The following resources related to this article are available online at http://stke.sciencemag.org.
This information is current as of 18 March 2011.

Article Tools	Visit the online version of this article to access the personalization and article tools: **http://stke.sciencemag.org/cgi/content/full/sigtrans;4/164/rs3**
Supplemental Materials	*"Supplementary Materials"* **http://stke.sciencemag.org/cgi/content/full/sigtrans;4/164/rs3/DC1**
Related Content	The editors suggest related resources on *Science*'s sites: **http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/146/ra80**
References	This article cites 68 articles, 22 of which can be accessed for free: **http://stke.sciencemag.org/cgi/content/full/sigtrans;4/164/rs3#otherarticles**
Glossary	Look up definitions for abbreviations and terms found in this article: **http://stke.sciencemag.org/glossary/**
Permissions	Obtain information about reproducing this article: **http://www.sciencemag.org/about/permissions.dtl**

System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation

Kristoffer T. G. Rigbolt,[1]* Tatyana A. Prokhorova,[1]* Vyacheslav Akimov, [1] Jeanette Henningsen, [1,2] Pia T. Johansen,[3] Irina Kratchmarova,[1] Moustapha Kassem,[3,4] Matthias Mann,[5,6] Jesper V. Olsen,[6†] Blagoy Blagoev[1†]

To elucidate cellular events underlying the pluripotency of human embryonic stem cells (hESCs), we performed parallel quantitative proteomic and phosphoproteomic analyses of hESCs during differentiation initiated by a diacylglycerol analog or transfer to media that had not been conditioned by feeder cells. We profiled 6521 proteins and 23,522 phosphorylation sites, of which almost 50% displayed dynamic changes in phosphorylation status during 24 hours of differentiation. These data are a resource for studies of the events associated with the maintenance of hESC pluripotency and those accompanying their differentiation. From these data, we identified a core hESC phosphoproteome of sites with similar robust changes in response to the two distinct treatments. These sites exhibited distinct dynamic phosphorylation patterns, which were linked to known or predicted kinases on the basis of the matching sequence motif. In addition to identifying previously unknown phosphorylation sites on factors associated with differentiation, such as kinases and transcription factors, we observed dynamic phosphorylation of DNA methyltransferases (DNMTs). We found a specific interaction of DNMTs during early differentiation with the PAF1 (polymerase-associated factor 1) transcriptional elongation complex, which binds to promoters of the pluripotency and known DNMT target genes encoding OCT4 and NANOG, thereby providing a possible molecular link for the silencing of these genes during differentiation.

INTRODUCTION

Human embryonic stem cells (hESCs) in culture represent a model system for studying cellular differentiation because they are pluripotent and can therefore differentiate into cells derived from all three embryonic germ layers (*1*). The self-renewal and pluripotent properties of these cells make them potential candidates for regenerative medicine applications aimed at replacing diseased or damaged tissue with functionally and lineage-specific differentiated hESCs (*2*, *3*). The emergence of protocols for reprogramming somatic cells to generate induced pluripotent stem (iPS) cells with properties essentially identical to those of embryo-derived ESCs has enhanced the promise of stem cell–based therapies (*4*, *5*).

To gain broader insight into the regulatory networks controlling pluripotency, self-renewal, and stem cell differentiation, system-wide studies of both the transcriptome (*6*) and the phosphoproteome have been performed (*7–9*). Improvements in fractionation, enrichment, in-

strumentation, and data processing have facilitated the application of mass spectrometry–based approaches for examining phosphorylation (*10*). Because the combination of global phosphoproteomics with stable isotope labeling by amino acids in cell culture (SILAC) (*11*) is an effective strategy for characterizing quantitative changes in phosphorylation sites at a global scale (*12–16*), we chose this technique for our studies. An application of two complementary peptide fragmentation techniques to map phosphorylation in pluripotent hESCs identified 10,844 sites (*7*). However, Swaney *et al.* did not apply quantitative strategies to follow changes in phosphorylation during differentiation, but provided insights into predominant linear kinase motifs present in hESCs in the pluripotent state (*7*). Van Hoof *et al.* used an approach-based on SILAC to quantify phosphorylation site dynamics during the first 4 hours after induction of differentiation with bone morphogenetic protein 4 (BMP4), which induces the cells to become trophoblasts, and identified 3067 phosphorylation sites, of which 1987 could be quantified (*9*). In addition to identifying previously unknown phosphorylation sites, this study revealed decreased activity of the cell cycle regulator cyclin-dependent kinase 2 (CDK2). Brill *et al.* applied a label-free quantification approach to compare the phosphoproteomes of undifferentiated hESCs and hESCs treated with retinoic acid for 4 days, which induces nonspecific differentiation to a heterogeneous population of cells (*8*). This resulted in the identification of 2546 phosphorylation sites on 1602 proteins, of which 929 showed differences between the two cellular states and revealed a role for platelet-derived growth factor (PDGF) in the maintenance of the hESC pluripotency.

To expand the understanding of the cellular regulation and circuitry involved in hESC self-renewal and differentiation, we applied quantitative proteomics and phosphoproteomics to profile dynamic changes

[1]Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. [2]Centre of Inflammation and Metabolism, University of Copenhagen, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark. [3]Molecular Endocrinology Laboratory, Department of Endocrinology, Odense University Hospital and Medical Biotechnology Center, Winslowsparken 25, DK-5000 Odense, Denmark. [4]Stem Cell Unit, Department of Anatomy, College of Medicine, King Saud University, 11461 Riyadh, Kingdom of Saudi Arabia. [5]Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. [6]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b, DK-2200 Copenhagen, Denmark.
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bab@bmb.sdu.dk (B.B.); jesper.olsen@cpr.ku.dk (J.V.O.)

in protein abundance and phosphorylation state during the first 24 hours after induction of lineage-independent (nondirected) hESC differentiation. To enable discrimination between treatment-specific and common events associated with initiation of differentiation, we performed parallel quantitative proteomics studies on hESCs undergoing two distinct differentiation protocols: initiation of nondirected differentiation by a diacylglycerol analog (*17–19*) and initiation by transfer into medium that was not conditioned by feeder cells, which are required for maintenance of stem cell renewal and pluripotency (*20*). These screens, combined with subsequent bioinformatic analysis, provide a source of information on the complex and dynamic processes underlying hESC pluripotency and nondirected differentiation. In addition to identifying numerous pre-

viously unknown phosphorylation sites on many proteins involved in signal transduction and transcriptional regulation, we observed dynamic changes in the phosphorylation status of DNA methyltransferases (DNMTs), which silence the genes encoding the transcription factors OCT4 (also known as POU5F1) and NANOG, that are crucial for self-renewal and pluripotency (*21, 22*). We also present evidence for a specific interaction between DNMTs and the polymerase-associated factor 1 (PAF1) transcriptional elongation complex and show that this interaction increased during differentiation. The changes in DNMT phosphorylation and its increased interaction with the PAF1 complex that occurred in response to differentiation may provide a link between DNMT regulation and stem cell identity.
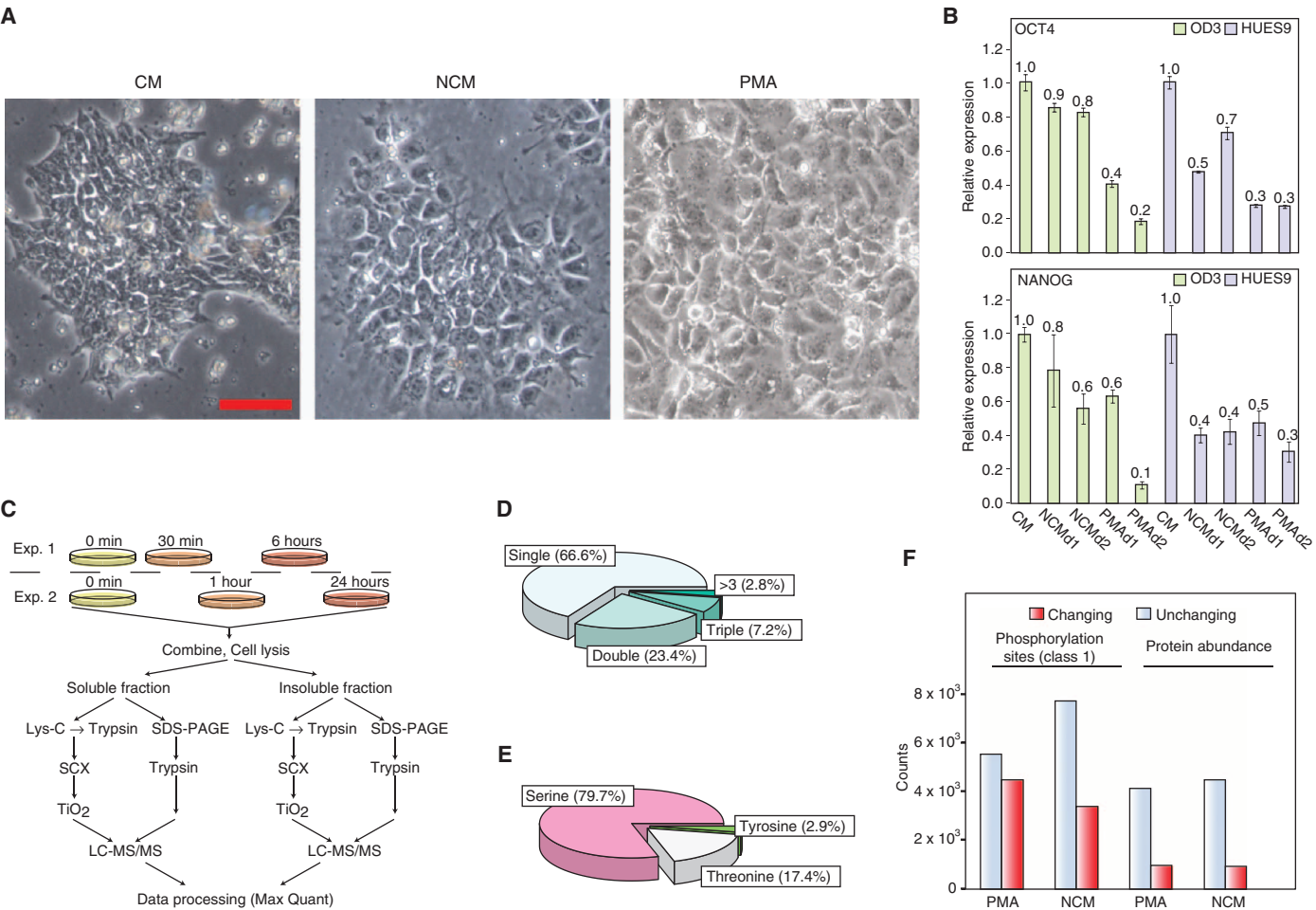

Fig. 1. Characteristics of the proteome and phosphoproteome of hESCs undergoing 24 hours of nondirected differentiation. (**A**) Phase-contrast images of HUES9 hESCs cultured in condition medium (CM) (left), treated with nonconditioned medium (NCM) (center), or treated with PMA (right) for 24 hours. Scale bar, 100 μm. (**B**) The abundance of transcripts encoding the self-renewal markers OCT4 and NANOG in untreated cells and cells treated with NCM or PMA in Odense-3 (green) and HUES9 (purple) cells was evaluated by real-time QPCR. Error bars indicate SD from three replicates. (**C**) Flow diagram of the quantitative proteomics and phosphoproteomics platform applied to early (within 24 hours) hESC differentiation. (**D**) Proportion of singly and multiply phosphorylated peptides in the com-

bined data sets of all time points of HUES9 cells treated with either NCM or PMA. (**E**) Distribution of identified phosphorylated serine, threonine, and tyrosine residues in the combined data sets of all time points of HUES9 cells treated with either NCM or PMA. (**F**) Number of class 1 phosphorylation sites that showed a change in status or the number of proteins that showed a change in abundance or were unchanged after PMA or NCM treatment of HUES9 cells at all time points. We included any phosphorylation sites that exhibited a twofold change or any protein that exhibited a change in abundance relative to the self-renewing hESCs with significance less than 0.01 at any time point as "Changing." Counts refers to the number of phosphorylation sites or the number of proteins, respectively.

RESULTS

Induction of nondirected hESC differentiation by nonconditioned medium or phorbol ester

We conducted global proteomics and phosphoproteomics experiments to discover proteins with altered abundance, phosphorylation status, or both during hESC differentiation to gain insight into the proteins and corresponding pathways involved in maintaining the self-renewing pluripotent state, as well as those directing the hESC transition toward differentiation. We used two distinct differentiation protocols to discriminate between events specific to a particular treatment and those generally associated with the induction of hESC differentiation. One treatment involved exposure of the cells to phorbol 12-myristate 13-acetate (PMA), a diacylglycerol analog that activates members of the protein kinase C (PKC) family of kinases and that is a potent inducer of hESC differentiation (17–19). We also induced differentiation by transferring hESC cultures from medium conditioned by mouse embryo fibroblasts (MEFs) to a nonconditioned medium (NCM), which induces differentiation of hESC due to the absence of factors, contributed to the medium by the feeder cells, that are required for stem cell maintenance (20). To ensure that the observed properties were not cell line–specific, we examined two different hESC lines: Harvard University embryonic stem cells 9 (HUES9) (23) and Odense-3 (19). Subjecting the hESCs to either treatment resulted in rapid changes in cellular morphology (Fig. 1A). Whereas the cells maintained in conditioned media preserved the morphology that was typical for undifferentiated hESCs of tightly packed colonies of round cells with high nucleus-cytoplasm ratio, the PMA- or NCM-treated cells became visibly flattened with lower nucleus-cytoplasm ratio and formed more diffuse colonies. Furthermore, despite the observed variability in the response of the cell lines to the two treatments, the expression of either one or both of the genes encoding the self-renewal markers OCT4 and NANOG was reduced within 24 hours (Fig. 1B).

To examine whether PMA and NCM treatment induced hESCs toward any lineage-specific differentiation, we performed quantitative polymerase chain reaction (QPCR) experiments to directly monitor the expression of 12 lineage-specific genes, 4 for each of the ectoderm, endoderm, or mesoderm lineages (fig. S1). Neither the HUES9 nor the Odense-3 cells demonstrated consistent preferential expression of multiple markers for a single germ layer. However, the abundance of the transcripts of some marker proteins, such as the ectodermal marker PAX6 and mesodermal marker BRACHYURY, displayed different trends in the two cell lines (fig. S1). This is consistent with previous observations that the propensity to differentiate to distinct lineages varies considerably among different established hESC cell lines (24). Additionally, some markers exhibited differences associated with the treatment paradigm. For example, the abundance of the transcripts for the endodermal marker SOX7 was increased in response to exposure to PMA, but not NCM (fig. S1). This variability in the marker expression profiles suggests that differentiation occurred randomly in both cell lines and that the two stimuli trigger hESCs toward two distinct, but lineage-independent, differentiation programs.

Temporal profiling of the hESC proteome and phosphoproteome during early differentiation

To profile the early phosphorylation events during hESC differentiation, we combined SILAC with strong cation exchange (SCX) fractionation, TiO_2 phosphopeptide enrichment, and high-accuracy and high-resolution liquid chromatography–tandem mass spectrometry (LC-MS/MS) (16) (fig. S2A). Four populations of HUES9 cells were "double-triple"–labeled with three isotopically distinct versions of lysine and arginine (16, 25) and
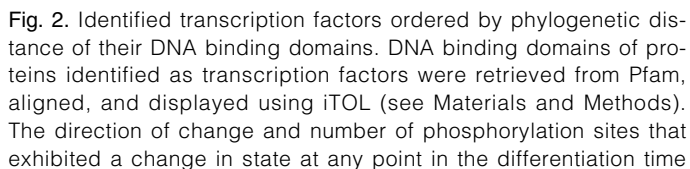
subjected to either PMA or NCM treatment for 0 min, ½ hour, 1 hour, 6 hours, or 24 hours. After harvesting, the cells from either treatment were combined into two pools each (one of the 0-min, ½-hour, and 6-hour samples, and one of the 0-min, 1-hour, and 24-hour samples) and further processed as described in Materials and Methods for proteome and phosphoproteome analyses, resulting in time course profiles with five points (Fig. 1C). Biological replica experiments for the 0-min, ½-hour, and 6-hour time points after both treatments were also performed (fig. S2, B and C, and table S3). Not including the replicate experiment, 378 samples were analyzed separately by LC-MS/MS. We processed all raw MS files together using the MaxQuant software suite (26), identified the peptide sequences with the Mascot search algorithm, and filtered the data set to a final false discovery rate (FDR) of less than 1% at the peptide level and then filtered those data at the same FDR of less than 1% at the protein level (27), resulting in the identification of 78,819 peptides with an effective peptide FDR of 0.28% and an average absolute mass error of 401 parts per billion (ppb) (fig. S2A).

The identified peptides mapped to 6521 proteins (table S1) and included 14,865 phosphopeptides (table S2) unique by sequence, of which most were singly or doubly phosphorylated (Fig. 1D). The 14,865 phosphopeptides mapped to 4335 proteins containing 23,522 phosphorylation sites, of which 15,004 could be assigned to a specific position within the protein with a probability of at least 0.75 (class 1 sites; median localization probability 0.996) (16). The 15,004 class 1 sites were composed of 11,954 phosphorylated serine (pSer) residues, 2609 pThr, and 441 pTyr sites (Fig. 1E). The biological replica experiments for the 0-min, ½-hour, and 6-hour time points produced 4667 proteins and 8205 phosphorylation sites, of which 5699 were class 1 sites (table S3).

The changes associated with differentiation in phosphorylation state were more extensive than the changes in protein abundance. We interpret these changes to reflect regulatory events that influence the activity or function of the identified proteins and the involvement of these proteins in the hESC differentiation process. However, we acknowledge that it is possible that some of these changes are functionally silent or not relevant to the differentiation process. We found that the phosphorylation status of 4504 (45% of 10,066) of the class 1 phosphorylation sites dynamically changed after PMA treatment, and the phosphorylation status of 3380 (30% of 11,104) dynamically changed after NCM treatment (Fig. 1F). That both treatments caused such a pronounced alteration of the phosphoproteome suggests that a large fraction of the cellular protein networks was engaged in the conversion of the hESCs from their self-renewing pluripotent state toward differentiation and that this process is complex. In the proteome data set, we observed changes in the abundance of 955 (19% of 5103) and 934 (17% of 5428) of the identified proteins after PMA and NCM treatment, respectively (Fig. 1F).

We classified all proteins and phosphorylation sites according to Gene Ontology (GO) (28) for "Molecular Function" and "Cellular Component" and found that they spanned a broad range of categories (fig. S3). For example, the proteome data set contained 223 kinases, 467 transcription regulators, and 137 receptors. Many of the class 1 phosphorylation sites are also present in these categories, including >650 sites on kinases, >1600 sites on transcription regulators, and >280 sites on receptors (fig. S3). All data are included in tables S1 and S2 from which we provide links to graphical displays of the corresponding temporal profiles and are also available in the Phosida database (http://www.phosida.com) (29).

Because changes in gene expression are critical for hESC differentiation, we examined in more detail the 216 transcription factors identified in the data sets (Fig. 2), which included OCT4, SOX2, SOX13, SOX15,

Fig. 2. Identified transcription factors ordered by phylogenetic distance of their DNA binding domains. DNA binding domains of proteins identified as transcription factors were retrieved from Pfam, aligned, and displayed using iTOL (see Materials and Methods). The direction of change and number of phosphorylation sites that exhibited a change in state at any point in the differentiation time course are represented by bar plots in the outer rim, and the bars in the inner rim indicate the change in protein abundance, which includes any change detected in either differentiation paradigm. All proteins are listed by their corresponding gene names. (Note: *POU5F1* is the gene name for OCT4 and *FOSL2* is the gene name for FRA2.) See also table S4.

UTF1 (undifferentiated embryonic cell transcription factor 1), FOXO1 (Forkhead box protein O1), FOXO3, FOXO4, SALL1 (Sal-like protein 1), SALL2, and SALL4. These families of transcription factors comprise some of the most well-known factors implicated in the processes of stem cell differentiation or stem cell self-renewal and pluripotency (21, 30–33). To evaluate whether there were consistent trends in terms of the changes in protein abundance or phosphorylation status among transcription factors that had homologous DNA binding domains, we ordered the transcription factors according to homology based on their annotated domains and presented the regulation of their expression and identified phosphorylation sites (Fig. 2). The abundance of only 26 and 20 of these

factors changed within 24 hours of PMA and NCM treatment, respectively (table S4). As expected, we observed a marked decrease in OCT4 abundance upon each of the treatments. Decreased abundance of SOX2, another pluripotency-associated marker, occurred after PMA treatment, but the abundance of this transcription factor was increased in NCM-treated cells. This observation may reflect that in addition to its functions in maintaining hESC pluripotency, SOX2 is also a neural stem cell marker (34) and is an indication that PMA and NCM induce distinct differentiation programs in hESCs. In addition, we identified four class 1 phosphorylation sites on SOX2, and whereas the phosphorylation of three of these (Ser^{246}, Ser^{249}, Ser^{251}) was decreased within 30 min of
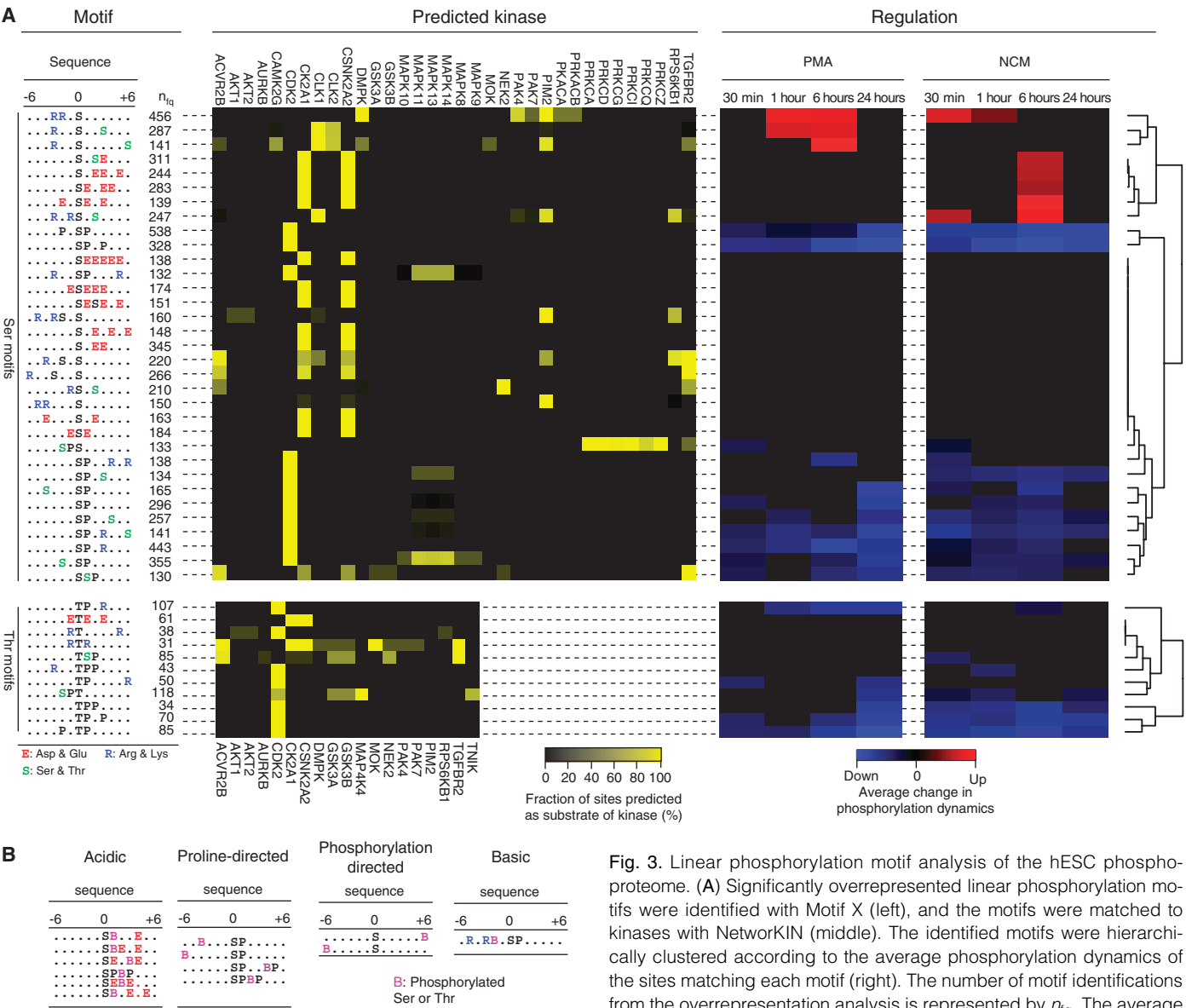

Fig. 3. Linear phosphorylation motif analysis of the hESC phospho-proteome. (**A**) Significantly overrepresented linear phosphorylation motifs were identified with Motif X (left), and the motifs were matched to kinases with NetworKIN (middle). The identified motifs were hierarchically clustered according to the average phosphorylation dynamics of the sites matching each motif (right). The number of motif identifications from the overrepresentation analysis is represented by n_{fq}. The average change in phosphorylation of sites matching the motif in each time point is color-coded blue for decreased or red for increased compared to undifferentiated cells. Only average phosphorylation changes with a mean significantly different from zero on confidence level 0.01 were included. See also figs. S4 and S5. (**B**) Motifs with two phosphorylation sites identified as overrepresented compared to their single phosphorylated peptides using Motif X.

NCM treatment, the phosphorylation of these sites was unchanged during the entire course of PMA treatment (table S4). Among the group of transcription factors that exhibited increased abundance were several members of the AP-1 family including c-JUN, JUNB, and FRA2 (also known as FOSL2), suggesting that these factors may play a role in the initiation of differentiation.

We identified 714 class 1 phosphorylation sites on transcription factors, including 282 sites not previously reported as phosphorylated on 106 distinct proteins (table S4). Almost half of all phosphorylation sites (327 of 714) on transcription factors changed more than twofold during differentiation, including sites on hESC markers, such as SOX2 and UTF1, indicating extensive direct involvement of this posttranslational modification in the control of gene expression in hESCs.

Distinct changes in the hESC kinome during nondirected differentiation

To identify specific kinases that are activated by the two treatments, we searched for the occurrence of predominant linear kinase motifs within the identified phosphorylation sites using Motif X (35) (see Materials and Methods for details). We identified 33 significantly enriched phosphoserine motifs and 11 phosphothreonine motifs (Fig. 3A, fig. S4, and table S5). To identify the kinases likely to be responsible for phosphorylating these motifs, we processed all of the amino acid sequences containing an enriched phosphorylation motif with the NetworKIN algorithm (36). We also examined how phosphorylation of specific motifs changed during hESC differentiation by testing whether the mean of ratios of the phosphorylation sites matching each motif was significantly different from zero in one or more of the four time points and two treatments. After clustering the motifs solely on the basis of any change in phosphorylation status without distinguishing those that increased from those that decreased, we observed a partition into groups with similar amino acid composition surrounding the central phosphorylation site. This shows that the motifs do indeed represent separate regulatory entities with distinct patterns of phosphorylation regulation (Fig. 3A and fig. S4).

This motif analysis revealed a group of motifs enriched in charged amino acid residues that became increasingly phosphorylated after both PMA and NCM treatment (Fig. 3A), suggesting that their corresponding kinases are critical nodes in the transition of hESCs from self-renewing pluripotent state toward a differentiated state. On the basis of NetworKIN predictions, a large proportion of the basic-rich motifs were predicted to be substrates of CDC-like kinase 2 (CLK2), p21-activated kinase 7 (PAK7), PIM2, or cAMP (adenosine 3′,5′-monophosphate)–dependent protein kinase (PKA) (Fig. 3A and table S6), whereas the acidic-rich motifs were predicted to be phosphorylated by members of the casein kinase family and myotonic dystrophy protein kinase (DMPK). Conversely, we observed a tendency toward decreased phosphorylation of motifs phosphorylated by proline-directed kinases, such as the mitogen-activated protein kinases (MAPKs) and the CDKs. These sites were consistently less phosphorylated, reaching minima after 6 hours of NCM treatment or continuing to decrease in phosphorylation for 24 hours after PMA treatment (Fig. 3A). Therefore, the attenuation of the activities of these kinases may be necessary for the progression of hESC differentiation.

We searched for enriched linear motifs on multiple phosphorylated peptides to examine whether the presence of a phosphorylation site created an alternative motif for additional phosphorylation in the proximity of the corresponding site. With this approach, we identified 12 significant double-phosphorylation motifs (Fig. 3B, fig. S5, and table S5). As previously noted for nonembryonic cells (37), most of these motifs are strongly acidic or contain a proline residue in the +1 position. We also

identified two previously unknown motifs, which did not have any sequence constraints apart from the presence of another phosphorylation site in either the −6 or the +6 position relative to the phosphorylation site. All the remaining double-phosphorylation motifs could be generated by distinct kinases independent of each other, despite the presence of one phosphorylation possibly influencing the frequency of phosphorylation at the other site, because these motifs appear to be concatenations of individual single-phosphorylation motifs (Fig. 3B and fig. S5).

Our data set included more than 200 protein kinases with 654 class 1 phosphorylation sites, including 195 previously unreported sites on 86 kinases (table S6). The changes in the abundance and phosphorylation status of all of the kinases that we identified are presented in Fig. 4, ordered on the basis of the similarity of their kinase domains (38). Through the motif analysis, we found extensive reduction of phosphorylation of sites predicted to be targeted by members of the CDK, MAPK, GSK (glycogen synthase kinase), and CDK-like (CMGC) family of kinases (Fig. 3A). From our proteomic analysis, we identified 29 members from this family spanning all subfamilies (Fig. 4 and table S6). CDK6 was the only kinase that exhibited a similar change (increase) in abundance in both treatments; the remainder of the kinases did not change their abundances in either of the differentiation paradigms or only exhibited a change in abundance in response to one of the treatments and no change in the other. Intriguingly, transcription of the gene encoding CDK6 is controlled by the master stem cell regulator NANOG (39) and hence might be expected to be down-regulated during differentiation. However, CDK6 displayed a more than threefold increase in abundance after PMA treatments and a twofold increase after NCM treatment relative to that of self-renewing undifferentiated cells maintained in conditioned medium (CM) (table S6), indicating that there are additional mechanisms controlling CDK6 abundance in hESCs.

In contrast to the few changes in the abundance of most CMGC family kinases, we identified regulated phosphorylation sites on a large number of these proteins. Among these phosphorylation events, we observed increased phosphorylation of Thr^{14} and Tyr^{15} on both CDK2 and CDC2 (also known as CDK1); phosphorylation of these sites inhibits the kinase activity of both proteins (40). This correlates with the observed decrease in phosphorylation sites predicted to be targets of these CDKs from the motif analysis (Fig. 3A). Tyr^{15} on CDK2 and CDC2 is phosphorylated by one of the master cell cycle regulators, WEE1, and this phosphorylation event inhibits cell cycle progression (41). After 6 hours of either treatment, the abundance of WEE1 peaked at about threefold higher amounts compared to that of the self-renewing hESCs (table S6). The expression of the gene encoding WEE1 is repressed by the pluripotency-associated transcription factor Krueppel-like factor 4 (KLF4) (42), and the abundance of KLF4 is reduced during differentiation (4). The reduction in KLF4 abundance may lead to increased abundance of WEE1 and subsequent CDK2 and CDC2 inhibition by Tyr^{15} phosphorylation, providing a link to the observed decreased phosphorylation of target substrates for these two CDK kinases (Fig. 3A). Our results are consistent with a model that the decrease in the abundance of pluripotency-related transcription factors that are associated with the progress toward somatic lineages may moderate the length of time spent in various phases of the cell cycle, and this effect on the cell cycle may be mediated by the WEE1-dependent inhibitory phosphorylation of CDK2 and CDC2 (Fig. 5). That these observations are similar after PMA and NCM treatment supports this model and suggests that it might be a universal regulatory circuit used during hESC differentiation. This example also illustrates the potential of integrating system-wide studies of protein and phosphorylation changes with our accumulating knowledge of ES cell fate determination.
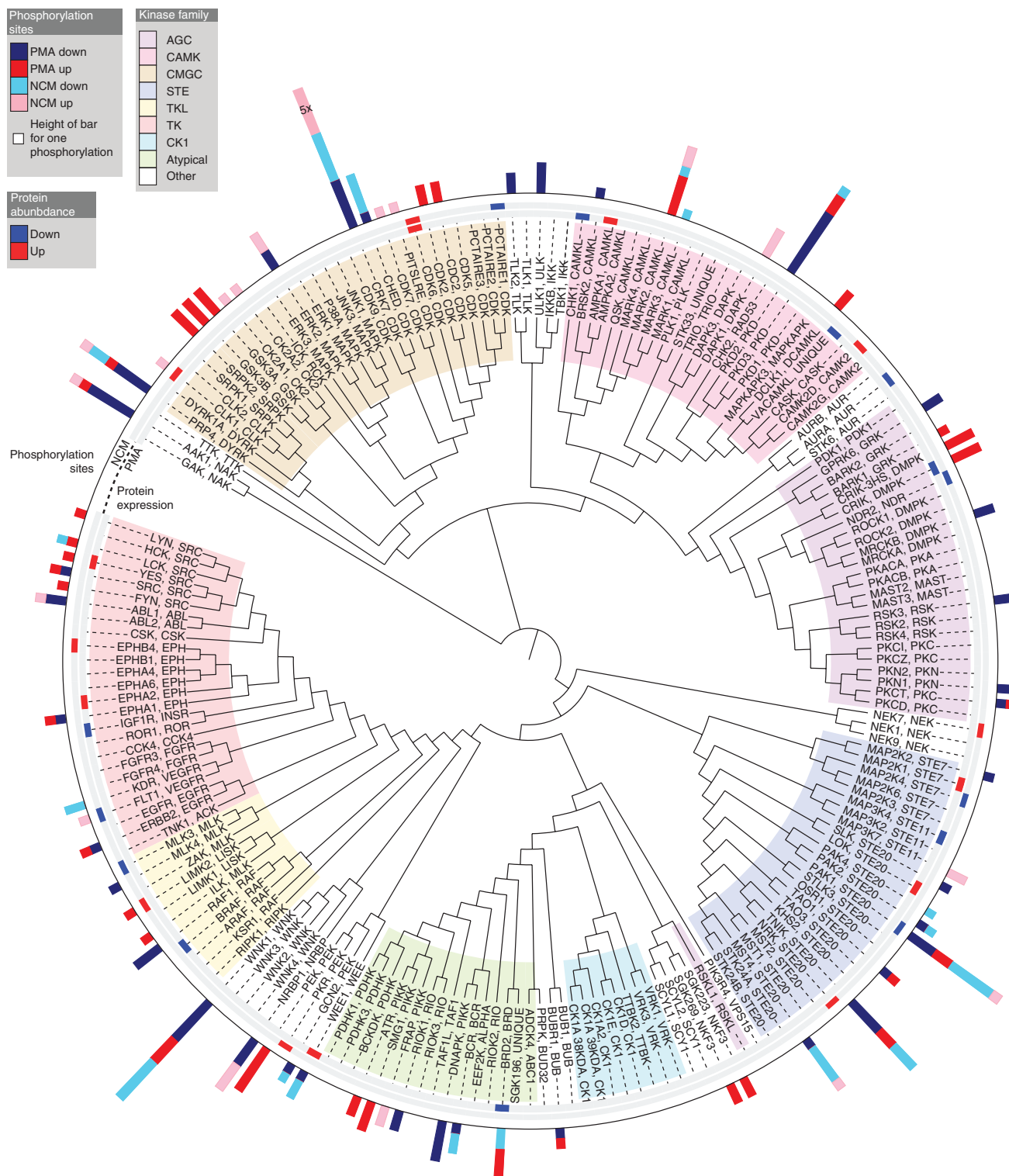
Fig. 4. Identified kinases ordered by phylogenetic distance of the kinase domain. Kinase domains of identified kinases were retrieved from KinBase, aligned, and displayed using iTOL (Materials and Methods). The direction of change and number of phosphorylation sites that exhibited a change in state at any point in the differentiation time course are represented by bar plots in the outer rim, and the bars in the inner rim indicate the change in protein abundance, which includes any change detected in either differentiation paradigm. All proteins are listed by their corresponding gene names. For the CDK CRK7, 5× in the pink bar represents a total of 20 class 1 sites on CRK7 with increased phosphorylation status after NCM treatments.
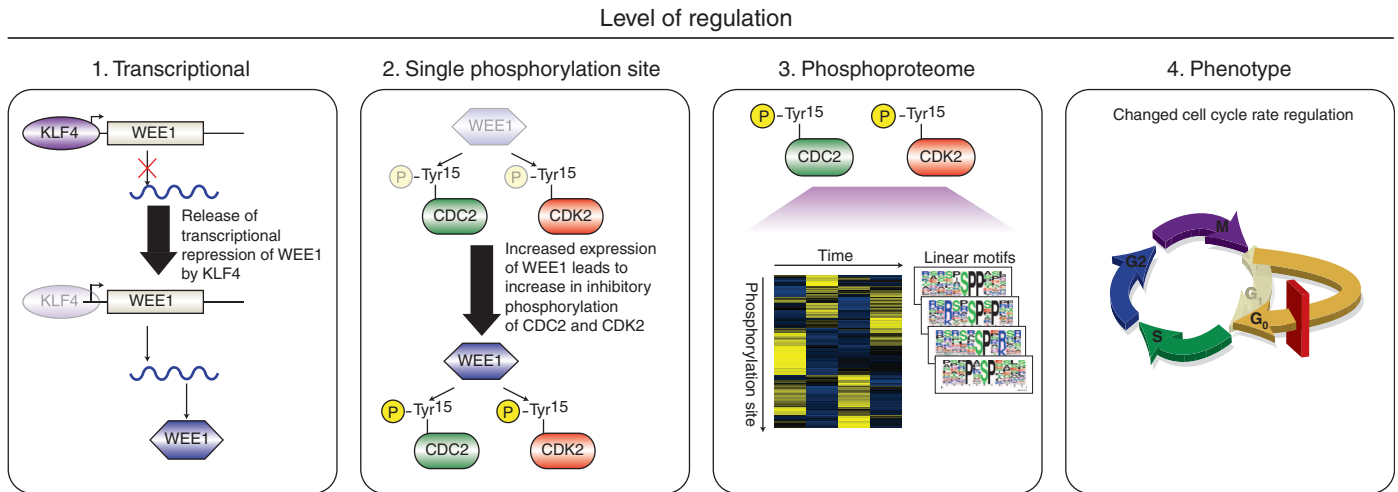
Fig. 5. A model for how an increase in the abundance of WEE1 (transcriptional regulation) and the phosphorylation of its targets CDC2 and CDK2 (single phosphorylation site regulation) alters the activity of these kinases to exert a change in the phosphoproteome and alter cell cycle progression.

Identification of a core hESC phosphoproteome and dynamic changes in the phosphorylation of critical transcriptional regulators

We induced hESC differentiation by two fundamentally different approaches to allow discrimination between treatment-specific events and events associated with initiation of hESC lineage-independent differentiation in general, which should allow us to identify the core phosphoproteome of nondirected differentiating hESCs. To compare the extent of similarity between the changes in each phosphorylation site after both treatments, we defined a quantitative measure, a similarity score (S-score) (see fig. S6 for details about S-score calculation). Briefly, for each phosphorylation site, the difference in the total degree of change after the two treatments is quantified by the area under the time-course profiles (the magnitude component), and the similarity of their temporal profiles is calculated on the basis of the Pearson correlation coefficient (the temporal component). These two values are then multiplied and the S-score is calculated as a $-\log_{10}$ of the resulting number, thereby providing an estimate of the similarity of each phosphorylation site between the two treatments, accounting for both the temporal pattern and the total change in the magnitude of phosphorylation.

When the temporal changes in the phosphorylation sites were plotted against the magnitude of the phosphorylation changes, phosphorylation sites with similar regulation profiles between the two treatments clustered in the upper left region of the heat map (Fig. 6A). This analysis shows that both the temporal and the magnitude components of the S-score needed to be taken into account to describe the similarity of regulation because rarely does the magnitude of the change correlate with the time of treatment (Fig. 6A). Calculated in this way, an S-score of 20, for example, corresponds to a Pearson correlation P value of 0.1 (meaning, a correlation at significance level 90%) and a difference in the magnitude of the phosphorylation change that is at most 10% between the two treatments. Therefore, we considered phosphorylation sites with S-score of 20 or higher as changing in a similar way after the two treatments and counted those as a "core hESC phosphoproteome" (all S-scores are listed in table S2). Plotting the S-score frequency for all phosphorylation sites that changed in response to the treatments

revealed that most changes were treatment-specific and that a smaller set of the changes was common to both treatments (Fig. 6, B and C; see also fig. S6B for representative examples of the temporal profiles covering the entire range of S-scores).

From the group of phosphorylation sites exhibiting changes in status within the core hESC phosphoproteome, we found similar S-scores for sites on various proteins previously implicated in hESC pluripotency and differentiation. Among these sites were three phosphorylations of Ser[609], Ser[610], and Ser[612] on the transcriptional regulator CTCF (CCCTC-binding factor), which were all dephosphorylated more than twofold by 30 min of treatment relative to the self-renewing hESCs (Fig. 6D). CTCF is a known regulator of c-MYC, and dephosphorylation of these sites on CTCF increases the repression of transcription from CTCF target genes that are also regulated by c-MYC (43). The transcription factor c-MYC is a well-established hESC marker and was included in the original group of four proteins required to produce iPS cells (4). Considering the decreased CTCF phosphorylation on these critical sites in two distinct protocols of hESC differentiation and its previously reported effect in repressing the expression of c-MYC target genes, dephosphorylation of these CTCF sites might constitute a key event in hESC differentiation. We also found decreased amounts of Ser[18] phosphorylation on UTF1 after induction of differentiation by either PMA or NCM treatment (Fig. 6D). UTF1 is an established hESC transcriptional regulator and influences the efficiency of human iPSc generation (44). Furthermore, both treatments resulted in dynamic changes in the phosphorylation of Ser[37] on SOX15, a protein that is homologous to SOX2. SOX15 interacts with OCT4 in mouse ESCs (mESCs) and, when associated with OCT4, binds promoters with known OCT4/SOX2 binding sites, such as in the gene encoding fibroblast growth factor 4 (FGF4) (31). In each of these examples of phosphorylation-state changes during differentiation, the changes were dynamic over the treatment period (Fig. 6D). However, overall, the sites on CTCF and UTF1 were dephosphorylated and remained lower than the amounts in the self-renewing hESCs. In contrast, the phosphorylation of Ser[37] on SOX15 was reduced by 30 min, but subsequently increased twofold higher than in the self-renewing hESCs by 24 hours (Fig. 6D). The core hESC phosphoproteome also
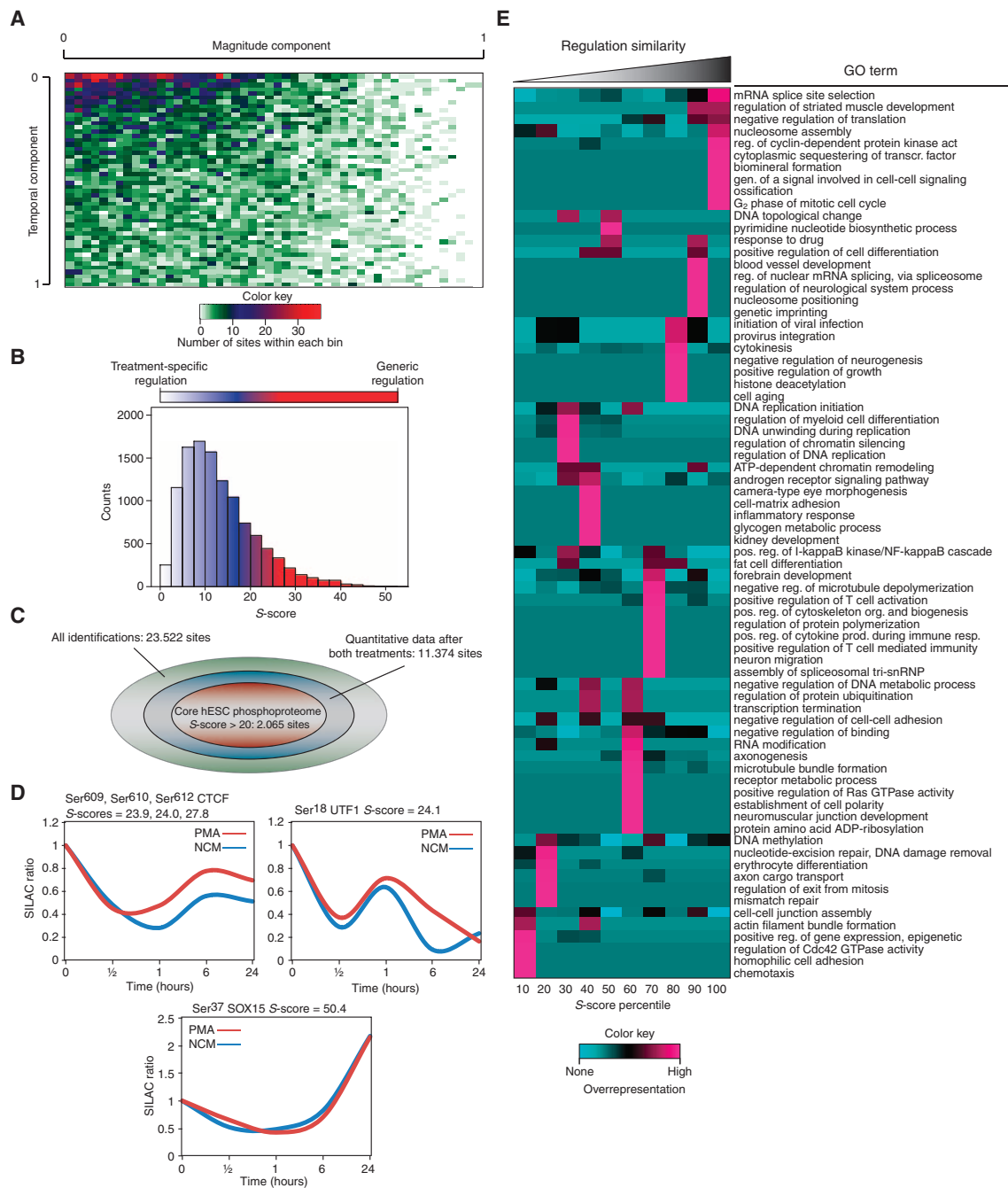
Fig. 6. A quantitative assessment by *S*-score of the similarity in the phosphoproteome after PMA or NCM treatment. (**A**) Distribution of the change in phosphorylation status of phosphorylation sites in the temporal and magnitude components of the *S*-score. The heat map is color-coded by the number of phosphorylation sites with a combination of temporal and magnitude component values, with a score of 0 reflecting perfect correlation and 1 reflecting no correlation between the response to the two treatments. (**B**) Histogram of the *S*-score occurrences ranging from treatment-specific changes (white) to changes observed in both treatments (red). Counts represents the number of phosphorylation sites exhibiting a particular range of *S*-score. (**C**) Overview of the results from *S*-score analysis showing that

2065 phosphorylation sites comprise the identified core hESC phosphoproteome. (**D**) Temporal profiles of phosphorylation sites on the indicated transcriptional regulators. (**E**) Phosphorylation sites were partitioned in 10 equally sized groups on the basis of *S*-score, and each group was tested for overrepresented GO Biological Process terms compared to the sites from the remaining nine groups. Enriched GO terms for the most treatment-specific sites are found at the left and the similarity increases toward the right. The results are color-coded to indicate the *S*-score percentile bins where the GO term is found overrepresented. See also fig. S6. GO terms use the following abbreviations: reg., regulation; act, activity; gen, generation; transcr., transcription; pos., positive.

contained coordinately changing phosphorylation sites on other transcription factors (for example, CUX2, TCF20, and ZNF423), kinases (for example, PAK1 and PAK2), and ubiquitin E3 ligases (for example, UBR4, HUWE1, and NEDD4L) (table S2). Thus, we have identified a set of phosphoproteins that exhibit coordinated changes in phosphorylation status in response to more than one nondirected differentiation protocol during the first 24 hours. Whether these changes are directing the transition from self-renewal and pluripotency to a differentiated state or whether these changes are the result of that transition remains to be determined. Nevertheless, these dynamically, but coordinately, changing phosphoproteins represent candidates for future studies of the maintenance of hESC pluripotency and self-renewal or differentiation.

To gain insight into processes associated with each treatment or with differentiation in general, we grouped the phosphorylation sites according to their S-scores in 10 bins with an equal number of sites and tested for overrepresented GO "Biological Process" terms in each of the bins versus the remaining nine bins (Fig. 6E). Treatment-specific processes are positioned on the left (sites with lowest S-scores), and events common to both treatments are found on the right (sites with highest S-scores). Phosphorylation sites on proteins associated with several GO terms related to cell cycle control changed similarly after the two treatments, because GO terms, such as "regulation of cyclin-dependent protein kinase activity" and "G_2 phase of mitotic cell cycle," were enriched within the sites with the 10% highest S-scores. Conversely, we found that phosphorylation sites on proteins involved in cell adhesion exhibited divergent changes in phosphorylation and were enriched in the left part of the heat map containing sites with low S-scores. This finding also correlates with our observation that the ability of hESCs treated with PMA attached more slowly and poorly to Matrigel compared to the cells treated with NCM.

Regulatory phosphorylations on epigenetic regulators

To uncover additional patterns in the changes in phosphorylation site status, we subjected the sites with changing status to unsupervised clustering with the fuzzy c-means algorithm (45). We performed this approach, which complements the S-score analysis, to capture potentially important events that were happening in both differentiation paradigms but were not coordinated in time. The quantitative kinetic profiles partitioned into 10 clusters with distinct patterns of regulation after induction of hESC differentiation, providing an unbiased summation of the dynamics of phosphorylation (Fig. 7A). We then tested for overrepresented GO Biological Process terms in each of the clusters compared with the group of proteins that did not exhibit changes in phosphorylation status (Fig. 7B). Biological processes related to differentiation and tissue development, such as "heart development," "in utero embryonic development," "hemopoiesis," and "tissue development," emerged as overrepresented from this analysis in the treatment-specific groups. However, several categories associated with epigenetic regulation, such as "chromatin remodeling," "nucleosome organization," and "DNA methylation," appeared among the GO terms common for both treatments (Fig. 7B).

At least three distinct epigenetic or noncoding elements contribute to the control of hESC pluripotency and differentiation, and these involve chromatin reorganization, DNA methylation, and microRNAs (miRNAs) (46, 47). LIN28 is part of the miRNA-processing machinery with an established role in the maintenance of hESC pluripotency and in reprogramming fibroblasts to iPS cells (48). Two LIN28 proteins exist in mammals, LIN28A and LIN28B, and we identified four class 1 phosphorylation sites on LIN28A (Ser[3], Ser[120], Ser[200], and Thr[202]) and three on LIN28B (Ser[241], Ser[244], and Ser[247]) (table S2). Of the seven identified class 1 LIN28 phosphorylation sites, only two (Ser[200] and Thr[202] on LIN28A) have been reported previously (9). The abundance of LIN28A and LIN28B did not change in response to either differentiation protocol (table S1), but the phosphorylation status of Ser[200] and Thr[202] on LIN28A was changed in opposite directions after the two treatments (table S2). Phosphorylation of these sites was gradually decreased by NCM treatment and increased by PMA treatment, showing a peak by 1 hour. These data may reflect treatment-specific regulation of miRNA processing that contributes to differentiation.

Another pivotal event in hESC maintenance and differentiation is the generation of specific histone marks (46). Among the critical components in this process are the histone methyltransferase MLL complexes (46) and the polycomb repressor complex 2 (PRC2); the latter interacts with the DNMTs (49). We identified 21 class 1 phosphorylation sites on PRC2 members, of which two-thirds have not been previously described, including sites on the core components EED (Thr[115]), EZH2 (Thr[383]), and SUZ12 (Ser[539]). Similarly, we identified 42 class 1 phosphorylation sites on members of the MLL1-MLL and MLL3-MLL4 complexes, of which 18 had not been identified previously. Of the previously unreported sites that exhibited a change in phosphorylation status in response to both treatments were Ser[161], Ser[165], and Ser[167] (table S2), which are adjacent to the AT-hook DNA binding domain, and phosphorylation of these sites may affect the interaction of MLL with DNA.

The importance of DNA methylation for hESC differentiation is well recognized (46). DNA methylation is primarily performed by three proteins: DNA (cytosine-5) methyltransferase 1 (DNMT1), DNMT3A, and DNMT3B. These are crucial for proper embryonic development (50, 51). Although all three interact with each other (49), these enzymes belong to two functional classes. DNMT1 is a maintenance methyltransferase responsible for preserving the methylation pattern in the nascent DNA strand during replication (46), whereas DNMT3A and DNMT3B are responsible for de novo DNA methylation, thereby creating the somatic methylation pattern of the organism during embryogenesis (46, 50). The hESC transition from the pluripotent state toward differentiation is accompanied by silencing of OCT4, NANOG, and other pluripotency-associated genes by de novo DNA methylation of their promoters (46). Furthermore, treatment with DNMT inhibitors improves the efficiency of the reprogramming of somatic cells for generation of iPS cells (52). We observed extensive changes in the phosphorylation of sites in the N-terminal regions of all three DNMTs (Fig. 8A). Despite the importance of DNMTs for hESC differentiation, only limited information is available about the involvement of phosphorylation of the de novo DNMTs in this process. Only 2 phosphorylation sites have been reported for DNMT3A (9, 14) and 7 for DNMT3B (8, 9), whereas we identified 11 phosphorylation sites on DNMT3A and 34 on DNMT3B. All of the phosphorylation sites that exhibited changes during differentiation clustered in a region of ~250 residues near the N termini of these proteins. This region mediates the association among the three DNMTs, and it also mediates the transcriptional repression activity of DNMTs by recruiting several chromatin-remodeling proteins (49). Of the identified phosphorylation sites on DNMT3B, 14 exhibited a change in phosphorylation status after PMA treatment and 17 after NCM treatment, suggesting that the biological activity of these proteins or their interactions with binding partners may be regulated by phosphorylation during hESC differentiation. Indeed, we hypothesized that changes in the phosphorylation of DNMTs may provide a direct link between phosphorylation-dependent signaling and DNA methylation, which represses the expression of such master pluripotency markers as OCT4 and NANOG (53) through epigenetic modulation of gene expression.
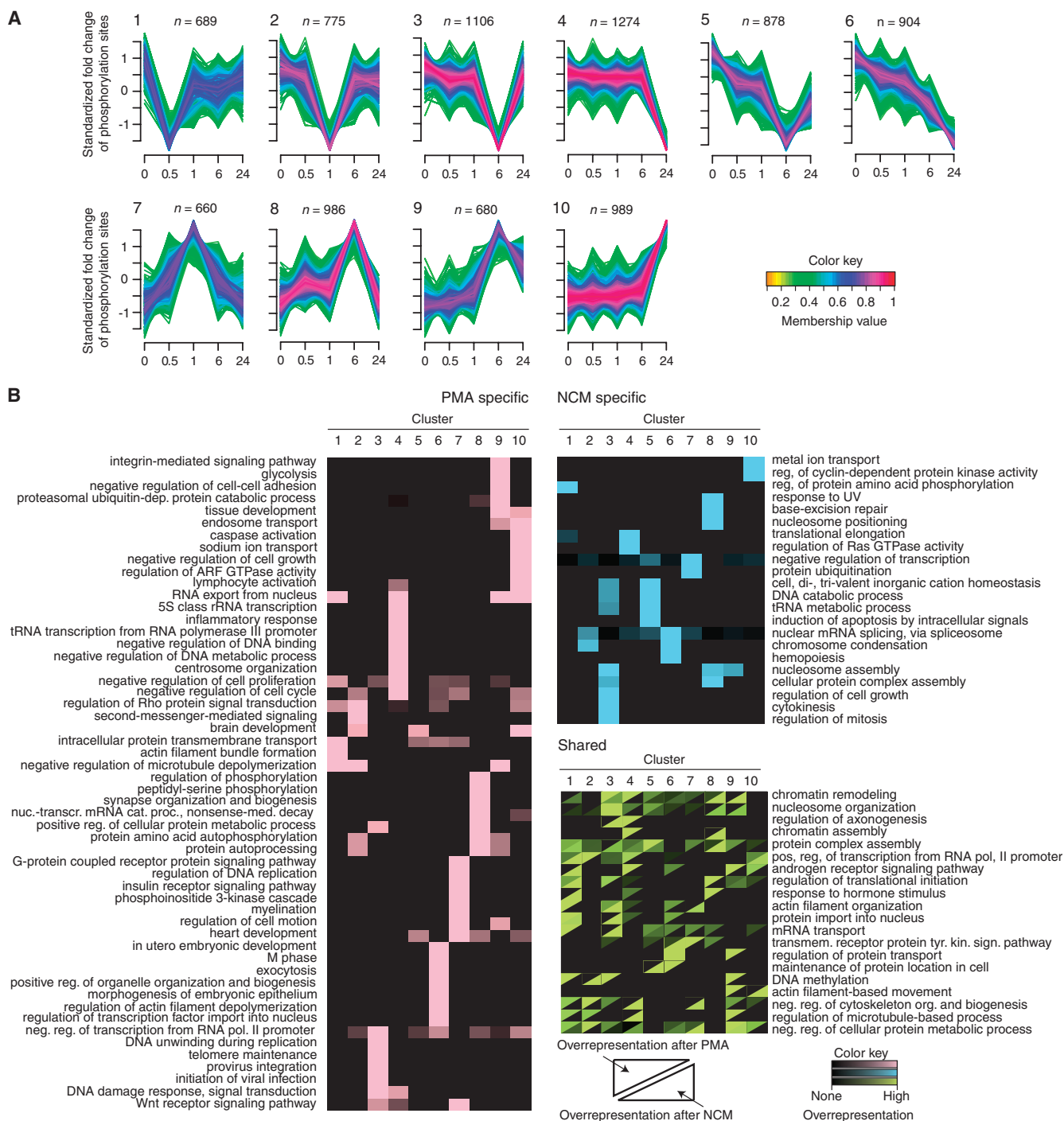
Fig. 7. Clustering of the hESC phosphoproteome by the Fuzzy *c*-means algorithm. (**A**) Phosphorylation sites that exhibited a change in status during the time course of either treatment were subjected to unsupervised clustering with the Fuzzy *c*-means algorithm. The ratio for each phosphorylation site was transformed and standardized (see Materials and Methods). Ten patterns of dynamic changes were evident, and the number of sites in each pattern is indicated by the *n* above each graph. Membership value represents how well the dynamic change of each phosphorylation site fits with the general profile of the cluster. (**B**) Phosphorylation sites in each cluster were sequentially tested against unchanging sites for significantly overrepresented GO Biological Process terms. Biological Process terms that were significantly enriched in each of the 10 clusters for PMA-specific, NCM-specific, and shared (common to both treatments) phosphorylation sites are summarized in the heat maps. GO terms use the following abbreviations: reg., regulation; neg., negative; pol., polymerase; nuc., nuclear; cat., catalytic; proc., processing; nonsense-med., nonsense-mediated; transmeb., transmembrane; tyr., tyrosine; kin., kinase; sign., signaling; transcr., transcription; pos., positive.

Interaction between DNMTs and the PAF1 complex upon induction of hESC differentiation

Because the N-terminal regions of DNMT3A and DNMT3B mediate protein-protein interactions, we examined whether the changes in phosphorylation of these regions, which accompanied the exit of hESCs from the pluripotent state, resulted in specific changes in the interactions of the DNMT complex. We used an unbiased proteomics approach combining SILAC labeling and affinity enrichment of DNMT3A and DNMT3B (54, 55) (fig. S7 and Materials and Methods). Two identical coimmunoprecipitation experiments were performed, one for DNMT3A and another for DNMT3B, with antibodies to the endogenous proteins. Most of the identified interaction partners that were specific to the early stages of hESC differentiation appeared associated with various aspects of RNA polymerase II (POL II)–mediated transcription, including three subunits of the POL II itself, POLR2H, POLR2C, and POLR2A (Fig. 8B) (see table S7 for a complete list of identified proteins). We observed an in-

creased interaction of DNMT3A and DNMT3B with all five members of the PAF1 complex (PAF1C)—RTF1, LEO1, CTR9, PAF1, and CDC73 (Fig. 8B)—and we confirmed the interactions for two of the members by reverse coimmunoprecipitation experiments (Fig. 8C). PAF1C has been implicated in a number of biological processes related to transcriptional regulation, including histone ubiquitylation, acetylation, and methylation (56), and here we connect this complex to DNA methylation. PAF1C binds to the promoters and regulates the expression of central pluripotency-associated genes, such as those encoding OCT4 and NANOG (57). Therefore, our finding that the DNMT complex associates specifically with PAF1C during early stages of hESC differentiation may provide a link to the silencing of these key regulators of pluripotency. The direct association between critical components of different types of epigenetic regulation supports the suggestion that several epigenetic processes, in combination with transcription factors, converge to maintain the pluripotent state of hESC and in turn specify differentiation (46).
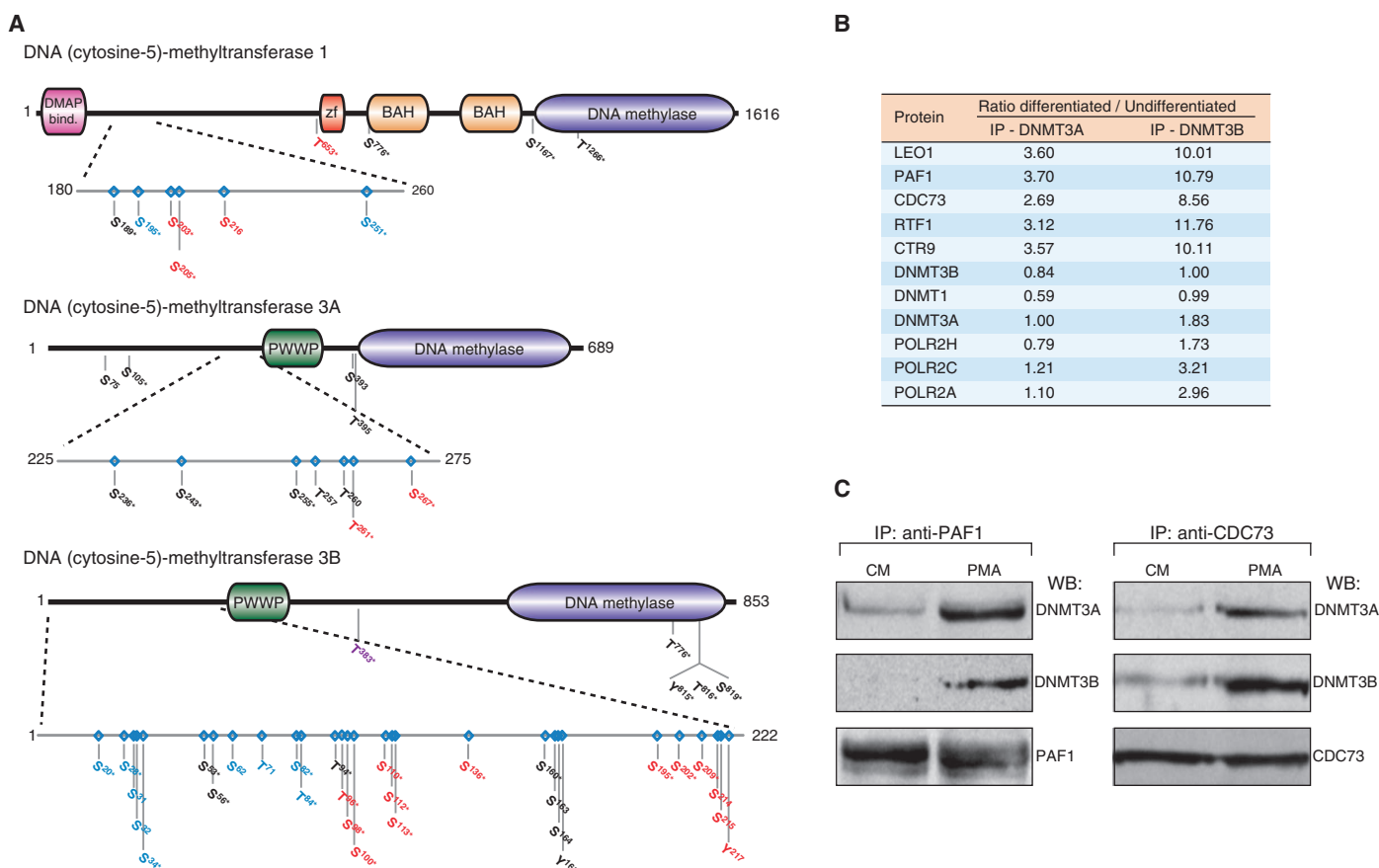

Fig. 8. Changes in the phosphorylation of DNMTs within the first 24 hours of hESC differentiation. (**A**) Identified phosphorylation sites on DNMT1, DNMT3A, and DNMT3B. Class 1 sites are indicated by asterisks (*), and coloring represents changes in phosphorylation status after PMA treatment (red), NCM treatment (blue), or both (purple). Black represents phosphorylation sites that do not exhibit change in phosphorylation during the differentiation period. DMAP bind., DMAP1-binding domain; zF, zinc finger CXXC domain; BAH, bromo adjacent homology domain; PWWP, PWWP domain (domain with conserved Pro-Trp-Trp-Pro motif).

(**B**) Search for DNMT binding partners by quantitative proteomics identifies specific interaction with the PAF1 complex in differentiating hESCs (see Materials and Methods and table S7). (**C**) Western blots of co-immunoprecipitation experiments show that PMA treatment for 6 hours promotes the interaction of DNMT3A and DNMT3B with the PAF1 complex members PAF1 and CDC73. IP, immunoprecipitation; WB, Western blotting; POLR2H, POLR2C, and POLR2A, constituents of the RNA POL II complex; RTF1, CTR9, PAF1, LEO1, and CDC73, constituents of the PAF1 complex.

DISCUSSION

We applied global quantitative proteomics and phosphoproteomics to characterize the phosphorylation dynamics and changes in protein abundance after induction of hESC differentiation with two distinct treatments. We developed and applied bioinformatics approaches to extract information regarding the regulation of the complex processes controlling hESC pluripotency and self-renewal and differentiation. Here, we have looked at phosphorylation, but an even more complex regulatory map would likely emerge if one were to take into account other posttranslational modifications like ubiquitylation, acetylation, or methylation, highlighting that protein abundance or gene expression changes are not only the end point in many regulatory circuits.

Comparing our results with those of Van Hoof et al., who applied quantitative proteomics to study hESC differentiation (of a different hESC cell line from the ones we studied) induced by BMP4 (9), showed good correlation between the two studies regarding some of the major findings. For example, both studies reported three serine phosphorylation sites on SOX2 and, in common with our data after PMA treatment, reported that there were no changes in the phosphorylation of these sites in the differentiating cells. Van Hoof et al. further demonstrated that these sites were important for SUMOylation of SOX2 because the replacement of these three serines with aspartic acid residues, thus mimicking constitutive phosphorylation, caused an increase in SOX2 SUMOylation. Van Hoof et al. did not quantitatively assess protein abundance but showed, by Western blotting, decreased SOX2 abundance in response to BMP4, which is similar to the decrease in the SOX2 abundance that we observed after PMA treatment. In contrast, we found that after NCM treatment, phosphorylation of SOX2 in the SUMOylation motif was decreased, and the abundance of SOX2 was increased. Although speculative, it seems likely that the increased amount of SOX2 after NCM treatment might be due to decreased phosphorylation in the SUMOylation motif and concomitant reduced degradation.

In further correlation with our results, Van Hoof et al. found an increase of the inhibitory phosphorylation of Tyr[15] on CDK2 and CDC2 after induction of differentiation. The same site was also identified by Brill et al. (8), but their data indicate that this phosphorylation might be decreased in hESCs differentiated by retinoic acid treatment. This may suggest that the required decrease in cell cycle progression is achieved through a distinct route, apart from CDK2 and CDC2 inactivation, when a direct modulation of gene expression by retinoic acid is used to induce hESC differentiation.

This global data set provides a wealth of candidates for hypothesis generation and subsequent "low-throughput" mechanism–focused studies to clarify the regulatory networks that control stem cell character and fate. By analyzing two different nondirected differentiation paradigms, we identified a core hESC phosphoproteome, which included the identification of putative regulatory phosphorylation events on DNMTs, transcription factors, and kinases, and provided insight into potential mechanisms controlling the cell cycle in hESCs during early differentiation events. We expect that this rich foundation of information associated with the processes underlying hESC "stemness" and differentiation will open entirely new leads and horizons for research in stem cell biology to clarify the regulatory events controlling hESC maintenance and differentiation.

MATERIALS AND METHODS

Cell culture

Both hESC lines used in this study—Odense-3 and HUES9—were maintained on Matrigel (BD Biosciences) in CM, which was medium conditioned by MEFs, essentially as described in "Protocols for Maintenance of Human Embryonic Stem Cells in Feeder Free Conditions" from Geron Corporation (http://www.geron.com). Briefly, the plastic surface of the cell culture dishes was coated with 1 ml of Matrigel (prediluted to ~0.3 mg/ml) per 10 cm^2. The MEFs CM was prepared as follows: hESC medium consisting of KnockOut Dulbecco's modified Eagle's medium (DMEM), 15% knockout serum replacement, 1% penicillin/streptomycin, 1% nonessential amino acids (NEAAs), 1% glutamax (all from Invitrogen), 0.5% human serum albumin, 0.1 mM β-mercaptoethanol (Sigma), and FGF2 (10 ng/ml; Invitrogen) was conditioned for 24 hours on irradiated MEFs plated at a density of 56,000 cells/cm^2 (0.3 ml of medium per square centimeter of MEFs) and frozen at −20°C. Immediately before use, CM was sterile-filtered and supplemented with FGF2 (10 ng/ml). For in vitro hESC differentiation experiments, cells were induced to differentiate by replacing the CM with NCM not containing FGF2 or by supplementing the CM with 50 nM PMA. Cell populations were cultured in this medium for 30 min, 1 hour, 6 hours, or 24 hours for the proteomics experiments and Western blotting analysis, and cells were harvested by trypsinization, washed twice with ice-cold phosphate-buffered saline (PBS), and frozen as pellets at −80°C. Cells for QPCR analysis were treated for 24 and 48 hours, as indicated in Fig. 1B and fig. S1, and were harvested directly into Trizol.

SILAC labeling and sample preparation

For the large-scale PMA and NCM experiments, CM was dialyzed three times against PBS using a membrane with a 3500-dalton cutoff, and once against DMEM containing 1% penicillin/streptomycin, 1% NEAA, glucose (4.5 g/liter), and 1% glutamax, but lacking arginine and lysine. To generate light, medium, and heavy SILAC-labeling media, we supplemented the medium with SILAC amino acids L-arginine (Arg0) and L-lysine (Lys0), L-lysine–^{2}H$_4$ (Lys4), and L-arginine–U-^{13}C$_6$ (Arg6), or L-lysine–U-^{13}C$_6$-^{15}N$_2$ (Lys8) and L-arginine–U-^{13}C$_6$-^{15}N$_4$ (Arg10) at a concentration of 28 mg/liter for the arginine and 146 mg/liter for the lysine amino acids. For coimmunoprecipitation experiments with DNMTs and for biological replica experiments, SILAC-labeling medium was prepared essentially as described (58). Briefly, hESC medium [consisting of DMEM lacking arginine and lysine, 15% knockout serum replacement, 1% penicillin/streptomycin, 1% NEAA, 1% glutamax (all from Invitrogen), 0.5% human serum albumin, 0.1 mM β-mercaptoethanol (Sigma), glucose (4.5 g/liter), L-proline (500 mg/liter), and FGF2 (10 ng/ml) (Invitrogen)] was supplemented with Arg0, Arg6, or Arg10 (84 mg/liter), and the corresponding Lys0, Lys4, or Lys8 (146 mg/liter) was conditioned on MEFs as described above and sterile-filtered, and FGF2 (10 ng/ml) was added immediately before use. For all experiments, hESCs were grown in the labeling media for five passages before further manipulation of the cells to achieve complete SILAC incorporation.

HUES9 cells were SILAC-labeled as described (19) with Arg0 and Lys0, Arg6 and Lys4, or Arg10 and Lys8 to achieve double-triple SILAC encoding (59). At 80 to 90% confluence, cells were either treated with 50 nM PMA or cultured in NCM. Differentially SILAC-labeled cells (5 × 10^7 cells per condition) were subjected to these treatments for 0 hour (untreated, Arg0/Lys0), ½ and 1 hour (Arg6/Lys4), or 6 and 24 hours (Arg10/Lys8). Differentially labeled cells were mixed 0 min/½ hour/6 hours and 0 min/1 hour/24 hours to generate four SILAC experiments (two treatments and two time point sets). Mixed cells were lysed in buffer containing 1% NP-40, 0.1% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 50 mM tris (pH 7.5), phosphatase inhibitors (1 mM sodium orthovanadate, 5 mM sodium fluorate, 5 mM β-glycerophosphate), and protease inhibitors (Complete tablets, Roche Diagnostics). Lysates were centrifuged at 4°C for 10 min at 15,000g, producing a soluble fraction and a pellet

containing chromatin and other insoluble material. Each of the pellets from the four SILAC experiments was resuspended in 8 M urea containing 1% N-octylglucoside, phosphatase inhibitors, and benzonase. Proteins from the soluble fractions were precipitated overnight at −20°C by four volumes of ice-cold acetone, pelleted by centrifugation for 2 min at 2000g, and redissolved in 8 M urea containing 1% N-octylglucoside and phosphatase inhibitors.

Cell lysate proteins from all samples prepared as described above were then reduced for 30 min by addition of dithiothreitol to a final concentration of 10 mM and alkylated with 55 mM iodoacetamide for 20 min in the dark at room temperature. Subsequently, a 100-μg aliquot was taken from each sample, separated by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) on a 4 to 12% NuPAGE gel (Invitrogen), and visualized by Colloidal Blue staining (Invitrogen) (Fig. 1C). Gel lanes were cut into 20 bands each, in-gel digested with trypsin (12.5 ng/μl; sequencing grade, Promega) overnight at 37°C (60), and used for the quantitative proteome analysis. The remainders of each sample were digested in solution at room temperature with endoproteinase Lys-C (Wako) for 4 hours followed by trypsin digestion overnight (16). Proteases were added at a ratio of 1:160 and 1:200 with respect to the total amount of protein for Lys-C and trypsin, respectively. After digestion, trifluoroacetic acid was added to the peptide solutions to a final concentration of 0.1%, and the solutions were clarified by centrifugation at 20,000g for 5 min. Peptide mixtures were then fractionated by SCX (16), enriched for phosphorylated peptides by TiO₂ (61, 62), and used for the quantitative phosphoproteomics analysis (see Supplementary Methods for additional details about phosphopeptide enrichment).

LC-MS/MS and data processing

All samples were analyzed by nanoscale C18 high-performance liquid chromatography (HPLC) coupled online to a mass spectrometer (16). For C18 HPLC, an EASY-nLC (Proxeon), Agilent 1100, or Agilent 1200 system was used to separate the peptides over a 90-min linear gradient from 8 to 24% acetonitrile (ACN) in 0.5% acetic acid with a flow rate of 250 nl/min. Mass spectrometry was performed in positive ion mode on an LTQ-FT Ultra, LTQ-Orbitrap, or LTQ-Orbitrap XL (all Thermo Scientific) mass spectrometer equipped with a nanoelectrospray source (Proxeon) using data-dependent acquisition to isolate and fragment the 10 most intense ions (see Supplementary Methods for additional details).

The MaxQuant software suite (26) for proteomics data processing was used to generate peak lists, which were searched against a concatenated target-decoy database with Mascot (http://www.matrixscience.com). Peptide identifications were also quantified and filtered for less than 1% false positives, and phosphorylation sites were localized and scored in MaxQuant. The identified 14,865 phosphopeptides contained 23,522 phosphorylation sites, of which 15,004 could be assigned to a specific residue within the peptide with a probability of at least 0.75 (class 1 sites; median localization probability 0.996) (16). However, the probability that the peptide was correctly identified and phosphorylated was still larger than 99% for all of the peptides covering the 23,522 phosphorylation sites. Furthermore, 71% of the identified phosphopeptides were sequenced by MS/MS more than once, and 58% of the reported phosphorylation site ratios are the result of more than one quantification event.

Phosphorylation sites were considered as changing during differentiation if they displayed at least a twofold change in their corresponding SILAC ratios. The legitimacy of the twofold change threshold for phosphorylation sites was estimated using the cases where the same phosphopeptide was identified in two distinct forms: one with and one without

additional modifications such as oxidized methionine. Comparing ratios of the phosphopeptides with these common modifications that are induced during sample preparation, we determined an average relative SD (RSD) of 13.4% for quantification of phosphopeptides. Even allowing that peptides for which a modified form is also detected are somewhat more abundant, the twofold regulation threshold is on average several SDs away from the mean ratio of 1, which corresponds to no change. In addition, only 1.5% of all phosphopeptide ratios displayed an RSD above 50%.

Sequence homology analysis of kinases and transcription factors

To group kinases and transcription factors in accordance with their sequence homology, we chose an approach essentially identical to the one described in (38). To analyze kinases, we extracted FASTA sequences of the kinase domain from all identified kinases from the KinBase resource (http://www.kinase.com) and aligned the sequences in ClustalX2 (63) using the default parameters for multiple alignment and bootstrapping. For visualization of the alignment, a phylogenetic tree was calculated with the neighbor-joining algorithm, exported in Phylip format, and uploaded to the Interactive Tree of Life (iTOL) tool (64). To illustrate the regulation at the protein and phosphorylation level of the kinases, we used the heat map and multivalue bar chart display options.

To perform the analysis for transcription factors, we used the DNA binding domains for all identified transcription factors. Whereas the kinase domains of kinases are generally well established, the DNA binding domain of many transcription factors has not been characterized. Therefore, we were required to adopt a more elaborate strategy. First, we extracted the identity and sequence of the DNA binding domains for the identified transcription factors from the Pfam database (65) in cases where this information was available. To construct a "template DNA binding domain," we aligned the amino acid sequence of DNA binding domains of the families HLH, HMG-box, bZIP, Forkhead, and Homeobox, corresponding to DNA binding domains from 50 proteins. Selecting the profile mode in ClustalX2, we then added the sequences of DNA binding domains not contained in the template-building domains and aligned these to the template profile. As the final alignment step, we added the nontemplate domain profile to the template profile and aligned the entire sequence of the transcription factors with no described DNA binding domain. To visualize the result, we combined all profiles and constructed a phylogenetic tree as described for the kinases.

Linear kinase motif analysis

We assessed the prevalence of particular amino acids in the proximity of phosphorylation sites by looking for specific kinase phosphorylation motifs. To avoid clouding of motifs with ambiguously assigned phosphorylation sites, we used only sites with a localization probability >0.75 (class 1) for the analysis. To search for overrepresented motifs, we used sequence windows of ±6 residues adjacent to the phosphorylation sites. Sequence windows to serve as the background population were created from the experimentally identified Ser and Thr residues in unphosphorylated peptides. Sequence windows also observed for phosphorylated sites were excluded from the background population, resulting in a final foreground population of 14,563 sequence windows and a background population of 126,368 sequence windows. A partially degenerate amino acid code was used, where the residues Arg and Lys, Glu and Asp, Gln and Asn, as well as Ser and Thr that were not part of a phosphorylation site, were grouped.

Motif X (35) was used to test for significantly overrepresented motifs, with the following parameters: P value threshold ≤1 × 10⁻⁶, minimum fold

increase ≥2, and an occurrence limit of at least ~1% of the sequence windows submitted to the analysis (n_{min} = 124 for pSer and 26 for pThr). The NetworKIN algorithm (*36*) version 2.0 was used to predict the kinases responsible for phosphorylating the sites matching the identified phosphorylation motifs. For each motif, the frequency of potentially responsible kinases was calculated, and kinases predicted to phosphorylate less than 5% of the sites matching the motif were discarded. For the remaining kinases, the frequency was normalized to 1 for the most frequently occurring kinase for each motif.

To evaluate the overall regulation of the phosphorylation sites matching the identified linear motifs, we extracted all regulated sites matching each motif and used Student's *t* test to test whether the mean of the log$_2$ ratios at any time point after both treatments was different from 0, at significance level 0.01. Motifs with no significant overall regulation were assigned a mean regulation value of 0, and the remaining values where normalized by assigning the maximum increase in phosphorylation a value of 1 and the maximum decrease in phosphorylation a value of −1.

Profile correlation of phosphorylation sites (*S*-score)

To quantitatively describe the resemblance between the temporal profiles observed after subjecting the cells to the two treatments and to identify the core hESCs phosphoproteome, we defined a similarity score (*S*-score). The *S*-score was based on the Pearson's product moment correlation coefficient, which gives a quantitative estimate of the association between the profiles obtained after the two treatments. The Pearson *P* value (*P* component) was calculated as

$$r = 1 - \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}$$

with *X* and *Y* being the SILAC ratios after the two treatments, $\bar{x}$ and $\bar{y}$ the means of *X* and *Y*, and σ_X and σ_X the SDs of *X* and *Y*. From *r*, a test statistic *T* can be approximated as

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \approx t_{n-2}$$

from which a probability of positive correlation can be calculated from the one-tailed *t* distribution with *n* − 2 degrees of freedom.

The Pearson coefficient is, however, not sufficient to evaluate the similarity between the profiles because the correlation factor in many cases will be different from 1. To account for this, we introduced a second parameter, a magnitude component (*M* component), which quantified the resemblance of the degree of regulation, without taking the pattern of the profiles into account. To quantify the magnitude of regulation over the time course, we log$_2$-transformed the ratios for each site and calculated the area defined by the absolute values of the transformed ratios and the *x* axis with the trapezoidal rule to approximate the definite integral as

$$\text{AUC} = \sum\limits_{i=2}^{n} \frac{y_i - y_{i-1}}{2}(x_i - x_{i-1})$$

with *Y* being the absolute value of the log$_2$-transformed SILAC ratios and *X* the position in the time course, in this case, 1 to 5. We chose to use the position in the time course (1, 2, 3, 4, 5) and not the actual temporal values (0, ½, 1, 6, 24) because the use of the actual values

would ascribe an unjustified importance to the 24-hour time point and render the early time points nearly insignificant.

To quantify the similarity between the extents of regulation after the two treatments, we calculated the *M* score as

$$M \text{ component} = \frac{\text{AUC}_1 - \text{AUC}_2}{\text{AUC}_1 + \text{AUC}_2}$$

with AUC$_1$ and AUC$_2$ being the areas under the curve for treatments 1 and 2.

Finally the *S*-score was calculated as

$$S\text{-score} = -10 \log_{10}(P \text{ component} \times M \text{ component})$$

and thus describes both the resemblance of the patterns of regulation and the resemblance between the degrees of regulation in the range from zero to infinity. Although the *S*-score provided a tool for quantifying the similarity of regulation and was based on the Pearson correlation coefficient, the obtained score cannot be interpreted as a statistical description, primarily due to two issues. First, the Pearson correlation coefficient assumption that the data are sampled from a normal distribution will not be met in all cases, and, second, the Pearson correlation coefficient is adjusted by the *M* score, which is not intended to be a statistical value. Thus, an *S*-score corresponding to a product of the *P* component and *M* component equal to 0.05 should not be interpreted as significant similar regulation at level 0.05. Regardless, the *S*-score provides a measurement of similarity, as seen in fig. S6, where representative profiles after the two treatments with increasing *S*-score are presented, and a clear intuitive increase in similarity is observed as the *S*-score increases.

Clustering of phosphorylation sites

Phosphorylation sites and protein identifications with quantitative data for all time points were extracted from the MaxQuant output and partitioned into a constant and a regulated group, based on criteria for greater than twofold change of the phosphorylation status of the site in at least one time point. Time course data for the changing phosphorylation sites were transformed to obtain symmetric values for equal increases and decreases in phosphorylation centered on zero by applying the equation, ratio$_{transformed}$ = (−1/SILAC ratio) + 1, for SILAC ratios <1, and subtracting 1 from SILAC ratios >1. The resulting transformed ratios were standardized by dividing the peptide SILAC ratio for each time point by the SD for this peptide and then subjecting to unsupervised clustering with the fuzzy *c*-means algorithm as implemented in the Mfuzz package (*45*, *66*) with a fuzzification parameter of 2 and 10 centers.

Identification of DNMT3A- and DNMT3B-interacting proteins

For coimmunoprecipitation experiments of DNMT3A and DNMT3B followed by LC-MS/MS analysis, HUES9 cells were grown on Matrigel in Arg0/Lys0- or Arg6/Lys4-labeling SILAC media as described above, and for Western blotting analysis, cells were grown in regular CM. To induce differentiation, we supplemented the Arg6/Lys4-labeled cells with 50 nM PMA for 6 hours before harvesting, whereas the Arg0/Lys0 cells were left untreated. For making nuclear extracts, cells were trypsinized, washed extensively in ice-cold PBS, and resuspended in 5× cell pellet volumes of ice-cold buffer A [10 mM Hepes KOH (pH 8.0), 1.5 mM MgCl$_2$, 10 mM KCl, 0.2% NP-40, and protease inhibitors]. The cell suspension was incubated on ice for 5 min and centrifuged at 2500*g* for 15 min. The pellet containing the crude nuclei was resuspended in five

excess volumes of buffer B (buffer A with 420 mM NaCl) and sonicated for 10 s at low power with a Misonix sonicator. Benzonase (Roche) was added to the lysate to a final concentration of 1000 U/ml, incubated on ice for 10 min, and sonicated as described above. Subsequently, the lysates were cleared by centrifugation and the supernatants were supplemented with 10% glycerol.

Antibodies to DNMT3A, DNMT3B, PAF1, and CDC73 were from Santa Cruz Biotechnology. For identification of DNMT3A-interacting proteins, 25 μg of an antibody that recognizes DNMT3A was bound to 15 μl of protein A/G beads for 1 hour at 4°C. Protein A/G beads with bound antibodies were then washed in buffer B and incubated with the corresponding Arg6/Lys4 or Arg0/Lys0 nuclear extracts at 4°C with end-over-end rotation. Immunoprecipitates from Arg0/Lys0- and Arg6/Lys4-labeled cells were washed five times with 1 ml of buffer B and mixed. Immunoprecipitated complexes were separated on 4 to 12% NuPAGE gels (Invitrogen) and subjected to in-gel digestion and LC-MS/MS analysis as described above. For identification of DNMT3B interaction partners, an identical coimmunoprecipitation experiment was performed with DNMT3B-specific antibodies. For reciprocal coimmunoprecipitation experiments of PAF1 and CDC73, the procedure and conditions were generally identical except that Western blotting was used for visualization of the coimmunoprecipitating DNMT3A and DNMT3B (as indicated on Fig. 8C). Western blotting experiments were performed essentially as described before (67).

Real-time QPCR

The real-time QPCR analyses were performed as described previously (68). Briefly, total RNA was purified with Trizol (Invitrogen) according to the manufacturer's instructions. One microgram of each RNA sample was treated with RQ1 RNase-Free DNase (Promega, M6101) to remove the presence of any genomic DNA before reverse transcription–PCR. Inactivation of the RQ1 DNase was done by heating samples at 75°C. Complementary DNA was synthesized with Random Hexamer primers (GE Healthcare) and M-MLV Reverse Transcriptase (Invitrogen). QPCR was performed with the SYBR Green JumpStart Taq ReadyMix (Sigma) and the Mx3000P QPCR instrument (Stratagene). To verify the specificity of a PCR product, we made a dissociation curve analysis at the end of each PCR run and confirmed amplification of a single product by ethidium bromide–stained 1.5% agarose gels. Expression of the target genes was all normalized to the expression of actin. Primers used for QPCR are provided in table S8.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/4/164/rs3/DC1
Materials and Methods
Fig. S1. Cells treated with NCM or PMA differentiate in a nondirected manner.
Fig. S2. General properties of the data sets.
Fig. S3. Frequency of protein and phosphorylation site identifications with selected GO Molecular Function and Cellular Component terms.
Fig. S4. Dynamic changes in phosphorylation organized by phosphorylation motif.
Fig. S5. Double-phosphorylation motifs.
Fig. S6. The S-score algorithm.
Fig. S7. Quantitative proteomics approach for identification of DNMT3A- and DNMT3B-interacting proteins.
Table S1. Protein identifications.
Table S2. Phosphorylation site identifications.
Table S3. Protein and phosphorylation site identifications for 30-min and 6-hour biological replicas.
Table S4. Identified transcription factors and transcription factor phosphorylation sites.
Table S5. Linear phosphorylation motifs.
Table S6. Identified kinases and kinase phosphorylation sites.
Table S7. Protein identifications from DNMT3A and DNMT3B immunoprecipitations.
Table S8. Primers used for real-time QPCR.
References

REFERENCES AND NOTES

1. R. L. Gardner, R. S. Beddington, Multi-lineage 'stem' cells in the mammalian embryo. *J. Cell Sci. Suppl.* **10**, 11–27 (1988).
2. S. D. Fouse, Y. Shen, M. Pellegrini, S. Cole, A. Meissner, L. Van Neste, R. Jaenisch, G. Fan, Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).
3. P. H. Lerou, G. Q. Daley, Therapeutic potential of embryonic stem cells. *Blood Rev.* **19**, 321–331 (2005).
4. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
5. M. Wernig, A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B. E. Bernstein, R. Jaenisch, In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
6. S. Assou, T. Le Carrour, S. Tondeur, S. Strom, A. Gabelle, S. Marty, L. Nadal, V. Pantesco, T. Reme, J. P. Hugnot, S. Gasca, O. Hovatta, S. Hamamah, B. Klein, J. De Vos, A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
7. D. L. Swaney, C. D. Wenger, J. A. Thomson, J. J. Coon, Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 995–1000 (2009).
8. L. M. Brill, W. Xiong, K. B. Lee, S. B. Ficarro, A. Crain, Y. Xu, A. Terskikh, E. Y. Snyder, S. Ding, Phosphoproteomic analysis of human embryonic stem cells. *Cell Stem Cell* **5**, 204–213 (2009).
9. D. Van Hoof, J. Munoz, S. R. Braam, M. W. Pinkse, R. Linding, A. J. Heck, C. L. Mummery, J. Krijgsveld, Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* **5**, 214–226 (2009).
10. B. Macek, M. Mann, J. V. Olsen, Global and site-specific quantitative phosphoproteomics: Principles and applications. *Annu. Rev. Pharmacol. Toxicol.* **49**, 199–221 (2009).
11. S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, M. Mann, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
12. J. Dengjel, V. Akimov, J. V. Olsen, J. Bunkenborg, M. Mann, B. Blagoev, J. S. Andersen, Quantitative proteomic assessment of very early cellular signaling events. *Nat. Biotechnol.* **25**, 566–568 (2007).
13. J. Dengjel, I. Kratchmarova, B. Blagoev, Receptor tyrosine kinase signaling: A view from quantitative proteomics. *Mol. Biosyst.* **5**, 1112–1121 (2009).
14. N. Dephoure, C. Zhou, J. Villen, S. A. Beausoleil, C. E. Bakalarski, S. J. Elledge, S. P. Gygi, A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10762–10767 (2008).
15. M. Hilger, T. Bonaldi, F. Gnad, M. Mann, Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol. Cell. Proteomics* **8**, 1908–1920 (2009).
16. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
17. M. J. Clemens, I. Trayner, J. Menaya, The role of protein kinase C isoenzymes in the regulation of cell proliferation and differentiation. *J. Cell Sci.* **103**, 881–887 (1992).
18. D. Phanstiel, J. Brumbaugh, W. T. Berggren, K. Conard, X. Feng, M. E. Levenstein, G. C. McAlister, J. A. Thomson, J. J. Coon, Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4093–4098 (2008).
19. T. A. Prokhorova, K. T. Rigbolt, P. T. Johansen, J. Henningsen, I. Kratchmarova, M. Kassem, B. Blagoev, Stable isotope labeling by amino acids in cell culture (SILAC) and quantitative comparison of the membrane proteomes of self-renewing and differentiating human embryonic stem cells. *Mol. Cell. Proteomics* **8**, 959–970 (2009).
20. C. Xu, M. S. Inokuma, J. Denham, K. Golds, P. Kundu, J. D. Gold, M. K. Carpenter, Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* **19**, 971–974 (2001).
21. I. Chambers, S. R. Tomlinson, The transcriptional foundation of pluripotency. *Development* **136**, 2311–2322 (2009).
22. S. Ohtsuka, S. Dalton, Molecular and biological properties of pluripotent embryonic stem cells. *Gene Ther.* **15**, 74–81 (2008).
23. C. A. Cowan, I. Klimanskaya, J. McMahon, J. Atienza, J. Witmyer, J. P. Zucker, S. Wang, C. C. Morton, A. P. McMahon, D. Powers, D. A. Melton, Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**, 1353–1356 (2004).
24. K. Osafune, L. Caron, M. Borowiak, R. J. Martinez, C. S. Fitz-Gerald, Y. Sato, C. A. Cowan, K. R. Chien, D. A. Melton, Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* **26**, 313–315 (2008).
25. B. Blagoev, M. Mann, Quantitative proteomics to study mitogen-activated protein kinases. *Methods* **40**, 243–250 (2006).
26. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

27. J. V. Olsen, M. Vermeulen, A. Santamaria, C. Kumar, M. L. Miller, L. J. Jensen, F. Gnad, J. Cox, T. S. Jensen, E. A. Nigg, S. Brunak, M. Mann, Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3 (2010).

28. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

29. F. Gnad, S. Ren, J. Cox, J. V. Olsen, B. Macek, M. Oroshi, M. Mann, PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8**, R250 (2007).

30. S. M. Kooistra, V. van den Boom, R. P. Thummer, F. Johannes, R. Wardenaar, B. M. Tesson, L. M. Veenhoff, F. Fusetti, L. P. O'Neill, B. M. Turner, G. de Haan, B. J. Eggen, Undifferentiated embryonic cell transcription factor 1 regulates ESC chromatin organization and gene expression. *Stem Cells* **28**, 1703–1714 (2010).

31. M. Maruyama, T. Ichisaka, M. Nakagawa, S. Yamanaka, Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J. Biol. Chem.* **280**, 24371–24379 (2005).

32. S. H. Park, H. Sakamoto, K. Tsuji-Tamura, T. Furuyama, M. Ogawa, Foxo1 is essential for in vitro vascular formation from embryonic stem cells. *Biochem. Biophys. Res. Commun.* **390**, 861–866 (2009).

33. J. Yang, C. Gao, L. Chai, Y. Ma, A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS One* **5**, e10766 (2010).

34. V. Episkopou, SOX2 functions in adult neural stem cells. *Trends Neurosci.* **28**, 219–221 (2005).

35. D. Schwartz, S. P. Gygi, An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398 (2005).

36. R. Linding, L. J. Jensen, A. Pasculescu, M. Olhovsky, K. Colwill, P. Bork, M. B. Yaffe, T. Pawson, NetworKIN: A resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 (2008).

37. J. Villén, S. A. Beausoleil, S. A. Gerber, S. P. Gygi, Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1488–1493 (2007).

38. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).

39. X. Zhang, I. Neganova, S. Przyborski, C. Yang, M. Cooke, S. P. Atkinson, G. Anyfantis, S. Fenyk, W. N. Keith, S. F. Hoare, O. Hughes, T. Strachan, M. Stojkovic, P. W. Hinds, L. Armstrong, M. Lako, A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J. Cell Biol.* **184**, 67–82 (2009).

40. Y. Gu, J. Rosenblatt, D. O. Morgan, Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15. *EMBO J.* **11**, 3995–4005 (1992).

41. C. H. McGowan, P. Russell, Human Wee1 kinase inhibits cell division by phosphorylating p34cdc2 exclusively on Tyr15. *EMBO J.* **12**, 75–85 (1993).

42. X. Chen, E. M. Whitney, S. Y. Gao, V. W. Yang, Transcriptional profiling of Krüppel-like factor 4 reveals a function in cell cycle regulation and epithelial differentiation. *J. Mol. Biol.* **326**, 665–677 (2003).

43. E. M. Klenova, I. V. Chernukhin, A. El-Kady, R. E. Lee, E. M. Pugacheva, D. I. Loukinov, G. H. Goodwin, D. Delgado, G. N. Filippova, J. León, H. C. Morse III, P. E. Neiman, V. V. Lobanenkov, Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* **21**, 2221–2234 (2001).

44. Y. Zhao, X. Yin, H. Qin, F. Zhu, H. Liu, W. Yang, Q. Zhang, C. Xiang, P. Hou, Z. Song, Y. Liu, J. Yong, P. Zhang, J. Cai, M. Liu, H. Li, Y. Li, X. Qu, K. Cui, W. Zhang, T. Xiang, Y. Wu, Y. Zhao, C. Liu, C. Yu, K. Yuan, J. Lou, M. Ding, H. Deng, Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* **3**, 475–479 (2008).

45. L. Kumar, M. E. Futschik, Mfuzz: A software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).

46. M. Bibikova, L. C. Laurent, B. Ren, J. F. Loring, J. B. Fan, Unraveling epigenetic regulation in embryonic stem cells. *Cell Stem Cell* **2**, 123–134 (2008).

47. V. K. Gangaraju, H. Lin, MicroRNAs: Key regulators of stem cells. *Nat. Rev. Mol. Cell Biol.* **10**, 116–125 (2009).

48. S. R. Viswanathan, G. Q. Daley, Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445–449 (2010).

49. G. D. Kim, J. Ni, N. Kelesoglu, R. J. Roberts, S. Pradhan, Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J.* **21**, 4183–4195 (2002).

50. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).

51. J. A. Yoder, N. S. Soman, G. L. Verdine, T. H. Bestor, DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J. Mol. Biol.* **270**, 385–395 (1997).

52. T. S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, A. Meissner, Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).

53. M. Imamura, K. Miura, K. Iwabuchi, T. Ichisaka, M. Nakagawa, J. Lee, M. Kanatsu-Shinohara, T. Shinohara, S. Yamanaka, Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6**, 34 (2006).

54. B. Blagoev, I. Kratchmarova, S. E. Ong, M. Nielsen, L. J. Foster, M. Mann, A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315–318 (2003).

55. J. Dengjel, I. Kratchmarova, B. Blagoev, Mapping protein-protein interactions by quantitative proteomics. *Methods Mol. Biol.* **658**, 267–278 (2010).

56. J. Kim, M. Guermah, R. G. Roeder, The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).

57. L. Ding, M. Paszkowski-Rogacz, A. Nitzsche, M. M. Slabicki, A. K. Heninger, I. de Vries, R. Kittler, M. Junqueira, A. Shevchenko, H. Schulz, N. Hubner, M. X. Doss, A. Sachinidis, J. Hescheler, R. Iacone, K. Anastassiadis, A. F. Stewart, M. T. Pisabarro, A. Caldarelli, I. Poser, M. Theis, F. Buchholz, A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4**, 403–415 (2009).

58. S. C. Bendall, C. Hughes, M. H. Stewart, B. Doble, M. Bhatia, G. A. Lajoie, Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell. Proteomics* **7**, 1587–1597 (2008).

59. K. T. Rigbolt, B. Blagoev, Proteome-wide quantitation by SILAC. *Methods Mol. Biol.* **658**, 187–204 (2010).

60. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).

61. M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, T. J. Jorgensen, Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886 (2005).

62. M. W. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. Heck, Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943 (2004).

63. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

64. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

65. R. D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010).

66. M. E. Futschik, B. Carlisle, Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol.* **3**, 965–988 (2005).

67. I. Kratchmarova, B. Blagoev, M. Haack-Sorensen, M. Kassem, M. Mann, Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472–1477 (2005).

68. J. Henningsen, K. T. Rigbolt, B. Blagoev, B. K. Pedersen, I. Kratchmarova, Dynamics of the skeletal muscle secretome during myoblast differentiation. *Mol. Cell. Proteomics* **9**, 2482–2496 (2010).