Published online 28 January 2017

Nucleic Acids Research, 2017, Vol. 45, No. 8 ***4479–4492***
doi: 10.1093/nar/gkx055

Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function

Tianxiang Hu[1], Wenhu Pi[1], Xingguo Zhu[1], Miao Yu[2], Hongseok Ha[1], Huidong Shi[2], Jeong-Hyeon Choi[3] and Dorothy Tuan[1,*]

[1]Department of Biochemistry and Molecular Biology, Augusta University, Augusta, GA 30912, USA, [2]Cancer Center, Augusta University, Augusta, GA 30912, USA and [3]Department of Biostatics and Epidemiology, Augusta University, Augusta, GA 30912, USA

Received November 20, 2016; Revised January 02, 2017; Editorial Decision January 19, 2017; Accepted January 23, 2017

ABSTRACT

LTR retrotransposons are repetitive DNA elements comprising ∼10% of the human genome. However, LTR sequences are disproportionately present in human long, non-coding RNAs (lncRNAs). Whether and how the LTR lncRNAs serve biological functions are largely unknown. Here we show that in primary human erythroblasts, lncRNAs transcribed from the LTR retrotransposons of ERV-9 human endogenous retrovirus activated transcription of key erythroid genes and modulated *ex vivo* erythropoiesis. To dissect the functional mechanism of ERV-9 lncRNAs, we performed genome-wide RNA and ChIRP analyses before and after global knockdown or locus-specific deletion of ERV-9 lncRNAs in human erythroblasts carrying ∼4000 copies of the ERV-9 LTRs and in transgenic mouse erythroblasts carrying a single copy of the primate-specific ERV-9 LTR in the 100 kb human β-globin gene locus. We found that ERV-9 lncRNAs acted in cis to stabilize assembly of the ERV-9 LTR enhancer complex and facilitate long-range LTR enhancer function in activating transcription of downstream, cis-linked globin genes. Our findings suggested that LTR lncRNAs transcribed from many of the 4000 copies of ERV-9 LTR retrotransposons acted by a similar cis mechanism to modulate LTR enhancer function in activating transcription of downstream genes critical to cellular processes including erythropoiesis.

INTRODUCTION

Retrotransposons including Alus, L1s and solitary long terminal repeats (LTRs) of human endogenous retroviruses (HERVs) comprise over 40% of the human genome (1–3). Retrotransposons are transcriptional regulators capable of initiating RNA synthesis, as the Alus contain RNA polymerase III promoters, the L1s RNA polymerase II (pol II) promoters and the solitary LTRs both pol II promoters and enhancers (1). Indeed, a recent survey of long, intergenic, noncoding lncRNAs (aka lincRNAs) finds that over 80% of human lncRNAs contain retrotransposon sequences (4). Although the solitary LTRs comprise ∼10% of the human genome and thus a minor portion of the retrotransposons, LTR sequences are disproportionally represented in the lncRNAs (4), likely due to presence of both promoters and enhancers in the LTRs (1,5). Whether and how the LTR lncRNAs perform biological functions are largely unknown.

The solitary ERV-9 LTR (aka LTR12 or pTR5) retrotransposons, derived from the ERV-9 HERV, are present in the human genome at ∼4000 copies distributed across all human chromosomes (2,6). The ERV-9 LTR contains the U3, R and U5 regions characteristic of retroviral LTRs but no retroviral genes (7). The U3 regions of ERV-9 LTRs span the retroviral enhancer and promoter, which contain high densities of CCAAT and GATA motifs that bind respectively NF-Y and GATA-1 and -2 to assemble an LTR enhancer-pol II complex (8). The LTR enhancer complex is active in erythroid progenitor cells (9); it initiates synthesis of ERV-9 lncRNAs from the 5′ end of the R region through the U5 region into the downstream genomic DNA (10–12) and activates transcription of downstream cis-linked genes by a long-range tracking and transcription (T&T) mecha-

*To whom correspondence should be addressed. Tel: +1 706 721 0272; Fax: +1 706 721 6608; Email: dtuanlo@augusta.edu
Present addresses:
Hongseok Ha, Department of Genetics, Human Genetic Institute of New Jersey, The State University of New Jersey, Piscataway, NJ 08854, USA.
Jeong-Hyeon Choi, Department of Applied Research, Marine Biodiversity Institute of Korea, Chungnam, South Korea

nism (13–14). The function, if any, of the ERV-9 lncRNAs was unknown.

To investigate the functional significance of ERV-9 lncRNAs, we depleted them globally by shRNAs in human erythroblasts cultured from peripheral blood $CD34^+$ stem/progenitor cells. We found that global depletion of ERV-9 lncRNAs inhibited *ex vivo* erythropoiesis. ERV-9 lncRNAs could act in cis to regulate transcription of cis-linked, key erythroid genes and/or they could act in trans by diffusing from their LTR templates to regulate transcription of target genes located on separate chromosomes. To dissect the cis versus trans mechanisms in the complex transcription network of erythropoiesis, we took advantage of a simpler BAC Tg mouse model (14), carrying a single copy of the primate-specific ERV-9 LTR (12,15) in the human β-globin gene locus (see Figure 1A). In both Tg mice and human erythroblasts, depletion of ERV-9 lncRNAs destabilized assembly of the ERV-9 LTR enhancer-pol II complex and decreased the looping/interaction frequencies of the LTR enhancer complex with the downstream locus, thereby suppressing transcription of the entire globin gene locus. Consistent with this cis mechanism of ERV-9 lncRNA function, locus-specific deletion of the ERV-9 LTR/lncRNAs from the β-globin gene locus by CRISPR-cas9 in erythroleukemia K562 cell line suppressed transcription of downstream, cis-linked globin genes. Our findings indicated that ERV-9 lncRNAs transcribed from many of the 4000 copies of the ERV-9 LTR retrotransposons could act in cis to facilitate long-range LTR enhancer function in activating transcription of downstream, key erythroid genes to coordinate the transcription network of erythropoiesis.

MATERIALS AND METHODS

For isolation and culture of human CD34+ cells and mouse fetal liver (FL) erythroblasts, construction of lentiviral vectors expressing shRNAs and transduction to deplete ERV-9 lncRNAs, analyses of total cellular RNAs by qRT-PCR, RNA-seq and microarrays, bioinformatics analysis of RNA-seq data, ChIP and 3C assays, see Supplementary Methods.

Homozygous deletion of locus-specific ERV-9 LTR in K562 cells by CRISPR-Cas9

For locus-specific deletion of the ERV-9 LTR, thus synthesis of locus-specific ERV-9 lncRNAs, from the β-globin gene locus in K562 cells by CRISPR-Cas9 (16), we designed two guide sgRNAs of 20 nucleotides (nt) targeting the 5′ and 3′ flanking DNAs of the ERV-9 LTR (sgRNA1: 5′TGGTAGTCTTTTTGCACTCCAGG3′; sgRNA2: 5′TAGAAGGGCACTCTGCCTTA-GGG3′), using an online CRISPR Design Tool (crispr.mit.edu). The DNA templates for the sgRNAs were separately inserted into the BbsI site immediately 5′ of the DNA encoding Cas9 binding, scaffold RNA in pSpCas9-2A-Puro (Addgene plasmid #62987). To delete the ERV-9 LTR by homology-directed repair (HDR), we synthesized by polymerase chain reaction (PCR) a donor DNA template of 1.2 kb that spanning DNAs immediately 5′ and 3′ of the ERV-9 LTR, without the ERV-9 LTR, in place of which was a single LoxP site of 34 bp (14). The pSpCas9-2A-Puro plasmids expressing respectively the 5′ and 3′ sgRNAs (0.5 μg each) were mixed with 100 ng of the repair DNA and transfected into 2×10^5 K562 cells in 12-well plates using Lipofectamine 2000. The transfected K562 cells were selected and expanded in culture medium containing puromycin (1.5 μg/ml) for 4 days. For homozygous deletion of the ERV-9 LTR, transfection of sgRNAs and donor template was repeated. Clonal lines were grown from single cells and harvested after ∼3 weeks' of culture for DNA analysis by PCR and Sanger sequencing to determine LTR deletion and RNA analysis by qRT-PCR and microarrays. Potential off-target sites (OTS's) for the sgRNAs were identified using CRISPR Design Tool. DNA fragments spanning OTS's were amplified by PCR from the DNA of the clonal lines and sequenced (McLab). No DNA deletions were detected at the top-ranked five potential OTS's (Hu *et al.*, in preparation).

ChIRP

The ChIRP assays utilizing a split probes strategy were carried out as described (17). Briefly, 4×10^7 Day 8 Tg mouse FL, 10^8 Day 5 Tg mouse spleen and 10^7 Day 13 human erythroblasts were cross-linked with 1% glutataldehyde; chromatin was extracted and fragmented by sonication (Misonic 3000) to 300–500 bp. Two separate pools of the sonicated chromatin were hybridized respectively to two separate pools of biotinylated oligonucleotide probes, the odd and even probes (synthesized by Integrated DNA Technologies). The chromatin containing ERV-9 lncRNA that hybridized to the probes were pulled down using streptavidin-conjugated magnetic beads. The pulled-down chromatin was eluted from the beads and used for RNA and DNA isolation and subsequent analyses by qRT-PCR, ChIRP-qPCR and ChIRP-seq. Sequences of Odd and Even probes: 1. AAAGCTTCCA-CAGT GTGGAA; 3. CAGTGTAGACCCAAACAGTG; 5. TG GCTTCAAGAGTGAAGCTG; 7. TGAACGTTCCAG CTCT-TAAG; 9. TCGAGTTCTTCCTTCTGATG; 11. TAAAGGCTGCGTGTGTGGAG; 13. TTCAAGAATG AAGCCGGGGA; 2. CAGCAGCAAGATTTATTGCA; 4. GTGAGCGTTACAGCTCATAA; 6. TCTGGAGTTG TTCGTTCCTC; 8. TCAGGAGTGAAGCTGCAGAC; 10. GTTCGTTCTGATGTTTGGAT; 12. TCGTGGTG AGTGTTACAGTT; 14. AATTGGTGGGTTCTTGGT TT. For ChIRP-seq, chromatin DNAs pulled down by separate pools of odd and even probes were constructed into two separate DNA libraries. The four DNA libraries, two each from Tg mouse and human erythroblasts that passed the quality checks were sequenced with Illumina HiSeq platform by Integrated Genomics Core on campus, with 100 base paired-end reads, 50 million reads per sample. For bioinformatics analysis of ChIRP-seq data, see Supplementary Methods.

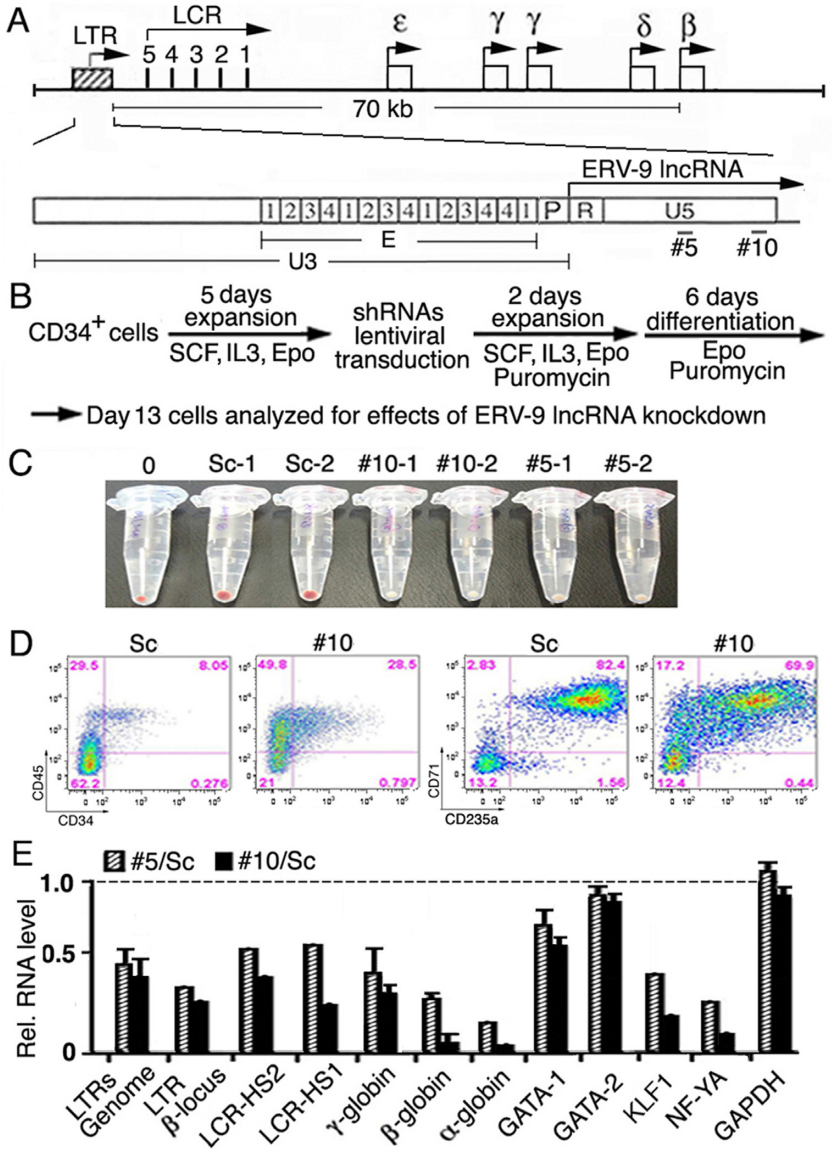

Figure 1. Depletion of ERV-9 lncRNAs suppressed transcription of key erythroid genes and impaired ex vivo erythropoiesis. (**A**) Top: map of the human β-globin gene locus. Hatched box: ERV-9 LTR; angled arrows: direction of transcription of ERV-9 LTR lncRNAs, LCR (locus control region) and globin genes. E: LTR enhancer; P: LTR promoter. Boxes 1, 2, 3 and 4: enhancer subunits 1–4 (see Supplementary Figure S1). (**B**) Culture scheme of D13 erythroblasts from human CD34+ stem/progenitor cells. (**C**) Images of Day 13 erythroblasts transduced in duplicates with scrambled, #10 and #5 shRNAs; 0, non-transduced D13 erythroblasts. (**D**) Day 13 cells transduced by scrambled and #10 shRNAs analyzed by FACS with early hematopoietic stem cell markers CD34 and CD45 and erythroid markers, transferrin receptor (CD71) and glycophorin A (CD235a). (**E**) qRT-PCR. LTRs (genome): genome-wide ERV-9 LTR RNAs; LTR (β-locus): β-globin ERV-9 lncRNAs. #5/Sc and #10/Sc: RNA levels of genes in D13 cells transduced with #5 or #10 shRNAs relative to those of the genes in cells transduced with scrambled shRNA, which were set at 1, as marked by the dotted horizontal line. Relative RNA levels were mean ± SEM of three RNA samples isolated from three independently transduced Day 13 cells.

RESULTS

Depletion of ERV-9 lncRNAs in human erythroblasts suppressed transcription of key erythroid genes and inhibited *ex-vivo* erythropoiesis

To investigate the functional significance of ERV-9 lncRNAs, we utilized RNA interference (18) to knock down ERV-9 lncRNAs. Among 11 recombinant lentiviruses expressing shRNAs targeting different R-U5 sub-regions in ERV-9 lncRNA, #5 and #10 shRNAs (see Supplementary Figure S1 for locations and sequences) were most effective in depleting ERV-9 lncRNA in pilot experiments (not shown). Therefore, these two shRNAs and a scrambled control shRNA were separately transduced into human erythroblasts (Figure 1B). This knockdown (KD) strategy of utilizing multiple shRNAs (19) was employed to distinguish the specific KD effect of ERV-9 lncRNAs generated by perfect alignment of #5 and #10 shRNAs with ERV-9 lncRNAs from the off-target effects potentially generated by partial alignment of the shRNAs with heterologous RNAs, since only the specific KD of ERV-9 lncRNA—the common target of #5 and #10 shRNAs—would generate the same

genotype and phenotype in cells transduced separately by the two shRNAs.

Indeed, KD of ERV-9 lncRNAs by #5 and #10 shRNAs generated the same phenotype: the Day13 erythroblasts transduced by both shRNAs did not turn red as the Day 13 cells transduced by the scrambled shRNA (Figure 1C). This phenotype indicated that KD of ERV-9 lncRNAs suppressed expression of globin genes, thus the hemoglobins that impart the red color to erythroid cells. (F)luorescence activated cell sorting (FACS) analysis showed that cells transduced with #10 shRNAs (and #5 shRNA) expressed early hematopoietic stem cell markers CD34 and CD45 at higher levels but erythroid markers, transferrin receptor (CD71) and glycophorin A (CD235a) at lower levels, as opposed to cells transduced with the scrambled shRNA, which expressed high levels of both CD71 and CD235a (Figure 1D), indicating that KD of ERV-9 lncRNAs impaired erythropoiesis. RNA analysis by qRT-PCR confirmed that KD of ERV-9 lncRNAs suppressed transcription of globin genes but also of key transcription factor (TF) genes GATA-1, NF-Y, KLF1 that regulate erythropoiesis and transcription of globin genes (Figure 1E).

Whole genome RNA-seq of transduced Day 13 cells showed that #5 and #10 shRNAs reduced by ≥2-fold the levels of ERV-9 lncRNAs transcribed from ∼500 ERV-9 LTRs (Figure 2A, first panel). This appeared reasonable since BLAST alignment showed that #10 shRNA (and #5shRNA) perfectly aligned with 1300 of the 4000 ERV-9 LTRs. Among these 1300 LTRs, some were associated with inactive genomic loci and did not transcribe ERV-9 lncRNAs and some were embedded in kilobase stretches of contiguous LTRs, Alu's and L1s, without any unique sequence tags for unambiguously mapping their chromosomal locations and were removed from bioinformatics analysis. Paradoxically, the shRNAs designed to KD ERV-9 LTR lncRNAs also increased the LTR lncRNAs transcribed from ∼700 ERV-9 LTRs (Figure 2A, first panel). One explanation could be that the upregulated ERV-9 lncRNAs were transcribed from ERV-9 LTRs located in indirectly responsive gene loci, whose transcription and that of the cis-linked ERV-9 LTRs were up-regulated by the downregulation of the directly responsive genes (see below).

KD of ERV-9 lncRNAs also respectively down- and upregulated ∼6000 genes by ≥2×: among them, 600 genes were down- and 750 genes upregulated by ≥2-fold with $P<0.05$ (Figure 2A, second and third panels), indicating that ERV-9 lncRNAs could act as transcriptional activators or repressors for different subsets of genes. Among the 600 downregulated genes were 60 key erythroid genes including the α- and β-globin and many TF genes; among the 750 upregulated genes were ∼20 early hematopoietic and 9 erythroid genes (Supplementary File S1: erv9rnaseq_EryGene_xlsx). Over 90% of these erythroid and hematopoietic genes were similarly down- or upregulated by both #5 and #10 shRNAs, thus by specific KD of ERV-9 lncRNAs (Figure 2A, fourth panel, lower left and upper right quadrants). Since ERV-9 lncRNAs acted predominantly as transcriptional activators for these key erythroid genes, KD of ERV-9 lncRNAs suppressed transcription of these genes and inhibited *ex vivo* erythropoiesis. Inhibition of erythropoiesis caused the erythroblasts to remain in more immature state (Figure 1D) and express higher levels of early hematopoietic genes. Thus, upregulation of these genes could result from indirect effect of ERV-9 lncRNA KD. In this study, we focused on the key erythroid genes that were down-regulated directly by ERV-9 lncRNA KD.

In the β-globin gene locus, KD of ERV-9 lncRNAs suppressed transcription of γ- and β-globin genes and intergenic regions including the locus control region (LCR) far upstream and the DNAs far downstream of the globin genes (Figure 2B). However, ε-globin gene, expressed at 30–1000-fold lower levels than γ- and β-globin genes, was activated (Figure 2B and C), possibly due to global changes in the levels of TFs specific to ε-promoter that overcame the direct, suppressive effect of ERV-9 lncRNA KD on ε-globin gene. Indeed, suppression of ε-globin gene was observed in the human globin transgenic locus in Tg mose erythroblasts, in which ERV-9 lncRNA KD did not globally change the TF levels (see Figure 3B and D). This locus-wide transcriptional suppression could result from ERV-9 lncRNAs transcribed from the β-globin locus ERV-9 LTR acting in cis or the ERV-9 lncRNAs transcribed from many other among the 4000 copies of ERV-9 LTRs acting in trans to regulate transcription of the β-globin locus. Moreover, GATA-1, KLF1 and NFYA genes also have cis-linked ERV-9 LTRs (Supplementary Figure S2) and were downregulated by ERV-9 lncRNAs KD (Figure 2C); these downregulated TFs could in turn cause transcriptional suppression of the β-globin locus.

In erythroblasts of Tg mice, depletion of ERV-9 lncRNAs suppressed transcription of the human β-globin gene locus but not mouse gene loci without cis-linked ERV-9 LTR

To dissect the complex transcriptional network of erythropoiesis coordinated by ERV-9 lncRNAs, we first investigated the cis- versus the trans-mechanism of ERV-9 lncRNAs, utilizing a Tg mice model carrying the 100 kb human β-globin locus with a single copy of cis-linked ERV-9 LTR (14). Tg FL erythroblasts were transduced with #5, #10 and scrambled shRNAs and cultured to Day 8 for analysis. Like Day 13 human erythroblasts, Day 8 Tg FL erythroblasts expressed the adult globin gene program (Supplementary Figure S3A). As in human erythroid cells (12), ERV-9 lncRNAs in Day 8 Tg erythroblasts were transcribed as long (>200 nt), non-coding, polyadenylated, nuclear RNAs (Supplementary Figure S4). Although the ERV-9 lncRNAs were located predominantly in the nucleus (Supplementary Figure S4C), the RISC complex was able to process the #5 and 10 shRNAs transcribed from the integrated lentiviral DNA to target the nuclear ERV-9 lncRNAs, as shown previously (18).

The Day 8 erythroblast populations transduced separately by #5 and #10 shRNAs exhibited similar degree of hemoglobinization (Figure 3A). qRT-PCR showed that depletion of ERV-9 lncRNAs by either #5 or #10 shRNA similarly reduced transcription of the human β-globin transgenic locus, without significantly affecting transcription of mouse globin genes and TF genes NFYA, GATA-1, -2 and KLF1 (Figure 3B). The same KD effects were observed also with a second line of Tg mouse (Supplementary Figure

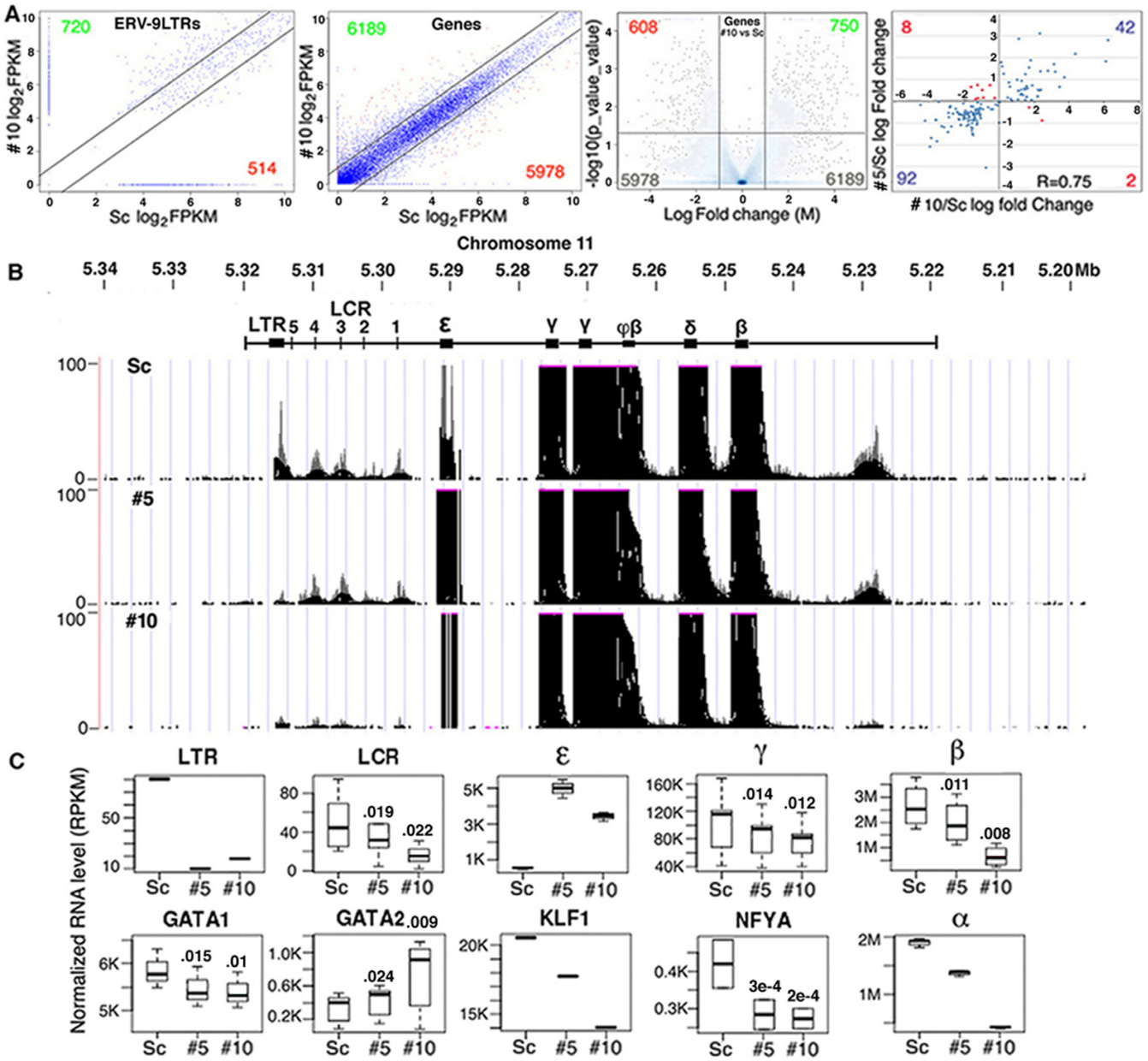

Figure 2. Transcriptome analysis of transduced Day 13 erythroblasts by whole genome RNA-seq. (**A**) **first panel:** scatter-plot: number of ERV-9 LTRs whose RNA levels were up- (in green) or downregulated (in red) by >2-fold due to ERV-9 lncRNA KD by #10 shRNA. FPKM: fragments per kilobase per million reads. Row of dots at bottom: ERV-9 lncRNAs, whose FPKMs were reduced to 0 by shRNA KD. **Second panel:** scatter-plot: Genes whose RNA levels were up- (in green) or downregulated (in red) by >2× due to #10 shRNA. **Third panel:** volcano plot of data in second panel. Genes to the left and right of the double vertical lines were respectively down- and upregulated by >2×. Genes above the horizontal line had >2× changes with *P*-values ≤0.05. **Fourth panel:** hematopoietic and erythroid genes whose #5/Sc log fold changes (Y-axis) were plotted versus #10/Sc log fold changes (X- axis). Genes in blue in the upper right and lower left quadrants were up- or downregulated by both #5 and 10 shRNAs; genes in red in the upper left and lower right quadrants were differently up- or downregulated by the shRNAs and were off-targets of the shRNAs (Supplementasry File S1). (**B**) Transcription profiles of human β-globin locus in Day 13 erythroblasts transduced with Sc, #5 and #10 shRNAs. Coordinates in megabases (Mb) were from tip of the short arm of chromosome 11. φβ: pseudo β-globin gene. (**C**) Box plots of normalized RNA sequence reads of key erythroid genes. Top and bottom of boxes: first and third quartiles of the FPKM values; whiskers: minimum and maximum FPKM values; horizontal bars, medium FPKM values. K and M: RPKM in thousands and millions respectively. *P*-values shown above the boxes were from paired, one tailed *t*-tests; for the other boxes: *P*-values were <2e-4 and not shown.

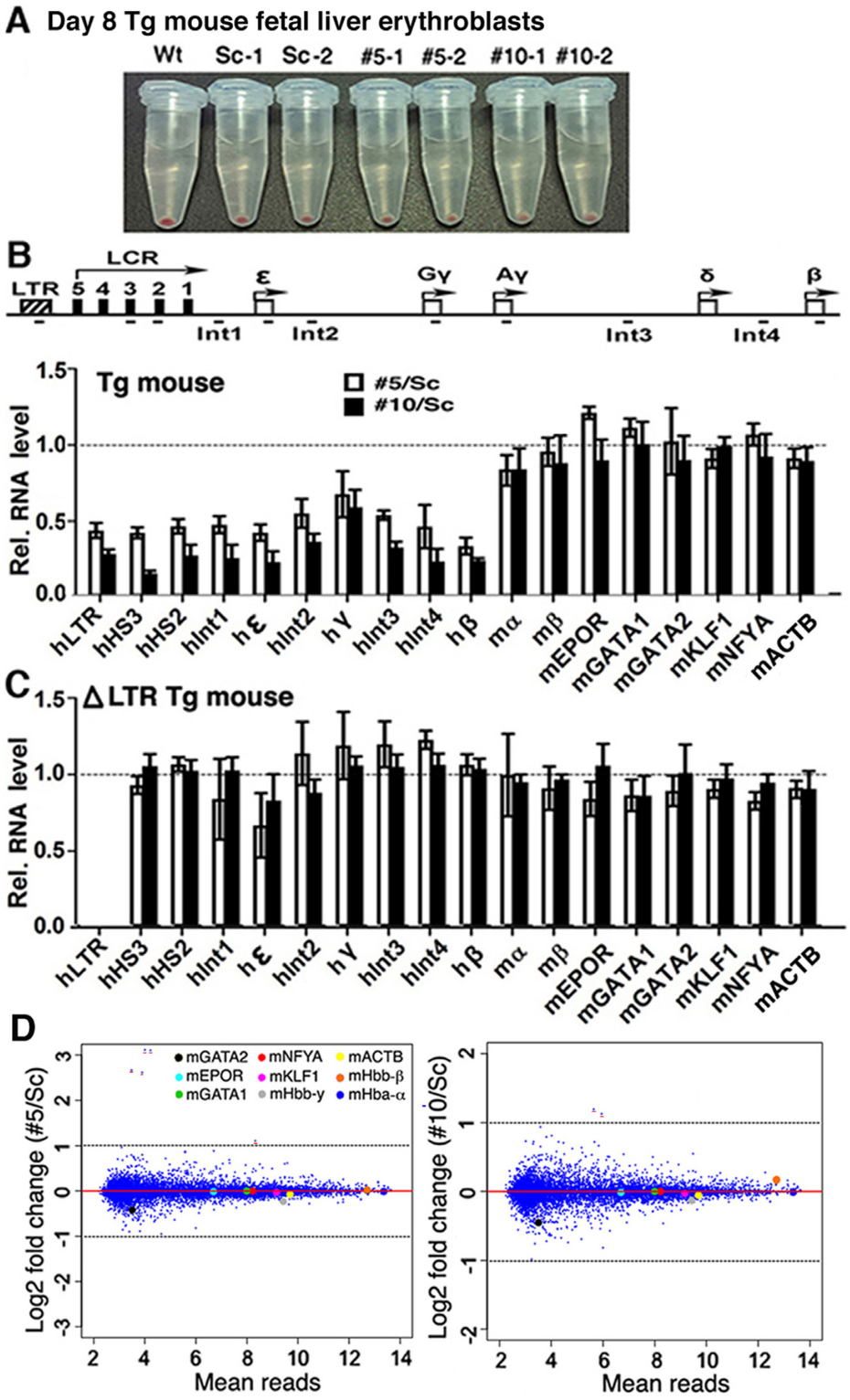

Figure 3. Depletion of ERV-9 lncRNA suppressed transcription of the human globin transgenic locus but not of mouse genes in Day 8 Tg FL erythroblasts. (**A**) Images of transduced Day 8 Tg FL erythroblasts. Annotations: same as Figure 1C; Wt, non-transduced Day 8 FL erythroblasts of wild-type mouse. (**B** and **C**) qRT-PCR of RNAs from Day 8 Tg FL erythroblasts (line #7) and ΔLTR (line #7). Bars under map: locations of PCR primer pairs. Prefixes h and m: human or mouse genes. Normalization of RT-PCR products were as in Figure 1D. Relative RNA levels, #5/Sc and #10/sc, were mean±SEM of three independently transduced cells. qRT-PCRs of an independent Tg line #16 (see Supplementary Figure S3B). (**D**) RNA analysis by microarrays. Y axis: Log2 fold change (fc) of #5/Sc and #10/Sc. X axis: mean sequence reads of each gene from reads of the gene in #5-, #10- & Sc- transduced Day 8 cells; numbers were exponents on a base of 2. Colored dots: log2 fold change of key mouse erythroid genes. The 5 and 2 genes with log2 fc>1 marked by red bars were upregulated due to off-target effects, as they were different genes with diverse mean sequence reads and were not commonly up-regulated by both shRNAs.

S3B). Thus, the KD phenotype and genotype in Tg mouse erythroblasts were generated by specific KD of ERV-9 lncRNAs.

To further confirm that transcriptional suppression of the human β-globin transgenic locus resulted not from off-target effects of #5 and #10 shRNAs, lentiviruses expressing #5 and #10 shRNAs were also transduced separately into FL erythroblasts of ΔLTR Tg mice. The ΔLTR Tg mouse was derived from the parental BAC Tg mouse, in which the ERV-9 LTR was deleted by cre-loxP mediated *in situ* recombination (14). In the transduced erythroblasts of ΔLTR Tg mice–in the absence of the ERV-9 LTR DNA template and therefore synthesis of ERV-9 lncRNAs, neither #5 nor #10 shRNA suppressed transcription of the β-globin transgenic locus (Figure 3C). Thus, specific KD of ERV-9 lncRNAs suppressed transcription of the cis-linked human β-globin gene locus but exerted no transcriptional effects in trans on mouse genes.

Genome-wide RNA analysis by microarrays confirmed the cis-mechanism of ERV-9 lncRNA function: KD of ERV-9 lncRNAs by #5 or #10 shRNAs did not commonly change the RNA levels of mouse genes including key mouse erythroid genes by ≥2× (Figure 3D). While changes in RNA levels of ≤1.5-fold could be background noise and insignificant, 6 and 8 mouse genes were respectively up- and downregulated by 1.5–1.9-fold by both #5 and #10 shRNAs (Supplementary Figure S5). Changes in these 14 genes, comprising 0.05% of the 26 000 mouse genes on the microarray, could result from chance events and not from specific KD by #5 and #10 shRNAs (Supplementary Table S1). These 14 genes did not affect transcription of the mouse globin genes (Supplementary Figure S5) or the human globin transgenes, as their protein products were not shown to do so (Supplementary Table S1). Thus, microarray data confirmed that in Tg mouse erythroblasts ERV-9 lncRNAs acted solely in cis to regulate transcription of the cis-linked human β-globin transgenic locus.

In Tg mouse erythroblasts, ERV-9 lncRNAs associated strongly in cis with the ERV-9 LTR and spread through the downstream locus to regulate transcription of cis-linked human globin genes; ERV-9 lncRNAs associated in trans with mouse genes exhibited no transcriptional activity

A cis-mechanism of ERV-9 lncRNA action suggested that ERV-9 lncRNAs activated transcription of the globin gene locus by co-localizing with the cis-linked globin genes and intergenic DNAs but not in trans with the mouse gene loci. To examine this possibility, we used ChIRP assays with a split probe strategy (17) to pull down the chromatin associated with ERV-9 lncRNAs in Day 8 Tg FL erythroblasts. As the two pools of odd and even probes contained different sequences (Figure 4A; 'Materials and Methods' section), the two respective pools of chromatin DNAs pulled down by off-target hybridization would be different. Only the chromatin DNA associated with ERV-9 lncRNAs would be commonly pulled down by both the odd and even probes.

The two pools of chromatin pulled down respectively by the odd and even probes were subsequently used for RNA and DNA isolation. RNA analysis by qRT-PCR showed that both pools of probes pulled down >50% of ERV-9 lncRNAs in the Day 8 erythroblasts (Figure 4B). ChIRP-qPCR showed that both pools of probes commonly pulled down ERV-9 LTR and the LCR and lower levels of globin genes and intergenic DNAs (Figure 4C).

To examine the global association patterns of the ERV-9 lncRNAs, we carried out ChIRP-seq with spleen erythroid cells of Tg mice (14), since the spleen erythroblasts expressed the same globin gene program as Day 8 FL erythroblasts (Supplementary Figure S3C and D) and were available in large cell numbers required in ChIRP-seq to produce sufficient sequence reads by deep sequencing. ChIRP-seq results showed that ERV-9 lncRNAs associated at a very high level with the 100 kb human β-globin locus, at log2_fold enrichment (fe) = 3.9 and 1.9, including or excluding the ERV-9 LTR respectively (Figure 4D). In contrast, the log2_fe of ERV-9 lncRNAs with Tg mouse DNAs, analyzed in 1 kb tiling of 100 kb DNA segments, were significantly lower and showed a normal Gaussian distribution, indicating potentially random trans-associations of ERV-9 lncRNAs with the mouse DNAs at significantly lower levels (Figure 4D). ERV-9 lncRNA occupancy profile of the human β-globin transgenic locus viewed with UCSC Genome Browser showed that ERV-9 lncRNAs associated at an extremely high level with the ERV-9 LTR, potentially through RNA/DNA sequence complementarity in the R-U5 region, and at much lower levels with LCR HS3 and HS2 sites, the intergenic DNAs and globin genes, but did not associate significantly with the mouse β-globin gene locus (Figure 4E). This finding was consistent with earlier results that KD of ERV-9 lncRNA suppressed transcription of the human but not the mosue globin genes (Figure 3B and D). Thus, ERV-9 lncRNAs associated strongly with the ERV-9 LTR and acted in cis to activate transcription of downstream human the globin genes but did not act in trans to regulate transcription of the mouse globin gene.

In human erythroblasts, ERV-9 lncRNAs associated strongly with ERV-9 LTRs and spread in cis to regulate transcription of downstream genes; these direct-responsive genes in turn regulated transcription of indirect-responsive genes without nearby, cis-linked ERV-9 LTRs

In D13 human erythroblasts, ChIRP-seq showed that the ERV-9 lncRNA occupancy profile at the β-globin locus was nearly identical to that of β-globin locus in Tg mouse erythroblasts containing only a single copy of the ERV-9 LTR (compare Figure 4E and F, left panels). Thus, in human erythroblasts, ERV-9 lncRNAs also associated at an extremely high level with the LTR DNA and spread in cis to associate at lower levels with multiple sites in the downstream globin locus, without apparent interference from ERV-9 lncRNAs transcribed and diffused from many of the 4000 ERV-9 LTRs. In the NFYA gene locus, where the nearest ERV-9 LTRs were megabases from the gene (Supplementary Figure S2D), ERV-9 lncRNAs did not associate at significant levels comparable to those of the β-globin gene locus (Figure 4F; Supplementary Figure S6A, compare log2_fe of β-globin and NFYA loci). This indicated that ERV-9 lncRNAs did not directly associate and regulate transcription of NFYA gene, although KD of ERV-9 lncRNAs significantly reduced transcription of the gene

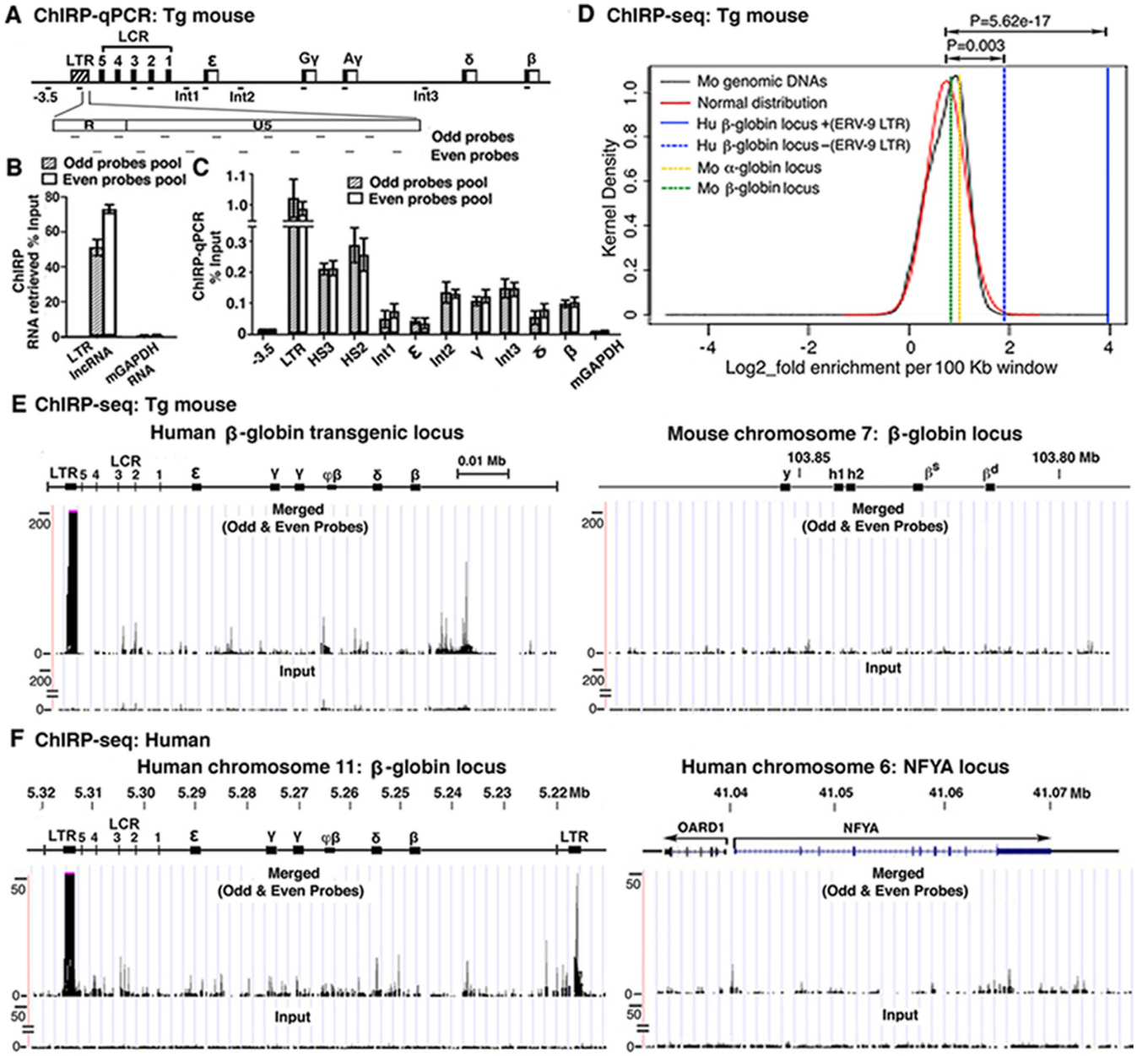

Figure 4. Occupancy profiles of ERV-9 lncRNAs in Tg mouse and human erythroblasts. (**A**) Locations of ChIRP probes and of primer pairs for ChIRP-qPCRs. −3.5: DNA site at 3.5 kb 5′ of the ERV-9 LTR. (**B**) Yield of ERV-9 lncRNA pulled down by ChIRP in Day 8 Tg FL erythroblasts. Amount of ERV-9 lncRNAs in input chromatin was set at 100%. Results were averages of two independent assays. (**C**) ChIRP-qPCR of pulled down chromatin DNA in Day 8 Tg FL erythroblasts. Level of PCR product from input DNA was set at 100. Values were mean±SEM of three independent assays. (**D**) Levels of ERV-9 lncRNA occupancies throughout the mouse genome and the human β-globin transgenic locus in 1 kb tiling of 100 kb DNA segments in Tg spleen erythroblasts (line #7). Log2_fold enrichment (fe) = log2 FPKM of DNA pulled down by the merged Odd and Even probes/FPKM of the same DNA segment in Input DNA. Vertical lines: log2_fe of human β-globin and mouse α- and β-globin gene loci. (**E** and **F**) ERV-9 lncRNA occupancy profiles at the human and mouse β-globin loci locus in Tg spleen erythroblasts and at human β-globin and NFYA locus in D13 human erythroblasts. Y axis: Normalized sequence reads of chromatin DNAs pulled down by the merged Odd and Even probes. (see http://ccc.gru.edu/tuan/erv9.chirpseq; ID: dorothytuan; password: tuanchoihu).

(Figure 2C). Thus, NFYA was likely an indirect-responsive gene down-regulated by genes that were down-regulated directly by ERV-9 lncRNA KD. Genome-wide, there were many potentially direct-responsive gene loci, which, like the β-globin gene locus, have cis-linked ERV-9 LTRs within 300 kb from the genes and locus-wide association of ERV-9 lncRNAs and whose transcription was suppressed by ≥2× due to ERV-9 lncRNA KD (File S2). These potential, direct-responsive genes included TF genes encoding TFs such as KLF1 and CCNDBP1 (Supplementary Figure S6 B and C). Each of these TFs could in turn regulate transcription of multiple target genes without cis-linked ERV-9 LTRs, like the indirect-responsive NFYA gene.

Locus-specific deletion of ERV-9 LTR and lncRNAs by CRISPR-cas9 from β-globin gene locus suppressed transcription of cis-linked globin genes and a few unlinked genes

To further examine a cis mechanism of ERV-9 lncRNA function shown above in Tg mouse and human erythroblasts by global KD of ERV-9 lncRNAs with the shRNAs, we used CRISPR-cas9 to delete the ERV-9 LTR, thus synthesis of locus-specific ERV-9 lncRNAs, from the β-globin gene locus in erythroleukemia K562 cells (Figure 5A and B). Immortalized K562 cells were chosen for the study, since primary human erythroblasts derived from CD34+ cells could be cultured robustly for only up to 2 weeks, which precluded selection and expansion of clonal lines carrying homozygous deletion of the ERV-9 LTR/lncRNAs. In three clonal, homozygous ΔLTR lines #2, 6 and 8, qRT-PCR and microarrays showed that deletion of the ERV-9 LTR/lncRNAs from the β-globin gene locus commonly suppressed transcription of downstream ε-, γ- and φβ−globin genes by ≥2× (Figure 5C, D and F), except for β−globin gene, which was not suppressed, since it was expressed at a very low level ∼2–5% that of the ε− and γ−globin genes (Figure 5D) and could exist in an inaccessible chromatin structure resistant to editing by the CRISPR-cas9 plasmids. The results confirmed that the ERV-9 LTR/lncRNAs acted in a cis mechanism to regulate transcription of downstream, cis-linked globin genes.

However, 11 and 8 non-globin genes were respectively, commonly down- and upregulated by ≥2× in the three clonal lines (Figure 5E) and none of them were OTSs predicted for the sgRNAs ('Materials and Methods' section). Transcriptional dis-regulation of these 19 genes did not appear to result from the down-regulation of globin genes since globin genes, unlike TF genes, do not regulate transcription of other genes, nor from CRISPR-cas9 mediated deletion of the ERV-9 LTR enhancer and promoter, both of which are cis-regulatory elements that do not regulate transcription of unlinked and un-coordinately expressed genes. Thus, ERV-9 lncRNAs transcribed from the β-globin ERV-9 LTR appeared to associate with and regulate in trans the transcription of these 19 genes, comprising 0.08% of the 27 000 genes on the microarray, potentially through chance encounter between the ERV-9 lncRNAs and the unlinkd gene loci. However, among the 11 commonly down-regulated non-globin genes, 7 gene loci contained ERV-9 LTRs within 300 kb from the genes (Supplementary Table S2). This indicated that the ERV-9 lncRNAs dissociated from the β-globin ERV-9 LTR could form RNA/DNA duplex with the R-U5 regions in the ERV-9 LTRs of these 7 gene loci to activate transcription of the cis-linked genes by a similar cis mechanism as in the β-globin gene locus. Consistent with this interpretation, human KLF1 gene with a cis-linked ERV-9 LTR at 300 kb from the gene (Supplementary Figures S2E and S6C) was commonly suppressed by ∼1.7× in the three LTR deletion clonal lines (Figure 5F). In contrast, mouse KLF1 gene without cis-linked ERV-9 LTR was not suppressed by ERV-9 lncRNA KD in Tg mouse erythroblasts (Figure 3B and D). As 4 among the 11 non-globin genes commonly downregulated ≥2× had no nearby cis-linked ERV-9 LTRs, ERV-9 lncRNAs transcribed from the globin locus ERV-9 LTR appeared to associate and regulate in trans these four genes.

An alternative interpretation of the results appeared unlikely that ERV-9 lncRNAs transcribed from the globin locus ERV-9 LTR associated with and acted exclusively by a trans mechanism even for the globin genes: the lncRNAs dissociated from the LTR template and diffused through the nucleoplasm to associate with and activate transcription of the globin genes. In the contrary, we found that forced expression of ERV-9 lncRNAs from integrated LTR-GFP plasmid in K562 cells (Supplementary Figure S4B) and ΔLTR clonal line #2 acted in cis to activate transcription of the linked GFP gene (Supplementary Figure S7B and C, middle panels) but did not reach and activate in trans transcription of the endogenous globin genes, which were transcribed at similar levels in control cells transfected by the GFP plasmid not expressing ERV-9 lncRNAs (Supplementary Figure S7D). This result was consistent with an earlier finding that forced expression of LCR HS2 enhancer lncRNAs (eRNAs) from transfected HS2-CAT reporter plasmids activated in cis the transcription of the linked CAT gene but did not act in trans to affect transcriptional activity of the endogenous ε-globin promoter (20). Together, the results indicated that ERV-9 lncRNAs transcribed from the globin ERV-9 LTR bound to and activated transcription of globin genes by a cis-mechanism; the ERV-9 lncRNAs through chance encounter also bound in trans to 19 unlinked gene loci to regulate either in cis or in trans transcription of these 19 genes.

ERV-9 lncRNAs stabilized looping of ERV-9 LTR with downstream LCR, globin genes and intergenic DNAs

To assess the functional significance of the strong association of ERV-9 lncRNAs with the ERV-9 LTR shown by ChIRP-seq (Figure 4) and the cis-mechanism of ERV-9 lncRNA function, we used chromosome conformation capture (3C) (21) to determine if depleting ERV-9 lncRNAs impaired the interaction/looping of the LTR with the downstream globin gene locus in Day 8 Tg FL and Day 13 human erythroblasts. The 3C results showed that KD of ERV-9 lncRNAs significantly lowered the co-localization frequencies of the ERV-9 LTR with the LCR, globin genes and intergenic DNAs (Figure 6A and B). Cloning and sequencing of the 3C junction fragments confirmed co-localization/looping of the ERV-9 LTR with downstream globin genes and intergenic DNAs (Supplementary Figure S8). Thus, the ERV-9 lncRNAs tightly bound to the ERV-9 LTR stabilized interaction/looping of the LTR DNA with downstream DNA sites throughout the locus.

ERV-9 lncRNAs stabilized assembly of the ERV-9 LTR enhancer–pol II complex and the interaction of the LTR enhancer complex with the downstream gene locus

To further assess the functional significance of ERV-9 lncRNAs, we next used chromatin immunoprecipitation (ChIP) to determine the effects of ERV-9 lncRNA KD on assembly of the ERV-9 LTR enhancer complex in Day 8 Tg erythroblasts. The ChIP-qPCR results showed that ERV-9 LTR bound very high levels of NF-YA, GATA-1, pol

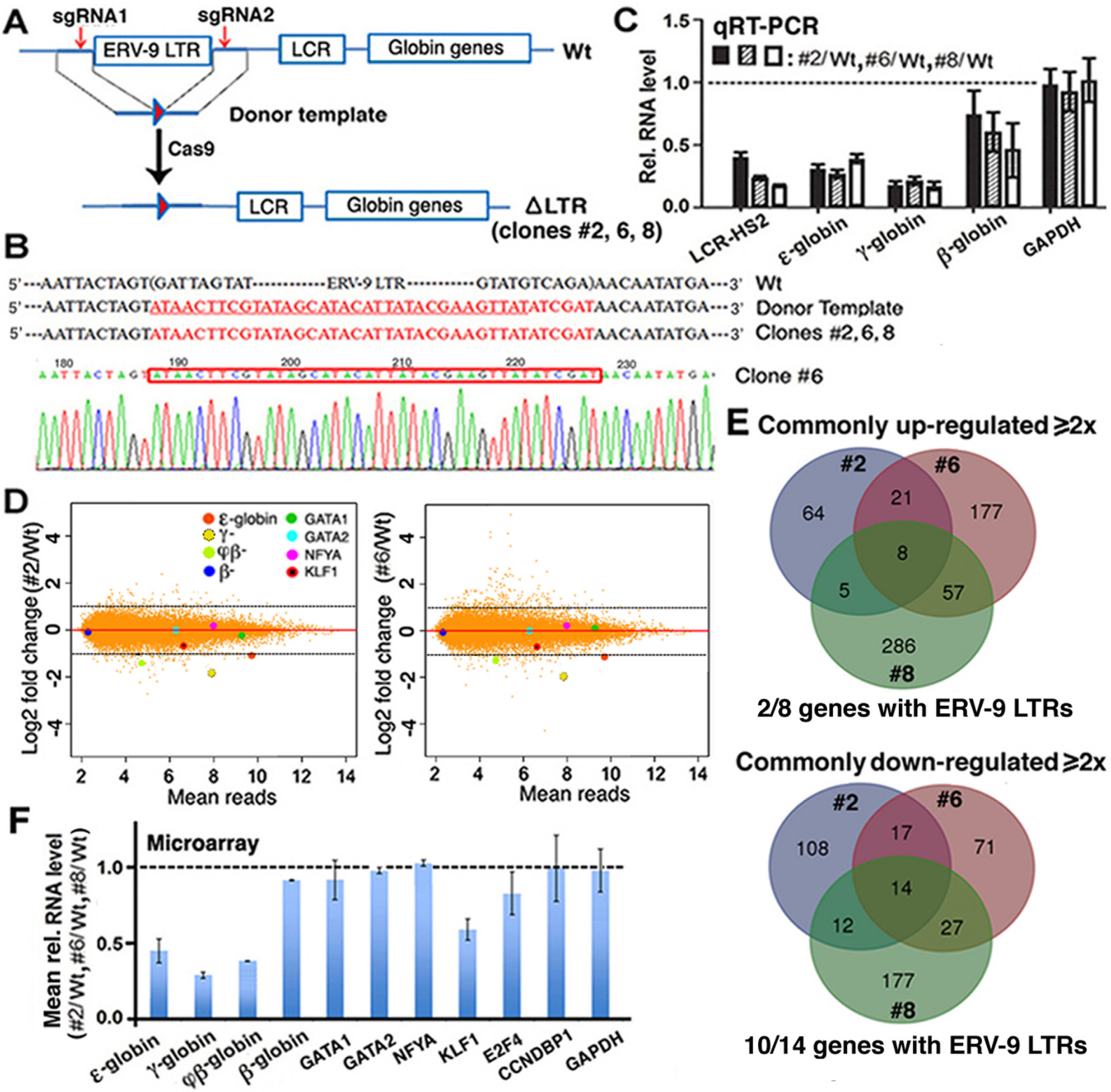

Figure 5. Locus-specific deletion of ERV-9 LTR/lncRNAs: ERV-9 lncRNAs transcribed from β-globin ERV-9 LTR associated and regulated in cis the transcription of downstream globin genes; ERV-9 lncRNAs also associated in trans and regulated either in cis or in trans the transcription of some unlinked genes. (**A**) Strategy for locus-specific deletion of ERV-9 LTR from β-globin gene locus by CRISPR-cas9 to generate clonal ΔLTR lines (see 'Materials and Methods' section). (**B**) Confirmation of ΔLTR clonal lines containing homozygous deletion of the ERV-9 LTR from β-globin gene locus by Sanger sequencing of PCR amplified DNA fragments spanning the ERV-9 LTR in Wt and ΔLTR clonal lines aligned with the sequence of the donor template. (**C**) qRT-PCR analysis of total cellular RNAs isolated from ΔLTR clonal lines #2, 6 and 8. Dotted line: RNA levels in Wt K562 cells serving as reference and set at 1. Values were mean of triplicate RT-PCR reactions. (**D**) Genome-wide RNA analysis by microarrays: Scatter plots of genes up- and downregulated by deletion of ERV-9 LTR and lncRNAs from β-globin locus: (clone #8, not shown). Dotted horizontal lines: Log2 fold change of +1 or −1, marking up- or downregulation by 2×. Colored dots: key erythroid genes. Other designations: same as Figure 3D. (**E**) Venn diagrams of commonly up- and downregulated genes in the 3 ΔLTR clones: 2 of 8 commonly upregulated genes and 10 of 14 commonly downregulated genes—3 globin genes and 7 other genes–have cis-linked ERV-9 LTRs within 300 kb of the genes (Supplementary Table S2). (**F**) Microarrays analysis of key erythroid genes in the 3 ΔLTR clones: Y axis: mean RNA levels of #2, 6 and 8 LTR deletion clones relative to those of Wt K562 cells, which were set at 1 as marked by the dotted horizontal line.

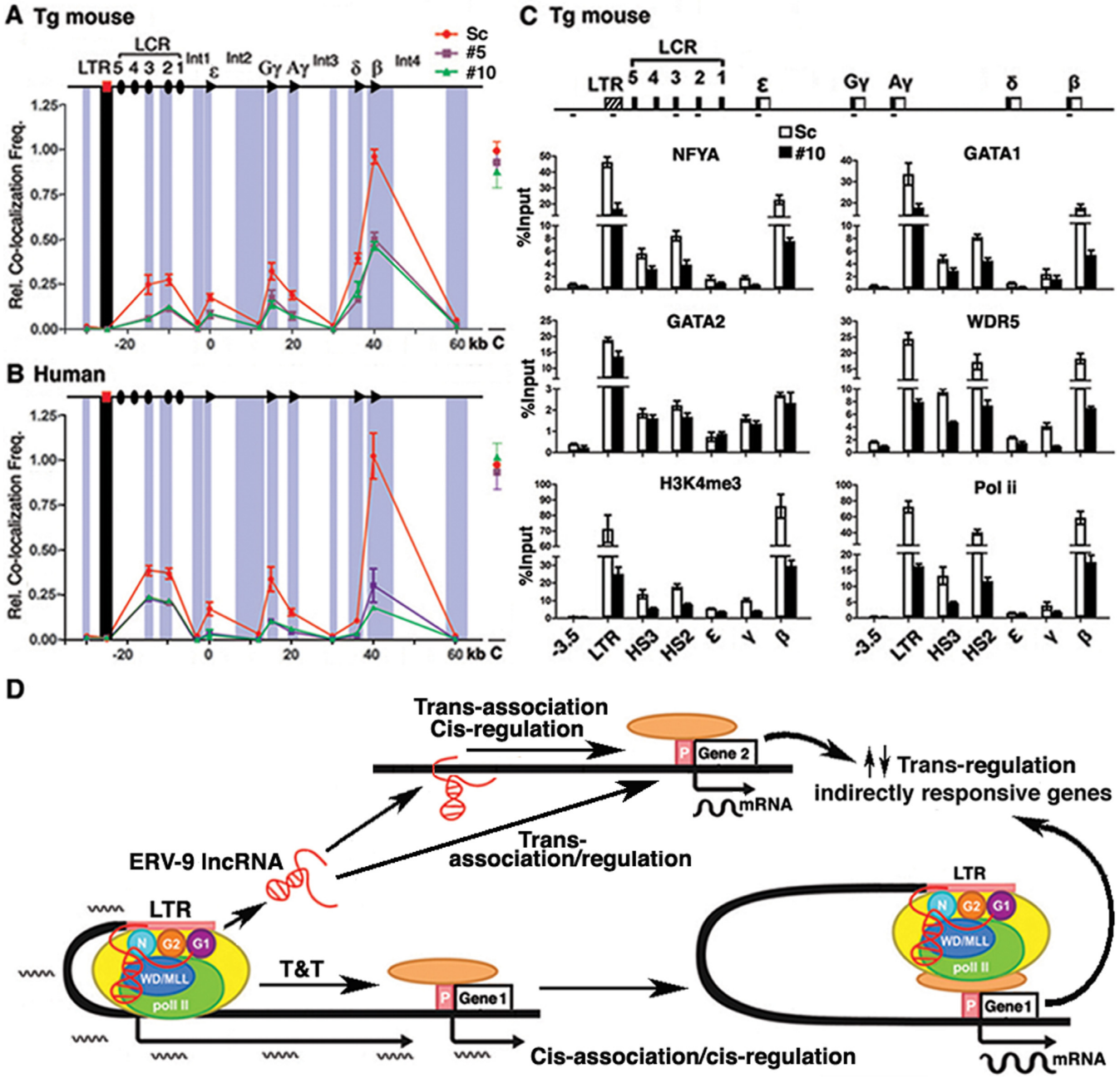

Figure 6. Depletion of ERV-9 lncRNAs destabilized assembly of the ERV-9 LTR enhancer complex and reduced looping frequencies of the LTR enhancer with the downstream gene locus; model of ERV-9 lncRNA function. (**A**) and (**B**) 3C assays of Day 8 Tg FL erythroblasts (line #7) and Day 13 human erythroblasts. Vertical colored bars: DNA fragments generated by Ase I cleavage; black bar: Ase I fragment spanning ERV-9 LTR, 3C anchor fragment. Looping frequency between the ERV-9 LTR and β-globin gene was set at 1; values were mean±SEM of three independent 3C assays. (**C**) Control co-localization frequency of mouse HS2-mouse β-globin gene in Tg erythroblasts and of human β-actin locus, first exon-last exon in human erythroblasts. C**:** ERV-9 lncRNA KD reduced occupancies of TFs on the ERV-9 LTR and the LCR and globin genes in Day 8 Tg erythroblasts (line #7). Bars under map: locations of ChIP-qPCR primer pairs. ChIP values were averages of two independent assays. (**D**) Model of ERV-9 lncRNA function: Lower graph: cis-association and cis-regulation. ERV-9 lncRNA transcribed from the ERV-9 LTR interacted with NF-Y (N), GATA-1 (G1), WDR5/MLL2 (WD/MLL) and pol II to assemble the LTR enhancer complex (Yellow oval) and tethered the LTR complex to downstream DNA sites and by a cis T&T mechanism through the intergenic DNAs to ultimately loop with the target gene. Squiggles: intergenic RNAs transcribed by the LTR enhancer complex and mRNA transcribed by the promoter complex (orange oval) of the cis-linked gene. Top graph: trans-association and trans- or cis-regulation. ERV-9 lncRNA dissociated from its LTR DNA template could associate in trans and regulate either in trans the transcription of unlinked genes loci or in cis the transcription of unlinked gene loci with nearby ERV-9 LTRs. Direct-responsive Genes 1 and 2–their RNA and protein products–in turn could up- or downregulate in trans the indirect-responsive genes without cis-linked ERV-9 LTRs.

II and the WDR5/MLL2 co-activator complex to assemble the LTR-pol II enhancer complex (Figure 6C). Depleting ERV-9 lncRNAs significantly reduced the *in vivo* occupancies of these TFs, co-factors and pol II at the ERV-9 LTR and multiple downstream sites and diminished the active chromatin mark H3K4me3 at these sites (Figure 6C). Thus, ERV-9 lncRNAs strengthened the interaction of the TFs and co-factors with the ERV-9 LTR enhancer and the downstream globin gene locus. Similarly in D13 erythroblasts, ChIP-qPCR showed that ERV-9 lncRNA KD reduced the *in vivo* occupancies of the TFs and pol II at the ERV-9 LTR enhancer and multiple downstream sites (Supplementary Figure S9A). In agreement, ChIP-seq showed that the ERV-9 LTR in the β-globin gene locus binds GATA-1, NF-Y and pol II and enhancer markers H3K4Me1 and p300 to assemble an active ERV-9 LTR enhancer complex in human erythroblasts and K562 cells (Supplementary Figure S9B). Thus, in both Tg mouse and human erythroblasts, ERV-9 lncRNAs stabilized assembly of the LTR enhancer/NF-Y/GATA-1/pol II complex and facilitated co-localization of enhancer-associated TFs and co-factors with the downstream globin gene locus.

Together, ChIRP, 3C and ChIP assays demonstrated that in the β-globin gene locus of Tg mouse and human erythroblasts, ERV-9 lncRNAs associated strongly and predominantly with their LTR template to stabilize assembly of the LTR enhancer/NF-Y, GATA-1, -2/pol II complex and strengthen looping of the LTR complex with multiple, downstream DNA sites to facilitate a cis T&T mechanism of the LTR enhancer complex through the intergenic DNAs to ultimately loop with and activate transcription of the globin genes (Figure 6D). Thus, ERV-9 lncRNA KD by shRNAs or its elimination by CRISPR-cas9 destabilized assembly of the ERV-9 LTR enhancer complex and weakened the cis T&T mechanism of the LTR complex in activating transcription of the downstream globin genes (Figures 1–3, 5). Consistent with the cis-mechanism of ERV-9 lncRNA function, forced expression of ERV-9 lncRNAs from transfected LTR-GFP plasmid did not rescue the transcriptional suppression of the globin genes in ΔLTR clonal cells (Supplementary Figure S7).

DISCUSSION

In this study, we showed that ERV-9 lncRNAs transcribed from ERV-9 LTR retrotranspsons served biological function in modulating *ex vivo* human erythropoiesis by regulating in cis the transcription of key erythroid genes such as the globin genes and master TF genes including KLF1 and CCNDBP1 genes (Figure 1, Supplementary Figures S2 and 6; Supplementary File S2). Global KD of ERV-9 lncRNAs in human erythroblasts directly suppressed transcription of these key erythroid genes with nearby ERV-9 LTRs, at ≤300 kb from the respective genes–the direct-responsive genes, and indirectly dis-regulated target genes without nearby ERV-9 LTRs, the indirect–responsive genes, to disrupt the transcription network of erythropoiesis and inhibit *ex vivo* erythropoiesis (Figures 1, 2 and 6D).

In support of a cis-mechanism of ERV-9 lncRNA function, ChIRP-seq showed that ERV-9 lncRNAs associated strongly with the LTR DNA template and weakly throughout the downstream globin gene locus in Tg mouse and human erythroblasts (Figure 4). Combination of RNA-seq and ChIRP-seq showed that genome-wide there were over 60 human gene loci, which contained cis-linked ERV-9 LTRs within 300 kb from the respective genes, associated with ERV-9 lncRNAs locus-wide and were transcribed locus-wide in the intergenic DNAs and the genes and whose transcription and that of the cis-linked ERV-9 LTRs were suppressed ≥2× by ERV-9 lncRNA KD (Supplementary Figure S6 and File S2). In these gene loci, as in the β-globin gene locus, the ERV-9 lncRNAs transcribed from their respective ERV-9 LTR templates could likewise regulate transcription of the downstream genes by a similar cis mechanism.

An established model of a cis-mechanism of lncRNA function is Xist lncRNA. It is transcribed from the Xist locus on X-chromosome and spread in cis to coat and inactivate the entire allelic X chromosome to achieve X chromosome dosage compensation in female cells (22,23). In addition, HOTTIP lncRNA synthesized from the HOXA gene locus acts also in cis to remodel chromatin structure and activate transcription of the linked HOXA genes (24). Whether the HOTTIP lncRNA acts through a T&T mechanism of long-range enhancer function is not known. In contrast, lncRNAs can act also in trans: The HOTAIR lncRNA transcribed from the HOXC locus diffuses through the nucleoplasm to associate with over 800 chromosomal sites with GA-rich sequences to trimethylate histone 3 lysine 27 and suppress the target gene domains (17). These examples demonstrate that whether acting in cis or in trans, the lncRNAs need to physically associate with their target gene loci. In the NFYA gene locus without nearby ERV-9 LTRs (Supplementary Figure S2D), ERV-9 lncRNAs were not found to physically associate at a significant level with the gene locus (Figure 4F and Supplementary Figure S6A). Thus, ERV-9 lncRNAs did not appear to directly regulate in cis or in trans the transcription of NFYA gene, even though KD of ERV-9 lncRNAs significantly reduced transcription of the gene (Figure 2C), The NFYA gene and many other human gene loci without nearby ERV-9 LTRs and without significant occupancy by the ERV-9 lncRNAs but were down-regulated by ≥2× were likely indirect-responsive genes whose transcription was regulated not by the ERV-9 lncRNAs but by the direct-responsive genes of ERV-9 lncRNAs (Figure 6D).

The cis-mechanism of ERV-9 lncRNA function was most clearly demonstrated In Tg mouse erythroblasts carrying a single copy of the primate-specific ERV-9 LTR of the human β-globin locus in the absence of the ERV-9 lncRNAs transcribed from other copies of the ERV-9 LTRs. In this simplified model system, ERV-9 lncRNA KD suppressed transcription of the cis-linked human globin genes but did not affect transcription of unlinked mouse globin and TF genes nor inhibited mouse erythropoiesis (Figure 3). The lack of trans regulation of mouse genes by ERV-9 lncRNAs could be due to species difference in that the mouse genome does not contain the primate specific ERV-9 LTRs, so the ERV-9 lncRNAs transcribed from the human globin transgenic locus were unable to form RNA/DNA duplex through sequence complementary with mouse gene loci to facilitate assembly of the LTR enhancer complex and a cis T&T mecha-

nism of enhancer function. Thus, the issue of species difference did not necessarily provide support in favor of a trans mechanism of ERV-9 lncRNA functiion. In support of this interpretation, in mice, the LTR lncRNAs transcribed from the murine IAP LTR retrotransposons have been shown to act in cis to regulate expression of the cis-linked Agouti and Axin genes (25,26). Thus, in both mouse and human genomes, LTR retrotransposons/lncRNAs could act in cis to regulate transcrirption of linked genes.

Taken together, our findings indicated that by acting in cis to directly regulate transcription of the direct-responsive, key erythroid genes with nearby ERV-9 LTRs and by acting indirectly and in trans through the RNA and protein products of the direct-responsive genes to regulate indirect-responsive genes without nearby ERV-9 LTRs, ERV-9 LTR lncRNAs modulated the complex transcription network of erythropoiesis (Figure 6D). LncRNAs play significant roles in cellular differentiation and development and dis-regulation of lncRNAs causes cancers and leukemia (27–31). Indeed, LTR lncRNAs have been reported to regulate the transcription network and maintain pluripotency in human ES cells and early embryos (32–35). However, the molecular mechanisms by which the LTR lncRNAs exert such regulatory activities are largely unknown. Our findings here indicate that LTR lncRNAs including ERV-9 lncRNAs transcribed from many of the 4000 copies of the ERV-9 LTRs may act in cis to stabilize assembly of the LTR enhancer complex and modulate a cis mechanism of long-range LTR enhancer function in activating transcription of downstream genes critical to cellular processes including embryogenesis, erythropoiesis and carcinogenesis.

ACCESSION NUMBERS

The RNA-seq and ChIRP-seq data have been submitted to GEO under accession numbers GSE93280 and GSE9328.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr R. Bollag of GRU Tumor Bank for human peripheral blood mononuclear cells, Dr M. Socolovsky for advice on culturing mouse FL erythroblasts, Dr C.S. Chang for performing whole-genome deep sequencing and Mr Robbie Smith for computer graphics in Figure 6D and Dr D. Browning for critical reading of the manuscript.
Author contributions: T.H. designed and carried out the experiments and analyzed the results and participated in writing the manuscript; W.P. and X.Z. participated in design and execution of 3C and ChIP; H.H. designed and performed pilot experiments on ERV-9 lncRNA KD; M.Y. performed FACS analysis; H.S. participated in designing ChIRP-seq and interpretation of the data; J.C. performed bioinfomatic analysis of RNA-seq and ChIRP-seq data; D.T. designed the experiments and wrote the manuscript.

FUNDING

National Institutes of Health [HL089519]. Funding for open access charge: Augusta University Extramural success award (ESA) Award.
Conflict of interest statement. None declared.

REFERENCES

1. Henikoff,S., Greene,E.A., Pietrokovski,S., Bork,P., Attwood,T.K. and Hood,L. (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science*, **278**, 609–614.
2. 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.
3. Sela,N., Mersch,B., Gal-Mark,N., Lev-Maor,G., Hotz-Wagenblatt,A. and Ast,G. (2007) Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol.*, **8**, R127.
4. Kelley,D.R. and Rinn,J.L. (2012) Transposable elements reveal a stem cell specific class of long noncoding RNAs. *Genome Biol.*, **13**, R107.
5. Jern,P. and Coffin,J. (2008). Effects of retrotransposons on host genome function. *Annu. Rev. Genet.*, **42**, 709–732.
6. Lower,R., Lower,J. and Kurth,R. (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci.*, **93**, 5177–5184.
7. Temin,H.M. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. *Cell*, **27**, 1–3.
8. Yu,X., Zhu,X., Pi,W., Ling,J., Ko,L., Takeda,Y. and Tuan,D. (2005) The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. *J. Biol. Chem.*, **280**, 35184–35194.
9. Pi,W., Yang,Z., Wang,J., Ruan,L., Yu,X., Ling,J., Krantz,S., Isales,C., Conway,S.J., Lin,S. *et al.* (2004) The LTR enhancer of ERV-9 human endogenous retrovirus is active in oocytes and progenitor cells in transgenic zebrafish and humans. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 805–810.
10. Long,Q., Bengra,C., Li,C., Kutlar,F. and Tuan,D. (1998) A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5′ boundary area of the human beta-globin locus control region. *Genomics*, **54**, 542–555.
11. Plant,K.E., Routledge,S.J. and Proudfoot,N.J. (2001) Intergenic transcription in the human beta-globin gene cluster. *Mol. Cell. Biol.*, **21**, 6507–6514.
12. Ling,J., Pi,W., Bollag,R., Zeng,S., Keskintepe,M., Saliman,H., Krantz,S., Whitney,B. and Tuan,D. (2002) The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. *J. Virol.*, **76**, 2410–2423.
13. Zhu,X., Ling,J., Zhang,L., Pi,W., Wu,M. and Tuan,D. (2007) A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res.*, **35**, 5532–5544.
14. Pi,W., Zhu,X., Wu,M., Wang,Y., Fulzele,S., Eroglu,A., Ling,J. and Tuan,D. (2010) Long-range function of an intergenic retrotransposon. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12992–12997.
15. Wilkinson,D.A., Mager,D.L. and Leong,J.C. (1994) Endogenous human retroviruses. In: Levy,J (ed). *The Retroviridae*, Plenum Press, NY, pp. 465–535.
16. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
17. Chu,C., Qu,K., Zhong,F.L., Artandi,S.E. and Chang,H.Y. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell*, **44**, 667–678.
18. Castel,S.E. and Martienssen,R.A. (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.*, **14**, 100–112.
19. Jackson,A.L. and Linsley,P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat. Rev. Drug. Discov.*, **9**, 57–67.

20. Tuan,D., Kong,S.M. and Hu,K. (1992) Transcription of the HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 11219–11223.
21. Dekker,J. (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat. Methods*, **3**, 17–21.
22. Cerase,A., Pintacuda,G., Tattermusch,A. and Avner,P. (2015) Xist localization and function: new insights from multiple levels. *Genome Biol.*, **16**, 166.
23. van Bemmel,J.G., Mira-Bontenbal,H. and Gribnau,J. (2016) Cis- and trans-regulation in X inactivation. *Chromosoma*, **125**, 41–50.
24. Wang,K.C., Yang,Y.W., Liu,B., Sanyal,A., Corces-Zimmerman,R., Chen,Y., Lajoie,B.R., Protacio,A., Flynn,R.A., Gupta,R.A. *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, **472**, U120–U158.
25. Morgan,H.D1, Sutherland,H.G., Martin,D.I. and Whitelaw,E. (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.*, **23**, 314–318.
26. Vasicek,T.J., Zeng,L., Guan,X.J., Zhang,T., Costantini,F. and Tilghman,S.M. (1997) Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics*, **147**, 777–786.
27. Yang,L., Froberg,J.E. and Lee,J.T. (2014) Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem. Sci.*, **39**, 35–43.
28. Paralkar,V.R., Mishra,T., Luan,J., Yao,Y., Kossenkov,A.V., Anderson,S.M., Dunagin,M., Pimkin,M., Gore,M., Sun,D. *et al.* (2014) Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*, **123**, 1927–1937.
29. Alvarez-Dominguez,J.R., Hu,W., Yuan,B., Shi,J., Park,S.S., Gromatzky,A.A., van Oudenaarden,A. and Lodish,H.F. (2014) Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood*, **123**, 570–581.
30. Batista,P. and Chang,H.Y. (2013) Long noncoding RNAs: cellular address codes in development and disease. *Cell*, **152**, 1298–1307.
31. Gibb,E.A., Brown,C.J. and Lam,W.L. (2011) The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer*, **10**, 38.
32. Macfarlan,T.S., Gifford,W.D., Driscoll,S., Lettieri,K., Rowe,H.M., Bonanomi,D., Firth,A., Singer,O., Trono,D. and Pfaff,S.L. (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**, 57–63.
33. Fort,A., Hashimoto,K., Yamada,D., Salimullah,M., Keya,C.A., Saxena,A., Bonetti,A., Voineagu,I., Bertin,N., Kratz,A. *et al.* (2014) Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.*, **46**, 558–566.
34. Ohnuki,M., Tanabe,K., Sutou,K., Teramoto,I., Sawamura,Y., Narita,M., Nakamura,M., Tokunaga,Y., Nakamura,M., Watanabe,A. *et al.* (2014) Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 12426–12431.
35. Goke,J., Lu,X., Chan,Y.S., Ng,H.H., Ly,L.H., Sachs,F. and Szczerbinska,I. (2015) Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*, **16**, 135–141.