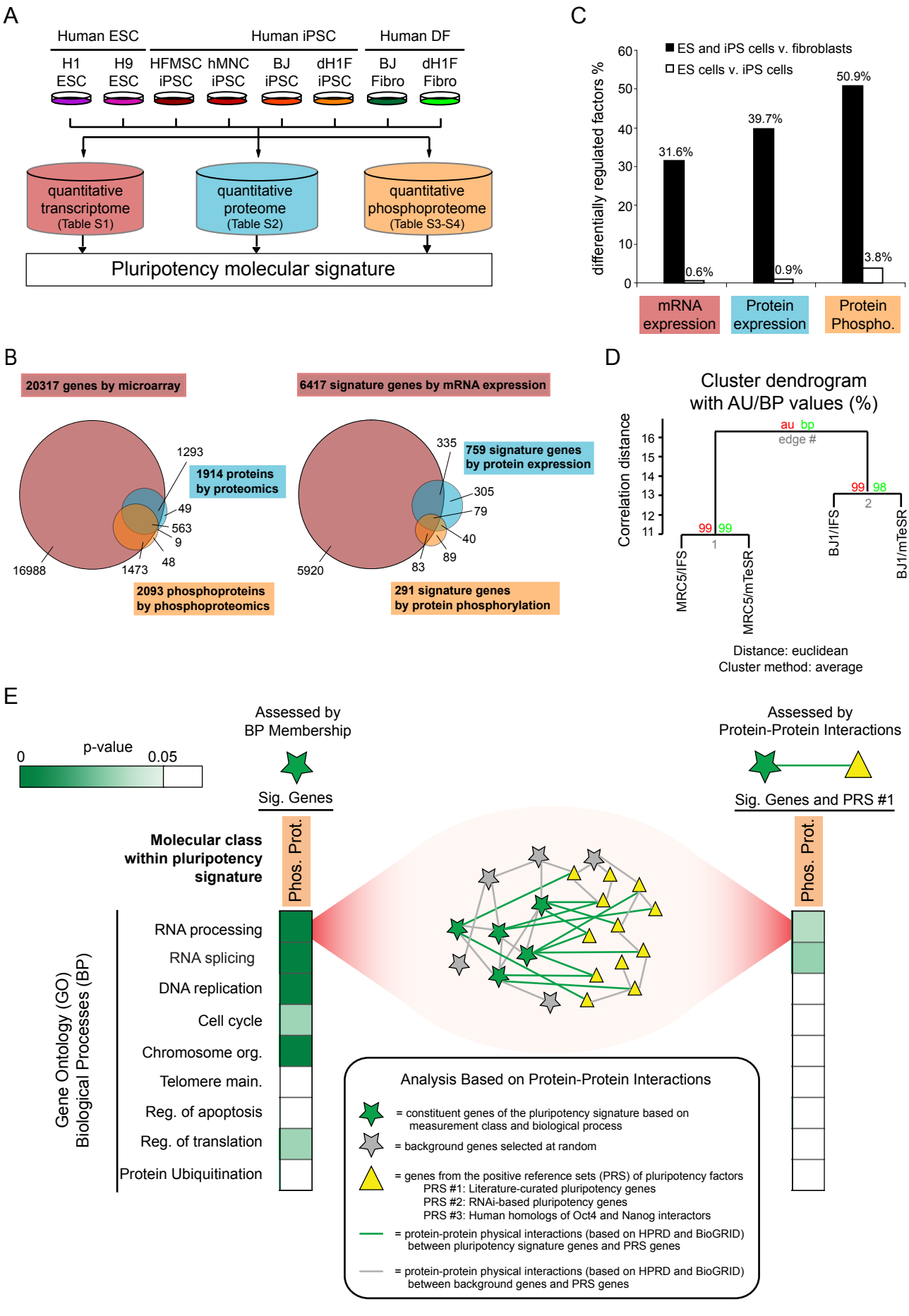
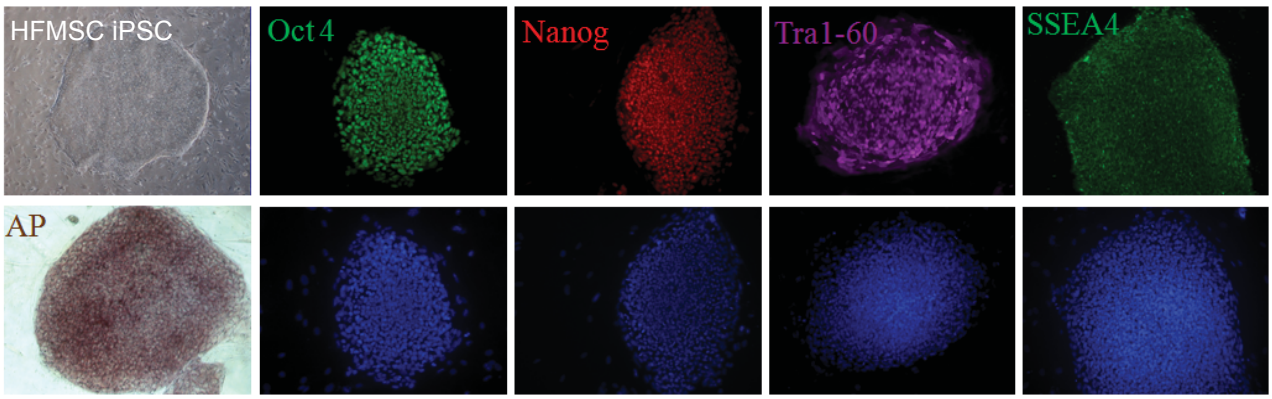
Lu Figure S1.

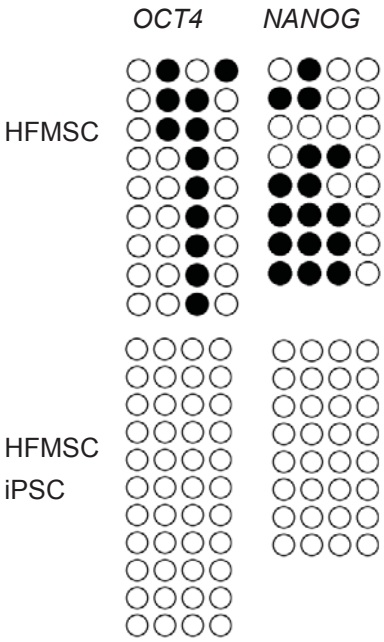

A
Human ESC
Human iPSC
Human DF
H1 ESC
H9 ESC
HFMSC iPSC
hMNC iPSC
BJ iPSC
dH1F iPSC
BJ Fibro
dH1F Fibro
quantitative transcriptome (Table S1)
quantitative proteome (Table S2)
quantitative phosphoproteome (Table S3-S4)
Pluripotency molecular signature
C
ES and iPS cells v. fibroblasts
ES cells v. iPS cells
differentially regulated factors %
31.6%
0.6%
39.7%
0.9%
50.9%
3.8%
mRNA expression
Protein expression
Protein Phospho.
B
20317 genes by microarray
6417 signature genes by mRNA expression
1914 proteins by proteomics
759 signature genes by protein expression
2093 phosphoproteins by phosphoproteomics
291 signature genes by protein phosphorylation
16988
1293
49
563
9
48
1473
5920
335
305
79
40
89
83
D
Cluster dendrogram with AU/BP values (%)
Correlation distance
au bp
edge #
MRC5/IFS
MRC5/mTeSR
BJ1/IFS
BJ1/mTeSR
Distance: euclidean
Cluster method: average
E
p-value
0
0.05
Assessed by BP Membership
Sig. Genes
Assessed by Protein-Protein Interactions
Sig. Genes and PRS #1
Molecular class within pluripotency signature
Phos. Prot.
Gene Ontology (GO) Biological Processes (BP)
RNA processing
RNA splicing
DNA replication
Cell cycle
Chromosome org.
Telomere main.
Reg. of apoptosis
Reg. of translation
Protein Ubiquitination
Analysis Based on Protein-Protein Interactions
= constituent genes of the pluripotency signature based on measurement class and biological process
= background genes selected at random
= genes from the positive reference sets (PRS) of pluripotency factors
PRS #1: Literature-curated pluripotency genes
PRS #2: RNAi-based pluripotency genes
PRS #3: Human homologs of Oct4 and Nanog interactors
= protein-protein physical interactions (based on HPRD and BioGRID) between pluripotency signature genes and PRS genes
= protein-protein physical interactions (based on HPRD and BioGRID) between background genes and PRS genes

Lu Figure S2.

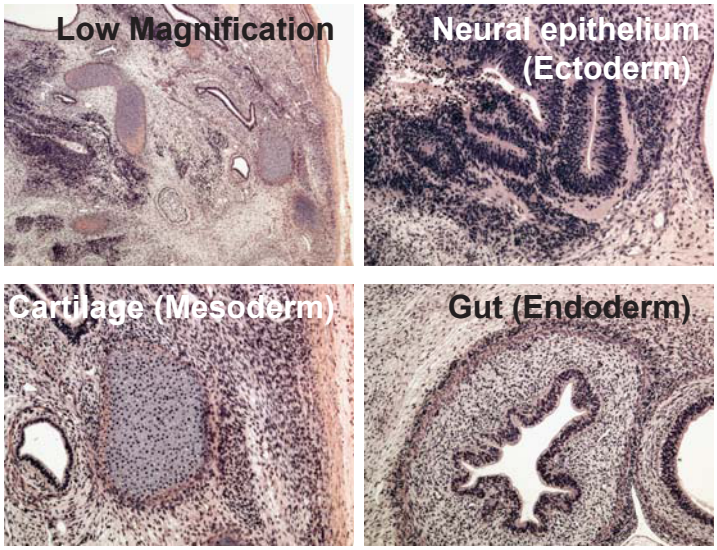
A



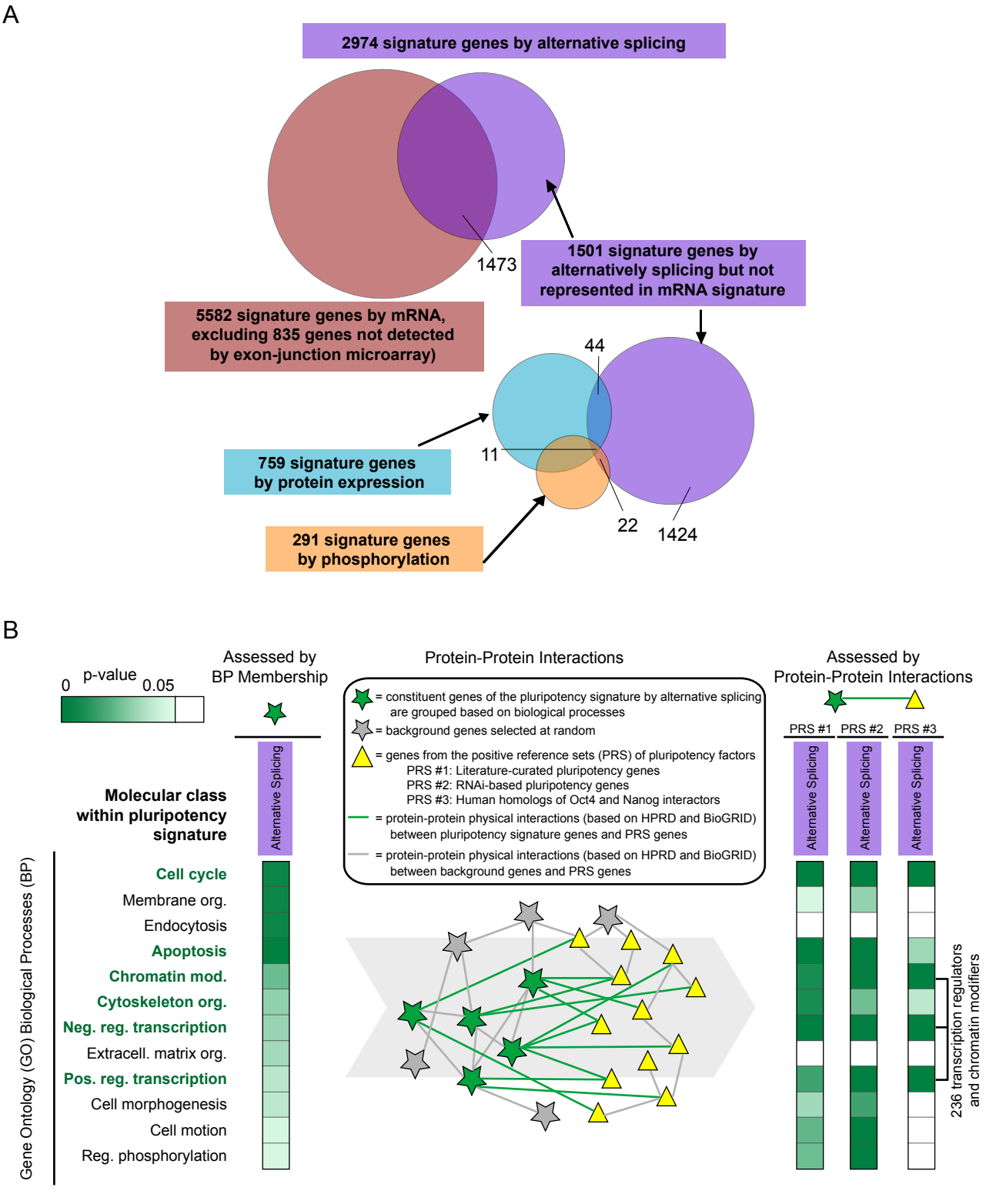
B



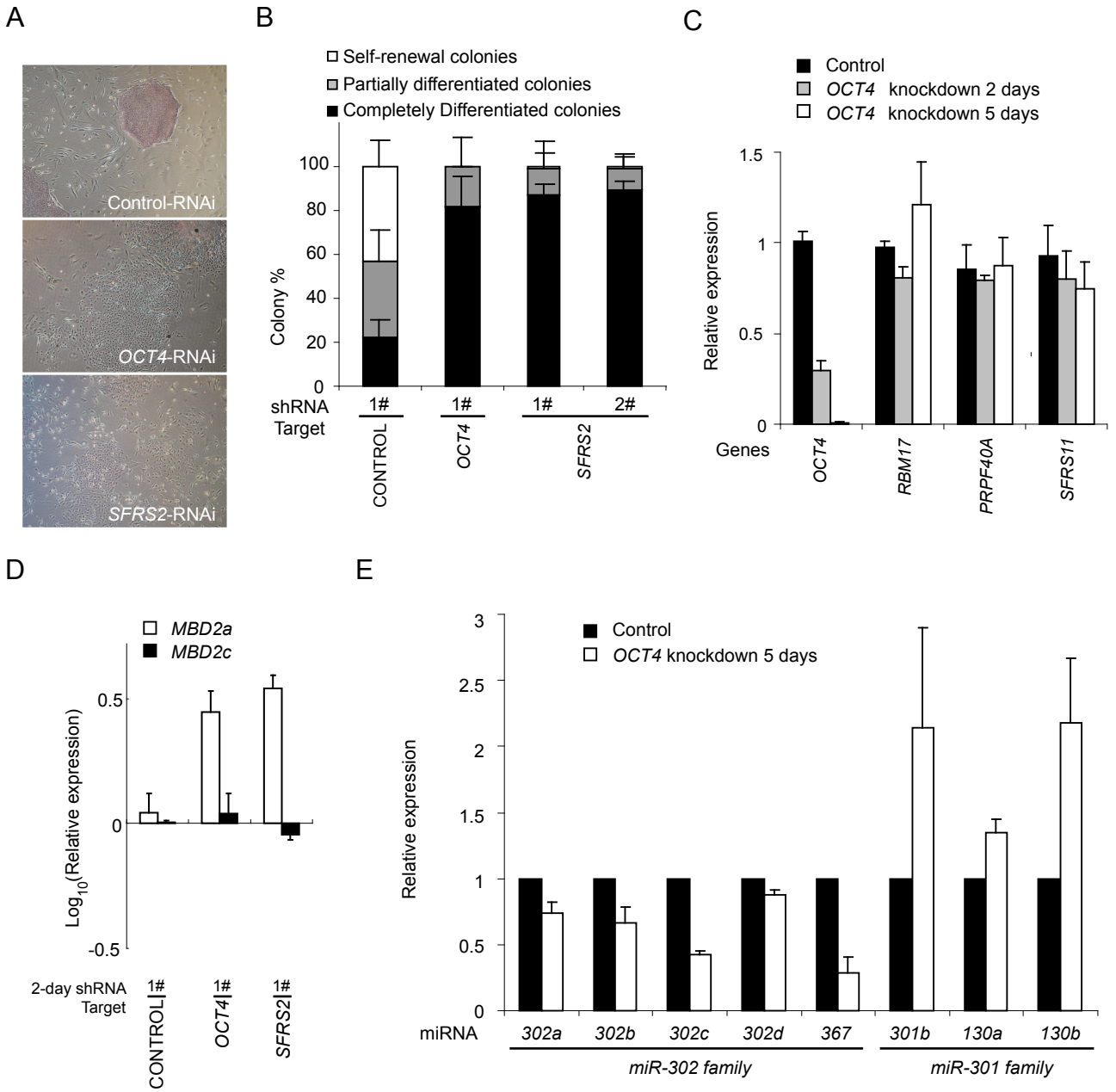
C



Lu Figure S3.


A
2974 signature genes by alternative splicing
1473
1501 signature genes by alternatively splicing but not represented in mRNA signature
5582 signature genes by mRNA, excluding 835 genes not detected by exon-junction microarray)
44
11
22
1424
759 signature genes by protein expression
291 signature genes by phosphorylation
B
p-value
0
0.05
Assessed by BP Membership
Protein-Protein Interactions
= constituent genes of the pluripotency signature by alternative splicing are grouped based on biological processes
= background genes selected at random
= genes from the positive reference sets (PRS) of pluripotency factors
PRS #1: Literature-curated pluripotency genes
PRS #2: RNAi-based pluripotency genes
PRS #3: Human homologs of Oct4 and Nanog interactors
= protein-protein physical interactions (based on HPRD and BioGRID) between pluripotency signature genes and PRS genes
= protein-protein physical interactions (based on HPRD and BioGRID) between background genes and PRS genes
Assessed by Protein-Protein Interactions
PRS #1
PRS #2
PRS #3
Alternative Splicing
Molecular class within pluripotency signature
Gene Ontology (GO) Biological Processes (BP)
Cell cycle
Membrane org.
Endocytosis
Apoptosis
Chromatin mod.
Cytoskeleton org.
Neg. reg. transcription
Extracell. matrix org.
Pos. reg. transcription
Cell morphogenesis
Cell motion
Reg. phosphorylation
236 transcription regulators and chromatin modifiers

Lu Figure S4.


A
Control-RNAi
OCT4-RNAi
SFRS2-RNAi
B
Self-renewal colonies
Partially differentiated colonies
Completely Differentiated colonies
Colony %
100
80
60
40
20
0
shRNA Target
1#
1#
1#
2#
CONTROL
OCT4
SFRS2
C
Control
OCT4 knockdown 2 days
OCT4 knockdown 5 days
Relative expression
1.5
1
0.5
0
Genes
OCT4
RBM17
PRPF40A
SFRS11
D
MBD2a
MBD2c
Log10(Relative expression)
0.5
0
-0.5
2-day shRNA Target
1#
1#
1#
CONTROL
OCT4
SFRS2
E
Control
OCT4 knockdown 5 days
Relative expression
3
2.5
2
1.5
1
0.5
0
miRNA
302a
302b
302c
302d
367
301b
130a
130b
miR-302 family
miR-301 family

Supplemental Figures S1-S4:

Figure S1. Human ESC and iPSC exhibit unique molecular signatures as compared to differentiated fibroblasts, related to Figure 1.

(A) Experimental strategy for identification of molecular signatures in pluripotent cells. Human ESC (H1 and H9 (Thomson et al., 1998)), human iPSC (BJ-iPSC and dH1F-iPSC (Park et al., 2008)), the corresponding parental fibroblast lines (BJ and dH1F), and two additional human iPSC from peripheral blood mononuclear cells (hMNC-iPSC (Loh et al., 2010)) and mesenchymal stem cells isolated from the dermal sheath/papilla of hair follicles (HFMSC-iPSC, pluripotency validated in Figure S2), respectively, were cultured and processed for comparative analysis of mRNA, protein, and phosphorylation. (B) Comparison of molecular profiles between pluripotent and differentiated cells indicated independent regulation at the level of gene, protein, and phosphorylation. Notably 28% and 41% of pluripotency signature genes were regulated independent of mRNA expression (red) at the level of protein expression (blue) or phosphorylation (orange), respectively (see Table S1 for a complete list of pluripotent signature genes). (C) The relative differences in molecular signatures are the greatest for pluripotent stem cells (ESC and iPSC) as compared to DF (31.6%, 39.7%, and 50.9% genes as measured by mRNA expression, protein expression, and phosphorylation, respectively). In contrast the relative differences in mRNA expression, protein expression, and phosphorylation between pluripotent cells (ESC versus iPSC) or between different cells of fibroblast origin, were more subtle. (D) Clustering analysis was performed based on the quantitative phosphoproteomic data derived from BJ and MRC5 fibroblasts cultured in conventional serum-containing medium (IFS) and mTeSR medium (Ludwig and A Thomson, 2007). Hierarchical clustering was followed by an uncertainty assessment using the pvclust function implemented in R (http://www.r-

project.org/). Calculation of p-values based on either Approximate Unbiased (AU) or Bootstrap Probability (BP) methods suggest that signaling is more strongly correlated with genetic background (MRC5 versus BJ1) as compared to culture condition. The Y-axis represents the metric for correlation distance between samples. (E) Assessment of physical interactions between constituent genes of the pluripotent signature ("Input genes", green stars) and positive reference sets (PRS) of pluripotent factors (yellow triangles). Input gene sets were derived from the intersection of GO biological process and molecular class (mRNA – red, protein – blue, and phosphoprotein – orange) of the pluripotency signature (see Fig. 1B). For clarity only phosphoprotein signature data are shown; in each analysis three different PRS were considered: PRS #1: 154 pluripotency genes derived from literature survey (see Table S2); PRS #2: 527 proteins that disrupted *OCT4* expression in a recent genome-wide RNA interference screen in human ESC (see Table S2); PRS #3: human homologs for 164 murine proteins identified as interactors of Oct4 or Nanog (see Table S2). Edges are based on human protein-protein interactions in HPRD and BioGRID. The number of interactions between pluripotency signature genes and PRS factors was compared to a random sampling of genes (grey stars) from the background of all genes detected within a Biological Pathway (BP pathway).

Figure S2. Human iPSC derived from mesenchymal stem cells (HFMSC-iPSC) are pluripotent, related to Figure 1.

(A) Cell colony morphology, alkaline phosphatase activity, and immunohistochemistry staining of OCT4 (green), NANOG (red), TRA1-60 (magenta) and SSEA4 (cyan), with DAPI nuclear staining (blue). (B) Bisulfite genomic sequencing of the *OCT4* and *NANOG* promoters reveals demethylation in the HFMSC-iPSC line. Each horizontal row of circles represents an individual

sequencing reaction for a given amplicon. Open and filled circles represent unmethylated and methylated CpGs dinucleotides, respectively. (C) Hematoxylin and eosin staining of teratomas derived from immunodeficient mice injected with HFMSC-iPSC show tissues representing all three embryonic germ layers.

Figure S3. Pluripotent human ESC have vastly different alternative mRNA splicing profiles compared to somatic fibroblasts, related to Figure 1.

(A) Comparison of molecular profiles for alternative splicing substrates between pluripotent and differentiated cells indicated that more than 50% (1501 out of 2974) of alternatively spliced genes (purple) are not represented in the gene transcripts (red) associated with pluripotency. Among these genes, the majority (95%, or 1424 out of 1501 genes) are not differentially regulated at the level of protein expression (blue) or phosphorylation (orange) between pluripotent and differentiated cells (see Table S3 for a complete list of alternatively regulated genes). (B, left) Multiple Gene Ontology (GO) biological processes were enriched within the set of genes subject to alternative splicing in human ESC as compared to DF. (middle) Pluripotent signature genes derived from each GO biological process were tested for physical interactions with PRS factors as described in Figure S1E and Supplementary Experimental Procedures. This analysis suggested an enrichment of multiple pathways, including cell cycle, apoptosis, chromatin modification, cytoskeleton organization, transcription regulation, morphogenesis, and phosphorylation regulation. Null distributions were derived from random sampling of gene sets, equivalent in size to the membership of each GO category, from the background of all pluripotency signature genes detected by exon-junction microarray. Abbreviations: ***Extracell.***,

extracellular; ***Mod.***, modification; ***Neg.***, negative; ***Org.,*** organization; ***Pos.***, positive; ***Reg.***, regulation.

Figure S4. OCT4, SFRS2, MBD2, and microRNAs form a putative positive feedback loop in human pluripotent stem cells, related to Figure 2.

(A-B) Depletion of the splicing factors *SFRS2* disrupted self-renewal in H1 ESC. The self-renewal status of H1 ESC was monitored via (A) colony morphology and alkaline phosphatase activity, and (B) colony numbers (data represented as mean +/- SEM). Cells were grown on mouse embryonic fibroblasts (MEF). (C) Unlike *SFRS2* (Fig. 2C), depletion of *OCT4* in H1 ESC for 5 days did not induce down-regulation of splicing factors *RBM17*, *PRPF40A*, and *SFRS11* (data represented as mean +/- SEM). (D) Lentiviral shRNA-mediated depletion of *OCT4* and *SFRS2* each led to increased *MBD2a* expression, but only *SFRS2* depletion led to a modest decrease in *MBD2c* expression after 2 days in H1 ESC, assessed by qRT-PCR (data represented as mean +/- SEM). (E) Depletion of *OCT4* in H1 ESC for 5 days led to decreased expression levels of *miR-302* cluster members (*miR-302a*, *miR-302b*, *miR-302c*, and *miR-367*), and increased expression levels of *miR-301* cluster members (*miR-301b*, *miR-130a*, *miR-130b*), as assessed by qRT-PCR (data represented as mean +/- SEM).

Supplemental Tables S1-S4:

Table S1. Summary of microarray, proteomic, phosphoproteomic, and normalized phosphoproteomic analyses in H1 ESC, H9 ESC, BJ iPSC, dH1F iPSC, HFMSC iPSC, hMNC iPSC, BJ DF, and dH1F DF, related to Figure 1.

Table S2. Summary of positive reference sets, related to Figure 1.
Tab 1, positive reference set 1 as reported pluripotency-associated genes (Adewumi et al., 2007; Chia et al., 2010; Ho et al., 2009; Nandan and Yang, 2009; Seki et al., 2010; Zhang et al., 2006, 2008);
Tab 2, positive reference set 2 as genome-wide RNAi screening positive hits in human ESC (Chia et al., 2010);
Tab 3, positive reference set 3 as human homolog genes for 164 interactors of Oct4 (van den Berg et al., 2010; Cheong et al., 2011; Liang et al., 2008; Pardo et al., 2010) and Nanog (Liang et al., 2008; Wang et al., 2006).

Table S3. Summary of protein-protein interaction significance and GO biological process enrichment analyses, related to Figure 1.
Tab 1, analysis between pluripotency signature genes in human pluripotent stem cells (ESC and iPSC) and positive reference sets in each measurement class;
Tab 2-5, GO biological process enrichment analysis from pluripotency signature genes in human ESC and iPSC by each measurement class;
Tab 6-9, protein-protein interaction significance analysis between pluripotency signature genes assigned to each GO biological process, by measurement class, and positive reference sets;
Tab 10, ranking of pluripotency signature genes belonging to the RNA splicing pathway.

Table S4. Summary of exon-junction microarray analysis in H1 ESC, BG01 ESC, and BJ DF, related to Figure 1.

Supplemental Experimental Procedures

Assay for teratoma formation. For teratoma formation, HFMSC iPS cells were resuspended in a mixture of DMEM, Matrigel and collagen (ratio of 2:1:1) and injected intramuscularly into immune compromised *Rag2*(-/-) *gammaC*(-/-) mice. Xenografted masses formed within 4 to 6 weeks and paraffin sections were stained with haematoxylin and eosin for all histological determinations.

Bisulfite genomic sequencing. Bisulfite treatment of genomic DNA (gDNA) was carried out using a CpGenome DNA Modification Kit (Chemicon) according to the manufacturer's protocol. Briefly, converted gDNA was amplified by PCR using OCT4 primer sets 2 and NANOG primer set 3 reported previously (Freberg et al., 2007). PCR products were gel purified and cloned into bacteria using the TOPO TA cloning kit (Invitrogen). Bisulfite conversion efficiency of non-CpG cytosines ranged from 80% to 99% for all individual clones for each sample.

Chromatin Immunoprecipitation (ChIP) assays. ChIP tests in 293T cells were performed using anti-FLAG gel (Sigma). Relative occupancy values (fold enrichments) were calculated by determining the IP efficiency (ratios of the amount of immunoprecipitated DNA to that of the input sample) and normalized to the level observed at a control region, which was defined as 1.0. Coordinates for the ChIP control region is chr12:7,948,510- 7,948,699 (hg19 genome build). ChIP primers are as follows:

ChIP_control_F (GAGGTCTCGTATTTGCTGCATCGTA);
ChIP_control R (GCTAATTTCCTTCTCCACCCCAACCA);
*SFRS2*_promoter_1F (CTCTCAGGCAGTTGCCTTCCGCGTG);
*SFRS2*_promoter_1R (GGCCAATCAGAAGGTTTCATTTCCG);
*SFRS2*_1kb_F (GTTTATGTTCACTGCCTTTCAAACT);
*SFRS2*_1kb_R (TGACACAAGAAATCAAACCTAATCA);

*SFRS2*_7kb_F (CAGTACCTTTGGTTTCACAGAAGAT);
*SFRS2*_7kb_R (AAACAACATTCTTAGCTCACAGACC);
*OCT4*_promoter_F (CGCCAGTTGTGTCTCCCGGT);
*OCT4*_promoter_R (CCCCAGTCCCACCCACTAGCC);
*NANOG*_promoter_F (CCCCACCTAGTCTGGGTTACTCTGC);
*NANOG*_promoter_R (GCGGGCACACCCCCTACTGA).

Coimmunoprecipitation (CoIP) assays. For protein CoIP experiments, HEK293T cells were transfected with control or MBD2-isoform expression plasmids for 24hrs, followed by lysis in RIPA buffer containing 0.5% NP-40.

Exon-junction microarray. To capture the genome-wide alternative splicing events in the human ESCs and fibroblasts, we utilized Affymetrix (http://www.affymetrix.com/estore/index.jsp) human exon junction arrays (HJAY) to interrogate both the exons and the observed exon-exon junctions for approximately 30,000 genes in the human genome. Using the content from ExonWalk, Ensembl, and Refseq (available at http://genome.ucsc.edu), Affymetrix designed the HJAY including exon probes (8-10 perfect match probes per exon) and exon-exon junction probes (8 probes per junction, -4 to +4) for more than 30,000 human genes on the array. Specifically, based on the NCBI36 human genome, there are 5,235,274 probes on the HJAY chips in total, among them there are 35,123 transcript clusters, 335,663 transcripts, 260,488 junctions (x8 probes), 249,240 exon clusters and 315,137 probe selection regions (x10 probes). The junction probes design provides the information on how exons are joined together, and the exon skipping events are measured directly, rather than as a decrease in exon signal, which allows for reciprocal change analyses. mRNA from BJ fibroblast, H1 hESC and BG01 hESC were prepared using the same method as that for the Affymetrix HGU133 plus 2.0 GeneChip arrays. Triplicate arrays were generated for each sample. CEL files from the human exon

junction arrays were processed and analyzed using AtlAnalyze (http://code.google.com/p/altanalyze/) by comparing the expression of reciprocal junctions in hESCs and fibroblasts using a linear regression based method (Salomonis et al., 2009, 2010; Sugnet et al., 2006). The HJAY array data have been deposited in the GEO database under the accession number GSE55673 (Table S4)

Immunoblots. For whole cell lysates in the blots, cells were lysed in RIPA buffer containing 1% NP-40, 1% sodium deoxycholate, and 0.1% SDS. For each blot, 30ug of whole cell lysates or 20% of IP INPUTs/eluates were loaded. Primary antibodies used include: MBD2 (Bethyl Laboratories), GAPDH (Cell Signaling Technology), HDAC1 (Cell Signaling Technology), HDAC2 (Cell Signaling Technology), RbAp46 (Santa Cruz), MTA2 (Santa Cruz), MI2 (Santa Cruz), and SIN3A (Santa Cruz).

Immunohistochemistry staining. HFMSC-iPS cells were fixed in 4% paraformaldehyde for 30 min, permeabilized with 0.2% Triton X-100 for 30 min, and blocked in 3% BSA in PBS for 2 hours. Cells were incubated with primary antibody overnight at 4 ºC, washed, and incubated with Alexa Fluor (Invitrogen) secondary antibody for 2 hours. SSEA3, SSEA4, TRA-1-60 and TRA-1-81 antibodies were obtained from Millipore. OCT3/4 and NANOG antibodies were obtained from Abcam. Alkaline phosphatase staining was performed according to the manufacturer's recommendations (Millipore).

Luciferase reporter assays for SFRS2 promoter region. To study the effects of OCT4 knockdown on the *SFRS2* promoter, 750 ng of human SFRS2-Luc plasmid containing wild-type,

mutated, or truncated *SFRS2* promoters, and 5 ng of plasmid containing Renilla luciferase (pRL-SV40; Promega), were transfected into clone 9 human iPSCs growing on 96-wells pre-coated with matrigel. Human *OCT4* siRNA (Dharmacon, ON-TARGETplus SMART pool) and Nontargeting Control (ON-TARGET plus Control) were transfected twice with RNAiMAX (Invitrogen) at 50nM. Firefly and Renilla luciferase activities were measured 48hr after 2nd siRNA transfection with the Dual Luciferase System (Promega). For human SFRS2-Luc, a 500bp (-340 to 155 relative TSS) fragment of the human SFRS2 promoter was cloned into pGL3-Basic plasmid (BglII and MluI sites) upstream of the firefly luciferase gene (Promega). The cloning primers used are as follows: for SFRS2-Luc, 5'-GCATATATAGATCTTGGGGACACTGGGAAAGGCCTTGC-3' and 5'-TTGGCTAGATCTGGAGTTCCGGGAATATCCTCC-3'. Site-directed mutagenesis mutate the putative Oct4 binding site from WT sequence of ATGCCAAT to CCCTTTTT (MT2) and CCCTGGGC (MT3). The Oct4 binding site was also removed from the 500bp promoter using the following primers, 5'-ATGTCGTGACGCGTGAACTTGGAGTATTGGCATGAA-3' and 5' ATGTCGTGACGCGTGCGTTTCGCGGGCTTTCCAA-3' to generate the 400bp and 300bp truncated SFRS2 promoter.

Luciferase reporter assays for 3'-UTR regions. miR-301 cluster members *miR-301b* and *miR-130b* recognizes motifs in *SFRS2* 3'-UTR region, while there are two predicted binding motifs (275-281bp, and 3220-3226bp) by *miR-302* family members in the 3'-UTR of *MBD2a* which are absent in *MBD2c*. MBD2a, MBD2c and SFSR2 3'UTR sequences were PCR amplified and cloned into psiCheckTM-2 vector (Promega). The simoltanous presence in the same plasmid of both renilla (RLuc) and firefly (FLuc) luciferase gene allowed normalization for transfection

efficiency. Mutant derivatives (MBD2a-mut and SFSR2-mut) were obtained by deletion of miR-302 and miR-301 binding sites, respectively. Luciferase constructs were co-transfected with miRNA-overexpressing constructs into Hela cells and 293T cells using FUGENE 6 (Roche, Basel Switzerland) transfection reagent following the manufacturers protocol. 48 hours after the transfection of the reporter constructs, cells were lysed and luciferase assays were performed using a Dual-Luciferase Reporter Assay System (Promega, Madison, WI) following the manufacturers specifications. Transfection of each construct was performed in triplicate in each assay and a total of three assays were performed. Empty vectors were transfected for normalization purposes. Ratios of Renilla luciferase values to Firefly luciferase ones were calculated for each experiment and triplicates were averaged. The average values of the tested constructs were normalized to the activity of the empty construct.

Microarray analysis. Gene expression profiling was carried out using Human Genome HGU133 plus 2.0 Affymetrix GeneChip arrays. In brief, cDNA (1.5 μg) was fragmented and hybridized to Affymetrix HGU133 plus 2.0 GeneChip arrays according to the manufacturer's instructions. Arrays were processed by the Coriell Institute Genotyping and Microarray Center (Camden, NJ). DNA chips were washed, stained and scanned using an Affymetrix Fluidics device and a GCS3000 scanner, and the images obtained were analyzed using GCOS software. The experiment was performed in duplicate for the dH1F fibroblast, BJ fibroblast, hESC and all the iPSC. Data normalization was calculated with the robust multichip average (RMA) algorithm (Irizarry et al., 2003) implemented in BioConductor (http://www.bioconductor.org/). The array data have been deposited in the GEO database under the accession number GSE55673, and the normalized data are provided (Table S1).

Phosphoproteomics analysis. Approximately 50μg (corresponding to ~5e5 cells) of total protein was isolated from each cell line, digested with trypsin. The resulting peptide solutions were labeled by 8-plex iTRAQ reagents (AB Sciex), combined, and desalted. The iTRAQ labeled peptide mixture was processed according to a previously-reported phosphopeptide enrichment protocol using Fe^{3+}-NTA magnetic resins (Ficarro et al., 2009). The phosphopeptide and supernatant (i.e., non-phosphorylated) fractions were analyzed by a 3D RP-SAX-RP LC-MS/MS platform (Ficarro et al., 2011) with a QSTART Elite mass spectrometer (AB Sciex). QSTAR data files were converted and searched with Mascot version 2.2.1 against a forward-reversed human (38190 entries) NCBI refseq database (downloaded Nov. 2009) with an appended cRAP (common repository of adventitious proteins) database of 752 entries. multiplierz software (Geryon edition 0.7.188) (Parikh et al., 2009) was used to organize and process the search results. From 217136 MS/MS spectra acquired through 214 LC-MS/MS runs (two biological replicates for phosphopeptide samples and three bioreplicates for quantitative protein expression analysis) 3435 proteins (Table S1), including 2093 that were phosphorylated (60.9%) (Table S1), were identified with a false discovery rate less than 1% at the peptide level. For quantitative phosphoproteomics analysis, the intensities of 1971 unique phosphorylated peptides on 572 proteins that were identified in both the phosphorylation-specific and the non-phosphorylation proteomics experiments were normalized and used for further analysis (Table S1).

Quantitative Reverse-Transcription-Polymerase Chain Reaction Assays. For quantitative RT-PCR assays, cDNA from about ~5e5 ESC was synthesized using SuperScript II reverse transcription kit (Invitrogen) with either oligo-dT (for mRNA expression analysis) or random

primers (for RNA immunoprecipitation analysis). Reactions were performed in a 96-well format using 500ng of cDNA and FAST SYBR GREEN Master mix (Invitrogen) on a StepOnePlus quantitative PCR system (AB). Primer sequences used for qRT-PCR are as follows:

POU5F1,
forward (AGCGAACCAGTATCGAGAAC), reverse (TTACAGAACCACACTCGGAC)
NANOG,
forward (TGAACCTCAGCTACAAACAG), reverse (TGGTGGTAGGAAGAGTAAAG)
SOX2,
forward (AGCTACAGCATGATGCAGGA), reverse (GGTCATGGAGTTGTACTGCA)
TBP,
forward (TTCGTGCCCGAAACGCCGAA), reverse (TTCTTCACTCTTGGCTCCTGTGCA)
SFRS2,
forward (AGGAGTGCTGCTCGGAGGTTACT), reverse (ACCTGCTCCAATACACGCCCAC)
RBM17,
forward (AGCAAGCGTGGCGGCAAGAT), reverse (CCCGCACCAACCATGTTCCTTAGT)
PRPF40A,
forward (AATCTGACTCTCCAGAATCC), reverse (GTATCCCAATTACCAGAATCCT)
SFRS11,
forward (GTGTCTGCTTTGTTAAGTTCC), reverse (ATCAGGAATAACTCCTTCTGC)
MBD2a,
forward(AGCAAGCCTCAGTTGGCAAGGT),reverse(TGTTCATTCATTCGTTGTGGGTCTG)
MBD2c,
forward (AGCAAGCCTCAGTTGGCAAGGT), reverse (TGAAAGCGCATGCCATGGTGCA)
*MBD2a*_overexp,
forward (GCGTGGGCCCTGGAAGCAAC), reverse (ATCTCTTCGGTGTCGGCGGC)
*MBD2c*_overexp,
forward (GCCCCTTGGCACGCTCTGTC), reverse (AGGGTGGGAAGCTCTGGTTCTGT)

Ranking of splicing factors within the pluripotent molecular signature. Splicing factors within the pluripotent molecular signature were individually ranked based on their relative levels of gene expression, protein expression, and protein phosphorylation in pluripotent cells compared to differentiated fibroblasts, along with their relative impact on *OCT4* expression when depleted in the context of a systematic functional genomics study performed in human ESC (Chia et al., 2010). Individual scores for each splicing factor were aggregated to yield a rank-ordered list.

Reprogramming assays. 25,000 BJ fibroblasts, infected with lentiviruses overexpressing MBD2 isoforms and tdTomato control and selected by hygromycin resistance, were plated in 12-well plates and infected over-night with lentiviral reprogramming factors (Addgene). Cells were re-plated 1:6 onto six-well plates after 6 days, and were fed with hESC medium daily until day 22, when plates were fixed with 4% paraformaldehyde for 30 minutes. Cells were then blocked with 5% FBS for 30 minutes, stained with biotin-anti-Tra-1-60 (1:250) (eBioscience) overnight at 4oC and streptavidin horseradish peroxidise (1:500) (Biolegend). Staining was developed with Vector labs DAB kit, and iPSC colonies quantified with ImageJ software.

RNA Immunoprecipitation (RIP). RIP tests were performed on H1 ES cells using anti-FLAG (Sigma). In brief, cells were cross-linked in 0.1% formaldehyde for 10min prior to harvest and lysis. Cell lysates were sonicated, immunoprecipitaed, and the eluates were reverse-crosslinked by incubating with 200mM NaCl at 65C for 1hr. RNA in each eluted sample was then reverse transcribed and measured by quantitative PCR. Relative occupancy values (also known as fold enrichments) were calculated by determining the IP efficiency (ratios of the amount of immunoprecipitated RNA to that of the input sample) and normalized to the level observed by immunoprecipitation using non-specific IgG, which was defined as 1.0. The RIP primers are listed:

*MBD2*_RIP_I_F (AAGCAAGCCTCAGTTGGCAAGGT);
*MBD2*_RIP_I_R (GAGAGGATCGTTTCGCAGTCTCTGT);
*MBD2*_RIP_II_F (AAACAGAGACTGCGAAACGATCCTC);
*MBD2*_RIP_II_F (GGGTATGGGGACATGCACGGG);
*MBD2*_RIP_III_F (TGCTGGGTACCTATAAAAGGGGCT);
*MBD2*_RIP_III_F (GCATGCCATGGTGCAGGACGA).

RNA-interference. For shRNA infections, *pLKO.1-puro-based* shRNA lentiviral constructs against target genes (Sigma) were added to H1 ESC with 5μg /ml protamine sulfate. Puromycin (Sigma) selection was introduced 1 day after infection at 0.8 μg/mL, and maintained for 5 days prior to harvest. *pLKO.1-puro-control* consisted of scrambled shRNA sequences was used as control. shRNA sequences used in this study are as follows:

POU5F1,
CCGGCCCTCACTTCACTGCACTGTACTCGAGTACAGTGCAGTGAAGTGAGGGTTTTT
SFRS2.1#,
CCGGGGCAAGCAGTGTAAACGGAGGCTCGAGCCTCCGTTTACACTGCTTGCCTTTTT
SFRS2.2#,
CCGGTATCGGCAAGCAGTGTAAACGCTCGAGCGTTTACACTGCTTGCCGATATTTTT

Analysis of protein-protein interactions between constituent genes of the pluripotency signature and positive reference set (PRS) genes. The total human protein-protein interaction (PPI) network, based on direct physical interactions in the BioGRID (Stark et al., 2006) and HPRD (Peri et al., 2004) databases, consisted of 128,069 unique edges (67,504 from BioGRID and 98,349 from HPRD) among 12,028 unique proteins (9,935 from BioGRID and 9,780 from HPRD). We estimated the significance of the number of interactions between two different protein sets (e.g., an input set corresponding to genes in the pluripotent molecular signature and a second protein set from one of the positive reference sets – PRS, see below) using a permutation test (Welch, 1990). We created three Positive Reference Sets (PRS) corresponding to literature-curated pluripotency-associated genes (PRS #1, Table S2), genes that scored as pluripotent factors in a recent genome-wide RNAi depletion study (PRS #2, Table S2), and human homologs for 164 murine proteins identified as interactors of Oct4 or Nanog (PRS #3, Table S2). For each analysis, the input set corresponded to either (i) genes of the pluripotency signature (e.g., based on detection of mRNA, protein, or phosphoprotein) or (ii) *n* genes sampled at random, without

replacement, from the background of all genes detected, where *n* was equal to the number of genes in each molecular class of the pluripotency signature. The latter was used to create a null distribution (10,000 iterations). Based on this analysis a p-value < 0.05 suggests that the number of interactions between genes of each molecular class in the pluripotency signature and the PRS is significantly higher as compared to interactions between the PRS and genes sampled at random (Table S3).

Analysis of the number of protein-protein interactions between pluripotency signature genes assigned to each GO biological process and positive reference set (PRS) genes. Based on the same protein-protein interaction (PPI) networks as described above, we estimated the significance of the number of interactions between the protein set corresponding to pluripotency signature genes assigned to each GO BP and those in one of three positive reference sets (PRS, see above) using the permutation test. For each analysis, the input set corresponded to either (i) the intersection of pluripotency signature genes from each measurement class and each GO BP or (ii) *n* genes sampled at random, without replacement, from the background of all pluripotency signature genes in the measurement class, where *n* was equal to the number of pluripotency signature genes assigned to each GO BP within each measurement class. The latter was used to create a null distribution (10,000 iterations). Based on this analysis a p-value < 0.05 suggests that the number of interactions between pluripotency signature genes of each GO BP and the PRS is significantly higher as compared to interactions between the PRS and genes sampled at random (Table S3).

Supplemental References

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., et al. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol *25*, 803–816.

Van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

Cheong, C.Y., Lon Ng, P.M., Ponnampalam, R., Tsai, H.-H., Bourque, G., and Lufkin, T. (2011). In silico tandem affinity purification refines an Oct4 interaction list. Stem Cell Res Ther *2*, 26.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Ficarro, S.B., Adelmant, G., Tomar, M.N., Zhang, Y., Cheng, V.J., and Marto, J.A. (2009). Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. Anal. Chem *81*, 4566–4575.

Ficarro, S.B., Zhang, Y., Carrasco-Alfonso, M.J., Garg, B., Adelmant, G., Webber, J.T., Luckey, C.J., and Marto, J.A. (2011). Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. Mol. Cell Proteomics *10*, O111.011064.

Freberg, C.T., Dahl, J.A., Timoskainen, S., and Collas, P. (2007). Epigenetic reprogramming of OCT4 and NANOG regulatory regions by embryonal carcinoma cell extract. Mol. Biol. Cell *18*, 1543–1553.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. U.S.A *106*, 5181–5186.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249–264.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., et al. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol. *10*, 731–739.

Loh, Y.-H., Hartung, O., Li, H., Guo, C., Sahalie, J.M., Manos, P.D., Urbach, A., Heffner, G.C., Grskovic, M., Vigneault, F., et al. (2010). Reprogramming of T cells from human peripheral blood. Cell Stem Cell *7*, 15–19.

Ludwig, T., and A Thomson, J. (2007). Defined, feeder-independent medium for human embryonic stem cell culture. Curr Protoc Stem Cell Biol *Chapter 1*, Unit 1C.2.

Nandan, M.O., and Yang, V.W. (2009). The role of Krüppel-like factors in the reprogramming of somatic cells to induced pluripotent stem cells. Histol. Histopathol *24*, 1343–1355.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Parikh, J.R., Askenazi, M., Ficarro, S.B., Cashorali, T., Webber, J.T., Blank, N.C., Zhang, Y., and Marto, J.A. (2009). multiplierz: an extensible API based desktop environment for proteomics data analysis. BMC Bioinformatics *10*, 364.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T.K.B., Chandrika, K.N., Deshpande, N., Suresh, S., et al. (2004). Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. *32*, D497–501.

Salomonis, N., Nelson, B., Vranizan, K., Pico, A.R., Hanspers, K., Kuchinsky, A., Ta, L., Mercola, M., and Conklin, B.R. (2009). Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. PLoS Comput. Biol. *5*, e1000553.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proc Natl Acad Sci U S A *107*, 10514–10519.

Seki, Y., Kurisaki, A., Watanabe-Susaki, K., Nakajima, Y., Nakanishi, M., Arai, Y., Shiota, K., Sugino, H., and Asashima, M. (2010). TIF1beta regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. Proc. Natl. Acad. Sci. U.S.A *107*, 10926–10931.

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535–539.

Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D., et al. (2006). Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput. Biol. *2*, e4.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364–368.

Welch, W.J. (1990). Construction of Permutation Tests. Journal of the American Statistical Association *85*, 693–698.

Zhang, J., Tam, W.-L., Tong, G.Q., Wu, Q., Chan, H.-Y., Soh, B.-S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat. Cell Biol *8*, 1114–1123.

Zhang, X., Zhang, J., Wang, T., Esteban, M.A., and Pei, D. (2008). Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. J. Biol. Chem *283*, 35825–35833.