



Published in final edited form as:

Nat Genet. 2016 January ; 48(1): 44–52. doi:10.1038/ng.3449.

The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming

Jens Durruthy-Durruthy^{1,2,3,9}, Vittorio Sebastiano^{1,2,3,9}, Mark Wossidlo^{1,2,3}, Diana Cepeda^{1,2,3}, Jun Cui^{1,2,3}, Edward J Grow², Jonathan Davila⁴, Moritz Mall⁴, Wing H Wong⁵, Joanna Wysocka^{6,7}, Kin Fai Au⁵, and Renee A Reijo Pera^{1,2,3,8}

¹Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California, USA

²Department of Genetics, Stanford University, Stanford, California, USA

³Department of Obstetrics and Gynecology, Stanford University, Stanford, California, USA

⁴Department of Pathology, Stanford University, Stanford, California, USA

⁵Department of Statistics, Stanford University, Stanford, California, USA

⁶Department of Chemical and Systems Biology, Stanford University, Stanford, California, USA

⁷Department of Developmental Biology, Stanford University, Stanford, California, USA

Abstract

Long intergenic noncoding RNAs (lincRNAs) are derived from thousands of loci in mammalian genomes and are frequently enriched in transposable elements (TEs). Although families of TE-derived lincRNAs have recently been implicated in the regulation of pluripotency, little is known of the specific functions of individual family members. Here we characterize three new individual TE-derived human lincRNAs, human pluripotency-associated transcripts 2, 3 and 5 (HPAT2, HPAT3 and HPAT5). Loss-of-function experiments indicate that HPAT2, HPAT3 and HPAT5 function in preimplantation embryo development to modulate the acquisition of pluripotency and the formation of the inner cell mass. CRISPR-mediated disruption of the genes for these lincRNAs

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to R.A.R.P. (renee.reijopera@montana.edu) or V.S. (vsebast@stanford.edu).

⁸Present address: Department of Cell Biology and Neurosciences, Montana State University, Bozeman, Montana, USA.

⁹These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

J.D.-D., V.S., W.H.W., K.F.A. and R.A.R.P. conceived the project, designed experiments and wrote the manuscript, with input from all authors. J.D.-D., V.S. and D.C. performed siRNA knockdown experiments. J.C. designed and tested CRISPR constructs. M.W. conducted the human embryo experiments. E.J.G. performed ChIP experiments. J.D. and M.M. performed RIP and immunoblot experiments. J.W. helped with manuscript writing.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Accession codes. ArrayExpress, E-MTAB-2994; Gene Expression Omnibus (GEO), GSE73725. ChIP-Seq data on NANOG in H9 are deposited in the Gene Expression Omnibus (GEO) database. Microarray data on H9 hESCs are deposited in the ArrayExpress database.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

in pluripotent stem cells, followed by whole-transcriptome analysis, identifies HPAT5 as a key component of the pluripotency network. Protein binding and reporter-based assays further demonstrate that HPAT5 interacts with the let-7 microRNA family. Our results indicate that unique individual members of large primate-specific lincRNA families modulate gene expression during development and differentiation to reinforce cell fate.

Recent studies have catalogued more than 10,000 lincRNAs in the human genome¹⁻⁴ and have found that TEs are present in more than two-thirds of mature lincRNA transcripts⁵, thus contributing to the lineage-specific diversification of vertebrate lincRNA repertoires. The functions of families of lincRNAs, defined by TE class, have been linked to diverse biological processes such as imprinting⁶, dosage compensation^{7,8}, regulation of developmental gene expression^{7,8}, chromatin modification⁹⁻¹¹, and stem cell pluripotency and differentiation in vertebrates¹². However, functional studies of individual lincRNAs remain challenging, in large part owing to the highly repetitive nature of the sequences and low expression levels, in combination with the absence of high-quality transcript annotation models that accurately define the genomic features of lincRNAs, including transcription start sites, splicing, polyadenylation sites and isoform abundance. As a result, TE-derived lincRNAs have been almost exclusively studied as an aggregate class of repetitive elements^{1-5,13-17}. One lincRNA TE class, human endogenous retrovirus-H (HERV-H), has been shown to be required for maintenance of the pluripotent state in human embryonic stem cells (hESCs)¹⁷. More recently, the activity of specific HERV classes, including HERV-H and HERV-K, has also been linked to human preimplantation embryo development^{18,19}. In addition, a recent study posited that hESC-specific TE-derived lincRNAs may not act as a single functional family, despite the sequence similarity of the component members, but instead may function individually to influence diverse physiological pathways²⁰. However, functional data on individual TE-derived lincRNAs are scarce.

We recently used a hybrid RNA sequencing technique to identify more than 2,000 new lincRNA transcript isoforms, of which 146 were specifically expressed in pluripotent hESCs¹³. We identified the 23 most abundantly expressed transcripts, confirmed specificity of expression in pluripotent cells and termed the corresponding genomic loci *HPAT1-HPAT23* (human pluripotency-associated transcripts 1-23). The sequence of one of the HPATs, *HPAT5*, was also described in 1987 (ref. 21). We obtained a consensus sequence of the 856-bp 5'-terminal part of the internal portion of HUERS-P1, an LTR8-containing retrotransposon. Cross-referencing the genomic sequence of *HPAT5* with the genomes of seven distinct primate species (baboon, chimpanzee, gibbon, gorilla, marmoset, orangutan and rhesus macaque) suggested that *HPAT5* is closely related to a genomic location on chromosome 6 in chimpanzee and gorilla, indicating that *HPAT5* was recently introduced into the primate lineage, approximately 5-9 million years ago²². Here we show that *HPAT1-HPAT23* encode TE-derived lincRNAs; that three HPATs (HPAT2, HPAT3 and HPAT5) may modulate cell fate in human preimplantation development; and that the molecular mechanism through which HPAT5 functions in hESCs is mediated via let-7.

RESULTS

HPAT1–HPAT23 gene structure

To further probe the identity and function of *HPAT1–HPAT23*, we began with sequence alignment and found that the majority of the *HPAT1–HPAT23* sequences comprise repetitive elements at the genome and transcript levels (Supplementary Fig. 1a–c), with these elements accounting for an average of 64.8% (range of 15–99%) of the total lincRNA sequence. Upon closer examination, we found that a large proportion of the repetitive sequences were derived from TEs in four major classes: short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeat/endogenous retrovirus (LTR/ERV) elements and DNA transposons. Members of the LTR/ERV class represented the largest fraction of genomic sequences (present in all HPATs; average of 44.6%, range of 4.9–97.9%; Supplementary Table 1). The HERV-H family, as expected, contributed greatly to the sequences of the HPATs (19 of 23 HPATs overlapped with the HERV-H sequence; Supplementary Table 1), as previously observed for other hESC-specific lincRNAs^{14,17,23,24}. Notably, we found that the exons of HPAT genes overlapped with TEs from all four classes, although LTR elements (of the HERV-H subclass) were most common, suggesting that this subclass may contribute most extensively to functional gene features⁵ (Supplementary Fig. 1d). In contrast, protein-coding genes that are highly expressed in hESCs such as *POU5F1* (also known as *OCT4*), *NANOG* and *SOX2*, showed little to no overlap of gene sequences with TE segments. Thus, we found through sequence alignment that all 23 HPAT genes are derived from TEs. Of these HPAT genes, 13 were derived exclusively from HERV-H elements (100% LTR/ERV sequence coverage), six partially aligned with HERV-H elements (51.7–92.8% sequence coverage) and four were derived from TEs other than HERV-H sequences (Supplementary Table 1).

HPAT expression *in vivo* in human embryos

We next profiled HPAT1–HPAT23 expression in single cells of human blastocysts (Fig. 1a). Of all the HPAT transcripts, three—HPAT2, HPAT3 and HPAT5—were expressed specifically in the inner cell mass (ICM) and not in trophoblast, with HPAT5 expressed at the highest levels (Fig. 1b–d). No or very low expression was detected for the remaining HPAT transcripts in human blastocysts (data not shown). In addition, we confirmed expression of HPAT3 and HPAT5 in human blastocysts via RNA FISH; note that the prevalence of repetitive sequences in HPAT2 did not allow RNA FISH of this transcript (Supplementary Fig. 1b). This analysis indicates that both HPAT3 and HPAT5 are expressed predominantly in the ICM ($n = 9$ blastocysts for HPAT3 and $n = 11$ blastocysts for HPAT5; Fig. 1e and Supplementary Fig. 2a), with expression overlapping that of OCT4 protein. Mouse blastocysts ($n = 3$) provided negative controls for both sets of probes.

To determine whether HPAT2, HPAT3 and HPAT5 function in human preimplantation development, we used short interfering RNAs (siRNAs) to reduce the levels of these transcripts. For this purpose, we injected a single blastomere of an embryo at the two-cell stage with a combination of the siRNAs specifically targeting each of the three HPAT transcripts and tracked development over time with time-lapse imaging and confocal microscopy, with reference to the sister ‘control’ blastomere. The knockdown efficiency of

the siRNAs, in comparison to scrambled siRNA used as a control, was tested before injections, with the results indicating a two- to fivefold reduction in transcript levels (Supplementary Fig. 2b). Coinjection with fluorophore- labeled dextran facilitated identification of descendent cells in the developing embryo (Fig. 1f and Supplementary Fig. 2c,d). We observed that blastomeres deficient for HPAT2, HPAT3 and HPAT5 did not contribute to the ICM ($n = 3$), whereas cells that originated from blastomeres injected with scrambled siRNA contributed to both the trophectoderm and ICM ($n = 3$). These data indicate that one or more of the three TE-derived lincRNAs (HPAT2, HPAT3 and HPAT5) likely contribute to formation of the pluripotent ICM. To our knowledge, these data provide the first evidence that lincRNAs may have a fundamental role *in vivo* in human embryogenesis.

Three HPATs are linked to the core pluripotency network

To probe the biological function of HPATs further, we used an *in vitro* assay of nuclear reprogramming to produce induced pluripotent stem cells (iPSCs). We began by examining gene expression during the transition from somatic cells to iPSCs. Studies were performed at single-cell resolution to aid in the identification of gene networks and allow reconstruction of network hierarchies during the establishment and maintenance of pluripotency²⁵. We collected 864 single cells at different time points of reprogramming, from fibroblasts (day 0) through days 2, 5, 7, 10 and 12 to fully established iPSCs (Supplementary Fig. 3a,b). We profiled the expression of 96 transcripts, including HPAT1–HPAT23 and genes expressed in somatic cells and pluripotent stem cells (Supplementary Table 2). Assay validation ensured specificity and reproducibility, with 578 single cells and 82 assays passing quality controls to result in a high-quality single-cell data matrix of 47,396 data points that was used for subsequent comprehensive bioinformatic analyses (Supplementary Fig. 3c–i). Results indicated that the expression of genes linked to pluripotency increased over the course of reprogramming, with expression peaking in fully established iPSCs. Conversely, as expected, reprogramming was accompanied by the silencing of fibroblast markers (Supplementary Fig. 4a,b). Over the course of reprogramming, three patterns of expression for HPAT transcripts were evident: (i) a gradual increase over the duration of reprogramming, (ii) activation in late stages of reprogramming and (iii) expression exclusive to fully reprogrammed iPSCs (Fig. 2a and Supplementary Fig. 4c). Further examination of the data via principal-component analysis (PCA) also demonstrated that two components accounted for the largest proportion (47.5%) of biological variability in the data and, when projected, identified distinct groups of single cells that coincided with the time point of collection. Unsupervised clustering illustrated in a heat map confirmed these observations (Supplementary Fig. 4d,e). Notably, to determine whether expression was simply a consequence of transcriptionally permissive chromatin during periods of transition from one cell fate to another, we examined whether HPAT transcripts were expressed in the transition from fibroblasts to induced neurons. We observed no expression of HPATs in the derivation of induced neurons from fibroblasts (Supplementary Fig. 4f, exemplified by HPAT2).

We extended our analysis of HPAT gene expression in the context of the establishment of pluripotency in single cells via bicluster and correlation analysis (Fig. 2b,c, Supplementary Fig. 4g, Supplementary Table 3 and Supplementary Note) and derived a Bayesian network

based on all pluripotency markers and HPATs, across all single cells (Fig. 2d). In agreement with previous reports²⁶, our Bayesian network identified (i) a triangle consisting of *POU5F1*, *NANOG* and *SOX2*, known to be the core pluripotency regulators^{27–29}, and (ii) the close association of *SALL4* with the core pluripotency network^{30,31}. Notably, we observed that the three new lincRNAs (HPAT2, HPAT3 and HPAT5) that emerged in our previous bicluster and correlation analysis, as well as preimplantation embryo studies, were closely linked to the core regulatory network of pluripotency and directly associated with *POU5F1*, *SOX2*, *SALL4* and *NANOG* (Fig. 2d). Note that *HPAT2* and *HPAT3* both map to chromosome 2 and are derived from HERV-H elements; *HPAT5* maps to chromosome 6 and comprises SINE and HUERS-P1 repeat elements (Supplementary Fig. 5d).

HPATs modulate reprogramming and hESC differentiation

To test whether HPAT2, HPAT3 and HPAT5 are required for nuclear reprogramming in induced pluripotency, we transiently knocked down their expression simultaneously over the course of mRNA-based reprogramming (Fig. 3a). Transient knockdown with siRNAs (Supplementary Fig. 5a (efficiencies) and Supplementary Table 4) inhibited reprogramming and resulted in formation of fewer TRA-1-60–positive cells and alkaline phosphatase–positive colonies, both markers for pluripotency, in comparison to control cells (Fig. 3b–d). Titration of siRNA molecules resulted in more pronounced phenotypes (Fig. 3e,f). To determine the effects of single HPATs, we reduced the levels of HPAT2, HPAT3 and HPAT5 transcripts individually and evaluated reprogramming efficiencies. Reduced expression of each individual HPAT appeared to negatively affect nuclear reprogramming; however, reduced expression of HPAT5 alone resulted in a statistically significant reduction in the number of alkaline phosphatase–positive colonies (Fig. 3h,i). To ensure reproducibility, experiments were performed twice with three replicates; we also ensured that the observed results were not a consequence of different cell proliferation rates (Fig. 3g).

Concurrently, we overexpressed all three HPATs (days 0 and 3; Supplementary Fig. 5b) and observed reciprocal phenotypes with elevated reprogramming efficiencies (Fig. 3a–f). We also reprogrammed BJ fibroblasts by inclusion of all three HPATs at different molar ratios in combination with *POU5F1* mRNA. Note that transfection with *POU5F1* mRNA alone never resulted in reprogramming. In contrast, although the majority of trials resulted in no reprogramming to pluripotency, when *POU5F1* was expressed with HPAT2, HPAT3 and HPAT5 at a molar ratio of 3:1:1:3 (Fig. 3j), alkaline phosphatase–positive clones were successfully derived.

On the basis of the data outlined above, we suspected that the HPAT genes might be under the transcriptional control of key transcription factors linked to pluripotency. Thus, we generated NANOG chromatin immunoprecipitation and sequencing (ChIP-seq) data from hESCs (H9) using 100-bp paired-end reads. NANOG is a transcription factor that occupies single-copy loci, as well as TEs including the ERV1 repeat family and LTR7 sequences in pluripotent stem cells^{14,24}. We observed that *HPAT2*, *HPAT3* and *HPAT5* (and all other HPAT genes) were specifically bound by NANOG (Supplementary Fig. 5c,d and Supplementary Table 5). We then compared our data set with the NANOG ChIP-seq data set generated by Kunarso *et al.*³² and observed extensive overlap (72%), indicating that the data

sets are very comparable. Indeed, almost every single HPAT gene promoter region was also found to be bound by NANOG in the Kunarso *et al.* data set (data not shown). Moreover, exogenous expression of NANOG activated HPAT2, HPAT3 and HPAT5 expression in fibroblasts in a methylation-dependent manner (Fig. 3k). Collectively, these results indicate that HPAT2, HPAT3 and HPAT5 expression is regulated by NANOG and can contribute directly to reprogramming and acquisition of pluripotency. HPAT5, relative to HPAT2 and HPAT3, appeared to have more pronounced effects and had a unique sequence derived from a previously uncharacterized ERV in pluripotent stem cells. Thus, we focused further efforts on the *HPAT5* gene using a series of complementary functional assays.

To determine whether HPAT5 modulates differentiation, we generated an H1 hESC line that stably overexpressed *HPAT5* (HPAT5-OE) under the control of the *EEF1A1* promoter (Supplementary Fig. 5e,f). Under self-renewal conditions, HPAT5-OE cells displayed normal pluripotent stem cell morphology and a high level of expression of *HPAT5* mRNA relative to the control line. Upon differentiation (forced via the removal of basal fibroblast growth factor (bFGF) and down-regulation of *POU5F1* expression) (Fig. 4a), ectopic overexpression of HPAT5 suppressed hESC differentiation, as indicated by increased and persistent expression of the pluripotency markers *POU5F1*, *NANOG* and *SOX2* (Fig. 4c) at days 3 and 6 after differentiation, in combination with delayed upregulation of genes indicative of differentiation of all three germ layers (Supplementary Fig. 5g). Conversely, control mCherry-overexpressing H1 hESCs readily differentiated under the same conditions and adopted a somatic cell-like morphology 3 d after transfection (Fig. 4b and Supplementary Fig. 5e,f).

HPAT5 binds to members of the microRNA processing machinery

Genes for lincRNAs have been shown to be involved in epigenetic regulation, by recruiting chromatin-remodeling complexes^{33,34}. To determine whether the *HPAT5* gene might function through interactions with chromatin-remodeling complexes or other RNA-binding proteins, we used protein microarray assays to globally screen interactions between HPAT5 and candidate proteins *in vitro*³⁵. This analysis yielded a list of candidates that bound *in vitro* to HPAT5 (Supplementary Fig. 6a–d, Supplementary Table 6 and Supplementary Note). TARBP2, a subunit of the RNA-induced silencing complex (RISC) that has a major role in the microRNA (miRNA) processing pathway³⁶, showed the most significant enrichment in binding. AGO2, a second subunit of RISC, was also among the proteins most significantly bound to HPAT5. This result was surprising, as it is commonly known that miRNAs are required to guide RISC to target RNAs, despite recent reports that have consistently observed AGO2-RNA associations independent of miRNAs³⁷.

We included HPAT2 and HPAT3 as controls and validated binding with proteins previously described, such as OCT4, to interact with HERV-H-containing sequences (Supplementary Fig. 6e) (ref. 17). In light of these results, we hypothesized that HPAT5 exerts post-transcriptional regulation of gene expression, possibly by binding to specific miRNAs through the miRNA-loading complex. To test this hypothesis, we used a bioinformatic approach (see URLs) aimed at predicting miRNA response elements (MREs) in the HPAT5 transcript. We included HPAT2 and HPAT3 as controls in our analysis to demonstrate

differences in predicted binding partners from HPAT5. linc-ROR was used as an additional control, as it harbors HERV-H–derived sequences at its 5' end, as do HPAT2 and HPAT3 (Supplementary Table 7). Our analysis showed that, whereas HPAT2 and HPAT3 were predicted to bind similar miRNAs, the list of miRNA potentially bound by HPAT5 was substantially different. Indeed, one entire miRNA family—the let-7 family—was predicted to bind to exon 2 of HPAT5 within an Alu element, a TE subclass of SINEs, but not to HPAT2 or HPAT3 (Supplementary Fig. 7a). This miRNA family has previously been shown to function in modulating hESC pluripotency and reprogramming, and its expression has been reported to be inversely correlated with expression of LINEs^{38–41}.

HPAT5 modulates let-7 expression

To validate that HPAT5 is indeed targeted *in vivo* by let-7, we constructed luciferase reporters encoding full-length HPAT5 in the 3'UTR of the luciferase gene (Fig. 4d). Reporters were cotransfected into HEK293 cells with two miRNA mimics (Hs-let-7a and Hs-let-7d). HPAT2, HPAT3 and linc-ROR were used as negative controls. The Hs-let-7a and Hs-let-7d mimics significantly reduced luciferase activity in comparison to scrambled miRNA for the HPAT5 reporter (Fig. 4e). No differences were observed in negative controls. Further, to test the specificity of binding of the Hs-let-7a and Hs-let-7d mimics to predicted target sites, we deleted both MREs in HPAT5 and constructed a mutant luciferase reporter (HPAT5-mutant) to measure its activity in HEK293 cells. Whereas the wild-type reporter showed significantly decreased luciferase activity when cotransfected with the Hs-let-7a and Hs-let-7d mimics, the mutant reporter was refractory to Hs-let-7a– and Hs-let-7d–driven reporter inhibition (the activity was comparable to that of control cells transfected with scrambled miRNA or no miRNA) (Fig. 4f). We expanded our analysis by introducing two different point mutations in the seed sequence (base pairs 4 and 6) of the Hs-let-7a and Hs-let-7d mimics and with two additional mutant reporters that compensated for the mutations in HPAT5 (Fig. 4g,h), confirming our previous results.

To further test whether, in addition to binding to mature let-7 miRNAs, HPAT5 might interfere with let-7 maturation (conversion from pri-miRNA to mature miRNA), we measured the expression levels of pre-let-7 miRNA in differentiated fibroblasts that transiently overexpressed HPAT5. Endogenous mature let-7 levels but not pre-let-7 levels were significantly downregulated 48 h after exogenous HPAT5 overexpression, indicating that HPAT5 does not suppress the transcription of let-7 or inhibit the maturation of pre-let-7 into let-7 (Supplementary Fig. 7b). In contrast, LIN28A overexpression (used as a positive control) resulted in significant downregulation of pre-let-7 expression, which leads to decreased mature let-7 levels, consistent with the literature^{42,43}.

Collectively, our results demonstrate that let-7 binds specifically to complementary sequences on HPAT5 and that point mutations within the seed sequence of let-7 can abolish this interaction. Further, we demonstrated that overexpression of HPAT5 in hESCs delayed induced differentiation (Fig. 4a–c) and that HPAT5 is functionally linked to the differentiation by interaction with let-7. To probe functional interaction further, we knocked out the endogenous genomic *HPAT5* locus in pluripotent stem cells (*HPAT5*-KO) using CRISPR/Cas9 (Supplementary Fig. 7c–e). Analysis of *HPAT5*-null cells showed increased

let-7 levels relative to wild-type cells (Supplementary Fig. 7f), although these levels were not sufficient to induce spontaneous hESC differentiation. These results are consistent with studies in wild-type mouse ESCs in which additional mechanisms exist to ensure protection of the pluripotent state by inhibiting let-7 activity or that of other stem cell-specific miRNAs^{38,44}.

To test whether *HPAT5* ablation has a significant effect on reprogramming efficiencies (as indicated by our experiments with siRNAs in Fig. 3h), we differentiated *HPAT5*-KO hESCs into fibroblasts⁴⁵ and reprogrammed them back into iPSCs using episomal vectors encoding the Yamanaka factors (Supplementary Fig. 7g). *HPAT5*-KO secondary fibroblasts reprogrammed with lower efficiencies than wild-type secondary fibroblasts, as indicated by lower percentages of alkaline phosphatase- and TRA-1-81-positive cells at day 24 after the initiation of reprogramming (Supplementary Fig. 7h). In addition, we assessed endogenous *HPAT5* and let-7 levels in cells that were transitioning between the fibroblast and iPSC states (Supplementary Fig. 7i). *HPAT5*-KO cells had significantly ($P = 0.0191$) higher let-7 levels at day 10 of reprogramming than wild-type controls. However, let-7 levels were, similar to in wild-type cells, significantly lower in *HPAT5*-depleted cells in comparison to the originating fibroblasts, indicating that additional mechanisms regulate endogenous let-7 expression during the acquisition of pluripotency (for example, expression of LIN28; ref. 43).

Our data suggest that *HPAT5* might be modulating the balance between pluripotency and differentiation by counteracting let-7 activity when let-7 is expressed at very high levels (for example, in somatic cells). To test this hypothesis, we overexpressed let-7 in *HPAT5*-KO and wild-type cells and examined gene expression changes 48 h after treatment. Exogenous overexpression of let-7 triggered differentiation in *HPAT5*-KO cells relative to wild-type cells, as determined by microarray analysis and examination of morphological changes (Fig. 5a and Supplementary Table 8).

To rescue the effects of let-7-mediated differentiation in *HPAT5*-KO cells, we transfected cells with an overexpression vector for wild-type *HPAT5*. We also overexpressed a mutant *HPAT5* transcript that lacked the predicted let-7-binding sites to test for specificity. Whereas overexpression of wild-type *HPAT5* provided partial rescue of let-7-mediated differentiation in *HPAT5*-KO lines, the *HPAT5* mutant did not do so but instead led to changes in the transcriptome similar to those observed in cells that overexpressed let-7 alone (Supplementary Fig. 8a and Supplementary Table 8). To determine whether differentiation in *HPAT5*-KO lines was triggered by downregulation of let-7 targets, we used cWords, a tool that identified enrichment of miRNA seed sequences among the entire list of ranked differentially expressed genes. Among the top-ranked seed sequences that were shared by the most significantly downregulated genes in *HPAT5*-KO cells was the let-7 seed (TACCTC) (Supplementary Table 5b). Similarly, the let-7 seed sequence was significantly enriched in *HPAT5*-KO cells that failed to rescue endogenous *HPAT5* depletion (Supplementary Fig. 8b–d). In contrast, *HPAT5*-KO cells in which wild-type *HPAT5* was rescued were not enriched for downregulated genes that shared a let-7 seed sequence, indicating that exogenous *HPAT5* successfully sequestered overexpressed let-7 to prevent it from downregulating its targets.

To further probe the regulatory role of HPAT5 in the expression of let-7 in hESCs, we overexpressed and knocked down HPAT5 in hESCs and assessed mature let-7 miRNA levels at 48 h after transfection. To control for non-specific binding, we again used the overexpression vector for mutant HPAT5. The expression levels of both mature let-7 miRNAs were inversely associated with the expression levels of HPAT5 (Fig. 5c). Specifically, we observed the most significant change in Hs-let-7d expression upon knockdown of HPAT5. In contrast, mutant HPAT5 did not have the same effect, suggesting that HPAT5 negatively regulates let-7 expression and activity through specific binding. In addition, when we overexpressed let-7 in hESCs, HPAT5 levels were downregulated in comparison to those in control samples (Supplementary Fig. 8e). Given that miRNA-lincRNA target pairs can be purified by immunoprecipitation of the RISC component AGO2 (ref. 46), we tested whether AGO2 would coprecipitate with HPAT5. RNA immunoprecipitation (RIP) followed by quantitative PCR (qPCR) confirmed the *in vivo* interaction between HPAT5 and AGO2 in the presence of let-7 in hESCs, in contrast to GAPDH (Fig. 5d and Supplementary Fig. 8f). Collectively, these results indicate that mature let-7 is able to directly bind and guide RISC to its target, HPAT5 (Fig. 5).

DISCUSSION

Recent data have linked one entire class of TE-derived lincRNAs (HERV-H) to a naive pluripotent hESC state *in vitro* that resembles pluripotent cells of the human ICM¹⁶. Yet, to our knowledge, this study is the first to probe the biological relevance of three individual TE-derived lincRNAs during human embryo development *in vivo* and to use CRISPR-mediated gene editing technology to investigate the mechanism of action of a single lincRNA in knockout human stem cells *in vitro*.

The importance of crosstalk between miRNAs and lincRNAs in the regulation of pluripotency in hESCs has been documented in several studies^{20,38}. Here we provide data that support crosstalk between let-7 and HPAT5. We observed that let-7 binding occurs within an Alu element in the second exon of HPAT5 (Supplementary Fig. 7a). The acquisition of a single base pair within this Alu element in HPAT5 has generated a let-7 seed sequence, likely conferring specificity to HPAT5. Other Alu elements, which have expanded tremendously in primate genomes, have an important role in human embryonic development and, more importantly, have been described to function in DNA binding and mRNA recognition when embedded in lincRNAs^{47–50}. Indeed, a recent study has shown that several human miRNAs and miRNA target sites are, in fact, derived from L1, Alu and MIR elements^{40,51}.

More recently, the activity of specific classes of retrovirus-derived lincRNAs has been linked to human preimplantation embryo development. LTR-driven expression of specific HERV families has been described in a stage-specific context during early human embryo development¹⁸. Two specific HERV families (HERV-H and HERV-K) have also been linked to human preimplantation development. Grow *et al.* described LTR element-driven reactivation of HERV-K, which, unlike most other HERVs, has retained multiple copies of intact ORFs that encode retroviral proteins¹⁹. In contrast, HERV-H expression may have regulatory roles in establishing and/or maintaining pluripotency, a hallmark of pluripotent

epiblast cells in blastocyst-stage embryos¹⁶. These findings, together with our results, demonstrate that different HERV elements have distinct roles in regulating fundamental biological processes, including the acquisition of pluripotency *in vivo* during embryogenesis. Dissecting the role of each single HERV element may be of paramount importance to understanding the specifics of human development. We anticipate that direct genetic dissection via genome editing, as demonstrated here, may find that many cell fate decisions, including those of human pre- and post-implantation development, are modulated by complex regulatory mechanisms that employ ‘recycled’ retroviral sequences that were introduced and modified during the course of evolution to confer human-specific dynamics to development.

URLs

Bioinformatic pipeline, <http://regrna2.mbc.nctu.edu.tw/>; Bioresearch Technologies, <http://www.biosearchtech.com/>; Massachusetts Institute of Technology CRISPR design tool, <http://crispr.mit.edu/>.

ONLINE METHODS

Cells

BJ human fibroblast cells (passage 6) were established from normal fetal foreskin, purchased from Stemgent and used for nuclear reprogramming toward iPSCs. Cells were tested for mycoplasma prior to use for experiments.

Cell culture

BJ fibroblast cells were cultured on plates coated with 0.2% gelatin (Sigma) in DMEM-FBS (DMEM + GlutaMAX (DMEM) supplemented with 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin). Fibroblasts were maintained in culture by changing the medium every 3 d and passaging cells at a 1:3 dilution when they were 80–90% confluent. hESCs (H9 and H1) and derived iPSCs were cultured on plates precoated with growth factor reduced Matrigel (BD Biosciences) in basal mTeSR1 medium (Stemcell Technologies) supplemented with 5× mTeSR1 supplement (Stemcell Technologies). Cells were maintained in culture by changing the medium daily and enzymatically passaging cells at a 1:2 to 1:5 dilution with prewarmed Accutase (Innovative Cell Technologies). Differentiated cells were removed and/or cleaned under a laminar flow dissection hood.

All cultures were maintained at 37 °C and 5% CO₂. BJ fibroblasts were frozen in 90% FBS (Gibco, Life Technologies) and 10% DMSO (Sigma-Aldrich). hESCs and iPSCs were frozen in Bambanker (Wako Chemicals). Tissue culture reagents and chemicals were purchased from Life Technologies, Sigma-Aldrich, Becton Dickinson and Company (BD) and Fisher Scientific unless otherwise stated.

Microarray analysis

Total mRNA was isolated from hESCs using the RNeasy kit (Qiagen). The quality of the total RNA was confirmed using an Agilent 2100 Bioanalyzer. Samples were sent to the Pan Facility at Stanford University for further processing. Biotinylated cRNA was prepared

according to the standard Affymetrix protocol from 6 µg of total RNA (GeneChip Whole-Transcript Sense Target-Labeling Assay, 701880 Rev.5, Affymetrix). The samples were then hybridized to the Human Gene 2.0 ST array. Probe arrays were washed and scanned with the Hewlett-Packard GeneArray Scanner G2500A. Raw data files were created by Command Console, the Affymetrix operating software program. The Affymetrix Expression Console Program was used to examine the Affymetrix Gene Array quality control factors for all samples in a project. Global scaling was used as the normalization method (RMA). Enrichment analysis was performed with cWords.

Statistical analysis

For single-cell analysis, individual cells were considered as biological replicates ($n = 578$). Calculated primer efficiencies were normally distributed, as determined by the Shapiro-Wilk test. For normally distributed data, we used the two-tailed Student's t test for significance calculations. Nonparametric statistical approaches were applied for data not following a normal distribution. Specifically, we chose the Kurskal-Wallis test for independent and unequally sized sample calculations. Statistical significance was set to $P < 0.05$ for gene expression analysis ($n > 3$) and the Plaid and CC bicluster algorithm, respectively. The Xmotif bicluster algorithm only resulted in bicluster formation with $P < 0.01$. Only Bayesian network connections with $P < 0.05$ are shown. Correlation analysis found only significant correlations with $P < 0.05$. For TRA-1-60- and alkaline phosphatase-positive colony count, sample counts were normalized to the total number of iPSC colonies within one experiment to decrease variability between experiments. Resulting values (for each experiment) were subjected to two-tailed Student's t test. Error bars represent standard deviation in all tests of statistical significance.

Assay performance validation

Primers were designed to span introns to avoid the amplification of possible contaminating genomic DNA. Each primer pair was tested before use for single-cell gene expression analysis for efficiency, sensitivity and specificity as well as to determine the expected melting temperature (T_m) for the specific amplicon for each assay. We used cDNA prepared from bulk total RNA extracted from BJ fibroblasts, hESCs (H1) and iPSCs (BJ.iPSCs). Pre-amplification was performed with 20 ng of total RNA, 50 nM of each primer pair and 1× TaqMan PreAmp Master Mix (Applied Biosystems) in a 20-µl total volume. The thermal cycling protocol comprised incubation at 95 °C for 10 min; 14 cycles of 95 °C for 5 s and 60 °C for 4 min; and holding at 4 °C. Samples were treated with Exonuclease I (NEB) at 37 °C for 30 min, and the reaction was inactivated by heating at 80 °C for 15 min. cDNA was diluted with DNA suspension buffer (Teknova) to a total volume of 100 µl. A 1:2 dilution series was prepared by mixing 30 µl of each cDNA sample with 60 µl of DNA suspension buffer. The diluted cDNA sample was subsequently diluted further, down to a 14× dilution of the original sample. The 15 cDNA samples, including a no-template control (DNA suspension buffer), were analyzed by qPCR using a 96.96 Dynamic Array Integrated Fluidic Circuit (IFC) and the BioMark HD (Fluidigm) according to the manufacturers' instructions. Each diluted sample was loaded in six technical replicates to determine the lower limit of technical noise of the instrument. Only sample-assay combinations with specific amplification were used for standard curve calculations of \log_{10} -transformed sample

dilution versus average C_t value. For each assay, efficiency was estimated from the slope of the standard curve using efficiency (E) = $10^{-1/\text{slope}} - 1$. Linear regression analysis depicted a precise quantitative response to the dilution series for 88 of 96 assays as R^2 values were between 0.97 and 0.99 (Supplementary Fig. 3c–e); thus, we excluded the eight assays with $R^2 < 0.97$. Using the primer efficiency distribution histogram, we calculated an average primer efficiency of 1.02 (102%) with s.d. = 0.06 (Supplementary Fig. 3d).

Single-cell quantitative PCR

We used the C1 Single-Cell Auto Prep System (Fluidigm) for single-cell capture and preamplification according to the manufacturer's instructions (protocol 100-4904). Briefly, we prepared a pool of all x primers (500 nM). We then prepared a lysis final mix, a reverse-transcriptase (RT) final mix and a preamplification (PreAmp) final mix and stored them on ice. Next, the C1 IFC chip for medium-size single cells (10 to 17 μm in diameter; barcode 1782x) was primed: 200 μl each of C1 collection reagent, preloading reagent, blocking reagent and wash buffer were loaded onto the chip, the chip was placed into the C1 Single-Cell Auto Prep System and the script 'Prime (1782x)' was run. Priming lasted 20 min, and cells were prepared in the meantime as follows. For days 0–7, when cells were still homogeneous in culture, cells were treated with Accutase (Innovative Cell Technologies) to generate a single-cell suspension, washed once and resuspended in Pluriton medium at a concentration of 250×10^3 cells/ml. For days 7–12 and iPSCs, colony-like structures were manually isolated and treated with Accutase, washed once and resuspended in Pluriton medium at a concentration of 250×10^3 cells/ml (see Supplementary Fig. 3b for phase-contrast images). Then, 12 μl of single-cell suspension was mixed with 8 μl of C1 Cell Suspension Reagent (Fluidigm). After priming, the blocking and priming solutions were removed and 10 μl of cell mix was loaded onto the C1 chip. The C1 chip was placed back into the instrument, and the script 'Cell Load (1782x)' was run. After cell capturing, the C1 chip was removed and single-cell capturing was evaluated on a microscope (Supplementary Fig. 3f). Empty capture sites were noted, and the C1 chip was loaded with collection reagent, 7 μl of lysis final mix, 7 μl of RT final mix and 24 μl of PreAmp final mix. The chip was placed back into the instrument, and the script 'PreAmp (1782x)' was run with the following settings: reverse transcription, 25 $^\circ\text{C}$ (600 s) and 42 $^\circ\text{C}$ (3,600 s); preamplification, 95 $^\circ\text{C}$ (600 s), 18 cycles of (95 $^\circ\text{C}$ (15 s) and 60 $^\circ\text{C}$ (240 s)) and 4 $^\circ\text{C}$ (hold). After preamplification, the C1 chip was removed from the instrument and 3 μl of cDNA (for each single cell) was isolated and diluted in 25 μl of DNA suspension buffer (Fluidigm). Preamplified samples were then subsequently used on the BioMark HD using the protocol 100-3488 and starting with the preparation of sample and assay mix (see "RNA isolation and gene expression analysis of bulk samples with quantitative PCR").

Determining limit-of-detection values

Because of the lognormal distribution described by Bengtsson *et al.*⁵³ and others, single-cell data are best viewed as expression level above detection limit on a log scale. For qPCR data, we determined the log base 2 and defined $\log_2(\text{expression}) = \text{LOD } C_t - C_t \text{ raw (of gene)}$, where LOD is the limit of detection. We used bulk RNA and the dilution series of generated cDNA samples to calculate LOD C_t as follows. Mean C_t values and s.d. for each assay (six replicates) were calculated for all serial dilutions. Average C_t values with s.d. > 1 determined

the threshold that was assigned to the LOD for each assay. We finally calculated the median of all LOD C_t values across all assays to determine a universal LOD C_t score of 25, which was used throughout this study.

Quality assessment and normalization of single-cell expression values

Melting curves were analyzed, and false positive signals were excluded (Supplementary Fig. 3g). Chip-to-chip variation was assessed with three IFCs (fibroblasts, fibroblasts transfected with GFP for 2 d and fibroblasts transfected with GFP for 5 d) to identify assays that significantly change across different IFC chips (Supplementary Fig. 3h). We excluded six assays (*DPPA4*, *HDAC3*, *HPAT11*, *HPAT13*, *INO80C* and *PRMT5*) for subsequent analysis because they did not correlate within an acceptable range between the three IFCs. The remaining 82 assays (Supplementary Table 2) led to only small observed variations in gene expression changes across all three IFCs, indicating their robustness across chip-to-chip variation as well as GFP versus non-transfected single cells and were used for subsequent analysis. Then, raw C_t values were converted to expression levels using $\log_2(\text{expression}) = \text{LOD } C_t - C_t \text{ raw (of gene)}$ with $\text{LOD } C_t = 25$. Values with $\log_2(\text{expression}) < 0$ were excluded. Genes expressed in fewer than 5% of single cells were eliminated as well. Single cells with $\log_2(\text{expression})$ values lower than 3 s.d. of an assay across all cells were labeled apoptotic and were excluded; 192 cells (two IFCs with GFP control) were removed from further analysis and 94 cells across all seven remaining IFCs were eliminated owing to the above-mentioned reasons, resulting in 578 cells. We normalized such that each cell had the same median $\log_2(\text{expression})$ value across all genes detected in that cell. This ensured that the normalization factor included data from all genes in the study. For this study, we generated a high-quality data matrix of 578 genes across 82 assays, resulting in 47,396 single-cell expression values that was used for data analysis (Supplementary Fig. 3i).

Alkaline phosphatase and TRA-1-60 staining

Alkaline phosphatase staining was performed using Vector Red Alkaline Phosphatase Substrate Kit I (Vector Laboratories) following the manufacturer's instructions.

StainAlive DyLight 488 anti-Human TRA-1-60 antibody (Stemgent) was diluted in fresh cell culture medium to a final concentration of 5 $\mu\text{g/ml}$. Old medium was aspirated and replaced with medium containing diluted antibodies. Cells were incubated for 30 min at 37 °C and 5% CO_2 . The medium with primary antibody was aspirated, and cells were washed gently twice with cell culture medium. Fresh cell culture medium was added, and cells were examined under a fluorescent microscope using the appropriate filters. Cells were kept in culture after examination. Representative images ($n > 3$) were acquired with the same microscope settings (gain, exposure and fluorescence excitation), and fluorescence intensity (emission) was measured with ImageJ and calculated against TRA-1-60-negative cells.

Derivation of mRNA-induced pluripotent stem cells

BJ fibroblasts were seeded at $1-4 \times 10^4$ cells per well of a six-well plate on wells coated with growth factor reduced Matrigel and cultured in Pluriton basal medium. After 24 h, Pluriton basal medium was replaced with conditioned Pluriton media from NuFF cells (human fibroblasts, GlobalStem) (Stemgent) supplemented with Pluriton supplement

(Stemgent) and B18R (200 ng/ml; eBioscience). Cells were transferred to a low-oxygen environment (5%) for higher reprogramming efficiency before the first transfection. After 2 h of equilibration in low-oxygen conditions, mRNA cocktail containing OSKM factors (*OCT4*, *SOX2*, *KLF4* and *MYC*) was transfected into cells, and transfection was repeated every 24 h until colony formation was observed, around day 12–14. Incubation of mRNA and transfection mix with cells was carried out for 4 h.

Primary iPSCs appeared around day 14 and were handpicked onto fresh culture dishes coated with Matrigel; the medium was replaced with mTeSR1 supplemented with 5× mTeSR1 supplement. Established iPSC lines were cultured under 20% oxygen conditions and were subjected to single-cell gene expression analysis. The pluripotency of iPSCs was assessed by teratoma formation (Supplementary Fig. 3a). Reprogramming efficiencies with our optimized feeder-free mRNA-based protocol ranged between 4–6% on the basis of initial cell seeding numbers and fully developed primary colonies.

Functional characterization of HPAT2, HPAT3 and HPAT5 during nuclear reprogramming was performed in replicate. Combined knockdown or overexpression of HPAT2, HPAT3 and HPAT5 was performed twice. Nuclear reprogramming with single HPATs was performed twice. Reprogramming with OCT4 and HPAT2, HPAT3 and HPAT5 was performed once with different molar ratios of all four factors. The molar ratio of 3:1:1:3 for OCT4, HPAT2, HPAT3 and HPAT5 was repeated for confirmation. Statistical significance between TRA-1-60–positive cell counts was assessed by fluorescence detection with ImageJ. Three representative areas for each experimental group were used to analyze fluorescence intensities, and Student's *t* test was applied. Statistical significance was measured in a similar fashion for iPSC colony counts on the basis of positive staining for alkaline phosphatase and cell number counts.

***In vitro* differentiation**

Neuronal differentiation. Differentiation was initiated from fibroblasts that were directly converted into induced neuronal cells as previously described⁵⁴ (Supplementary Fig. 4d).

RNA isolation and gene expression analysis of bulk samples with quantitative PCR

Gene expression analysis was performed using a micro-fluidic platform (Fluidigm) following the manufacturer's protocol *Single-Cell Gene Expression Using EvaGreen DNA Binding Dye* (protocol 100-3488), with some modifications adjusted for bulk samples. Briefly, CellsDirect 2× Reaction Mix (Life Technologies), SuperScript III RT Platinum Taq Mix (Invitrogen), 4× Primer Mix (200 nM) and DNA suspension buffer (Teknova) were added to a total volume of 9 µl. Cells in bulk were collected, washed and counted; the cell suspension was adjusted to a concentration of 50,000–100,000 cells/ml. One microliter of cell suspension was added to each reaction, and the following thermal cycling protocol was set: reverse transcription, 50 °C for 15 min; inactivation of reverse transcriptase/activation of Taq, 95 °C for 2 min for 18 cycles, 95 °C for 15 s and 60 °C for 4 min, 4 °C for infinity. ExoSAP-IT treatment removed unused material and was performed at 37 °C for 30 min (digestion) and 80 °C for 15 min (inactivation). The reaction was diluted 1:5 in DNA suspension buffer and stored at –20 °C or immediately used for sample pre-mix. Sample

pre-mix, sample and assay mix were prepared according to the manufacture's instructions (Fluidigm). Dynamic Arrays IFCs (96.96 or 48.48) were primed with control line fluid, and the chip was loaded with assay and sample mixes using the HX IFC controller (Fluidigm). RT-PCR was performed on the BioMark HD (Fluidigm) with the following two-step fast cycling protocol (EvaGreen): 95 °C (2 min) followed by 40 cycles of 95 °C (5 s), 60 °C (20 s) and melting curve generation.

Teratoma formation

The pluripotency of derived iPSCs was evaluated with teratoma formation assays. Cells (one well of a 12-well plate) were collected in 30 μ l of culture medium and injected into the kidney capsule of SCID mice (adult (10 weeks), female; Stanford Assurance Number A3213-01, protocol number 16146). After 3–4 weeks, teratomas were dissected and fixed overnight in 4% paraformaldehyde diluted in PBS.

Fixed samples were sent to AML Laboratories for paraffin embedding, sectioning and staining with hematoxylin and eosin. Sections were then examined for the presence of tissue representatives of all three germ layers.

ChIP-seq

ChIP assays were performed from approximately 1×10^7 cells per experiment, according to a previously described protocol with slight modifications²⁷. Briefly, cells were cross-linked with 1% formaldehyde for 10 min at room temperature, and formaldehyde was quenched by the addition of glycine to a final concentration of 0.125 M. Chromatin was sonicated to an average size of 0.5–2 kb, using a Bioruptor (Diagenode). Protein G dynal beads (50–75 μ l; Invitrogen) were used to capture 3–5 μ g of antibody in phosphate citrate buffer, pH 5.0 (2.4 mM citric acid, 5.16 mM Na_2HPO_4) for 30 min at 27 °C. Antibody-bead complexes were rinsed two times with PBS and added to sonicated chromatin; samples were rotated at 4 °C overnight. Ten percent of chromatin was reserved as 'input' DNA. Magnetic beads were washed and chromatin was eluted, followed by reversal of the cross-linking and DNA purification. Resultant ChIP DNA was dissolved in TE buffer. Results were verified with qPCR of seven selected regions (Supplementary Fig. 5c and Supplementary Table 5).

Overexpression of *NANOG* in 5-Aza-2'-deoxycytidine-treated fibroblast cells

Fibroblast cells were cultured in DMEM supplemented with 10% FBS to which was added 5-Aza (5 mM final concentration diluted in DMSO; Sigma-Aldrich) 48 h before transfection. mRNA encoding *NANOG* was trans-fected (300 ng/ μ l) into cells with RNAiMAX (Life Technologies) according to the manufacturer's instructions. Control cells were transfected with mRNA encoding eGFP or were mock transfected. Untreated control cells were supplemented with DMSO only. Cells were collected 48 h after transfection.

Design of siRNAs and transfection

siRNA duplexes targeting unique regions of the new transcripts were selected on the basis of low seed frequency using the siDESIGN Center tool from Thermo Scientific Dharmacon RNAi Technologies. A pool of two to four target-specific siRNAs for each transcript was used for transfections. All siRNAs, including POU5F1/OCT4 SMART pool siRNA

(L-019591-00), siGLO Green transfection indicator (D-001630) and scrambled siRNA (siGENOME Non-Targeting siRNA Pool #1, D-001206-13) were purchased from Thermo Scientific. siRNA transfections were carried out using Lipofectamine RNAiMAX reagent, as recommended by the manufacturer's protocol for reverse transfection. Briefly, for transfections in 24-well plates, 50 pmol of a mixture of target-specific siRNA duplexes was diluted in 0.1 ml of Opti-MEM serum-free medium (Life Technologies) for each well and added to Matrigel-coated plates. Then, 1.5 μ l of Lipofectamine RNAiMAX was added to the wells containing siRNA, and complexes were incubated at room temperature for 15–30 min. H1 hESCs, which had been pretreated with 10 μ M ROCK inhibitor Y-27632 for 1 h, were dissociated using Accutase and plated in the wells containing the siRNA-RNAiMAX complexes at 1×10^5 cells per well in 0.4 ml of mTESR1 medium supplemented with 2 μ M thiazovivin. For transfection in other vessels, the above protocol was scaled up or down accordingly. All transfections were performed in three replicate wells. Medium was changed every day, and cells were collected on day 2 after transfection.

Design of overexpression vectors

The sequences of all isoforms for HPAT2, HPAT3 and HPAT5 (identified by Au *et al.*¹³) were assembled with Gibson Assembly Cloning technology (NEB). Briefly, gBlocks (gene fragments) were synthesized by Integrated DNA Technologies (IDT), and individual gBlocks were assembled to one gene transcript with Gibson Assembly technology followed by an amplification reaction with Phusion DNA polymerase according to the manufacturer's instructions. Amplified genes were ligated into the pENTR/D-TOPO vector (Life Technologies) for the Multisite Gateway system. Clones were transformed into One-Shot Competent *Escherichia coli*, DNA was purified and sequenced, and positive clones were used for a recombination reaction with the Gateway destination vector (pcDNA-DEST40). Subsequent transformation into One-Shot Competent *E. coli* followed by DNA purification and sequencing for verification of correct cloning resulted in overexpression vectors for each isoform of HPAT2, HPAT3 and HPAT5.

LincRNA labeling with Cy5 and protein microarray analysis

We performed lincRNA labeling followed by probing on Protoarrays (Life Technologies) as previously described³⁵ with the following modification: *in vitro* transcription of HPAT2, HPAT3 and HPAT5 was performed with the MegaScript kit (Ambion) according to the manufacturer's instructions.

Luciferase reporter construction

Full-length transcripts of HPAT2, HPAT3 and HPAT5 and from linc-ROR (including mutated versions of HPAT5) were assembled from individual gBlocks (IDT), amplified with specific primers and directionally ligated into the pMIR-REPORT miRNA Expression Reporter Vector (Invitrogen). Clones were transformed into One-Shot Competent *E. coli*, DNA was purified and sequenced, and positive clones were used for luciferase activity assays.

Luciferase reporter transfection and dual-luciferase assays

HEK293 cells (15×10^4) were seeded into each well of a 96-well plate, incubated overnight and cotransfected with 80 ng of reporter or mutant pMIR-REPORT construct, 8 ng of internal control pRL-TK *Renilla* luciferase vector and indicated miRNA mimics (final concentration of 50 nM) with 1 μ l of Lipofectamine 2000 (Invitrogen). hESCs (90×10^4) were seeded into each well of a 24-well plate, incubated overnight and cotransfected with 200 ng of reporter or mutant pMIR_REPORT construct, 20 ng of internal control pRL-TK *Renilla* luciferase vector and indicated miRNA mimics (final concentration of 50 nM) with 2 μ l of Lipofectamine 2000. Lysates were collected 48 h after transfection, and reporter activity was measured with the Dual-Luciferase Assay (Promega). Data were normalized by dividing firefly luciferase activity by *Renilla* luciferase activity, according to the manufacture's instructions.

microRNA quantification of let-7a and let-7d expression levels by quantitative RT-PCR

RNA was extracted from hESCs (H1) with the miRNease Mini kit (Qiagen) according to the manufacturer's instructions. cDNA was synthesized from 1 μ g of total RNA using the miScript Reverse Transcription kit (Qiagen). Quantification of miRNAs was performed using the miScript SYBR Green PCR kit (Qiagen) and the Hs-let-7a and Hs-let-7d miScript Primer Assays (Qiagen), according to the manufacturer's instructions, on an ABI 7300 Real-Time PCR System. qRT-PCR was carried out using normalization to Hs-RNU6-2 as recommended. Differences in fold expression with regard to controls were calculated from triplicate C_t values following the 2^{-C_t} method. HPAT5 lincRNA levels were quantified in parallel according to the manufacturer's instructions. Experiments were carried out twice.

Differentiation of hESCs

Lentivirus-transduced H1 hESCs selected in 8 μ g/ml neomycin for 1 week were used for all of our differentiation experiments. H1 hESCs were maintained as feeder-free cultures on Matrigel under pluripotent conditions in hESC complete culture medium composed of DMEM/F12 and supplemented with 2.5 mM L-glutamine, 15 mM HEPES (Invitrogen), 20 μ g/ml insulin (Invitrogen), 64 μ g/ml L-ascorbic acid-2-phosphate (Sigma), 140 ng/ml sodium selenite (Sigma), 10.7 μ g/ml transferrin (Sigma), 100 ng/ml recombinant human FGF2 (Peprotech) and 2 ng/ml recombinant human transforming growth factor (TGF)- β (Peprotech).

For differentiation experiments, 24 h before transfection, cells were plated at 8×10^4 cells per well in Matrigel-coated 24-well plates. Transfections were performed with 50 pmol of scrambled or *POU5F1*-targeted siRNAs or with let-7a and let-7d miRNAs using Lipofectamine RNAiMAX according to the manufacturer's protocol. Four hours after transfection, the siRNA complexes were removed and cells were induced to differentiate by the removal of FGF2. The differentiation medium was changed daily, and cells were either collected at day 3 after transfection or passaged with Accutase, retransfected and collected at day 6. Transfections were performed in three replicate wells each for days 3 and 6. Cells were collected for gene expression analysis by direct lysis on the plate using the CellsDirect One-Step qRT-PCR kit (Invitrogen, 11753).

In vitro fibroblasts differentiated from hESCs were generated as previously described⁴⁵.

RNA immunoprecipitation

RIP was performed as previously described⁵⁵.

Immunoblot analysis

Immunoblotting after SDS-PAGE separation was performed on whole-cell protein samples (input) and immunoprecipitation eluates resuspended in SDS sample buffer. Antibodies to α -tubulin (tubulin; T6074) and Ago2 (AGO2; SAB4200085) were purchased from Sigma, and secondary horseradish peroxidase (HRP)-conjugated antibodies were purchased from Jackson ImmunoResearch.

Source and procurement of human embryos

Supernumerary human blastocysts from successful *in vitro*-fertilized (IVF) cycles, donated for basic research, were obtained with written informed consent from the Stanford University RENEW Biobank. Deidentification was performed according to the Stanford University Institutional Review Board-approved protocol entitled The RENEW Biobank (10466), and the molecular analysis of the embryos was in compliance with institutional regulations.

Thawing of human embryos

Human embryos frozen at the blastocyst stage were thawed using Quinn's Advantage Thaw kit (CooperSurgical) according to the manufacturer's protocol. In brief, cryocontainers were removed from liquid nitrogen and kept at room temperature for 30 s before incubating them in a 37 °C water bath until thawed. Thawed embryos were incubated in a 0.5 M and 0.2 M sucrose solution for 10 min each, washed in freeze-thaw diluent solution for an additional 10 min and kept for 2–4 h in Quinn's Advantage Cleavage Medium (CooperSurgical) supplemented with 10% serum protein substitute under mineral oil (Sigma) at 37 °C with 6% CO₂, 5% O₂ and 89% N₂.

Single-cell gene expression analysis of human embryos

Assays were validated as mentioned above (see "Assay performance validation"). We used the C1 Single-Cell Auto Prep System for single-cell capture and preamplification according to the manufacturer's instructions (protocol 100-4904). A total of four C1 chips were run. On each chip, 8–10 pooled human blastocysts were loaded that were previously enzymatically treated for single-cell dissociation. Only human blastocysts with morphologically distinct phenotypes (recognizable ICM and trophectoderm) were used for pooling and subsequent analysis. Hierarchical cluster analysis and PCA identified single cells that originated from either the ICM or trophectoderm on the basis of specific gene expression. This information was used to measure the gene expression of HPATs in each compartment.

Knockdown, immunofluorescence and RNA FISH on human and mouse preimplantation embryos

Knockdown of HPAT2, HPAT3 and HPAT5 was carried out in early human two-cell embryos. Briefly, siRNAs against HPAT2, HPAT3 and HPAT5 or scrambled siRNA were injected in one blastomere of two-cell embryos coinjected with dextran-tetramethylrhodamine (Invitrogen; 100 µg/ml). After cultivation to the blastocyst stage, successfully developed human blastocysts were analyzed for tetramethylrhodamine-positive cells and imaged live with a fluorescence microscope. For immunofluorescence analysis, blastocysts were briefly washed in M2 medium, and the zona pellucida was removed by treatment with acidic tyrodes solution. Embryos were fixed for 20 min in 3.7% paraformaldehyde in PBS at 4 °C and permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature. Fixed blastocysts were blocked overnight at 4 °C in 1% BSA, 0.1% Triton X-100 in PBS. Next, embryos were stained for 3 h with antibody to Oct4 (goat polyclonal, Santa Cruz Biotechnology, sc-8628) and antibody to Sox2 (goat polyclonal, Santa Cruz Biotechnology, sc-17320). After several washes in blocking solution, embryos were incubated at room temperature with anti-goat or anti-rabbit secondary antibody for 1.5 h coupled with Alexa Fluor 647 (10 µg/ml) or Alexa Fluor 488 (10 µg/ml; Molecular Probes). After incubation in DAPI (10 min, 1 µg/ml), blastocysts were mounted on slides with a small drop of Vectashield (Vector Laboratories) mounting medium. Blastocysts were then analyzed on a Zeiss LSM510 Meta inverted laser-scanning confocal microscope. ImageJ software was used to compute z stacks (~20 stacks with 0.4 µm per sample) for immunofluorescence images.

We performed single-molecule RNA FISH against HPAT3 and HPAT5 as previously described⁵⁶. Briefly, we designed fluorescently labeled (CAL Fluor Red 610) oligonucleotides specifically targeting each transcript (Biosearch Technologies; see URLs). Because of the repetitive sequence character, we custom designed the probes for HPAT3 and HPAT5 and maximized the probe number that specifically recognized only HPAT3 or HPAT5. Adherent cells were washed and fixed with 3.7% formaldehyde in 1× PBS for 10 min at room temperature. Cells were then washed twice with 1× PBS and permeabilized with 70% ethanol for 2 h at 4 °C. Probes were diluted in hybridization buffer (100 mg/ml dextran sulfate and 10% formamide in 2× SSC) at three different final concentrations initially to determine the optimal concentration (5 mM). Cells were washed once with wash buffer (10% formamide in 2× SSC) for 5 min at room temperature. Wash buffer was removed, and diluted probes were applied and samples were incubated overnight at 37 °C in a humidified chamber. Cells were washed once with wash buffer at 37 °C for 30 min and washed again (with wash buffer containing 5 ng/ml DAPI) to counterstain the nuclei at 37 °C for 30 min. Wash buffer was removed, and one drop of Vectashield mounting medium was applied to the slides, which were mounted with a coverglass. The coverglass was sealed with nail polish, and slides were imaged using a fluorescence confocal or bright-field microscope and the appropriate filters.

CRISPR design and HPAT5-KO derivation

CRISPR guide RNAs (gRNAs) were designed using the online CRISPR design tool from the Massachusetts Institute of Technology (see URLs). Candidate gRNAs with the highest score

were chosen for each genomic region. Oligonucleotides for these gRNAs were synthesized and cloned into plasmid pX459 (Addgene 48139) carrying both Cas9 and gRNA expression cassettes, with one modification of the original plasmid in which the *Cas9* promoter was replaced by the *EEF1A1* promoter. The cutting efficiency of each gRNA construct was validated by transfecting HEK293T cells and sequencing the target regions in the genome. CRISPR pairs were nucleofected into iPSC.BJ-WT cells and plated as single cells. Single cells were clonally expanded and isolated for PCR to test for successful *HPAT5* deletion and sequencing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

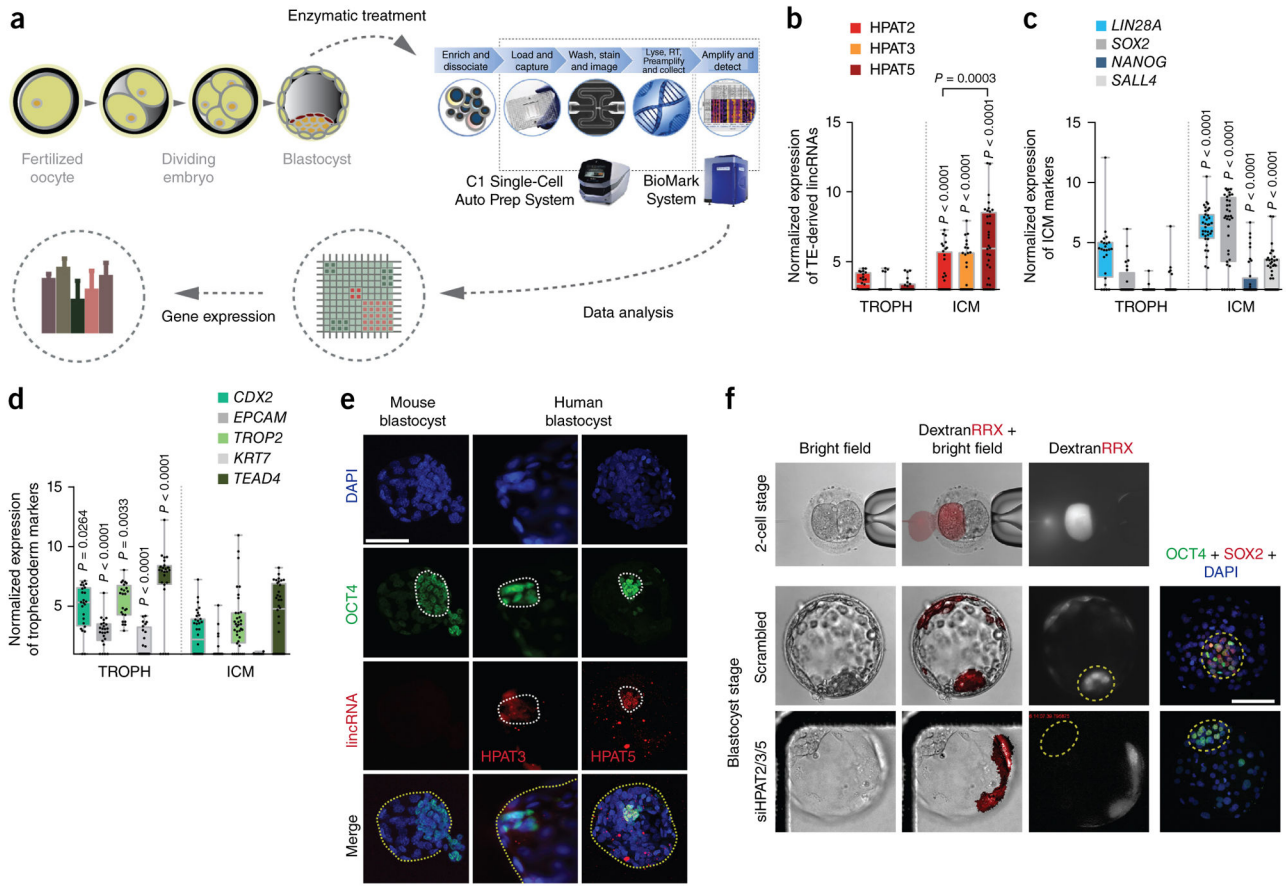
We thank members of the Reijo Pera laboratory for thoughtful feedback and comments regarding manuscript preparation. We thank Z. Siprashvili for assistance in the protein microarray experiments. We thank S. Marro for the neural transdifferentiation experiments and L.B. Torrez for the siRNA experiments. We thank G. Glinsky for thoughtful feedback regarding data assessment. This work was funded by grants U54-1U54HD068158-01 (Stanford University Center for Reproductive and Stem Cell Biology), U01-1U01HL100397-01 (Basic and Translational Research of iPSC-based Hematologic and Vascular Therapies), R01HG006018 (US National Institutes of Health) and RB3-2209 (California Institute of Regenerative Medicine). No federal funding was used for human embryo studies.

References

1. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012; 22:1775–1789. [PubMed: 22955988]
2. Jia H, et al. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA.* 2010; 16:1478–1487. [PubMed: 20587619]
3. Ørom UA, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 2010; 143:46–58. [PubMed: 20887892]
4. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013; 154:26–46. [PubMed: 23827673]
5. Kapusta A, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 2013; 9:e1003470. [PubMed: 23637635]
6. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008; 322:750–756. [PubMed: 18974356]
7. Martin L, Chang HY. Uncovering the role of genomic “dark matter” in human disease. *J Clin Invest.* 2012; 122:1589–1595. [PubMed: 22546862]
8. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA.* 2008; 105:716–721. [PubMed: 18184812]
9. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA.* 2009; 106:11667–11672. [PubMed: 19571010]
10. Koziol MJ, Rinn JL. RNA traffic control of chromatin complexes. *Curr Opin Genet Dev.* 2010; 20:142–148. [PubMed: 20362426]
11. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell.* 2007; 129:1311–1323. [PubMed: 17604720]
12. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem.* 2012; 81:145–166. [PubMed: 22663078]

13. Au KF, et al. Characterization of the human ESC transcriptome by hybrid sequencing. *Proc Natl Acad Sci USA*. 2013; 110:E4821–E4830. [PubMed: 24282307]
14. Kelley D, Rinn J. Transposable elements reveal a stem cell–specific class of long noncoding RNAs. *Genome Biol*. 2012; 13:R107. [PubMed: 23181609]
15. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. 2011; 147:1537–1550. [PubMed: 22196729]
16. Wang J, et al. Primate-specific endogenous retrovirus–driven transcription defines naive-like stem cells. *Nature*. 2014; 516:405–409. [PubMed: 25317556]
17. Lu X, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*. 2014; 21:423–425. [PubMed: 24681886]
18. Göke J, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*. 2015; 16:135–141. [PubMed: 25658370]
19. Grow EJ, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*. 2015; 522:221–225. [PubMed: 25896322]
20. Fort A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet*. 2014; 46:558–566. [PubMed: 24777452]
21. Harada F, Tsukada N, Kato N. Isolation of three kinds of human endogenous retrovirus–like sequences using tRNA^{Pro} as a probe. *Nucleic Acids Res*. 1987; 15:9153–9162. [PubMed: 2825129]
22. Rogers J, Gibbs RA. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet*. 2014; 15:347–359. [PubMed: 24709753]
23. Loewer S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. 2010; 42:1113–1117. [PubMed: 21057500]
24. Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology*. 2012; 9:111. [PubMed: 23253934]
25. Moignard V, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*. 2013; 15:363–372. [PubMed: 23524953]
26. Buganim Y, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*. 2012; 150:1209–1222. [PubMed: 22980981]
27. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005; 122:947–956. [PubMed: 16153702]
28. Gagliardi A, et al. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J*. 2013; 32:2231–2247. [PubMed: 23892456]
29. Loh YH, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*. 2006; 38:431–440. [PubMed: 16518401]
30. Wu Q, et al. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem*. 2006; 281:24090–24094. [PubMed: 16840789]
31. Zhang J, et al. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of *Pou5f1*. *Nat Cell Biol*. 2006; 8:1114–1123. [PubMed: 16980957]
32. Kunarso G, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*. 2010; 42:631–634. [PubMed: 20526341]
33. Alcidi EA, Tsukiyama T. ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. *Genes Dev*. 2014; 28:2348–2360. [PubMed: 25367034]
34. Zhu Y, Rowley MJ, Bohmdorfer G, Wierzbicki ATA. SWI/SNF chromatin-remodeling complex acts in noncoding RNA–mediated transcriptional silencing. *Mol Cell*. 2013; 49:298–309. [PubMed: 23246435]
35. Siprashvili Z, et al. Identification of proteins binding coding and non-coding human RNAs using protein microarrays. *BMC Genomics*. 2012; 13:633. [PubMed: 23157412]

36. Chendrimada TP, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*. 2005; 436:740–744. [PubMed: 15973356]
37. Li J, et al. Identifying mRNA sequence elements for target recognition by human Argonaute proteins. *Genome Res*. 2014; 24:775–785. [PubMed: 24663241]
38. Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*. 2010; 463:621–626. [PubMed: 20054295]
39. Worringer KA, et al. The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. *Cell Stem Cell*. 2014; 14:40–52. [PubMed: 24239284]
40. Ohms S, Lee SH, Rangasamy D. LINE-1 retrotransposons and let-7 miRNA: partners in the pathogenesis of cancer? *Front Genet*. 2014; 5:338. [PubMed: 25339972]
41. Ohms S, Rangasamy D. Silencing of LINE-1 retrotransposons contributes to variation in small noncoding RNA expression in human cancer cells. *Oncotarget*. 2014; 5:4103–4117. [PubMed: 24980824]
42. Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol*. 2009; 16:1021–1025. [PubMed: 19713958]
43. Heo I, et al. Lin28 mediates the terminal uridylation of *let-7* precursor microRNA. *Mol Cell*. 2008; 32:276–284. [PubMed: 18951094]
44. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet*. 2007; 39:380–385. [PubMed: 17259983]
45. Hockemeyer D, et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell*. 2008; 3:346–353. [PubMed: 18786421]
46. Karginov FV, et al. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci USA*. 2007; 104:19291–19296. [PubMed: 18042700]
47. Giordano J, et al. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput Biol*. 2007; 3:e137. [PubMed: 17630829]
48. Tanaka Y, Chung L, Park IH. Impact of retrotransposons in pluripotent stem cells. *Mol Cells*. 2012; 34:509–516. [PubMed: 23135636]
49. Holdt LM, et al. Alu elements in *ANRIL* non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through *trans*-regulation of gene networks. *PLoS Genet*. 2013; 9:e1003588. [PubMed: 23861667]
50. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. *Nature*. 2011; 470:284–288. [PubMed: 21307942]
51. Spengler RM, Oakley CK, Davidson BL. Functional microRNAs and target sites are created by lineage-specific transposition. *Hum Mol Genet*. 2014; 23:1783–1793. [PubMed: 24234653]
52. Lazzeroni L, Owen A. Plaid models for gene expression data. *Stat Sin*. 2002; 12:61–86.
53. Bengtsson M, Stahlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res*. 2005; 15:1388–1392. [PubMed: 16204192]
54. Pang ZP, et al. Induction of human neuronal cells by defined transcription factors. *Nature*. 2011; 476:220–223. [PubMed: 21617644]
55. Goff LA, et al. Ago2 immunoprecipitation identifies predicted microRNAs in human embryonic stem cells and neural precursors. *PLoS One*. 2009; 4:e7192. [PubMed: 19784364]
56. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods*. 2008; 5:877–879. [PubMed: 18806792]

**Figure 1.**

Molecular single-cell gene expression and functional analyses of HPAT2, HPAT3 and HPAT5 during human embryo development. **(a)** Overview of single-cell gene expression analysis for human blastocysts. A total of 24 blastocysts were pooled and run on four C1 chips. **(b–d)** Single-cell gene expression analyses. RT, reverse transcribe. **(b)** Three HPATs had significantly higher expressed in ICM than in trophoderm (TROPH). **(c,d)** ICM **(c)** and trophoderm **(d)** markers are also shown. Box plots are shown for each group. The whiskers are the minimum and maximum data points. The bottom and top of each box are the first and third quartiles, respectively. $n = 46$ single trophoectoderm cells and $n = 67$ single ICM cells. **(e)** Immunohistochemistry and RNA FISH for OCT4 (green) and lincRNAs (red), respectively, in human blastocysts. Sections are counterstained with DAPI (blue). ICMs are circled by dotted white lines. Entire human or mouse blastocysts are circled by dotted yellow lines in the merged images. lincRNA signal was specific to the ICM. Speckles (red) are nonspecific and are observed in all human blastocysts. Mouse blastocysts initiated hatching when fixed. Images are representative ($n = 9$ human blastocysts for HPAT3, $n = 11$ human blastocysts for HPAT5 and $n = 3$ mouse blastocysts; $n = 2$ independent experiments). Scale bar, 100 μ m. **(f)** Blastomeres with reduced expression of HPAT2, HPAT3 and HPAT5 during human embryo development did not contribute to ICM. The presence of ICM was validated with OCT4 and SOX2 staining. ICMs are circled by

yellow dashed lines ($n = 3$ blastocysts). RRX, Rhodamine Red-X; siHPAT2/3/5, combination of siRNAs targeting HPAT2, HPAT3 and HPAT5. Scale bar, 100 μm .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

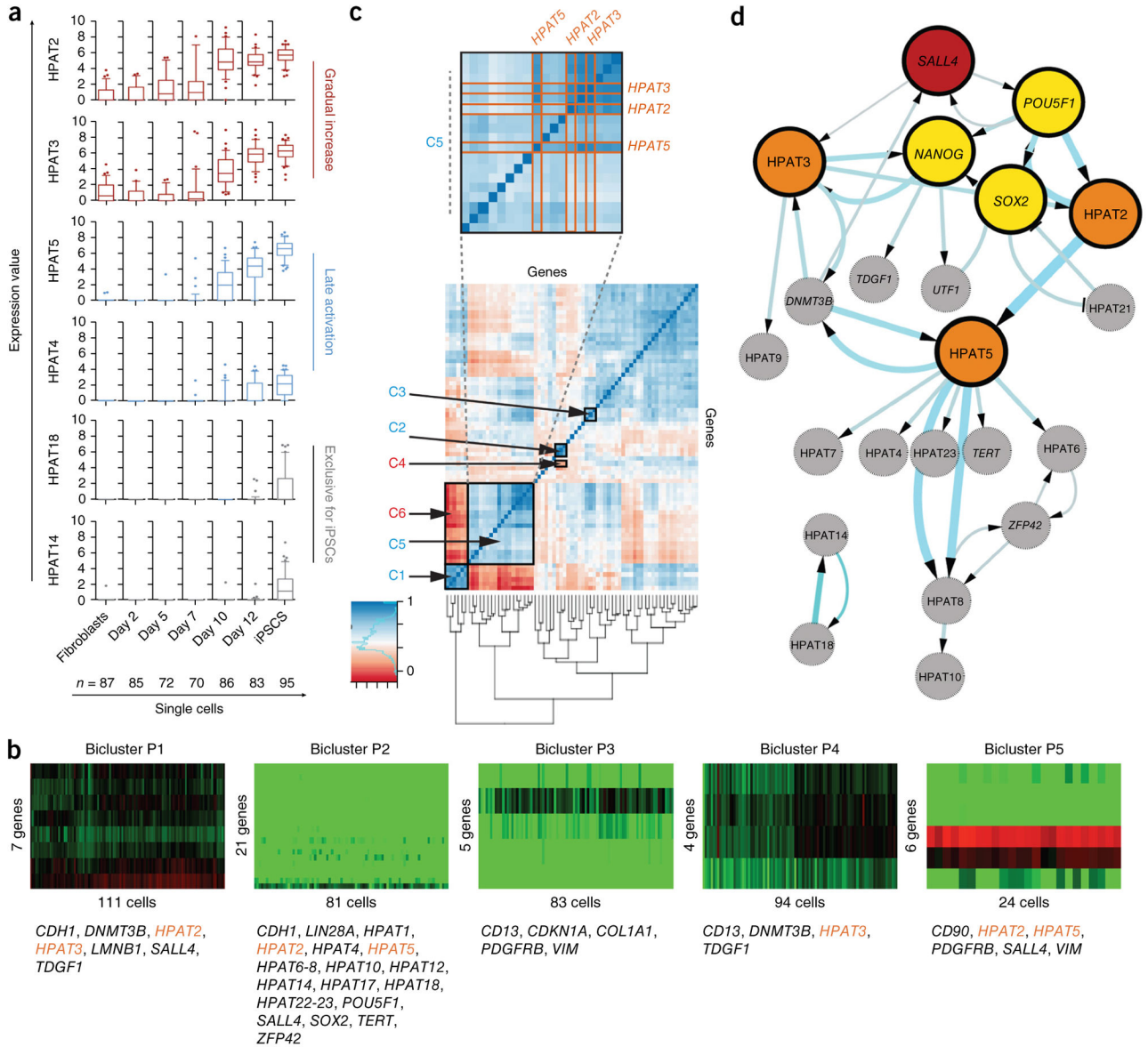


Figure 2. Single-cell expression analysis of HPAT transcripts during nuclear reprogramming. **(a)** Dynamics in single-cell expression of HPAT transcripts during nuclear reprogramming shown as box plots. HPATs are grouped according to activation pattern (with two examples representing each group). **(b)** Bicluster analysis identifies five biclusters, P1–P5, and correlates *HPAT2*, *HPAT3* and *HPAT5* (gene names in orange) with core pluripotency markers (method by Lazzeroni and Owen⁵²). **(c)** Correlation analysis across single cells identifies positively and negatively correlated gene pairs (see supplementary table 3 for additional details). **(d)** Bayesian network analysis predicts a central role for HPAT2, HPAT3 and HPAT5 (orange circles) within the core regulatory network (yellow circles). The hierarchical view predicts that *SALL4* (red circle) triggers a cascade of key pluripotency gene activation. Arrow thickness and circle size increase with the confidence level of

interactions in the calculated network. Only pluripotency genes and lincRNAs were included. Data in **a–d** represent $n = 578$ single cells.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

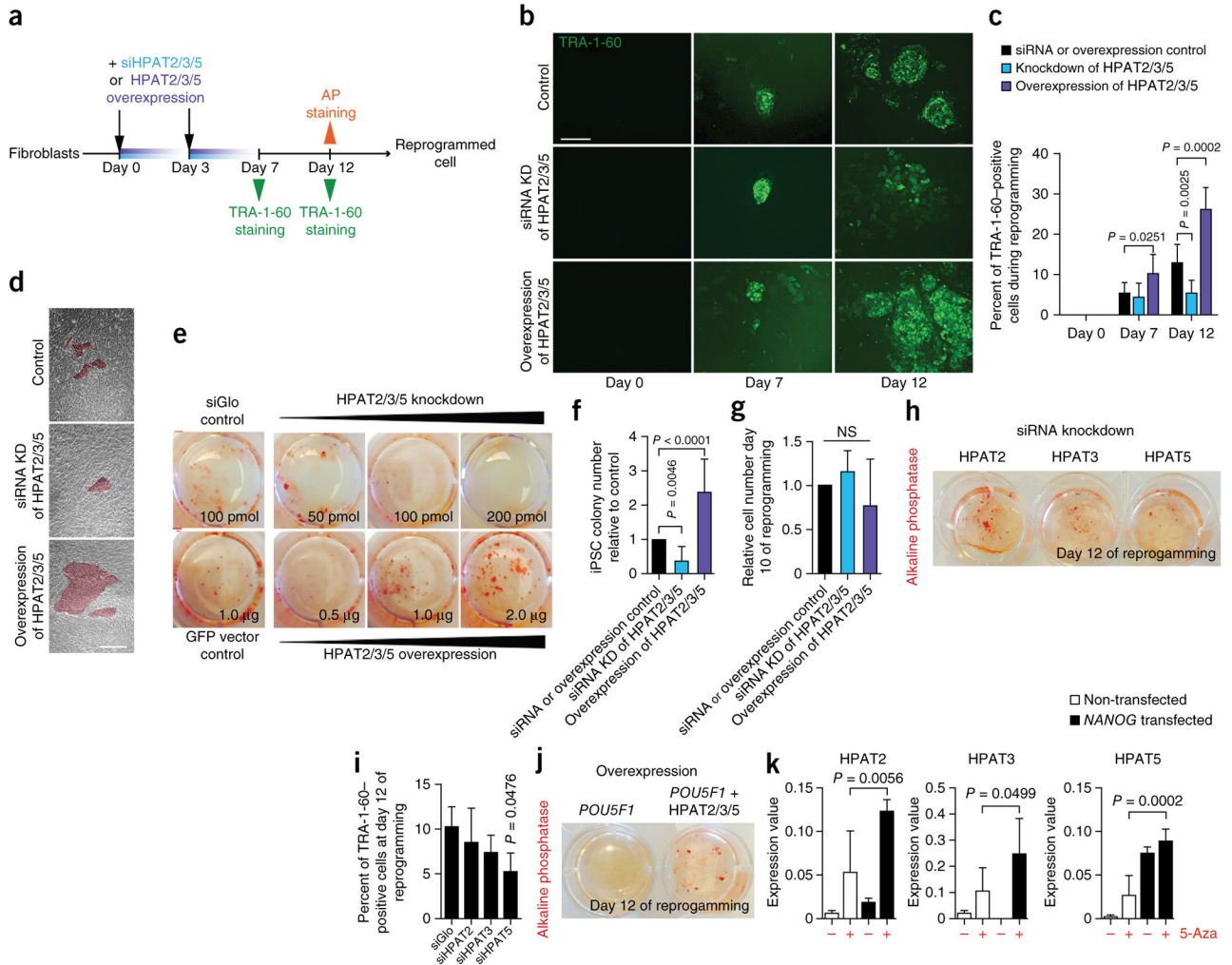


Figure 3. Functional analyses of HPAT2, HPAT3 and HPAT5 during nuclear reprogramming. **(a)** Experimental scheme of functional analysis of HPAT2, HPAT3 and HPAT5 (HPAT2/3/5) during iPSC reprogramming. AP, alkaline phosphatase. **(b)** Immunostaining of TRA-1-60 during reprogramming with HPAT2, HPAT3 and HPAT5 in combination. Representative images are shown ($n = 8$). KD, knockdown. Scale bar, 100 μm . **(c)** Calculated percentage of TRA-1-60–positive cells at different points during reprogramming ($n = 8$). Data are represented as means + s.e.m. **(d)** Representative images of colony size appearance at day 12 during reprogramming ($n = 8$). Scale bar, 100 μm . **(e)** Alkaline phosphatase staining at day 12. Shown are the wells of a six-well plate from one experiment ($n = 2$ independent experiments). **(f)** iPSC colony number counted on the basis of alkaline phosphatase staining ($n = 3$). Data are represented as means + s.e.m. **(g)** Cell number relative to control cells during reprogramming ($n = 3$). Data are represented as means + s.e.m. NS, not significant. **(h)** Alkaline phosphatase staining at day 12 during reprogramming ($n = 2$ for each condition). **(i)** Percentage of TRA-1-60–positive cells at day 12 of single knockdown of HPAT2, HPAT3 or HPAT5. siGlo was used as a control. Data are represented as means +

s.e.m. (j) Reprogramming with *POU5F1* and HPAT2, HPAT3 and HPAT5. *POU5F1* only is used as a control. (k) Epigenetic and gene expression analysis of *HPAT2*, *HPAT3* and *HPAT5*. *NANOG* mRNA was transfected into BJ fibroblasts treated or not with 5-Aza-2'-deoxycytidine (5-Aza) with gene expression measurement at 48 h ($n = 6$). Data are represented as means + s.e.m.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

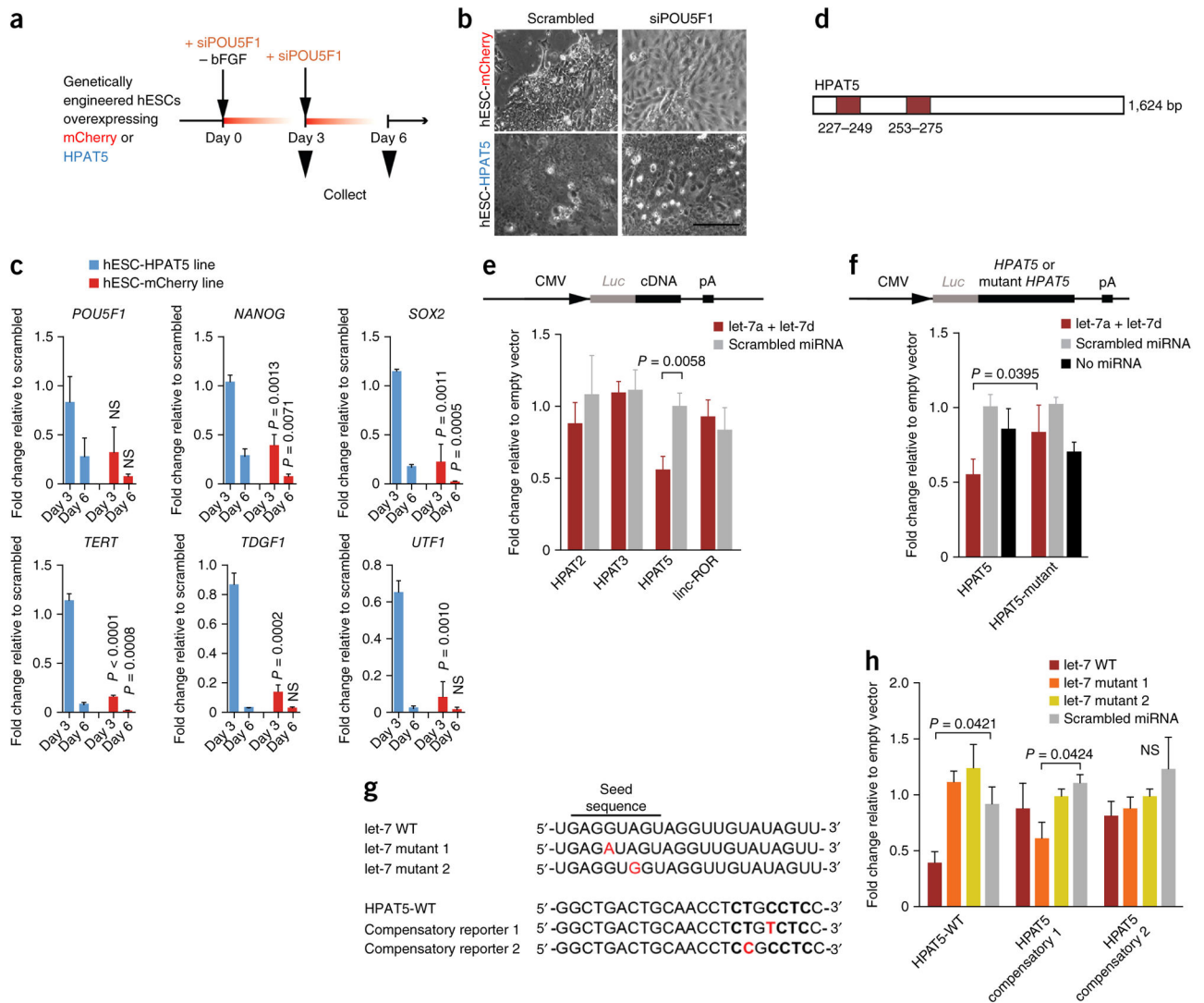


Figure 4. HPAT5 binds directly to let-7. (**a-c**) HPAT5-OE hESCs suppress differentiation mediated by siRNA to *POU5F1* and bFGF removal. (**a**) Outline of the protocol. (**b,c**) The expression of key pluripotency markers decreases with delay in comparison to the mCherry-OE control line. *P* values are calculated for the comparison of the mCherry-OE and HPAT5-OE lines on the same days (**c**), with this evidence supported by morphological changes (after day 3) (**b**) ($n = 2$ independent experiments). Scale bar, 50 μ m. (**d**) Two predicted let-7 binding positions in HPAT5 identified by RegRNA 2.0 (miRanda). (**e,f**) Target validation using luciferase reporters in HEK293 cells. The relative luciferase activity (shown as fold change relative to empty vector) was assayed 48 h after cotransfection of cells with the indicated miRNAs or control (scrambled miRNA) together with wild-type reporter (**e**) or the let-7 miRNAs together with wild-type or mutant reporter (**f**). (**g**) The point mutations (in red) introduced into two mutant let-7 mimics and two mutant HPAT5 reporters for compensatory analysis. WT, wild type. (**h**) Analysis of the effects of the point mutations using luciferase reporters in HEK293 cells. Relative luciferase activity (shown as fold change relative to empty vector) of

wild-type or mutant reporters was assayed 48 h after cotransfection of cells with the indicated miRNAs or scrambled miRNA. Representative results from $n = 2$ (**c,e,f,h**) independent experiments; $n = 3$ samples; data are shown with s.e.m.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

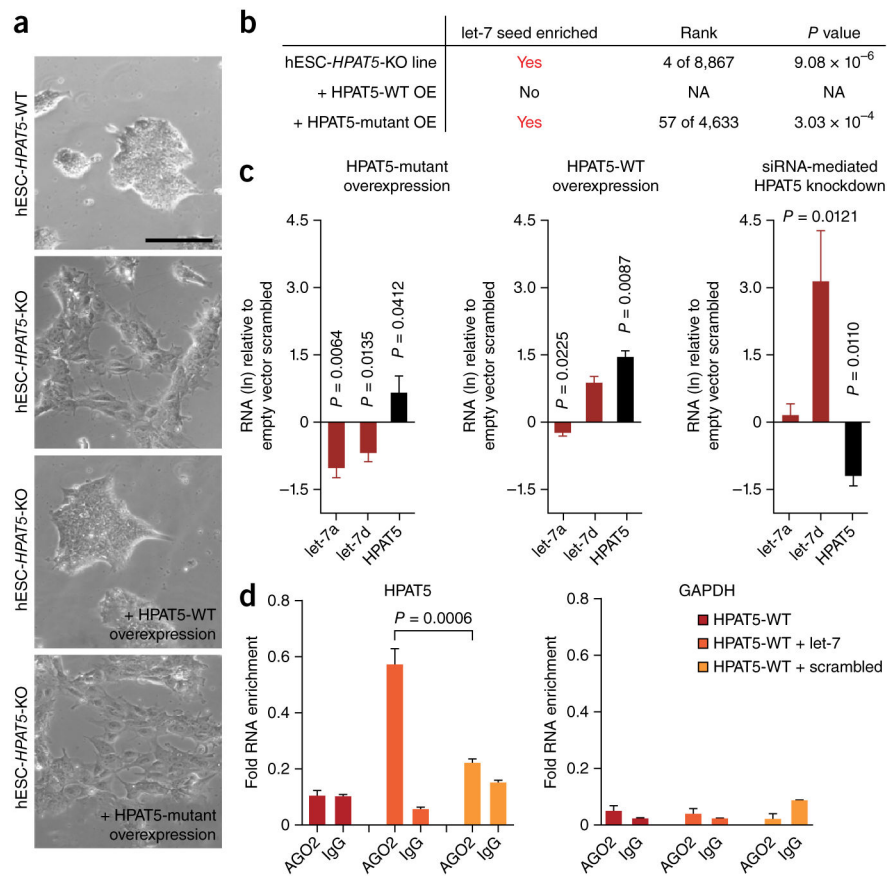


Figure 5. HPAT5 regulates let-7 in hESCs during differentiation. **(a)** Morphological changes 48 h after let-7 overexpression in *HPAT5*-WT and *HPAT5*-KO lines. **(b)** Table listing findings from enrichment analysis with cWords (supplementary table 8). NA, not applicable. **(c)** HPAT5 regulates let-7 activity. Expression levels of mature let-7a and let-7d in undifferentiated hESCs (H1) transiently transfected for 48 h with wild-type (WT) or mutant HPAT5 or with knockdown with siHPAT5. Empty vector or scrambled miRNA was used as a negative control. RNA levels (ln) and *P* values are shown relative to negative controls. *n* = 3 samples; data are shown with s.e.m. **(d)** RIP-qPCR showing interaction between AGO2 and HPAT5 but not GAPDH in hESCs transfected with let-7. *n* = 3 samples; data are shown with s.e.m.