constructs were generated by cloning the following fragments of the *Tal1* locus (numbering refers to the *Tal1* locus database entry AJ131017) downstream from the luciferase gene in pGL2 (Promega) or downstream from an SV40 minimal promoter containing a *lacZ* reporter cassette. The +18/+19 regions of *Tal1* were cloned as a 1.1-kb NcoI/HindIII fragment from 78,359 to 79,480 and an SspI site in the middle of this region was used to generate the shorter 480-bp NcoI/SspI fragment from 78,359 to 78,840 and the 640-bp SspI/HindIII fragment from 78,840 to 79,480 for the +18 and +19 constructs, respectively. The mutation constructs were made as previously described (Gottgens et al. 2002b) and verified by sequencing. Stable transfections and luciferase assays were performed as described (Gottgens et al. 1997). Data are presented as a representative plot of three biological replicates performed in quadruplicate plus/minus standard deviation unless otherwise stated.

ES cell culture and targeting

The HM-1 ES cell line was a kind gift from David Melton (Magin et al. 1992). ES cells were cultured on mitotically inactivated primary embryonic fibroblasts in standard ES cell media supplemented with LIF. For the introduction of transgenes at the *Hprt* locus ~1×10^8 ES cells were electroporated with 100 µg of Pme I linearized DNA (0.8Kv,3uF Gene Pulser, Bio-Rad) followed by selection in HAT media (Sigma). HAT-resistant colonies were picked after 7–10-d growth in HAT-containing media. Targeted ES cells were differentiated in 90-mm petri dishes in IMDM supplemented with 15% FCS, 2 mM L-glutamine, 300 µg/mL transferrin, 4×10^{-4} M MTG, 50 µg/mL ascorbic acid, and 5% PFHM-II. Harvested EBs were either fixed and stained for 2 h with X-gal as for the transgenic embryos or disrupted with trypsin for 3 min, and 1×10^6 cells lysed and assayed for *lacZ* activity using the synthetic β-galactosidase substrate ONPG as described (Bockamp et al. 1995).

In vivo DMS footprinting

Footprinting was performed as previously described (Follows et al. 2003). Briefly cells or genomic DNA were incubated at room temperature in 0.2% DMS solution in PBS for 5 min. The reaction was stopped with multiple washes with ice-cold PBS. Following cell lysis and DNA extraction, DNA was cleaved with 0.1 M piperidine at 90°C for 10 min and analyzed by ligation-mediated PCR. PCR products were labeled by primer extension using ^{32}P-labeled nested primers and analyzed on 6% denaturing polyacrylamide gels. Linker and primer sequences are available on request. Experiments were performed on three separate occasions with material from two independent DMS-treated cell and genomic DNA preparations.

Comparison of MIR repeats with genomic annotation

Five sets of genome annotation corresponding to the human genome (NCBI v35/hg17/May 2004) were obtained for this study, including a genome-wide set of coordinates for MIR family repeats to be compared against four other data sets, namely: two collections of DNase I-hypersensitive sites, histone methylation sites, and random genomic regions. All MIR family repeats were obtained using the UCSC genome table browser (http://genome.ucsc.edu). Genome coordinates were output in the BED format, selecting the group: "Variation and Repeats", track: "RepeatMasker", and table: "rmsk" (n = 591,703). Two sets of DNase I hypersensitivity site annotation were extracted using the table browser. The first ENCODE-based set, by Sabo et al. (2004), was retrieved by selecting the group: "ENCODE Chromosome, Chromatin and DNA Structure", track: "UW Dnase-array", and table: "DNase I HSs" (n = 2768). The second genome-wide set, by Boyle et al. (2008), was retrieved by selecting the group: "Expression and Regulation", track: "Duke DNase Sites", and table:

"dukeDnaseCd4Sites" (n = 95,723). Histone methylation sites were extracted from the supplementary Web site relating to the work of Barski et al. (2007). "Summary BED files" for H3K4me1, H3K4me3, and H3K27me3 modifications were downloaded. A Perl script was used to convert the files from "variable step" to BED format and extract all regions containing a minimum number of tags, resulting in the most significant ~2000 regions per histone modification. The minimum numbers of tags for H3K4me1, H3K4me3, and H3K27me3 modifications represented in the extracted data where 42, 87, and 9 respectively. The genome coordinates were converted from human version hg18 to hg17 using the "liftOver" conversion function of Galaxy (Giardine et al. 2005; http://g2.trac.bx.psu.edu/).

To allow the calculation of Z-scores (to evaluate the chance of randomly finding an overlap between the ENCODE DNase I-hypersensitive sites and MIR repeats), 1000 sets of 2768 random genome coordinates, each of 250 nucleotides in length, were generated using a Perl script. A second Perl script was used to calculate the percentage overlap of each random genome coordinate set (extended by 500 flanking nucleotides each side, creating a set of extended coordinates) with the MIR repeat coordinates. The mean and standard deviation of the percentage overlaps for the 1000 comparisons were then calculated.

The comparison of the MIR repeat coordinates to the other data sets was performed using Galaxy. The coordinates of the two collections of DNase I-hypersensitive sites and histone methylation sites were extended by 500 flanking nucleotides each side, creating two sets of extended coordinates. The "intersection" (by at least one nucleotide) between the extended coordinates and the MIR family coordinates was determined.

Acknowledgments

This work was supported by the Cambridge MIT institute, Medical Research Council (MRC), Leukemia Research Fund (LRF), Biotechnology and Biological Sciences Research Fund (BBSRC), and the Wellcome Trust. We thank David Melton (University of Edinburgh) for the HM-1 cell line and Stephen Duncan (Medical College of Wisconsin) for the *Hprt* targeting vector. We thank Mike Chapman, Darren Grafham, and Team 47 at the Wellcome Trust Sanger Institute for the platypus sequences, and Sandie Piltz and Michelle Hammett for generating the transgenic embryos. We also thank Josette-Renée Landry and John Pimanda for the critical reading of this manuscript.

References

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823–837.

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441:** 87–90.

Bockamp, E.O., McLaughlin, F., Murrell, A.M., Gottgens, B., Robb, L., Begley, C.G., and Green, A.R. 1995. Lineage-restricted regulation of the murine SCL/TAL-1 promoter. *Blood* **86:** 1502–1514.

Bockamp, E.O., McLaughlin, F., Gottgens, B., Murrell, A.M., Elefanty, A.G., and Green, A.R. 1997. Distinct mechanisms direct *SCL/tal-1* expression in erythroid cells and CD34 positive primitive myeloid cells. *J. Biol. Chem.* **272:** 8781–8790.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132:** 311–322.

Britten, R.J. and Davidson, E.H. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46:** 111–138.

Bronson, S.K., Plaehn, E.G., Kluckman, K.D., Hagaman, J.R., Maeda, N., and Smithies, O. 1996. Single-copy transgenic mice with chosen-site integration. *Proc. Natl. Acad. Sci.* **93:** 9067–9072.

Chan, W.Y., Follows, G.A., Lacaud, G., Pimanda, J.E., Landry, J.R., Kinston, S., Knezevic, K., Piltz, S., Donaldson, I.J., Gambardella, L., et al. 2007. The paralogous hematopoietic regulators *Lyl1* and *Scl* are coregulated by Ets and GATA factors, but *Lyl1* cannot rescue the early $Scl^{-/-}$ phenotype. *Blood* **109:** 1908–1916.

Davidson, E.H. 2006. *The regulatory genome: Gene regulatory networks in development and evolution*. Academic Press, San Diego.

Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. 2005. Transcriptional regulation of the SCL locus: Identification of an enhancer that targets the primitive erythroid lineage in vivo. *Mol. Cell. Biol.* **25:** 5215–5225.

Dexter, T.M., Allen, T.D., Scott, D., and Teich, N.M. 1979. Isolation and characterisation of a bipotential haematopoietic cell line. *Nature* **277:** 471–474.

Donaldson, I.J. and Gottgens, B. 2006. TFBScluster web server for the identification of mammalian composite regulatory elements. *Nucleic Acids Res.* **34:** W524–W528.

Donaldson, I.J., Chapman, M., and Gottgens, B. 2005a. TFBScluster: A resource for the characterization of transcriptional regulatory networks. *Bioinformatics* **21:** 3058–3059.

Donaldson, I.J., Chapman, M., Kinston, S., Landry, J.R., Knezevic, K., Piltz, S., Buckley, N., Green, A.R., and Gottgens, B. 2005b. Genome-wide identification of *cis*-regulatory sequences controlling blood and endothelial development. *Hum. Mol. Genet.* **14:** 595–601.

Donehower, L.A., Slagle, B.L., Wilde, M., Darlington, G., and Butel, J.S. 1989. Identification of a conserved sequence in the non-coding regions of many human genes. *Nucleic Acids Res.* **17:** 699–710.

Donnelly, S.R., Hawkins, T.E., and Moss, S.E. 1999. A conserved nuclear element with a role in mammalian gene regulation. *Hum. Mol. Genet.* **8:** 1723–1728.

Dunn, C.A., Medstrand, P., and Mager, D.L. 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human β1,3-galactosyltransferase 5 in the colon. *Proc. Natl. Acad. Sci.* **100:** 12841–12846.

Eguchi, M., Eguchi-Ishimae, M., Green, A., Enver, T., and Greaves, M. 2005. Directing oncogenic fusion genes into stem cells via an SCL enhancer. *Proc. Natl. Acad. Sci.* **102:** 1133–1138.

Elefanty, A.G., Robb, L., Birner, R., and Begley, C.G. 1997. Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells. *Blood* **90:** 1435–1447.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. 2003. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130:** 4217–4227.

Follows, G.A., Tagoh, H., Lefevre, P., Hodge, D., Morgan, G.J., and Bonifer, C. 2003. Epigenetic consequences of AML1-ETO action at the human c-FMS locus. *EMBO J.* **22:** 2798–2809.

Follows, G.A., Dhami, P., Gottgens, B., Bruce, A.W., Campbell, P.J., Dillon, S.C., Smith, A.M., Koch, C., Donaldson, I.J., Scott, M.A., et al. 2006. Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. *Genome Res.* **16:** 1310–1319.

Follows, G.A., Janes, M.E., Vallier, L., Green, A.R., and Gottgens, B. 2007. Real-time PCR mapping of DNaseI-hypersensitive sites using a novel ligation-mediated amplification technique. *Nucleic Acids Res.* **35:** e56. doi: 10.1093/nar/gkm108.

Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D., and Jurka, J. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* **17:** 992–1004.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **15:** 1451–1455.

Gilbert, N. and Labuda, D. 1999. CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl. Acad. Sci.* **96:** 2869–2874.

Gothert, J.R., Gustin, S.E., Hall, M.A., Green, A.R., Gottgens, B., Izon, D.J., and Begley, C.G. 2005. In vivo fate-tracing studies using the Scl stem cell enhancer: Embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. *Blood* **105:** 2724–2732.

Gottgens, B., McLaughlin, F., Bockamp, E.O., Fordham, J.L., Begley, C.G., Kosmopoulos, K., Elefanty, A.G., and Green, A.R. 1997. Transcription of the SCL gene in erythroid and CD34 positive

primitive myeloid cells is controlled by a complex network of lineage-restricted chromatin-dependent and chromatin-independent regulatory elements. *Oncogene* **15:** 2419–2428.
Gottgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M., et al. 2000. Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat. Biotechnol.* **18:** 181–186.
Gottgens, B., Gilbert, J.G., Barton, L.M., Grafham, D., Rogers, J., Bentley, D.R., and Green, A.R. 2001. Long-range comparison of human and mouse SCL loci: Localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. *Genome Res.* **11:** 87–97.
Gottgens, B., Barton, L.M., Chapman, M.A., Sinclair, A.M., Knudsen, B., Grafham, D., Gilbert, J.G., Rogers, J., Bentley, D.R., and Green, A.R. 2002a. Transcriptional regulation of the stem cell leukemia gene (*SCL*)—Comparative analysis of five vertebrate SCL loci. *Genome Res.* **12:** 749–759.
Gottgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E.C., Stanley, M., Sanchez, M.J., Ciau-Uitz, A., Patient, R., and Green, A.R. 2002b. Establishing the transcriptional programme for blood: The SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J.* **21:** 3039–3050.
Gottgens, B., Broccardo, C., Sanchez, M.J., Deveaux, S., Murphy, G., Gothert, J.R., Kotsopoulou, E., Kinston, S., Delaney, L., Piltz, S., et al. 2004. The *scl* +18/19 stem cell enhancer is not required for hematopoiesis: Identification of a 5′ bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. *Mol. Cell. Biol.* **24:** 1870–1883.
Grosveld, F. 1999. Activation by locus control regions? *Curr. Opin. Genet. Dev.* **9:** 152–157.
Hambor, J.E., Mennone, J., Coon, M.E., Hanke, J.H., and Kavathas, P. 1993. Identification and characterization of an *Alu*-containing, T-cell-specific enhancer located in the last intron of the human CD8α gene. *Mol. Cell. Biol.* **13:** 7056–7070.
Jordan, I.K., Rogozin, I.B., Glazko, G.V., and Koonin, E.V. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19:** 68–72.
Jurka, J., Zietkiewicz, E., and Labuda, D. 1995. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucleic Acids Res.* **23:** 170–175.
Katti, M.V., Sakharkar, M.K., Ranjekar, P.K., and Gupta, V.S. 2000. TRES: Comparative promoter sequence analysis. *Bioinformatics* **16:** 739–740.
Kazazian Jr., H.H. 2004. Mobile elements: Drivers of genome evolution. *Science* **303:** 1626–1632.
Koschmieder, S., Gottgens, B., Zhang, P., Iwasaki-Arai, J., Akashi, K., Kutok, J.L., Dayaram, T., Geary, K., Green, A.R., Tenen, D.G., et al. 2005. Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis. *Blood* **105:** 324–334.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
Landry, J.R., Rouhi, A., Medstrand, P., and Mager, D.L. 2002. The Opitz syndrome gene *Mid1* is transcribed from a human endogenous retroviral promoter. *Mol. Biol. Evol.* **19:** 1934–1942.
Landry, J.R., Kinston, S., Knezevic, K., Donaldson, I.J., Green, A.R., and Gottgens, B. 2005. Fli1, Elf1, and Ets1 regulate the proximal promoter of the *LMO2* gene in endothelial cells. *Blood* **106:** 2680–2687.
Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K., et al. 2008. *Runx* genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. *Blood* **111:** 3005–3014.
Levine, M. and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424:** 147–151.
Lowe, C.B., Bejerano, G., and Haussler, D. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl. Acad. Sci.* **104:** 8005–8010.
Magin, T.M., McWhir, J., and Melton, D.W. 1992. A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Res.* **20:** 3795–3796.
Mahony, S., Auron, P.E., and Benos, P.V. 2007. DNA familial binding profiles made easy: Comparison of various motif alignment and clustering strategies. *PLoS Comput. Biol.* **3:** e61. doi: 10.1371/journal.pcbi.003006.
Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res.* **12:** 1483–1495.
Medstrand, P., van de Lagemaat, L.N., Dunn, C.A., Landry, J.R., Svenback, D., and Mager, D.L. 2005. Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet. Genome Res.* **110:** 342–352.
Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447:** 167–177.
Misra, R.P., Bronson, S.K., Xiao, Q., Garrison, W., Li, J., Zhao, R., and Duncan, S.A. 2001. Generation of single-copy transgenic mouse embryos directly from ES cells by tetraploid embryo complementation. *BMC Biotechnol.* **1:** 12. doi: 10.1186/1472-6750-1-12.
Murphy, G.J., Gottgens, B., Vegiopoulos, A., Sanchez, M.J., Leavitt, A.D., Watson, S.P., Green, A.R., and Frampton, J. 2003. Manipulation of mouse hematopoietic progenitors by specific retroviral infection. *J. Biol. Chem.* **278:** 43556–43563.
Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., and Miller, W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* **17:** 413–421.
Ogilvy, S., Ferreira, R., Piltz, S.G., Bowen, J.M., Gottgens, B., and Green, A.R. 2007. The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. *Mol. Cell. Biol.* **27:** 7206–7219.
Pimanda, J.E., Chan, W.Y., Donaldson, I.J., Bowen, M., Green, A.R., and Gottgens, B. 2006. Endoglin expression in the endothelium is regulated by Fli-1, Erg, and Elf-1 acting on the promoter and a –8-kb enhancer. *Blood* **107:** 4737–4745.
Pimanda, J.E., Donaldson, I.J., de Bruijn, M.F., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J.R., Green, A.R., Tannahill, D., et al. 2007a. The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc. Natl. Acad. Sci.* **104:** 840–845.
Pimanda, J.E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W.Y., Wilson, N.K., Landry, J.R., Wood, A.D., Kolb-Kokocinski, A., Green, A.R., et al. 2007b. Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl. Acad. Sci.* **104:** 17692–17697.
Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. 1996. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86:** 47–57.
Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. 1996. The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J.* **15:** 4123–4129.
Rosenbauer, F., Wagner, K., Kutok, J.L., Iwasaki, H., Le Beau, M.M., Okuno, Y., Akashi, K., Fiering, S., and Tenen, D.G. 2004. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* **36:** 624–630.
Sabo, P.J., Humbert, R., Hawrylycz, M., Wallace, J.C., Dorschner, M.O., McArthur, M., and Stamatoyannopoulos, J.A. 2004. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl. Acad. Sci.* **101:** 4537–4542.
Sanchez, M., Gottgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. 1999. An SCL 3′ enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. *Development* **126:** 3891–3904.
Sanchez, M.J., Bockamp, E.O., Miller, J., Gambardella, L., and Green, A.R. 2001. Selective rescue of early haematopoietic progenitors in $Scl^{-/-}$ mice by expressing *Scl* under the control of a stem cell enhancer. *Development* **128:** 4815–4827.
Santangelo, A.M., de Souza, F.S., Franchini, L.F., Bumaschny, V.F., Low, M.J., and 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.
Silberstein, L., Sanchez, M.J., Socolovsky, M., Liu, Y., Hoffman, G., Kinston, S., Piltz, S., Bowen, M., Gambardella, L., Green, A.R., et al. 2005. Transgenic analysis of the stem cell leukemia +19 stem cell enhancer in adult and embryonic hematopoietic and endothelial cells. *Stem Cells* **23:** 1378–1388.
Sinclair, A.M., Gottgens, B., Barton, L.M., Stanley, M.L., Pardanaud, L., Klaine, M., Gering, M., Bahn, S., Sanchez, M., Bench, A.J., et al. 1999. Distinct 5′ SCL enhancers direct transcription to developing brain, spinal cord, and endothelium: Neural expression is mediated by GATA factor binding sites. *Dev. Biol.* **209:** 128–142.
Smit, A.F. and Riggs, A.D. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23:** 98–102.
Steidl, U., Steidl, C., Ebralidze, A., Chapuy, B., Han, H.J., Will, B., Rosenbauer, F., Becker, A., Wagner, K., Koschmieder, S., et al. 2007. A distal single nucleotide polymorphism alters long-range regulation

of the *PU.1* gene in acute myeloid leukemia. *J. Clin. Invest.* **117:** 2611–2620.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **19:** 530–536.
Xi, H., Shulha, H.P., Lin, J.M., Vales, T.R., Fu, Y., Bodine, D.M., McKay, R.D., Chenoweth, J.G., Tesar, P.J., Furey, T.S., et al. 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3:** e136. doi: 10.1371/journal.pgen.0030136.
Xie, X., Kamal, M., and Lander, E.S. 2006. A family of conserved noncoding elements derived from an ancient transposable element. *Proc. Natl. Acad. Sci.* **103:** 11659–11664.

Received February 4, 2008; accepted in revised form June 5, 2008.