ELSEVIER

Experimental Hematology

Experimental Hematology 2011;39:866–879

Retinoblastoma-binding proteins 4 and 9 are important for human pluripotent stem cell maintenance

Michael D. O'Connor[a,b,c], Elizabeth Wederell[a,*], Gordon Robertson[d,*], Allen Delaney[d,*], Olena Morozova[d,*], Steven S.S. Poon[e,*], Damian Yap[e,*], John Fee[e], Yongjun Zhao[d], Helen McDonald[d], Thomas Zeng[d], Martin Hirst[d], Marco A. Marra[d], Samuel A.J.R. Aparicio[e], and Connie J. Eaves[a]

[a]*Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada;* [b]*School of Medicine, University of Western Sydney, NSW, Australia;* [c]*Molecular Medicine Research Group, University of Western Sydney, NSW, Australia;* [d]*Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada;* [e]*Department of Molecular Oncology, British Columbia Cancer Agency, Vancouver, BC, Canada*

(Received 9 March 2011; revised 13 May 2011; accepted 20 May 2011)

Objective. **The molecular mechanisms that maintain human pluripotent stem (PS) cells are not completely understood. Here we sought to identify new candidate PS cell regulators to facilitate future improvements in their generation, expansion, and differentiation.**

Materials and Methods. **We used bioinformatic analyses of multiple serial-analysis-of-gene-expression libraries (generated from human PS cells and their differentiated derivatives), together with small interfering RNA (siRNA) screening to identify candidate pluripotency regulators. Validation of candidate regulators involved promoter analyses, Affymetrix profiling, real-time PCR, and immunoprecipitation.**

Results. **Promoter analysis of genes differentially expressed across multiple serial-analysis-of-gene-expression libraries identified E2F motifs in the promoters of many PS cell-specific genes (e.g., *POU5F1*, *NANOG*, *SOX2*, *FOXD3*). siRNA analyses identified two retinoblastoma binding proteins (RBBP4, RBBP9) as required for maintenance of multiple human PS cell types. Both RBBPs were bound to RB in human PS cells, and E2F motifs were present in the promoters of genes whose expression was altered by decreasing *RBBP4* and *RBBP9* expression. Affymetrix and real-time PCR studies of siRNA-treated human PS cells showed that reduced *RBBP4* or *RBBP9* expression concomitantly decreased expression of *POU5F1*, *NANOG*, *SOX2*, and/or *FOXD3* plus certain cell cycle genes (e.g., *CCNA2*, *CCNB1*), while increasing expression of genes involved in organogenesis (particularly neurogenesis).**

Conclusions. **These results reveal new candidate positive regulators of human PS cells, providing evidence of their ability to regulate expression of pluripotency, cell cycle, and differentiation genes in human PS cells. These data provide valuable new leads for further elucidating mechanisms of human pluripotency.** © **2011 ISEH - Society for Hematology and Stem Cells. Published by Elsevier Inc.**

Given appropriate physiologic stimulation, human pluripotent stem (PS) cells can differentiate into every cell type found in the body. Moreover, if kept under appropriate laboratory conditions, human PS cells can retain this latent developmental potential while proliferating rapidly. Great strides have been made in recent years in our understanding of the molecular mechanisms responsible for human PS cell self-renewal and differentiation, but additional knowledge is essential to increasing the efficiency with which pluripotent human cells can be generated and expanded. These issues, in turn, are prerequisite to using human PS cells for realizing anticipated advances in human developmental biology, biotechnology, and medicine.

*Drs. Wederell, Robertson, Delaney, Morozova, Poon and Yap contributed equally to this work.

Offprint requests to: Michael D. O'Connor, Ph.D., University of Western Sydney, Locked Bag 1797, Penrith, NSW, Australia, 2751.; E-mail: m.oconnor@uws.edu.au

Supplementary data associated with this article can be found in the online version at 10.1016/j.exphem.2011.05.008.

0301-472X/$ - see front matter.
doi: 10.1016/j.exphem.2011.05.008

Much of our current knowledge about the molecular mechanisms that maintain the status of human PS cells is based on studies of POU5F1, NANOG, SOX2, and FOXD3. These four transcription factors are believed to act as a self-reinforcing autoregulatory network that simultaneously suppresses the expression of genes required for differentiation [1–6]. The central role that these four factors play in maintaining PS cells has also led to their broad use as markers of the pluripotent state, thereby enabling refinement of conditions for human PS cell maintenance. *POU5F1* expression, in particular, has been widely used for this purpose. Along with other genes such as *SOX2*, *NANOG*, *KLF4,* and *c-MYC*, *POU5F1* has also proven to be a key element in protocols for reprogramming somatic cells to a pluripotent state [7–9]. However, most of these genes have also been found to be replaceable, reflecting the complexity and redundancy that exist in the cell signaling pathways involved in maintaining PS cells in an undifferentiated state. For example, *Nr5a2* has been shown to be capable of replacing *Pou5f1* for the reprogramming of mouse cells [10]. Similarly, omitting *cMYC* from the reprogramming cocktail reduces the generation of induced PS (iPS) cells with tumorigenic properties, but at the expense of reprogramming efficiency [11]. Taken together, these studies underscore the need for more detailed understanding of the genes involved in human PS cell maintenance to enable further exploitation of the multiple opportunities that human PS cells offer.

A variety of human PS cells exist, including human embryonal carcinoma (EC) cells [2,12], embryonic stem (ES) cells [13], and iPS cells [7–9]. Human ES and iPS cells are the most likely PS cell types to have clinical applications, however, human EC cells have for almost 3 decades provided a useful model to study pluripotency, as they are more easily propagated and genetically modified, yet possess similar properties to ES and iPS cells [12]. In the present study, we used examples of each of these three human PS cell types to look for novel regulators of human pluripotency. As a first step, we looked for shared transcription factor binding sequences in the promoter regions of genes that we found were expressed in undifferentiated human ES and EC cells [14]. These studies suggested the family of E2F transcription factors to be likely candidate regulators of many human PS cell-specific genes. In parallel, we undertook experiments to screen small interfering RNA (siRNAs) for elements that would inhibit the rapid proliferation of human EC cells as a surrogate indicator of a loss of their pluripotent status. From this latter study, we identified 23 genes (6 known, 17 novel), including the gene encoding the retinoblastoma (RB)-binding protein, RBBP9, as candidate pluripotency regulators. Subsequent experiments indicated that both RBBP9 and another RB-binding protein (RBBP4) appear to play similar, but not identical, roles in maintaining human EC, ES, and iPS cell pluripotency.

Materials and methods

Cells

H9 human ES cells were purchased from WICELL (Madison, WI, USA) and NTera2D1 human EC cells were obtained from the ATCC (Manassas, VA, USA). The CA1 human ES cells were provided by A. Nagy [15] and MSC-iPSC1 human iPS cells were provided by G. Daley [8]. Approval for use of these cells as described was obtained from the Canadian Stem Cell Oversight Committee and the Research Ethics Board of the University of British Columbia.

Cell culture, colony assay, and flow cytometry

The H9 human ES cells used to generate the novel Tag-Seq serial analysis of gene expression (SAGE) libraries hs0212, hs0046, and hs0048 were maintained using mouse embryonic fibroblasts using media containing Knock-Out Serum Replacer (Invitrogen, Carlsbad, CA, USA) and 10 ng/mL FGF 2 (StemCell Technologies, Vancouver, BC, Canada) as described previously [16]. All other human ES and iPS cell cultures were maintained in mTeSR1 (StemCell Technologies) [16]. Human EC cells were passaged every 3 or 4 days using TrypLE (Invitrogen), and plated on tissue culture plastic in DS medium consisting of Dulbecco's modified Eagle's medium (StemCell Technologies) containing 10% fetal bovine serum (StemCell Technologies). For bulk differentiation cultures using all-trans retinoic acid (RA; Sigma, St Louis, MO, USA), the medium was replaced with DS medium supplemented with 10^{-5} M RA, and cells were cultured for at least 4 days with medium changes as necessary. Alkaline phosphatase-based colony-forming cell (CFC) assays were performed as described previously [16]. Consistent with previously published results, no alkaline phosphatase-negative colonies were seen in any of the assays.

Flow cytometric assessment of $SSEA3^+$ cells was performed using an anti-SSEA3 antibody (Developmental Studies Hybridoma Bank, Iowa City, IA, USA) and a fluorescein isothiocyanate–conjugated anti-mouse IgM secondary antibody (Becton Dickinson, San Jose, CA, USA). Data was collected using a FACSCalibur (Becton Dickinson) and analyzed using FlowJo software (Tree Star, Ashland, OR, USA) as described previously [17]. Gates for positive cells were set to exclude $>99\%$ of events detected when the primary antibody was omitted from the staining protocol.

Gene expression profiling, PASTAA analysis, and GO analysis

LongSAGE [18] and Tag-Seq SAGE [19] gene expression libraries were prepared as described previously. Tag-Seq libraries were processed to filter out the majority of error tags using an analysis of the frequency of one-off tags and mapping to genome and transcriptome resources [19]. The libraries used are available via the Gene Expression Omnibus database as part of record GSE14 of the Cancer Genome Anatomy Project resource [20] as follows: she10, HES3 human ES cells; she11, HES4 human ES cells; she13, H7 human ES cells; she14, H14 human ES cells; she15, H13 human ES cells; she16, H1 human ES cells; she17, H1 human ES cells; she19, BG01 human ES cells; shes2, H9 human ES cells; shes9, HSF6 human ES cells; shs11, H1 human ES cell-derived erythromegakaryocytic progenitors; shs12, H1 human ES cell-derived enriched primitive hematopoietic

multipotent progenitors; shs13, H1 human ES cell-derived enriched primitive hematopoietic myeloid progenitors; cg643, normal adult bulk pancreas; cg647, mammary gland antibody purified; cg648, normal substantia nigra; cg655, normal liver vascular endothelium.

LongSAGE tags were mapped to genes using the SAGE Genie tool. The resulting dataset was subjected to seriation analysis as described [21]. To identify transcription factor binding sites in groups of genes, we used the PASTAA Web server [14], which ranks genes by estimating the overall affinity of a position weight matrix (PWM) for sequence regions that are defined relative to the transcriptional start site of each gene in a list. Gene ontology (GO) analysis was performed using the program DAVID Bioinformatics Suite [22].

For Affymetrix profiling, 12 samples (3 × *RBBP4* and 3 × *B2M* harvested 72 hours after siRNA treatment; plus 3 × *RBBP9* and 3 × *B2M* harvested 48 hours after siRNA treatment) were analyzed using HuGene 1.0 st arrays (Affymetrix, Santa Clara, CA, USA). Labeling and hybridization to the arrays was performed by the Ramaciotti Centre for Gene Function Analysis (University of New South Wales, Sydney, Australia) as per manufacturer's instructions. Analysis of the resulting data was performed using the GenePattern software suite [23], with the top 10% of up- and down-regulated genes shown.

siRNA transfection

Single cell suspensions were plated into 96-well plates for siRNA screens (Greiner, Frickenhausen, Germany) and 6-well plates (Becton Dickinson) for Western blotting analyses, using a medium appropriate for the cell type being evaluated (i.e., mTeSR1, DS medium, or DS + 10^{-5} M RA). Cells were allowed to attach overnight before being transfected for 18 to 24 hours using DharmaFect3 (DF3; Dharmacon, Lafayette, CO, USA) and Dharmacon siGENOME siRNA at the concentration indicated. For the 319 gene-siRNA screens, triplicate test wells distributed over three separate 96-well plates were used for each treatment, and each treatment consisted of equimolar pools of four siRNAs per gene. For follow-up studies, we used either equimolar pools of four siRNA for each gene, or a combination of the two most effective siRNAs for each gene. After transfection the medium was replaced with a cell-appropriate medium and the cells cultured for an additional 2 to 3 days before final assessment. To approximate the cell output, wells were stained with 1 μg/mL Hoechst 33342 (Sigma) and the fluorescence intensity of each well was measured using a plate reader. Positive-effect control conditions (i.e., *POU5F1* and *NANOG* siRNAs), as well as negative-effect control conditions (i.e., *B2M* siRNAs or DF3 reagent only), were included in each plate along with the test siRNAs to monitor for consistency in relative response and to correct for plate-to-plate variations. Two methods were used to infer cell densities from the fluorescence intensity values: traditional data normalization [24] and linear modeling [25], as described here.

Normalization of siRNA screen data

For traditional normalization of the data [24], the background fluorescence level obtained from unstained wells or stained wells without added cells was first subtracted from the fluorescence values of all other wells for each plate. The base 2 logarithm of these background corrected values was then calculated to compensate for variations expected in exponentially increasing numbers and, thus, obtain a more representative measure for each siRNA over the different ranges of response evaluated. To compensate for plate-to-plate variations, the median log value for the DF3 reagent-only control wells was subtracted from the log values of all the wells. This resulted in the normalized value for the DF3-only control wells being close to zero, and enabled the normalized value for each siRNA-treated well to indicate the relative effect on cell number. Comparisons of significance in response to siRNA treatments against the *B2M* siRNA-negative control wells were then made using a single-tailed *t*-test. A further adjustment to the set of *p*-values for the 26 siRNA targets was made to control for the false discovery rate that can arise by chance from multidata comparisons [24].

For assessing the data using a linear model method, we first log-transformed the measured values, as this gave the most uniform and least residual error distribution for the model estimate. The linear model includes a predictor random variable to compensate for experiment-to-experiment effects, and another predictor random variable to compensate for plate-to-plate differences in the plate-reader's (optimal) gain settings used to capture the fluorescence measurements. For each siRNA treatment, we obtained a model estimate of its relative response versus the control *B2M* siRNA wells in the same plate [25]. We then performed an analysis of variance of the linear modeled data with a model where the targeted siRNA does not have a different response from the *B2M* test siRNA (the null hypothesis). Finally, the set of *p* values for the 26 siRNA targets from the analysis of variance tests was adjusted to control for the false discovery rate that can arise by chance from multidata comparisons [24]. The rank ordering of the siRNAs with respect to the relative response for both the linear and traditional normalization methods are very similar: i.e., siRNAs that exhibited a large differential response in the linear model also exhibited a large difference in the traditional normalization method. The results from the linear model are reported here.

RNA purification and reverse transcription

RNA was extracted and purified using Absolutely RNA kits (Stratagene, Cedar Creek, TX, USA). Reverse transcription using Superscript II (Invitrogen) and quantitative real-time PCR using Sybr-Green (Applied Biosystems, Warrington, WA, USA) was performed as described [17] using the following primer sets (forward primer, reverse primer):

GAPDH (CCCATCACCATCTTCCAGGAG, CTTCTCCATGGTGGTGAAGACG); *POU5F1* (GTGGAGGAAGCTGACAACAA, CTCCAGGTTGCCTCTCACTC); *NANOG* (AACTGGCCGAAGAATAGCAA, CATCCCTGGTGGTAGGAAGA); *SOX2* (CATGGCAATCAAAATGTCCA, TTTCACGTTTGCAACTGTCC); *FOXD3* (CAACCGCTTCCCCTACTACA, TTGACGAAGCAGTCGTTGAG); *B2M* (GAATTCACCCCCACTGAAAA, CGAGACATGTAAGCAGCATCA); *RBBP4* (GTTGATGCTCACACTGCTGAA, GATCCCACAAGGCAACAGTC); and *RBBP9* (ACATCAGACTTGGGGGATGA, GGGTCGTCAGTAGAGCCAAA).

Protein sample preparation, coimmunoprecipitation, and Western blotting

Cells were collected using TrypLE for 10 minutes, centrifuged, and the pellets washed with phosphate-buffered saline (StemCell Technologies) before being resuspended in protein lysis buffer

consisting of: 50 mM HEPES, 100 mM NaF, 10 mM $Na_4P_2O_7$, 2 mM Na_3VO_4, 2 mM EDTA, 2 mM $NaMoO_4.2H_2O$, 10% glycerol, 2 mM PMSF (Sigma), 0.1% NP-40 alternative (Calbiochem, Gibbstown, NJ, USA), and protease inhibitors (Sigma). For co-immunoprecipitation, ~300 μg of protein lysate was incubated on a rocker overnight at 4°C with either anti-human RB antibody (Becton Dickinson) or Ig control antibody (Dako, Glostrup, Denmark). Protein-G agarose beads (Thermo Fisher Scientific, Rockford, IL, USA) were washed in protein lysis buffer, added to each sample, and incubated on a rocker for 2 hours at 4°C. Samples were then centrifuged and the beads washed using protein lysis buffer. Proteins were eluted by incubating for 10 minutes in NuPAGE sample buffer/reducing agent (Invitrogen) at 95°C before Western analysis. For Western blotting, samples were separated on 4% to 12% NuPAGE Bis-Tris gels (Invitrogen), transferred to PVDF membrane (Millipore, Billerica, MA, USA), blocked in phosphate-buffered saline containing 0.1% Tween-20 (Sigma) and 5% bovine serum albumin (Invitrogen), then incubated overnight with one of the following anti-human antibodies: anti-RB (Becton Dickinson), anti-RBBP4 (Bethyl Laboratories, Montgomery, TX, USA), anti-RBBP9 (ProteinTech, Chicago, IL, USA), or anti-GAPDH (Sigma). Blots were scanned to file using a Canon Pixma MP160 digital scanner (Canon, Ontario, Canada) and are presented without digital manipulation. Densitometric analysis of Western blots was performed using an LAS-4000 (FujiFilm Corporation, Tokyo, Japan) and accompanying Image Reader LAS-4000 software, and the levels of RBBP4 and RBBP9 protein compared between siRNA treatments relative to GAPDH protein levels.

Statistical analyses

Tests for statistical significance were performed using the single-tailed two-sample *t*-test, single-tailed paired two-sample *t*-test, or analysis of variance, as appropriate. Statistical significance was assigned to p values <0.05.

Results

Comparative analyses of gene expression libraries prepared from human PS cells

Three new Tag-Seq SAGE libraries were generated as described previously [19] from undifferentiated human H9 ES cells (p41; Fig. 1A, C; hs0046); undifferentiated NTera2D1 human EC cells (Fig. 1E; hs0212); and from human H9 ES cells (p41) harvested 96 hours after addition of RA to induce their differentiation (Fig. 1B, D; hs0048). Confirmation of the respective undifferentiated and

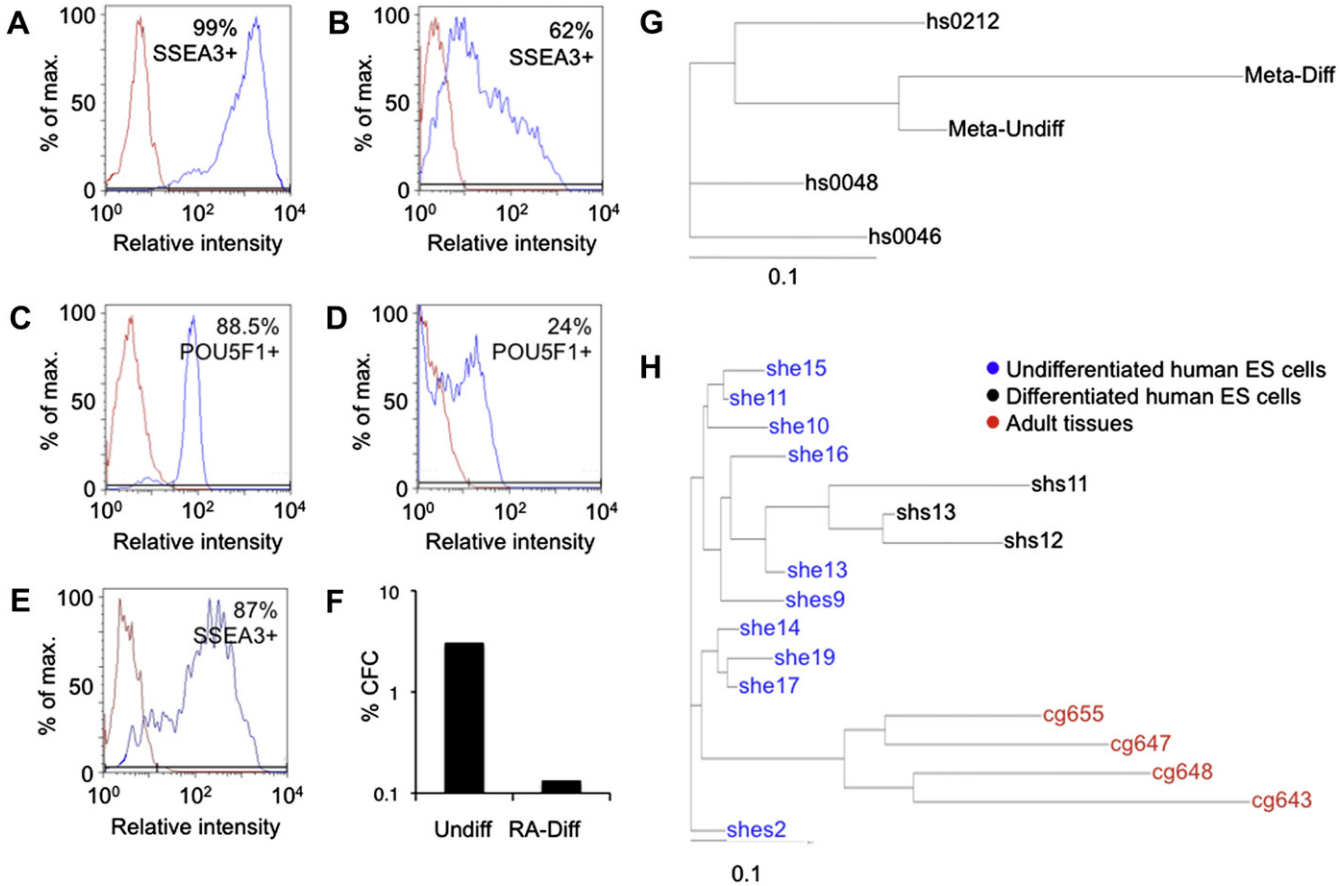


Figure 1. Biologic characterization of human PS cell SAGE libraries. **(A–E)** Flow cytometry data for undifferentiated **(A, C, E)** and RA-treated **(B, D)** H9 human ES cells **(A–D)** and NTera2D1 EC cells **(E)**. **(F)** Alkaline phosphatase CFC frequencies for the same human ES cells shown in **(A–D)**. **(G, H)** Tag type and frequency comparisons for the Tag-Seq SAGE libraries generated here (hs0212, hs0046, hs0048) versus previously established LongSAGE meta-libraries **(G)**, as well as LongSAGE libraries used for seriation analysis **(H)**.

differentiating states of the H9 cell populations was established by flow cytometric analyses and assays for alkaline phosphatase-positive CFCs performed as described previously [16]. These studies showed that most of the undifferentiated H9 and NTera2D1 cells expressed high levels of SSEA3 and/or POU5F1 (Fig. 1A, C, E), whereas most of the RA-treated H9 cells expressed undetectable or much lower levels of these antigens (Fig. 1B, D). The frequency of CFCs in the H9 human ES cells was also greatly diminished (22-fold) in the RA-treated H9 cells (Fig. 1F).

These three Tag-Seq SAGE libraries were sequenced to a depth that yielded between 4 and 13 $\times$ 10^6 total tags that mapped to the human genome, and are included in the Gene Expression Omnibus database as part of record GSE14 of the Cancer Genome Anatomy Project resource [20]. The Table in Supplementary Table E1 [online only, available at www.exphem.org] shows a summary of the tag frequencies for *POU5F1*, *NANOG*, *SOX2*, and *FOXD3* by comparison to previously published values for LongSAGE libraries prepared from H9 and other undifferentiated human ES cell lines [26]. The absolute level of tags for these four pluripotency genes in the Tag-Seq libraries is up to 80-fold higher than in the published LongSAGE libraries. We then compared the frequencies and distribution of other tags in the Tag-Seq SAGE libraries with data from equivalently sized meta-libraries generated in silico from the published LongSAGE data sets (obtained from nonpluripotent as well as pluripotent cell populations; Fig. 1G). This comparison showed the tag representation in the two new undifferentiated human PS cell Tag-Seq SAGE libraries reported here (hs0046 and hs0212) to be highly correlated, both with each other and with the tags in the undifferentiated human ES cell LongSAGE meta library.

The depth of the three Tag-Seq SAGE libraries made it possible to identify 319 genes whose transcripts were present at various levels in the Tag-Seq libraries prepared from both of the undifferentiated cell types (see Supplementary Table E2; online only, available at www.exphem.org). These 319 genes have been previously found to be involved in transcription, chromatin maintenance, and membrane receptor signaling.

Use of seriation to identify developmentally restricted gene subsets

As a first approach to identifying which of the 319 genes might be most tightly linked to the maintenance of pluripotency, we undertook a further analysis of their representation in 17 previously published LongSAGE libraries prepared from human cells that span a broad time course of development. These additional libraries were derived from H9 and other human ES cells before and after induction of their differentiation, as well as various cell populations from several adult human tissues (total = 17 libraries; Fig. 1H). For this analysis, we employed a heuristic approach (termed *seriation*) that identifies groups of genes with similar expression levels in different libraries (termed *Supercontigs*) [21] from pair-wise correlations of tag frequencies in the libraries analyzed. We thus identified three Supercontigs (see Supplementary Table E3; online only, available at www.exphem.org). One contained 114 genes whose expression was mostly highly restricted to undifferentiated human ES cells (Supercontig 1, Fig. 2A). This group contained tags for *POU5F1*, *NANOG*, *SOX2,* and *FOXD3* and other genes whose expression is known to decrease upon human ES cell differentiation. The average level of expression of these 114 genes in all 17 LongSAGE libraries is shown in Figure 2B. A second group (Supercontig 2) consisted of 145 genes whose expression was generally higher in differentiated cells. The third subset (Supercontig 3) contained the remaining 60 genes whose expression patterns did not fit within either of the first two groups.

Identification of E2F motifs in promoters of genes expressed in human PS cells

We then used PASTAA software to analyze the promoters of the genes contained in these three Supercontigs, to look for transcription factors that might participate in regulating their expression [14]. PASTAA interrogates groups of coexpressed genes and ranks their likelihood of being regulated by a previously established transcription factor PWM. Two separate PASTAA analyses were performed on each of the three Supercontigs. One involved interrogating a region extending 10-kb upstream from the transcription start site (distal analysis); the other interrogated a region $\pm$400 bp on each side of the transcription start site (proximal analysis). In assessing the results obtained, we took into consideration the fact that the PWM-based motif identification method may not distinguish between specific members of transcription factor families that bind to the same sequence, even though this methodology may reliably identify a common binding site shared by members of a particular transcription factor family.

PASTAA analysis of Supercontig 1 (i.e., the undifferentiated human PS cell-specific genes) showed SOX (p = 0.0018) and FOX (p = 0.041) PWMs to be highly ranked in the distal analysis, and NANOG (p = 0.013) and OCT/POU (p = 0.024) PWMs to be highly ranked in the proximal analysis (Fig. 2C, D; also see Supplementary Table E4; online only, available at www.exphem.org). Notably, the PASTAA data predicted binding of these four core pluripotency transcription factors to their own and each others' promoters (see Supplementary Table E4; online only, available at www.exphem.org), as expected [27]. PASTAA analysis of Supercontig 2 also predicted a FOX motif to be present in the distal promoter region of these genes (p = 0.0185) and a NANOG motif in the proximal promoter region (p = 0.00786; see Supplementary Table E4; online only, available at www.exphem.org). Analysis of Supercontig 3 predicted FOX motifs in both the distal (p = 0.0404)

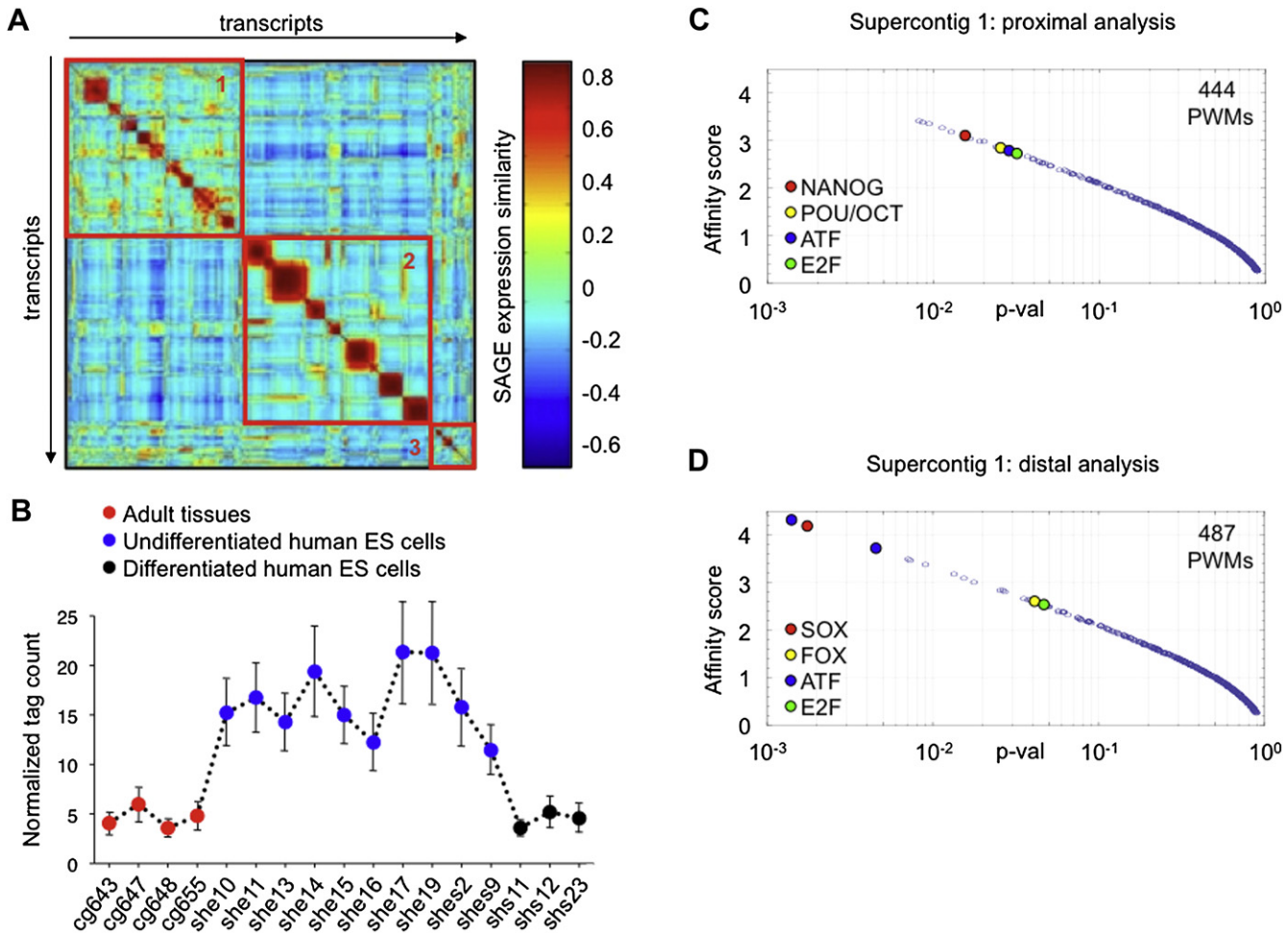


Figure 2. Analysis of differentially expressed genes and bioinformatic identification of E2F transcription factors as regulators in human PS cells. **(A)** Seriation analysis revealing three Supercontigs of co-expressed genes. **(B)** Average expression level for Supercontig1 genes across 17 previously established LongSAGE libraries. **(C, D)** Proximal and distal promoter analyses for Supercontig1 genes [multiple ATF motifs are identified in (**D**)].

and proximal ($p = 0.0327$) promoter regions of these genes, but no NANOG, OCT/POU, or SOX motifs in either the distal or proximal promoter regions (see Supplementary Table E4; online only, available at www.exphem.org).

The demonstrated ability of the PASTAA analyses to identify known transcriptional regulators of human PS cells (i.e., NANOG, POU5F1, SOX2, and FOXD3), encouraged us to investigate other novel candidates identified in the same analyses. E2F and ATF transcription factors were among the top rankings in both the distal and proximal promoter regions of genes in Supercontig 1 (see Supplementary Table E4; online only, available at www.exphem.org). ATF PWMs were ranked 1st and 4th in the distal analysis ($p = 0.0014$ and 0.0046) and 21st in the proximal analysis ($p = 0.029$), while E2F PWMs were ranked 21st in the distal analysis ($p = 0.045$) and 22nd in the proximal analysis ($p = 0.032$). Notably, E2F and ATF binding sites were found in the promoters of *NANOG, POU5F1*, *SOX2*, and *FOXD3* (see Supplementary Table E4; online only, available at www.exphem.org). Analysis of Supercontig 2 revealed E2F ($p = 0.0324$), but not ATF, motifs in the distal promoter regions of these genes and both E2F ($p = 0.00138$ to 0.0449) and ATF ($p = 0.0465$) motifs in the proximal promoter regions. For Supercontig 3, no evidence of either E2F or ATF motifs in the distal promoter regions was found, but E2F ($p = 0.0373$) and ATF ($p = 0.000819$ to 0.0431) motifs were present in the proximal promoter regions.

An analysis of published chromatin-immunoprecipitation (ChIP) data generated using antibodies against E2F1, E2F4, and E2F6 provided further evidence that these transcription factors are bound to the promoters of active genes in human PS cells [28]. The ChIP data showed that the promoters of *NANOG*, *POU5F1*, *SOX2,* and *FOXD3* were all variably enriched in bound E2F1, E2F4, and E2F6 in NTera2 EC cells, but this was not the case for any of the four nonpluripotent cell lines analyzed in the same way (see Supplementary Table E5; online only, available at www.exphem.org).

A similar ChIP experiment performed on undifferentiated human ES cells [27] showed E2F4 bound to the promoter of *FOXD3* (see Supplementary Table E5; online only, available at www.exphem.org). Taken together, these findings strongly suggested a role for E2Fs in regulating human PS cells.

Identification of RBBPs as candidate regulators of human PS cells

As a complementary approach to the identification of candidate regulators of human PS cells, we undertook a knockdown screen using siRNAs for each of the 319 genes initially identified from the Tag-Seq libraries (Fig. 3A). We chose the NTera2D1 human EC cells to evaluate the effects of these siRNAs because these cells share morphological, phenotypic, molecular, and functional properties with human ES and iPS cells [2,12], but are much easier to genetically manipulate. To detect an altered pluripotent state, we used Hoechst staining to measure the number of cells present in triplicate cultures 4 days after the addition of the siRNAs at a final concentration of 50 nM. This surrogate biological end point was chosen based on the assumption that loss of essential pluripotency maintenance factors would cause a marked and immediate prolongation of the cell cycle transit time, or detachment, and/or death of targeted cells. Support for this assumption was provided by the prior successful application of a similar end point to identify eight new pluripotency regulators in a RNA interference screen of mouse ES cells [29]. Preliminary experiments using positive control siRNAs (against the known pluripotency regulators *NANOG* and *POU5F1*), as well as negative control siRNAs (against *B2M* that does not regulate

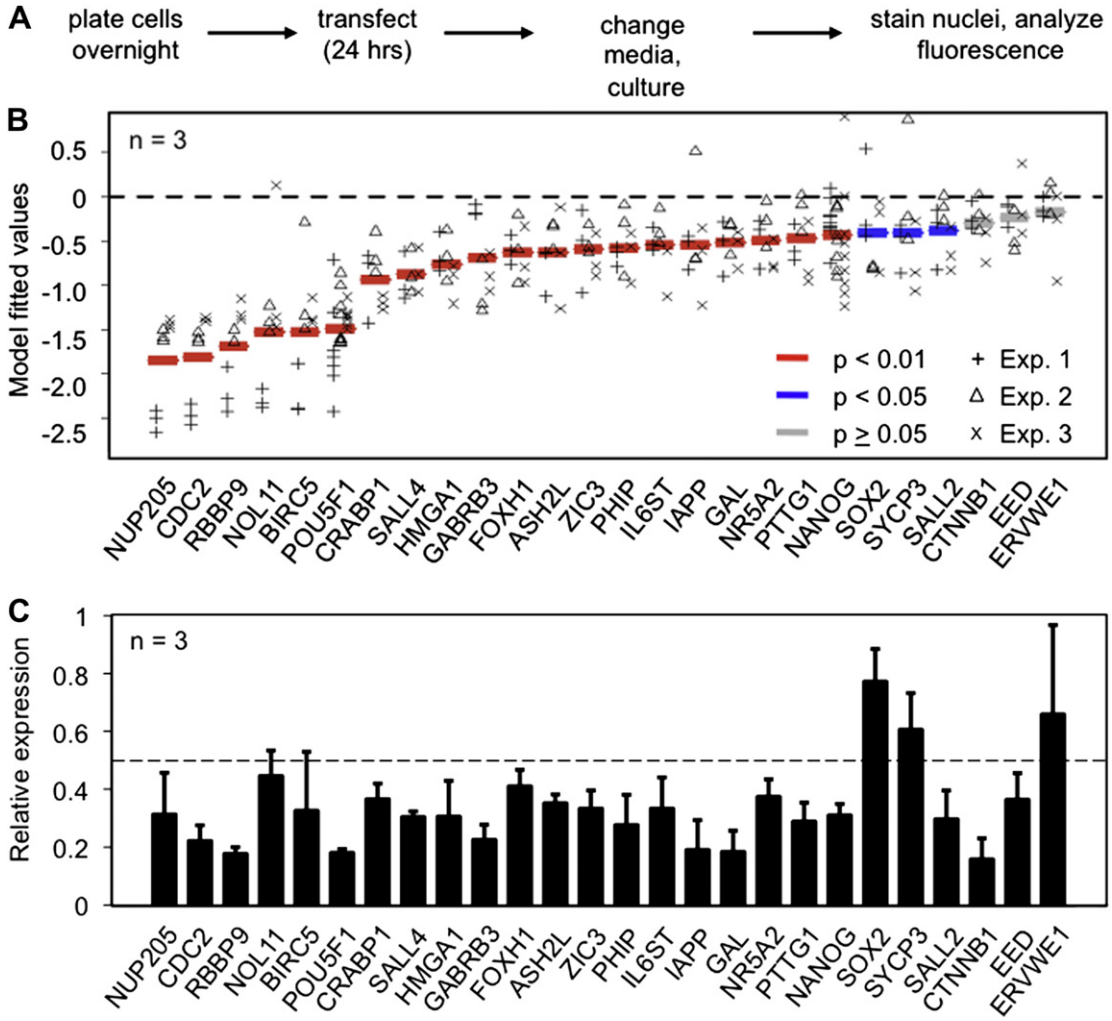


Figure 3. Identification of genes involved in NTera2D1 cell maintenance. **(A)** Diagram of the siRNA screening strategy used for the 319-gene siRNA screens, and the re-screening of the 26 preliminary hits. siRNA effects were determined through measurement of cell numbers (via Hoechst fluorescence) 4 days after siRNA treatment, in comparison to the negative control (*B2M* siRNA). **(B)** Rank ordering of the linear modeled data obtained from siRNA-based investigation of the 26 preliminary hits (3 experiments, 3 replicates/experiment: +, Δ, x). Each data point with the corresponding mean is shown for each treatment. *POU5F1* and *NANOG* have more than nine wells, as these were used as positive controls in each experiment. (**C**) Reduction of transcript levels by siRNAs tested in **(B)**. Most show decreases of >50%. Data shown as mean ± standard error of mean.

pluripotency), demonstrated the specificity that could be expected using this approach. These preliminary experiments also made it possible to optimize the reagent concentrations and exposure times used.

Two initial experiments were performed with siRNAs against all 319 genes (see Supplementary Figure E1; online only, available at www.exphem.org). The results identified 26 preliminary hits. These hits were then rescreened in another three experiments, at which stage the level of siRNA-mediated mRNA transcript downregulation was assessed. Most of these 26 siRNA treatments specifically decreased messenger RNA levels from the target gene by ≥50% (Fig. 3C), and 23 of the 26 also significantly reduced the number of cells present after 4 days below that measured in control cells treated with siRNAs against *B2M* (Fig. 3B). The 23 confirmed hits are known to be involved in regulating a range of biological processes, including transcription (ASH2L, FOXH1, SALL4, ZIC3, NANOG, SOX2, HMGA1, POU5F1, NR5A2); cell cycle progression (SYCP3, PTTG1, CDC2 BIRC5, RBBP9); signal transduction (CRABP1, GABRB3, IAPP); and apoptosis (BIRC5, IAPP, CDC2).

One of the genes whose inhibited expression was associated with the largest reduction of NTera2D1 cell maintenance was the RB-binding protein, RBBP9. This observation was of particular interest given that RB is a well-established binding partner and regulator of E2F transcription factor activity [30]. In addition, our PASTAA analyses had identified E2F transcription factors as potential regulators of gene expression in PS cells. Interestingly, re-examination of the H9 human ES cell Tag-Seq SAGE libraries showed little change in RBBP9 expression following induction of differentiation by exposure to RA. However, another RB-binding protein (RBBP4) was found to be expressed at a five-fold higher level in the undifferentiated H9 cells as compared to their differentiating derivatives.

siRNA-mediated loss of RBBPs induces changes in human PS cell functional properties

These findings prompted us to design a further series of experiments to investigate the potential role of RBBP4 and/or RBBP9 in human PS cell maintenance. As a first step, we asked whether the findings obtained in NTera2D1 EC cells would be replicated in human ES and iPS cells. Accordingly, we transiently transfected all three cell types with siRNAs for *B2M*, *RBBP4,* and *RBBP9* and then examined the amount of protein and total cells present 4 days later. Western blots showed RBBP4 and RBBP9 were specifically and consistently decreased across all 3 human PS cell types in response to the appropriate siRNA treatment (Figs. 4A, 5A). For the Western blots shown, densitometry revealed an average decrease of 60% in RBBP9 protein levels ($p = 0.017$) and 30% in RBBP4 protein levels ($p = 0.004$), relative to protein levels in the same cells treated with *B2M* siRNA, across all three human PS cell types. This decrease in RBBP4 and RBBP9 protein levels was also accompanied by a significant decrease in the number of cells present 4 days after siRNA exposure (Figs. 4B, 5B). siRNA-mediated inhibition of RBBP9 also reduced the yield of cells in RA-differentiated derivatives obtained from each of the cell types tested (Fig. 4B), whereas loss of RBBP4 had little or no effect on these RA-differentiated cells (Fig. 5B). This latter finding is consistent with the observation that *RBBP4,* but not *RBBP9* was down-regulated after RA induction of differentiation in H9 cells.

Treatment of human CA1 (ES) cells with siRNA against *RBBP4, RBBP9,* or *B2M* showed that a loss of either RBBP4 and RBBP9 was also associated with a significant and specific decrease in the frequency of CFCs able to form alkaline phosphatase-positive colonies, consistent with a loss of their prior pluripotent status (Figs. 4C, 5C). No change in cell or colony morphology was detected with any siRNA treatment.

To determine whether reduced sequestration of RB protein might be one of the consequences of reduced RBBP4 and RBBP9 protein levels, we next asked whether either actually binds RB in undifferentiated human PS cells. Gels containing RB that had been immunoprecipitated from NTera2D1 cell lysates gave 2 distinct bands (Fig. 6A) corresponding to the expected sizes of hyper- and hypophosphorylated RB protein [31–33]. Western blotting of these preparations with specific antibodies against RBBP4 and RBBP9 showed both of these proteins had coprecipitated with the RB protein (Fig. 6A).

siRNA-mediated RBBP loss results in gene expression changes indicative of differentiation

Treatment of NTera2D1 cells with siRNA to RBBP9 resulted in a significant reduction in *FOXD3* expression (Fig. 6B), whereas the siRNA-mediated decrease in *RBBP4* caused a rapid and significant decrease in expression of *POU5F1*, *NANOG*, *SOX2,* and *FOXD3* (within 24 to 48 hours, Fig. 6C). The specificity of these effects was again confirmed by their comparison to the lack of effects seen in the same cells treated with siRNA to *B2M*, or the transfection reagent alone (Fig. 6D).

To investigate accompanying changes in the expression of other genes, Affymetrix analyses were performed on RNA extracted from cells harvested 48 hours after being treated with siRNAs for *RBBP9* and 72 hours after being treated with siRNAs for *RBBP4* (or siRNAs for *B2M* as a control). These times were selected to enable gene expression changes to be examined when the greatest effects on known pluripotency genes had been seen. The results of these Affymetrix analyses are presented in Supplementary Figure E2 (online only, available at www.exphem.org). Notably, *RBBP4* and *RBBP9* transcripts were among

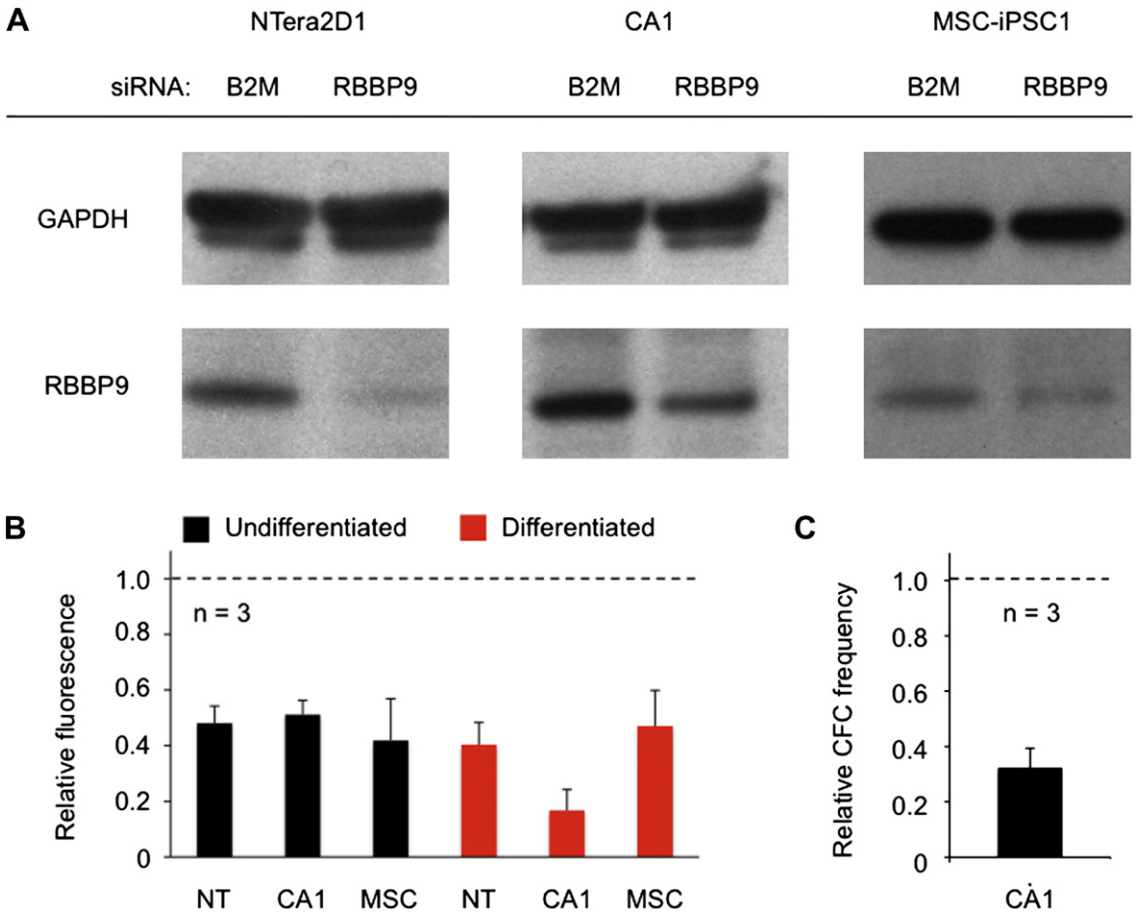


Figure 4. Effect of *RBBP9* siRNA treatment on undifferentiated and differentiating human PS cells. **(A)** Western blots of lysates from undifferentiated human PS cells 48 hours after siRNA treatment. **(B)** Reduced numbers of undifferentiated and differentiating human EC (NTera2D1, NT), ES (CA1), and iPS (MSC) cells detected 4 days after siRNA treatment as measured by Hoechst fluorescence. **(C)** CFC assays performed on human ES cells, and initiated 2 days after siRNA treatment, show reduced numbers of pluripotent cells detected after treatment with *RBBP9* siRNA compared to *B2M* siRNA. Data shown as mean ± standard error of mean ($p < 0.05$).

the most, if not the most, significantly down-regulated transcripts after *RBBP* siRNA treatment. Comparison of the overall effects of these siRNA treatments showed that approximately 10% of the genes whose expression increased after loss of RBBP4 and RBBP9 were shared. A similar degree of overlap was seen with the genes whose expression was induced to decrease. Level 5 GO analysis performed on the genes whose expression decreased after loss of RBBP4 showed an enrichment for genes involved in RNA processes and processing ($p = 9.4 \times 10^{-10}$ to 0.0018), and also in genes involved in cell cycle control, including well-known cell cycle regulators ($p = 0.0028$ and 0.0048; Table 1). Similarly, genes whose expression decreased when RBBP9 was reduced were associated with RNA and protein metabolism ($p = 2.7 \times 10^{-8}$ to 0.033), as well as cell cycle-related processes ($p = 0.0049$ and 0.006; Table 2). These analyses also confirmed that *NANOG* was among the most highly down-regulated genes after *RBBP4* siRNA treatment, and *FOXD3* was similarly highly down-regulated after *RBBP9* siRNA treatment.

Conversely, GO analysis of the genes whose expression increased after loss of RBBP4 showed an enrichment for genes involved in the regulation of GTPases ($p = 0.00034$), protein modification, and protein metabolic processes ($p = 0.0058$ to 0.048), and also in genes involved in nerve development (e.g., *EPHB1*, *EPHB2, FYN, SEMA3F, SEMA4D, SLIT2,* and *SMAD2,* Table 1). GO analysis of the genes whose expression increased after loss of RBBP9 showed an enrichment for genes involved in transcriptional regulation ($p = 6.6 \times 10^{-4}$ to 4.4×10^{-2}), plus genes involved in organ morphogenesis and tissue development ($p = 8 \times 10^{-4}$ to 0.044; Supplementary Figure E2; online only, available at www.exphem.org In particular, this included genes involved in neurogenesis (e.g. *EFNB3, EPHB1, NEUROD4, NEUROG3, NOTCH3, OLIG1, OLIG2, PAX6,* and *RUNX3; p* = 8×10^{-4} to 0.034, Table 2). Notably among these neurogenesis-related genes, increased expression of *EPHB1* was common to both *RBBP4* and *RBBP9* siRNA treatments.

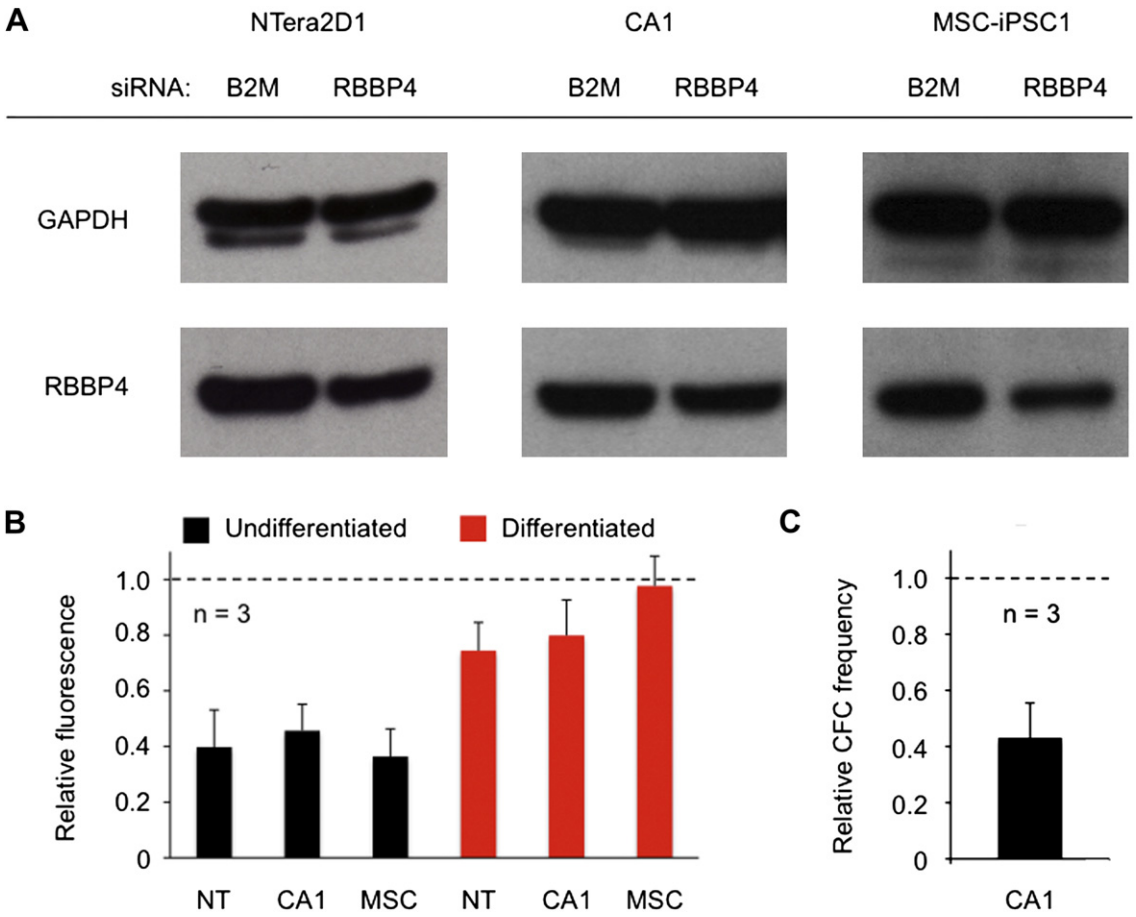


Figure 5. Effect of *RBBP4* siRNA treatment on undifferentiated human PS cells. **(A)** Western blots of lysates from undifferentiated human PS cells 72 hours after siRNA treatment. **(B)** Reduced numbers of undifferentiated human EC (NTera2D1, NT), ES (CA1), and iPS (MSC) cells, as measured by Hoechst fluorescence, 4 days after initiating the siRNA treatment, without a significant effect on differentiating cells. **(C)** CFC assays performed on human ES cells, and initiated 2 days after siRNA treatment, show reduced numbers of pluripotent cells detected after treatment with *RBBP4* siRNA compared to *B2M* siRNA. Data shown as mean ± standard error of mean ($p < 0.05$).

Promoter analysis of genes regulated by RBBP4 and RBBP9

Finally, we again used PASTAA distal and proximal analyses to investigate transcription factors that might be involved in mediating the gene expression changes seen after treatment of NTera2D1 cells with siRNAs to *RBBP4* and *RBBP9*. Both E2F and ATF motifs were among the top PWMs predicted in the distal and/or proximal regions of both the up- and down-regulated genes in both treatment groups, including those genes whose expression was commonly up- or down-regulated in the 2 groups (Supplementary Table E6; online only, available at www.exphem.org).

We then repeated an interrogation of the same ChIP data analyzed above to determine whether in undifferentiated human PS cells, E2Fs were bound to the promoters of genes whose expression increased after treatment with siRNAs to *RBBP4* or *RBBP9*. To do this, we focused on genes involved in neurogenesis, as we had found these to be commonly up-regulated after siRNA-mediated loss of either RBBP. The ChIP data showed no evidence of differential E2F1 or E2F4 binding in the PS cells compared to four nonpluripotent cell types for any of the seven neurogenesis genes examined whose expression increased after *RBBP4* siRNA treatment. For E2F6, binding at the promoters of three of these seven genes (*EPHB1*, *SEMA4D*, and *SLIT2*) was noted in the ChIP data for the pluripotent cells (Supplementary Table E5; online only, available at www.exphem.org). The ChIP data also showed no differential binding of E2F1 to the promoters of any of the nine neurogenesis genes examined whose expression increased after *RBBP9* siRNA treatment. Similarly, E2F4 was differentially bound to the promoter of only one of these nine genes (*OLIG2*) in PS cells. For E2F6, binding at the promoters of four of these nine genes (*EPHB1*, *EFNB3, NEUROG3, OLIG2*, see Supplementary Table E5; online only, available at www.exphem.org) was noted in the ChIP data for the PS cells compared to the four non-pluripotent cell types.

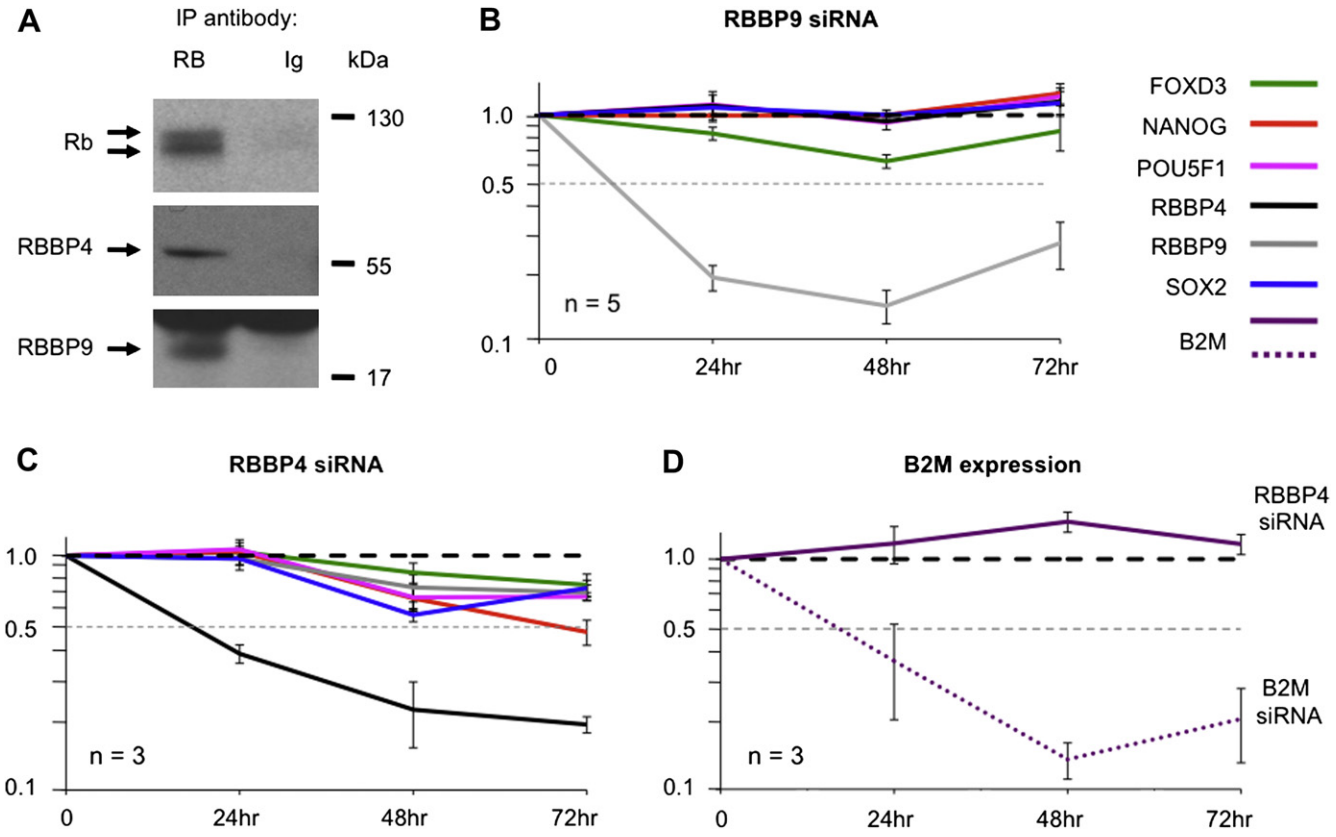


Figure 6. RBBP4/9 associate with RB and regulate human PS cell gene expression. **(A)** RBBP4 and RBBP9 co-immunoprecipitate with RB. **(B)** *RBBP9* siRNA treatment decreases *FOXD3* expression after 48 hours. **(C)** *RBBP4* siRNA treatment decreases *FOXD3, NANOG, POU5F1,* and *SOX2* expression after 72 hours. **(D)** *B2M* siRNA, but not *RBBP4* siRNA, decreases *B2M* expression. Data shown as mean ± standard error of mean ($p < 0.05$).

Discussion

Here, we demonstrate the utility of high-resolution transcriptome analyses together with bioinformatics and siRNA screens to discover candidate new regulators of the pluripotent state of human EC, ES, and iPS cells. These methods, both individually and in combination, consistently identified previously known critical pluripotency genes including *POU5F1, NANOG, SOX2,* and *ZIC3* [1–6,34,35]. The detection of *IL6ST* and *BIRC5* (i.e., *SURVIVIN*) provides further evidence of the biological relevance of the siRNA strategy used, as a signaling pathway involving IL6ST has been shown to be functional in NTera2D1 cells [36], and BIRC5 is required for teratoma formation by human ES cells [37]. Further examination of the other genes identified will help to discriminate those that participate in regulatory networks that are shared by all types of human PS cells. On the other hand, candidates that prove to be required only by EC cells may provide novel information about pathways important in tumor biology. Some of the genes identified here may also be useful in somatic cell reprogramming protocols, as suggested by the demonstrated reprogramming activities of *NANOG, POU5F1,* and *Nr5a2* [10], genes which were all identified as regulating human PS cells in the present study.

Of particular interest is the evidence we obtained indicating a key role of two RB-binding proteins, RBBP4 and RBBP9, in the maintenance of all three types of human PS cells studied. Affymetrix analysis, real-time PCR, and Western blotting showed the *RBBP* siRNA treatments caused specific, rapid, and significant decreases in the targeted mRNAs and proteins. This coincided with a significant decrease in cell turnover and/or cell adhesion and viability. PS cells are known to share a characteristically short cell cycle and high rate of proliferation, which are both affected by differentiation stimuli [29]. Consistent with this, RBBP9 expression in rat liver epithelial cells leads to a hyperproliferative state (due to its ability to sequester RB from RB/E2F1 complexes, thus releasing free/active E2F1 to stimulate cell cycle progression) [38]. The decreased cell numbers obtained as a result of reducing RBBP4 and RBBP9 expression are thus consistent with the disruption of a complex shared network that links maintenance of pluripotency with a specific program for controlling the cell cycle progression of PS cells. The ability to generate colonies of alkaline phosphatase-positive cells in low-density cultures (CFC assay) is another characteristic feature of pluripotent human ES and iPS cells that is rapidly lost when these cells begin to differentiate [16]. Notably,

Table 1. GO analysis of genes with altered expression after *RBBP4* siRNA treatment

Decreases with *RBBP4* siRNA: GO biological process level 5 "M phase of mitotic cell cycle" ($p = 0.0028$)	Increases with *RBBP4* siRNA: GO biological process level 5 "regulation of small GTPase-mediated signal transduction" ($p = 0.0065$)
C13orf34	*ABR*
CCNA1	*ACAP2*
CCNA2	*ADAP1*
CCNB1	*ARHGEF7*
CDC20	*ASAP3*
CDCA5	*FGD1*
CDCA8	*FGD5*
CENPV	*GEFT*
DCTN1	*PLCE1*
FBXO5	*PSD3*
FGF4	*RASA2*
HORMAD2	*RASGRP1*
NCAPH	*RGL3*
NEK3	*RGNEF*
NOLC1	*S1PR1*
NUDC	*SGSM3*
PPP5C	*SMAP2*
RAN	*SYNGAP1*
SIRT7	*TBC1D25*
SKA2	*TBC1D2B*
SSSCA1	*TBC1D5*
ZC3HC1	*TBC1D8*
ZW10	*TBC1D8B*

Table 2. GO analysis of genes with altered expression after *RBBP9* siRNA treatment

Decreases with *RBBP9* siRNA: GO biological process level 5 "mitosis" ($p = 0.0049$)	Increases with *RBBP9* siRNA GO biological process level 5 " neurogenesis" ($p = 0.0008$)	
ANAPC4	*AGRN*	*NDE1*
ANAPC7	*ATL1*	*NEFL*
BUB1	*ATN1*	*NOTCH3*
BUB3	*ATXN10*	*NRTN*
CCNA1	*AZU1*	*NTN3*
CDC26	*BAI1*	*ONECUT2*
CEP63	*BAIAP2*	*PAX6*
CLASP2	*BAX*	*PCSK1*
DCTN2	*BMP7*	*PHOX2A*
KATNA1	*CDK5R1*	*PIP5K1C*
KIF15	*CHRNB2*	*POU3F1*
LMLN	*CNP*	*POU4F2*
MIS12	*DRD2*	*PPT1*
NEK9	*DVL1*	*PTK2*
NIPBL	*EPHB1*	*RELA*
NUP37	*EVX1*	*RPGRIP1*
NUP43	*GBX2*	*RXRA*
OIP5	*GRIN1*	*SLITRK1*
PDS5A	*GRIN3A*	*SOX11*
STAG2	*ID3*	*STMN3*
VCPIP1	*IGF1R*	*TGIF2*
ZWILCH	*MAP1S*	*TNR*
	MC1R	*TP73*
	MNX1	*XRCC6*
	NAPA	

forced reduction of *RBBP4* and *RBBP9* expression caused a rapid and significant decrease in CFC frequency for human ES and iPS cells.

The Affymetrix and real-time PCR data presented here demonstrated that the cells present after *RBBP4* and *RBBP9* siRNA treatments had significantly reduced levels of core pluripotency gene transcripts (particularly *NANOG* following *RBBP4* siRNA treatment, and *FOXD3* following *RBBP9* siRNA treatment). Treatment with *RBBP4* and *RBBP9* siRNA also decreased expression of cell cycle genes, including some known to be regulated by E2F transcription factors (e.g., *CCNA2* and *CCNB1*). Additionally, loss of RBBP4 and RBBP9 was associated with increased expression of multiple differentiation regulators, particularly those involved in neurogenesis (e.g., *EPHB1, EPHB2, FYN, SEMA3F, SEMA4D, SLIT2, SMAD2* after treatment with *RBBP4* siRNA, and *EPHB1, EFNB3, NEUROD4, NEUROG3, NOTCH3, OLIG1, OLIG2, PAX6, RUNX3* after treatment with *RBBP9* siRNA). Taken together, these findings suggest that maintained expression of both RBBP4 and RBBP9 facilitates human PS cell maintenance by promoting or permitting expression of a combination of known pluripotency genes and cell cycle regulators, and by inhibiting expression of genes required for induction and propagation of specific differentiation programs.

The present studies also provide several lines of evidence that the roles of RBBP4 and RBBP9 in maintaining human PS cells may be mediated, at least in part, through interactions with the RB-E2F pathway that is known to control cell cycle progression [30], differentiation programs in various cell types including mouse and human ES cells [39–42], and somatic cell reprogramming [42]. Both RB-binding proteins can bind RB [38,43], as shown here for NTera2D1 cells. Bioinformatic data presented here and elsewhere [44–46] predicted E2F motifs to be present within the promoters of core pluripotency regulators, as well as a large number of other genes expressed in PS cells or early differentiating cells. Moreover, the promoters of genes whose expression was altered in PS cells treated with *RBBP4* or *RBBP9* siRNAs also contained E2F motifs, suggesting involvement of E2F factors as mediators of these effects.

Traditionally, E2F1 has been viewed as a transcriptional activator of cell cycle genes, whereas E2F4 and E2F6 are viewed as transcriptional repressors (E2F6 acting in an RB-independent manner). Inactivation of E2F1 or activation of E2F4, both of which could occur as a result of RB release after loss of RBBP4 or RBBP9, could thus conceivably result in decreased transcription of pluripotency genes. In support of this concept, we found that a number of genes known to be regulated by E2F transcription factors

(e.g., *CCNA2*, *CCNB1*) showed decreased expression in PS cells treated with *RBBP4* siRNAs. Additional insights were provided by analysis of published ChIP data. These confirmed selective binding of E2F1, E2F4, and E2F6 to the promoters of *NANOG, POU5F1, SOX2,* and *FOXD3* in human ES and EC cells, but not in four separate nonpluripotent cell lines [27,28]. Strikingly, in the same comparison, the PS cells showed no preferential binding of E2F1 to the promoters of any of the 15 neurogenesis genes examined whose expression increased after *RBBP* siRNA treatment. Similarly, for only one of these genes was E2F4 preferentially enriched at the promoter in PS cells, whereas the RB-independent E2F6 was enriched at 6 of the 15 gene promoters in PS cells compared to nonpluripotent cells. For genes whose expression increased after *RBBP* siRNA treatment, it is conceivable that RB release could activate ATF transcription factors to increase gene expression (via the ATF motifs identified here in their promoters), analogous to that shown for RB and ATF2 increasing TGF-β2 expression [47].

It is likely, however, that the consequences of *RBBP4* and *RBBP9* siRNA treatments involve more complex mechanistic changes that have many differences as well as some similarities. The fact that only about 10% of the genes with altered expression after *RBBP4* and *RBBP9* siRNA were shared underscores this possibility, in spite of the different treatment times used to generate Affymetrix profiles (i.e., 48 hours after addition of the *RBBP9* siRNAs and 72 hours after addition of the *RBBP4* siRNAs). Additional mechanisms could include the known involvement of *RBBP4* in NuRD and/or PRC2 complexes [48,49], both of which have been implicated in PS cell function [50,51], or through a recently identified serine hydrolase activity of RBBP9 [52]. Taken together, these findings strongly support the hypothesis that RBBP4 and RBBP9 play non-redundant and essential roles in maintaining the regulatory programs that define human PS cells. The present study sets the stage for future investigations aimed at defining the molecular mechanisms of RBBP4 and RBBP9 action in pluripotency maintenance, the initiation of differentiation and, potentially, somatic cell reprogramming.

Acknowledgments

This work was funded by grants from the Canadian Stem Cell Network in partnership with StemCell Technologies, Inc. (Vancouver, BC, Canada); the Canadian Institutes of Health Research (CIHR, Ottawa, ON, Canada); the National Institutes of Health (Bethesda, MD, USA), and Genome BC/Canada (Vancouver, BC, Canada) with core infrastructure support from the Michael Smith Foundation for Health Research (MSFHR, Vancouver, BC, Canada), and the British Columbia Cancer Foundation (Vancouver, BC, Canada). M.D.O. held a CIHR Industrial Postdoctoral Fellowship with matching funds from StemCell Technologies. O.M. was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) fellowship and is a Junior Trainee of the Michael Smith Foundation for Health Research. M.A.M. is a Senior Scholar of the MSFHR. S.A. is a Canada Research Chair in Molecular Oncology.

The authors are grateful to Steven McKinney in the Department of Molecular Oncology for his advice on the statistical evaluation of the data from the siRNA screen, and to Glenn Edin, Michael Li, Sarah Stevenson, Raymond Kwok, and Austin Hammond for expert technical support.

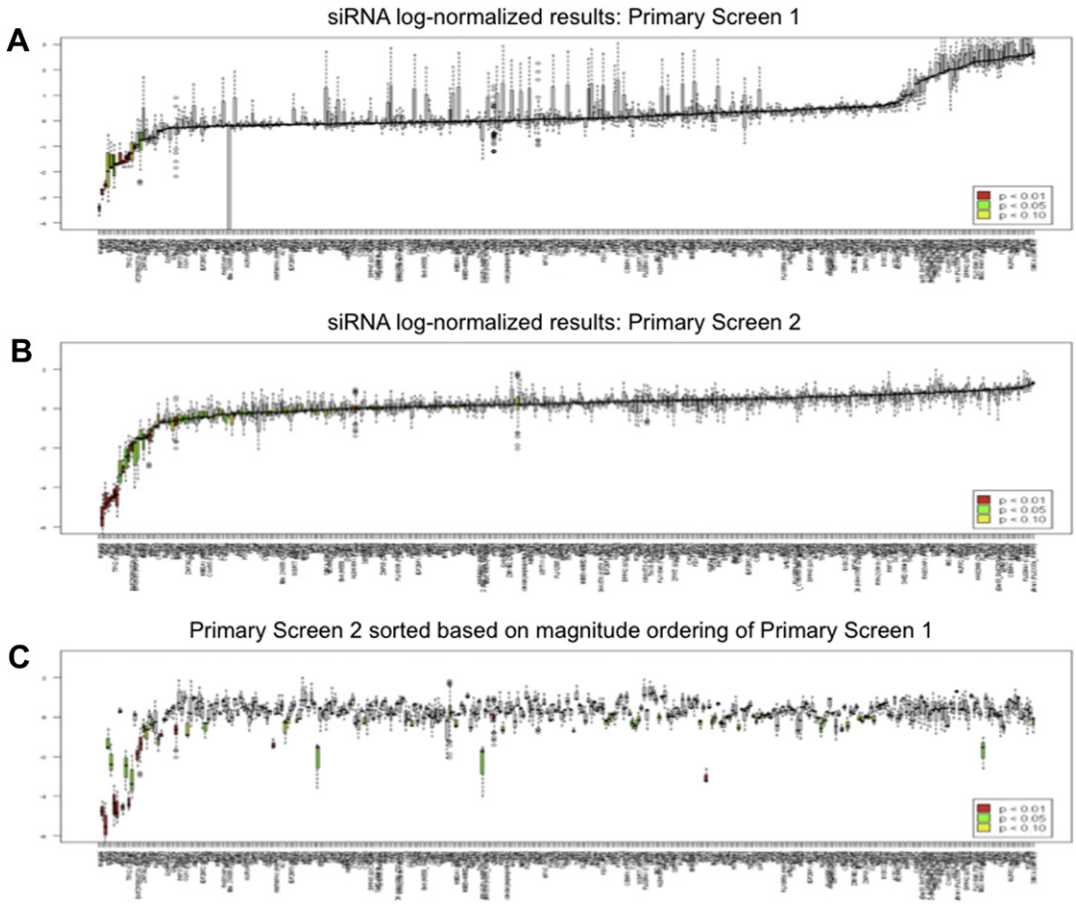
Conflict of interest disclosure

No financial interest/relationships with financial interest relating to the topic of this article have been declared.

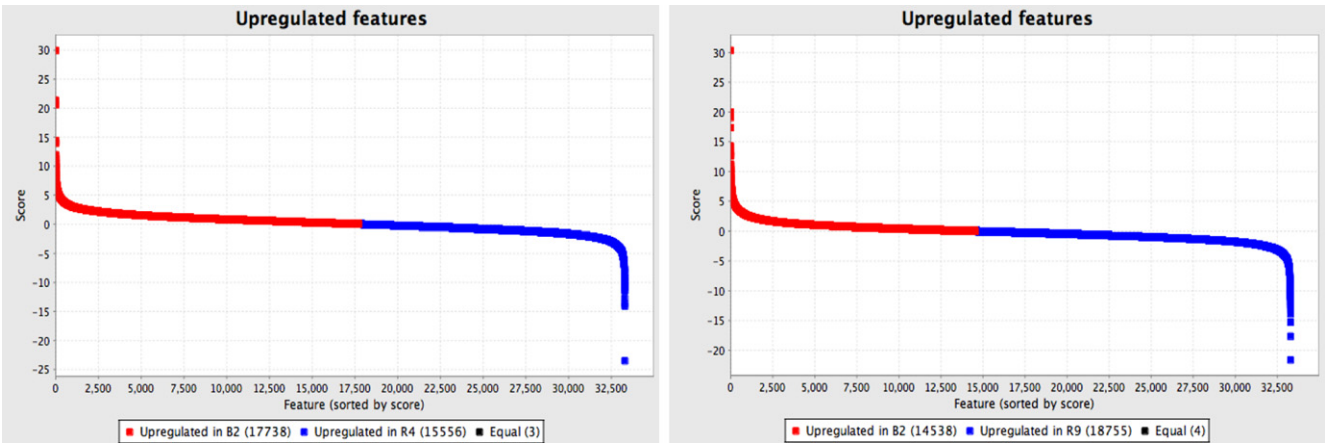
References

1. Boyer LA, Mathur D, Jaenisch R. Molecular control of pluripotency. Curr Opin Genet Dev. 2006;16:455–462.
2. Chambers I, Smith A. Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene. 2004;23:7150–7160.
3. Chen L, Daley GQ. Molecular basis of pluripotency. Hum Mol Genet. 2008;17:R23–R27.
4. Cole MF, Young RA. Mapping key features of transcriptional regulatory circuitry in embryonic stem cells. Cold Spring Harb Symp Quant Biol. 2008;73:183–193.
5. Orkin SH, Wang J, Kim J, et al. The transcriptional network controlling pluripotency in ES cells. Cold Spring Harb Symp Quant Biol. 2008;73:195–202.
6. Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res. 2007;17:42–49.
7. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–872.
8. Park IH, Zhao R, West JA, et al. Reprogramming of human somatic cells to pluripotency with defined factors. Nature. 2008;451:141–146.
9. Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318:1917–1920.
10. Heng JC, Feng B, Han J, et al. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell. 2010;6:167–174.
11. Nakagawa M, Koyanagi M, Tanabe K, et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol. 2008;26:101–106.
12. Andrews PW, Matin MM, Bahrami AR, Damjanov I, Gokhale P, Draper JS. Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. Biochem Soc Trans. 2005;33:1526–1530.
13. Thomson JA, Itskovitz-Eldor J, Shapiro SS, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998;282:1145–1147.
14. Roider HG, Manke T, O'Keeffe S, Vingron M, Haas SA. PASTAA: identifying transcription factors associated with sets of co-regulated genes. Bioinformatics. 2009;25:435–442.
15. Adewumi O, Aflatoonian B, et alInternational Stem Cell Initiative. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol. 2007;25:803–816.
16. O'Connor MD, Kardel MD, Iosfina I, et al. Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells. Stem Cells. 2008;26:1109–1116.
17. Ungrin M, O'Connor MD, Eaves CJ, Zandstra PW. Phenotypic analysis of human embryonic stem cells. Curr Protoc Stem Cell Biol. 2007;2:3.1–3.25.

18. Khattra J, Delaney AD, Zhao Y, et al. Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines. Genome Res. 2007;17:108–116.
19. Morrissy AS, Morin RD, Delaney A, et al. Next-generation tag sequencing for cancer gene expression profiling. Genome Res. 2009; 19:1825–1835.
20. Boon K, Osorio EC, Greenhut SF, et al. An anatomy of normal and malignant gene expression. Proc Natl Acad Sci U S A. 2002;99:11287–11292.
21. Morozova O, Morozov V, Hoffman BG, Helgason CD, Marra MA. A seriation approach for visualization-driven discovery of co-expression patterns in Serial Analysis of Gene Expression (SAGE) data. PLoS One. 2008;3:e3205.
22. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
23. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet. 2006;38:500–501.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995; 57:289–300.
25. Van Belle G, Fisher LD, Heagerty PJ, Lumley T. In: Biostatistics: a methodology for the health sciences. Hoboken, NJ: John Wiley & Sons; 2004.
26. Hirst M, Delaney A, Rogers SA, et al. LongSAGE profiling of nine human embryonic stem cell lines. Genome Biol. 2007;8:R113.
27. Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122:947–956.
28. Xu X, Bieda M, Jin VX, et al. A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. Genome Res. 2007;17:1550–1561.
29. Ivanova N, Dobrin R, Lu R, et al. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006;442:533–538.
30. Dimova DK, Dyson NJ. The E2F transcriptional network: old acquaintances with new faces. Oncogene. 2005;24:2810–2826.
31. Savatier P, Huang S, Szekely L, Wiman KG, Samarut J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene. 1994;9:809–818.
32. Fluckiger AC, Marcy G, Marchand M, et al. Cell cycle features of primate embryonic stem cells. Stem Cells. 2006;24:547–556.
33. Filipczyk AA, Laslett AL, Mummery C, Pera MF. Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells. Stem Cell Res. 2007;1:45–60.
34. Do JT, Scholer HR. Regulatory circuits underlying pluripotency and reprogramming. Trends Pharmacol Sci. 2009;30:296–302.
35. Lim LS, Loh YH, Zhang W, et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol Biol Cell. 2007;18:1348–1358.
36. He W, Gong K, Zhu G, Smith DK, Ip NY. Membrane distal cytokine binding domain of LIFR interacts with soluble CNTFR in vitro. FEBS Lett. 2002;514:214–218.
37. Blum B, Bar-Nur O, Golan-Lev T, Benvenisty N. The anti-apoptotic gene survivin contributes to teratoma formation by human embryonic stem cells. Nat Biotechnol. 2009;27:281–287.
38. Woitach JT, Zhang M, Niu CH, Thorgeirsson SS. A retinoblastoma-binding protein that affects cell-cycle control and confers transforming ability. Nat Genet. 1998;19:371–374.
39. Bieda M, Xu X, Singer MA, Green R, Farnham PJ. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res. 2006;16:595–605.
40. McClellan KA, Slack RS. Specific in vivo roles for E2Fs in differentiation and development. Cell Cycle. 2007;6:2917–2927.
41. Dannenberg JH, van Rossum A, Schuijff L, te Riele H. Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. Genes Dev. 2000;14:3051–3064.
42. Ruiz S, Panopoulos AD, Herrerias A, et al. A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. Curr Biol. 2011;21:45–52.
43. Qian YW, Wang YC, Hollingsworth RE Jr, Jones D, Ling N, Lee EY. A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. Nature. 1993;364:648–652.
44. Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008;133:1106–1117.
45. Muller FJ, Laurent LC, Kostka D, et al. Regulatory networks define phenotypic classes of human stem cell lines. Nature. 2008;455: 401–405.
46. Sharov AA, Ko MS. Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA Res. 2009;16:261–273.
47. Kim SJ, Wagner S, Liu F, O'Reilly MA, Robbins PD, Green MR. Retinoblastoma gene product activates expression of the human TGF-beta 2 gene through transcription factor ATF-2. Nature. 1992;358:331–334.
48. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 2002;16:2893–2905.
49. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999; 13:1924–1935.
50. Wang J, Rao S, Chu J, et al. A protein interaction network for pluripotency of embryonic stem cells. Nature. 2006;444:364–368.
51. Liang J, Wan M, Zhang Y, et al. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol. 2008;10:731–739.
52. Bachovchin DA, Brown SJ, Rosen H, Cravatt BF. Identification of selective inhibitors of uncharacterized enzymes by high-throughput screening with fluorescent activity-based probes. Nat Biotechnol. 2009;27:387–394.



Supplementary Figure E1. Data from siRNA screens performed on the 319 genes identified in Supplementary Table E2.



Supplementary Figure E2. Affymetrix gene expression data and GO analysis resulting from RBBP4 and RBBP9 siRNA treatments.