Stem Cell Reports, Volume *2*

Supplemental Information

NANOG is Multiply Phosphorylated and Directly Modified by ERK2 and CDK1 In Vitro

Justin Brumbaugh, Jason D. Russell, Pengzhi Yu, Michael S. Westphall, Joshua J. Coon, and James A. Thomson

Inventory of supplemental information

Supplemental Figure Legends

Figure S1. NANOG is not amenable to mass spectrometery via traditional tryptic methods, related to Figure 1

The primary sequence for human NANOG is shown. Trypsin sites are highlighted in red. Note that cleavage sites are clustered rather than spread throughout the protein, largely generating peptides that are either two short or too long for detection by mass spectrometry.

Figure 2. Representative spectra for all NANOG phosphorylation sites reported in this study, related to Figure 1

Neutral loss (asterisk) is annotated. To view individual spectra in more detail, please visit the following website to download a PDF: http://scor.chem.wisc.edu/data/Brumbaugh_et_al_Supplemental_Fig_2_individual.pdf

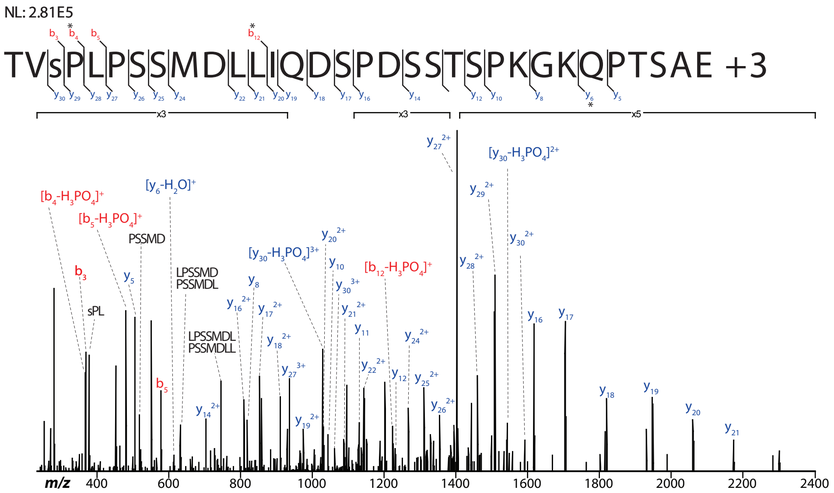
Figure S3. Proof of concept for normalization of NANOG quantitation using non-phosphorylated *E. coli* peptides and quantitation for additional NANOG phosphorylation sites identified by MAKS, related to Figure 3

(A) One microgram of NANOG protein preparation was diluted at the indicated ratios to simulate pipetting error and sample loss. Reporter ion intensities were corrected based on non-phosphorylated *E. coli* peptides. Raw data is shown for comparison. The dashed line represents an ideal correction. Note that this technique normalizes a 20-fold deviation to within 30% of the intended value.

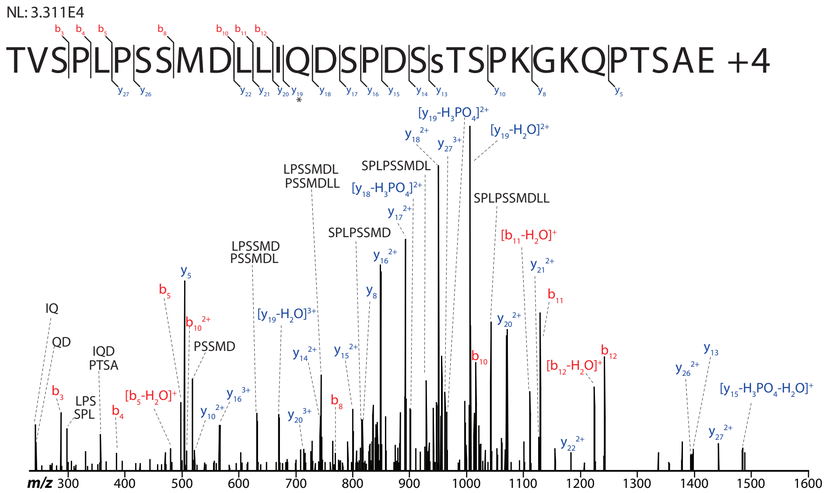
(B) Quantitation for Ser65 and Ser23 on human NANOG. Bar graphs represent the normalized abundance of at least three PSMs for each peptide across three independent experiments. Error bars represent standard deviation calculated from at least three PSMs across three independent experiments. For reference, the dashed line indicates the level expected if the contribution for all channels were equal.

Supplemental Figure S1; Brumbaugh et al.

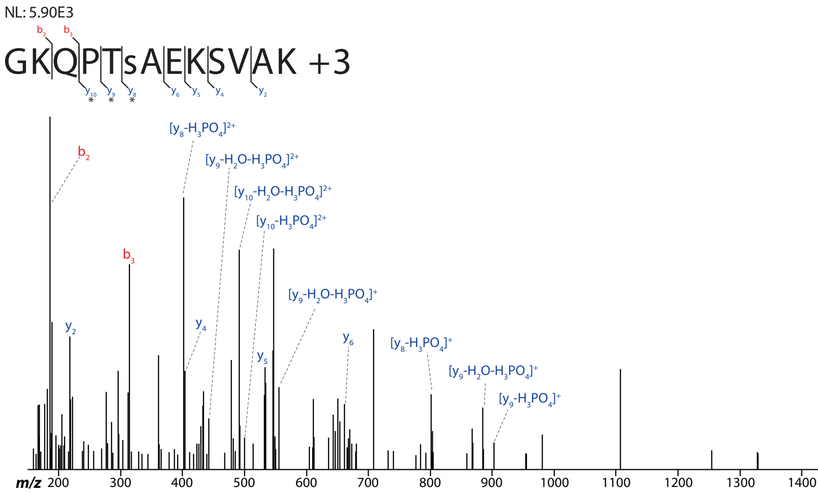
NL: 2.81E5



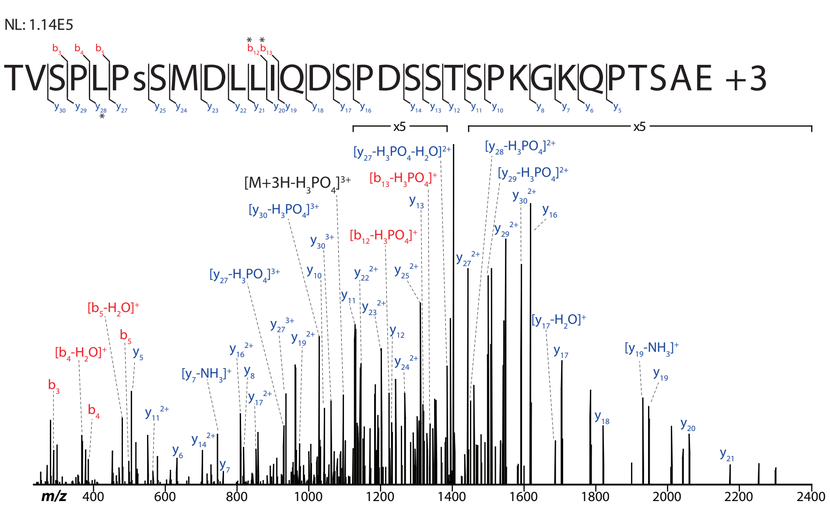
NL: 3.311E4



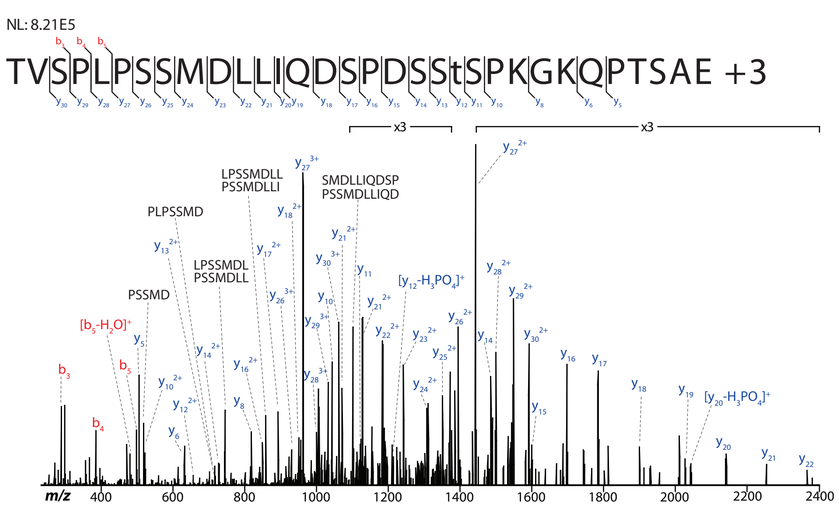
NL: 5.90E3



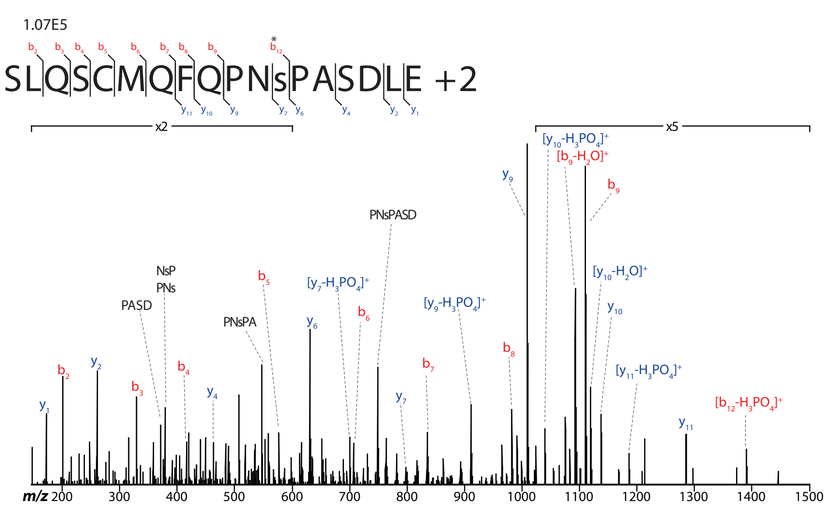
NL: 1.14E5



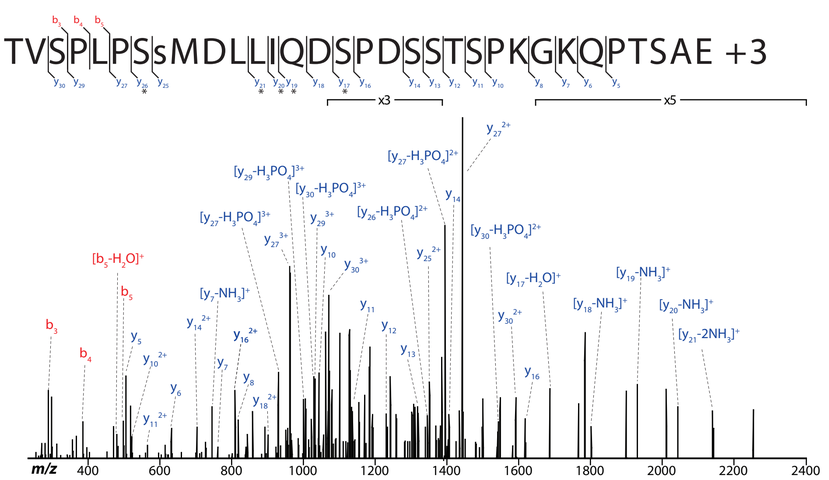
NL: 8.21E5



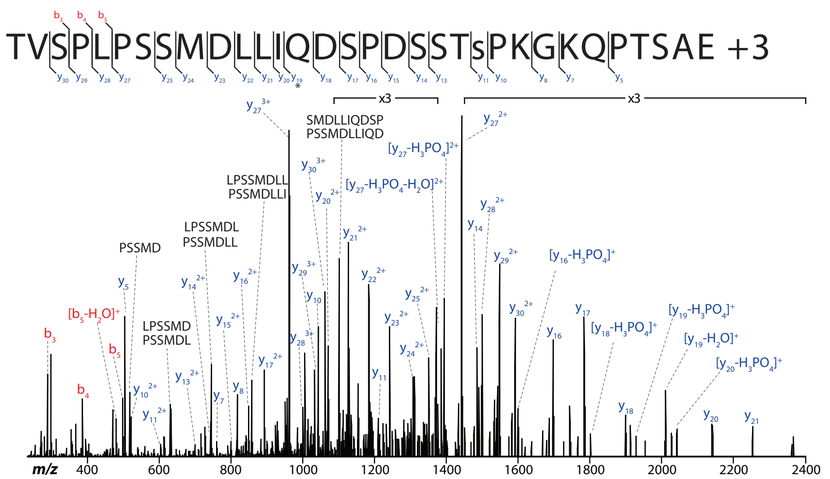
1.07E5



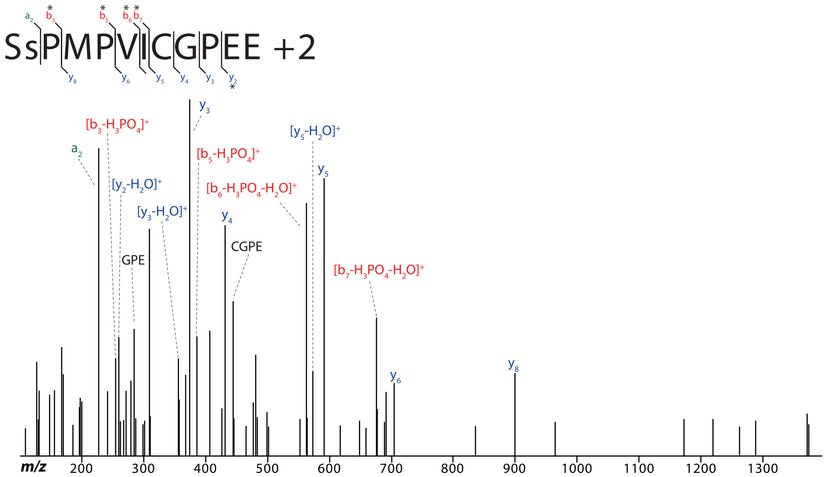
NL: 6.36E4



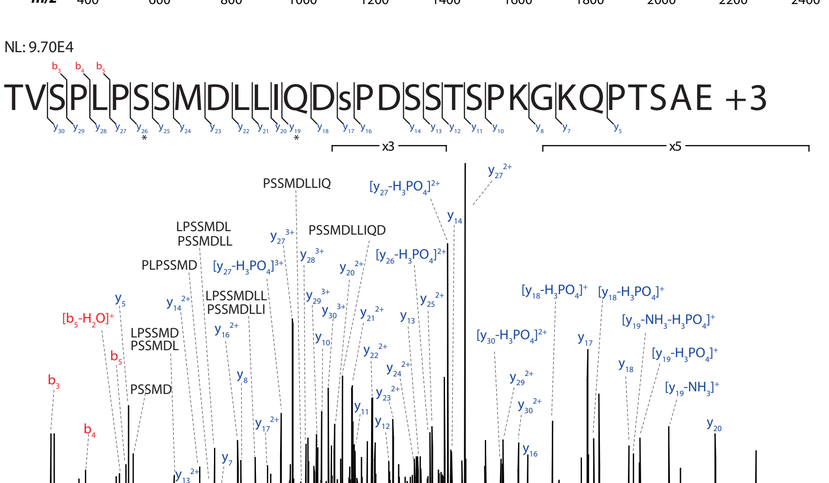
NL: 2.72E5



NL: 3.17E3



NL: 9.70E4



NL: 1.65E4

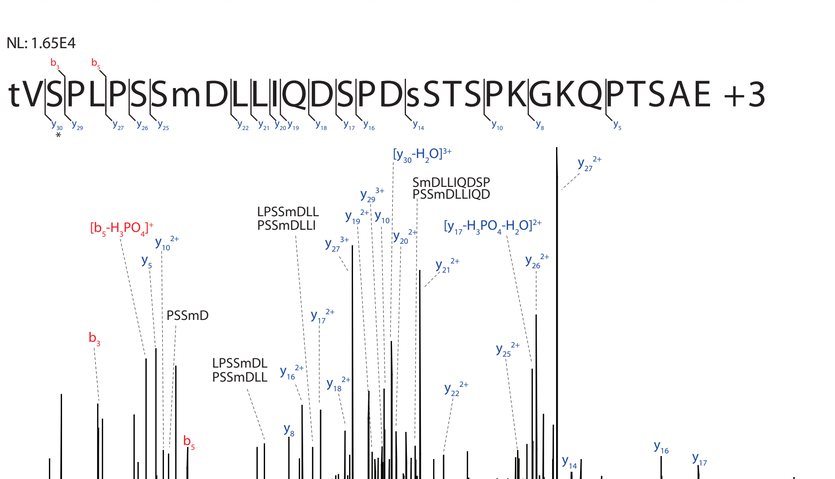


Figure S2: Brumbaugh et al.

Supplemental Figure S3; Brumbaugh et al.

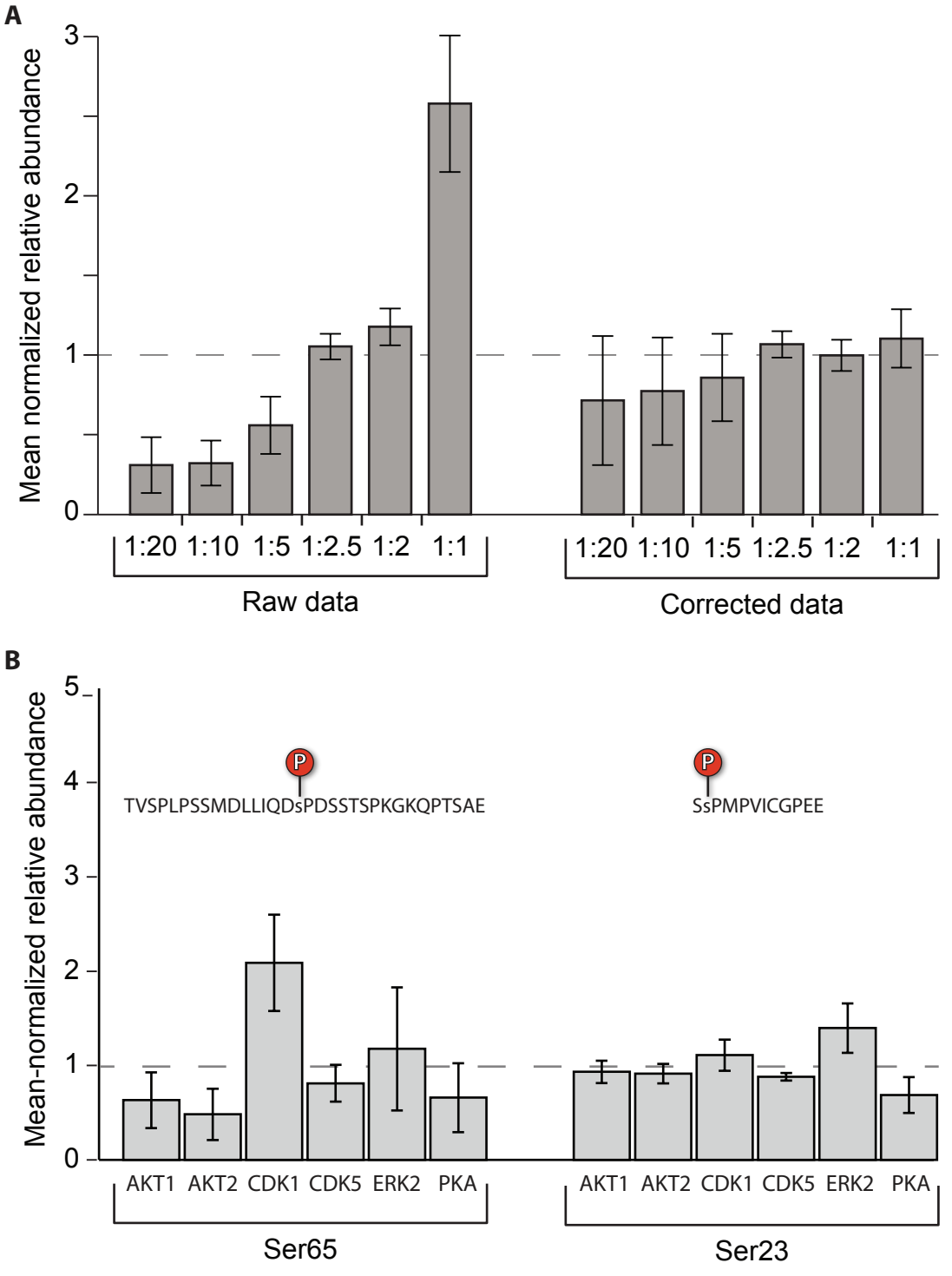


Table S1.

Localized		Non-localized	
Modification	Peptide spectral match	Modification	Peptide spectral match
phosphorylation of S:258	SLQSCMQFQPNsPASDLE	phosphorylation with neutral loss on S:22 or S23	SsPMPVICGPEE
phosphorylation of S:52	TVsPLPSSMDLLIQDSPDSSTSPKGKQPTSAE	phosphorylation of T:70 or s71	TVSPLPSSMDLLIQDSPDSStSPKGKQPTSAE
phosphorylation of S:56	TVSPLPsSMDLLIQDSPDSSTSPKGKQPTSAE	phosphorylation of S:69, t70, or s71	TVSPLPSSMDLLIQDSPDSsTSPKGKQPTSAE
phosphorylation of S:57	TVSPLPSsMDLLIQDSPDSSTSPKGKQPTSAE	phosphorylation of T:50, oxidation of M:9, and phosphorylation of S:68, S:69, T:70, or S:71	tVSPLPSSmDLLIQDSPDsSTSPKGKQPTSAE
phosphorylation of S:65	TVSPLPSSMDLLIQDsPDSSTSPKGKQPTSAE		
phosphorylation of S:71	TVSPLPSSMDLLIQDSPDSSTsPKGKQPTSAE		
phosphorylation of T:70	TVSPLPSSMDLLIQDSPDSStSPKGKQPTSAE		
phosphorylation with neutral loss on S:79	GKQPTsAEKSVAK		
phosphorylation of T:50, oxidation of M:9, and phosphorylation of S:68, S:69, T:70, or S:71	tVSPLPSSmDLLIQDSPDsSTSPKGKQPTSAE		

Endogenous NANOG phosphopeptide identifications with relevant details, related to Figure 1

Table S2. Phosphopeptides (< 1% false discovery rate) from control substrates demonstrate that recombinant kinases are active, related to Figure 3.

Standard substrates, MBP and Histone H1, were subjected to separate kinase assays. Only unique identifications are shown.

(Please see corresponding Xcel file)

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Purification of endogenous, human NANOG

Approximately 10^8 human ES cells (line H1) were grown using a feeder-independent system (Chen et al., 2011), harvested using TrypLE (Gibco), and washed three times in PBS (Gibco). High detergent lysis buffer [50mM Tris-HCl, pH8, 150mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1mM EDTA, 1×protease inhibitor mixture (Roche Diagnostics), 1×PhosSTOP phosphatase inhibitors (Roche Diagnostics), 0.01U/μL benzonase (Novagen)] was applied to the pelleted cells. The resulting lysate was incubated at 4° C for 10 minutes with shaking. Cell debris was removed by centrifugation at 20,000×*g* and anti-NANOG antibody (1:100 dilution, Cell Signaling) was incubated with the cleared lysate for one hour at 4° C. Protein G beads (GE Healthcare) were added for an additional hour at 4° C. The beads were then washed twice with lysis buffer and four times with 50mM Tris-HCl, pH8, and 150mM NaCl. Protein was reduced by adding DTT to a final concentration of 5 mM and incubated for 30 minutes at 37º C. Samples were cooled to 23º C and alkylated using iodoacetamide at a final concentration of 10 mM for 30 minutes in the dark. This reaction was quenched by bringing each sample to 10 mM DTT. Protein was

then digested at 37º C using either trypsin (Promega) or GluC (Roche Diagnostics) at an enzyme to protein ratio of 1:200. Trypsin digestion was limited to one hour while GluC digestions were incubated for 14 hours. Following digestion, samples were de-salted using tC-18 solid phase extraction columns (Waters).

Phosphopeptides were enriched essentially as described (Phanstiel et al., 2011). Briefly, magnetic beads (Qiagen) were washed three times with water, three times with 40 mM EDTA (pH 8.0) for 30 minutes with shaking, and three times with water again. Beads were then incubated with 100 mM $FeCl_3$ for 30 minutes with shaking. Finally, beads were suspended in 1 mL 80% acetonitrile/0.1% TFA and washed three more times with 80% acetonitrile/0.1% TFA. Samples were suspended in 80% acetonitrile/0.1% TFA and incubated with beads for 30 minutes with shaking. Beads were washed six times with 200 µL 80% acetonitrile/0.1% TFA, and eluted using 1:1 acetonitrile:5% NH_4OH in water. Eluted phosphopeptides were acidified immediately with 4% formic acid, lyophilized to ~10 µL.

Cloning/protein expression

The coding sequence for NANOG was amplified from the pSIN-EF2_NANOG_Pur plasmid (Yu et al., 2007) (primers: AAAAAGCTAGCATGAGTGTGGATCCAGCTTGTCC; GAGCTCAAGCTTCGAATTCGATTCATC). The PCR fragment was then cloned into a modified pGEX4T vector; the resultant construct (pGEX4T-NANOG) was

expressed in Rosetta2 (DE3)pLysS *E. coli* cells (Novagen) and purified as previously reported (Brumbaugh et al., 2012).

Chromatographic separations

Capillary LC columns were constructed in house from 75 µm i.d. x 360 µm o.d. fused silica capillary. Analytical columns were constructed with an integrated emitter by use of a laser puller (P-2000, Sutter Instrument Company, Novarto, CA) followed acid etching of the tip as previously described (McAlister et al., 2012). Analytical columns were slurry-packed with Magic C18AQ reversed-phase material (3 µm, 200 Å, Michrom Bioresources Inc.) to 15 cm in length. Pre-columns incorporated cast chemical frits and were packed to 5 cm in length with the same stationary phase (Ficarro et al., 2009). Nanoflow separations were conducted on a nanoACQUITY UPLC system (Waters Corp.) using a vented-style configuration for sample loading. Mobile phase A consisted of 0.2% formic acid and mobile phase B was acetonitrile with 0.2% formic acid. Samples were loaded using 100% mobile phase A at 1.5 µL/min for 10 minutes. Gradient elution was optimized for each enzymatic digest, but generally consisted of a linear increase from 7.5% to 35% mobile phase B over 60 minutes at 300 nL/min.

Mass spectrometry

Tandem mass spectrometry was performed on two generations of orbitrap mass spectrometers, an LTQ Orbitrap Velos, and an Orbitrap Elite (Thermo Fisher Scientific). MS automatic gain control (AGC) was set at a target value of $1x10^6$ charges and the MS/MS AGC target was varied from $1\text{-}3x10^5$ charges. A

resolving power of $6x10^5$ was used for MS while a resolving power of $7.5x10^3$ (Orbitrap Velos) or $1.5x10^4$ (Orbitrap Elite) was used for MS/MS. Data-dependent MS/MS was performed with an isolation width of 3 Th. For ETD experiments, charge-state dependent reaction time scaling was used such that a plus 3 precursor was reacted for ~ 60 ms. For HCD experiments, the normalized collision energy was varied from 34-38%. Dynamic exclusion was enabled with a +/- 10 ppm tolerance with the repeat count varied from 1 to 3 and the exclusion duration varied from 10-30 s. Unassigned and singly charged precursors were excluded from MS/MS selection and the monoisotopic mass selection option was enabled. For HCD, a fixed first mass of *m/z* 115 was used to ensure mass analysis of the TMT reporter tag region (*m/z* 126-131). For targeted experiments, HCD collision energy and ETD reaction times were optimized based upon the results of scouting runs with those parameters varied.

Spectral Processing

Raw data was processed using COMPASS (Coon OMSSA Proteomic Analysis Software Suite), freely available at http://www.chem.wisc.edu/~coon/software.php (Wenger et al., 2011). The TMT-6plex spectral cleaning option was selected as was ETD spectral preprocessing when applicable for creating DTA files (Good et al., 2009). DTAs were searched using the Open Mass Spectrometry Search Algorithm (OMSSA, v2.18) using a concatenated target-decoy human database (IPI v3.86) combined with a target-decoy *E. coli* database (K12/MG1655)(Geer et al., 2004). Carbamiodomethylation of cysteine residues and TMT modification of lysine

residues and peptide N-termini were searched as fixed modifications and methionine oxidation was searched as a variable modification. For HCD, phosphorylation with neutral losses from serine and threonine and phosphorylation of tyrosine were searched as variable modifications. For ETD, phosphorylation of serine, threonine, and tyrosine were considered. The precursor search tolerance was set at +/- 1.5 Th (average mass), while the product ion tolerance was +/- 0.01 Th (monoisotopic mass). The command to utilize the precursor charge state input was enabled. Up to 3 enzymatic missed cleavages were allowed. Results were filtered to a 1% false discovery rate (FDR) using the FDR Optimizer program in COMPASS (Elias and Gygi, 2007; Wenger et al., 2011).

SUPPLEMENTAL REFERENCES

Brumbaugh, J., Hou, Z., Russell, J.D., Howden, S.E., Yu, P., Ledvina, A.R., Coon, J.J., and Thomson, J.A. (2012). Phosphorylation regulates human OCT4. Proc Natl Acad Sci U S A *109*, 7162-7168.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E.*, et al.* (2011). Chemically defined conditions for human iPSC derivation and culture. Nature Methods *8*, 424-429.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207-214.

Ficarro, S.B., Zhang, Y., Lu, Y., Moghimi, A.R., Askenazi, M., Hyatt, E., Smith, E.D., Boyer, L., Schlaeger, T.M., Luckey, C.J.*, et al.* (2009). Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. Anal Chem *81*, 3440-3447.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X.Y., Shi, W.Y., and Bryant, S.H. (2004). Open mass spectrometry search algorithm. J Proteome Res *3*, 958-964.

Good, D.M., Wenger, C.D., McAlister, G.C., Bai, D.L., Hunt, D.F., and Coon, J.J. (2009). Post-acquisition ETD spectral processing for increased peptide identifications. J Am Soc Mass Spectrom *20*, 1435-1440.

McAlister, G.C., Russell, J.D., Rumachik, N.G., Hebert, A.S., Syka, J.E.P., Geer, L.Y., Westphall, M.S., Pagliarini, D.J., and Coon, J.J. (2012). Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. Anal Chem *84*, 2875-2882.

Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., Probasco, M.D., Bailey, D.J., Swaney, D.L., Tervo, M.A., Bolin, J.M., Ruotti, V.*, et al.* (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nature Methods *8*.

Wenger, C.D., Phanstiel, D.H., Lee, M.V., Bailey, D.J., and Coon, J.J. (2011). COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. Proteomics *11*, 1064-1074.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.