

Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency

Marc Friedli,¹ Priscilla Turelli,¹ Adamandia Kapopoulou,¹ Benjamin Rauwel,¹ Nathaly Castro-Díaz,¹ Helen M. Rowe,^{1,5} Gabriela Ecco,¹ Carmen Unzu,² Evarist Planet,¹ Angelo Lombardo,³ Bastien Mangeat,⁴ Barbara E. Wildhaber,² Luigi Naldini,³ and Didier Trono¹

¹School of Life Sciences and "Frontiers in Genetics" National Program, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; ²Pediatric Surgery Laboratory, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland; ³San Raffaele Telethon Institute for Gene Therapy and Vita Salute San Raffaele University, 20132 Milan, Italy; ⁴Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland

Endogenous retroelements (EREs) account for about half of the mouse or human genome, and their potential as insertional mutagens and transcriptional perturbators is suppressed by early embryonic epigenetic silencing. Here, we asked how ERE control is maintained during the generation of induced pluripotent stem cells (iPSCs), as this procedure involves profound epigenetic remodeling. We found that all EREs tested were markedly up-regulated during the reprogramming of either mouse embryonic fibroblasts, human CD34⁺ cells, or human primary hepatocytes. At the iPSC stage, EREs of some classes were repressed, whereas others remained highly expressed, yielding a pattern somewhat reminiscent of that recorded in embryonic stem cells. However, variability persisted between individual iPSC clones in the control of specific ERE integrants. Both during reprogramming and in iPSCs, the up-regulation of specific EREs significantly impacted on the transcription of nearby cellular genes. While transcription triggered by specific ERE integrants at highly precise developmental stages may be an essential step toward obtaining pluripotent cells, the broad and unspecific unleashing of the repetitive genome observed here may contribute to the inefficiency of the reprogramming process and to the phenotypic heterogeneity of iPSCs.

[Supplemental material is available for this article.]

The forced expression of a combination of transcription factors such as POU5F1 (also known as OCT4), KLF4, and SOX2 can result in the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka 2006; Takahashi et al. 2007). The efficiency of this process varies according to the cells chosen as starting material and the protocols used for their modification, but without further manipulation it does not exceed a few percent and it implies a latency of 1 wk to several weeks, suggesting that a cascade of events, some of which are probably stochastic, is required for full reprogramming (Jaenisch and Young 2008; Hanna et al. 2009; Yamanaka 2009). During this period, a complex sequence of still incompletely characterized epigenetic changes takes place, whereby the expression of pluripotency genes is ultimately induced whereas that of differentiation genes is repressed (Koche et al. 2011; Polo et al. 2012).

The development of a totipotent fertilized egg into a nascent embryo is a paradigm for the reverse process. It too stems from epigenetic mechanisms, which are essential not only for the establishment of specialized lineages but also for events such as imprinting and, most importantly, for the silencing of endogenous retroelements (EREs). EREs collectively account for more than half of the genome of either humans or mice, with thousands to millions

of copies of endogenous retroviruses (ERVs), long interspersed elements (LINEs), short interspersed elements (SINEs; which in humans include *Alu* repeats), or SVAs (SINE-VNTR-*Alu*, a hominoid-specific ERE family) (de Koning et al. 2011). These genetic invaders, which multiply by the copy-and-paste process that defines retrotransposons, are targeted during the first few days of embryogenesis by silencing mechanisms notably involving their recognition by sequence-specific protein- or RNA-based repressors and the secondary recruitment of heterochromatin-inducing complexes (Rowe and Trono 2011). Histone methylation, histone deacetylation, and DNA methylation ensue, which inactivate their potential as insertional mutagens (Kaer and Speek 2013; Shukla et al. 2013) and repress their *cis*-acting transcriptional components, which would otherwise activate neighboring genes via promoter or enhancer effects (Bourque et al. 2008; Kunarso et al. 2010; Macfarlan et al. 2011; Rebollo et al. 2012; Schmidt et al. 2012; Chuong et al. 2013; Rowe et al. 2013).

The epigenetic control of EREs is a rigorously orchestrated process, with some of these elements never expressed, and others transcribed in low-cellularity embryos or in embryonic stem cells (ESCs) to be silenced only later (Lane et al. 2003; Macfarlan et al. 2011). It has recently emerged that transient transcription driven

⁵Present address: Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London, London WC1E 6BT, UK

Corresponding author: didier.trono@epfl.ch

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.172809.114>.

© 2014 Friedli et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

by specific coopted ERE integrants that briefly evade repression seems critical to the identity of the pluripotent state both in mice (Macfarlan et al. 2012) and humans (Santoni et al. 2012; Lu et al. 2014). This layered repression process likely reflects at least in part the orderly action of cognate repressors, including KRAB-containing zinc finger proteins (KRAB-ZFPs), which together with their cofactor TRIM28 (also known as KAP1 or TIF1B) are key to the early embryonic control of a broad spectrum of retrotransposons (Wolf and Goff 2007; Matsui et al. 2010; Rowe et al. 2010). Interestingly, KRAB-ZFPs, encoded in the hundreds by the genomes of both mouse and human, are widely expressed in ESCs but subsequently adopt highly tissue-, stage-, and cell-specific patterns of expression (Barde et al. 2013; Corsinotti et al. 2013). Whether the same holds true for other yet-to-be-identified early embryonic controllers of EREs is unknown. However, it is likely that any lag between the de-repression of specific ERE integrants and the reactivation of their sequence- or class-specific repressors will open a window of opportunity for ERE-originating transcriptional perturbations and, at least for the small fraction of these elements still endowed with retrotransposition ability (Finnegan 2012), for insertional mutagenesis. Accordingly, the present study was undertaken to examine how the transcriptional control of EREs is maintained during the reprogramming of either murine or human somatic cells, and whether it is fully reestablished in induced pluripotent stem cells.

Results

Global ERE de-repression during the reprogramming of mouse and human cells

In order to assess the control of EREs during reprogramming to pluripotency, we transduced mouse embryonic fibroblasts (MEFs) harboring a *Pou5f1*-GFP transgene with a lentiviral vector expressing *POU5F1* (*OCT4*), *KLF4*, and *SOX2* (hereafter called OKS) as a single polycistronic transcript from a spleen focus forming virus (SFFV) promoter (Supplemental Fig. S1A; Pasi et al. 2011). As expected, OKS-transduced MEFs formed GFP-positive colonies after ~12 d (Supplemental Fig. S1B) and silenced the OKS vector (Supplemental Fig. S1C). We picked and replated a series of iPSC clones and verified that they expressed *Nanog* at similar levels as ESCs (Supplemental Fig. S1D) and could be differentiated into embryoid bodies (EBs) (Pasi et al. 2011), indicating successful reprogramming. We then used real-time quantitative PCR (RT-QPCR) to measure the expression of families of EREs at various times of the reprogramming process and in the resulting iPSCs. We consistently observed marked increases in the transcript levels of all tested EREs in independent reprogramming experiments (Fig. 1A–D). The timing of the up-regulation was variable, usually initiating 6–10 d post-transduction with a burst of expression occasionally observed around day 2. In all iPSC clones tested, LINE1 and the ERV MusD remained highly expressed, whereas IAP (intracisternal A particle, another ERV) exhibited a fully repressed state, so that for these families of retroelements expression patterns were roughly comparable between iPSC and ESC, as previously noted (Wissing et al. 2012). This is consistent with a model whereby the *trans*-acting factors controlling IAPs, but not those responsible for silencing LINE1 and MusD, are reactivated during the late stages of reprogramming. MERV1, another ERV, exhibited little change in expression during reprogramming itself, but rose sharply in iPSCs, where its levels were markedly above those measured in ESCs. Rather than a defect in controlling factors, this could reflect higher fractions of cells cycling into an early post-zygotic-like state where the MERV1 long terminal repeat (LTR), which serves as

promoter for many genes restricted to the 2/4 cell stage of embryonic development, is particularly active (Macfarlan et al. 2012).

We next assessed the control of EREs during the reprogramming of human somatic cells. Upon transduction of cord blood CD34⁺ cells with the OKS vector, iPSC clones were efficiently obtained (Supplemental Fig. S1B), which exhibited a morphology comparable to that of ESCs, expressed a full set of pluripotency genes, induced the formation of teratomas when injected into immunodeficient mice, and were karyotypically euploid (Supplemental Fig. S1E–H). However, de-repression of all tested human ERE families was detected by RT-QPCR during the reprogramming process, with sharp increases around day 19 post-transduction across three independent experiments performed with cells from different donors (Fig. 1E,H; Supplemental Fig. S2). Similar to their murine counterparts, human LINE1s remained highly expressed in pluripotent cells, as did SVAs (Fig. 1E,F). In contrast, the LTR-containing HERVK, including its HERVK14ci strain, was silenced in iPSCs, mimicking the behavior of IAPs in the murine system (Fig. 1G,H).

To confirm and extend these observations, we performed RNA deep sequencing (RNA-seq) at multiple time points of an independent reprogramming experiment, including in our analysis the parental CD34⁺ cells, six of the resulting iPSC clones, and a human ESC clone (H1) (for RNA-seq data, see Supplemental Tables 1–3, and GEO accession number GSE57866). A comparison of the transcriptomes of cells harvested at day 19 post-transduction with that of the starting population confirmed the de-repression of multiple EREs, and further identified specific subclasses of dysregulated retrotransposons (Fig. 2A), with HERVH, HERVK, and their associated LTRs LTR7Y and LTR5-Hs displaying the most pronounced up-regulation (Fig. 2B). In iPSCs, some of these elements, such as HERVK and S71 (LTR6b), displayed a repressed state comparable to that found in ESCs, but others (e.g., HERVW, LTR17, LTR7Y, L1Hs) did not or did only partially (Fig. 2B,C). Of note, while significant expression of HERVH was detected in iPSCs and ESCs, as previously reported (Santoni et al. 2012), close to a third of the HERVH integrants up-regulated at day 19 were not among those ultimately detected in these cells (data not shown). Finally, in addition to the d19 peak of expression, we occasionally observed a much earlier burst around d2–d4 for some retroelements (e.g., HERVH and HERVK) (Fig. 2C). We further RNA-sequenced eight independent human ES cell lines (UCLA1-6 [Diaz Perez et al. 2012], H19, and an independent H1 sample) to verify that the elevated ERE expression levels detected in individual iPSC clones were not a general feature of pluripotent cells. We found very little heterogeneity between all tested human ES cell lines, which sharply contrasted with the inappropriate control of EREs in iPSC clones and even more with their marked up-regulation during the reprogramming process (Fig. 2C). To ascertain that the observed unleashing of EREs during reprogramming is not restricted to fibroblasts or blood cells, we assessed ERE expression during reprogramming of human liver cells. We thus reprogrammed primary human hepatocytes with OKS or with a vector that also included MYC (OKSM) (Sommer et al. 2010) and measured ERE-specific transcription by real time quantitative PCR. Confirming the generality of this phenomenon, all ERE classes recapitulated the burst of expression during reprogramming previously observed in MEFs and CD34⁺ cells (Fig. 2D).

Deregulation of ERE-close gene transcripts during reprogramming

We next asked whether the up-regulation of ERVs that occurred during reprogramming impacted the expression of nearby genes.

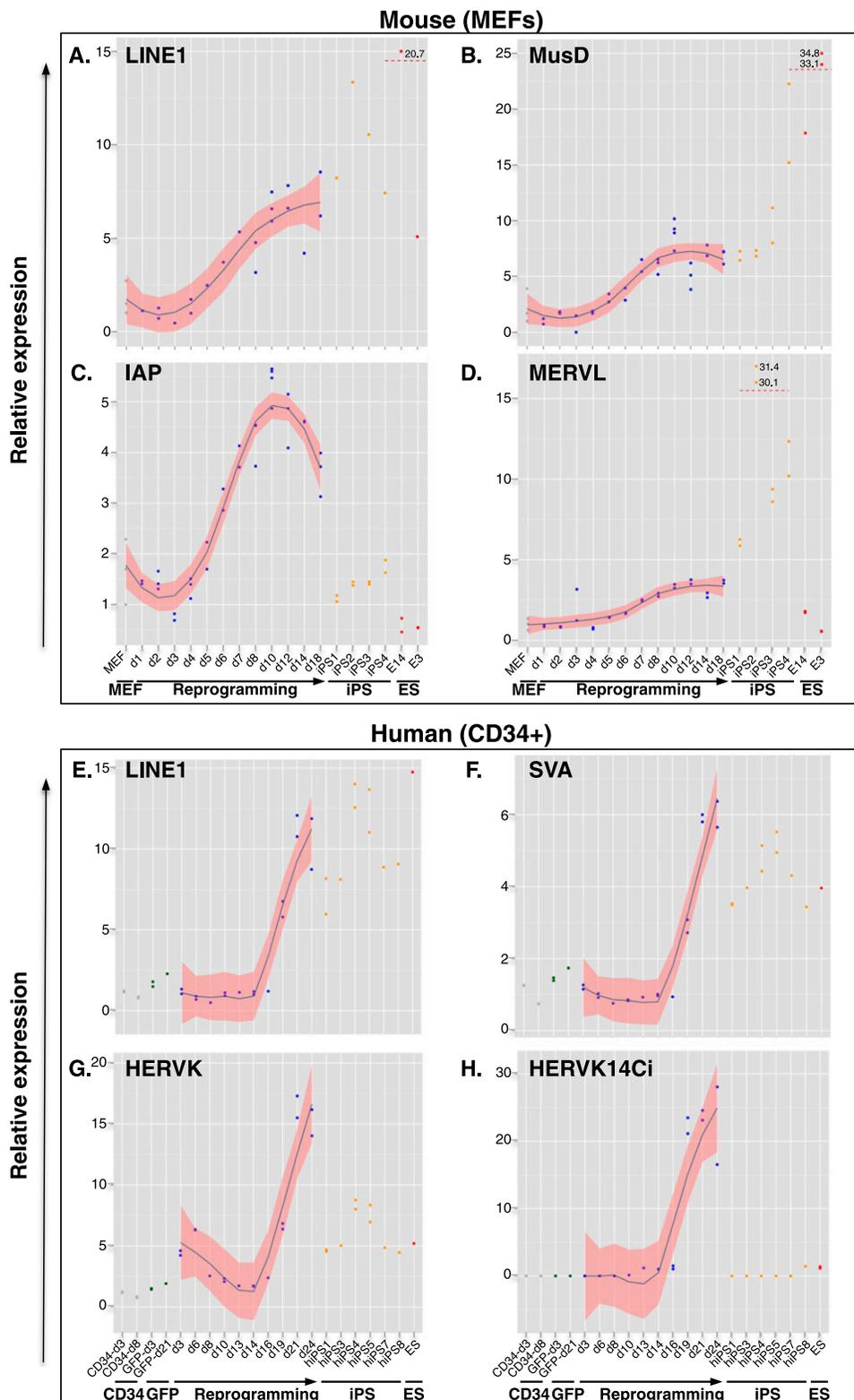


Figure 1. Global up-regulation of EREs during reprogramming. qPCR quantification of transcripts from indicated murine (A–D) and human (E–H) EREs during reprogramming of MEFs and cord blood CD34⁺ cells, respectively. Expression levels are indicated for parental cells (gray dots), cells transduced with an SFFV-GFP control vector (green dots), OKS-induced reprogramming time points (blue dots), individual iPS clones (orange dots), and ES cells (red dots). For the smoothing pattern across OKS-induced reprogramming time points, we computed a locally weighted polynomial regression (LOESS) with a 95% confidence interval.

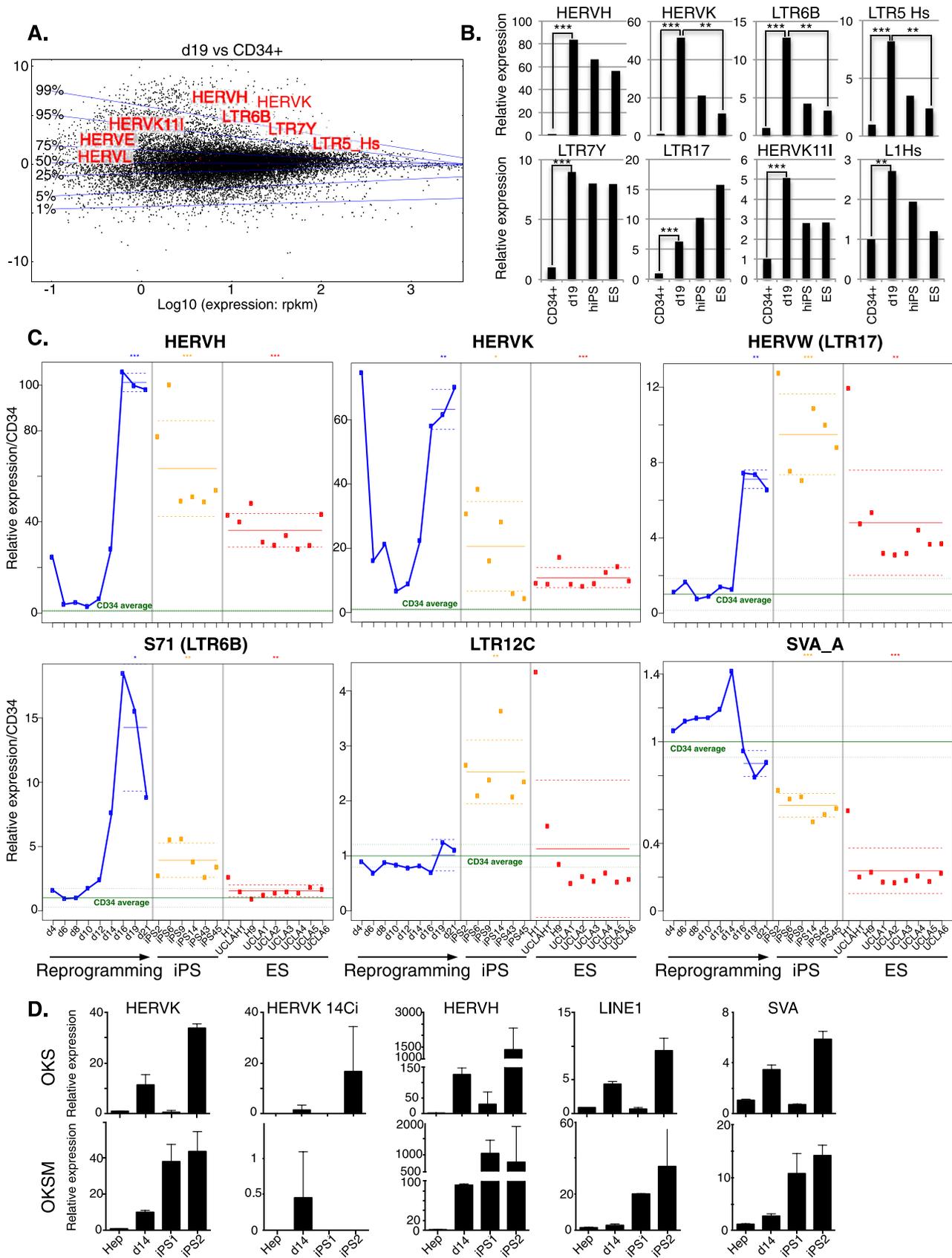


Figure 2. (Legend on next page)

Using a twofold change cutoff and an adjusted P -value of <0.05 , we identified 3703 genes up-regulated at d19 in OKS-transduced CD34⁺ cells compared with nontransduced cells. Genes near up-regulated HERVH (<40 kb between TSS and HERV, $n = 365$ genes) were far more likely ($P = 2.603 \times 10^{-05}$) to be up-regulated than genes more distant from these elements (>80 kb between TSS and HERV, $n = 18,145$ genes) (Fig. 3A). A similar analysis could not be performed for HERVK due to insufficient numbers. However, two additional lines of evidence confirmed that it was the HERV that influenced the gene and not the reverse. First, for both HERVH and HERVK, expression was not influenced by distance from an up-regulated gene (Fig. 3A). Second, intragenic HERVs were not more often up-regulated than their intergenic counterparts (Fig. 3B); in fact, the reverse trend was observed ($P = 0.024$ for HERVK and $P = 4.26 \times 10^{-10}$ for HERVH).

Heterogeneity of human iPSC clones in repression of specific EREs and induction of prototypic ERE controllers

Importantly, when we compared levels of expression of individual EREs between iPSC clones derived from a single donor and issued from the same reprogramming experiment, we noticed striking differences, notably for HERVH, HERVK, and L1Hs (e.g., cf. clone 6 and clone 43 or cf. clone 2 and clone 45) (Figs. 2C, 4A). HERVK, for instance, was fivefold to ninefold more expressed in iPSC and iPSC6 compared with iPSC43, iPSC45, and human ES cells (Fig. 4A, top), whereas L1Hs was threefold more expressed in iPSC2 than in iPSC14. Interestingly, several members of the KRAB-ZFP gene family (e.g., ZNF492, ZNF649, ZNF208) exhibited marked differences in expression between iPSC clones (Supplemental Figs. S3A, S4C). A direct comparison of the transcriptomes of two iPSC clones confirmed that specific EREs and some KRAB-ZFPs were among the most discordant transcripts, suggesting that a failure to reactivate sequence-specific repressors during reprogramming might account for the lack of silencing of their target EREs in iPSCs (Supplemental Fig. S3B). Consistent with this hypothesis, *Trim28* expression increased gradually during reprogramming, and *Trim28* knockout MEFs failed to reprogram (data not shown). Noteworthy, known post-transcriptional controllers of retroelements (e.g., APOBEC3A and SAMHD1) (Bogerd et al. 2006; Hrecka et al. 2011; Laguette et al. 2011) were transiently induced during reprogramming (Supplemental Fig. S3A), indicating the activation of at least some genome defense mechanisms along this process.

Incomplete control of specific EREs activates the transcription of nearby genes in iPSCs

Upon scoring the expression of specific ERE integrants, we again detected considerable heterogeneity between iPSC clones (Fig. 4A,

bottom). For example, a HERVH located on chromosome 8 was four- to fivefold more expressed in iPSC6 compared with other iPSC clones and to human ES, while a HERVK inserted on chromosome 22 was four- to eightfold more expressed in clones 2, 6, and 14 compared with clones 43, 45, and human ESCs (Fig. 4A, bottom). EREs can epigenetically affect the local genomic landscape owing to their content in a variety of *cis*-acting regulatory sequences (Bourque et al. 2008; Kunarso et al. 2010; Rebollo et al. 2012; Schmidt et al. 2012; Chuong et al. 2013; Rowe et al. 2013). Correspondingly, we found numerous instances where lack of repression of a specific ERE was accompanied by up-regulation of the adjacent gene (Fig. 4B,C; Supplemental Fig. S4A,B). For example, the *HHLA1* gene, situated next to the above-mentioned chromosome-8 HERVH, was significantly expressed only in iPSC6 (Spearman correlation $P = 0.93$, $P = 0.007$), while expression of the *PRODH* gene correlated perfectly with that of the adjacent chromosome-22 HERVK among iPSC clones (Spearman correlation $P = 0.96$, $P = 0.003$) (Fig. 4B,C). Likewise, expression of *KLKB1* and *C9orf129* paralleled that of HERVs situated nearby (Supplemental Fig. S4A,B). Interestingly, with the exception of the HERV near *PRODH* (for which mapping of ChIP-seq reads was not possible), all these EREs were previously identified as bound by TRIM28 and adorned with the H3K9me3 repressive mark in human ESCs (Fig. 4B,C; Supplemental Fig. S4A,B; data not shown). Noteworthy, when differentiated to EBs or neural committed, some ERE de-repressed iPSC clones regained control of specific integrants while others remained uncontrolled (Supplemental Fig. S5). For example, the ERE near *PRODH* in iPSC6 was still locked in the highly expressed state after neural commitment even though expression of *PRODH* itself was reduced about threefold (Supplemental Fig. S5).

To further explore this phenomenon, we examined the state of the chromatin at several of these loci by chromatin immunoprecipitation followed by quantitative PCR (ChIP-PCR). We found histone marks typical of active enhancers (H3K4me1 and H3K27ac) near the LTRs of the HERVs situated next to *HHLA1* and *PRODH* in iPSCs exhibiting an up-regulation of the corresponding ERE–gene pairs, but not in clones where these units were fully repressed (Fig. 4D). Conversely H3K9me3, a repressive mark that is a hallmark of KRAB/TRIM28-mediated silencing (Schultz et al. 2002), displayed the opposite pattern at these loci.

Discussion

These data demonstrate that the reprogramming of somatic cells to induced pluripotent stem cells is accompanied by a marked de-repression of endogenous retroelements from all known classes. It has recently been suggested that transcription from EREs is important to drive ES-specific transcripts, in particular MERVL in

Figure 2. De-repression of individual EREs during reprogramming and in iPSCs. (A) MA-plot comparing RNA-seq-determined transcriptome of day 19 (d19) OKS-transduced vs. parental CD34⁺ cells. Transcripts (RefSeq) are plotted in black with the ratio (d19/CD34⁺) on the y-axis and expression levels on the x-axis. Representative up-regulated Repbase families are shown in red. Transversal blue lines depict magnitude of gene deregulation (e.g., only 1% of genes lie above the 99% line). (B) Expression levels of indicated HERVs in parental (average of three samples) or d19 OKS-transduced CD34⁺ cells, human iPSC cells (average of six clones from the same reprogramming experiment), and the H1 ES cell line (average of two samples). Fold changes compared with CD34⁺ triplicates and P -values are calculated using the DESeq package (Anders and Huber 2010). (C) Relative expression of indicated HERVs during reprogramming of CD34⁺ cells, in six resulting iPSC clones (orange dots) and nine independent samples of hES cells (red dots). Green line indicates parental cells (average of three, same donor); dotted green lines, plus and minus standard deviation; and solid blue line, reprogramming time points. The horizontal (blue, orange, and red) solid lines show the mean of each group of samples. Dotted lines show a 95% confidence interval for each mean. We performed a Wilcoxon test for each mean to test if it was different from one. (D) qPCR quantification of transcripts from indicated human EREs during reprogramming of primary hepatocytes. (Top) Reprogramming with OKS. (Bottom) Reprogramming with OKSM. (Hep) Average of four nontransduced hepatocyte samples.

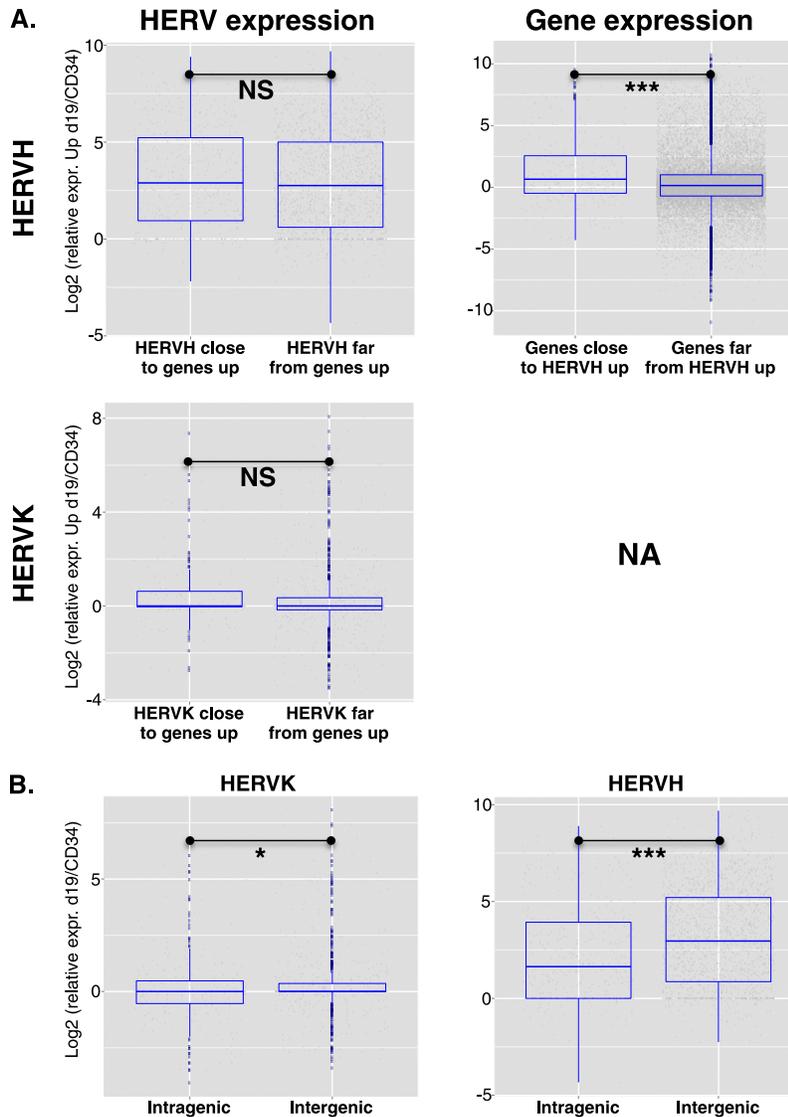


Figure 3. Up-regulation of ERE-close genes during reprogramming. (A) Using a twofold change cutoff and an adjusted P -value of <0.05 , 3703 genes were found up-regulated in CD34⁺ cells at day 19 post-OKS transduction, compared with untransduced cells. (Left) Relative expression of HERVH (top) or HERVK (bottom) integrants situated close (<40 kb) or far (>80 kb) from these genes. No significant difference was detected (Wilcoxon test $P = 0.14$ for HERVK, 0.11 for HERVH). (Right) Reverse analysis, revealing that HERVH-close transcripts ($n = 365$ genes, or 695 transcripts) were more likely to be up-regulated than HERVH-distant transcripts ($n = 18145$ genes, or $30,053$ transcripts). (NA) Not available due to insufficient numbers. (B) Expression levels of intragenic vs. intergenic HERVHs or HERVKs at d19 of CD34⁺ cells reprogramming showing no bias toward intragenic HERVs. In fact, the reverse trend was observed, with more intergenic than intragenic HERVs up-regulated at d19 of reprogramming.

murine cells (Macfarlan et al. 2012) and HERVH in human cells (Santoni et al. 2012; Lu et al. 2014). However, while we indeed found numerous HERVH integrants up-regulated in human ES cells compared with cord blood CD34⁺ cells, we also noted that more than a third of the members of this group found to be activated at day 19 of reprogramming were fully repressed in ESCs or iPSCs. Thus the possible requirement of HERVH-mediated transcription during iPS reprogramming and to maintain the pluripotent state comes at a price, as control is broadly released on this family of elements. Furthermore, HERVKs, including HERVK14ci, were markedly up-regulated during reprogramming but fully

repressed in pluripotent cells. For most EREs, activation reached its peak shortly before the emergence of fully reprogrammed iPSCs. Our results suggest that this phenomenon might be due at least partly to a lack of synchronization between erasure of repressive chromatin marks at EREs on the one hand and reactivation of sequence-specific *trans*-repressors, for instance, KRAB-ZFPs, on the other hand. Importantly, at that stage, genes situated near up-regulated EREs had a greater chance of being themselves induced, consistent with ERE-based promoter or enhancer effects (Rowe et al. 2013). It could be that this contributes to the inefficiency of reprogramming, if it results in the stochastic activation of genes affecting the path to pluripotency (Polo et al. 2012). It will be interesting to determine whether ERE activation also occurs when reprogramming efficiency is increased by depletion of the MBD3 repressor (deterministic reprogramming) (Rais et al. 2013), by coexpression of CEBPA (Di Stefano et al. 2014), or following nuclear transfer. These faster reprogramming methods may be accompanied by timely reactivation of cognate KRAB-ZFPs, which may minimize aberrant ERE reactivation. ERE de-repression is predicted to result only rarely in mutagenic transposition events, owing both to the paucity of retroelements still endowed with transposition capacity (Quinlan et al. 2011; Finnegan 2012) and to the presence of restriction factors blocking their spread at a post-transcriptional level (Wolf and Goff 2008). However, the transcriptional perturbation of ERE-close genes may confer iPSCs or their progeny with phenotypic anomalies difficult to detect through conventional assays, such as blockade of differentiation to particular lineages, predisposition to oncogenic changes, aberrant release of bioactive molecules, or altered immunogenicity. Supporting this model, a recent comparative analysis of 49 iPSC lines derived from several human tissues detected an aberrant up-regulation of some LTR7/HERVH transcripts and neighboring genes, including *HHLAI1*, in several differentiation-defective clones (Koyanagi-Aoi et al. 2013). Furthermore, the potential for more distal phenotypic anomalies resulting from inappropriate ERE-induced gene activation is illustrated by the recent observation that the schizophrenia-linked *PRODH* gene (Kempf et al. 2008) is controlled by the nearby HERVK (Suntsova et al. 2013), which we found deregulated in some iPSC clones. Our findings thus warrant an in-depth survey of the genomic, transcriptional, and epigenetic state of the repetitive genome of iPSC clones, whether these are to be used for basic research or are envisioned for clinical applications.

Methods

Reprogramming

MEFs: Primary *Pou5f1*-GFP MEFs were prepared from E12 embryos (<http://jaxmice.jax.org/strain/008214.html>) and reprogrammed by transduction with OKS using four HCT116-transducing units (HC-TU) per cell as previously described (Pasi et al. 2011). CD34⁺ cells from human cord blood were obtained and prepared as previously described (Barde et al. 2013), before transducing 250,000 cells with OKS using 100 HC-TU per cell. After 5 d, cells were switched to mTeSR1 medium (Stemcell Technologies no. 05859) and grown on a mouse fibroblast feeder layer until reprogrammed colonies emerged (~21 d). Individual human iPSC clones were then picked and expanded. Primary human hepatocytes were isolated from liver biopsies as previously described (Birraux et al. 2009) and plated on Matrigel or collagen before being transduced with OKS or OKSM using 20 HC-TU per cell. After 5 d, cells were switched to mTeSR1 medium (Stemcell Technologies no. 05859) and grown until reprogrammed colonies emerged (~25 d).

qPCR

Total RNA from cells at different reprogramming time points, iPSC clones, and ES cells was extracted with TRIzol reagent (Life Technologies no. 15596-018) and PureLink micro-to-midi total RNA Purification System (Life Technologies no.12183018). cDNA was prepared with SuperScript II reverse transcriptase and real-time PCR quantification was performed with FastStart Universal SYBR Green Master (Rox; Roche no. 04913914001). Normalization was done to two or three housekeeping genes (mouse: *Gapdh*, *Cox6a1*, *Tfrc*; human: *TFRC*, *B2M*).

Chromatin immunoprecipitation (ChIP)-PCR

Ten million iPSCs or ESCs were immunoprecipitated as previously described (Barde et al. 2013), with TRIM28- (Abcam ab10483), H3K9me3- (Diagenode pAb-056-050), H3K27ac- (Abcam ab4729), or H3K4me1- (Diagenode pAb-037-050) specific antibodies. SYBR green qPCR was performed to quantify enrichment at specific loci.

RNA-seq

RNA-seq was performed on an Illumina HiSeq 2000 sequencer (single-read 100-cycles assay). The library was generated from 250 ng total RNA using the TruSeq RNA Sample Preparation v2 kit (Illumina). Raw reads (100-bp single-end) were mapped to the human transcriptome (RefSeq), to the human genome (hg19 assembly), and to Repbase consensus sequences using the Bowtie short-read aligner (Langmead et al. 2009) and allowing up to three mismatches, and counts were normalized to the transcript length and to the total number of reads (RPKM). Differentially expressed RefSeq transcripts and ERs were defined using the DESeq Bioconductor package (Anders and Huber 2010).

Data access

The RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE57866.

Acknowledgments

We thank S. Offner, S. Verp, and C. Raclot for technical assistance; the staff of the Lausanne University Hospital delivery room for help in obtaining cord blood; the University of Lausanne Genomics core

facility for sequencing; the EPFL histology core facility, Alessandra Piersigilli, and Anne-Laure Rougemont-Pidoux for histological sections and analyses; and Vital-IT for computing. This work was financed through grants from the Swiss National Science Foundation to D.T., from the Louis Jeantet Foundation and the Novartis Consumer Health Foundation to B.E.W., from Fondazione Telethon to L.N., and from the European Research Council to D.T and L.N. (ERC 268721, KRAB'n'KAP, and ERC 249845, TARGETINGGENETHERAPY, respectively).

References

- Anders S, Huber W 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.
- Barde I, Rauwel B, Marin-Florez RM, Corsinotti A, Laurenti E, Verp S, Offner S, Marquis J, Kapopoulou A, Vanicek J et al. 2013. A KRAB/KAP1-miRNA cascade regulates erythropoiesis through stage-specific control of mitophagy. *Science* **340**: 350–353.
- Birraux J, Menzel O, Wildhaber B, Jond C, Nguyen TH, Chardot C. 2009. A step toward liver gene therapy: efficient correction of the genetic defect of hepatocytes isolated from a patient with Crigler-Najjar syndrome type 1 with lentiviral vectors. *Transplantation* **87**: 1006–1012.
- Boger HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. 2006. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* **34**: 89–95.
- Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**: 1752–1762.
- Chuong EB, Rumi MA, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**: 325–329.
- Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM, Mouscaz Y, Deplancke B, Trono D. 2013. Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS ONE* **8**: e56721.
- de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**: e1002384.
- Di Stefano B, Sardina JL, van Oevelen C, Collombet S, Kallin EM, Vicent GP, Lu J, Thieffry D, Beato M, Graf T. 2014. C/EBP α poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**: 235–239.
- Diaz Perez SV, Kim R, Li Z, Marquez VE, Patel S, Plath K, Clark AT. 2012. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum Mol Genet* **21**: 751–764.
- Finnegan DJ. 2012. Retrotransposons. *Curr Biol* **22**: R432–R437.
- Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creighton MP, van Oudenaarden A, Jaenisch R. 2009. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**: 595–601.
- Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**: 658–661.
- Jaenisch R, Young R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**: 567–582.
- Kaer K, Speek M. 2013. Retroelements in human disease. *Gene* **518**: 231–241.
- Kempf L, Nicodemus KK, Kolachana B, Vakkalanka R, Verchinski BA, Egan MF, Straub RE, Mattay VA, Callicott JH, Weinberger DR, et al. 2008. Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. *PLoS Genet* **4**: e1000252.
- Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A. 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**: 96–105.
- Koyanagi-Aoi M, Ohnuki M, Takahashi K, Okita K, Noma H, Sawamura Y, Teramoto I, Narita M, Sato Y, Ichisaka T, et al. 2013. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc Natl Acad Sci* **110**: 20569–20574.
- Kunarto G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have wired the core regulatory network of human embryonic stem cells. *Nat Genet* **42**: 631–634.
- Laguette N, Sobhian B, Casartelli N, Ringard M, Chable-Bessia C, Seregal E, Yatim A, Emiliani S, Schwartz O, Benkirane M. 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**: 654–657.

- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**: 88–93.
- Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21**: 423–425.
- Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, et al. 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* **25**: 594–607.
- Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**: 57–63.
- Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927–931.
- Pasi CE, Dereli-Oz A, Negrini S, Friedli M, Fragola G, Lombardo A, Van Houwe G, Naldini L, Casola S, Testa G, et al. 2011. Genomic instability in induced stem cells. *Cell Death Differ* **18**: 745–753.
- Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al. 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**: 1617–1632.
- Quinlan AR, Boland MJ, Leibowitz ML, Shumilina S, Pehrson SM, Baldwin KK, Hall IM. 2011. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9**: 366–373.
- Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. 2013. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**: 65–70.
- Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. 2012. Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol* **13**: R89.
- Rowe HM, Trono D. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411**: 273–287.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**: 237–240.
- Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. 2013. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* **23**: 452–461.
- Santoni FA, Guerra J, Luban J. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9**: 111.
- Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**: 335–348.
- Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ III. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**: 919–932.
- Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, et al. 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**: 101–111.
- Sommer CA, Sommer AG, Longmire TA, Christodoulou C, Thomas DD, Gostissa M, Alt FW, Murphy GJ, Kotton DN, Mostoslavsky G. 2010. Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. *Stem Cells* **28**: 64–74.
- Suntsova M, Gogvadze EV, Salozhin S, Gaifullin N, Eroshkin F, Dmitriev SE, Martynova N, Kulikov K, Malakhova G, Tukhbatova G, et al. 2013. Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc Natl Acad Sci* **110**: 19472–19477.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–676.
- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861–872.
- Wissing S, Munoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC. 2012. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet* **21**: 208–218.
- Wolf D, Goff SP. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**: 46–57.
- Wolf D, Goff SP. 2008. Host restriction factors blocking retroviral replication. *Annu Rev Genet* **42**: 143–163.
- Yamanaka S. 2009. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**: 49–52.

Received January 21, 2014; accepted in revised form May 23, 2014.