Regulation of transcriptional elongation in pluripotency and cell differentiation by the PHD-finger protein Phf5a

Alexandros Strikoudis 1,2,3, Charalampos Lazaris 1,2,3,4, Thomas Trimarchi 1,2,3, Antonio L. Galvao Neto 1,5, Yan Yang 1,2, Panagiotis Ntziachristos 1,2,3,9, Scott Rothbart 1,6, Shannon Buckley 1,2,3,9, Igor Dolgalev 1,2,3,4,7, Matthias Stadtfeld 1,8, Brian D. Strahl 6, Brian D. Dynlacht 1,2, Aristotelis Tsirigos 1,2,4,10 and Iannis Aifantis 1,2,3,10

Pluripotent embryonic stem cells (ESCs) self-renew or differentiate into all tissues of the developing embryo and cell specification factors are necessary to balance gene expression. Here we delineate the function of the PHD-finger protein 5a (Phf5a) in ESC self-renewal and ascribe its role in regulating pluripotency, cellular reprogramming and myoblast specification. We demonstrate that Phf5a is essential for maintaining pluripotency, since depleted ESCs exhibit hallmarks of differentiation. Mechanistically, we attribute Phf5a function to the stabilization of the Paf1 transcriptional complex and control of RNA polymerase II elongation on pluripotency loci. Apart from an ESC-specific factor, we demonstrate that Phf5a controls differentiation of adult myoblasts. Our findings suggest a potent mode of regulation by Phf5a in stem cells, which directs their transcriptional programme, ultimately regulating maintenance of pluripotency and cellular reprogramming.

Modulation of gene expression is crucial for stem cell self-renewal or cell specification. The Paf1 transcriptional complex (Paf1C) is central to these processes dictating RNA-PolII function and deposition of histone modifications7,8. Paf1C plays important roles in development and is necessary for differentiation9-11; however, details regarding its function in stem cells remain unknown. Here we demonstrate that Phf5a is a potent regulator of Paf1C stability and chromatin binding. Furthermore, we show that it is essential for ESC self-renewal, and cellular reprogramming and found that RNA-PolII elongation of pluripotency genes is defective after Phf5a depletion. Beyond its role in ESCs, we found Phf5a to regulate muscle specification, suggesting additional functions in adult stem cells. We conclude that Phf5a is crucial for regulating RNA elongation of genes controlling pluripotency and cell differentiation.

RESULTS

Phf5a depletion leads to loss of ESC pluripotency and inhibits reprogramming

To delineate the role of Phf5a in pluripotency, we investigated its expression levels during mouse ESC differentiation. We utilized the

---

1Department of Pathology, NYU School of Medicine, New York, New York 10016, USA. 2Laura & Isaac Perlmutter Cancer Center, NYU School of Medicine, New York, New York 10016, USA. 3Helen L. & Martin S. Kimmel Center for Stem Cell Biology, NYU School of Medicine, New York, New York 10016, USA. 4Center for Health Informatics and Bioinformatics, NYU School of Medicine, New York, New York 10016, USA. 5Ronald O. Perelman Department of Dermatology, NYU School of Medicine, New York, New York 10016, USA. 6Department of Biochemistry and Biophysics and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, North Carolina 27599, USA. 7Genome Technology Center, Office of Collaborative Science, NYU School of Medicine, New York, New York 10016, USA. 8Department of Cell Biology, NYU School of Medicine, New York, New York 10016, USA. 9Present addresses: Department of Biochemistry and Molecular Genetics, Northwestern School of Medicine, Chicago, Illinois 60611, USA (P.N.); Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA (S.B.). 10Correspondence should be addressed to A.T. or I.A. (e-mail: aristotelis.tsirigos@nyumc.org or iannis.aifantis@nyumc.org)

Received 18 November 2015; accepted 15 September 2016; published online 17 October 2016; DOI: 10.1038/ncb3424

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Phf5a is required for maintenance of ESC self-renewal. (a) Western blot analysis of Phf5a, Nanog and Oct4 proteins during ESC differentiation (unprocessed original scans of blots are shown in Supplementary Fig. 7). D0, day 0. (b) Histogram fluorescence-activated cell sorting (FACS) plots representing loss of GFP fluorescence in Nanog-GFP transcriptional reporter ESCs following knockdown with control shRNA (shControl) or Phf5a shRNA (shPhf5a). (c) Alkaline phosphatase staining of ESCs following knockdown with shControl or shPhf5a (two different hairpins). Scale bars, 100 μm. (d) Heatmap of Affymetrix microarrays for differentially expressed genes of Nanog-GFP ESCs following knockdown with shControl or shPhf5a, respectively. Red, upregulated genes; blue, downregulated genes. Q value < 0.05. Fold change (log2) >1.5. (e,f) Bar graphs showing expression levels by qRT-PCR of Phf5a, pluripotency markers (e) and differentiation markers (f), respectively, following shPhf5a knockdown in ESCs. n = 6 biologically independent replicates (see Supplementary Table 5). Phf5a, Nanog, Pou5f1, Sox2, Zfp42 and Nr0b1: *P < 0.0001. Gata6, Gata4, Nkx2-5, Sox1 and Meox1, **P = 0.0001; Brachyury: NS, not significant; P = 0.5632, two-sided Student’s t-test, values represent the mean ± s.d. (g) GO circle plot displaying gene-annotation enrichment analysis. Blue and red indicate downregulated or upregulated gene-associated GO terms, respectively, relative to the z score of the analysis. (h) GO chord plot displaying relationships between several representative downregulated and upregulated GO terms and associated genes. Distinct categories linked to pluripotent or differentiated cells cluster separately. FC, fold change. (i) Western blot analysis of pluripotency factors following CRISPR–Cas9-mediated Phf5a depletion in ESCs (unprocessed original scans of blots are shown in Supplementary Fig. 7).

Nanog-GFP (NG) reporter ESC line, a faithful indicator of self-renewal. Phf5a expression, both at mRNA and protein levels, is high in pluripotent ESCs, but becomes rapidly downregulated following differentiation (Fig. 1a and Supplementary Fig. 1b). Consistent with a possible role in preserving self-renewal, knockdown of Phf5a led to a significant loss of Nanog-GFP fluorescence (Fig. 1b). This was accompanied by morphological changes (Supplementary Fig. 1c), and cells exhibited considerably reduced alkaline phosphatase (AP) staining, an additional marker of pluripotency (Fig. 1c). Importantly, we did not observe increased levels of cell death or apoptosis following Phf5a knockdown (Supplementary Fig. 1d,e), suggesting that its loss leads to differentiation without effects on viability.

We next determined the transcriptional response of Phf5a silencing by gene expression array analysis revealing two sets of genes significantly up- or downregulated (Fig. 1d and Supplementary Table 1). Changes in expression of selected genes were validated with qRT-PCR. Consistent with ESC differentiation, pluripotency markers were significantly downregulated, whereas lineage markers...
were upregulated with the exception of several mesoderm markers (Fig. 1e,f). To further characterize gene expression profiles of Phf5a-depleted ESCs, we performed RNA-seq followed by Gene Ontology (GO) and gene-annotation enrichment analysis. Downregulated genes associated with GO terms related to stem cell maintenance, chromatin organization and cell division, whereas upregulated genes associated with GO terms related to early morphogenesis. Transcriptional profiling therefore strongly supports the notion that Phf5a loss triggers ESC differentiation.

Figure 2 Phf5a regulates ESC pluripotency and cellular reprogramming. (a,b) Comparison of mass (a) and size (b) of teratomas generated in Severe combined immunodeficiency (SCID) mice following injection of doxycycline-induced ESCs engineered to express shControl or shPhf5a cassettes from the Col1a1 locus. n=4 biologically independent replicates (see Supplementary Table 5). \( P=0.001 \) two-sided Student’s t-test, values represent the mean ± s.d. (c) Comparison of Phf5a transcript levels between differentiated fibroblasts and pluripotent stem cells by qRT–PCR. n=4 biologically independent replicates (see Supplementary Table 5). ESCs: \( P=0.0035 \), iPSCs: \( P=0.0013 \), two-sided Student’s t-test, values represent the mean ± s.d. (d) Western blot analysis of Phf5a protein in differentiated fibroblasts and pluripotent stem cells (unprocessed original scans of blots are shown in Supplementary Fig. 7). (e,f) Alkaline phosphatase (AP) staining and comparison of AP-positive ESC-like colony number (f), respectively, of reprogrammable OKSM MEFs on day 14 post-initial doxycycline induction following shPhf5a knockdown. n=4 biologically independent replicates (see Supplementary Table 5). \( P=0.001 \), two-sided Student’s t-test, values represent the mean ± s.d.

To validate our findings, we first ruled out potential off-target effects. Phf5a silencing using individual shRNAs, siRNAs and CRISPR–Cas9-mediated gRNAs, targeting distinct regions of its transcript, resulted in identical downregulation of pluripotency markers Oct4 and Nanog, as well as decreased AP staining (Fig. 1i and Supplementary Fig. 1g–j). Examination of Phf5a levels in ESCs from different backgrounds revealed no significant differences during differentiation (Supplementary Fig. 1k,l). Finally, we engineered inducible knockdown ESC lines by introducing miR–30–shPhf5a cassettes in the Col1a1 locus of cells constitutively expressing the M2rtTA transactivator, and were able to faithfully reproduce differentiation phenotypes following addition of doxycycline (Supplementary Fig. 2a–d). Using these lines, we also investigated the role of Phf5a in pluripotency in vivo. Injection of inducible ESCs in immunodeficient mice post-induction with doxycycline significantly inhibited the ability to form teratomas (Fig. 2a,b). Examination by haematoxylin and eosin stain as well as immunohistochemistry showed reduction of mesoderm markers compared with ectoderm or endoderm (Supplementary Fig. 2e–g). Specifically, we noticed apparent histological differences, including depletion of skeletal muscle formation, which we validated using desmin immunohistochemistry (Supplementary Fig. 2g). To further document defects in mesoderm differentiation, we examined in vitro differentiation towards the mesoderm lineage using the reporter ESC line Dppa4–RFP/Brachyury–GFP. We found that shPhf5a depletion results in loss of self-renewal (Dppa4–RFP reduction), however, without an increase of mesoderm differentiation (Brachyury–GFP gain) (Supplementary Fig. 2h). Last, we were unable to detect upregulated Brachyury protein levels or additional mesoderm markers (Supplementary Fig. 2i,j). These results suggest that Phf5a loss results in failure of self-renewal maintenance while affecting lineage skewing, demonstrating its importance in stem cells.

These phenotypes prompted us to further investigate its role in pluripotency and cellular reprogramming. We compared Phf5a expression in pluripotent or differentiated cells and found that it correlated with the pluripotent state (Fig. 2c,d). Consistent with that, analysis of proliferation profiles in ESCs or primary mouse embryo fibroblasts (MEFs) following shPhf5a silencing resulted only...
in ESC self-renewal defects (Supplementary Fig. 2k,l). Finally, Phf5a overexpression during ESC differentiation maintained pluripotent marker expression (Supplementary Fig. 2m,n).

We reasoned that Phf5a expression might also regulate induced pluripotent stem cell (iPSC) formation. We silenced Phf5a in reprogrammable primary MEFs, engineered to express the Oct4, Klf4, Sox2 and c-Myc (OKSM) reprogramming cassette\(^{17,18}\) and interrogated its effects 14 days post-induction of reprogramming factors. We observed that Phf5a loss resulted in a dramatic reduction of reprogrammed fibroblasts (Fig. 2e and Supplementary Fig. 2o). We noticed a significant decrease in the absolute number of AP-positive ESC-like colonies (Fig. 2f), suggesting that Phf5a silencing leads to decreased efficiency of reprogramming. Last, we examined transgene-independent expression of endogenous markers of pluripotency\(^{19}\) and found that Phf5a deficiency blocked their upregulation (Supplementary Fig. 2p). Collectively, we conclude that Phf5a is necessary for ESC self-renewal and efficient iPSC generation as its silencing results in aberrant initiation of differentiation and a block to reprogramming.

Phf5a interacts with the Paf1 complex

The impact of Phf5a knockdown on ESCs and iPSCs prompted us to investigate its functional role propagating pluripotency. We first explored its intracellular localization. We engineered inducible Phf5a-expressing or control ESC lines, fractionated cytoplasmic and nuclear extracts following doxycycline induction and confirmed a primarily nuclear localization in ESCs (Supplementary Fig. 3a). Immunofluorescence confirmed nuclear accumulation of Phf5a (Supplementary Fig. 3b). Given its internal PHD-finger motif, we hypothesized a possible role in transcription and chromatin regulation.

To identify its molecular function we undertook an unbiased approach by purifying Phf5a from ESCs and analysing its interacting partners by mass spectrometry (Fig. 3a and Supplementary Table 2). Among the top interacting proteins we found 3 out of 6 subunits of the Paf1 transcriptional complex (Paf1C). The mammalian Paf1C, which consists of the subunits Ctr9, Rtf1, Leo1, Paf1, Cdc73 and Wdr61, has been implicated in transcriptional regulation and deposition of histone modifications\(^2\). Since mass spectrometry suggested close association of Paf1C with Phf5a and since Paf1C deletion also results in ESC differentiation\(^{21,22}\), we decided to further investigate this interaction.

Although we initially identified specific Paf1C subunits as Phf5a binding partners, we validated interactions with the entire complex (Fig. 3b). Similar to Phf5a, Paf1C subunits are downregulated during ESC differentiation (Supplementary Fig. 3c). We engineered inducible knock-in ESC lines expressing the subunits Cdc73 and Wdr61 and confirmed binding with Phf5a (Supplementary Fig. 3d). Furthermore, we validated Phf5a-Paf1C interactions using tandem-affinity purification (Supplementary Fig. 3e), as well as endogenous protein immunoprecipitations in ESCs (Fig. 3c). Importantly, we found that DNA or RNA does not mediate this interaction since it persists after extensive nuclease treatment (Supplementary Fig. 3f). These results suggest a robust interaction between Phf5a and Paf1C.

We next attempted to specify whether distinct Paf1C subunits mediate this interaction. We in vitro translated Paf1, Cdc73 and Wdr61 subunits and interrogated binding to purified recombinant Phf5a. We observed interactions with Cdc73 and Wdr61 subunits but not Paf1 (Fig. 3d), suggesting that Phf5a binds to a subset of Paf1C subunits directly. Next, we performed glycerol gradient density sedimentation analysis from ESCs and identified that Phf5a forms high-molecular-weight complexes and co-fractionates with Paf1C under native conditions (Fig. 3e). Furthermore, since Paf1C is implicated in transcriptional elongation, we confirmed Phf5a interaction with RNA-PolII (Supplementary Fig. 3g). Moreover, in density sedimentation analysis the elongating form of RNA-PolII co-sedimented with Phf5a and Paf1C (Fig. 3e), indicating active engagement during transcription elongation. In contrast, we were unable to detect interaction with the initiation factor TFIIID (Supplementary Fig. 3e), absent in elongating complexes, suggesting Phf5a association with PolII at specific transcription stages. These findings establish interaction of Phf5a with Paf1C and suggest its possible role in transcriptional regulation in ESCs.

Phf5a depletion leads to Paf1C destabilization and loss of binding to its target genes

The direct interaction between Phf5a and Paf1C together with differentiation phenotypes after knockdown suggested an intimate connection to Paf1C function. To explore this functional association, we first compared gene expression profiles of Phf5a- and Paf1-depleted ESCs and found similar gene expression patterns (Supplementary Fig. 4a and Supplementary Table 3). Since Phf5a was previously implicated in alternative exon recognition in malignant cells\(^23\), we also investigated whether shPhf5a or shPaf1 knockdown results in aberrant alternative splicing in ESCs. We used multivariate analysis of transcript splicing\(^24\), and found a small number of splicing differences, concluding that their loss does not result in overt changes in splicing patterns (Supplementary Table 4).

We next studied Paf1C stability following loss of Phf5a. We re-examined Paf1C composition using density sedimentation analysis in the presence or absence of Phf5a. We found that Phf5a knockdown leads to Paf1C distribution towards lower molecular-weight fractions, indicating destabilization (Supplementary Fig. 4b–e). In contrast, other protein complexes, such as Swi/Snf and NELF, remain unaltered (Supplementary Fig. 4f). This suggested loss of interaction among Paf1C subunits and we confirmed significant decrease between subunit associations following Phf5a silencing (Fig. 4a).

These findings suggested decreased Paf1C binding to target genes after Phf5a loss. We directly interrogated Paf1C occupancy in ESCs using Leo1, Cdc73 and Paf1 chromatin precipitation (ChIP)-sequencing in the presence or absence of Phf5a. We identified \(\sim4,200\) high-stringency targets of Paf1C in ESCs (Fig. 4b) with Cdc73- and Paf1-bound genes constituting subsets of Leo1-bound genes. We observed that Paf1C peaks virtually disappear following Phf5a silencing, supporting its critical role in complex function (Fig. 4c,d). We dissected peak localization and determined occupancy among promoters, UTRs, coding and intergenic regions. Consistent with regulation of active gene expression\(^25\), we noticed that Leo1 peaks fall mostly within gene bodies and promoters (Fig. 4e). Further dissection revealed that downregulated genes engage Leo1 mostly within gene bodies (42%), whereas upregulated genes utilize Leo1 mostly on promoters (47%) (Supplementary Fig. 4g). This suggests that Paf1C sub-complexes may differentially localize on their targets.
to control gene expression. However, using gene-set enrichment analysis we found that Paf1 targets are enriched in pluripotency genes and correlate with ESC signatures (Fig. 4f). These Paf1C targets include Pou5f1, Esrrb, Sall4, Prdm14 and other well-described pluripotency genes (Fig. 4g). Finally, to test direct Phf5a binding to the self-renewal network we engineered a Tet-inducible haemagglutinin (HA)–tagged Phf5a line and performed ChIP-seq using HA-epitope purification. Tagged Phf5a, Wdr61 (positive control) and GFP (negative control) were transiently expressed in engineered Tet-inducible ESC lines following addition of doxycycline. Bait proteins are tagged (marked with an asterisk) and migrate slower than endogenous proteins (unprocessed original scans of blots are shown in Supplementary Fig. 7). (. Endogenous protein immunoprecipitations for Phf5a and Paf1C subunits in ESCs. Solid or dashed lines illustrate systems-generated pathway of Phf5a interacting proteins following pulldown assay using HA-immunoprecipitation and western blot analysis (unprocessed original scans of blots are shown in Supplementary Fig. 7). (e) Endogenous protein immunoprecipitations for Phf5a and Paf1C subunits in ESCs (unprocessed original scans of blots are shown in Supplementary Fig. 7).

**Figure 3** Phf5a physically associates with the Paf1 complex. (a) Ingenuity systems-generated pathway of Phf5a interacting proteins following purification and mass spectrometry in ESCs. Solid or dashed lines illustrate established direct or indirect interactions, respectively. (b) Validation of Phf5a interactions with the Paf1 complex in ESCs using Flag-Phf5a purification. Tagged Phf5a, Wdr61 (positive control) and GFP (negative control) were transiently expressed in engineered Tet-inducible ESC lines following addition of doxycycline. Bait proteins are tagged (marked with an asterisk) and migrate slower than endogenous proteins (unprocessed original scans of blots are shown in Supplementary Fig. 7). (c) Endogenous protein immunoprecipitations for Phf5a and Paf1C subunits in ESCs (unprocessed original scans of blots are shown in Supplementary Fig. 7). (d) Paf1-complex subunits Paf1, Cdc73 and Wdr61 were cloned in HA-tag-expressing vectors and subjected to *in vitro* transcription and translation. Phf5a protein was expressed and purified from bacteria. *In vitro* binding of HA-tagged subunits and Phf5a was interrogated by a pulldown assay using HA-immunoprecipitation and western blot analysis (unprocessed original scans of blots are shown in Supplementary Fig. 7). (e) Phf5a interacting proteins from ESCs were subjected to glycerol gradient sedimentation followed by fractionation and western blot analysis resulting in overlapping distributions of Phf5a and Paf1C subunits. A control analysis for GFP is shown in the lower panel (unprocessed original scans of blots are shown in Supplementary Fig. 7).

**Loss of Phf5a leads to RNA-PolII promoter-proximal pausing**

The Paf1C facilitates transcriptional elongation and is required for maximal levels of Ser2-P-PolII (refs 27,28). Since our studies suggest that Phf5a controls Paf1C binding on self-renewal genes, we hypothesized that its loss might lead to their aberrant elongation. We first investigated Ser2-P-PolII levels following Phf5a loss and found them significantly lower (Fig. 5a and Supplementary Fig. 5a). We next interrogated effects on RNA-PolII stalling and regulation of elongation. Posing of activated RNA-PolII constitutes a rate-limiting step in gene expression and pause release is important for embryonic development and somatic cell reprogramming. We profiled nascent RNAs using global-run-on sequencing (GRO-seq) and calculated the elongation pausing index in ESCs following shPhf5a knockdown. We initially observed that downregulated genes were paused (Fig. 5b,c) in contrast to upregulated genes (Fig. 5d) signifying a difference between the two sets. We further dissected pausing ratios for multiple GO subsets. Downregulated categories exhibited significantly elevated promoter-proximal pausing compared with upregulated ones (Fig. 5e). In addition, downregulated gene read density shows significant decrease within gene bodies (Fig. 5f). Similar to shPhf5a, shPaf1 knockdown leads to profound RNA-PolII pausing on downregulated genes (Fig. 5g) analogous to flavopiridol treatment. In contrast, ESCs differentiated in the absence of LIF did not exhibit increased RNA-PolII stalling (Supplementary Fig. 5b–d), suggesting specific roles for Paf1C/Phf5a in maintenance of pluripotency. Indeed, we found that...
Figure 4 Phf5a controls interactions among Paf1C subunits and its silencing abrogates Paf1C recruitment on pluripotency genes in ESCs. (a) Western blot analysis of Paf1C subunit immunoprecipitations following knockdown with shControl or shPhf5a, showing loss of interactions between different Paf1C members following Phf5a depletion. Two different shRNA hairpins are shown. Left panel, blot for Leo1; right panel, blot for Cdc73; IP, immunoprecipitation. IB, immunoblot (unprocessed original scans of blots are shown in Supplementary Fig. 7). (b) Venn diagram showing the number of genes bound by individual Paf1C subunits in ESCs using ChIP-seq with antibodies against endogenous Leo1, Cdc73 and Paf1 proteins. (c) Heatmap representations of normalized read density for Leo1 binding in ESCs following shControl or shPhf5a silencing. (d) Venn diagrams showing the numbers of genes bound by Leo1, Cdc73 and Paf1 in ESCs in the presence or absence of Phf5a. (e) Binding profiles for genomic distribution of Leo1 peaks (upstream, promoter, coding region, 5'UTR, 3'UTR, downstream and intergenic) in ESCs, showing preferential (32%) binding within gene bodies. (f) Gene-set enrichment analysis plots showing significant enrichment of the top Paf1 targets for genes linked to embryonic stem cell signatures. FDR, false discovery rate; NES, normalized enrichment score. (g) Snapshots of Leo1 and Cdc73 binding on representative pluripotency gene targets (Sall4, Klf4, Zfp42, Esrrb, Sox2, Pou5f1 and Nanog) in the presence (blue) or absence (red) of Phf5a, respectively.

~50% of downregulated genes are direct ChIP-seq Paf1C targets and show promoter-proximal pausing after shPhf5a knockdown, compared with upregulated targets (Supplementary Fig. 5e).

To further study elongation dynamics we performed ChIP-seq for initiating (Ser5-phosphorylated) and elongating (Ser2-phosphorylated) RNA-PolIII in the presence or absence of Phf5a. We calculated the RNA-PolIII pausing index and found that Paf1C targets and self-renewal genes, such as Nanog, Pou5f1, Sox2, Klf4, Fbxo15, Myc, Esrrb and others, exhibit significant stalling (Fig. 5h). We also found elevated levels of Ser5-RNA-PolIII near gene promoters for related
Figure 5 Phf5a controls transcriptional elongation and RNA-PolII pause release of pluripotency genes in ESCs. (a) Western blot analysis of total PolII and Ser2-phosphorylated RNA-PolII in ESCs following shControl or shPhf5a knockdown or ESCs differentiated in the absence of leukaemia inhibitory factor (LIF; unprocessed original scans of blots are shown in Supplementary Fig. 7). (b,c) Scatter plot representing pausing indices of downregulated (b) or upregulated genes (c), respectively, 72 h following shControl or shPhf5a knockdown using GRO-seq analysis. Read density of 500 bp downstream of promoters (5’ density) was normalized to read density in the rest of the gene bodies (3’ density). Pausing index = 5’ density/3’ density. Grey, all genes; blue, downregulated genes; red, upregulated genes. (d) Box plot showing pausing index ratios after GRO-seq analysis for downregulated (blue) or upregulated (red) genes, respectively, following shPhf5a knockdown. Only downregulated genes exhibit significant promoter-proximal pausing after shPhf5a depletion. n = 3 biologically independent replicates, non-parametric Wilcoxon signed rank test. (e) Box plot showing pausing index ratios after GRO-seq analysis for specific GO terms. Blue, downregulated and red, upregulated GO term categories, respectively. Only downregulated GO terms exhibit significant promoter-proximal pausing after shPhf5a depletion. n = 3 biologically independent replicates, non-parametric Wilcoxon signed rank test (f) Comparison of GRO-seq read density profiles of genes 72 h following shControl or shPhf5a knockdown in ESCs. RPKM, reads per kilobase per million total reads; TSS, transcription start site. (g) Box plot representing comparison of log2 pausing index for downregulated genes 72 h following shControl, shPhf5a, shPaf1 knockdown, or flavopiridol-treated ESCs, respectively, using GRO-seq analysis. Flavopiridol treatment is used as a positive control of pause release block. n = 3 biologically independent replicates, non-parametric Wilcoxon signed rank test. (h) Scatter plot representing RNA-PolII pause index for Paf1C targets and pluripotency genes based on normalized Ser5 (on TSSs)/Ser2 (on gene bodies) read density ratio of RNA-PolII ChiP-Seq in ESCs following shControl or shPhf5a silencing, respectively. In box plots (d,e,g) the central mark is the median, and the edges of the box are the first and third quartiles. Whiskers extend to the most extreme non-outlier data points.
GO term genes, in contrast to decreased Ser2-RNA-PolII in their corresponding gene bodies (Supplementary Fig. 5f,g).

Finally, we interrogated whether Phf5a affects histone modifications associated with Paf1C function\(^7\). We performed ChIP-sequencing for H3K4me3, H3K79me2 and H3K36me3 in the presence or absence of Phf5a. We found its depletion negatively affected elongation-associated histone modifications H3K79me2 and H3K36me3 on Paf1C targets and self-renewal genes, but not the promoter-associated mark H3K4me3 (Fig. 6a). H3K79me2 and H3K36me3 profiles revealed a pronounced loss from gene bodies after shPhf5a knockdown (Fig. 6b). Dissection of H3K79me2 revealed that Paf1C targets have similar profiles to downregulated genes, compared with upregulated ones (Fig. 6c). Last, we examined occupancy on pluripotency genes and found diminished H3K79me2 and H3K36me3 levels in their gene bodies, but not H3K4me3 levels on their promoters, compared with housekeeping genes (Fig. 6d). These data suggest that Phf5a loss affects Paf1C functions including regulation of RNA-PolII elongation and histone mark occupancy.

**Phf5a regulates myogenic differentiation**

We next asked whether Phf5a functions specifically in ESCs or it controls differentiation in additional systems. Since Phf5a depletion results in aberrant mesoderm differentiation and muscle formation in teratomas, and Pafl is also implicated in cardiomyocyte specification in zebrafish\(^10\), we decided to study Phf5a function in muscle. Myoblasts self-renew; however, in differentiation conditions, they commence myogenic programmes and fuse, forming elongated, multinucleated myotubes (Supplementary Fig. 6a). We initially depleted Paf1C subunits in C2C12 myoblasts and verified its role in myotube differentiation (Supplementary Fig. 6b). We then depleted Phf5a and found that it also compromises their ability to differentiate. We observed that myoblasts fail to upregulate the differentiation marker myosin heavy chain (MHC) (Supplementary Fig. 6c,d). Additionally, Phf5a silencing results in maintenance of Pax7, a marker of myoblast self-renewal (Fig. 7a). To support our findings in C2C12 cells, we generated a Tet-inducible RNAi mouse model by knocking-in individual shPhf5a hairpins in the Col1a1 locus of ESCs (Fig. 7b). We generated Rosa26\(^{rtTA}\)Col1a1\(^{ TRE}\).shPhf5a mice and crossed them to EIIA-Cre mice to drive hairpin expression (Fig. 7b). We generated primary mouse myoblasts from these animals, and observed defects in myotube differentiation following Phf5a silencing (Fig. 7c). We observed upregulation of Phf5a and PaflC during primary myoblast differentiation (Supplementary Fig. 6e) and were also able to mimic effects of RNAi depletion on myotube differentiation using CRISPR-Cas9 strategies (Fig. 7d,e and Supplementary Fig. 6f,g). These results demonstrate that loss of Phf5a blocks myogenic differentiation.

To further investigate how Paf1C/Phf5a control mechanisms of muscle differentiation, we performed ChIP-sequencing for Leol in myoblasts and myotubes. We found an increased number of Leol-bound genes during muscle differentiation from 700 genes in myoblasts, to more than 2,700 in myotubes (Fig. 7f). GO analysis of Leol-bound genes identified multiple chromatin- and transcription-associated GO terms in myoblasts, such as chromatin assembly, nucleosome organization and others, compared with muscle-specific GO terms in differentiated myotubes, such as actin organization, muscle development, muscle organization and myofibril assembly (Supplementary Fig. 6h,i). Examples for Leol binding include Hist1 cluster genes in myoblasts, Myog, Myo1c and Myom3 in myotubes and many others (Fig. 7g).

Finally, we performed ChIP-sequencing in myotube differentiation after shPhf5a knockdown. We found that Leol binding was abolished from its myotube targets (Fig. 7h and Supplementary Fig. 6j). Interestingly, we identified ~1,000 Leol targets after shPhf5a knockdown; however, almost none of them is associated with myogenic processes (Fig. 7h). In addition, we performed GO analysis of Leol targets after shPhf5a knockdown and found genes associated with neurogenesis, instead of myogenesis (Supplementary Fig. 6j). Specifically, we identified neurological system process, cognition, sensory perception of smell, and sensory perception of chemical stimuli GO terms after shPhf5a knockdown (Supplementary Fig. 6j). We observed loss of Leol binding in specific myogenic genes, but gain in neurogenic ones such as olfactory, taste and neurotransmitter receptors, G-protein-coupled receptors, ion channels and many others (Fig. 7i). Finally, to investigate whether Phf5a loss affects Paf1C stability we directly interrogated interactions among Paf1C subunits after its knockdown and found dissociation of the core subunits Leol and Cdc73, resulting in complex disruption (Fig. 7f). In conclusion, these studies confirm that Phf5a is an essential regulator of myoblast differentiation and suggest that it stabilizes Paf1C in chromatin promoting myogenic programmes.

**DISCUSSION**

Since their early characterization, PHD-finger proteins were recognized as tethering molecules recruiting or stabilizing protein complexes\(^34^-^36\) leading to tight regulation of gene expression\(^27\). Here we characterize in detail the functions of Phf5a in ESCs, iPSCs and myoblasts and attribute its requirement for self-renewal to the binding and stabilization of the Paf1C. We found that Phf5a depletion resulted in ESC differentiation and inhibition of cellular reprogramming. Phf5a and Paf1C associate strongly, are recruited on actively transcribed pluripotency genes and positively regulate RNA-PolII elongation.

Previous efforts to study Paf1C functions focused entirely on promoter occupancy using ChIP-on-chip assays\(^21\), or in relation to DNA methylation\(^28\). Our studies characterize how Phf5a affects elongation dynamics in self-renewing ESCs and we integrated a combination of approaches to elucidate its functions. First, we characterized promoter-proximal pausing ratios using nascent RNA profiling following Phf5a depletion. Second, utilizing RNA-PolIII ChIP-sequencing we calculated travelling ratios on pluripotency genes and Paf1C targets. Last, we determined how elongation-specific histone marks change in response to Paf1C alterations. Our results suggest that Phf5a directly regulates Paf1C stability, facilitating pause release and productive elongation of the self-renewal network. Elongation of upregulated genes in the absence of Paf1C is due to indirect effects. Direct Paf1C/Phf5a targets include master regulators of pluripotency, and it would be intriguing to further study mechanisms of Paf1C/Phf5a recruitment and possible cooperation with transcription factors\(^22\).

Furthermore, we found that Phf5a functions are not limited to ESCs. Using in vivo and in vitro models we demonstrated that Phf5a is essential for differentiation of myoblasts to myotubes. These functions are Paf1C-dependent as loss of Phf5a expression leads to significant decrease of Paf1C occupancy at myogenic genes.
Figure 6 Phf5a regulates the deposition of histone marks characteristic of transcriptional elongation in pluripotency gene loci. (a) Box plots representing log2 fold change of normalized read density for H3K4me3, H3K79me2 and H3K36me3 ChIP-seq in ESCs following shControl or shPhf5a silencing. Plots represent comparisons of all expressed transcripts in ESCs with direct Paf1 targets around transcription start sites (H3K4me3) or gene bodies (H3K79me2 and H3K36me3). n = 3 biologically independent replicates, non-parametric Wilcoxon signed rank test. (b) Normalized read density profiles around TSSs (H3K4me3) or gene bodies (H3K79me2 and H3K36me3) on Paf1C targets and pluripotency genes in ESCs in the presence (blue) or absence (red) of Phf5a. (c) Box plots showing log2 fold change in H3K79me2 occupancy on gene bodies of target genes. H3K79me2 occupancy is increased in upregulated genes; however, H3K79me2 occupancy is decreased in downregulated genes compared with all expressed genes. n = 3 biologically independent replicates, non-parametric Wilcoxon signed rank test. (d) Snapshots of representative H3K4me3, H3K79me2 and H3K36me3 density tracks on pluripotency genes or control loci under conditions of shControl (blue) or shPhf5a silencing (red). In box plots (a,c) the central mark is the median, and the edges of the box are the first and third quartiles. Whiskers extend to the most extreme non-outlier data points.

and de novo targeting at neurogenic ones. These findings are consistent with Paf1C regulating cardiac specification and heart morphogenesis in zebrafish\textsuperscript{10} and deregulation on muscle organ formation in Phf5a-depleted C. elegans\textsuperscript{6}. Collectively, we conclude that Phf5a mediates Paf1C functions to orchestrate myogenic differentiation.
Processes that balance self-renewal or cell specification can be deregulated in cancer. Paf1C can also act as an oncogene and its amplification or overexpression is implicated in tumour formation. 8,39 Although Paf1C stimulates transcriptional elongation in vitro and in vivo 7,26,40, it was recently suggested to also suppress RNA-PolII transcription in cancer. 41 Our data suggest that Paf1C is active in ESCs,
directly promoting elongation of pluripotency networks. However, Paf1C may function in a cell-type-specific manner influenced by chromatin accessibility.

Recent studies in human malignancies implicated Phf5a in endometrial cancer and glioblastoma\(^2,3\). Although Phf5a is suggested to interact with ATP-dependent helicases and the U2 snRNP spliceosome\(^2,3\), its silencing surprisingly affects exon recognition and splicing only in glioblastoma stem cells but not their normal counterpart neural stem cells\(^3\). Despite extensive sequencing studies, we did not find significant splicing defects in Phf5a-depleted ESCs, enforcing the notion that these are cell-type specific. It would be intriguing to speculate that splicing phenotypes are absent from non-malignant cells (ESCs and neural stem cells) but present in glioblastoma or other malignancies, opening the way to investigate this distinction as a potential vulnerability in cancer. It is interesting to interrogate whether aberrant Phf5a expression correlates with deregulated Paf1C functions in human disease, and since Phf5a loss inhibits proliferation of cancer cells\(^4\), its targeting might be an alternative therapeutic option.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

**ACKNOWLEDGEMENTS**

We thank all members of the Aifantis laboratory for useful comments and discussions throughout the duration of this project; Z. Gao and P. Voigt for experimental help with glycerol gradients and helpful discussions; K. J. Armache and P. De Ioannes Fernandez for help with protein purification, helpful discussions and manuscript preparation; H.-J. Fehling for the Dpp4r-RFP/Brachyury-GFP reporter ESC line; A. Nagy and the NYU Genome Technology Center (supported in part by National Institutes of Health (NIH)/National Cancer Institute (NCI) grant P50CA16087-30) for expertise with sequencing experiments; the NYU Histology Core (SP50CA16087-31) for assistance; C. Loomis and L. Chariboga for immunohistochemistry experiments; and H. Li and T. Liu at the Center for Advanced Proteomics Research, New Jersey School of Medicine for mass spectrometry. This work has used computing resources at the High Performance Computing Facility of the Center of Health Informatics and Bioinformatics at the NYU Medical Center. A.S. is supported by the NYSTEM institutional NYU Stem Cell Training Grant (C026880). I.A. is supported by the NIH (5R01CA133759, RO1CA190379, ROICAI96455, SROCA173636, ROI1CA194923) and the NYSTEM programme of the New York State Health Department (NYSTEM-N11G-235).

**AUTHOR CONTRIBUTIONS**

I.A. and A.S. designed the experiments and wrote the manuscript. A.S. performed the experiments. A.T., C.L., T.T. and I.D. designed and performed the analysis of genome-wide data. P.N. provided expertize in sequencing experiments. A.T., C.L., T.T. and I.D. designed and performed the analysis of genome-wide data. P.N. provided expertize in sequencing experiments. A.S. is supported by the NYSTEM-N11G-235.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

**REFERENCES**

1. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
2. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
3. Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835–850 (2011).
4. Buckley, S. M. et al. Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* **11**, 783–798 (2012).
5. Trappe, R. et al. Identification and characterization of a novel murine multigene family containing a PHD-finger-like motif. *Biochem. Biophys. Res. Commun.* **293**, 816–826 (2002).
6. Trappe, R., Schulze, E., Rzymski, T., Frode, S. & Engel, W. The Caenorhabditis elegans ortholog of human Phf5a shows a muscle-specific expression domain and is essential for C. elegans morphogenetic development. *Biochem. Biophys. Res. Commun.* **297**, 1049–1057 (2002).
7. Jaehning, J. A. The Paf1 complex platform or player in RNA polymerase II transcription? *Biochim. Biophys. Acta* **1799**, 379–388 (2010).
8. Tomson, B. N. & Arndt, K. M. The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta* **1829**, 116–125 (2013).
9. Akanuma, T., Koshida, S., Kawamura, A., Kishimoto, Y. & Takada, S. Paf1 complex homologues are required for Notch-regulated transcription during somite segmentation. *EMBO Rep.* **8**, 858–863 (2007).
10. Langenbacher, A. D. The PAF1 complex differentially regulates cardiomyocyte specification. *Dev. Biol.* **353**, 19–28 (2011).
11. Nguyen, C. T., Langenbacher, A., Hsieh, M. & Chen, J. N. The PAF1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish. *Dev. Biol.* **341**, 167–175 (2010).
12. Schnabel, C. et al. Smarccl1/Sfi155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells** **27**, 2979–2991 (2009).
13. Walter, W., Sanchez-Cabo, F. & Ricote, M. GDplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914 (2015).
14. Premirurit, P. K. et al. A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* **145**, 145–158 (2011).
15. Lian, X. et al. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nat. Protoc.* **8**, 162–175 (2013).
16. Shelton, M. et al. Derivation and expansion of PAX7-positive muscle progenitors from human and mouse embryonic stem cells. *Stem Cell Rep.* **3**, 516–529 (2014).
17. Madan, B. et al. The pluripotency-associated gene Dpp4p is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol. Cell. Biol.* **29**, 3186–3203 (2009).
18. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reproducible mouse strain from gene-targeted embryonic stem cells. *Nat. Methods* **7**, 53–55 (2010).
19. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28 (2006).
20. Sommer, C. A. et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells** **27**, 543–549 (2009).
21. Ding, L. et al. 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).
22. Ponnusamy, M. P. et al. RNA polymerase II associated factor 1/PD2 maintains self-renewal by its interaction with Oct3/4 in mouse embryonic stem cells. *Stem Cells** **27**, 3001–3011 (2009).
23. Hubert, C. G. et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
24. Sood, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5953–E5601 (2014).
25. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
26. Kim, J., Guernah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
27. Mueller, C. L., Porter, S. E., Hoffman, M. G. & Jaehning, J. A. The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol. Cell* **14**, 447–456 (2004).
28. Nordick, K., Hoffman, M. G., Betz, J. L. & Jaehning, J. A. Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by disruption of Paf1 from RNA polymerase II. *Eukaryot. Cell* **7**, 1158–1167 (2008).
29. Smith, E. & Shilatifard, A. Transcriptional elongation checkpoint control in development and disease. *Genes Dev.* **27**, 1079–1088 (2013).
30. Levine, M. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 503–511 (2011).
31. Liu, L. et al. Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15**, 574–588 (2014).
32. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
33. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
34. Breiz, M. The PHF finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31**, 35–40 (2006).
35. Li, H. et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
36. Pena, P. V. et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).
37. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
38. Rigbolt, K. T. et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* **4**, rs3 (2011).
39. Chaudhary, K., Deb, S., Moniaux, N., Ponnusamy, M. P. & Batra, S. K. Human RNA polymerase II-associated factor complex: dysregulation in cancer. *Oncogene* **26**, 7499–7507 (2007).
40. Yu, M. et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).
41. Chen, F. X. et al. PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015 (2015).
42. Falck, E. & Klinga-Levan, K. Expression patterns of Phf5a/PHF5A and Gja1/GJA1 in rat and human endometrial cancer. *Cancer Cell Int.* **13**, 43 (2013).
43. Rzymski, T., Grzmil, P., Meinhardt, A., Wolf, S. & Burfeind, P. PHF5A represents a bridge protein between splicing proteins and ATP-dependent helicases and is differentially expressed during mouse spermatogenesis. *Cytogenet. Genome Res.* **121**, 232–244 (2008).
44. Nijhawan, D. et al. Cancer vulnerabilities unveiled by genomic loss. *Cell* **150**, 842–854 (2012).
METHODS

Mouse ESC and iPSC culture and OKSM MEF reprogramming. Mouse ESCs, iPSCs and ‘reprogrammable’ OKSM MEFs were cultured under standard conditions as described previously [1] using recombinant LIF, on gelatin-coated plates or feeder-MEFs, respectively. For viral transductions, ESCs and OKSM MEFs were transduced with pLKO.1-puro backbone lentiviruses and selected with puromycin. For reprogramming experiments, ESC-like colonies were enumerated 14 days post-initiation with doxycycline.

Culture of C2C12 myoblasts, and myotube differentiation. C2C12 myoblast cells were cultured as described previously [2]. For differentiation, C2C12 cells were grown to confluence followed by culturing in differentiation media (DME/F12 supplemented with 2% horse serum) for 72 h or up to 120 h before analysis.

Immunofluorescence. ESCs expressing Flag-Phf5a or C2C12 cells were fixed with 4% formaldehyde in PBS for 20 min at room temperature, washed once with PBS and permeabilized for 20 min at room temperature using Block Solution (5% normal goat serum, 0.1% Triton X-100 in PBS). Cells were incubated overnight at 4°C with Flag antibody (1:500), Desmin (1:500) or MHC serum (1:500) in Block Solution, washed three times with PBS, incubated for 1 h at room temperature with 1:1,000 secondary rabbit anti-mouse Alexa594-conjugated antibody in Block Solution and DAPI.

Animal experiments. Female C57Bl/6 mice (6–8 weeks old) were obtained from the National Cancer Institute. For teratoma assays, non-obese diabetic/severe combined immunodeficient NOD/MtBomTac-Prkdc<sup>−/−</sup> (NOD-SCID) mice were obtained from Taconic. KH2 ESCs (10<sup>4</sup>) engineered to express inducible shControl or shPhf5a hairpins were injected subcutaneously into NOD-SCID mice, were under 0.1 mg doxycycline diet throughout the course of the assay and were euthanized 3 weeks after induction when tumours were collected. Tumours were fixed in formalin followed by embedding in paraffin, sectioned and stained for histological analysis with haematoxylin and eosin or immunohistochemistry stainings following standard procedures. For generation of Tet-inducible knockdown animals, Rosa26<sup>tm120</sup>Cola1<sup>tm120</sup>shPhf5a-targeted ESCs were purchased from Mirimex. Mice were generated by injection into tetraploid blastocysts at the NYU Medical Center Rodent Genetic Engineering Core. Engineered mice were crossed to EIIA-Cre recombinase mice to drive Rosa26<sup>tm120</sup> expression. For in vivo Col1a1<sup>tm120</sup>shPhf5a cassette expression, mice were placed on 0.1 mg doxycycline diet. Isolation and growth of primary myoblasts was performed as described previously [3]. Briefly, neonatal mice were euthanized and limb muscle was dissected from skin and minced in ice-cold dissociation buffer C, beads were washed with 200 column volumes of buffer C and HA-tagged proteins were eluted with 20 μl Flag magnetic beads, washed with 1× buffer E (IBA), and boiled in SDS-loading buffer. For cytoplasmic and nuclear fractionation, 1× KH2 ESCs were lysed as above and cytoplasmic and nuclear extracts were prepared as described previously using a glass homogenizer [4]. For mass spectrometry, peptides were analysed by LCMS/MS on an Orbitrap Q Exactive (Thermo Fisher) using the precursor ion for Phf5a (148,571±1264 Da). For tandem affinity purification, elutions were bound to 50 μl Flag magnetic beads, washed with 1× buffer E (IBA), and boiled in SDS-loading buffer. For cytoplasmic and nuclear fractionation, 1× KH2 ESCs were lysed as above and cytoplasmic and nuclear extracts were prepared as described previously using a glass homogenizer [5]. For mass spectrometry, peptides were analysed by LC-MS/MS on an Orbitrap Q Exactive (Thermo Fisher) using the precursor ion for Phf5a (148,571±1264 Da).

Glycerol gradient density sedimentation analysis. Targeted ESCs (5×10<sup>4</sup>) were induced, lysed, and tagged proteins were purified as above and concentrated to a final volume of 100 μl. Glycerol density (15–35%) were prepared by mixing with each 4 ml of each 4.5 ml top-open tubes (Beckman) and fractions were prepared as described before using TCA precipitation [6].

Phf5a purification from bacteria and in vitro pulldown. Phf5a ORF was cloned into the pgEX-6P-1 vector (GE Healthcare) (gift from K. J. Armache, NYU School of Medicine, USA) at 1:10,000. For western blots, antibodies were used as described before.

NATURE CELL BIOLOGY

DOI: 10.1038/ncb3424

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
total of 12 h and 100 μl of material was removed and snap-frozen every 3 h. Nucleic acids were phenol/chloroform extracted, ethanol precipitated and visualized in a 2% agarose gel.

Flow cytometry analysis. Apoptosis and cell death was detected using Annexin-V APC-conjugated detection kit (BD Biosciences) along with 7-AAD following the manufacturer's protocol on a BD LSRFortessa (BD Biosciences) flow cytometer.

Real-time quantitative PCR with reverse transcription (qRT–PCR) and microarray analysis. Total RNA was harvested from cells using the Qiagen RNeasy Kit (Qiagen) and 2 μg was used for cDNA synthesis using the High-Capacity RNA-to-cDNA Kit (Applied Biosystems), qRT-PCR was carried out using LightCycler SYBR green mix (Roche) using a LightCycler 480 II (Roche). For microarray analysis RNA samples were hybridized to Affymetrix MouseGeno2.0 Chips and scanned at the NYU Genome Technology Center. CEL files were loaded into GeneSpring (Agilent). Feature intensities for each probe set were condensed into a single intensity value.

GRO-seq and library preparation. Analysis of nascent RNAs using global run-on experiments was performed as described previously46. Briefly, nuclei were isolated in swelling buffer (10 mM Tris- HCl pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂), lysed twice in lysis buffer (10 mM Tris- HCl pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂, 10% glycerol, 0.5% NP-40) and snap-frozen in freezing buffer (50 mM Tris pH 8.0, 40% glycerol, 5 mM MgCl₂, 0.1 mM EDTA). For run-on reaction, an equal volume of reaction buffer was added to thawed nuclei (10 mM Tris pH 8.0, 5 mM MgCl₂, 300 mM KCl, 500 mM ATP, 500 mM GTP, 5 μM CTP, 500 μM BrUTP, 1 mM dithiothreitol, 100 μM S- nonat, 1% Sarkosyl), mixed and incubated at 30°C for 50 min. The reaction was stopped with Trizol reagent and RNA was phenol/chloroform extracted and ethanol precipitated. RNA was heated in fragmentation buffer (20 mM Tris pH 8.0, 100 mM KCl, 6.25 mM MgCl₂, 1 mM dithiothreitol), DNase treated and purified using Zymo RNA Clean & Concentrator (Zymo Research) using the >17nt protocol. Run-on RNA was immunoprecipitated using BSA-blocked BrDU beads (Santa Cruz) in binding buffer (SSPE 0.5X, 1 mM EDTA, 0.05% Tween-20) for 1 h at 4°C, washed and eluted in elution buffer (5 mM Tris pH 7.5, 300 mM NaCl, 20 mM dithiothreitol, 1 mM EDTA, 1% SDS) at 65°C for 20 min. Nascent RNA was further phenol/chloroform extracted and sequencing libraries were prepared.

Data sources and computational pipelines. The samples were run using Illumina HiSeq2000. Raw reads were aligned against the mouse genome assembly mm10/GRCm38. Alignments were performed using Bowtie v1.0.0 (ref. 53). MACS 1.4.2 (ref. 54) or MACS 2.0 was used for peak calling in the case of ChiP-Seq data, while the suite GenomicTools version 2.8.1 (ref. 55) was used for genome binning, genomic annotations and the construction of occupancy profiles, both in the case of polymerases and histone marks. For plotting, R version 3.2.0 was used (R Core Team (2016). R: A language and environment for statistical computing, R Foundation for Statistical Computing) (https://www.R-project.org), along with the Venn Diagram package for the generation of Venn diagrams and ggplot2 for the generation of boxplots, density profiles and scatter plots with ggrepel. For the generation of heatmaps depicting the binding profiles of Pac1C components, deepTools was used.

ChIP-seq and library preparation. ChIP experiments were performed as described previously. Antibodies coupled to magnetic beads were added to precleared chromatin and incubated for 12–16 h. Beads were washed and eluted according to protocol. ChIP-seq libraries for Pac1C components, PolII samples and histone marks were prepared using the Illumina TruSeq system, including end repair, A-tailing, adaptor ligation and PCR amplification. AMPure XP beads (Beckman Coulter, A63880) were used for DNA cleaning in each step of the process. Raw images generated by Illumina HiSeq2000 using default parameters were processed by CASAVA to remove the first and last bases and then they were used to generate sequence reads in fastq format. Reads were aligned to the mm10 mouse genome using Bowtie with the standard parameters (except for –m 1 to report only unique alignments). MACS version 1.4.2 (ref. 54) was used to perform peak calling using the parameter values listed below.

In the case of ChIP-seq experiments for Pac1C components and Phf5a, MACS v1.4.2 was used and the parameter values were: –nomodel, –shiftsize=120, –local 5000, –local 50000, –P 10−4.

In the case of polymerase and histone ChIP-seq experiments, MACS v2.0 was used with the following parameter values: (a) –nomodel, (b) –broad, (c) –shiftsize=120, (d) –local 5000, (e) –local 50000, (f) –local 500000.

All PolII and histone ChIP-seq experiments were performed in triplicates and peaks present at least in two out of the three triplicates were used for downstream analysis. The files with the aligned reads were converted to wig format using GenomicTools and then to bigwig format using the corresponding UCSC tool.

Peak characterization. Peaks were assigned to the following categories based on their genome-wide distributions: Upstream: this category includes all peaks that fall within 1–3 kb upstream of the transcription start site (TSS), Promoter: it includes all peaks that fall within 1 kb upstream of the TSS, Gene body: it includes all peaks that fall within the 5′ UTR, the coding region of genes and the 3′ UTR, Downstream: it includes all peaks that fall within 3 kb from the transcription end site (TES), Distal Intergenic: All the peaks that fall within the remaining genomic loci. The peak characterization was performed using ChiPSseeker and custom in-house scripts.

Calculation of PolII pausing index. The calculation of PolII pausing index was performed as described previously. Specifically, the PolII densities were initially calculated as reads per kilobase per million to normalize for region length and number of reads. The initiating region was defined as the area between 30 bp upstream of the TSS to 300 bp after the TSS, while the elongating region was from +300 bp to the end of the gene. We calculated the PolII pausing index by dividing the PolII densities for the initiating region versus the elongating region. We used Wilcoxon's non-parametric test to compare the distributions of the fold-changes (log, scale) in the pausing index in the control (shControl) and the samples, where either Phf5a (shPhf5a) or Paf1 (shPaf1) were downregulated.

Gene set enrichment analysis. For GSEA analysis, the online version of the GSEA tool was used26. The peaks that were identified as significant after peak calling were ranked according to peak score (from highest to lowest) and they were used as input for GSEA. MSigDB v5.0 (updated in April 2015) was used for GSEA.

Gene ontology. Gene ontology (GO) analysis was performed on the basis of ChIP-Seq data and gene expression data by using the DAVID tool v6.7 (https://david.ncifcrf.gov). Visualization of the GO results was performed using the R package GOpol26. The reported z-score was calculated as described in GOpol and shows the trend (increasing/decreasing) of the corresponding GO category in terms of gene expression.

Statistics and reproducibility. Sample sizes and reproducibility for each figure are denoted in the figure legends. Unless otherwise noted, data are representative of at least three biologically independent experiments. For mouse experiments, no statistical method was used to predetermine sample size. Furthermore, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. Statistical significance between conditions was assessed by two-tailed Student's t-tests. Error bars represent s.d., and significance between conditions is denoted. Raw data from independent replicate experiments can be found in the Statistics Source Data (Supplementary Table 5).

Cell lines. No cell lines used in this study were found in the databases of commonly missidentified cell lines that are maintained by ICLAC and NCBI Biosample. The following sources of cell lines were used: KH2 ESCs, OKSM MEFs and iPSCs: gift from K. Hochmedinger, Harvard Stem Cell Institute, USA; CCE and Nanog-GFP: gift from I. Lemiischka, Mount Sinai School of Medicine, USA; MK6: gift from S. Y. Kim, Rodent Genetic Engineering Core, NYU School of Medicine, USA; Dppa4-RFP/Brachyury-GFP ESCs: gift from H. J. Fehling, University Clinics Ulm, Germany; C2C12 gift from B. Dynlacht, NYU School of Medicine, USA. The cell lines were not authenticated. The cell lines were routinely tested for mycoplasma contamination.

Primary accessions. Gene Expression Omnibus (GEO) GSE63974.

Data availability. The next-generation sequencing data that support the findings of this study in Figs 1 and 4–7 have been deposited in the Gene Expression Omnibus (GEO) database under the accession code GSE63974. Statistics source data have been provided as Supplementary Table 5. All other data supporting the findings of this study are provided from the corresponding author on request.

45. Reavey, L. et al. Regulation of hematopoietic stem cell differentiation by a single ubiquitin ligase-substrate complex. Nat. Immunol. 11, 207–215 (2010).
46. Blais, A., van Oeveren, C. J., Maguero, R., Acosta-Avea, D. & Dynlacht, B. D. Retinoblastoma tumor suppressor protein-dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. J. Cell Biol. 179, 1399–1412 (2007).
47. Springer, M. L., Rando, T. A. & Blau, H. M. in Current Protocols in Human Genetics (eds Haines, J. L. et al.) Ch. 13, Unit 13–14 (John Wiley, 2002).
48. Gloeckner, C. J., Boldt, K. & Ueffing, M. in Current Protocols in Protein Science (eds Coligan, J. E. et al.) Ch. 19, Unit 19–20 (John Wiley, 2009).
49. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**, 465–477 (2005).

50. Zhu, B. et al. The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev.* **19**, 1668–1673 (2005).

51. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).

52. Gao, Z. et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344–356 (2012).

53. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

54. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

55. Ntziachristos, P. et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* **18**, 298–301 (2012).

56. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**, 35 (2011).

57. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).

58. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

59. Ntziachristos, P. et al. Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).

60. Yu, G., Wang, L. G. & He, Q. Y. ChiPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

61. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

62. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
**Supplementary Information**

DOI: 10.1038/nbib3424

**Supplementary Figure 1** Phf5a depletion results in aberrant ESC differentiation. (a) Alignment of Phf5a protein sequence from different organisms. Green: Conserved cysteine residues, which coordinate Zn²⁺. Orange: Nuclear localization signal (NLS). (b) Relative expression of Phf5a transcript by qRT-PCR in ESC differentiation following siRNA knockdown. (c) Western blot analysis of Oct4 expression in shControl, shPhf5a-1, and shPhf5a-2. (d) FACS plots showing levels of apoptosis and cell death following Phf5a depletion in ESCs. Proteasome inhibition by MG132 is used as positive control. (e) Western blot analysis of total and cleaved Caspase-3 following Phf5a knockdown in ESCs and differentiation. Proteasome inhibition by MG132 is used as positive control. (f) Gene Ontology (GO) analysis of significantly downregulated or upregulated genes, respectively, after RNA-seq analysis in ESCs following knockdown using siRNA transfections. (g) Schematic of Phf5a mRNA indicating distinct target regions for shRNAs, gRNAs and siRNA used in the study. (h) Western blot analysis of Oct4 levels following Phf5a knockdown in ESCs. (i) Alkaline phosphatase (AP) staining for differentiation of ESCs following knockdown using siRNA transfections. Scale bars: 100 μm. (j) Western blot analysis of pluripotency factors following Phf5a knockdown in ESCs. (k) Bar graphs showing expression levels of Phf5a in ESCs of different genetic backgrounds. n=6 biologically independent replicates (see Supplementary Table 5). MK6 and CCE: D2 and D4 **p=0.0001, respectively.

*© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.*
Silencing of Phf5a affects ESC pluripotency and IPS generation. (a) Schematic of Tet-inducible mir-30 shRNA expression cassette at the targeted Col1a1 locus in ESCs. (b) Brightfield and fluorescent images of ESCs targeted with shControl and shPhf5a mir-30 shRNA expression cassettes at the Col1a1 locus following the addition of doxycycline. Scale bars, 100 μm. (c) Comparison of Phf5a expression levels by qRT-PCR in ESCs targeted with shControl and shPhf5a mir-30 shRNA expression cassettes at the Col1a1 locus following the addition of doxycycline. n=3 biologically independent replicates (see Supplementary Table 5). shPhf5a-2 hairpin: **p=0.0072 and shPhf5a-3 hairpin *p=0.008, respectively, two-sided Student’s t-test, values represent the mean ± s.d. (d) Western blot analysis of Brachyury protein levels following induction of mesoderm differentiation in the presence or absence of Phf5a. (i) Levels of several mesoderm markers following shPhf5a depletion. n=3 biologically independent replicates (see Supplementary Table 5). Brachyury: **p=0.0078, Msgn: n.s: non-significant p=0.0694, Nkx2-5: **p=0.0010, n.s: non-significant Isil: p=0.2223, two-sided Student’s t-test, values represent the mean ± s.d. (k and l) Proliferation assays of ESCs (k) or MEFs (l), respectively, following shControl or shPhf5a depletion. n=3 biologically independent replicates (see Supplementary Table 5). ESCs Day2 and Day4: **p=0.0001, respectively, MEFs Day2 and Day4: n.s: non-significant, p=0.4169 and p=0.8769, respectively, two-sided Student’s t-test, values represent the mean ± s.d. (m and n) Expression levels of pluripotency factors (m) and western blot for Nanog protein levels (n), respectively, during ESC differentiation, in the presence or absence of ectopic expression of Phf5a. n=3 biologically independent replicates (see Supplementary Table 5). Phf5a: **p=0.0016, Nanog: **p=0.0451, Pou5f1: **p=0.0335, Sox2: p=0.0485, Zfp42: n.s: non-significant p=0.0908, NrOb1: p=0.0209, two-sided Student’s t-test, values represent the mean ± s.d. (o) Representative iPSC colony morphology and AP-staining following shPhf5a knockdown. Scale bars, 100 μm. (p) Western blot analysis of reprogramming markers in OKSM MEFs following shPhf5a knockdown, on day 14 post-induction doxycycline.
Supplementary Figure 3 Phf5a interacts with the Paf1C. (a) Western blot analysis of Phf5a and control tagged-proteins following doxycycline induction and cytoplasmic or nuclear fractionation in ESCs. (b) Flag immunofluorescence showing nuclear localization of tagged-Phf5a following doxycycline induction in ESCs. Scale bars, 100 μm. (c) Western blot analysis of Paf1C subunits in ESC differentiation. (d) Western blot analysis of interacting proteins following doxycycline induction and streptactin purification of Paf1C subunits or GFP control in ESCs. (e) Western blot analysis of Phf5a- or control-interacting proteins following doxycycline induction and tandem affinity purification using strep-tagII and flag tags in ESCs. (f) Representative time point analysis for presence of nucleic acids following lysis of ESCs in the presence of benzonase nuclease prior to immunoprecipitation in ESCs.
**Supplementary Figure 4** Phf5a regulates Paf1C subunit composition. (a) Venn diagrams showing number of all significantly differentially expressed genes, as well as significantly downregulated and upregulated genes using RNA-sequencing after shPhf5a and shPaf1 depletion in ESCs. (b) Western blot of fractions following glycerol gradient sedimentation analysis of Paf1C subunits using one-step purification from ESCs in the presence or absence of shPhf5a knockdown. (c-e) Graphs showing quantification of percent distribution for the Paf1C subunits Paf1, Wdr61 and Cdc73 shown in (c), in the presence or absence of Phf5a, respectively, in ESCs. (f) Western blot of fractions following glycerol gradient sedimentation analysis of Swi/Snf complex subunit Smarca4 and NELF complex subunit NELF-A using one-step purification from ESCs in the presence or absence of shPhf5a knockdown. (g) Binding profiles for genomic distribution of Leo1 peaks (upstream, promoter, coding region, 5'UTR, 3'UTR, downstream and intergenic) in ESCs, showing preferential binding (42%) within gene bodies of downregulated genes, but preferential binding (31%) within promoters of upregulated genes. (h) A Tet-inducible “knock-in” tagged Phf5a ESC line was used to perform ChIP-sequencing of Phf5a using a HA-epitope in the presence of absence of doxycycline. Snapshots of Phf5a binding on representative pluripotency gene targets are shown (Pou5f1, Sall4, Prdm14, Esrrb) in the presence (blue) of absence (gray) of doxycycline.
Supplementary Figure 5 Phf5a regulates Paf1C functions on transcriptional elongation in ESCs. (a) Western blot analysis of Ser2-phospho- and Total RNA PolII in ESCs following CRISPR-Cas9 mediated Phf5a depletion. (b) Comparison of GRO-seq read density profiles of downregulated genes 72h following shControl or differentiated ESCs in the absence of LIF. RPKM: Reads Per Kilobase per Million total reads. (c) Comparison of GRO-seq read density profiles of downregulated genes 72h in all conditions tested (shControl, shPhf5a or differentiated ESCs in the absence of LIF). RPKM: Reads Per Kilobase per Million total reads. (d) Box plot showing comparison of log2 pausing index for downregulated genes, 72h following shControl, shPhf5a, ESC differentiated in the absence of LIF, or flavopiridol-treated ESCs, respectively, using GRO-seq analysis. Flavopiridol treatment is used as a positive control of pause-release block. n=3 biologically independent replicates, Wilcoxon signed rank test non-parametric. (e) (Upper) Venn diagrams between Leo1 ChIP-sequencing targets in ESCs and either downregulated or upregulated genes using RNA-sequencing after shPhf5a knockdown. (Lower) Box plot showing pausing index ratios after GRO-seq analysis for the direct Leo1 targets shown above. Only downregulated targets exhibit significant promoter-proximal pausing. n=3 biologically independent replicates, Wilcoxon signed rank test non-parametric. (f and g) Comparison of read density profiles for Ser5- (on TSSs) or Ser2- (on gene bodies) phosphorylated RNA PolII, respectively. 

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
**Supplementary Information**

P-Value t-test, values represent the mean ± s.d. n=3 biologically independent replicates (see Supplementary Table 5). Phf5a: p=0.0193, Paf1: p=0.0001, Cdc73: p=0.0031, Leo1: p=0.0378, Wdr61: p=0.0238, Rtf1: n.s, non-significant, p=0.1103, two-sided Student's t-test, values represent the mean ± s.d.

**(a)** Brightfield images of self-renewing myoblasts or 72h-differentiated myotubes. (b) Western blot analysis of self-renewal and differentiation markers in C2C12 cells following knockdown of Paf1C subunits and 72h differentiation. (c) Myocin heavy chain (MHC) immunofluorescence on differentiated myotubes for 72h following shControl or shPhf5a knockdown. Scale bars, 100μm.

**(d)** Quantification of myocin heavy chain (MHC) immunofluorescence intensity of *Rosa26rtTA*/*Col1a1TRE* shPhf5a primary myotubes following addition of doxycycline. n=3 biologically independent replicates (see Supplementary Table 5). **p<0.0001, one-tailed Student's t-test, values represent the mean ± s.d.**

**(e)** Relative expression of Phf5a and multiple Paf1C subunits using qRT-PCR showing higher levels of most of Paf1C subunits in primary mouse myotubes. *n=3* biologically independent replicates (see Supplementary Table 5). Phf5a: p=0.0193, Paf1: p=0.0001, Cdc73: p=0.0031, Leo1: p=0.0378, Wdr61: p=0.0238, Rtf1: n.s, non-significant, p=0.1103, two-sided Student's t-test, values represent the mean ± s.d.

**(f)** Desmin immunofluorescence on 72h differentiated myotubes following CRISPR-Cas9 depletion of Phf5a. Scale bars, 100μm.

**(g)** Western blot analysis of the differentiation marker myocin heavy chain following CRISPR-Cas9 depletion of Phf5a. (h-j) Gene Ontology (GO) analysis of Leo1-bound genes after ChIP-sequencing in myoblasts (h) and myotubes (i) under shControl conditions, or myotubes (j) after shPhf5a depletion, respectively.
Supplementary Figure 7 Scans of unprocessed key blots. (a) Scans from Figure 1a (b) Scans from Figure 1i (c) Scans from Figure 2d (d) Scans from Figure 3b (e) Scans from Figure 3c (f) Scans from Figure 3d (g) Scans from Figure 3e (h) Scans from Figure 4a (i) Scans from Figure 5a (j) Scans from Figure 7a (k) Scans from Figure 7e (l) Scans from Figure 7j.
Supplementary Table Legends

Supplementary Table 1 Affymetrix microarray expression data of ESCs following shControl or shPhf5a knockdown, respectively.

Supplementary Table 2 Phf5a interacting proteins identified following purification and mass spectrometry in ESCs.

Supplementary Table 3 Differentially expressed genes in ESCs using RNA-sequencing following shPhf5a and shPaf1 knockdown, respectively.

Supplementary Table 4 Analysis of alternative splicing using Multivariate Analysis of Transcript Splicing (rMATs).

Supplementary Table 5 Statistics source data.