Citrullination regulates pluripotency and histone H1 binding to chromatin

Maria A. Christophorou[1,*], Gonçalo Castelo-Branco[1,3*], Richard P. Halley-Stott[1,4], Clara Slade Oliveira[1,5], Remco Loos[6], Aliaksandra Radzisheuskaya[7,8], Kerri A. Mowen[9], Paul Bertone[6,7,10], José Silva[7,8], Magdalena Zernicka-Goetz[1], Michael L. Nielsen[11], John Gurdon[1,4] and Tony Kouzarides[1,2]

Affiliations:

[1] The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, United Kingdom

[2] Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, United Kingdom

[3] Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

[4] Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

[5] EMBRAPA Dairy Cattle Research Center, Juiz de Fora, Brazil

[6] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, United Kingdom

[7] Wellcome Trust - Medical Research Council Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, United Kingdom

[8] Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, United Kingdom

[9] Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA

[10] Genome Biology and Developmental Biology Units, European Molecular Biology Laboratory,

Meyerhofstraβe 1, 69117 Heidelberg, Germany

[11] Department of proteomics, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health Sciences, Blegdamsvej 3b, DK-2200 Copenhagen, Denmark

[*] These authors contributed equally to this work.

Author contributions:

M.A.C., G.C-B. and T.K. conceived the idea for this project, designed experiments and wrote the manuscript with the help of all the authors. G.C-B. and M.A.C performed ES cell transductions, established transgenic pre-iPS and ES cell lines and performed gene expression analyses. M.A.C carried out mutagenesis, protein isolation, biochemical and chromatin immunoprecipitation experiments, and performed citrullination analyses with the help of K.A.M.; G.C-B performed reprogramming experiments, with the help of J.S. and A.R. M.L.N. and M.A.C. performed Mass Spectrometric analyses. R.P.H-S and M.A.C. performed PADI4 treatments of permeabilized cells and subsequent chromatin compaction analyses, with the help of J.B.G. C.S.O and M.Z-G designed and performed mouse embryo experiments. R.L. and P.B. performed bioinformatic analyses of microarray data. T.K. supervised the study.

Abstract

Citrullination is the post-translational conversion of an arginine residue within a protein to the non-coded amino acid citrulline[1]. This modification leads to the loss of a positive charge and reduction in hydrogen bonding ability. It is carried out by a small family of tissue-specific vertebrate enzymes called peptidylarginine deiminases (PADIs)[2] and is associated with the development of diverse pathological states such as autoimmunity, cancer, neurodegenerative disorders, prion diseases and thrombosis[2,3]. Nonetheless, the physiological functions of citrullination remain ill-defined, though citrullination of core histones has been linked to transcriptional regulation and the DNA damage response [4-8]. PADI4 (or PAD4/PADV), the only PADI with a nuclear localization signal[9], was previously shown to act in myeloid cells where it mediates profound chromatin decondensation during the innate immune response to infection[10]. Here we show that the expression and enzymatic activity of PADI4 are also induced under conditions of ground state pluripotency and during reprogramming. PADI4 is part of the pluripotency transcriptional network, binding to regulatory elements of key stem cell genes and activating their expression. Its inhibition lowers the percentage of pluripotent cells in the early mouse embryo and significantly reduces reprogramming efficiency. Using an unbiased proteomic approach we identify linker histone H1 variants, which are involved in the generation of compact chromatin[11], as novel PADI4 substrates. Citrullination of a single arginine residue within the DNA binding site of H1 results in its displacement from chromatin and global chromatin decondensation. Together, these results uncover a role for citrullination in the regulation of pluripotency and provide new mechanistic insights into how citrullination regulates chromatin compaction.

Pluripotent cells have the capacity to self-renew and differentiate into all somatic and germ cell lineages and, hence, possess immense therapeutic potential for a multitude of medical conditions. Their generation by reprogramming of differentiated somatic cells has been achieved by nuclear transfer, cell fusion and transduction of transcription factors, such as Oct4, Sox2, Klf4 and c-Myc[12]. Pluripotent cells have a distinctly open chromatin structure that is essential for unrestricted developmental potential[13,14] and reprogramming involves an almost complete epigenetic resetting of somatic cells[13]. The ability of PADI4-mediated histone citrullination to induce chromatin decondensation in neutrophils[10] prompted us to ask whether it can play a role in pluripotency, where chromatin decondensation is also necessary. To investigate this we first assessed the expression of *Padi4* in the embryonic stem cell line ES Oct4-GIP (ES), the neural stem cell line NSO4G (NS) and in induced pluripotent stem (iPS) cells derived NSO4G (see Methods). *Padi4* is expressed in pluripotent ES and iPS but not multipotent NS cells (Fig.1a). Culture of ES cells in 2i/LIF medium establishes a ground state of pluripotency[15]. This leads to the down-regulation of lineage specific markers and the up-regulation of pluripotency factors, as well as rapid induction of *Padi4* (Fig. 1b). The pattern of *Padi4* expression follows closely that of *Nanog,* an essential transcription factor for the transition to ground state pluripotency[16](Fig 1a,b). While other PADIs are expressed in pluripotent cells, PADI4 is the only one whose expression clearly associates with naïve pluripotency (Extended Data Fig. 1a,b). Citrullination of histone H3 (H3Cit), a modification shown previously to be carried out specifically by PADI4[17], is detectable in ES and iPS cells (Extended Data Fig. 1c), indicating that PADI4 is also enzymatically active. H3 and global citrullination are undetectable in NS cells (Extended Data Fig. 1c,d).

To determine the kinetics of PADI4 activation during the establishment of pluripotency, we examined RNA and protein samples collected daily during the course of reprogramming of NSO4G into iPS cells[16]. *Padi4* is rapidly induced in NS cells after transduction of reprogramming

factors but only becomes active to citrullinate H3 after introduction of 2i/LIF, closely following the onset of *Nanog* expression (Fig. 1c). These observations strongly suggested that PADI4 activity is associated with ground state pluripotency and prompted us to examine whether *Padi4* is part of the pluripotency transcriptional network.

First, we asked whether the reprogramming factors regulate *Padi4* expression, utilizing the ZHBTc4.1 and 2TS22C cell lines where Oct4 and Sox2, respectively, can be deleted acutely in response to doxycycline treatment (see Methods). Deletion of Oct4, but not Sox2, led to a decrease in *Padi4* mRNA levels (Extended Data Fig. 1e). Furthermore, while Oct4 and Klf4 occupy the *Padi4* promoter in ES, but not NS cells, Sox2 is bound in both cell types (Extended Data Fig. 1f). To understand the effects of PADI4 on transcriptional regulation in pluripotent cells, we analyzed the transcriptome of ES cells upon PADI4 over-expression and inhibition. Several key pluripotency genes are up-regulated in response to PADI4 over-expression (Fig. 1d, Extended Data Figure 2a and Supplementary Table 2), including *Klf2, Tcl1, Tcfap2c*, and *Kit*. *Tcl1* was previously identified as the only regulator of self-renewal up-regulated in ground state pluripotency[15] and over-expression of *Tcl1* or *Tcfap2c* positively influence this process[18]. Gene Ontology (GO) analysis of this dataset indicates an enrichment of genes involved in stem cell development and maintenance (Fig. 1e). In addition, knockdown of *Padi4* in mES cells leads to decreased expression of *Tcl1* and *Nanog,* which is rescued by exogenous expression of RNAi-resistant human PADI4 (Fig. 1f, Extended Data Fig. 2b,c). These genes are under the control of PADI4 enzymatic activity since treatment with the chemical inhibitor Cl-amidine, which disrupts citrullination by PADI4[19], down-regulates their expression (Fig. 1g). Chromatin immunoprecipitation (ChIP) analysis indicated that H3Cit is present on regulatory regions of *Tcl1* and *Nanog* in ES and iPS cells, but not NS cells (Fig. 1h and Extended Data Fig. 3a). Accordingly, exogenously expressed human PADI4 localizes to and is enzymatically active on these regions, as well as regulatory regions of *Klf2* and *Kit* in ES cells (Extended Data Fig.

3b,c). In contrast to PADI4 over-expression, treatment of ES cells with Cl-amidine led to up-regulation of differentiation markers such as *Prickle1, EphA1* and *Wnt8a* and down-regulation of pluripotency markers such as *Klf5* (Extended Data Fig. 4a,b, and Supplementary Table 3), in addition to *Nanog* and *Tcl1* (Fig. 1g). GO analysis of this dataset indicated enrichment in genes involved in cell differentiation (Extended Data Fig. 4c). *Pou5f1 (Oct4)*, *Klf4*, *Sox2* and *c-Myc*, were not affected by PADI4 modulation (Extended Data Fig. 2a and Supplementary Tables 2 and 3). Cumulatively, the above results place PADI4 within the pluripotency transcriptional network, suggesting that it acts downstream of some of the cardinal reprogramming factors to regulate a specific subset of pluripotency genes.

Prompted by the above observations, we investigated whether PADI4 is necessary for pluripotency, as assessed during reprogramming (Extended Data Fig. 5a) and in the early stages of embryo development. NSO4G cells express a GFP reporter under the control of the Oct4 regulatory sequences, which is activated upon acquisition of pluripotency[16], allowing us to trace reprogrammed cells. Knock-down of *Padi4* in NSO4G-derived pre-iPS cells impaired the ability of the cells to establish H3Cit upon switch to 2i/LIF medium and led to a clear reduction in reprogramming (Fig. 2a,b, Extended Data Fig. 5b-e and Extended Data time-lapse video). Consistent with this, levels of *Tcl1* and *Nanog* were not elevated upon reprogramming to the same extent as in control cells (Fig. 2b and Extended Data Fig. 5f). Cl-amidine treatment led to a dramatic reduction of reprogramming efficiency and H3Cit (Fig. 2c and Extended Data Fig. 5g-i), suggesting that the catalytic activity of PADI4 is important for the induction of pluripotency.

Padi4 expression and H3Cit are detected in the early embryo[20,21] and *Padi4-null* mice are born in lower numbers than would be expected by Mendelian inheritance[22], suggesting that PADI4 loss affects embryonic development. To assess the role of PADI4 in early development, we cultured mouse embryos in Cl-amidine-containing medium from the 2-cell stage and throughout

pre-implantation development (see Methods and Extended Data Fig. 6a-c). Using 200μM Cl-amidine resulted in a complete developmental arrest of the embryos at the 8-cell stage (Extended Data Fig. 6a). We therefore used the maximum dose of Cl-amidine that reduced H3Cit (Extended Data Fig. S6b,c) but did not induce arrest (10μM). This led to a reduced percentage of pluripotent Nanog-positive epiblast cells and an increased percentage of differentiated trophectoderm cells at the blastocyst stage (Fig. 2d,e and Extended Data Fig. 6d,e). Time-course analyses of the cleavage patterns and cell fate decisions in early embryos showed that Cl-amidine increased the number of symmetric cell divisions at the expense of asymmetric divisions at the 8 to 16- and 16 to 32-cell transitions (Fig. 2f,g). This resulted in 16-cell stage embryos with fewer inner cells (destined for pluripotency) and greater numbers of outer cells (destined for differentiation into trophectoderm, reviewed in [23]) (Extended Data Fig. 6f). Treatment with another PADI4 inhibitor, TDFA[24], but not the HDAC inhibitor TSA, had similar effects (Extended Data Fig. 7, 8). These results indicate that PADI4 activity also promotes the maintenance of pluripotent cells in the early mouse embryo.

To elucidate the molecular mechanisms by which PADI4 regulates pluripotency, we aimed to identify PADI4 substrates in the chromatin fraction of mES cells using Stable Isotope Labeling of Amino acids in Cell culture (SILAC) (Fig. 3a and Extended Data Fig. 9a,b). Among the identified PADI4 substrates were Atrx, Dnmt3b, Trim28 and variants of linker histone H1 (Fig. 3b-e, Extended Data Fig. S9c-f, S10 and Supplementary Table 4), all of which can impact pluripotency. Histone H1 stabilizes the nucleosome and facilitates chromatin condensation, a state that is less permissive to processes that require access to the DNA, such as transcription[11]. The identified citrullinated H1 peptides correspond to, and are common between, variants H1.2, H1.3 and H1.4 (Fig. 3d,e), while an additional peptide corresponds to the same residue in H1.5 (Extended Data Fig. 10a,b). In ES cells, H1.2, H1.3 and H1.4 are required for chromatin compaction[25], while their depletion leads to increased expression of pluripotency

genes such as *Nanog* and stalls them in a self-renewal state with impaired differentiation capability[26]. Notably, H1 is more loosely bound to chromatin in ES cells than in differentiated cells[14] and its genomic localization in cancer cells was shown to anti-correlate with that of PADI4[7].

Mass spectrometric analysis accounted for all arginine residues within H1 but indicated that Arg54 (H1R54) is the only site citrullinated by PADI4 (Fig. 3e and Extended Data Fig. 9c). Indeed, we found that while H1.2 is citrullinated in ES cells (already by endogenous PADI4, and significantly increased upon PADI4 over-expression), it is refractory to modification when Arg54 is mutated (Fig. 3f). Similar results were obtained in *in vitro* citrullination assays (Extended Data Fig. 11a). H1R54 lies within the globular domain of H1 (Extended Data Fig. 11b), which is highly conserved among the linker histone family and is necessary for interaction with nucleosomal DNA[27,28]. To test whether H1R54 is necessary for binding of H1 to nucleosomes, we mutated and assessed it in nucleosome-binding assays. Figure 3g shows that an R54A mutant, which mimics the charge change that accompanies citrullination, is impaired for nucleosome binding. An R54K mutant, which retains the positive charge, is impaired to a lesser extend (Fig. 3g) suggesting that H1R54 is important for electrostatic interactions between H1.2 and the nucleosome.

The above results open up the possibility that PADI4 may affect chromatin compaction in pluripotent cells. To test this hypothesis, we first assessed whether citrullination by ectopic PADI4 can lead to decondensation of differentiated cell chromatin. Recombinant PADI4 was added to permeabilized and stabilized differentiated C2C12 mouse myoblast nuclei (Fig. 4a). This protocol ensures stabilization of the nuclear component while allowing the free diffusion of non-chromatin bound nuclear proteins into the extra-nuclear fraction, and their collection by washing. Incubation with active PADI4 (Extended Data Fig. 12a,b), leads to the eviction of H1

from the chromatin and its diffusion out of the permeabilized nucleus (Fig. 4b). The evicted H1 is citrullinated on R54, as determined by mass spectrometry (Extended Data Figure 12c,d). Consistent with this, PADI4-treated cells showed evidence of decondensed chromatin, as determined by nuclear swelling, diffuse DAPI staining and increased sensitivity to micrococcal nuclease (Fig. 4c,d and Extended Data Figure 12e). Similar results were observed when *PADI4* was over-expressed in C2C12 cells (Fig. 4e) or NS cells (data not shown). To monitor if PADI4 can affect H1 binding on pluripotent cell chromatin, we performed ChIP-qPCR analyses of H1.2 on the regulatory regions of *Tcl1* and *Nanog* and found that it is stabilized upon *Padi4* knockdown (Fig. 4f). The ability of PADI4 to disrupt the binding of H1 to nucleosomal DNA provides a novel mechanistic example of how citrullination regulates protein function and chromatin condensation.

The work described above identifies citrullination of chromatin components by PADI4 as a feature of pluripotency (Fig. 4g), in addition to its previously described role in the myeloid lineage. One of the reasons for the restricted expression pattern of PADI4 may be the requirement for an open chromatin state in these cell types. The selective expression characteristics and the inducible nature of the catalytic activity of PADI4, suggest that it is under tight spatial and temporal regulation, giving it a unique status among chromatin modifying enzymes. As such, inappropriate PADI4 activity may have deleterious consequences, which may explain its activation in cancers of varying origin during progression to malignancy[29]. Indeed, citrullination is a common feature of several unrelated diseases, suggesting that strict regulation is likely to be a requirement for the physiological function of all PADIs. During review of this manuscript, Coonrod and colleagues suggested that PADI2, thought to be mainly cytoplasmic, can also citrullinate histones and lead to transcriptional activation[30]. This opens the intriguing possibility that other PADIs may mediate nuclear events in specific contexts, including

in pluripotent cells. Further research into the function and targets of PADIs is likely to shed light into the etiology of several pathologies.

Supplemental Information

Complete experimental methods and references in Methods file.
Extended Data figure legends in “Extended Data figure legends” file.
Primer sequences in Supplementary Table S1.
Complete microarray data for PADI4 over-expression in ES cells in Supplementary Table S2.
Complete microarray data for PADI4 inhibition in ES cells in Supplementary Table S3.
Complete Mass Spec dataset in Supplementary Table S4.
Time-lapse video of reprogramming experiments in Supplementary video file.

Acknlowledgements

This work was funded by programme grants from Cancer Research UK (T.K.) and EMBL (P.B., R.L.). R.P.H-S and JBG are supported by the Medical Research Council [G1001690] and the Wellcome Trust. G.C.-B. was funded by EMBO (Long-Term Post-Doctoral Fellowship), European Union (FP7 Marie Curie Intra-European Fellowship for Career Development) and Swedish Research Council. M.A.C was funded by an EMBO Long-Term Post-Doctoral Fellowship and a Human Frontier Science Programme Long-Term Post-Doctoral Fellowship. C.S.O was supported by FAPESP (Foundation for Research Support of the State of São Paulo) and mouse embryo work was supported by the Wellcome Trust programme grant to M.Z.G. M.L.N was partly supported by the Novo Nordisk Foundation Center for Protein Research, the Lundbeck Foundation, and by and the European Commission's 7th Framework Programme HEALTH-F7-2010-242129/SYBOSS. We would like to thank Sri Lestari, Alistair Cook and Cynthia Hill for technical assistance; Paul Thompson for kindly providing the TDFA compound; GSK Epinova for Cl-amidine; Till Bartke for the kind gift of histone octamers and help with nucleosome pull-down assays; Andrew Finch for help with FPLC chromatography; Agnieszka Jedrusik for help with embryo work; Rachael Walker at the Flow Cytometry Core Facility at Wellcome Trust Centre for Stem Cell Research, University of Cambridge and Thor Theunissen for help with the flow cytometry; and members of the Kouzarides laboratory for critical discussions of the work. 2TS22C cells were kindly provided by Dr. Hitoshi Niwa, at the RIKEN Center for Developmental Biology, Kobe, Japan. The ChIP grade H1.2 antibody was a generous gift from Prof. Arthur Skoultchi, Albert Einstein School of Medicine, New York, USA.

References

1 Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *BioEssays : news and reviews in molecular, cellular and developmental biology* 25, 1106-1118, doi:10.1002/bies.10357 (2003).

2 Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim Biophys Acta* 1829, 1126-1135, doi:10.1016/j.bbagrm.2013.07.003 (2013).

3 Martinod, K. *et al.* Neutrophil histone modification by peptidylarginine deiminase 4 is critical for deep vein thrombosis in mice. *Proc Natl Acad Sci U S A* 110, 8674-8679, doi:10.1073/pnas.1301059110 (2013).

4 Hagiwara, T., Nakashima, K., Hirano, H., Senshu, T. & Yamada, M. Deimination of arginine residues in nucleophosmin/B23 and histones in HL-60 granulocytes. *Biochemical and biophysical research communications* 290, 979-983, doi:10.1006/bbrc.2001.6303 (2002).

5 Cuthbert, G. L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* 118, 545-553, doi:10.1016/j.cell.2004.08.020 (2004).

6 Wang, Y. *et al.* Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* 306, 279-283, doi:10.1126/science.1101400 (2004).

7 Zhang, X. *et al.* Genome-wide analysis reveals PADI4 cooperates with Elk-1 to activate c-Fos expression in breast cancer cells. *PLoS genetics* 7, e1002112, doi:10.1371/journal.pgen.1002112 (2011).

8 Tanikawa, C. *et al.* Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nature communications* 3, 676, doi:10.1038/ncomms1676 (2012).

9 Asaga, H., Nakashima, K., Senshu, T., Ishigami, A. & Yamada, M. Immunocytochemical localization of peptidylarginine deiminase in human eosinophils and neutrophils. *Journal of leukocyte biology* 70, 46-51 (2001).

10 Neeli, I., Khan, S. N. & Radic, M. Histone deimination as a response to inflammatory stimuli in neutrophils. *J Immunol* 180, 1895-1902 (2008).

11 Buttinelli, M., Panetta, G., Rhodes, D. & Travers, A. The role of histone H1 in chromatin condensation and transcriptional repression. *Genetica* 106, 117-124 (1999).

12 Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* 465, 704-712, doi:10.1038/nature09229 (2010).

13 Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nature reviews. Molecular cell biology* 12, 36-47, doi:10.1038/nrm3036 (2011).

14 Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* 10, 105-116, doi:10.1016/j.devcel.2005.10.017 (2006).

15 Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* 149, 590-604, doi:10.1016/j.cell.2012.03.026 (2012).

16 Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr Biol* 21, 65-71, doi:S0960-9822(10)01584-8 [pii]
10.1016/j.cub.2010.11.074 (2011).

17 Darrah, E., Rosen, A., Giles, J. T. & Andrade, F. Peptidylarginine deiminase 2, 3 and 4 have distinct specificities against cellular substrates: novel insights into

autoantigen selection in rheumatoid arthritis. *Annals of the rheumatic diseases* 71, 92-98, doi:10.1136/ard.2011.151712 (2012).

18 Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151, 1617-1632, doi:10.1016/j.cell.2012.11.039 (2012).

19 Luo, Y. *et al.* Inhibitors and inactivators of protein arginine deiminase 4: functional and structural characterization. *Biochemistry* 45, 11727-11736, doi:10.1021/bi061180d (2006).

20 Brahmajosyula, M. & Miyake, M. Localization and expression of peptidylarginine deiminase 4 (PAD4) in mammalian oocytes and preimplantation embryos. *Zygote*, 1-11, doi:10.1017/S0967199411000633 (2011).

21 Kan, R. *et al.* Potential role for PADI-mediated histone citrullination in preimplantation development. *BMC developmental biology* 12, 19, doi:10.1186/1471-213X-12-19 (2012).

22 Li, P. *et al.* PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. *J Exp Med* 207, 1853-1862, doi:10.1084/jem.20100239 (2010).

23 Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* 10, 467-477, doi:10.1038/nrg2564 (2009).

24 Jones, J. E. *et al.* Synthesis and screening of a haloacetamidine containing library to identify PAD4 selective inhibitors. *ACS chemical biology* 7, 160-165, doi:10.1021/cb200258q (2012).

25 Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123, 1199-1212, doi:10.1016/j.cell.2005.10.028 (2005).

26 Zhang, Y. *et al.* Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS genetics* 8, e1002691, doi:10.1371/journal.pgen.1002691 (2012).

27 Goytisolo, F. A. *et al.* Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J* 15, 3421-3429 (1996).

28 Brown, D. T., Izard, T. & Misteli, T. Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat Struct Mol Biol* 13, 250-255, doi:10.1038/nsmb1050 (2006).

29 Chang, X. *et al.* Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC cancer* 9, 40, doi:10.1186/1471-2407-9-40 (2009).

30 Zhang, X. *et al.* Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor alpha target gene activation. *Proc Natl Acad Sci U S A* 109, 13331-13336, doi:10.1073/pnas.1203280109 (2012).

Figure Legends

Figure 1: PADI4 expression and activity are features of pluripotent cells

(a,b) qRT-PCR for *Padi4* and *Nanog* expression in ES, NS and iPS cells **(a)**, and in ES cells upon culture in 2i/LIF for one passage **(b)**. *Pou5f1, Olig2* and *Pax6* are presented as controls. Expression normalized to *Ubiquitin (UbC)*. Error bars: standard error of the mean of three biological replicates.

(c) qRT-PCR for *Padi4* and *Nanog* expression and H3Cit immunoblot during the course of reprogramming (see also Extended Data Fig. 5a). Loading control: total histone H3. Representative of four experiments.

(d) Heat map of the genes regulated upon *hPADI4* over-expression in mES cells, as determined by microarray analysis. Displayed values are normalized log intensities, minus the mean expression of the gene across the four samples. Hierarchical clustering based on correlation.

(e) Gene Ontology for Biological Process (GOBP) analysis of the above microarray dataset. *P*-value is corrected for multiple testing using Benjamini and Hochberg False Discovery Rate (FDR).

(f,g) qRT-PCR for *Tcl1* and *Nanog* expression in mES cells after transient knock-down with *Padi4* or control (Ctrl) shRNAs, and over-expression of human *PADI4* or control vector (pPB CTRL) **(f)**, and after treatment with 200μM Cl-amidine **(g)**. Expression normalized to *UbC*. Error bars: standard error of the mean of three biological replicates.

(h) ChIP-qPCR for H3Cit on regulatory regions of *Tcl1* and *Nanog* in mES, NS and iPS cells. Error bars: standard deviation of three technical qPCR replicates. Representative of three experiments.

Asterisks denote difference with ES cells (a) or media (b), Control (f, g) and between samples (f); - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001, by ANOVA (a,f) or t-test (b,g).

Figure 2: Citrullination and PADI4 regulate pluripotency during reprogramming and early embryo development

(a) Flow cytometry analysis and phase contrast/fluorescence images for the assessment of Oct4-GFP reporter expression after reprogramming of pre-iPS cells stably expressing *Padi4* and Ctrl shRNAs. Representative of four independent experiments. Time-lapse video in supplementary data online.

(b) qRT-PCR for expression of *Tcl1*, *Nanog* and *Padi4* at the end of the above reprogramming assay. Error bars: standard error of the mean of triplicate samples.

(c) Flow cytometry analysis and phase contrast/fluorescence images for the assessment of Oct4-GFP reporter expression after reprogramming assay of pre-iPS cells treated with 200μM Cl-amidine. Representative of three independent experiments.

(d) E4.5 blastocysts from 2-cell stage embryos treated with 10μM Cl-amidine. SOX17 (primitive endoderm marker, red), Cdx2 (trophectoderm marker, green), Nanog (epiblast marker, white) and HOECHST 33342 (blue).

(e) Distribution of inner cell mass versus trophectoderm cells in E3.5 blastocyst treated as above.

(f,g) Time-lapse analysis of embryos in 10μM Cl-amidine from 2-cell stage. Number of symmetric versus asymmetric divisions at the 8 to 16-cell transition **(f)** and type of divisions at the 16 to 32-cell transition **(g)**. Error bars: standard error of the mean.

Statistical significance was determined by unpaired t-test (b), or Mann Whitney test upon non-normal distribution (e-g). Asterisks denote difference with Control; - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001.

Figure 3: PADI4 citrullinates Arg54 on linker histone H1 and affects its binding to nucleosomal DNA

(a) Experimental strategy for screening for PADI4 citrullination substrates in the chromatin fraction of ES cells.

(b) Scatter plot representing the overall fold change for all identified citrullination sites. Red diamonds: PADI4-regulated citrullinations.

(c) Table representing the 40 most highly regulated PADI4 substrates, their individual citrullination sites and the $\log_2$ SILAC ratio. Complete dataset in Supplementary Table 4.

(d) Quantification of citrullination site R54 on H1.2 through differential regulation of the triply charged peptide ERSGVSLAALKK.

(e) Fragmentation spectra of the triply charged and heavy SILAC labeled LysC peptide ERSGVSLAALKK surrounding Arginine 54 of H1.2. The y and b series indicate fragments at amide bonds of the peptide.

(f) Citrullination immunoblot of wild-type and R54A mutant GFP-tagged H1.2 expressed and pulled-down from ES cells expressing *PADI4* or control vector (Mock). Control for the efficiency of the pull-down: GFP.

(g) Nucleosome pull-down assay using wild-type and R54-mutant H1.2. WB: Western Blot.

Figure 4: PADI4 evicts histone H1 from chromatin and affects chromatin condensation

(a) Schematic representation of treatment of C2C12 myoblast nuclei with recombinant PADI4.

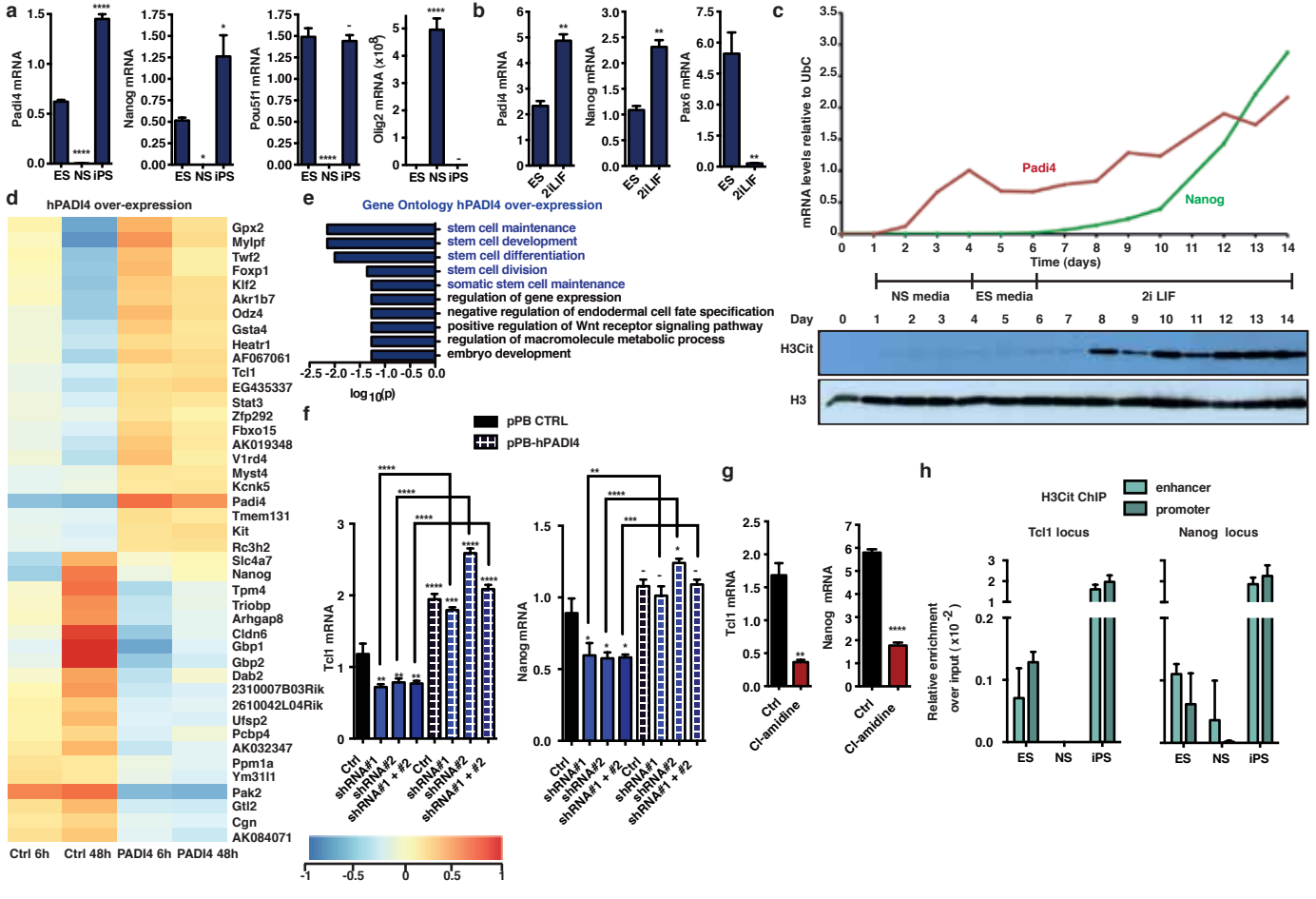
(b) Immunoblot analysis of the wash fraction after the above treatment, for histone H1.2.

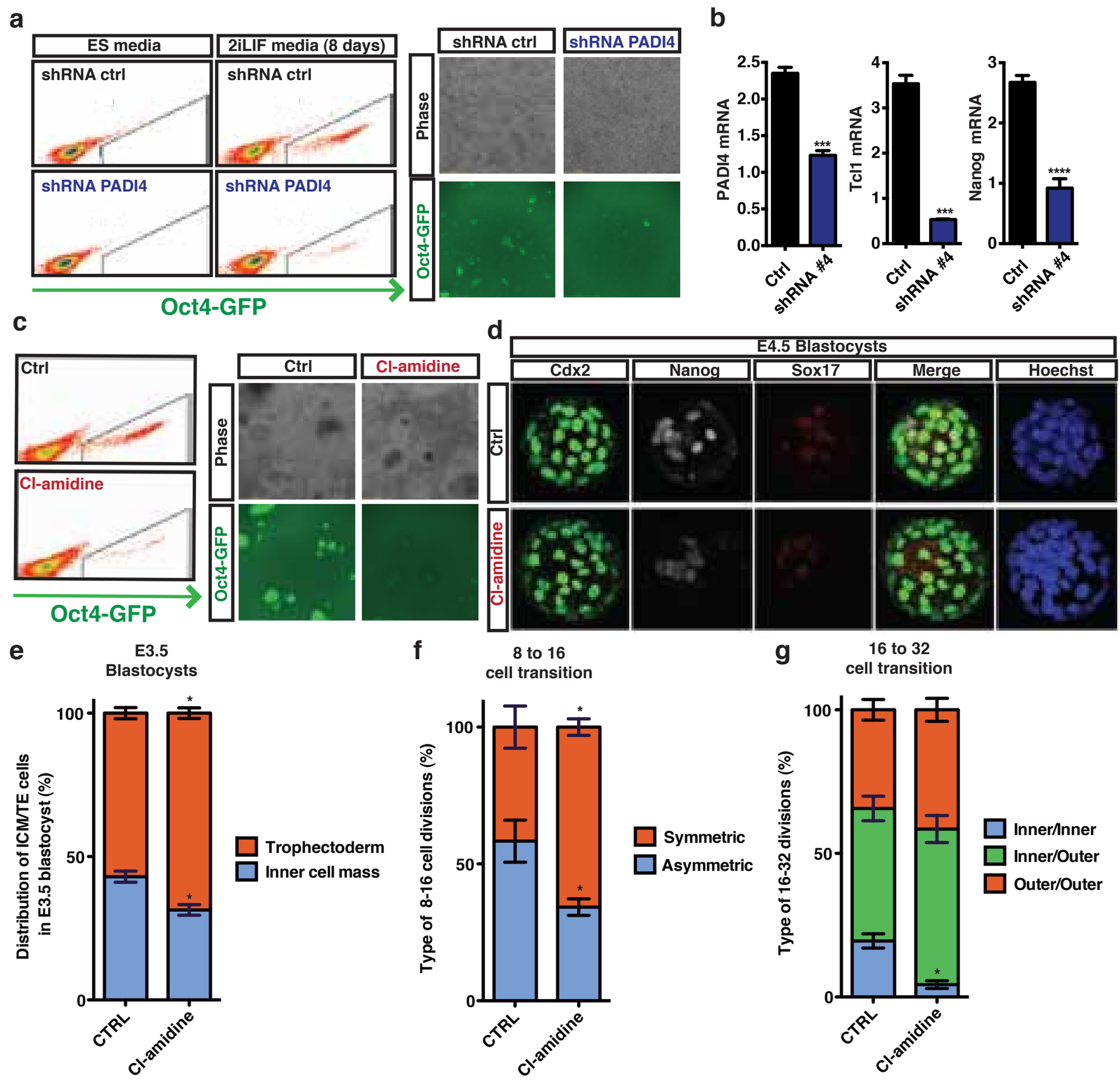
(c,d) Quantification of nuclear volume **(c)** and representative DAPI fluorescence **(d)** upon treatment of permeabilized C2C12 nuclei with recombinant PADI4. Error bars: standard error of the mean. Statistical significance determined by unpaired student t-test.

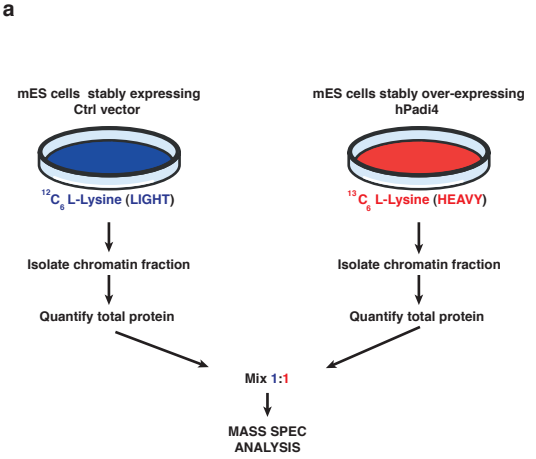
(e) Micrococcal nuclease digestion of C2C12 cells overexpressing an empty vector (Mock) or hPADI4.

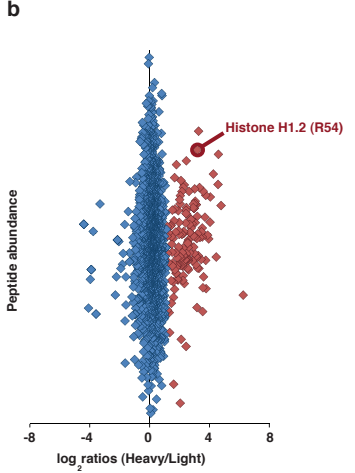
(f) ChIP-qPCR for H1.2 on the regulatory regions of *Tcl1* and *Nanog* in mES cells stably expressing *Padi4* or Ctrl shRNA. Error bars: standard error of the mean of three technical qPCR replicates. Representative of two experiments.

(g) Proposed model for the role of PADI4 in the regulation of pluripotency.


a
Padi4 mRNA
Nanog mRNA
Pou5f1 mRNA
Olig2 mRNA (x10^3)
ES NS iPS
b
Padi4 mRNA
Nanog mRNA
Pax6 mRNA
ES 2iLIF
c
mRNA levels relative to UbC
Padi4
Nanog
Time (days)
NS media
ES media
2i LIF
Day 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
H3Cit
H3
d
hPADI4 over-expression
Gpx2
Mylpf
Twf2
Foxp1
Klf2
Akr1b7
Odz4
Gsta4
Heatr1
AF067061
Tcl1
EG435337
Stat3
Zfp292
Fbxo15
AK019348
V1rd4
Myst4
Kcnk5
Padi4
Tmem131
Kit
Rc3h2
Slc4a7
Nanog
Tpm4
Triobp
Arhgap8
Cldn6
Gbp1
Gbp2
Dab2
2310007B03Rik
2610042L04Rik
Ufsp2
Pcbp4
AK032347
Ppm1a
Ym31l1
Pak2
Gtl2
Cgn
AK084071
Ctrl 6h
Ctrl 48h
PADI4 6h
PADI4 48h
e
Gene Ontology hPADI4 over-expression
stem cell maintenance
stem cell development
stem cell differentiation
stem cell division
somatic stem cell maintenance
regulation of gene expression
negative regulation of endodermal cell fate specification
positive regulation of Wnt receptor signaling pathway
regulation of macromolecule metabolic process
embryo development
log10(p)
f
pPB CTRL
pPB-hPADI4
Tcl1 mRNA
Nanog mRNA
Ctrl
shRNA#1
shRNA#2
shRNA#1 + #2
g
Tcl1 mRNA
Nanog mRNA
Ctrl
Cl-amidine
h
H3Cit ChIP
enhancer
promoter
Tcl1 locus
Nanog locus
Relative enrichment over input (x10^-2)
ES NS iPS

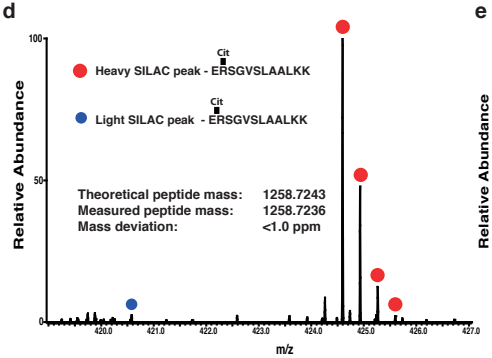

a
ES media
2iLIF media (8 days)
shRNA ctrl
shRNA ctrl
shRNA PADI4
shRNA PADI4
Oct4-GFP
shRNA ctrl
shRNA PADI4
Phase
Oct4-GFP
b
PADI4 mRNA
Tcl1 mRNA
Nanog mRNA
***
***
****
Ctrl
shRNA #4
c
Ctrl
Cl-amidine
Oct4-GFP
Ctrl
Cl-amidine
Phase
Oct4-GFP
d
E4.5 Blastocysts
Cdx2
Nanog
Sox17
Merge
Hoechst
Ctrl
Cl-amidine
e
E3.5
Blastocysts
Distribution of ICM/TE cells in E3.5 blastocyst (%)
100
50
0
Trophectoderm
Inner cell mass
CTRL
Cl-amidine
f
8 to 16
cell transition
Type of 8-16 cell divisions (%)
Symmetric
Asymmetric
g
16 to 32
cell transition
Type of 16-32 divisions (%)
Inner/Inner
Inner/Outer
Outer/Outer

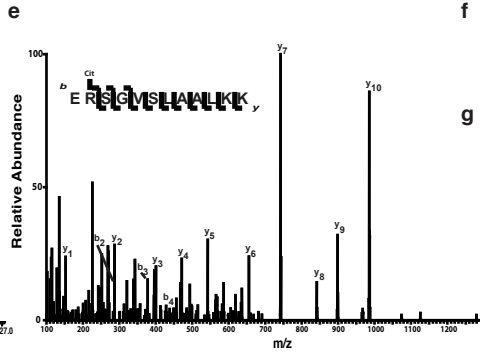


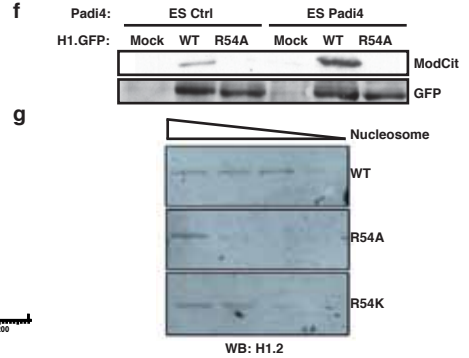


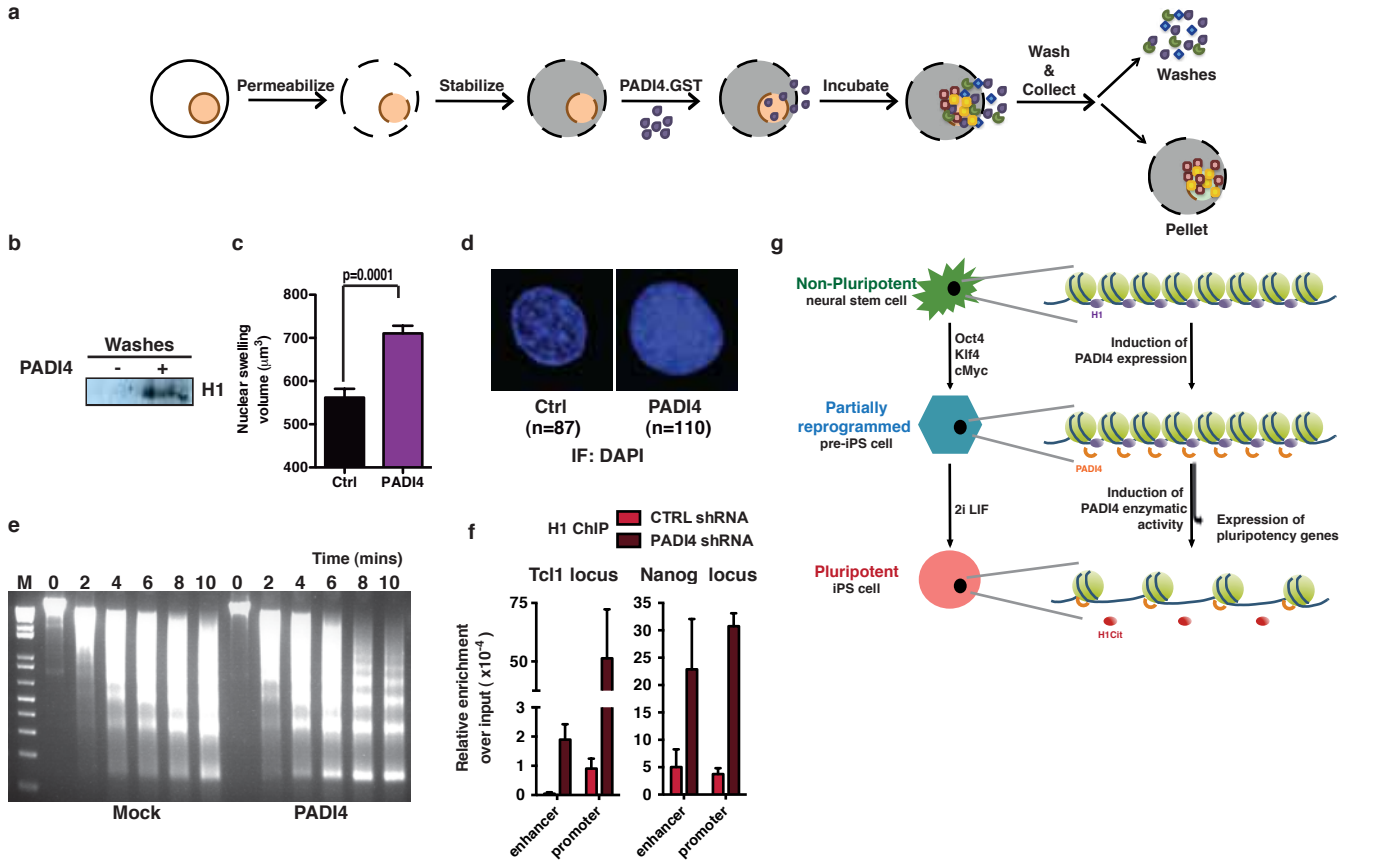
c

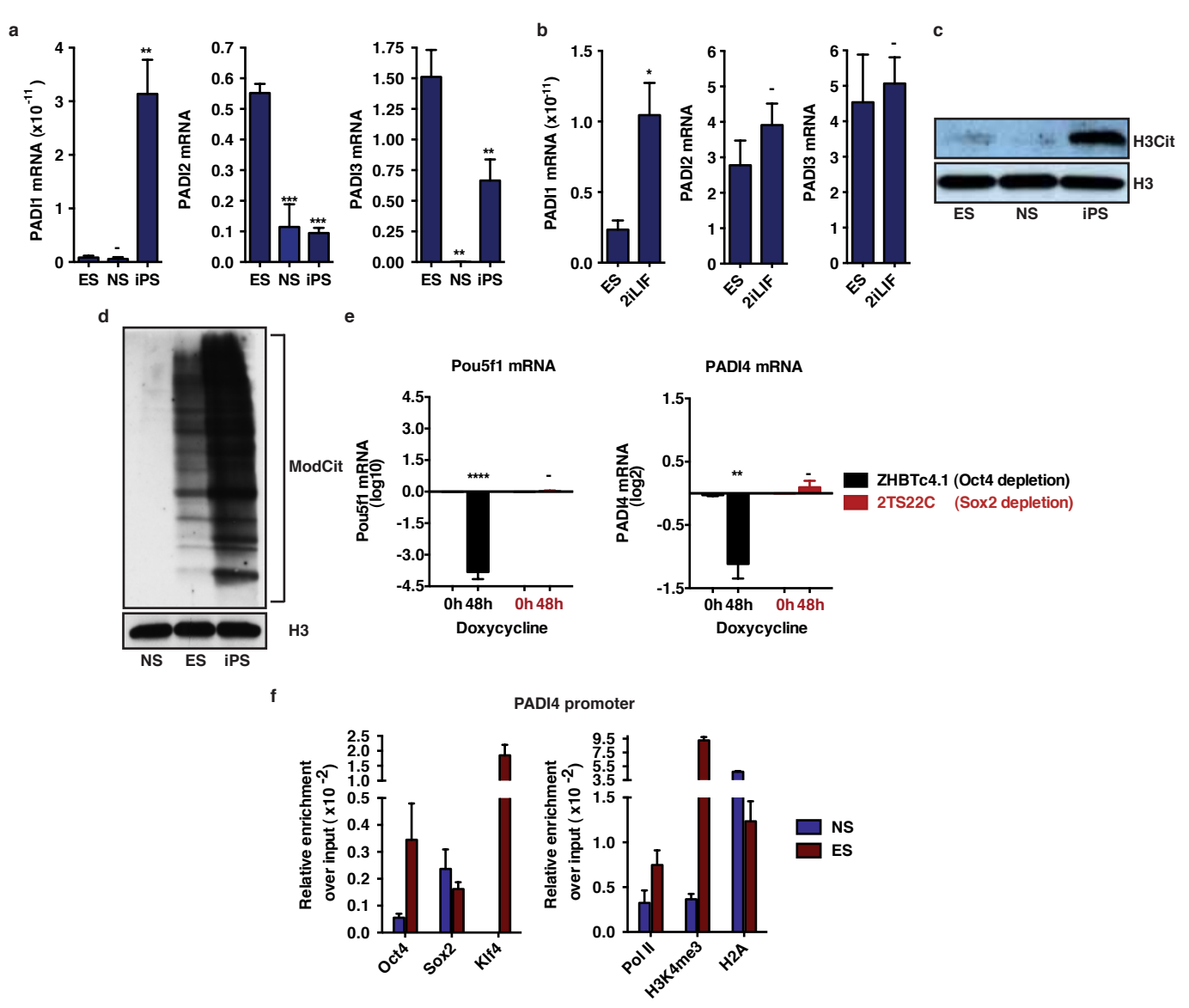
Gene name	Citrulline site	Log2 SILAC ratio	Gene name	Citrulline site	Log2 SILAC ratio
Nop2	R148	6.25	Rrp1b	R678	3.48
Sf3b1	R157	4.78	Srrm1	R7	3.45
Alyref	R140	4.60	Utp3	R375	3.37
Mybbp1a	R1322	4.57	Rrp15	R9	3.34
Rpl10	R32	4.45	Ftsj3	R390	3.30
Surf6	R106	4.26	Hnrnpu	R231	3.27
Ccdc86	R408	4.06	Ftsj3	R774	3.25
Rpl19	R38	4.04	Rps19bp1	R7	3.22
Psip1	R515	3.95	Rab11b	R4	3.19
Rpl4	R300	3.88	Hist1h1c	R54	3.19
Atrx	R1063	3.77	Rps11	R22	3.19
Mki67ip	R203	3.75	Phf16	R715	3.17
Utp14a	R431	3.75	Rpl19	R16	3.16
Rrs1	R273	3.70	Bclaf1	R802	3.15
Rpl13a	R59	3.60	Trim28	R470	3.12
Cdca8	R91	3.58	Trim28	R472	3.12
Utp14a	R586	3.56	Rpl23a	R41	3.10
Nop2	R86	3.55	Rpl19	R5	3.10
Rpl13a	R140	3.52	Tmpo	R319	3.05
Dnmt3b	R415	3.52	Matr3	R588	3.05



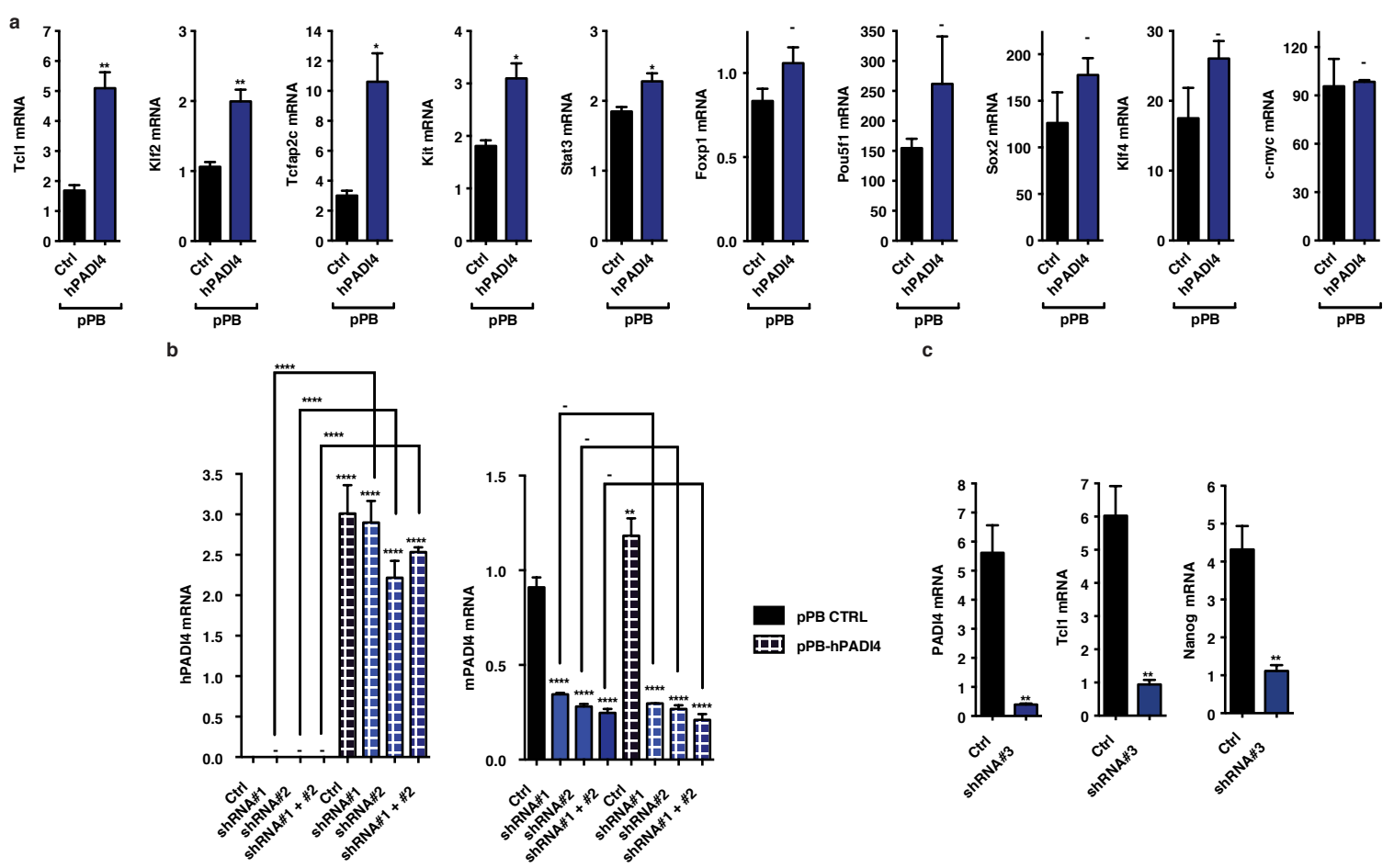




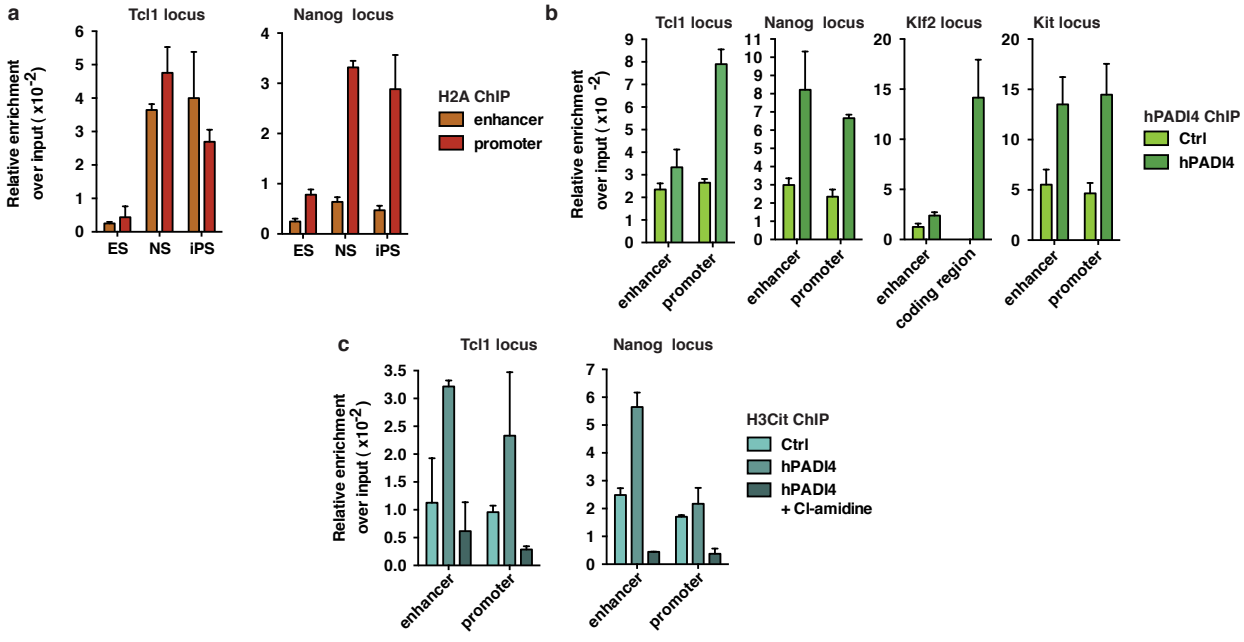

a
Permeabilize
Stabilize
PADI4.GST
Incubate
Wash
&
Collect
Washes
Pellet
b
Washes
PADI4
-
+
H1
c
p=0.0001
Nuclear swelling volume (μm3)
800
700
600
500
400
Ctrl
PADI4
d
Ctrl
(n=87)
PADI4
(n=110)
IF: DAPI
e
Time (mins)
M 0 2 4 6 8 10 0 2 4 6 8 10
Mock
PADI4
f
H1 ChIP
CTRL shRNA
PADI4 shRNA
Tcl1 locus
Nanog locus
Relative enrichment over input (x10-4)
75
50
3
2
1
0
35
30
25
20
15
10
5
0
enhancer
promoter
enhancer
promoter
g
Non-Pluripotent
neural stem cell
H1
Oct4
Klf4
cMyc
Induction of
PADI4 expression
Partially
reprogrammed
pre-iPS cell
PADI4
2i LIF
Induction of
PADI4 enzymatic
activity
Expression of
pluripotency genes
Pluripotent
iPS cell
H1Cit

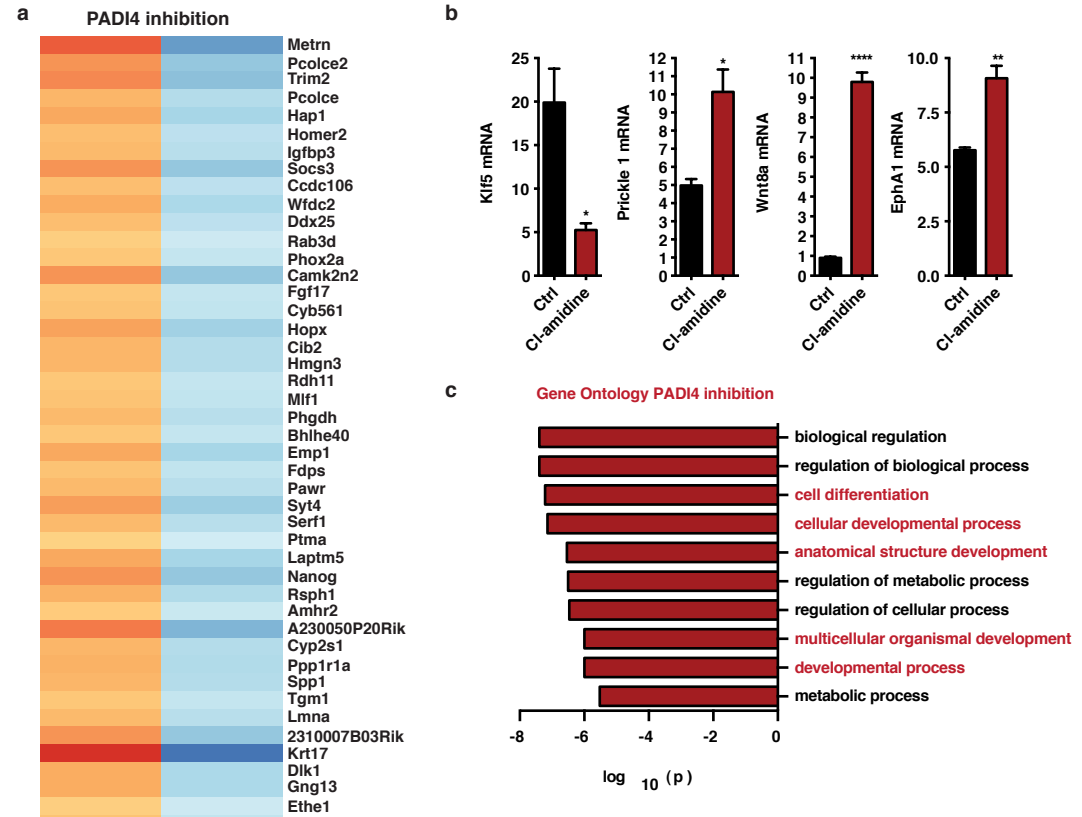


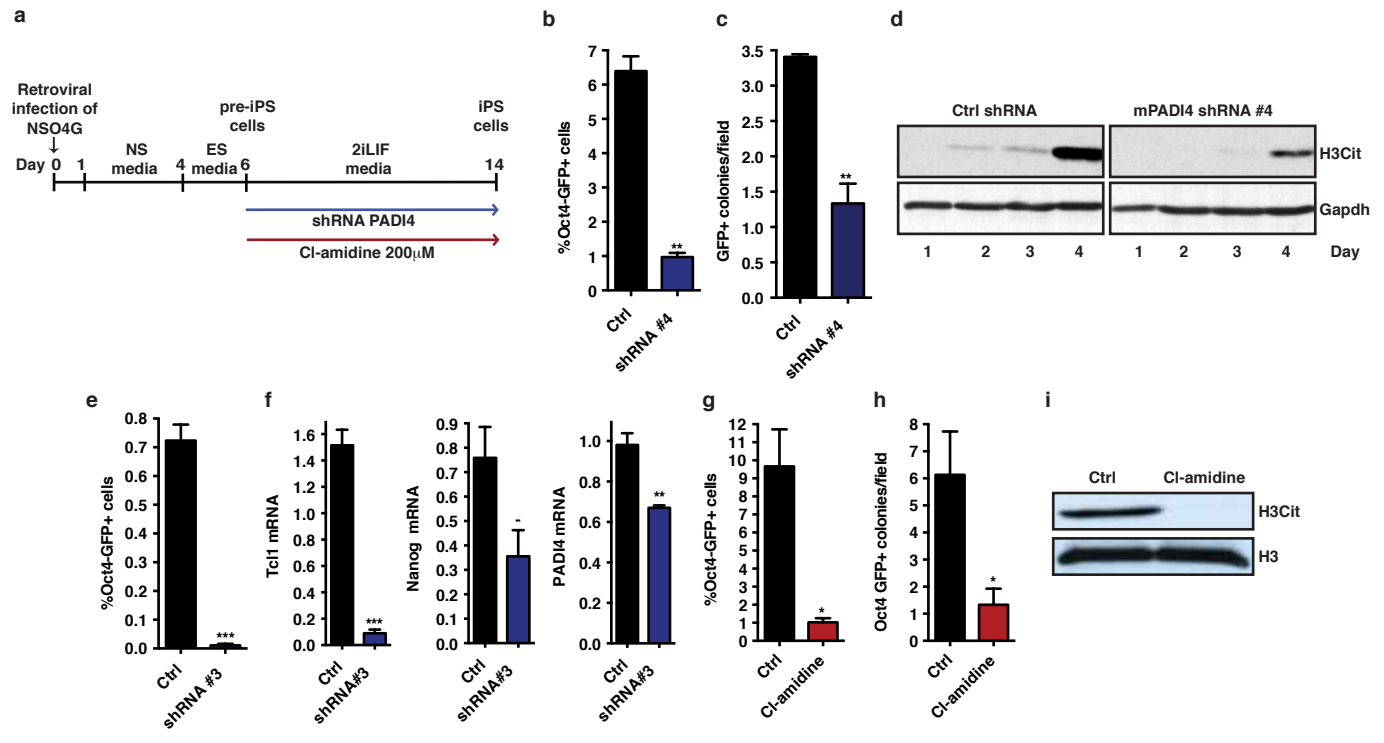
Kouzarides_Extended Data Figure 1



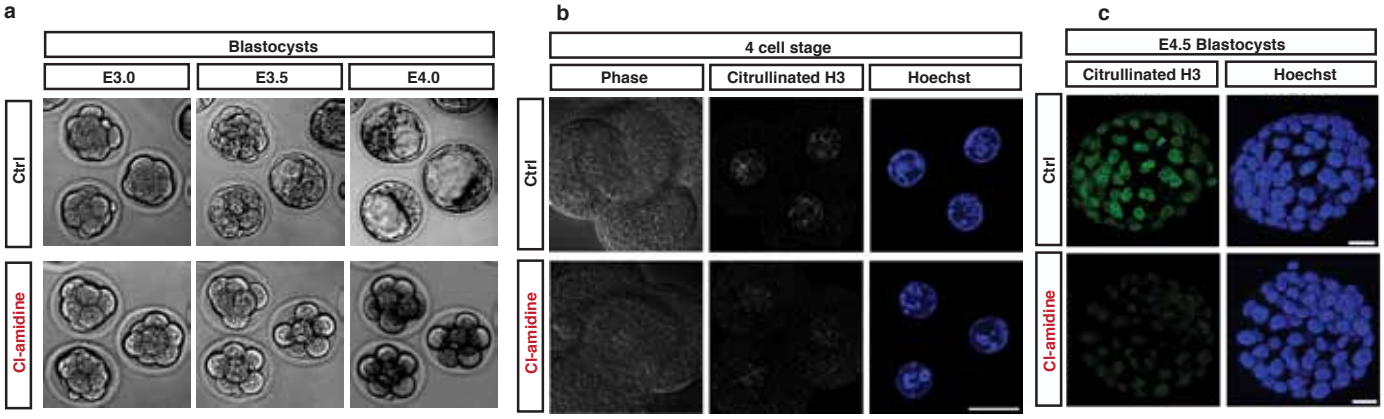
Kouzarides_Extended Data Figure 2


a
Tcl1 locus
Nanog locus
Relative enrichment over input (x10-2)
ES
NS
iPS
H2A ChIP
enhancer
promoter
b
Tcl1 locus
Nanog locus
Klf2 locus
Kit locus
Relative enrichment over input (x10 -2)
enhancer
promoter
coding region
hPADI4 ChIP
Ctrl
hPADI4
c
Tcl1 locus
Nanog locus
Relative enrichment over input (x10-2)
H3Cit ChIP
Ctrl
hPADI4
hPADI4
+ Cl-amidine


a
PADI4 inhibition
Metrn
Pcolce2
Trim2
Pcolce
Hap1
Homer2
Igfbp3
Socs3
Ccdc106
Wfdc2
Ddx25
Rab3d
Phox2a
Camk2n2
Fgf17
Cyb561
Hopx
Cib2
Hmgn3
Rdh11
Mlf1
Phgdh
Bhlhe40
Emp1
Fdps
Pawr
Syt4
Serf1
Ptma
Laptm5
Nanog
Rsph1
Amhr2
A230050P20Rik
Cyp2s1
Ppp1r1a
Spp1
Tgm1
Lmna
2310007B03Rik
Krt17
Dlk1
Gng13
Ethe1
Tcfap2a
Klf5
Sgk
Fbxl16
Calca
Krt42
Pde1b
Tmem107
Gna11
Gbp3
Raet1b
Dusp4
Hbp1
Svop
Prickle1
Egr1
Sall3
Pmaip1
Wnt8a
Epha1
2310061J03Rik
Armcx1
Lpar4
P2ry5
Gse1
Clk1
Ctrl
Cl-amidine
-1
-0.5
0
0.5
1
Value
b
Klf5 mRNA
Prickle 1 mRNA
Wnt8a mRNA
EphA1 mRNA
Ctrl
Cl-amidine
c
Gene Ontology PADI4 inhibition
biological regulation
regulation of biological process
cell differentiation
cellular developmental process
anatomical structure development
regulation of metabolic process
regulation of cellular process
multicellular organismal development
developmental process
metabolic process
log 10 (p)



Kouzarides_Extended Data Figure 5



d

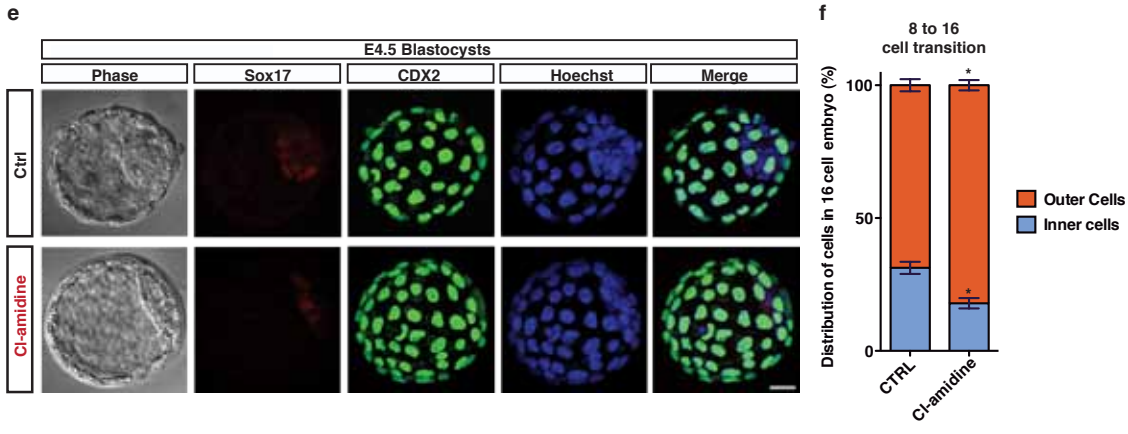
Lineage commitment in E4.5 blastocysts developed in the presence of Cl-amidine

	Number of EPI cells (mean±SE)	Percentage of EPI cells (mean±SE)	Number of TE cells (mean±SE)	Percentage of TE cells (mean±SE)	Number of PE cells (mean±SE)	Percentage of PE cells (mean±SE)	Total (mean±SE)
Control	10,00±0,73	12,37±0,91	60,15±3,23	77,10±1,43	6,61±0,76	7,60±0,78	77,84±3,94
Cl-amidine	10,26±0,63	9,48±0,40*	82,29±4,30*	80,79±0,87*	8±0,43	7,59±0,44	101,33±4,69*

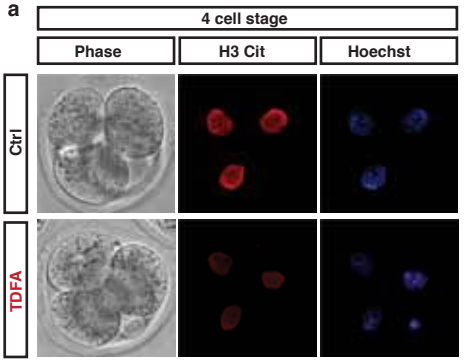
Data from 3 replicates and 50 E4.5 blastocysts (n = 24-26 per group)

*Asterisks indicate difference within the same column (Unpaired T-test, P<0.05)

EPI: epiblast, TE: trophectoderm, PE: primitive endoderm



a



b Lineage commitment in E4.5 blastocysts developed in the presence of TDFA

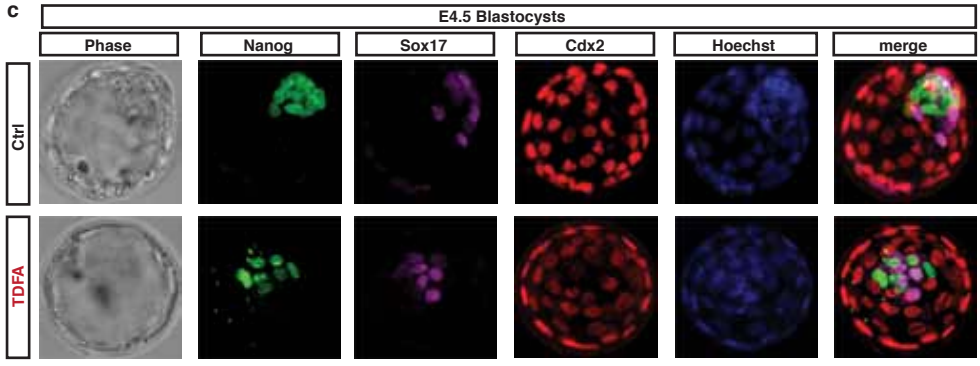
	Percentage of EPI cells (mean±SE)	Percentage of TE cells (mean±SE)	Percentage of PE cells (mean±SE)
Control	12,60±0,79	77,02±0,94	10,38±0,60
TDFA	10,43±0,62*	78,70±0,68	10,87±0,62

Data from 3 replicates and 60 E4.5 embryos (n=28-32 per group)

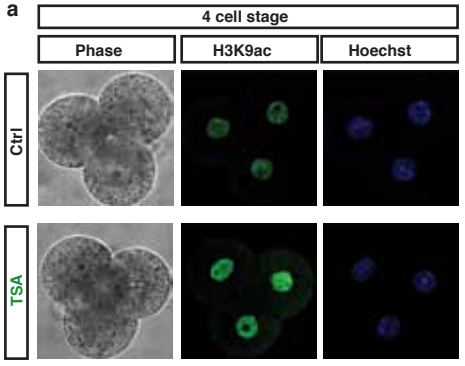
* Asterisks indicate difference within the same column (Mann Whitney test, P<0.05)

EPI: epiblast, TE: trophoectoderm , PE: primitive endoderm

c



a



b

Lineage commitment in E4.5 blastocysts developed in the presence of TSA

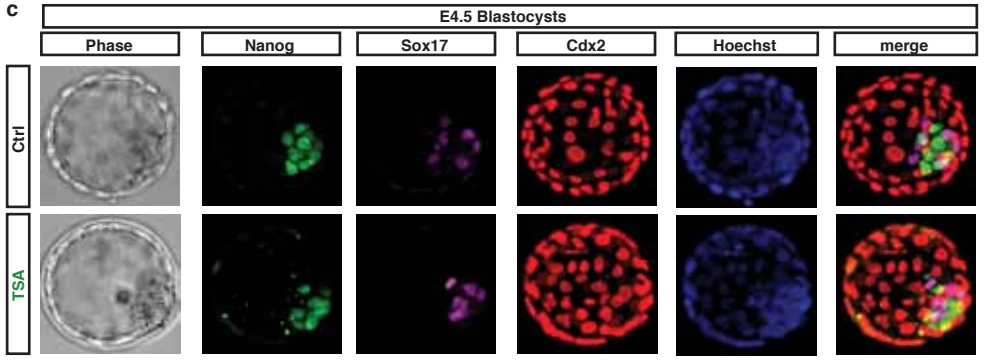
	Percentage of EPI cells (mean±SE)	Percentage of TE cells (mean±SE)	Percentage of PE cells (mean±SE)
Control	10,06±0,67	77,29±0,86	12,64±0,61
TSA	10,71±0,80	79,34±0,91	9,95±0,60*

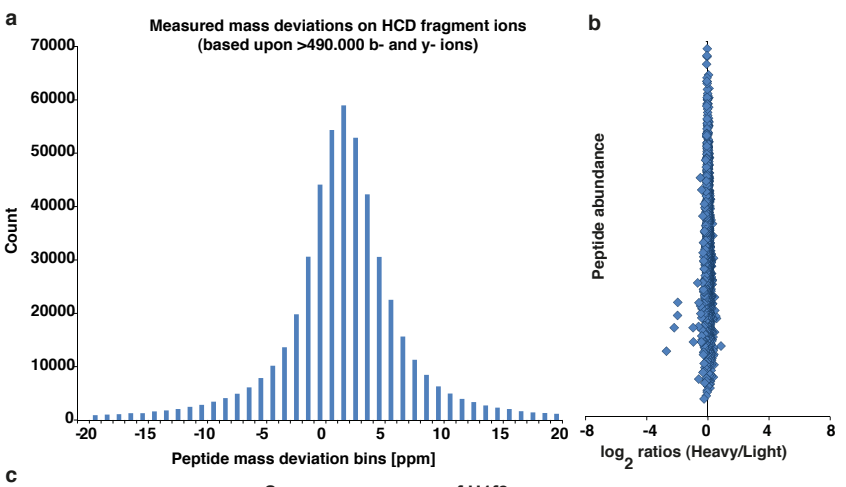
Data from 2 replicates and 32 E4.5 embryos (n=15-17 per group)

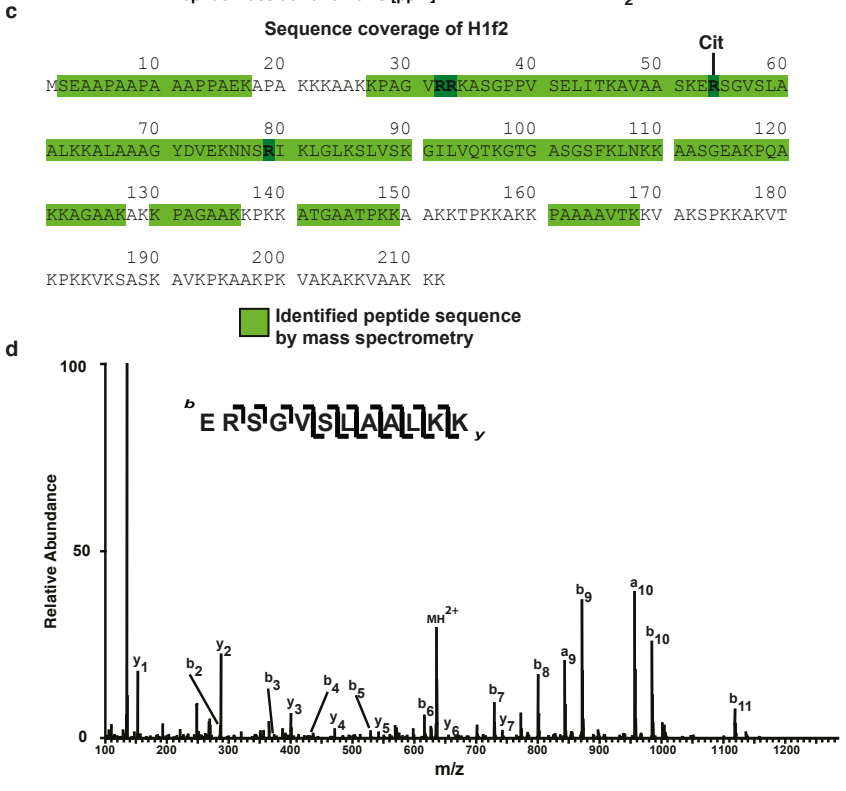
* Asterisks indicate difference within the same column (unpaired T-test, P<0.05)

EPI: epiblast, TE: trophoectoderm , PE: primitive endoderm

c







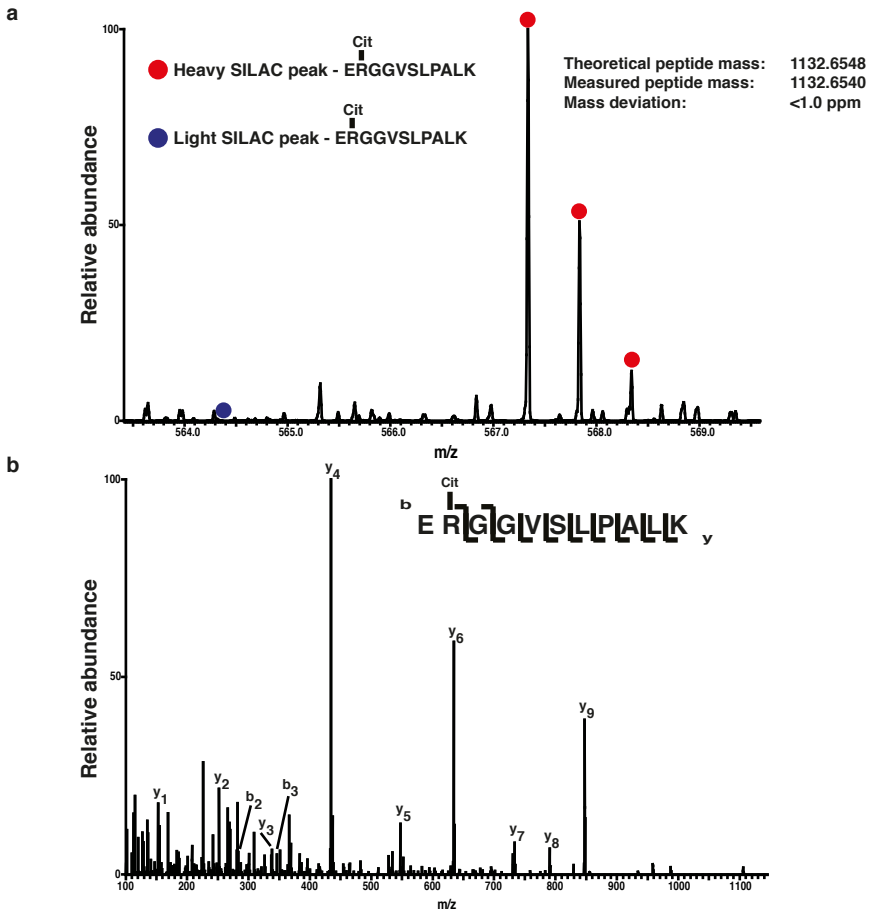
e

b-ion	Theo.Mass	Mass	PPM
b2	287.1355	287.1343	4.18
b3	374.1676	374.1649	7.22
b4	431.1890	431.1849	9.51
y-ion	**Theo.Mass**	**Mass**	**PPM**
y1	153.1335	153.1325	6.53
y2	287.2486	287.2474	4.18
y3	400.3326	400.3315	2.75
y4	471.3698	471.3683	3.18
y5	542.4069	542.4055	2.58
y6	655.4909	655.4880	4.42
y7	742.5230	742.5204	3.50
y8	841.5914	841.5886	3.33
y9	898.6128	898.6099	3.23
y10	985.6449	985.6422	2.74

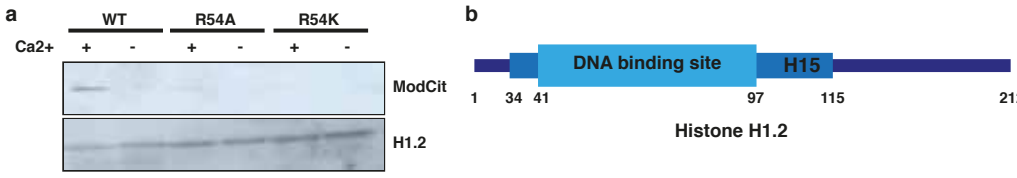
f

b-ion	Theo.Mass	Mass	PPM
b2	286.1515	286.1502	4.54
b3	373.1836	373.1808	7.50
b4	430.2050	430.2068	4.18
b5	529.2734	529.2703	5.86
b6	616.3055	616.3045	1.62
b7	729.3895	729.3867	3.84
b8	800.4266	800.4236	3.75
b9	871.4638	871.4616	2.52
b10	984.5478	984.5458	2.03
b11	1118.6629	1118.6654	2.23
y-ion	**Theo.Mass**	**Mass**	
y1	153.1335	153.1326	5.88
y2	287.2486	287.2474	4.18
y3	400.3326	400.3318	2.00
y4	471.3698	471.3679	4.03
y5	542.4069	542.4033	6.64
y6	655.4909	655.4884	3.81
y7	742.5230	742.5139	12.26

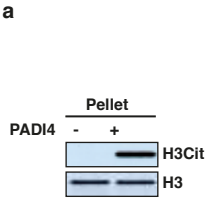
Kouzarides_Extended Data Figure 9

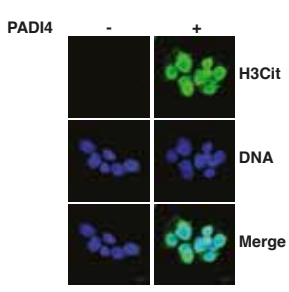


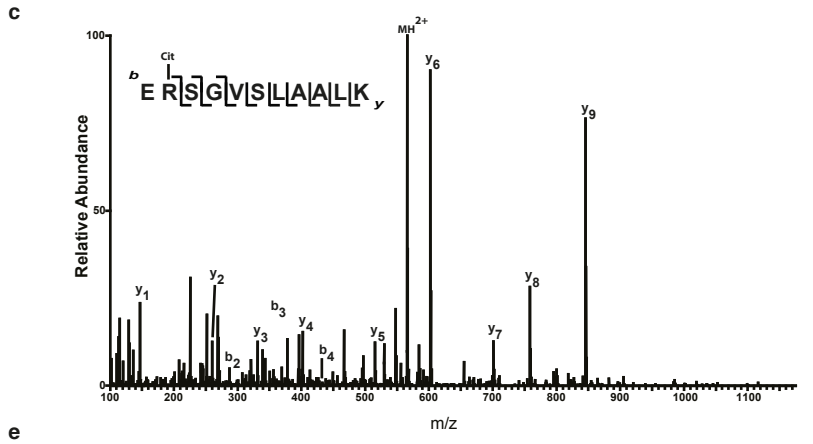
Kouzarides_Extended Data Figure 10



Kouzarides_Extended Data Figure 11

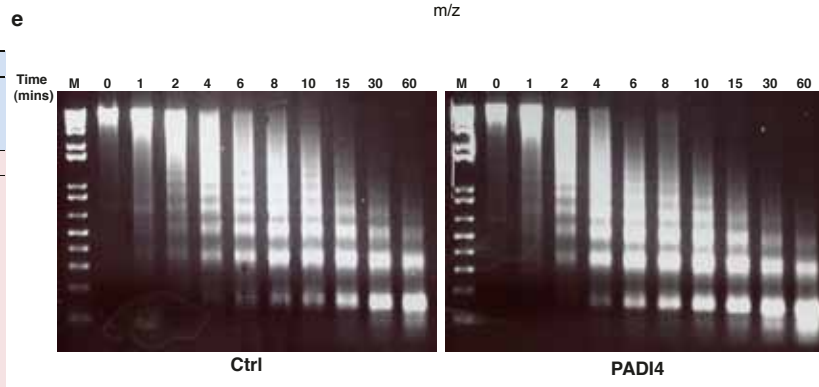






d

b-ion	Theo.Mass	Mass	PPM
b2	287.1355	287.1351	1.39
b3	374.1676	374.1667	2.41
b4	431.1890	431.1902	2.78
y-ion	**Theo.Mass**	**Mass**	**PPM**
y1	147.1134	147.1128	4.08
y2	260.1974	260.1969	1.92
y3	331.2345	331.2339	1.81
y4	402.2716	402.2711	1.24
y5	515.3557	515.3554	0.58
y6	602.3877	602.3871	1.00
y7	701.4562	701.4558	0.57
y8	758.4776	758.4770	0.79
y9	845.5096	845.5089	0.83



Kouzarides_Extended Data Figure 12

Extended Data Figure Legends

Extended Data Figure 1:

(a) Transcript levels for *Padi1, Padi2 and Padi3* in ES, NS and iPS cells, as assessed by qRT-PCR. *Padi6* was undetectable in all three cell types. Expression normalized to endogenous levels of *Ubiquitin (UbC)*. Error bars represent the standard error of the mean of three biological replicates.

(b) Transcript levels of *Padi1, Padi2 and Padi3* in ES cells upon switch to 2i containing medium for one passage, as assessed by qRT-PCR. *Padi6* was undetectable in both conditions. Expression normalized to *UbC*. Error bars represent the standard error of the mean of three biological replicates.

(c) Immunoblot analysis of H3Cit levels in ES, NS and iPS cells. Total H3 is presented as loading control.

(d) Immunoblot analysis of total citrullination the levels in ES, NS and iPS cells, using an antibody against Modified Citrulline (ModCit) which recognizes peptidylcitrulline irrespective of amino acid sequence. Total H3 is presented as loading control.

(e) ZHBTc4.1 and 2TS22C ES cell lines were treated with 1μg/ml doxycycline for 48 hours, resulting in depletion of Oct4 or Sox2 (data not shown). *Padi4* mRNA was significantly reduced upon Oct4, but not Sox2 knockdown, as assessed by qPCR. Error bars represent standard error of the mean of four biological replicates.

(f) ChIP-qPCR for Oct4, Sox2, Klf4, RNA polymerase II (PolII), H3K4me3 and H2A on the promoter of *Padi4* in mES and NS cells. For each cell condition, the signal is presented as fold enrichment over Input and after subtracting background signal from the beads. Error bars represent the standard deviation of three technical qPCR replicates.

Asterisks denote difference with ES cells (a) or media (b), and 0h time point (e); - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 by ANOVA (a) or t-test (b,e).

Extended Data Figure 2:

(a) Validation of selected targets from the PADI4 over-expression microarray dataset by qRT-PCR. Expression of *Pou5f1*, *Sox2*, *Klf4* and *c-Myc* is not affected by PADI4 over-expression. Expression normalised to *UbC*. Error bars presented as standard error of the mean of three biological replicates.

(b) Transcript levels of mouse *Padi4* and human *PADI4* in mES cells after transient knock-down with *Padi4* or control (Ctrl) shRNA, and over-expression of human *PADI4* or control vector (pPB CTRL), as assessed by qRT-PCR. Expression normalized to *UbC*. Error bars represent the standard error of the mean of three biological replicates.

(c) Transcript levels of mouse *Padi4, Tcl1* and *Nanog* in mES cell clones stably expressing *Padi4* or control (Ctrl) shRNA, as assessed by qRT-PCR. Expression

normalized to *UbC*. Error bars represent the standard error of the mean of three biological replicates.

Asterisks denote difference with Ctrl (a,b,c) and between samples (b); - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 by ANOVA (b) or t-test (a,c).

Extended Data Figure 3:

(a) Representative ChIP-qPCR for H2A on regulatory regions of *Tcl1* and *Nanog* in mES, NS and iPS cells (corresponding to Fig. 1h). For each cell condition, the signal is presented as fold enrichment over Input and after subtracting background signal from the beads. Error bars represent the standard deviation of three technical qPCR replicates.

(b) ChIP-qPCR for hPADI4 on regulatory regions of *Tcl1, Nanog, Klf2 and Kit*, which are up-regulated by hPADI4 over-expression, in mES cells stably expressing hPADI4. For each cell condition, the signal is presented as fold enrichment over Input and after subtracting background signal from the beads. Error bars represent the standard deviation of three technical qPCR replicates.

(c) Representative ChIP-qPCR for H3Cit on regulatory regions of *Tcl1* and *Nanog* in mES cells stably expressing hPADI4 and treated with 200μM Cl-amidine for 48h. For each cell condition, the signal is presented as fold enrichment over Input and after subtracting background signal from the beads. Error bars represent the standard deviation of three technical qPCR replicates.

Extended Data Figure 4:

(a) Heat map of the top 70 genes that showed differential expression after PADI4 inhibition in ES cells by with 200μM Cl-amidine for 48h, as determined by microarray analysis. Displayed values are normalized log intensities, minus the mean expression of the gene across the two samples. Hierarchical clustering based on correlation.

(b) Validation of selected targets from the above microarray dataset by qRT-PCR. Expression normalised to *UbC*. Error bars presented as standard error of the mean of three biological replicates. Asterisks denote difference with Ctrl; - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 by t-test.

(c) Gene Ontology for Biological Process (GOBP) analysis for the most regulated gene categories within the microarray dataset of Cl-amidine treatment in mES cells. *P*-value is corrected for multiple testing using Benjamini and Hochberg FDR.

Extended Data Figure 5:

(a) Scheme of reprogramming of neural stem cells to pluripotent state. NSO4G cells were retrovirally transduced with Oct4, Klf4 and c-Myc. After 6 days, partially reprogrammed pre-iPS cells arose. For shRNA experiments, pre-iPS cells were stably transfected with Ctrl or PADI4 shRNA and then full reprogramming was performed in the presence of 2iLIF media for 8 days. For PADI4 enzymatic inhibition,

pre-iPS cells were immediately changed to 2iLIF media in the presence of the inhibitor Cl-amidine for 8 days.

(b) Quantification of flow cytometry analysis for the assessment of Oct4-GFP reporter expression in a reprogramming assay using pre-iPS cells stably expressing *Padi4* shRNA #4 and Ctrl shRNA. Error bars represent standard error of the mean of triplicate samples within a representative from four reprogramming experiments.

(c) Quantification of Oct4-GFP positive colonies in the reprogramming assay where pre-iPS cells were *Padi4* shRNA #4 versus control (*see* Fig. 2a), after time-lapse image acquisition with Biostation CT. Error bars represent standard error of the mean of triplicate samples within a representative reprogramming experiment. Time-lapse video in supplementary data online.

(d) Immunoblot analysis of H3Cit in pre-iPS cells treated with 2i, after *Padi4* knock-down (PADI4 shRNA #4) versus control cells (Ctrl shRNA). 2i-containing medium was added on day 2. Total histone H3 presented as loading control.

(e) Quantification of flow cytometry analysis for the assessment of Oct4-GFP reporter expression in a reprogramming assay using pre-iPS cells stably expressing *Padi4* shRNA #3 and Ctrl shRNA. Error bars represent standard error of the mean of triplicate samples.

(f) qRT-PCR analysis for the expression of *Tcl, Nanog* and *Padi4* mRNAs at the end of the above reprogramming assay (e). Error bars represent standard error of the mean of triplicate samples.

(g) Quantification of flow cytometry analysis for the assessment of Oct4-GFP reporter expression in a reprogramming assay were treated with 200μM Cl-amidine. Error bars represent standard error of the mean of triplicate samples within a representative from three reprogramming experiment.

(h) Quantification of Oct4-GFP positive colonies in the reprogramming assay where pre-iPS cells were treated with 200μM Cl-amidine (*see* Fig. 2c) after time-lapse image acquisition with Biostation CT. Error bars represent standard error of the mean of triplicate samples within a representative reprogramming experiment.

(i) Immunoblot analysis for the presence of H3Cit at the end of the above reprogramming assay (g). Total histone H3 presented as loading control.

Asterisks denote difference with Control; - not significant, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 by t-test

Extended Data Figure 6:

(a) Embryos at 2-cell stage were treated with 200μM Cl-amidine and snapshots were taken at E3.0, E3.5 and E4.0. 200μM Cl-amidine embryos arrested at 8-cell stage, while Control embryos continued development to form blastocysts. Phase contrast images are shown.

(b) Embryos at 2-cell stage were treated with 10μM Cl-amidine for 12 hours, fixed and stained for H3Cit at the 4-cell stage. Phase contrast, H3Cit (white) and HOECHST 33342 (blue) images are shown. Bar represents 20μm.

(c) Embryos at E3.5 were treated with 10μM Cl-amidine for 24 hours, fixed and stained for H3Cit at E4.5. H3Cit (green) and HOECHST 33342 (blue) images are shown. Bar represents 20μm.

(d) Table with quantifications of lineage commitment in E4.5 blastocysts treated with 10μM Cl-amidine from the 2-cell stage. Asterisks denote difference with Control, unpaired t-test, * = $p<0.05$. n=3 (50 embryos).

(e) Embryos were cultured in medium supplemented with 10μM Cl-amidine from 2-cell stage and through preimplantation development. E4.5 blastocysts were fixed and stained for SOX17 (primitive endoderm marker, red), Cdx2 (trophectoderm marker, green) and HOECHST 33342 (blue). Bar represents 20μm.

(f) Time-lapse analysis of distribution of inner and outer cells at the 8 to 16-cell transition, upon culturing of embryos with medium containing 10μM Cl-amidine from 2-cell stage. Error bars represent standard error of the mean. Statistical significance was determined by unpaired t-test or Mann Whitney test upon non-normal distribution. Asterisks denote difference with Control; * $P \leq 0.05$.

Extended Data Figure 7:

(a) Embryos at 2-cell stage were treated with 100μM TDFA for 12 hours and fixed and stained for H3Cit at 4-cell stage. H3Cit and HOECHST 33342 images are shown.

(b) Table representing the percentage of cells committed to each embryonic lineage in E4.5 blastocysts upon treatment of embryos at 2-cell stage with 100μM TDFA. Bars represent mean percentage (±SEM). Asterisks denote difference with Control, Mann-Whitney test, * = $p<0.05$. n=3 (60 embryos).

(c) Embryos at 2-cell stage were treated with 100μM TDFA and fixed at embryonic day E4.5. Phase contrast, Nanog (green), Sox17 (purple), Cdx2 (red) and HOECHST 33342 (blue) images are shown.

Extended Data Figure 8:

(a) Embryos at 2-cell stage were treated with 10nM TSA for 12 hours, and fixed and stained for H3K9ac at 4-cell stage. H3K9ac and HOECHST 33342 images are shown.

(b) Table representing the percentage of cells committed to each embryonic lineage in E4.5 blastocysts upon treatment of embryos at 2-cell stage with 10nM TSA. Bars represent mean percentage (±SEM). Asterisks denote difference with Control, unpaired t-test, * = $p<0.05$. n=2 (32 embryos).

(c) Embryos at 2-cell stage were treated with 10nM TSA and fixed at embryonic day E4.5. Phase contrast, Nanog (green), Sox17 (purple), Cdx2 (red) and HOECHST 33342 (blue) images are shown.

Extended Data Figure 9:

(a) Histogram demonstrating the mass accuracies of all fragment ion masses used for identifying citrullinated peptides in our HCD MS/MS spectra. >490.000 y- and b-ion masses are depicted. The average absolute mass accuracy for all of these fragment ions is 3.97 ppm.

(b) Scatter plot representing SILAC ratios in ES cells cultured in $^{13}C_6$ L-Lysine (HEAVY) and LIGHT medium separately, to assess extend and quality of SILAC labeling. No significant outliers are observed, indicating equal labeling.

(c) Peptide coverage of histone H1 by LC-MS analysis. Detected peptides are highlighted in light green and cover >60% of H1. While all arginine residues of Histone H1 (highlighted in dark green) were accounted for by the analysis, Arg54 was the only one found citrullinated.

(d) Fragmentation spectra of the unmodified LysC peptide ERSGVSLAALKK surrounding Arginine 54 of H1.2 (unmodified counterpart of citrullinated peptide depicted in Fig. 3d). The y and b series indicate fragments at amide bonds of the peptide.

(e) Fragment ion table (expected and observed masses for detected y and b ions) for the identified H1R54 citrullination of peptide ERSGVSLAALKK on histone H1.2 (as shown in Fig. 3d). All measured fragment ions were detected with mass accuracies <10ppm, unambiguously identifying that the detected peptide sequence harbors a citrullination at position R54.

(f) Theoretical and measured b- and y-ion fragment masses for the corresponding unmodified and heavy SILAC labeled H1.2 peptide, as presented in (d) above.

Extended Data Figure 10:

(a) MS spectrum of Histone H1.5 in a SILAC proteomic screen for identification of PADI4 substrates. Linker histone H1.5 is deiminated by PADI4, as identified by a highly increased SILAC ratio of the heavy labeled identified peptide (marked by a red dot).

(b) Fragmentation spectra of the doubly charged LysC peptide ERGGVSLPALK surrounding Arginine 54 of H1.5. The y and b series indicate fragments at amide bonds of the peptide, unambiguously verifying the citrullinated peptide.

Extended Data Figure 11:

(a) Mutation of R54 renders histone H1.2 refractory to deimination. Immunoblot analysis of recombinant histone H1.2 using an antibody that detects all deimination events (ModCit). Wild type and R54-mutant H1.2 were treated with recombinant PADI4, in the presence of activating calcium. Only wild-type H1.2 can be deiminated, indicating that R54 is the only substrate of PADI4 in H1.2. No-calcium reactions presented as negative controls. Total H1.2 presented as loading control.

(b) Schematic representation of the position of R54 within the globular domain linker histone H1.2.

Extended Data Figure 12:

(a) Immunoblot analysis of the “Pellet” fraction of C2C12 permeabilised cells treated with recombinant PADI4. Presence of H3Cit species indicates PADI4 activity. Total H3 is presented as a control for equal use of starting material in the two experimental conditions.

(b) Immunofluorescence analysis of C2C12 nuclei after treatment with recombinant PADI4. Presence of H3Cit species indicates PADI4 activity. DNA is visualised by staining with DAPI.

(c) Fragmentation spectra of the citrullination site R54 on the evicted H1.2 peptide ERSGVSLAALK (corresponding to Fig. 4b). The evicted Histone H1 is citrulinated at R54.

(d) Theoretical and measured b- and y-ion fragment masses for the citrullinated H1.2 peptide (peptide sequence ERSGVSLAALK) evicted after treatment of C2C12 cells with recombinant hPADI4 (corresponding to Fig. 4b).

(d) Micrococcal nuclease digestion of C2C12 nuclei after treatment with recombinant PADI4, as described in Fig. 4a. M= size marker.

Supplementary Methods

Cell Culture

NSO4G neural stem cells[1] were cultured in RHB-A media (Stem Cell Sciences, Cambridge, UK), supplemented with penicillin/streptomycin (Life Technologies) and 10ng/ml bFGF and EGF (PeproTech). ES Oct4-GIP cells[2] and E14 ES cells were cultured in GMEM supplemented with 10% fetal calf serum (FCS) for ES cells (Biosera), 0.1mM non-essential amino acids, penicillin/streptomycin, 2mM L-glutamine, 1mM sodium pyruvate, 0.1mM beta-mercaptoethanol and 10^6 units/L leukemia inhibitory factor (LIF) (ESGRO, Millipore), or in 2i/LIF media, based on GMEM and containing 10% Knock-Out Serum Replacement (Life Technologies), 1% FCS for ES cells (Biosera), 0.1mM non-essential amino acids, penicillin/streptomycin, L-glutamine, sodium pyruvate, 0.1mM beta-mercaptoethanol, 1μM PD0325901 (AxonMedChem, Groningen, The Netherlands), 3μM CHIR99021 (AxonMedChem, Groningen, The Netherlands) and 10^6 units/L LIF (ESGRO, Millipore). Plat-E packaging cells were grown in DMEM media (Life Technologies) supplemented with 10% FCS, 1μg/ml puromycin, 10μg/ml blasticidin and penicillin/streptomycin. iPS cells were maintained in 2i/LIF media. 1μg/ml puromycin was added to iPS and ES Oct4-GIP cultures during expansion. ZHBTc4.1[3] and 2TS22C[4] ES cell lines were expanded in ES cell media and treated with 1μg/ml doxycycline for 48h prior to RNA extraction and qRT-PCR analysis. Complete knock-down of Oct4 in the ZHBTc4.1 cell line and of Sox2 in the 2TS22C ES cell line was confirmed by western blot (data not shown). All cells were grown at 37°C with 5% (Biostation) or 7.5% CO_2.

Over-expression of human PADI4 or shRNA against mouse Padi4 in mES cells

Human *PADI4* was inserted in ES E14 cells using the piggyBac transposon system[5]. The Gateway system was used to clone *hPADI4* into the piggyBac vector using the following primers:

PADI4_AttB1_F:

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGCCCAGGGGACATTGA

TCCG;

PADI4_AttB2_R:

GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGGCACCATGTTCCACC.

1µg PB-CAG-Ctrl or PB-CAG-hPADI4 vectors were transfected with 2µg piggyBac transposase (PBase) expression vector, pCAGPBase, by nucleofection according to the manufacturers' instructions (Lonza). ES E14 cells constitutively expressing the hygromycin resistance gene and *hPADI4* were selected and expanded in media containing 200µg/ml hygromycin.

For mouse *Padi4* knock-down experiments, ES E14 cells were transfected with Lipofectamine 2000 (Life Technologies) or by nucleofection with pRFP-C-RS HuSH shRNA RFP vectors (Origene) containing either the scrambled shRNA cassette TR30015 (Ctrl), mouse *Padi4* targeting shRNAs FI516326 (shRNA#1) or FI516328 (shRNA#2), or with Mission RNAi pLKO.1-puro vectors (Sigma), containing either the non-targeting shRNA SHC002 (Ctrl) or mouse *Padi4* targeting shRNA TRCN000101833 (shRNA#3). Where applicable ES cell lines were generated after selection with 1µg/ml puromycin.

Reprogramming

Reprogramming was performed as described in Silva et al., 2009 and Theunissen et al, 2011[6,7]. For retroviral supernatant preparation, 9 µg of pMXs-Oct4, pMX-Klf4 and pMXs-c-Myc were transfected FuGENE 6 into 1-2x10^6 Plat-E cells in separate 10cm dishes. After 24h incubation, the media was replaced with DMEM+10% FCS and penicillin/streptomycin. Virus-containing supernatants from Plat-E cultures were filtered through a 0.22-µm cellulose acetate filter, mixed in equal ratios, and 4µg/ml polybrene was added. 2 mls of the final viral mix was then added to previously plated

$1.2x10^5$ NSO4G cells, in 6-well plates coated with gelatin. After 1 day, the media was replaced with NS cell culture medium. After 3 days incubation, the media was changed to ES cell serum-containing media (see above). At this stage, pre-iPS colony formation was evident, but none of the colonies was positive for GFP expression (GFP+). After 2 days and to complete reprogramming, medium was replaced with 2i/LIF. Cells were maintained in 2i/LIF for 8 days, with media change every 2 days. Oct4-GFP+ colonies were counted at day 7 in 5-9 selected fields per well, either at the microscope or after time-lapse image acquisition on the Biostation CT (Nikon, Japan). The percentage of GFP+ cells was determined by flow cytometry at day 8, using a Dako Cytomation CyAN ADP high-performance cytometer and Summit software, as described in Silva et al., 2009 and Theunissen et al, 2011[6,7]. Statistical significance was determined by two-tailed unpaired t-test. For the time course experiments, cells were collected with 350µl RLT buffer (Qiagen) or 2X Laemmli buffer, from individual wells in consecutive days and after a PBS wash.

200µM Cl-amidine was added at the time of media exchange to 2i/LIF and replenished every 2 days. For *Padi4* knock-down experiments, pre-iPS cells were maintained in ES cell serum-containing media and cell lines were generated after transfection with lipofectamine 2000 with the Mission RNAi pLKO.1-puro vectors, containing either a non targeting shRNA (SHC002) or mouse *Padi4* targeting shRNAs TRCN000101833 (shRNA#3) or TRCN000101834 (shRNA#4). Following puromycin selection, control and *Padi4* knock-down pre-iPS cell lines were generated. Importantly, puromycin treatment abolished all non-transfected pre-iPS cells and Ctrl and *shPadi4* pre-iPS cell lines were Oct4-GFP negative (Fig. 2A). For reprogramming experiments, $1x10^5$ cells were plated in individual wells (6-well plates) in triplicate and without puromycin, and media was changed to 2i/LIF after day 1. Cells were maintained in 2i/LIF for 8 days, with media change every 2 days, and assessed for Oct4-GFP as previously described.

qRT-PCR

Primer design Genbank and Ensembl cDNA sequences were used to design gene specific primers in Primer 3[8] or in the Universal ProbeLibrary Assay Design Center (Roche Applied Science). The specificity of PCR primers was determined via the in-Silico PCR (UCSC Genome Browser) and Primer-BLAST (NCBI) web-based tools. Oligonucleotides were obtained from Sigma. Primer sequences can be found in Supplementary Table 1.

Reverse Transcription Total RNA was isolated from ES cells, NSO4G and cells during reprogramming, using the RNeasy extraction kit with in-column DNAse treatment (Qiagen). Total RNA was reverse transcribed with the High-Capacity cDNA Reverse Transcription Kit for 1 hour (Applied Biosystems). Samples were aliquoted equally into positive and negative (RT-) reactions. Prior to qPCR analysis, samples were diluted 5- or 10-fold with DNAse/RNAse free dH2O (Ambion).

qPCR qPCR reactions were performed in duplicate or triplicate for each sample. Each PCR reaction had a final volume of 10-20μl and 2.5-5μl of diluted cDNA or ChIP DNA. RT- samples were assayed to discount genomic DNA amplification. Fast SYBR Green Master Mix or TaqMan Fast Universal PCR Master Mix (Applied Biosystems) were used, according to the manufacture's instructions. A melting curve was obtained for each PCR product after each run, in order to confirm that the SYBR Green signal corresponded to a unique and specific amplicon. Random PCR products were also run in a 2-3% agarose gel to verify the size of the amplicon. Standard curves were generated for each real-time PCR run using serial 3-fold dilutions of a sample containing the sequence of interest. Their plots were used to convert CTs (number of PCR cycles needed for a given template to be amplified to an established fluorescence threshold) into arbitrary quantities of initial template per sample. Expression levels were then obtained by dividing the quantity by the value of housekeeping genes, such as ubiquitin (*UbC*). *UbC* assays were run every time samples were frozen/thawed. Statistical analysis was performed in Prism 6 using

one-way ANOVA analysis of variance with Holm-Sidak's multiple comparisons test or two-tailed unpaired t-test.

Gene expression analysis

Mouse WG-6 Expression BeadChip microarrays (Illumina) were processed at the Cambridge Genomic Services, Department of Pathology, University of Cambridge. Three biological replicates were assayed for each condition. Illumina microarray probes were matched to gene identifiers according to the re-annotation of the microarray platform[9]. For both the Padi4 overexpression and the Cl-amidine inhibition experiments, normalization was performed using the lumi[10] R package. Limma[11] was used for differential expression analysis, with Benjamini-Hochberg (FDR) adjusted p-values <0.05 considered significant. Gene ontology (GO) enrichment analysis was performed using GOstats[12], and results adjusted for multiple testing using the Benjamini-Hochberg procedure (FDR).

Immunoblot analysis

For immunoblot analysis, cell monolayers or pellets were resuspended in 2X Laemmli buffer, incubated for 5 mins at 95°C and passed 10 times through a 21G needle to shear genomic DNA. In the case of Trichloroacetic acid precipitated proteins, pellets were resuspended in buffer and boiled as above. Proteins were separated by SDS–PAGE, transferred to nitrocellulose membrane (Millipore) using wet transfer and incubated in blocking solution (5% BSA in TBS containing 0.1% Tween) for 1h at room temperature. Membranes were incubated with primary antibody at 4^{o}C overnight and appropriate HRP-conjugated secondary antibody for 2h at room temperature. Membranes were then incubated for enhanced chemiluminescence (ECLH; GE Healthcare) and proteins were detected by exposure to X-ray film. Primary antibodies, diluted in blocking solution were used against citrullinated histone H3 (α-H3Cit, Abcam, ab5103 at 1:50,000 dilution), unmodified

histone H3 (α-H3, Abcam, ab10799 at 1:2,000), linker histone H1 (α-H1, Santa Cruz Biotechnology, sc-34464 at 1:200), GFP (α-GFP, Abcam, ab290 at 1:5000) and Gapdh (α-Gapdh, Abcam, ab9485 at 1:2500). Citrulline-containing proteins were modified on the membrane and detected using Anti-Modified Citrulline detection kit (Millipore) as per manufacturer's instructions.

Chromatin immunoprecipitation

ChIP-IT Express (Active Motif) was used according to the supplier's recommendations. Cells were cross-linked using 1% formaldehyde for 10 min at room temperature. Formaldehyde was quenched by a 5-minute incubation with glycine, cells were rinsed twice with cold PBS, collected by scraping and pelleted at 2500 rpms for 10 minutes at 4°C. Frozen pelleted cells were thawed and resuspended in lysis buffer, rotated for 30 minutes at 4°C, dounced and centrifuged at 5000 rpms for 10 mins at 4°C. Pelleted nuclei were resuspended in shearing buffer. Chromatin was then sonicated using a Bioruptor™ 200 (Diagenode), high frequency, 0.5 min/0.5min, for 10 minutes twice. Sonicated chromatin was analysed in 1% agarose gel, to confirm efficient sonication. Input was collected for further analysis. 5-15 µg of chromatin was incubated with 2µg of rabbit IgG as control (Abcam, ab6742), α-PADI4 (Sigma, P4749), α-Klf4 (R&D Systems, AF3158), α-Oct4 (Santa Cruz Biotechnology, sc-5279X (mouse) or sc-8628X (goat)), α-Sox2 (Santa Cruz Biotechnology, sc-17320), α-H3Cit (Abcam, ab5103), α-H2A (Abcam, ab18255), α-H3K4me3 (Millipore, #17-614), α-RNA polymerase II (Millipore, clone CTD4H8, #05-623) or α-H1.2 (a-H1C, a kind gift from Prof. Arthur Skoultchi, Albert Einstein School of Medicine) for 1 hour at 4°C and subsequently with protein G magnetic beads. After overnight immunoprecipitation at 4°C, beads were washed 3 times with ChIP buffer 1 and 2 times with ChIP buffer 2. After elution and reverse cross-link (95°C for 15 minutes), samples were treated with proteinase K for 1 hour. Purified DNA and 1% input were analysed by Fast SYBR Green Master Mix qPCR,

using serial 4-fold dilutions of the concentrated input for standard curves and triplicates per sample. Occupancy is plotted as fold enrichment over input and after subtracting background signal from the beads. Primers are listed in Supplementary Table 1.

Embryo Collection and Culture

Embryos were collected into M2 medium (incl. 4 mg/ml BSA) from superovulated C57Bl6xCBA females mated with C57Bl6xCBA or H2B-EGFP[13] males as previously described[14]. Embryos were cultured in KSOM (incl. 4 mg/ml BSA) under paraffin oil in 5% CO_2 at 37.5°C. PADI4 inhibitors (Cl-amidine, 10μM or 200μM, and TDFA, 100μM) and HDAC inhibitor (TSA, Sigma, 10nM) were added to final KSOM from 2-cell stage embryos (44 h after hCG) onwards. The concentration for inhibitors was determined by titration, and was set as the lowest dose that leads to inhibition of enzyme activity and allowed embryonic development to blastocyst stage. Animals were maintained in the Animal Facility of the Gurdon Institute at a 12:12 light cycle and provided with food and water *ad libitum*. All experiments were conducted in compliance with Home Office regulations.

Immunofluorescence Staining and Analysis of Embryos

Embryos were fixed in 4% PFA for Nanog, Sox17, Cdx2 and H3K9 acetylated staining, and in methanol for H3Cit staining. Immunofluorescence staining was carried out as described by Jedrusik et al, 2008[15]. Primary antibodies used were as follows: rabbit α-H3Cit (Abcam, 1:100), rabbit α–H3K9ac (Upstate, 1:100), rabbit α-Nanog (Cosmo Bio Co., 1:200), goat α-Sox17 (R&D systems, 1:200), mouse α-Cdx2 (Cdx2-88) (Biogenex, 1:200). Secondary antibodies used were AlexaFluor 568-conjugated donkey anti-goat, AlexaFluor 488–conjugated donkey anti-mouse and AlexaFluor 633–conjugated donkey anti-rabbit, and AlexaFluor 488-conjugated anti-rabbit and Texas Red-conjugated goat anti-rabbit (Invitrogen, Carlsbad, CA), at

1:400. Collection of 4- cell embryos and E4.5 blastocysts was performed 56 h and 106 h after hCG, respectively. Confocal microscopy was performed and images analyzed using a 40/1.4 NA oil DIC Plan-Apochromatic lens on an inverted Zeiss 510 Meta confocal microscope (Thornwood, NY). Confocal sections were taken every 2 µm through the whole embryo. To objectively measure the fluorescence levels of H3Cit, individual cells were outlined manually in Image J, and the intensity of the fluorescent signal was recorded for each z-stack (3 measurements per nucleus). Only cells in the same z-stacks were compared. Statistical analysis was performed with two-tailed unpaired t-test or Mann Whittney test (for non-normal distributions). Fluorescence and DIC Z-stacks of embryos from late 2-cell (52 h after hCG) to blastocyst stage were collected on 15 focal planes every 15 min for 72 h of continuous embryo culture. The images were processed as described by Bischoff et al., 2008. All cells were followed in 4D using SIMI Biocell software (http://www.simi.com/en/products/biocell/index.html/)[16] as previously described (Bischoff et al., 2008)[14].

SILAC labeling and Mass Spectrometry

ES cells were labeled in culture using the Mouse Embryonic Stem Cell SILAC Protein Quantitation Kit (Pierce) for at least 6 passages. Extracted proteins were resuspended in Laemmli sample buffer, and resolved on a 4-20% SDS-PAGE (NuPAGE, Life Technologies). The gel was stained with Coomassie Blue, cut into 20 slices and processed for mass spectrometric analysis using standard in gel procedure[17]. Briefly, cysteines were reduced with dithiothreitol (DTT), alkylated using chloroacetamide (CAA),[18] and finally the proteins were digested overnight with endoproteinase Lys-C and loaded onto C18 StageTips prior to mass spectrometric analysis.

All MS experiments were performed on a nanoscale HPLC system (EASY-nLC from Thermo Scientific) connected to a hybrid LTQ–Orbitrap Velos[19] equipped with a nanoelectrospray source (Thermo Scientific). Each peptide sample was auto-sampled and separated on a 15 cm analytical column (75cm inner diameter) in-house packed with 3- m C18 beads (Reprosil Pur-AQ, Dr. Maisch) with a 2h gradient from 5% to 40% acetonitrile in 0.5% acetic acid. The effluent from the HPLC was directly electrosprayed into the mass spectrometer.

The MS instrument was operated in data-dependent mode to automatically switch between full-scan MS and MS/MS acquisition. Survey full-scan MS spectra (from m/z 300–1,700) were acquired in the Orbitrap mass analyser with resolution R = 30,000 at m/z 400 (after accumulation to a 'target value' of 1,000,000 in the linear ion trap). The ten most intense peptide ions with charge states ≥2 were subsequently isolated to a target value of 50,000 using predictive automatic gain control (pAGC) and fragmented by higher-energy collisional dissociation (HCD) in the octopole collision cell using normalized collision energy of 40%. The ion selection threshold was set to 5,000 counts for HCD and the maximum allowed ion accumulation times was set to 500 ms for full scans and 250 ms for HCD. All HCD fragment ion spectra were recorded in the Orbitrap mass analyzer with a resolution of 7,500 at m/z 400. For all full scan measurements a lock-mass ion from ambient air (m/z 445.120025) was used for internal calibration when present, as described[20].

Identification of peptides and proteins

Mass spectrometry data analysis was performed with the MaxQuant software suite (version 1.2.6.20) as described[21] supported by Andromeda (www.maxquant.org) as the database search engine for peptide identifications[22]. We followed the step-by-step protocol of the MaxQuant software suite[23] to generate MS/MS peak lists that were filtered to contain at most six peaks per 100 Da interval and searched by

Andromeda against a concatenated target/decoy[24] (forward and reversed) version of the Uniprot human database version (70.101 forward protein entries). Protein sequences of common contaminants such as human keratins and proteases used were added to the database. The initial mass tolerance in MS mode was set to 7 ppm and MS/MS mass tolerance was set to 20 ppm. Cysteine carbamidomethylation was searched as a fixed modification, whereas protein N-acetylation, oxidized methionine, deamidation of Asparagine and Glutamine, and citrullination of Arginines were searched as variable modifications. A maximum of two miscleavages was allowed while we required strict LysC specificity. Peptide assignments were statistically evaluated in a Bayesian model on the basis of sequence length and Andromeda score. We only accepted peptides and proteins with a false discovery rate of less than 1%, estimated on the basis of the number of accepted reverse hits.

Native purification of wild-type and mutant linker histone H1.2

Linker histone H1.2 was expressed from a pET 28b(+) vector, kindly provided by the laboratory of Robert Schneider (MPI fuer Immunobiologie, Freiburg). The vector was mutated using Quickchange® site-directed mutagenesis protocol (Stratagene) for the expression of R54A and R54K mutants. After transformation into expression strain BL21(DE3)-RIL, pre-cultures were diluted 1:100 into 3L of $LB_{Kan/Cam}$ medium at 37°C, 200rpm. At OD_{600} of 0.6, 0.2mM IPTG was added and the bacteria grown for another 2h under the same conditions. Bacterial pellets were harvested by centrifugation (15 min, 4,000xg, 4°C), frozen in liquid nitrogen and stored at -80°C. Linker histone expression was assessed by SDS-PAGE and Coommassie Blue stain. Pellets were thawed on ice, resuspended in 25 mL lysis buffer [20mM hepes, pH 7.6; 100mM NaCl; 5mM EDTA; 1X complete protease inhibitors (Roche); 1mM DTT; 0.2mM PMSF] per litre of culture and lysed through an Avestin high-pressure homogenizer. Lysates were cleared by centrigugation (20,000xg, 40 min, 4°C) and 4M ammonium sulfate added slowly to the supernatant to a final concentration of 2M. The mixture

was incubated on an end-to-end rotator at 4°C for 20 mins before the precipitated proteins were removed by centrifugation (20,000xg, 40 min, 4°C). The supernatant was filtered through a 0.2μm syringe filter and loaded onto a 20 mL HiPrep 16/10 Phenyl FF (low sub) hydrophobic interaction column (Amersham Biosciences), equilibrated in HEMG buffer [20 mM Hepes, pH 7.6; 50μM EDTA; 6.25 mM $MgCl_2$; 0.5 mM DTT; 10% glycerol] containing 2M ammonium sulfate. The sample was injected via a 50 mL super loop at a flow rate of 0.5 ml/min, while the flow through, which contained H1.2, was collected in 5 mL fractions. Ten column volumes of HEMG buffer were used to flush the remaining H1.2 from the column. Fractions were analysed by SDS-PAGE and H1.2-containing fractions were pooled and dialyzed overnight into HEG buffer [20 mM Hepes, pH 7.6; 50μM EDTA; 10% glycerol] containing 100 mM NaCl, using 6-8 kDa MWCO dialysis tubing. The protein solution was centrifuged (20,000 × g, 40 min, 4°C) to remove precipitated proteins and filtered as above. The solution was loaded onto a 6 mL Resource S column, pre-equilibrated with HEG buffer, and proteins eluted with a gradient of 100 mM to 1M NaCl-containing HEG buffer (0.5 mL/mig, 20CV) in 500 μL fractions. Protein fractions were analysed and pooled for the presence of H1.2 as above and dialyzed in HEG buffer with 100 mM NaCl using dialysis tubing with 10 kDa MWCO. Samples were filtered as above and loaded onto a Mono S column (Amersham Biosciences) equilibrated with HEG buffer containing 100 mL NaCl. 500 μL fractions were collected with a gradient of 100 mM to 1M NaCl-containing HEG buffer (0.5 mL/mig, 20CV) and analysed on a 15% SDS-PAGE. Fractions containing H1.2 were pooled, dialyzed as above into 50% glycerol and stored at -20°C.

H1.GFP and Nucleosome pull-down assays

GFP-tagged linker histone H1.2 (H1.GFP) was expressed in ES cells stably over-expression *hPADI4* or control vector by transfection (Lipofectamine 2000) of a pEGFP-H1.2 vector, containing either the wild-type or the R54A mutant sequence, or

with empty pEGFP vector. Transfection efficiency was assessed by visualization of GFP using fluorescence microscopy and determined to be >90%. 24h after transfection cells were harvested and H1.GFP was pulled down using GFP-TRAP conjugated magnetic beads (Chromotek) as per manufacturer's instructions, and subjected to immunoblot analysis.

For *in vitro* nucleosome pull-downs, nucleosomes containing purified and refolded histone octamers and biotinylated 601 DNA were assembled as described previously[25] and immobilized on Dynabeads Streptavidin MyOne T1 (Invitrogen) for 2h at 4^{o}C. Nucleosome-loaded beads in varying amounts were incubated on an end-to-end rotator with wild-type or mutant linker histone H1.2 in Binding buffer [20 mM HEPES, pH 7.9; 150 mM NaCl; 0.2 mM EDTA, 20% Glycerol; 0.1% NP40; 1 mM DTT, and complete protease inhibitors (Roche)] for 4 hr at 4^{o}C. After five washes in binding buffer, the beads were collected and bound proteins were eluted in sample buffer and subjected to Immunoblot analysis.

Purification of recombinant PADI4.GST and in vitro deimination assays

Recombinant human PADI4.GST was expressed from pGEX6p constructs in LB_{Kan} media, induced with 0.1 mM IPTG at 25^{o}C, purified using glutathione-sepharose resin, eluted using a 25mM glutathione solution and stored in 50% glycerol at -20^{o}C. *In vitro* deimination of linker histone H1.2 was carried out in deimination buffer [50 mM HEPES, pH 7.5; 2 mM DTT; in the presence of absence of 10 mM $CaCl_2$] at 37^{o}C for 1h, using the active enzyme. Samples were resuspended in sample buffer for immunoblot analysis.

Treatment of permeabilized cells with recombinant PADI4

Cells were resuspended in a cold solution of 80μg/ml digitonin in SuNaSP[26], mixed gently but thoroughly and incubated on ice for 3mins. Complete permeabilization was

checked by Trypan Blue uptake of sub-aliquots of the suspension and the reaction was stopped by addition of SuNaSP containing 3% BSA. Permeabilized cells were then pre-incubated in wash buffer [20mM Hepes-KOH, 75mM KCl, 1.5mM $MgCl_2$, 5mM $CaCl_2$] with 0.2% Triton X-100 for 30mins at 37°C. Nuclei are washed in wash buffer and incubated in reaction buffer [20mM HEPES-KOH, 75mM KCl, 1.5mM $MgCl_2$, 5mM $CaCl_2$, 2mM DTT] with or without recombinant PADI4 for 1.5hrs at 37°C. The nuclei were washed several times in wash buffer to remove excess PADI4 and released proteins that were not stably bound within the nucleus. The "Washes" fraction contains all proteins that are no longer stably bound, including recombinant hPADI4. The "Pellet" contains all other nuclear-retained proteins. The wash supernatants were concentrated by trichloroacetic acid protein precipitation and subjected to Immunoblot analysis. The washed nuclei were washed a further time in PBS for immunoblot analysis or in MNase buffer and used for DNA compaction assays.

Micrococcal nuclease digestion

Cell pellets were washed twice and resuspended in Micrococcal nuclease (MNase) buffer [15mM Tris, pH 7.5; 15mM NaCl; 60mM KCl; 0.34M sucrose; 0.5mM spermidine; 0.15mM spermine; 0.25mM PMSF; 0.1% β-mercaptoethanol]. 1mM of $CaCl_2$ was added and the suspension was divided into equal aliquots, which were kept on ice. MNase was added to a final concentration of 0.5U/mL and samples were incubated at 25^{o}C in a heat block for varying amounts of time. The reaction was stopped by addition of 0.5M EDTA and 0.5M EGTA and DNA purified using a QIAGEN® PCR purification kit. DNA was quantified and 1500ng of each sample were loaded on a 1.5% agarose gel containing ethidium bromide.

References:

1 Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997-1001, doi:nature04914 [pii] 10.1038/nature04914 (2006).

2 Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545-548, doi:10.1038/nature729 (2002).

3 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-376, doi:10.1038/74199 (2000).

4 Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-635, doi:10.1038/ncb1589 (2007).

5 Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473-483, doi:10.1016/j.cell.2005.07.013 (2005).

6 Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737, doi:S0092-8674(09)00969-6 [pii] 10.1016/j.cell.2009.07.039 (2009).

7 Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr Biol* **21**, 65-71, doi:S0960-9822(10)01584-8 [pii] 10.1016/j.cub.2010.11.074 (2011).

8 Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386 (2000).

9 Barbosa-Morais, N. L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **38**, e17, doi:10.1093/nar/gkp942 (2010).

10 Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548, doi:10.1093/bioinformatics/btn224 (2008).

11 K., S. G. in *'Bioinformatics and Computational Biology Solutions using R and Bioconductor'* (ed V. Carey R. Gentleman, S. Dudoit, R. Irizarry, W. Huber) 397-420 (Springer, New York , 2005).

12 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:10.1093/bioinformatics/btl567 (2007).

13 Hadjantonakis, A. K. & Papaioannou, V. E. Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice. *BMC biotechnology* **4**, 33, doi:10.1186/1472-6750-4-33 (2004).

14 Bischoff, M., Parfitt, D. E. & Zernicka-Goetz, M. Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. *Development* **135**, 953-962, doi:10.1242/dev.014316 (2008).

15 Jedrusik, A. *et al.* Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes & development* **22**, 2692-2706, doi:10.1101/gad.486108 (2008).

16 Schnabel, R., Hutter, H., Moerman, D. & Schnabel, H. Assessing normal embryogenesis in Caenorhabditis elegans using a 4D microscope: variability of development and regional specification. *Developmental biology* **184**, 234-265, doi:10.1006/dbio.1997.8509 (1997).

17 Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* **1**, 2856-2860, doi:10.1038/nprot.2006.468 (2006).
18 Nielsen, M. L. *et al.* Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nature methods* **5**, 459-460 (2008).
19 Olsen, J. V. *et al.* A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **8**, 2759-2769, doi:10.1074/mcp.M900375-MCP200 (2009).
20 Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4**, 2010-2021 (2005).
21 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:nbt.1511 [pii]
10.1038/nbt.1511 (2008).
22 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805, doi:10.1021/pr101065j (2011).
23 Cox, J. *et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **4**, 698-705, doi:nprot.2009.36 [pii]
10.1038/nprot.2009.36 (2009).
24 Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* **4**, 207-214, doi:10.1038/nmeth1019 (2007).
25 Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-484, doi:10.1016/j.cell.2010.10.012 (2010).
26 Halley-Stott, R. P. *et al.* Mammalian nuclear transplantation to Germinal Vesicle stage Xenopus oocytes - a method for quantitative transcriptional reprogramming. *Methods* **51**, 56-65, doi:10.1016/j.ymeth.2010.01.035 (2010).