Supporting Information

Swaney *et al.* 10.1073/pnas.0811964106

SI Text

Cellular Localization and Processes. Of the 4,339 proteins identified in this study, about half (2,762) had GO annotations for cellular localization. We submitted these genes to cellular localization analysis using FatiGO+ [www.babelomics.org/ (Fig. S1) (1)]. Over 40% of these genes were categorized as nuclear proteins, nearly 15% greater than the percentage of proteins within the IPI database that were localized to the nucleus. This result correlates well with other large-scale phosphoproteomics reports and suggests an enrichment of phosphorylation within the nucleus (2). Another 37.5% were localized to the cytoplasm, and 28.3% to membranes. In addition, a relatively large percentage (8.8%) of the phosphorylation events occurred on actin cytoskeletal proteins, which is not surprising because those make up a large percentage of the cytoplasmic content of the cell. These genes were also analyzed for biological process and molecular function. It was found that 27.2% of these genes were involved in the regulation of DNA-dependent transcription and that 22.8% are specifically involved in DNA binding. The large percentage of nuclear proteins identified here shows promise for future studies examining phosphorylation events critical to the regulation of genes that impact human ES cell state.

Cell Culture and Protein Harvesting. Human ES cells (line H1) maintained in feeder-independent media were expanded to 5×10^8 cells (3). Cells were then individualized with 0.05% trypsin-ETDA, the trypsin was neutralized with 10% FBS in DMEM, and, finally, washed twice with ice-cold PBS and pelleted. The pelleted cells were resuspended in 10 mL of lysis buffer containing 8 M urea, 30 mM NaCl, 40 mM Tris (pH 8), 6.4 mM sodium pyrophosphate, 1 mM sodium orthovanadate, complete mini ETDA-free protease inhibitor (Roche Diagnostics), and phosSTOP phosphatase inhibitor (Roche Diagnostics). Lysis was achieved by sonication, and the solution was spun at $30{,}000 \times g$ at 4 °C for 10 min. The supernatant was then stored at −80 °C.

Digestion. An aliquot of 10 mg of human ES cell protein was incubated in 2.5 mM DTT for 35 min at 50 °C to reduced cysteine residues, followed by alkylation via incubation for 30 min with 7 mM iodoacetamide at room temperature. The alkylation was capped by incubation in 2.5 mM DTT for 15 min at room temperature. The sample was then diluted to a final concentration of 1.5 M urea, 25 mM Tris, and 1 mM $CaCl_2$. Trypsin Gold (100 μg; Promega) was added and the sample digested for 45 min at 37 °C. After cooling to room temperature, the sample was quenched by the addition of TFA to 0.5%, desalted on a 500 mg tC_{18} SepPak cartridge (Waters), and the eluate lyophilized.

Fractionation. Digested peptides were dissolved in 1 mL of SCX buffer A [5 mM KH_2PO_4, 30% acetonitrile (pH 2.65)]. This was split in half, and each half was separately fractionated. For each fractionation procedure, 500 μL of the digested peptide mixture was injected onto a polysulfoethyl aspartamide column (9.4 × 200 mm; PolyLC) attached to a Surveyor LC quaternary pump (Thermo Electron) running at 3.0 mL/min. Peptides were detected by using a PDA detector (Thermo Electron). The eluate was collected in 4-min intervals over the following gradient: 2 min of isocratic buffer A, followed by a linear gradient of 0–15% buffer B from 2 to 5 min [5 mM KH_2PO_4, 30% acetonitrile, 350 mM KCl (pH 2.65)], followed by a linear gradient of 15–100% buffer B from 5 min to 35 min. Buffer B was held at 100% for 10 min, after which there was a 7-min transition from buffer B to buffer C [50 mM KH_2PO_4, 500 mM KCl (pH >7.5)]. Buffer C and buffer D (nanopure water) were used to wash the column. After pooling fractions from the 2 replicates, all fractions were lyophilized and desalted on 100-mg tC_{18} SepPak cartridges before phosphopeptide enrichment (Waters).

Phosphopeptide Enrichment. A solution of 2 M methanolic acid was generated by the dropwise addition of acetyl chloride to anhydrous methanol (Grace). Between 1 and 4 mL of methanolic acid was then added to desalted fractions and then sonicated in 15-min intervals over the course of 2 h. Fractions were then lyophilized to dryness and then resuspended in 300 μL of a reconstitution buffer (RB). The RB consisted of 1:1:1 mixture of methanol/acetonitrile/water in 0.01% acetic acid and adjusted to pH 3.4. Each reconstituted fraction was split in half and separately enriched for phosphopeptides via immobilized metal affinity chromatography (IMAC) (4, 5). A fused silica column (560 μm I.D. × 10 cm) packed with POROS MC-20 resin (Applied Biosystems) was stripped with 100 μL of 40 mM ETDA over the course of 5 min. After washing with 100 μL of nanopure water over the course of 5 min, the column was charged with a solution of 100 mM $FeCl_3$ at a flow rate of 20 μL/min for 7 min. The RB was then passed through the column at the same flow rate for 5 min to reequilibrate the IMAC column before sample loading. The sample was then loaded at a flow rate of ≈2 μL/min, followed by a wash with the RB at the same flow rate for 5 min, and then for 5 min at 20 μL/min. Next, a wash of 20 μL of 42 mM NaCl in 50% acetonitrile, 15% methanol, and 35% water, all in 0.01% acetic acid was passed over the column over the course of 1 min, followed by a 20 μL of 50 mM NaCl in 75% acetonitrile, 10% methanol, and 15% water, all in 0.01% acetic acid over the course of the next minute. After equilibration of the column with 40 μL 0.02% formic acid at 20 μL/min, peptides were eluted from the IMAC column using 95 μL of 50 mM phosphate buffer (pH 8.0). The eluates of each half of a given fraction were quickly acidified to a final acid concentration of 0.2% formic acid and immediately pooled.

nHPLC. Phosphopeptide-enriched SCX fractions were loaded onto a precolumn via a Waters nanoACQUITY UPLC system and separated on a 12-cm C_{18} analytical column. All columns were self-prepared as described in ref. 6. Peptides were separated via the following gradient: 1.4% to 4.2% acetonitrile in 0.2% formic acid in 15 min, then up to 31.5% acetonitrile from 15 to 80 min.

Mass Spectrometry. All experiments were performed on a hybrid linear ion trap–orbitrap mass spectrometer (Thermo Fischer Scientific) that had been modified to allow for ETD. The details of this modification are described in refs. 7 and 8. nHPLC eluates were directly sampled via an integrated electrospray emitter operating a 2.0 kV (6). Each experiment consisted of MS^1 analysis in the orbitrap mass analyzer followed by 6 data-dependent MS/MS events with mass analysis in the ion trap. The type of dissociation in each MS/MS event was either CAD or ETD. SCX fractions, with the exception of fraction 2, were sampled 3 times via CAD, twice via ETD, and 3 times via a decision tree-based acquisition, as described in ref. 9. SCX fraction 2 was analyzed in the same manner with the exception that it was analyzed 1 fewer time via both ETD and CAD. For all experiments QIT MS/MS AGC targets of 10,000 ions were used, precursors were dynamically excluded for 30 sec, and only

peptides with assigned charge states of 2 or greater were selected for MS/MS interrogation.

Database Searching. OMSSA (Open Mass Spectrometry Search Algorithm) was used to search all tandem MS spectra against a concatenated version of the human IPI database (version 3.37) (10). OMSSA parameters were set to include the following static modifications: +57 Da on cysteine residues (carbamidomethylation), +14 Da on aspartic acid, glutamic acid, and the c-terminus. Differential modification of +16 Da on methionine residues (oxidation), +80 Da (phosphorylation) on serine, threonine, or tyrosine were also used. A fragment ion mass tolerance of 0.5 Da and a precursor mass tolerance of 4.5 Da were used for searches. All identifications were trimmed using an in-house program to a 1% FDR through filtering of peptides by identification score, precursor *m/z* accuracy (50 ppm maximum), also the precursor *m/z* was adjusted for the possible selection of the ^{13}C isotopic peak.

Phosphosite Localization. Datasets were analyzed with a custom application written in-house, developed in C# with Microsoft Visual Studio 2005 using the Microsoft .NET Framework 2.0. For each phosphopeptide identified by OMSSA, the software generated all possible phosphopeptide positional isomers with the correct number of phosphorylations to yield the mass of the precursor. For each form, theoretical fragments were calculated and compared to experimental tandem mass spectra from the Thermo Scientific LC-MS/MS .raw data file, using the XRawfile COM library (XRawfile2.dll) for data access. For ETD spectra, singly charged *c*- or *z*•-type fragment ions were considered as well as doubly charged fragment ions of 500 *m/z* or greater. For CAD, all *b*- and *y*-type fragment ions of charge state less than the precursor were considered. A theoretical fragment was considered matched if an experimental peak was within ±0.3 *m/z* (CAD) or ±0.4 *m/z* (ETD) of the calculated fragment *m/z*. A relative ion intensity threshold 0.3% and 0.4% of the base peak was used for CAD and ETD, respectively. The software determined the number of matching fragments, the number of bonds cleaved, and whether or not bonds were cleaved in the −3 to +3 positions relative to each phosphorylation. The software also determined whether or not the phosphopeptide positional isomer identified by OMSSA was the best form (if it had the highest number of matching fragments out of all positional isomers) and if the phosphorylations were localized in the best form (if there are not multiple forms with the same number of highest matching fragments).

Motif Analysis. Unique peptide identifications via all dissociation methods were submitted to the Motif-X algorithm (11). The human IPI database was used as the background. All submitted peptide sequences were centered around the site of phosphorylation and extended to 6 aa on either side using the human IPI database. All motifs were extracted at a significance of $P < 10^{-6}$ and a minimum frequency of 100, 30, and 8 were used for phosphoserine, threonine, and tyrosine motifs, respectively. A second motif extraction was also performed in which all ETD phosphopeptide identifications were extracted from a background of 18,869 human phosphopeptides identified from CAD analyses that were downloaded from the PHOSIDA database (www.phosida.com). These motifs were extracted at a significance of $P < 10^{-6}$, and a minimum frequency of 80 phosphoserine or 20 phosphothreonine motifs.

1. Al-Shahrour F, *et al.* (2006) BABELOMICS: A systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* 34(suppl 42):W472–W476.
2. Olsen JV, *et al.* (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635–648.
3. Ludwig TE, *et al.* (2006) Feeder-independent culture of human embryonic stem cells. *Nat Methods* 3:867–867.
4. Ficarro SB, *et al.* (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 20:301–305.
5. Ndassa YM, Orsi C, Marto JA, Chen S, Ross MM (2006) Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J Proteome Res* 5:2789–2799.
6. Martin SE, Shabanowitz J, Hunt DF, Marto JA (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 72:4266–4274.
7. McAlister GC, *et al.* (2008) A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap–orbitrap mass spectrometer. *J Proteome Res* 7:3127–3136.
8. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ (2007) Implementation of electron-transfer dissociation on a hybrid linear ion trap–orbitrap mass spectrometer. *Anal Chem* 79:3525–3534.
9. Swaney DL, McAlister GC, Coon JJ (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 5:959–964.
10. Geer LY, *et al.* (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3:958–964.
11. Schwartz D, Gygi SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 23:1391–1398.

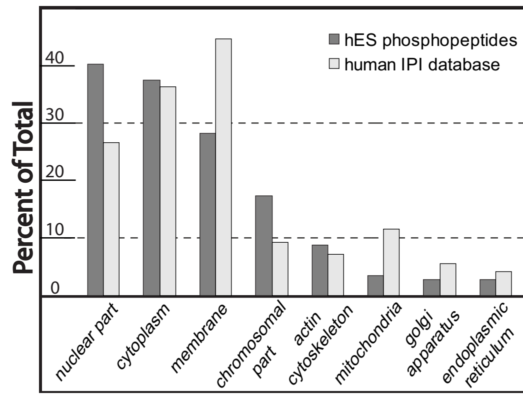

Fig. S1. Gene ontology analysis of the identified phosphoproteins.

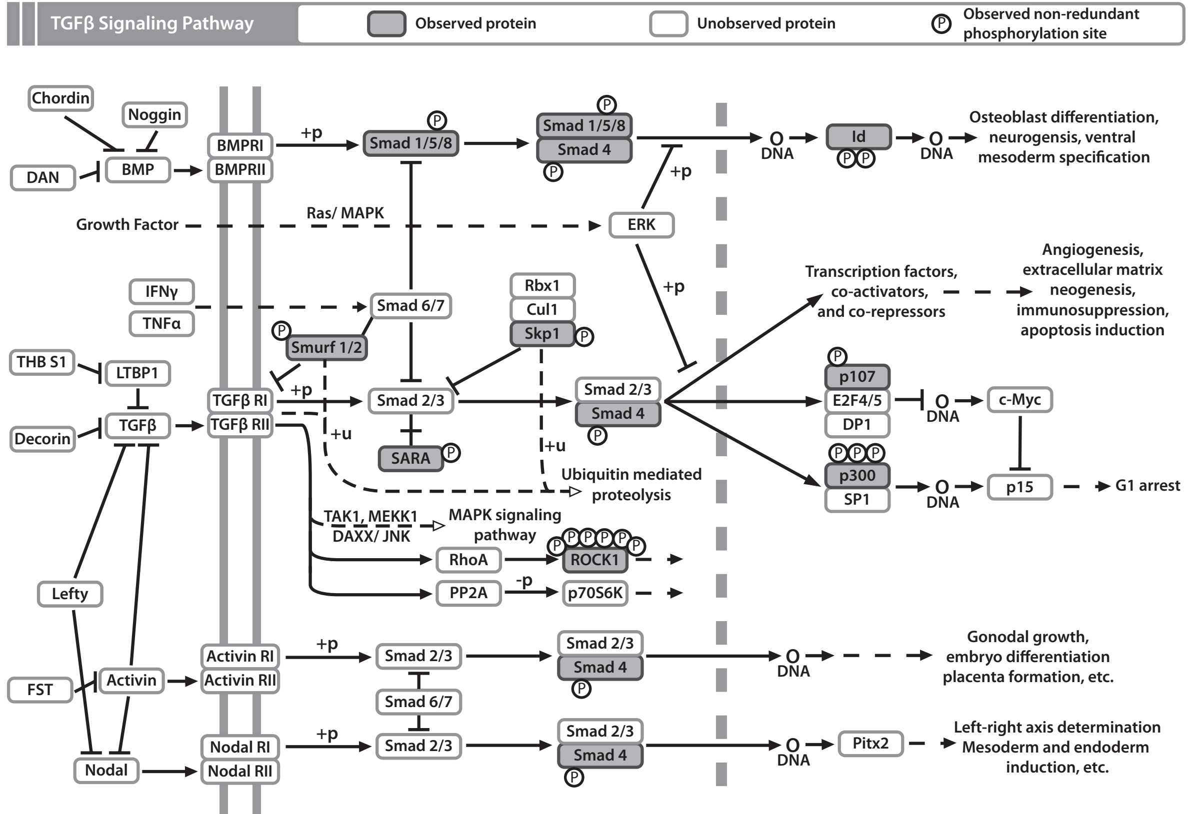

Fig. S2. Identification of phosphorylation sites residing on proteins of the TGFβ signaling pathway. This pathway is suspected to be influential in the maintenance of pluripotency in human ES cells due to an interaction with known pluripotency factors such as OCT4, SOX2, and NANOG. The number of nonredundant phosphorylation sites are displayed on each protein.

Table S1. Motifs extracted from all CAD and ETD data while using the human IPI database as a background

Motif	Motif score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
. . . RR.s.	32	215	3,571	5,073	1,081,450	12.83
. . . R..s.	16	355	2,130	41,570	991,289	3.97
. . . . . . s..E. . .	16	125	674	50,722	672,225	2.46
.R.R..s.	32	113	2,784	3,904	1,059,734	11.02
. . . . R.s.	16	270	1,391	45,660	842,015	3.58
. . . . . . sP.K. . .	32	238	4,008	3,250	1,089,478	19.91
. . . . . . sP.R. . .	32	144	3,356	4,238	1,076,377	10.9
. . . . . . tP.K. . .	32	85	1,050	2,124	687,043	26.19
. . . . . . tP.R. . .	30.81	48	906	2415	682,196	14.97
. . . . R.t.	14.42	83	615	32,893	634,123	2.6
. . . . . . sP.	16	459	2,671	61,173	1,055,830	2.97
. . . . . . tP.	16	140	755	39,050	673,173	3.2
. . . . . . t. . . . S.	6.06	65	376	49,490	533,004	1.86
. . . RS.s.	32	215	3,212	6,942	1,072,139	10.34
. . . . . . sP. . . . K	32	122	2,997	2,770	1,065,197	15.65
. . . . . . sP. . . K.	32	91	2,875	2,693	1,062,427	12.49
. . . . SPt.	32	69	824	4,759	677,932	11.93
. . . . . . sP. . . R.	32	199	3,770	4,778	1,086,228	12
..RR..s.	29.78	82	2,212	3,368	994,657	10.95
. . . . . . tP. . . . K	27.79	34	858	1,849	679,781	14.57
. . . K..s.	16	152	953	43,148	752,956	2.78
. . . . . . s..R. . .	16	168	1,121	43,399	796,355	2.75
. . . . . . s..K. . .	16	127	801	37,583	709,808	2.99
. . . . . . tSP. . . .	22.18	31	532	4,411	601,230	7.94
R. . . . . s.	16	89	549	33,832	621,503	2.98
. . . . . . t. . . . K.	9.39	58	434	30,690	563,694	2.45
R. . . . . t.	10.37	67	501	33,125	596,819	2.41
. . . R..tP.	32	59	965	2,723	684,919	15.38
. . . R..sP.	32	320	4,328	5,433	1,094,911	14.9
. . . . R.s.S. . . .	25.68	80	1,775	5,652	949,719	7.57
. . . . S.s.	16	304	1,695	102,052	944,067	1.66

A mimimum of 100 Ser and 30 Thr phosphorylation sites and a significance of $P > 10{-}6$ was required for motif extraction.

Table S2. Motifs extracted from all ETD data while using the CAD-identified peptides from the PHOSIDA database as a background extraction

Motif	Motif score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
. . . RR.s.	28.92	178	3,078	237	17,588	4.29
. . . R..s.	16	448	2,718	1,310	17,023	2.14
.R.R..s.	32	256	3,801	216	18,317	5.71
. . . . R.s.	16	350	1,775	995	13,322	2.64
. . . . . . sP.K. . .	32	244	4,045	370	18,687	3.05
. . . . . . sP.	16	495	2,270	2,391	15,713	1.43
. . . . . . sP. . . . K	32	122	3,200	145	17,733	4.66
. . . . . . sP. . . K.	32	161	3,545	246	18,101	3.34
. . . . . . sP. . . R.	32	283	4,328	457	19,144	2.74
. . . . . . tP. . . . K	30	46	678	46	6,165	9.09
. . . K..s.	16	190	1,207	810	11,536	2.24
. . . . . . s..R. . .	16	218	1,425	791	12,327	2.38
. . . . . . s..K. . .	16	138	858	502	10,029	3.21
R. . . . . s.	16	159	1,017	697	10,726	2.41
R. . . . . t.	11.4	72	525	334	5,957	2.45
. . . R..t.	8.54	67	388	425	5,285	2.15
.L.R..s.	25.71	94	2,812	126	17,149	4.55
R..R..s.	30.61	101	3,301	90	17,823	6.06
. . . . . . sP. . . . R	23.88	88	2,900	202	17,351	2.61
. . . . . . t. . . K..	8.25	37	283	206	4,646	2.95
. . . . . . tP. . . . R	26.06	41	590	67	6,063	6.29
. . . . . . tP..K..	26.87	42	632	56	6,119	7.26
..K. . . tP.K. . .	48	24	702	2	6,167	105.42
K. . . . . s.	16	104	720	504	9,527	2.73
. . . . . . t..K. . .	9.98	65	453	338	5,623	2.39
..K. . . tP.	20.02	24	549	39	5,996	6.72
.K. . . . t.	7.36	38	321	214	4,860	2.69
. . . R..s.K. . . .	32	83	3,384	32	17,855	13.69

A mimimum of 80 Ser and 20 Thr phosphorylation sites and a significance of $P > 10-6$ was required for motif extraction.

Table S3. A list of several cellular pathways and the number of genes identified in pathway

Pathway	No. of genes identified
Regulation of actin cytoskeleton	49
MAPK signaling pathway	40
Insulin signaling pathway	42
Tight junction	35
Cell cycle	37
Focal adhesion	37
Wnt signaling pathway	28
ErbB signaling pathway	24
Adherens junction	22
Notch signaling pathway	15
mTOR signaling pathway	13
TGF-β signaling pathway	12
Hedgehog signaling pathway	9

Table S4. Phosphorylation sites identified on pluripotency regulating transcription factors in human ES cells

Transcription factor	Peptide	Dissociation method
Oct4	RTsIENRVR	ETD
Sox2	sEASSsPPVVTSSSHSR	CAD
Sox2	SEASssPPVVTSSSHSR	CAD
Sox2	SEAsSSPPVVTSSSHSR	CAD
Sox2	SEASSsPPVVTSSSHSR	ETD
Sox2	SEASsSPPVVTSSSHSR	CAD

Lower-case lettering indicates the phosphorylated residue.