Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging reveals post-transcriptional regulation of proteins required for ES cell self-renewal

Robert N. O’Brien¹, Zhouxin Shen¹, Kiyoshi Tachikawa¹, Pei Jen Angel Lee¹, Steven P. Briggs¹,²

¹Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093-0380 United States of America
²To whom correspondence should be addressed: sbriggs@ucsd.edu

ATRA: all-trans retinoic acid
ES cell: Embryonic stem cell
EC cell: Embryonal carcinoma cell
FDR: false discovery rate
MRM: Multiple reaction monitoring
PTR: Post-transcriptional regulation
shRNA: Short hairpin RNA
Embryonic stem cells and embryonal carcinoma cells share two key characteristics: pluripotency (the ability to differentiate into endoderm, ectoderm, and mesoderm) and self-renewal (the ability to grow without change in an untransformed, euploid state). Much has been done to identify and characterize transcription factors that are necessary or sufficient to maintain these characteristics. Oct4 and Nanog are necessary to maintain pluripotency; they are down-regulated at the mRNA level by differentiation. There may be additional regulatory genes whose mRNA levels are unchanged but whose proteins are destabilized during differentiation. We generated proteome-wide, quantitative profiles of ES and EC cells during differentiation, replicating a microarray-based study by Aiba et al. who triggered differentiation by treatment with 1 μM all-trans retinoic acid (ATRA). We identified several proteins whose levels decreased during differentiation in both cell types but whose mRNA levels were unchanged. We confirmed several of these cases by RT-PCR and western blot. Racgap1 (also known as mgcRacgap) was particularly interesting because it is required for viability of pre-implantation embryos and hematopoietic stem cells, and it is also required for differentiation. To confirm our observation that Racgap1 declines during RA-mediated differentiation, we employed multiple reaction monitoring (MRM), a targeted mass-spectrometry-based quantitation method, and determined that Racgap1 levels decline by half during RA-mediated differentiation. We knocked down Racgap1 mRNA levels using a panel of 5 short hairpin RNAs (shRNAs). This resulted in a loss of self-renewal that correlated with the level of knockdown. We conclude that Racgap1 is post-transcriptionally regulated during blastocyst development to enable differentiation by inhibiting ES cell self-renewal.
INTRODUCTION

ES cells are interesting due to their pluripotent developmental potential and capability for self-renewal. Pluripotency is the ability to differentiate into cells representing or derived from, any of the three primordial germ layers that make up a post-implantation embryo. This property is shared with embryonal carcinoma (EC), a cancer of the testis. Self-renewal is distinct from proliferation in that the daughter cells are identical to the mother cells at both the genetic and epigenetic levels (4).

Aggregation and treatment of mouse ES and EC cells with retinoic acid (RA) has been shown to increase the proportion of ES and EC cells that differentiate into the neuroectodermal lineage (4, 5), though the rate at which these cells differentiate is quite different (1). Differentiation by RA treatment and aggregation is a standard first step in many mouse ES cell differentiation protocols, and is sufficient to efficiently drive mouse EC cells of line P19 to neural differentiation (1).

The last few years have lead to many advances in our understanding of the undifferentiated state of ES cells. The transcription factor Oct4, also called Oct3/4, was considered to be the key determinant and biomarker of undifferentiated cells. Oct4 was later found to dimerize with another transcription factor, Sox2 (6, 7). Around that time, another factor, Nanog was identified as being necessary for the undifferentiated ES cell state. More recent work from Yamanaka (8, 9) and Thompson (10) have identified combinations of transcription factors including Oct4, Sox2, Nanog, Lin28, cMyc and Klf4 which are sufficient to reprogram somatic cells into ES-like induced pluripotent stem cells (iPS) cells.

While the transcriptional networks that establish and maintain the pluripotent state have been extensively studied, there has been relatively little work identifying how signaling from the environment affects cells as they exit the undifferentiated state. Recent studies of early differentiation have identified Caspase 3 as a negative regulator of Nanog and Ronin protein stability (11, 12). Loading of ribosomes in undifferentiated ES cells is subject to extensive post-transcriptional control (13) resulting in changes in protein levels during differentiation that are not observed at the mRNA levels. Our study compares the proteome and transcriptome of pluripotent cells to identify cases where the proteins but not the mRNAs are enriched before or after differentiation.

There have been several deep proteomic analyses of human or mouse ES cells. Graumann et al. identified 5,111 proteins and quantified subcellular localization of proteins in undifferentiated mouse ES cells (14). SILAC was used in the study for a self to self comparison. Swaney et al. described 11,995 phosphopeptides from human ES cells (15, 16). Our study identified 5,489 proteins and provided quantitative comparisons of 4,986 proteins in mouse ES cells before and after differentiation. In addition our study compared embryonal carcinoma cells to ES cells, which share the properties of pluripotency and self-renewal but are otherwise distinct.

We designed our experiments to replicate the cell lines and differentiation
protocols published by Aiba et al. 2006, in order to compare our measurements of the proteome with their measurements of the transcriptome (1). To achieve the necessary coverage and sensitivity we used the MUDPIT LC ESI MSMS method (17, 18) to identify peptides, and in vitro iTRAQ mass tag labeling to measure their relative abundance.

iTRAQ labeling was done after protein extraction and protease digestion, therefore growth in labeling media was not required and >90% of the peptides from any sample were tagged. iTRAQ offers the advantage of multiplexing with 4 or 8 different mass tags; we used 4 tags to enable 4 samples to be compared directly within the same on-line chromatography run.

We confirmed biomarkers that distinguish undifferentiated pluripotent cells, as predicted by prior analysis of the transcriptome. More importantly, we report protein biomarkers that are under post-transcriptional regulation including several cases involving proteins that are essential for self-renewal. We discovered that Racgap1, which is enriched at the protein level but not the mRNA level in pluripotent cells, is necessary for ES cell self-renewal.

EXPERIMENTAL PROCEEDURES

Cell Culture—EC P19 cells (ATCC no. CRL-1825) were cultured as described in Aiba et al. (1) Cells were grown in DMEM supplemented with 10% Heat inactivated-Fetal Bovine Serum (HI-FBS) and 100 units penicillin/streptomycin. EC cells were subcultured 1:10 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, EC cells were plated on bacteriological plates in tissue culture media supplemented with 1 µM RA (10 mM stock of ATRA dissolved in 100% ethanol) as described in Aiba et al 2006(1). Media was replaced on day 2, EBs were collected for processing on day 4. Two biological replicates were performed for the EC whole cell experiments.

Nuclei from EC cells before and after differentiation were purified using a commercial kit (Pierce #78833) following the manufacture’s protocol. This worked well for EC cells but fractionations of ES cells and ES derived EBs failed due to the tough, fibrous nature of the ES-derived EBs. Three biological replicates were performed for the EC nuclei experiments.

129/SvEv mouse ES cells from ATCC were cultured as described in Aiba et al. Cells were grown in DMEM supplemented with 15% HI-FBS and 100 units pen/strep, 1000 units of LIF, 1mM Sodium Pyruvate, 1 mM NEAA and 55 µM β-mercaptoethanol. Cells were subcultured 1:5 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, 129/SvEv mouse ES cells were plated on bacteriological plates in ES culture media minus LIF as described in Aiba et al 2006(1) media was replaced on day 2, day 4 and day 6. Cells were treated with 1µM RA on days 4-8. Three biological replicates were performed for the ES whole cell experiments.

Cell lysis, reduction and trypsin digestion—100 µL cell pellets were lysed
in 250 µL lysis buffer (2% (w/v) RapiGest (Waters, 186002122), 1mM EDTA, and 50mM Hepes buffer (pH 7.2)). Cysteines were reduced and alkylated using 1 mM Tris(2-carboxyethyl) phosphine (TCEP, Fisher, AC36383)) at 95°C for 5 minutes followed by 2.5 mM iodoacetamide (Fisher, AC12227) at 37°C in the dark for 15 minutes. Protein concentrations were measured using Bradford assay (Pierce). Proteins were digested with trypsin (Roche, 03 708 969 001) at 37°C with an enzyme-to-substrate ratio (w:w) = 1:50 overnight.

**iTRAQ labeling of peptides**—For iTRAQ (Applied Biosystems, Foster City, CA) derivatization an aliquot of each digested sample (100 µg of total protein) was treated with one tube of one of the iTRAQ reagents in 70% isopropanol at pH 7.2 for 2 hours at room temperature. Labeled samples were dried down in a vacuum concentrator. 100 µL of water was added to each tube to dissolve the peptides. Samples tagged with 4 different iTRAQ reagents were pooled together. 1% trifluoroacetic acid (TFA), was added to a final pH of 1.4 to precipitate RapiGest. Samples were incubated at 4°C overnight and then centrifuged at 16,100 g for 15 minutes. Supernatant was collected and centrifuged through a 0.22 µM filter and was used for LC-MS/MS analysis. iTRAQ labeling efficiency was calculated by searching the MS/MS data specifying 4 possible iTRAQ modifications: 1) fully labeled; 2) n-terminus-labeled only; 3) lysine-labeled only; and 4) non-labeled. Using the above protocol we obtained higher than 90% iTRAQ labeling efficiency for all datasets (Table I). For all three comparisons (EC cell, EC nuclei, and ES cell before and after differentiation), the peptides from undifferentiated cells were labeled with 114 and 115 iTRAQ tags, while the peptides from differentiated cells were labeled with 116 and 117 iTRAQ tags.

**On-line separation of peptides by HPLC**—An Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA) delivered a flow rate of 300 nL per minute to a 3-phase capillary chromatography column through a splitter. Using a custom pressure cell, 5 µm Zorbax SB-C18 (Agilent) was packed into fused silica capillary tubing (200 µm ID, 360 µm OD, 20 cm long) to form the first reverse phase column (RP1). A 5 cm long strong cation exchange (SCX) column packed with 5 µm PolySulfoethyl (PolyLC, Inc.) was connected to RP1 using a zero dead volume 1 µm filter (Upchurch, M548) attached to the exit of the RP1 column. A fused silica capillary (100 µm ID, 360 µm OD, 20 cm long) packed with 5 µm Zorbax SB-C18 (Agilent) was connected to SCX as the analytical column (the second reverse phase column). The electro-spray tip of the fused silica tubing was pulled to a sharp tip with the inner diameter smaller than 1 µm using a laser puller (Sutter P-2000). The peptide mixtures were loaded onto the RP1 using the custom pressure cell. Columns were not re-used. Peptides were first eluted from the RP1 to the SCX column using a 0 to 80% acetonitrile gradient for 150 minutes. The peptides were fractionated by the SCX column using a series of salt gradients (from 10 mM to 1 M ammonium acetate for 20 minutes), followed by high resolution reverse phase separation using an acetonitrile gradient of 0 to 80% for 120 minutes. Typically it takes 4 days (38 salt fractions) for each full proteome analysis.

**Tandem Mass Spectrometry analysis**—Spectra were acquired using a
LTQ linear ion trap tandem mass spectrometer (Thermo Electron Corporation, San Jose, CA) employing automated, data-dependent acquisition. The mass spectrometer was operated in positive ion mode with a source temperature of 150°C.

The full MS scan range of 400-2000 m/z was divided into 3 smaller scan ranges (400-800, 800-1050, 1050-2000) to improve the dynamic range. Both CID (Collision Induced Dissociation) and PQD (Pulsed-Q Dissociation) scans of the same parent ion were collected for protein identification and quantitation. Each MS scan was followed by 4 pairs of CID-PQD MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 minute was used to improve the duty cycle of MS/MS scans. About 20,000 MS/MS spectra were collected for each salt step fractionation.

Data analysis—The raw data were extracted and searched using Spectrum Mill v3.03 (Agilent). The CID and PQD scans from the same parent ion were merged together. MS/MS spectra with a sequence tag length of 1 or less were considered to be poor spectra and were discarded. The remaining MS/MS spectra were searched against the International Protein Index (IPI) mouse database (v3.31, 56,555 protein sequences). The enzyme parameter was limited to fully tryptic peptides with a maximum miscleavage of 1. All other search parameters were set to default settings of Spectrum Mill (carbamidomethylation of cysteines, iTRAQ modification, +/- 2.5 Da for precursor ions, +/- 0.7 Da for fragment ions, and a minimum matched peak intensity (SPI%) of 50%). A concatenated forward-reverse database was constructed to calculate the in-situ false discovery rate (FDR). The total number of protein sequences in the combined database is 113,110. Cutoff scores (Table II) were dynamically assigned to each dataset to maintain the false discovery rate less than 1% at the protein level. Only proteins with 2 or more unique peptides were validated and selected for following quantitative analysis. Proteins that share common peptides were grouped to address the database redundancy issue. The proteins within the same group shared the same set or subset of unique peptides.

Data access—All mass spectra used in this study will be publicly available at the Proteome Commons Tranche website (https://proteomecommons.org/tranche/) at the time of publication.

Multiple Reaction Monitoring (MRM) analysis of Racgap1—Undifferentiated mES 129 cells and differentiated mEB samples were lysed and digested by the protocol described above. MRM experiments were performed on an Agilent 6410 triple quadrupole mass spectrometer equipped with Agilent 1200 nano-flow LC, Agilent 1200 micro auto-sampler, and HPLC-Chip Cube MS interface. Two synthetic peptides from Racgap1 (VSSLGPVTTPFEQLVK and TTDNLGTPQNTGGMR) were purchased from Eton Bioscience, Inc. One synthetic peptide (VAPEEHPVLLTEAPLNPK) from Actin and one from GAPDH (VPTPNVSVDLTCR) were also purchased for the purpose of data normalization. Method optimization was performed to determine the optimal transitions and collision energies for each protein. Synthetic peptide standards were spiked into samples at a high level (1pmol peptide per µg sample) for method optimization. Each sample was analyzed both with and without spike-in
standards. The Racgap1 peptide standards were spiked in at 2fmol per µg total peptide, which is close to the endogenous Racgap1 level. There were 4 total samples (mES no spike-in, mES with spike-in, mEB no spike-in, and mEB with spike-in). Each sample was analyzed 6 times. A total of 24 LC/MRM runs were performed and the order of injections was randomized to minimized systematic errors. 1µg of digested peptides were injected by Agilent 1200 auto-sampler and subject to a 40-minute reverse phase separation. The intensities from the best transitions (Racgap1: 864.5->1258.7, collision energy = 28, dwell time = 50ms; beta-actin: 977.5->1291.8, collision energy = 31, dwell time = 50ms) were used to calculate the absolute amount of Racgap1.

Quantification by iTRAQ mass tagging reagent—Protein iTRAQ intensities were calculated by summing the peptide iTRAQ intensities from each protein group. Peptides shared among different protein groups were removed before quantitation. A minimal total iTRAQ reporter ion intensity (sum of all 4 channels) of 100 was used to filter out low intensity spectra. Isotope impurities of iTRAQ reagents were corrected using correction factors provided by the manufacturer (Applied Biosystems).

Median normalization was performed to normalize the protein iTRAQ reporter intensities in which the log ratios between different iTRAQ tags (115/114, 116/114, 117/114) are adjusted globally such that the median log ratio is zero.

Quantitative analysis was performed on the normalized protein iTRAQ intensities. Protein ratios between undifferentiated and differentiated cells were calculated by taking the ratios of the total iTRAQ intensities from the corresponding iTRAQ reporters. T-test (two tailed, paired) was used to calculate the p-values. Proteins with more than 50% change and p-values less than 0.05 were considered significantly changed after differentiation.

Reanalysis of microarray data—Microarray data from Aiba et al. were obtained from pubmed GEO datasets. Data were processed by normalizing expression values to reference data and generating undifferentiated to differentiated ratios of mRNA measurements from each replicate of the experiment. Ratios were averaged in excel using the geomean function, and the student’s t-test was performed to obtain p values. Protein and mRNA data were combined in Microsoft access.

RT-PCR—RNA was isolated from undifferentiated and differentiated cells using Trizol (invitrogen cat #15596-026) according to manufacturer’s instructions. RNA quality was checked by agarose gel electrophoresis, and 2µg total RNA was treated with DNAase Turbo (ambion catalog #AM2238) according to manufacture’s instructions. After DNase inactivation, cDNA was synthesized by reverse transcription with/without Superscript II (invitrogen) and 10 µM random hexamer primers. RT PCR reactions were performed using the following primer sequences:

Beta actin: F-GATCTGGGCACCACACCTTCTACAATG R-CGTACATGGCTGGGTGTGAG, Oct4: F-CTCCCGAGGAGTCCCAGGACAT, R-GATGGTGATTGGCTGCTGAACACCT, Pdlim7: F-GCACTCAGGAGCAGGCACGATGG, R-CCTCCGGCGCGTGAGCCG,
Dpysl2: F-GTGACGCCCAAGACGGTGAC, R-ATGTTGTCGTCAATCTGAGCACCAG, H2afy F- GGAGAAGAAGGGCGGCAAGG, R-GGCCTGCACTAATAGCAGCTC, Utf1 F- GGTTGCGCCGCCGTCTACTG, R-GCAGGGGCAGGTTCGTCATTTTC, Racgap1 F-TCCCTATGATCCACFCTACAGAGAGTG, R-GCGCTCCACCACCTTG, Sall4 F-GGAGAGAAGCCTTTCGTGTG R-CTCTATGGCCAGCTTCCTTC. All RT-PCR reactions were performed at minimal cycle numbers necessary to observe bands (25 cycles for UTF1 and actin; 27 for others) avoid saturation of the PCR products.

**Western blotting**—Cells were collected in RIPA buffer and proteins were quantified by Bradford. Protein samples were prepared in 1X SDS loading buffer containing β-ME, sonicated and run on 10% tris-glycine protein gels, transferred to PVDF membranes. Membranes were blocked with 5% W/V non-fat dry milk in PBS-0.05% Tween and probed with the appropriate antibodies at concentrations from 1:100 to 1:1000. Specific antibodies: αUTF1 (RB chemicon, 1:1000), αSall4 (Santa Cruz 1:1000), PDLIM7 (RB 1:1000 Chemicon), H2afy (1:100 Santa Cruz), Dpysl2 (1:1000 abnova), β-Actin (MS SCBT 1:1000) was used as loading control in all cases.  

**Cellular Immunofluorescence**—Undifferentiated cells and differentiated cells were plated on coverslips coated with poly-d-lysine and cultured overnight. Cells were fixed in 4% paraformaldehyde, permeablized with PBS-0.2% Triton, blocked with 10% BSA in PBS-0.1% triton for 10 hour and stained with the appropriate pair of primary antibodies diluted 1:1000 (Oct4 and NeuN or UTF1 and βIII-Tubulin) in blocking buffer or left in blocking buffer, for secondary alone controls. Cells were washed 3X for 15 minutes in PBS 0.1% Triton and incubated with secondary antibodies from molecular probes diluted 1:1000 (antiMouse-534 and antirabbit-488). Slides were washed 3X and mounted in vectashield (vector labs) with dapi.

**Knockdown of Racgap1**—10⁵ undifferentiated ES cells were transfected via lipofection with Lipofectamine 2000 (1 µl in 50 µl optimem and 1 µg plasmid DNA). Each well was transfected with one of 5 Racgap1 targeting shRNA vectors from open Biosystems, control plasmid (pCAG) or mock (no DNA). Cells were grown in undifferentiated conditions in the absence of pen-strep for 24 hours and then selected in 1µg/ml puromycin for 24 hours. After this time, cells were passaged onto 6-well plates, puromycin was removed and the cells were allowed to form colonies for 3 days. Cells were fixed in 4% PFA, blocked in 10% BSA and immunostained with a rabbit anti-Oct4 antibody from Santa Cruz biotechnology (H-134) diluted 1:100 in 10% BSA. Cells were then stained with the Pierce DAB staining kit (cat #36000). Colonies were counted; standard deviation, mean and student’s t-test were quantified.

**RESULTS**

*Differentiation of ES and EC cells results in a loss of pluripotency biomarkers*—To better understand the protein makeup of pluripotent cells, and to identify differences between the transcriptome and the proteome, we replicated
the experimental design of a transcriptome study from Aiba et al. 2006 (1). The experimental design consisted of differentiating ES or EC cells by aggregation and retinoic acid treatment resulting in formation of embryoid bodies that are composed of cells that have lost expression of pluripotency markers and are in various stages of entering the neuroectodermal lineage. Specifics of the two protocols are illustrated in Fig. 1.

To verify that undifferentiated cells expressed pluripotency biomarkers and that differentiated cells had lost these markers, we used immunofluorescence to assay key biomarkers (Fig. S1). Undifferentiated ES cells strongly expressed Oct4 and UTF1, but did not express the neuroectodermal marker β III-Tubulin or the neuron-specific marker NeuN (Fig. S1A). Differentiated ES cells lost expression of Oct4 and UTF1, had no up-regulation of β III-Tubulin, and weak expression of NeuN (Fig. S1B). In contrast, undifferentiated EC cells strongly expressed β III-Tubulin along with Oct4 and UTF1 (Fig. S1C). Differentiated EC cells also maintained residual levels of UTF1 protein (Fig. S1D).

**Validated mass spectra**—We recorded approximately 6.8 million MS/MS spectra. After merging and filtering, we validated a total of 386,374 spectra using the cutoff scores listed in Table I. Among the validated spectra, 580 were from the decoy database, indicating a FDR of 0.15% at the spectrum level. There were a total of 36,794 unique peptides from the validated spectra, among them 132 were from the decoy database, resulting in a FDR of 0.36% at the peptide level. Different charge states from the same peptide were considered as one peptide.

**Efficiency of iTRAQ mass tagging**—iTRAQ (19, 20) can be used for multiplexed peptide profiling of up to four different samples. This approach labels samples with four independent reagents of the same mass that, upon fragmentation in MS/MS, give rise to four unique reporter ions (m/z = 114–117) that are subsequently used to quantify the four different samples, respectively.

iTRAQ reagents originally were not usable in ion trap instruments because of the “one third rule”. To illustrate, fragment ions of a m/z 900 parent will not be detected below m/z 300, and this would normally prevent the detection of the iTRAQ reporter ions. To overcome this limit, a new collision-activated fragmentation technique, called pulsed-Q dissociation (PQD), was invented which enables routine and reliable measurement of ions down to 50 m/z (21). We have developed and successfully applied a combined CID/PQD scan approach for peptide identification and quantitation (22, 23).

To determine the iTRAQ labeling efficiency, MS/MS spectra were searched against a concatenated forward and reverse mouse protein database searching for four possible iTRAQ modifications: 1) unmodified, 2) fully modified, 3) modified on the N-terminus or 4) modified on lysine side chains only. In all cases, labeling of tryptic peptides with the iTRAQ reagent was at least 92% efficient. The iTRAQ labeling efficiency of ES cell samples was 98.5%, while EC whole cell and ES nuclear samples were labeled at an efficiency of 92% and 97.7%, respectively (Table II).

**Managing protein redundancy**—A problem that any large-scale proteomics study must deal with is the redundancy in protein databases. In many cases,
tryptic peptides can be shared between different isoforms of a protein derived from the same locus, or from homologous proteins or even unrelated proteins. We address this redundancy by assigning proteins sharing the same peptides to a "protein group". Each protein group has at least one unique peptide. Proteins within the group may contain additional peptides. If two proteins share peptides and contain peptides not present in the other, they will be assigned to separate protein groups. A “group leader” heads protein groups that contain more than one protein accession number. The group leader is the protein with the highest identification score (the sum of the identification scores of the identified peptides composing the protein), which usually is the protein containing the largest number of peptides. In the remaining text, the term “protein” should be taken to mean "protein group leader". For iTRAQ quantitation, peptides that are shared between protein groups were removed and only peptides that are unique to each group were used.

iTRAQ mass tags enable relative quantitation of proteins—Analysis of 129/SvEv ES cells before and after retinoic acid mediated differentiation identified 4,053 proteins with a false discovery rate (FDR) of 0.56% (23 proteins from the decoy database out of 4,053 proteins identified) (Table III, supplemental Table S6). Removing all single peptide hits lowered the FDR to 0.1% (4 proteins from the decoy database out of 3,613 proteins with >1 peptide identified; Fig. S2). After removing proteins with low iTRAQ reporter ion intensities (sum of all channels below 100), we calculated the relative abundance of 3,566 proteins before and after differentiation. Similarly, we calculated the ratios of 3,801 proteins from EC cells with an FDR of 0.2% (9 proteins from the decoy database; Table III). The distribution of fold changes in all three datasets is illustrated in Fig. 2.

To increase the depth of our proteome coverage, we purified nuclei from cells before and after differentiation, using a fractionation kit (Pierce #78833). This worked well for EC cells but fractionations of ES cells and ES derived EBs failed due to the tough, fibrous nature of the ES-derived EBs.

Analysis of nuclei from EC cells enabled 3,669 protein ratios to be measured using the same criteria described above.

iTRAQ mass tag ratios reveal conserved protein biomarkers of pluripotency—To identify proteins in pluripotent cell types we first tried a two-tailed paired student’s t-test with a p-value cutoff of 0.05 and a fold-change cutoff of 1.5-fold. Supplemental Table S1 summarizes the results. Only 4 proteins were classified as specific to the pluripotent cells across the three datasets (Fig. 3A); 21 proteins were specific for differentiated cells across all datasets (Fig. 3B). The proteins present before differentiation in all three datasets include the adhesion protein Cdh1 (E-Cadherin) and the nuclear importin Kpna2. Of the 21 proteins enriched across all three RA-differentiated cells, 6 are metabolic enzymes. This list excludes a large number of proteins known to be enriched in pluripotent cells, such as Oct4 (24-31), UTF1 (32-37) and Dnmt3b (38-41). These and other known pluripotency-specific proteins were identified in our profiles. They met the threshold of >1.5 fold enrichment before differentiation in all replicates across all three datasets (Supplemental Table S2). We found that
the t-test was not an adequate statistical measure of reproducibility of enrichment. Variability in the dynamic range of quantitative measurements across experiments resulted in poor p-values in cases where enrichment was reproducibly observed across all three replicates. For this reason, we decided that the t-test is too stringent for identifying proteins with real fold differences. Instead we proceeded with a fold change cutoff to identify proteins with two or more unique peptides that reproducibly fall into the category of >50% enriched before or after differentiation across all 3 biological replicates.

Using fold change alone, we identified 101 proteins >1.5 fold enriched before differentiation in both ES and EC cells in all three datasets (Fig. 3C). These include known pluripotency associated proteins Oct4, UTF1, Dnmt3b, Cadherin 1 (42-50) (embryonic cadherin), and Tcf3 (51). Fold change also identified 181 proteins enriched after retinoic acid mediated differentiation in all three experiments (Fig. 3D, Supplemental Table S3). These proteins include the retinoic acid response proteins Rbp1, Crabp1 and Crabp2 (52-59), and Cadherin 2 (neural cadherin) (60-65); the embryonic patterning gene HoxB6 (66-70); and metabolic proteins. Thus, identification of proteins with >1.5 fold change during differentiation of two pluripotent cell types across three experiments reveals a group of proteins that includes many already known to be associated with pluripotency or differentiation, and additional proteins not previously associated with pluripotency.

**Biological processes enriched before and after differentiation**—To better understand the proteins enriched in cells before or after differentiation, we used the DAVID Bioinformatics Resource to identify gene ontology (GO) terms associated with the undifferentiated or differentiated states ([http://david.abcc.ncifcrf.gov/home.jsp](http://david.abcc.ncifcrf.gov/home.jsp)) (71, 72). The GO terms relating to biological processes enriched before differentiation in all three datasets included ribosome biogenesis, one-carbon metabolism, and amine biosynthesis (Fig. S4A). The GO terms relating to biological processes enriched after RA and aggregation-mediated differentiation all related to oxidative phosphorylation of glucose via the TCA cycle (Fig. S4B).

**Comparison of GO terms enriched in proteome and transcriptome data**—To compare the insights provided by genome-wide mRNA and protein profiles, we used the DAVID Bioinformatics Resource to identify GO terms associated with our protein dataset and with the published mRNA data from Aiba et al. ([http://david.abcc.ncifcrf.gov/home.jsp](http://david.abcc.ncifcrf.gov/home.jsp)) (71, 72). In all cases mRNA data produced more GO terms than the protein data, due to the greater depth of coverage. GO annotations of proteins were very similar to the annotations of mRNAs, especially in undifferentiated ES cells (Fig. S4). There were more differences between the mRNA and protein data of differentiated ES cells as well as undifferentiated and differentiated EC cells. However in all cases, the annotations assigned to the proteins were >50% shared with the annotations assigned to the transcripts indicating that most pathways observed were not subject to extensive post-transcriptional regulation.

**Identification of post-transcriptional regulation in pluripotent cells**—Our experiments were modeled on the global transcriptome study of Aiba et al., using
cell lines and differentiation conditions that replicate their study. We reanalyzed the data from Aiba et al. and compared the resulting mRNA measurements directly to our proteome measurements. Direct comparisons of mRNA and protein ratios for ES and EC cells are illustrated as scatterplots in Figs. 4A and 4B, respectively. Of the 282 proteins enriched before or after differentiation, 196 had corresponding data from the microarray; when these were compared, 95 of 196 (48%; Supplemental Table S4), agreed with our protein data. These gene products are therefore putative markers of pluripotency observed in independent experiments at both the mRNA and protein levels. On the other hand, 40 of 196 genes (20%) are enriched at the protein level but not at the mRNA level (Supplemental Table S5). These genes may be subject to post-transcriptional regulation during RA mediated differentiation of pluripotent cells. Mechanisms such as translational repression by miRNAs, destabilization by ubiquitylation or other mechanisms of proteolysis could explain the discrepancies between changes in protein and mRNA levels.

To confirm the discrepancy between protein and mRNA changes, we tested 4 selected cases (Pdlim7, H2afy, Dpysl2, and Sall4) by RT-PCR and by western blot (Fig. S5). These genes were chosen from candidate genes whose protein levels are changed but whose mRNA levels were unchanged and which had commercial antibodies available. In each of the 4 cases, the results of the western blots confirmed our iTRAQ data: Sall4 and Pdlim7 were enriched before differentiation, while Dpysl2 and H2afy were enriched after differentiation. The RT-PCR assays did not, in every case, confirm the observations from the microarray studies: Sall4, Dpysl2, and H2afy all changed at the mRNA level in correspondence with the observed protein changes, while Pdlim7 mRNA was unchanged before and after differentiation, confirming that Pdlim7 is subject to post-transcriptional regulation during RA mediated differentiation.

Racgap1 is necessary for self-renewal and is degraded in differentiating cells—A 5th case, Racgap1, was selected for further study. Antibodies were not available for western blots so we developed an MRM assay to confirm the iTRAQ data.

We first confirmed that Racgap1 mRNA is unchanged during differentiation by RT-PCR (Fig. 5A), and then used MRM to measure transitions from several peptides, described in Materials and Methods. As shown in Fig. 5B, MRM confirmed that Racgap1 is enriched at the protein level before the differentiation of ES cells.

We tested whether Racgap1 is necessary for ES cell self-renewal, using a library of 5 shRNAs from Open Biosystems (www.openbiosystems.com). We transfected three replicates of ES cells with the vectors, selected for transfected cells after 24 hours, passaged the cells onto 6-well plates, and allowed them to grow. After three days, plates were immunostained with rabbit anti-Oct4 (1:100, Santa Cruz), and DAB stained (Pierce cat #36000) to quantify Oct4+ colonies (Fig. 5C). Cells transfected with the most effective vectors (shRNAs #2, 3 and 5) resulted in the fewest colonies, significantly less than either the control (p=0.00077, 0.00037 and 0.00305 respectively), or the ineffective shRNA 1 (p=0.022, 0.024 and 0.048 respectively). A separate experiment using HEK293
Proteins associated with the undifferentiated state—Proteins enriched in undifferentiated cells in all three datasets included proteins known to be associated with pluripotency (Oct4, UTF1, Tcf3 and Dnmt3b). We did not detect Nanog or Ronin (Thap11) likely because of their low expression levels and lack of tryptic cleavage sites. To better understand and categorize the types of proteins enriched before differentiation in all three datasets, we used the Gene Functional Classification tool from DAVID to characterize the types of proteins enriched before differentiation (http://david.abcc.ncifcrf.gov/home.jsp; Fig. S3) (71, 72). The proteins fell into general classes: rRNA processing and ribosome assembly, general RNA binding, transcriptional regulators, ATP dependant helicases, and serine/threonine kinases (all involved in the cell cycle).

Proteins associated with RA mediated differentiation—In addition to RA-responsive genes such as Crabp1 and 2, we used the Gene Functional Classification tool from DAVID to identify classes of proteins enriched after RA-mediated differentiation in all three datasets. These clusters included proteins involved in the TCA cycle and general catabolism/metabolism, a family of dihydropyrimadinase-like proteins 2,3,4,5 involved in axon guidance and dendrite projection, cytoskeletal proteins, Ca^{2+} binding proteins, chromatin structural proteins of the H1/H5 family, and small GTPases involved in signal transduction.

Comparison between the proteome and transcriptome of pluripotent and differentiated cells—The results of this study suggest that the transcriptome is not a reliable predictor of the proteome. This is in agreement with several other proteome analyses, in multiple systems, using multiple techniques that have concluded that the proteome and transcriptome correlate only weakly, if at all (73-76). Our data confirm that the core transcription factors associated with pluripotency (Oct4, Sox2, Utf1, Rex1 and Sall4) change similarly at the protein and mRNA levels.

Putative cases of post-transcriptional regulation—A potential benefit of coupled proteome and transcriptome measurements is the ability to identify post-transcriptional regulation. We replicated an experimental design from Aiba et al. 2006, and by comparing our quantitative iTRAQ data with their quantitative Agilent microarray data, we were able to identify cases where our datasets disagreed. We chose four of these genes to confirm by RT-PCR and western blot. In all cases, western blots confirmed our proteome data, but in three of the cases, changes in protein levels were accompanied by changes in mRNA levels, contradicting the microarray-based measurements. The one exception was Pdlim7, which is indeed unchanged at the mRNA level during the differentiation of EC cells, but is down-regulated at the protein level during differentiation of EC cells. Pdlim7 is thought to be a protein scaffold that assembles signaling
molecules including Protein Kinase C (PKC) on the actin cytoskeleton to mediate signaling during development.

Sampath et al. (13) predicted that a 14-3-3 protein, Ywhab, would be more abundant in differentiated ES cells even though its mRNA levels were unchanged. Our iTRAQ data and the re-analyzed Aiba data confirm their prediction, making it likely that Ywhab is post-transcriptionally regulated.

Role of Racgap1 in ES cell self-renewal—Previous studies have determined that Racgap1 is necessary for division in some cell types. It is phosphorylated by Aurora kinase B and becomes a RhoA GAP rather than a Rac1 or Cdc42 GAP (77-80). There seem to be several other mechanisms to limit its activity toward Cdc42 and Rac (81). During mouse development, Racgap1 is expressed in the brain/ectoderm. Racgap1 is necessary for normal development of the pre-implantation embryo. Embryos that lack Racgap1 develop past the zygote stage but fail to form an inner cell mass. Conditional knockout of the gene in hematopoietic stem cells or B cells blocks proliferation and differentiation and leads to apoptosis, but this is apparently independent of the GAP activity of the protein (3, 82).

We observed that Racgap1 mRNA is expressed at the same level in undifferentiated and differentiated ES cells, but our protein measurements showed two-fold enrichment in undifferentiated cells. Knockdown of Racgap1 by shRNA caused a significant loss of Oct4+ colony formation. This phenotype may be due to the fact that the gene is necessary for proliferation of the inner cell mass. The mechanism by which Racgap1 is post-transcriptionally regulated is unknown. A search of the micro-RNA target database (found at www.microRNA.org) (83) identified 57 putative miRNA binding sites, but none of the 10 top miRNAs are expressed in the neural lineage. Therefore, it is unlikely that the predicted miRNAs are responsible for the post-transcriptional regulation of Racgap1.

Protein level changes in pathways during differentiation of pluripotent cells: Ribosome assembly complexes—The functional annotation clusters from undifferentiated cells include ribosome biogenesis and rRNA binding proteins. These proteins are enriched in undifferentiated ES and EC cells. Initially, we hypothesized that the cells are rapidly proliferating and thus need to produce many ribosomes. However, the Sampah et al. study determined that ES cells contain fewer ribosomes per cell than cells from EBs (as measured by rRNA content normalized to genomic DNA content) and the ribosomes that are present in ES cells are less loaded with mRNA than cells from EBs (13). Why, then do we see an enrichment of proteins associated with rRNA synthesis and regulation before differentiation?

Many of these proteins, apart from roles in rRNA synthesis and ribosome assembly, also regulate cell proliferation. Rps19bp1 (84), Gnl3 (85, 86), and Nolc1 (86) all inhibit p53 mediated growth arrest, while Bop1 inhibits growth of non-pluripotent 3T3 cells via the G1/S checkpoint which is absent in ES cells (87). Npm3 (88) and Noc3l (89-91) regulate chromatin structure/replication and are associated with several rapidly proliferating cell types. Nola3 (Nop10) is also a member of the telomerase complex, which is essential for extensive
proliferation (92). These alternate functions suggest that the ribosome protein enrichment is due partially to their recruitment to perform other tasks in proliferating cells in addition to increased ribosome synthesis activity.

**Protein level changes in pathways during differentiation of pluripotent cells:** *Metabolism*—The most striking biological process upregulated in differentiating ES and EC cells is an increase in oxidative glucose metabolism. This response might be counterintuitive, however ES cells, and many rapidly dividing cancer cell types, use glycolysis (93-95) as their primary energy source (Warburg effect), whereas oxidative phosphorylation of glucose via the TCA cycle is increasingly recognized to be associated with quiescent cell types such as neurons (96-98).

**Protein level changes in pathways during differentiation of pluripotent cells:** *Retinoic acid signaling*—ES and EC cells were aggregated and treated with retinoic acid, leading to an increase in the level of RA response proteins. ES and EC cells upregulated cellular retinoic acid proteins 1 and 2 (Crabp2, Crabp1) and retinol binding protein 1 (Rbp1). RA signaling leads to increases in Tgfb1 signaling (5, 99, 100). Although we did not see Tgfb1, we did see changes in its receptor (enriched in differentiated EC cells) and in downstream targets of Tgfb1 such as Smad5 (101) (enriched in the nucleus of differentiated EC cells). We also saw an increase in Smad5 in the nuclear fraction of differentiated cells, suggesting that the cells are indeed subject to increased Tgf beta signaling in response to retinoic acid.

**Protein level changes in pathways during differentiation of pluripotent cells:** *Signaling pathways activated during differentiation*—There were several signaling molecules enriched before differentiation. These include Aurora kinase A, Aurora kinase B, Bub1, and Polo-like kinase 1. These proteins are all involved in cell division (102-110), and likely reflect the rapid proliferation associated with pluripotent cells.

Differentiated cells are enriched with several types of signaling proteins including the 14-3-3 proteins Ywhab, Ywhaz, and Ywhae. These proteins bind phosphoproteins and mediate their translocation, activity, or degradation causing changes in development, apoptosis, and metabolism (111, 112). Interestingly, these proteins have been shown to interact with two proteins that are enriched before differentiation in all three datasets: Foxo1 and Tjp2, suggesting that the 14-3-3 proteins may be regulating these proteins during differentiation. Differentiated cells upregulated Gsk3b protein levels; this kinase is involved in neuronal differentiation (113) and is a negative regulator of the pro-pluripotency AKT signaling pathway (114, 115).

**Protein level changes in pathways during differentiation of pluripotent cells:** *Adhesion*—Another group of proteins that change during differentiation are proteins involved in cell adhesion. Undifferentiated ES and EC cells express higher levels of the Embryonic Cadherin (Cdh1) as well as tight junction associated protein Tjp 2 and junction plackoglobin Jup. During differentiation, cells begin to express N-cadherin (Cdh2) and downregulate Jup and Tjp2.

Interestingly, undifferentiated EC cells, but not ES cells, express three laminin subunits that make up the Laminin-511 complex. Recent publications
have identified exogenous Laminin-511 as a potent positive regulator of pluripotency in mouse and human ES cells (116). Mouse ES cells grown on a matrix of Laminin-511 no longer require external LIF in order to remain undifferentiated. EC cells may maintain their undifferentiated state in the absence of external signals such as LIF due to autocrine Laminin-511 signaling.

Protein level changes in pathways during differentiation of pluripotent cells: Transcriptional regulators and chromatin remodeling—Aggregation and RA treatment of ES and EC cells results in down-regulation of several transcription factors and chromatin modifying proteins. These include the well-known transcription factors Oct4, Sall4, UTF1, and Tcf3 plus the chromatin modifying protein Dnmt3b. Aggregation and RA treatment caused upregulation of histone family proteins H1/H5 and of transcription factors Hmgb3, Dach1, and Hoxb6.

Both UTF1 and H1/H5 have been implicated in epigenetic transcriptional regulation by remodeling and stabilizing condensed chromatin (117-120). Changes in these factors likely reflect changes in chromatin state during differentiation, resulting in loss of expression of undifferentiated state-specific proteins such as Oct4, Sall4, and Utf1 and induction of neural differentiation-specific proteins such as Crabp1, Crabp2, Nestin, Hoxb6, and N-cadherin.

REFERENCES


pluripotency in mouse and primate embryonic stem cells. *Oncogene* 25.


Fig. 1. **Overview of experimental design and sample collection.** A, ES cells of line 129/SvEv from ATCC cells were grown in undifferentiated conditions in the presence of LIF as described in the publication from Aiba et al. or differentiated by aggregation in bacteriological plates in the absence of 1 µM ATRA for 4 days, and then in the presence of 1 µM ATRA for 4 days as described in material and methods. Media was refreshed at differentiation day 2 and cells were collected at the indicated timepoints by treatment with versene and washed in 10 mM Hepes buffered saline (pH 7.4) and frozen at –80. Cells were dissolved in 2% rapigest, reduced, trypsinized and labeled with the iTRAQ mass-tagging reagent. B, P19 EC cells from ATCC cells were grown in undifferentiated conditions as described in the publication from Aiba et al. or differentiated by aggregation in bacteriological plates in the presence of 1 µM all-trans retinoic acid (ATRA) for 4 days as described in material and methods. Media was refreshed at differentiation day 2 and cells were collected at the indicated timepoints by treatment with versene and washed in 10 mM Hepes buffered saline (pH 7.4) and frozen at –80. Cells were dissolved in 2% rapigest, reduced, trypsinized and labeled with the iTRAQ mass-tagging reagent.
A

mESC (day 0)  

EB4 + RA4 (day 5)

Aggregate. Withdraw LIF

Culture in presence of 1μM ATRA

Day 0 1 2 3 4 5 6 7 8

• Collect  
• Stain  
• Process  

• Collect  
• Stain  
• Process
B

mEC (day 0)

P19 EB4RA4 (day 4)

Day
0 1 2 3 4

 Aggregate
Culture in presence of 1μM ATRA

- Collect
- Stain
- Process

- Collect
- Stain
- Process
Fig. S1. A, Immunofluorescence of ES cells before differentiation confirmed their undifferentiated state. Cells were positive for the pluripotency markers UTF1 and Oct4 and negative for both βIII Tubulin and NeuN. B, Immunofluorescence of ES cells after differentiation confirmed that the cells had largely differentiated. Cells were mainly negative for the pluripotency markers UTF1 and Oct4, however unlike in the case of the EC cells, a subset of the cells in the embryoid bodies maintained expression of Oct4 and UTF1. C, Immunofluorescence of EC cells before differentiation confirmed their undifferentiated state. Cells were positive for the pluripotency markers UTF1 and Oct4. EC cells were also positive for the neuroectodermal marker βIII Tubulin, likely reflecting their predisposition to differentiate into a neuroectodermal lineage. D, Immunofluorescence of EC cells after differentiation confirmed that the cells had differentiated, losing pluripotency markers UTF1 and Oct4, but not yet terminally differentiated into mature neurons as evidenced by the lack of NeuN staining.
Table I

*Filtering Criteria for auto-validation of database search results (spectrum score, spectrum SPI*%). SPI: Scored Peak Intensity*

<table>
<thead>
<tr>
<th>MS/MS spectrum cutoff scores (score, SPI%)</th>
<th>1+ peptide</th>
<th>2+ peptide</th>
<th>3+ peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;14, &gt;50%</td>
<td>&gt;12, &gt;50%</td>
<td>&gt;14, &gt;50%</td>
</tr>
</tbody>
</table>
Table II
Summary of whole proteome data sets, total validated spectra identified and iTRAQ labeling efficiency

<table>
<thead>
<tr>
<th></th>
<th>Total validated spectra</th>
<th>Un-modified</th>
<th>Fully iTRAQ modified</th>
<th>N-term iTRAQ modified only</th>
<th>Lysine iTRAQ modified only</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES cells</td>
<td>Spectra #</td>
<td>55,638</td>
<td>861</td>
<td>52,262</td>
<td>1927</td>
</tr>
<tr>
<td></td>
<td>Total Intensity</td>
<td>1.52E+10</td>
<td>1.17E+8</td>
<td>1.52E+10</td>
<td>4.7E+8</td>
</tr>
<tr>
<td>EC cells</td>
<td>Spectra #</td>
<td>94,399</td>
<td>7,529</td>
<td>60,811</td>
<td>23,528</td>
</tr>
<tr>
<td></td>
<td>Total Intensity</td>
<td>4.67E+10</td>
<td>2.42E+9</td>
<td>3.13E+10</td>
<td>1.2E+10</td>
</tr>
<tr>
<td>EC Nuclei</td>
<td>Spectra #</td>
<td>49,746</td>
<td>1,139</td>
<td>41,641</td>
<td>6,601</td>
</tr>
<tr>
<td></td>
<td>Total Intensity</td>
<td>7.32E+9</td>
<td>7.18E+7</td>
<td>6.41E+9</td>
<td>8.01E+8</td>
</tr>
</tbody>
</table>
Table III
*Summary of whole proteome data sets, replicates, peptides identified and total proteins groups identified*

<table>
<thead>
<tr>
<th></th>
<th>Biol. reps</th>
<th>Total proteins</th>
<th>Reverse database proteins</th>
<th>Forward database proteins</th>
<th>FDR</th>
<th>iTRAQ Labeling efficiency (N-term, lysine or both)</th>
<th>Proteins with total iTRAQ reporter intensity &gt;100</th>
</tr>
</thead>
<tbody>
<tr>
<td>129/SvEV ESCs</td>
<td>3</td>
<td>4053</td>
<td>23</td>
<td>4030</td>
<td>0.56%</td>
<td>98.5%</td>
<td>3566</td>
</tr>
<tr>
<td>EC cells</td>
<td>2</td>
<td>4501</td>
<td>9</td>
<td>4492</td>
<td>0.20%</td>
<td>92%</td>
<td>3801</td>
</tr>
<tr>
<td>EC Nuclei</td>
<td>3</td>
<td>4046</td>
<td>16</td>
<td>4030</td>
<td>0.39%</td>
<td>97.7%</td>
<td>3569</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>5523</td>
<td>34</td>
<td>5489</td>
<td>0.62%</td>
<td>95.2%</td>
<td>4986</td>
</tr>
</tbody>
</table>
Fig. 2. **Protein quantification.** Distribution of fold-change values for each dataset. Histogram shows a normal distribution of fold-changes of proteins before and after differentiation.
Distribution of protein ratios enriched before and after differentiation

Protein (group) Count

Enrichment before or after differentiation

129 Sv/EV ES cells
P19 nuclei
P19 whole cell

[Bar chart showing the distribution of protein ratios enriched before and after differentiation for 129 Sv/EV ES cells, P19 nuclei, and P19 whole cell.]
Fig. S2. *Proportion of single-peptide identifications in three datasets.* Single peptide hits make up a small percentage of overall data. Pie charts show the number of peptides identified that contribute to each identified protein group. In all cases single peptide identifications comprise <15% of the overall data. A, ES cell data. B, EC whole cell. C, EC nuclear.
Unique peptides identified per protein group 129 SV/EV ES cells

- 1 Peptide: 10%
- 2 Peptides: 12%
- 3 Peptides: 9%
- 4 Peptides: 9%
- >4 Peptides: 60%
B

Unique peptides identified per protein group P19 Whole cell

1 Peptide 14%  
2 Peptides 13%  
3 Peptides 10%  
4 Peptides 9%  
>4 Peptides 54%
Unique peptides identified per protein group P19 Nuclei

- 1 Peptide: 11%
- 2 Peptides: 11%
- 3 Peptides: 10%
- 4 Peptides: 10%
- > 4 Peptides: 58%
Fig. 3. **Overview of proteins enriched before and after differentiation using various criteria.**

A, To identify proteins generally enriched in pluripotent cells, we performed a student’s t-test on proteins >50% enriched in undifferentiated cells as described in the methods. Proteins with p value <0.05 and fold change of 50% were considered significantly enriched. Venn diagram in Fig. 3A represents the overlap of proteins that make the statistical and fold-change cutoff. 

B, To identify proteins generally enriched in cells differentiated with RA, we performed a student’s t-test on proteins >50% enriched in differentiated cells as described in the methods. Proteins with p value <0.05 and fold change of 50% were considered significantly enriched. Venn diagram in Fig. 3B represents the overlap of proteins that make the statistical and fold-change cutoff. 

C, To increase the number of proteins that we identify as generally enriched in pluripotent cells, we ignored statistical significance and used 50% enrichment as a fold-change cutoff. Venn diagram in Fig. 3C represents the overlap of proteins that make the fold-change cutoff. These proteins include known factors such as UTF1, Tcf3 and Oct4 (Pou5f1).

D, To increase the number of proteins that we identify as generally enriched in RA treated cells, we ignored statistical significance and used 50% enrichment as a fold-change cutoff. Venn diagram in Fig. 3D represents the overlap of proteins that make the fold-change cutoff. These proteins include known RA response factors such as Crab1 and Crab2 as well as the neural differentiation associated proteins Hoxb6 and N-cadherin.
Protein groups, >1 peptide, >50% enriched before differentiation, p<0.05
Protein groups, >1 peptide, >50% enriched after differentiation, p<0.05
Protein groups, >1 peptide identified 
>50% enriched before differentiation
Protein groups, >1 peptide identified >50% enriched after differentiation
Fig. 4. **Protein data to confirms some, but not all, transcript-level observations from Aiba et al.**

**A,** Scatterplot of 129/SvEv ES gene product ratios measured at the mRNA and protein levels. Only gene products observed at the protein and mRNA level were included in this graph. Pearson’s correlation coefficient was calculated as 0.362. Axes represent Log2 scale ratios with positive numbers representing proteins enriched before differentiation and negative numbers representing proteins enriched after differentiation. Points falling in the first or third quadrants represent cases of agreement between protein and mRNA data. Points falling in the second or fourth quadrants represent cases of disagreement and putative cases of post-transcriptional regulation. Points falling on axes represent proteins or transcripts whose levels are unchanged in the corresponding dataset.

**B,** Scatterplot of EC gene product ratios measured at the mRNA and protein levels. Only gene products observed at the protein and mRNA level were included in this graph. Pearson’s correlation coefficient was calculated as 0.409. Axes represent Log2 scale ratios with positive numbers representing proteins enriched before differentiation and negative numbers representing proteins enriched after differentiation. Points falling in the first or third quadrants represent cases of agreement between protein and mRNA data. Points falling in the second or fourth quadrants represent cases of disagreement and putative cases of post-transcriptional regulation. Points falling on axes represent proteins or transcripts whose levels are unchanged in the corresponding dataset.
Comparison of mRNA and protein measurements of 129 Sv/EV ES cells before and after differentiation

mRNA ratio before and after differentiation (Log2 scale)

R=0.362
Comparison of mRNA and protein measurements of whole P19 cells before and after differentiation

mRNA ratio before and after differentiation (Log2 scale)

R=0.409
Fig. S3. **GO terms enriched in all 3 datasets**. To identify consistently overlapping GO terms present in all three datasets, we graphed GO_BP_ALL terms identified with p<0.05 in all three datasets before and after differentiation. 
A, GO_BP_ALL terms enriched before differentiation in all three datasets. 
B, GO_BP_ALL terms enriched after differentiation in all three datasets.
GO BP ALL Terms significantly (p<0.05) enriched before differentiation in protein datasets
GO BP ALL Terms significantly \((p<0.05)\) enriched after RA-mediated differentiation in protein datasets.

Bar chart showing enrichment of different GO terms across ES, P19 whole cell, and P19 nuclei samples.
Fig. S4. **Comparison of GO annotations associated with pluripotent cells measured at the mRNA and protein levels.** As a general way to understand the classes of molecules enriched in protein and mRNA data from ES and EC cells, we used the DAVID analysis tool to identify GO terms associated with ES and EC cells before and after differentiation. 

A, GO biological process terms associated with ES and EC cells before and after differentiation. There is a large degree of overlap in GO terms between mRNA and protein data in all 4 conditions, however the overlap is markedly more between mRNA and protein data of undifferentiated ES cells, with only 2 GO terms from the protein data not occurring in the mRNA data. 

B, GO Molecular function terms associated with ES and EC cells before and after differentiation. There is a large degree of overlap in GO terms between mRNA and protein data in all 4 conditions, however the overlap is markedly more between mRNA and protein data of undifferentiated ES cells, with only 2 GO terms from the protein data not occurring in the mRNA data.
A

GO PANTHER_BP_ALL: Genes enriched

- Terms associated with proteins enriched before differentiation of ES cells
- Terms associated with proteins enriched after differentiation of ES cells
- Terms associated with proteins enriched before differentiation of EC cells
- Terms associated with proteins enriched after differentiation of EC cells

- GO annotation present in mRNA data only
- GO annotation present in both protein and mRNA data
- GO annotation present in protein data only
Terms associated with proteins enriched before differentiation of ES cells

Terms associated with proteins enriched after differentiation of ES cells

Terms associated with proteins enriched before differentiation of EC cells

Terms associated with proteins enriched after differentiation of EC cells

GO annotation present in mRNA data only

GO annotation present in both protein and mRNA data

GO annotation present in protein data only
Fig. S5. **Confirmation of cases of post-transcriptional regulation.**
Western blot and RT PCR were performed on EC cells differentiated as in the initial experiment and described in the methods. Undifferentiated and differentiated cells were collected and split for processing for western blotting (lysis in 1X SDS loading buffer with 1M β-ME) or RT-PCR (RNA purification in Trizol, DNase 1 treatment and Reverse transcription of 1 µg DNA using random hexamer primers and invitrogen superscript II reverse transcriptase). RT-pcr image colors are inverted for clarity. In all cases, the western blot results confirm the proteome measurements, however RT-PCR confirmed post-transcriptional regulation of Pdlim7 and Dpysl2 but fail to confirm post-transcriptional regulation of H2afy, Sall4 protein, where the mRNA results agree with the protein data.
<table>
<thead>
<tr>
<th>β-Actin</th>
<th>Utf1</th>
<th>Sall4</th>
<th>Pdlim7</th>
<th>H2afy</th>
<th>Dyps12</th>
<th>RT-PCR</th>
<th>Western</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://example.com/image1.png" alt="Image" /></td>
<td><img src="https://example.com/image2.png" alt="Image" /></td>
<td><img src="https://example.com/image3.png" alt="Image" /></td>
<td><img src="https://example.com/image4.png" alt="Image" /></td>
<td><img src="https://example.com/image5.png" alt="Image" /></td>
<td><img src="https://example.com/image6.png" alt="Image" /></td>
<td><img src="https://example.com/image7.png" alt="Image" /></td>
<td><img src="https://example.com/image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Fig. 5. **Racgap1 is a post-transcriptionally regulated protein necessary for mES cell self-renewal.** Another case of putative post-transcriptional regulation came from the protein Racgap1. Although the protein did not make it into the final list of putative cases of post-transcriptional regulation, the protein was an early candidate for a post-transcriptionally regulated protein necessary for self-renewal of ES cells. A, To confirm that Racgap1 protein levels are unchanged during differentiation, we performed RT-PCR analysis on ES cells and ES cells differentiated as in the initial experiment (aggregated in absence of LIF 4 days, grown in suspension in the presence of 1 µM ATRA for 4 more days). B, We employed multiple reaction monitoring (MRM) and confirmed that Racgap1 is significantly (p=0.00167) enriched in ES cells before differentiation. RT-PCR image colors are inverted for clarity. C, Knockdown of Racgap1 results in significant reduction of Oct4+ ES colonies. Experiments were performed in triplicate and the student’s t-test was used to calculate significance using values from control cells and cells transfected with an ineffective shRNA (#1). p-values are listed in the table under the chart.
Racgap1 mRNA is unchanged during differentiation of ESCs into EBs

<table>
<thead>
<tr>
<th></th>
<th>ESC</th>
<th>EB8 RA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racgap1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
B

Racgap1 protein levels measured by MRM in mouse ESCs and EBs

![Bar chart showing concentration of Racgap1 protein in ESCs and EBs](image)

- **ES**
  - Concentration: 4.5 mols/μg total sample
  - p-value: 0.00167
- **EB**
  - Concentration: 2.75 mols/μg total sample
  - ***** p-value: 0.00001

Cell type
p-values resulting from student's t-test performed against:

<table>
<thead>
<tr>
<th>Construct</th>
<th>Control</th>
<th>shRNA1</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>0.1547</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1547</td>
</tr>
<tr>
<td></td>
<td>0.1547</td>
<td>0.0217 *</td>
</tr>
</tbody>
</table>

**Control** | N/A | 0.1547 | 0.0007 *** | 0.0004 *** | 0.0032 *** | 0.0030 *** | 0.0004 ***

**shRNA1** | 0.1547 | N/A | 0.0217 * | 0.0238 * | 0.0742 | 0.0484 * | 0.0168 *
Supplemental Tables:

Table S1. **Proteins enriched in all three datasets.** Proteins enriched before differentiation, defined as >50% enriched (pink) and after differentiation defined as >50% enriched (green) with student’s T-test p-values < 0.05.

Table S2. **Examples of pluripotency and RA-response proteins reproducibly enriched but that do not make the p-value cutoff in all three datasets.** The student’s T-test excludes many known pluripotency and RA associated proteins that are reproducibly enriched in all three datasets. The table gives several examples. The high p values are consistently due to variability of the amount that the ratios exceed the 50% fold change cutoff.

Table S3. **Proteins enriched before differentiation in all three datasets.** Table represents those proteins reproducibly enriched before and after differentiation in all three datasets. We propose these proteins as protein fingerprints of the undifferentiated and the RA-differentiated states of pluripotent cells.

Table S4. **Proteins and mRNA markers of the undifferentiated and RA-mediated differentiated state.** Protein and mRNA ratios whose fold-changes agree.

Table S5. **Cases of putative post-transcriptional regulation.** Proteins reproducibly enriched before or after differentiation in all three protein datasets, but in neither mRNA dataset.

Table S6. **Proteins that were identified and quantified in this study.** Column A to Column AF are the normalized iTRAQ intensities from each protein. Column AG shows the protein group numbers. Proteins with the same group number share the same set or subset of peptides. For iTRAQ quantitation, peptides that are shared between protein groups were removed and only peptides that are unique to each group were used.