**Supplementary Materials and Methods:**

***Cell Culture and Transfection***

E14Tg2a (E14) and J1 cells were obtained from Mutant Mouse Research Resource Centers and American Type Culture Collection, respectively. Oct4GiP cells were kindly provided by Dr. Austin Smith. ESCs were maintained on gelatin-coated plates in ESGRO complete plus clonal grade medium (Millipore, Billerica, MA).

For all transfections, ESCs were cultured in gelatin-coated plates in M15 medium: DMEM (Invitrogen, Grand Island, NY) supplemented with 15% FBS, 10 mM 2-mercaptoethanol, 0.1 mM nonessential amino acids (Invitrogen), 1x EmbryoMax nucleosides (Millipore, Billerica, MA), 1,000 U of ESGRO (Millipore, Billerica, MA). Lipofectamine 2000-mediated transfections were performed as previously described (Zheng et al, 2012). "Mock" transfections were performed using lipofectamine but without siRNAs. "siControl" transfections were performed using siRNA against luciferase (CGTACGCGGAATACTTCGA, synthesized by Dharmacon, Thermo Scientific).

For knock-down assessment, cells were harvested two days post-transfection for qRT-PCR or on day 3 for western analysis as previously described (Zheng et al, 2012). For lineage marker and microarray analysis, cells were harvested for RNA extraction and quantitative real-time polymerase chain reaction (qRT-PCR) on day 4, post-transfection. Similarly, for alkaline phosphatase (AP) staining, cells were stained for AP activity with the Alkaline Phosphatase Detection Kit (Millipore, Billerica, MA) on day 4.

For EB differentiation, E14Tg2a cells were transfected with control or Fip1-siRNA2. 48 hrs post transfection, cells were collected and plated in ultralow-attachment 24-well plate (Corning) at 3 x $10^5$/well in M15 without LIF, and this time point was referred to as day-0 for EB formation. Medium was changed every other day, and EBs were collected at day-2, 4, 6 for RT-qPCR analysis.

For colony formation assay, E14Tg2a cells were transfected with lipid-only (mock), non-targeting siRNA (control), Fip1-siRNA2 or Fip1-siRNA3. 24 hrs post transfection, cells were plated at clonal density in 6-well plates. Colonies were stained with alkaline phosphatase staining kit (Stemgent) on day-8, and ~ 200 colonies were counted for each transfection. Colonies were categoried as undifferentiated, partially differentiated, or differentiated based on morphology and alkaline phosphatase staining.

HeLa cells were grown in DMEM plut 10% fetal bovine serum under standard conditions. For Fip1 knockdown, a pSUPERIOR.puro construct containing shRNA sequences based on Openbiosystems TRCN0000074420 was transfected into HeLa cells using Lipofectamine 2000 and stable transfectants were selected in the presence of 1μg/mlof puromycin. Colonies with optimal

knockdown efficiency were identified by Western analysis and used for further analysis.

### *Oct4GiP Reporter Assay*

Oct4GiP reporter assay was carried out as described previously (Zheng & Hu, 2012). For each experimental condition, 3-6 independent transfections were carried out.

### *Reprogramming*

C57BL6 MEFs or C57BL6x129SVA Oct4-GFP MEFs (P0) were plated at $5x10^4$ cells/well in 12-well plate. Cells were transduced with pGIPZ-NT or pGIPZ-Fip1-shRNA5 (RHS4346, RMM4431-200361996, Openbiosystems, Thermo Scientific; Fig. 4 and fig. S10) or with pHAGE-Mir-FF or pHAGE-Mir-Fip1-shRNA5 virus (fig. S8) the next day at a MOI of 4. Media was changed 24 hrs after virus transduction and cells were re-plated at 1:4 the next day in 12-well plate. 72 hrs after initial virus transduction, cells were transduced with STEMCCA virus (Sommer et al, 2009) at a MOI of 4, and this time point was referred to as “day-0” for the reprogramming process. Cells were re-plated on gelatin-coated plates at day-1 by a 1:4 passage in a 12-well plate. Medium was changed to M15 on day-2. Cells were fed every other day until day-12. Cells were AP stained and iPS colonies were counted for 6 independent trials of reprogramming. Results report average ± SEM.

### *Luciferase Assay*

ESCs were transfected in triplicate in 24-well format, using the lipid-mediated protocol, as described above. For Fig. 3A, pMIR-REPORT constructs containing various 3’ UTRs were co-transfected with p*Renilla* at a ratio of 1:50. For Fig. 2E and fig. S4B, pPASPORT constructs containing various PASs were transfected. Cells were grown for two days in M15. Lysates were prepared using 150ul lysis buffer per well and otherwise prepared following the Dual Luciferase Reporter Assay System (Promega, Madison, WI). Luciferase reads were collected using a Bio-Tek Synergy HT Multi-Mode Microplate Reader and analyzed by Gen5 software (BioTek, Winooski, VT).

### *Quantitative RT-PCR*

Total RNAs were prepared from cells with the Isolate RNA Mini kit (Bioline, Taunton, MA). For reverse transcription (RT) prior to gene expression analysis, cDNAs were generated using the iScript (Bio-Rad, Hercules, CA) for lineage marker analysis or SuperScript II (Invitrogen, Grand Island, NY) kits with oligo-d(T) priming for APA analysis, according to manufacturer’s instructions. Primers were designed and tested using a linear dilution series of input cDNAs. qRT-PCR were performed for expression analysis and APA site choice validation, respectively. All primers used will be provided in the final submission.

### *Western Blotting*

The following antibodies were used for Western blotting analyses: Fip1 (Bethyl Laboratories, A301-462A); CPSF100 (Bethyl, A301-581A); CPSF73 (Bethyl, A301-090A); CPSF30 (Bethyl, A301-584A); CFIm25/Nudt21 (Santa Cruz, sc-81109); eRF1/ETF1 (Genetex, GTX108271); Ncaph2 (Genetex, GTX45270); WWP2 (Bethyl, A302-935A); β-actin (Sigma, St. Louis, MO; A1978); GAPDH (Santa Cruz Biotechnology, sc-32233).

***Microarray Analysis***

ESCs were transfected with Fip1-2 and Fip1-3 siRNAs (siFip1-2 and siFip1-3), as described above, in triplicate. Gene expression analysis was conducted using Affymetrix Whole Mouse Genome oligo arrays (Affymetrix) following the Affymetrix 1-color microarray-based gene expression analysis protocol. Raw CEL files were processed using relevant R/Bioconductor packages as described previoulsy (Freudenberg et al, 2012). Genes were considered differentially expressed if their FDR $\leq 0.01$ and fold change $\geq 2$. Since global gene expression changes due to Fip1 KDs, using Fip1-2 and Fip1-3 siRNAs, were highly similar, we treated them as biological replicates in the differential expression analysis. Microarray data has been submitted (accession number GSE47907).

For ESC EB differentiation time course, raw expression data was downloaded from GEO (accession GSE3749) and re-processed using the RMA methodology and Entrez gene based re-annotated CDFs (version 13) to summarize probes. A linear model fit with the time points as factors was computed using the R limma package and variance estimates were adjusted using an empirical Bayes method. Heatmap for Log2 expression values of all genes or the 311 Fip1 APA target genes were plotted in R and genes were ordered based on their expression level in ESCs.

***Direct RNA Sequencing and data analysis***

Direct RNA sequencing (DRS) was performed by Helicos BioSciences, and DRS reads were aligned to the mouse reference genome (mm9) using the index-DP genomic tool in Helisphere (Helicos BioSciences). Only uniquely mapped reads with a minimum mapped length of 25 and an alignment score of 4.0 were kept. We further filtered reads that arose from internal poly(A) priming, as previously described (Yao et al, 2012). For the remaining reads, we designated their 5' ends as the corresponding poly(A) sites (PASs). To construct a consensus PAS annotation for downstream analysis, we clustered all individual PASs within a 40 nt window on the same chromosome strand and calculated a weighted coordinate as the designated PAS, as described previously (Yao et al, 2012). All DRS data has been deposited to NCBI SRA database (accession number: SRP025988).

Statistics for the two samples are included in the table below:

| DRS | ESC | Fip1 kd |
|---|---|---|
| Total reads | 4,013,145 | 4,151,182 |
| Mapped reads | 2,794,692 | 2,926,370 |
| Mapped to 3' UTR | 2,419,997 | 2,467,101 |

| Mapped Peaks | 289,059 | 292,984 |
| --- | --- | --- |
| Peaks mapped to 3’ UTR | 51,004 | 52,395 |
| Expressed genes (cut-off=1) | 22,388 | 22,587 |
| Genes with APA (cut-off=1) | 17,312 | 17,484 |
| Expressed genes (cut-off=10) | 10,229 | 10,312 |
| Genes with APA (cut-off=10) | 3,161 | 3,415 |

***Alternative polyadenylation analysis***
All DRS read clusters were mapped to ensemble genes. Next, we used Fisher-exact test to compare the ratio of the DRS read counts of one PAS to the sum of the read counts of all other PASs within the same gene. The p-values were adjusted by the Benjamini-Hochberg method to control false discovery rate (FDR). PASs with an FDR <=1.0e-4 were defined as significantly changed PASs. To create the scatterplot shown in Fig. 2A, we selected two PASs with the smallest p-values for genes with multiple poly(A)s, and the PAS closer to the transcript start site is designated as the proximal PAS and the other as distal PAS. We then calculated the corresponding proximal/distal ratio. PAS pairs with an FDR <=1.0e-4 and |log10 (ES/Fip1 KD-proximal/distal ratio)| >0.2 are highlighted. Genes with proximal-to-distal shifts are highlighted in blue, while those with distal-to-proximal shifts are highlighted in red.

***In vitro cleavage/polyadenylation assay***
PAS sequence (-100nt to +100nt of the cleavage site) of interest was cloned into pBluescript and RNA substrate was synthesized by in vitro transcription in the presence of $^{32}$P-$\alpha$-UTP. In vitro cleavage/polyadenylation was performed with HeLa nuclear extract under standard conditions, as previously described (Yao et al, 2012). For Fig. 2D, nuclear extract was prepared from control HeLa cells and a HeLa cell line that stably expresses a Fip1-specific shRNA.

***Alternative Polyadenylation Validation by RT-qPCR***
For respective detection of the common and extended 3’ UTR in Mock and Fip1 depleted samples, primers were designed to detect each species. UTR locations were determined using the annotations from the UCSC genome browser and the DRS data. Primer efficiency was first determined using a linear dilution series with cDNAs from mESCs (primer efficiency was determined by a slope of -3.33±0.2). Mock, siControl, siFip1-2, and siFip-3 samples from day 4 post-transfection were harvested and RNA was extracted as described above. RT was performed as described above using First-Strand Synthesis (Invitrogen, Grand Island, NY). Primer sequences will be listed in the final submission.

***Mapping Fip1-RNA interactions by iCLIP-seq***
Fip1 iCLIP-seq was carried out as previously described (Yao et al, 2012). For the analysis shown in Fig. 5E, for all genes in each group, we first normalize the iCLIP signals by DRS read counts. For distal PAS, the DRS read count for the distal PAS was used. For proximal PAS, the total read count for both proximal

and distal PAS were used as Fip1 iCLIP signals are from both the short and the long APA isoform. Then we divided the cUTR or aUTR into 200 bins. Normalized the iCLIP signals within the same bin from all genes in the same group were added up and then divided by the total number of genes within each group. The final normalized iCLIP signal (per gene) (y axis) is plotted against the position within the 3' UTR (percentage of UTR, x axis). The iCLIP map in Supplementary Figure S26 (bottom panel) was made in the same way. For Supplementary Figure S26 (top panel), the iCLIP signals at the same position within the -100nt to +100nt region were added up and normalized as described above.

***MicroRNA target site analyis of APA genes***

Predicted microRNAs sites were downloaded from "TS miRNA sites" track on the UCSC Genome Browser. For APA genes whose proximal and distal PASs are located in the same terminal exons (212 of the 311 P2D genes, 24 of the 63 D2P genes), we count number of microRNA target sites in the cUTR and aUTR separately, then each counted number are normalized by the corresponding cUTR or aUTR length to get the microRNA target site density (per kb).

The sequences of all DNA and RNA oligos used this study are listed in Supplementary Table 4.

**Reference:**

Freudenberg JM, Ghosh S, Lackford BL, Yellaboina S, Zheng X, Li R, Cuddapah S, Wade PA, Hu G, Jothi R (2012) Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res* 40(8): 3364-3377

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456(7221): 464-469

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G (2009) Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* 27(3): 543-549

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A* 109(46): 18773-18778

Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, Jothi R, Hu G (2012) Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. *Stem Cells* 30(5): 910-922

Zheng X, Hu G (2012) Oct4GiP reporter assay to study genes that regulate mouse embryonic stem cell maintenance and self-renewal. *J Vis Exp*(63)