LETTER

doi:10.1038/nature25784

# The SMAD2/3 interactome reveals that TGFβ controls m$^6$A mRNA methylation in pluripotency

Alessandro Bertero[1]†*, Stephanie Brown[1]*, Pedro Madrigal[1,2], Anna Osnato[1], Daniel Ortmann[1], Loukia Yiangou[1], Juned Kadiwala[1], Nina C. Hubner[3], Igor Ruiz de los Mozos[4], Christoph Sadée[4], An-Sofie Lenaerts[1], Shota Nakanoh[1], Rodrigo Grandy[1], Edward Farnell[5], Jernej Ule[4], Hendrik G. Stunnenberg[3], Sasha Mendjan[1]† & Ludovic Vallier[1,2]

**The TGFβ pathway has essential roles in embryonic development, organ homeostasis, tissue repair and disease[1,2]. These diverse effects are mediated through the intracellular effectors SMAD2 and SMAD3 (hereafter SMAD2/3), whose canonical function is to control the activity of target genes by interacting with transcriptional regulators[3]. Therefore, a complete description of the factors that interact with SMAD2/3 in a given cell type would have broad implications for many areas of cell biology. Here we describe the interactome of SMAD2/3 in human pluripotent stem cells. This analysis reveals that SMAD2/3 is involved in multiple molecular processes in addition to its role in transcription. In particular, we identify a functional interaction with the METTL3–METTL14–WTAP complex, which mediates the conversion of adenosine to N$^6$-methyladenosine (m$^6$A) on RNA[4]. We show that SMAD2/3 promotes binding of the m$^6$A methyltransferase complex to a subset of transcripts involved in early cell fate decisions. This mechanism destabilizes specific SMAD2/3 transcriptional targets, including the pluripotency factor gene *NANOG*, priming them for rapid downregulation upon differentiation to enable timely exit from pluripotency. Collectively, these findings reveal the mechanism by which extracellular signalling can induce rapid cellular responses through regulation of the epitranscriptome. These aspects of TGFβ signalling could have far-reaching implications in many other cell types and in diseases such as cancer[5].**

Activin and NODAL, two members of the TGFβ superfamily, play essential roles in cell fate decision in human pluripotent stem cells (hPSCs)[6–8]. Activin–NODAL signalling is necessary to maintain pluripotency, and inhibition of this pathway drives differentiation towards the neuroectoderm lineage[6,9,10]. Activin–NODAL signalling also cooperates with BMP and WNT pathways to drive mesendoderm specification[11–14]. We therefore used the differentiation of hPSCs into definitive endoderm as a model system to investigate the SMAD2/3 interactome during a dynamic cellular process. To allow a comprehensive and unbiased examination of the proteins that interact with SMAD2/3, we developed an optimized SMAD2/3 co-immunoprecipitation protocol that is compatible with mass-spectrometry analyses (Extended Data Fig. 1a, b and Supplementary Discussion). By examining human embryonic stem cells (hESCs) and hESCs that have been induced to differentiate to endoderm (Fig. 1a), we identified 89 interacting partners of SMAD2/3 (Fig. 1b, Extended Data Fig. 1c, d and Supplementary Table 1). Eleven of these proteins interacted with SMAD2/3 either in hESCs or the differentiating cells but not in both (Extended Data Fig. 1e), suggesting that the SMAD2/3 interactome is largely conserved across these two cell types (Supplementary Discussion). Notably, this list includes known SMAD2/3 transcriptional and epigenetic cofactors (including FOXH1, SMAD4, SNON, SKI, EP300, SETDB1 and CREBBP[3]). We also performed functional experiments on FOXH1, EP300, CREBBP and SETDB1, which uncovered essential functions of these SMAD2/3 transcriptional and epigenetic cofactors in hPSC fate decisions (Extended Data Figs 2, 3 and Supplementary Discussion).

These proteomic experiments also show that SMAD2/3 interacts with complexes involved in functions that have, to our knowledge, not previously been associated with TGFβ signalling (Fig. 1b and Extended Data Fig. 1f), such as ERCC1–XPF and DAPK3–PAWR. which are involved in DNA repair and apoptosis, respectively. We also identified several factors involved in mRNA processing, modification and degradation (Fig. 1b), such as the METTL3–METTL14–WTAP complex (involved in deposition of N$^6$-methyladenosine (m$^6$A)), the PABP-dependent poly(A) nuclease complex (PAN, involved in mRNA decay), the cleavage factor complex CFIm (involved in pre-mRNA 3′ end processing) and the NONO–SFPQ–PSPC1 factors (involved in RNA splicing and nuclear retention of defective RNAs). Overall, these results suggest that SMAD2/3 could be involved in a large number of biological processes in hPSCs, including not only transcriptional and epigenetic regulation, but also non-canonical functions of SMAD2/3.

To further investigate these functions, we examined the function of activin–NODAL signalling in m$^6$A deposition. m$^6$A is the most common RNA modification, and it regulates multiple aspects of mRNA biology including decay and translation[4,15–19]. However, it is not known whether this is a dynamic event that can be modulated by extracellular cues. Furthermore, whereas m$^6$A is known to regulate haematopoietic stem cells[20,21] and the transition between the naive and primed pluripotency states[22,23], its function in hPSCs and during germ layer specification is unknown. We first validated the interaction of SMAD2/3 with METTL3–METTL14–WTAP by co-immunoprecipitation followed by western blotting with both hESCs and human induced pluripotent stem cells (hiPSCs; Fig. 2a and Extended Data Fig. 4a, b). Inhibition of SMAD2/3 phosphorylation blocked this interaction (Fig. 2b and Extended Data Fig. 4c). Proximity ligation assays (PLA) also demonstrated that the interaction occurs in the nucleus (Fig. 2c, d). These observations suggest that SMAD2/3 and the m$^6$A methyltransferase complex interact and that this interaction depends on activin–NODAL signalling.

To investigate the functional relevance of this interaction, we assessed the transcriptome-wide effects of inhibition of activin–NODAL signalling on the deposition of m$^6$A by performing nuclear-enriched m$^6$A-methylated-RNA immunoprecipitation followed by deep sequencing (NeMeRIP–seq; Extended Data Fig. 5a–d, and Supplementary Discussion). Consistent with previous reports[17,19,24], deposition of m$^6$A onto exons was enriched around stop codons and transcription start sites, and occurred at a motif corresponding to the m$^6$A-consensus sequence (Extended Data Fig. 5e–g). Assessment of differential m$^6$A deposition revealed that inhibition of activin–NODAL

[1]Wellcome Trust–MRC Cambridge Stem Cell Institute, Anne McLaren Laboratory and Department of Surgery, University of Cambridge, Cambridge CB2 0SZ, UK. [2]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [3]Department of Molecular Biology, Radboud University, Nijmegen 6525GA, The Netherlands. [4]Francis Crick Institute and Department of Molecular Neuroscience, University College London, London NW1 1AT, UK. [5]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. †Present addresses: Department of Pathology, University of Washington, Seattle 98109, Washington, USA (A.B.); Institute of Molecular Biotechnology, Vienna 1030, Austria (S.M.)
*These authors contributed equally to this work.

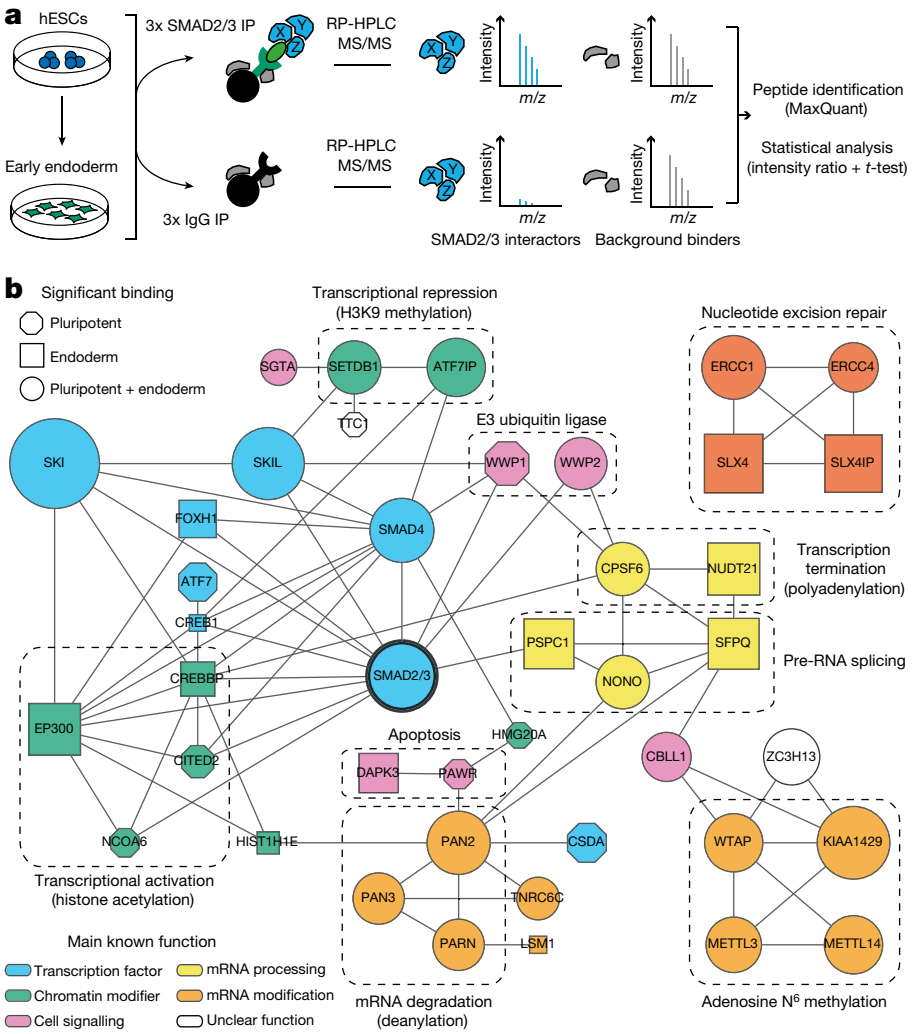



**Figure 1 | Identification of the SMAD2/3 interactome. a**, Experimental approach. IP, immunoprecipitation. RP-HPLC, reversed-phase high-performance liquid chromatography; MS/MS, tandem mass spectrometry. **b**, Interaction network of all known protein–protein interactions between selected SMAD2/3 partners identified in pluripotent and endoderm cells ($n = 3$ co-immunoprecipitations; one-tailed $t$-test, permutation-based false discovery rate (FDR) $< 0.05$). Nodes describe: (1) the lineage in which the proteins were significantly enriched (shape); (2) the significance of the enrichment (size is proportional to the maximum $-\log P$ value); and (3) the function of the factors (colour). Complexes of interest are marked.

signalling predominantly resulted in decreased m$^6$A levels in selected transcripts (Supplementary Table 2; mean absolute $\log_2$ fold-change of 0.56 and 0.35 for m$^6$A decrease and increase, respectively). Decreases in m$^6$A deposition were mostly observed on peaks located near stop codons (Extended Data Fig. 5h), where m$^6$A deposition has been reported to decrease the stability of mRNAs[16,24,25]. Transcripts with reduced m$^6$A levels after inhibition of activin–NODAL signalling largely and significantly overlapped with genes bound by SMAD2/3 ($P < 2.88 \times 10^{-18}$; Extended Data Fig. 5i), including well-known transcriptional targets such as *NANOG*, *NODAL*, *LEFTY1* and *SMAD7* (Fig. 2e and Extended Data Fig. 5j). Accordingly, activin–NODAL-sensitive m$^6$A deposition was largely associated with transcripts that rapidly decreased in abundance during the exit from pluripotency triggered by inhibition of activin–NODAL signalling (Extended Data Fig. 6a). Transcripts that behaved in this fashion were enriched in pluripotency regulators and factors involved in the activin–NODAL signalling pathway (Supplementary Table 3). On the other hand, the expression of a large number of developmental regulators associated with activin–NODAL-sensitive m$^6$A deposition remained unchanged following inhibition of activin–NODAL signalling (Extended Data Fig. 6a–c and Supplementary Table 3). Considered together, these findings show that activin–NODAL signalling can regulate m$^6$A deposition on a number of specific transcripts.

We then examined the molecular mechanisms that underlie the regulation of m$^6$A deposition by activin–NODAL signalling. RNA immunoprecipitation experiments on nuclear RNAs showed that inhibition of activin–NODAL signalling impaired binding of WTAP to several m$^6$A-labelled transcripts including *NANOG* and *LEFTY1* (Fig. 2f and Extended Data Fig. 4d, e), whereas SMAD2/3 itself interacted with these transcripts in the presence of activin–NODAL signalling (Fig. 2g and Extended Data Fig. 4e). Thus, SMAD2/3 appears to promote the recruitment of the m$^6$A methyltransferase complex to nuclear RNAs. Notably, recent reports have established that m$^6$A deposition occurs co-transcriptionally and involves nascent pre-RNAs[16,20,26]. Considering the broad overlap between SMAD2/3 transcriptional targets and transcripts showing activin–NODAL-sensitive m$^6$A deposition (Extended Data Fig. 5i), we hypothesized that SMAD2/3 could facilitate co-transcriptional recruitment of the m$^6$A methyltransferase complex onto nascent transcripts. Consistent with this notion, inhibition of activin–NODAL signalling mainly resulted in downregulation of m$^6$A, not only on exons, but also on pre-mRNA-specific features such as introns and exon–intron junctions (Extended Data Fig. 6d–i and Supplementary Table 2). Moreover, we observed a correlation in activin–NODAL sensitivity across m$^6$A peaks within the same transcript (Extended Data Fig. 6j), suggesting that SMAD2/3 regulates m$^6$A deposition at the level of the genomic locus rather than on a specific mRNA peak. Nevertheless, we did not detect stable and direct binding of the m$^6$A methyltransferase complex to DNA (Extended Data Fig. 4f). Thus, co-transcriptional recruitment might rely on indirect and dynamic interactions with chromatin. Considering all these results, we propose a model in which activin–NODAL signalling promotes co-transcriptional m$^6$A deposition by facilitating the recruitment of the m$^6$A methyltransferase complex onto nascent mRNAs (Fig. 2h).

To understand the functional relevance of this regulation in the context of hPSC cell-fate decisions, we performed inducible knockdown of the subunits of the m$^6$A methyltransferase complex[27] (Extended Data

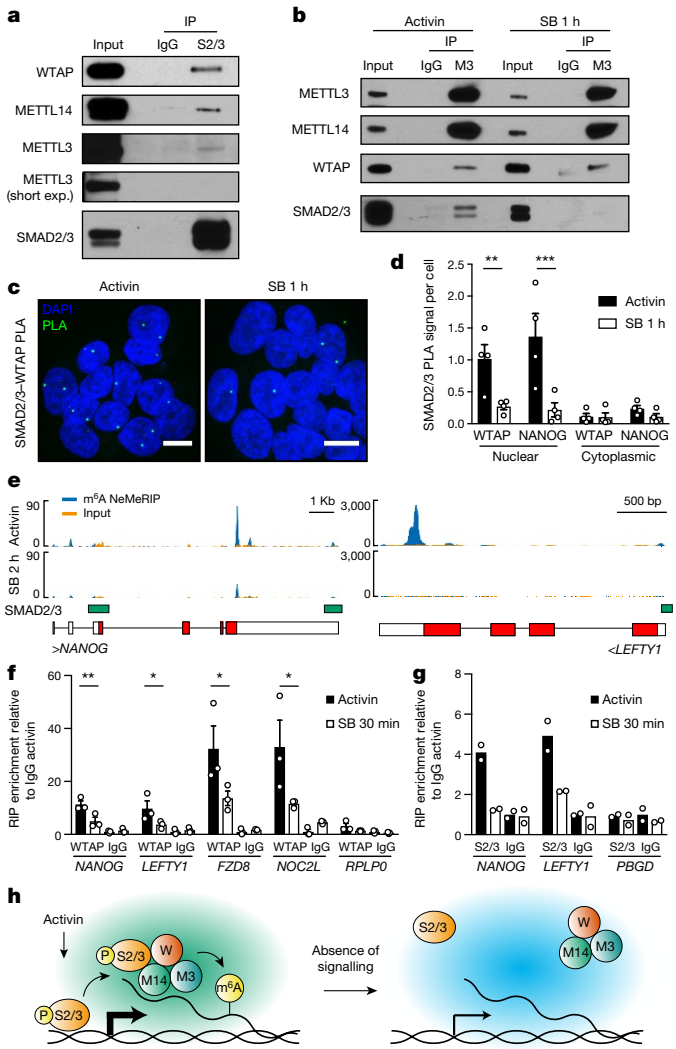



**Figure 2 | Activin–NODAL signalling promotes m$^6$A deposition on specific regulators of pluripotency and differentiation. a**, **b**, Western blots of SMAD2/3 (S2/3), METTL3 (M3) or control (IgG) immunoprecipitations from nuclear extracts of hESCs (representative of three experiments). Input is 5% of the material used for immunoprecipitation. In **b**, immunoprecipitations were performed on hESCs maintained in the presence of activin or treated for 1 h with the activin–NODAL inhibitor SB431542 (SB). For gel Source Data, see Supplementary Fig. 1. **c**, Proximity ligation assays (PLA) for SMAD2/3 and WTAP in hESCs maintained in the presence of activin or SB431542 (representative of two experiments). Scale bars, 10μm. DAPI, nuclei. **d**, PLA quantification; the known SMAD2/3 cofactor NANOG was used as positive control[10]. Mean ± s.e.m., $n = 4$ PLAs; two-way analysis of variance (ANOVA) with post hoc Holm–Sidak comparisons. **e**, Representative results of nuclear-enriched m$^6$A-methylated RNA immunoprecipitation followed by deep sequencing (m$^6$A NeMeRIP–seq; $n = 3$ cultures, replicates combined for visualization). Signal represents read enrichment normalized per million mapped reads and library size. GENCODE gene annotations (red, coding exons; white, untranslated exons; all potential exons are shown and overlaid) and SMAD2/3-binding sites from chromatin immunoprecipitation followed by deep sequencing (ChIP–seq) data[30] are shown. **f**, **g**, RNA immunoprecipitation (RIP) experiments for WTAP, SMAD2/3 or IgG control in hESCs maintained in the presence of activin or treated with SB431542. *RPLP0* and *PBGD* were used as negative controls as they do not contain m$^6$A. In **f**, mean ± s.e.m., $n = 3$ cultures. Two-way ANOVA with post hoc Holm-Sidak comparisons. In **g**, bars show mean, $n = 2$ cultures. **h**, Model of the mechanism by which SMAD2/3 promotes m$^6$A deposition. P, phosphorylation; W, WTAP; M14, METTL14.

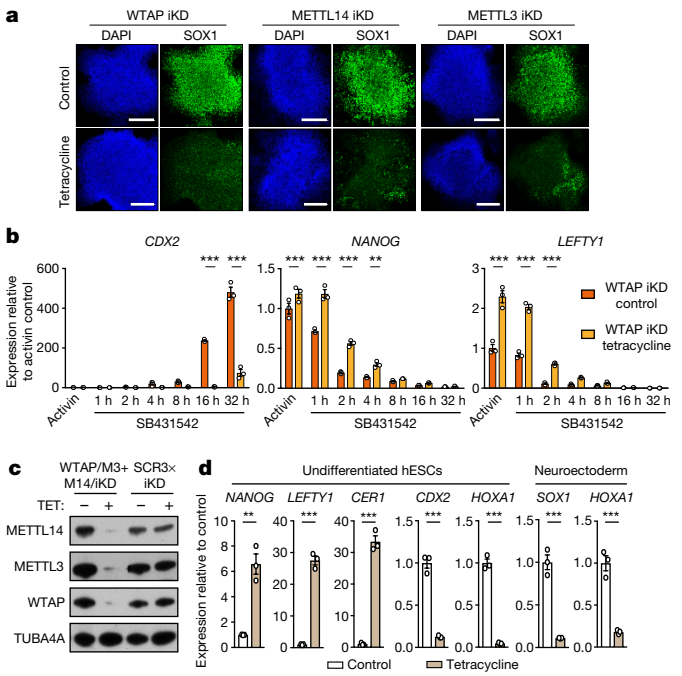



**Figure 3 | The m$^6$A methyltransferase complex antagonizes activin–NODAL signalling in hPSCs to promote timely exit from pluripotency. a**, Immunofluorescence for the neural marker SOX1 following neuroectoderm differentiation of tetracycline (TET)-inducible knockdown (iKD) hESCs (representative of two experiments). Control, no tetracycline; scale bars, 100μm. **b**, Quantitative PCR (qPCR) analyses in WTAP-iKD hESCs with inhibition of activin–NODAL signalling using SB431542 treatments with indicated duration. Mean ± s.e.m., $n = 3$ cultures. Two-way ANOVA with post hoc Holm-Sidak comparisons. **c**, Western blot validation of multiple inducible knockdown (MiKD) hESCs (iKD of WTAP, METTL3 and METTL14). Cells expressing three copies of the scrambled shRNA (SCR3×) were used as negative control. **d**, qPCR analyses in undifferentiated MiKD hESCs, or following differentiation of MiKD hESCs to neuroectoderm. Mean ± s.e.m., $n = 3$ cultures. Two-tailed *t*-test.

Fig. 7a, b). As expected, reducing expression of WTAP, METTL14 or METTL3 decreased m$^6$A deposition (Extended Data Fig. 7c, d); however, prolonged knockdown did not affect pluripotency (Extended Data Fig. 7e, f). We also found that expression of m$^6$A methyltransferase complex subunits was necessary for neuroectoderm differentiation induced by the inhibition of activin–NODAL signalling without being necessary for activin-driven endoderm specification (Fig. 3a and Extended Data Fig. 8a–c). Notably, Activin–NODAL is known to block neuroectoderm induction by promoting NANOG expression[28], whereas NANOG is required for the early stages of endoderm specification[13]. Accordingly, we found that NANOG transcript and protein were upregulated, and the stability of *NANOG* mRNA increased when m$^6$A methyltransferase activity was impaired (Fig. 3b and Extended Data Fig. 9a–c). These results show that m$^6$A deposition decreases the stability of *NANOG* mRNA, facilitating its downregulation upon loss of activin–NODAL signalling, and thereby facilitating exit from pluripotency and neuroectoderm specification (Extended Data Fig. 9d). Additional transcriptomic analyses showed that WTAP knockdown resulted in global upregulation of genes that were transcriptionally activated by SMAD2/3 in hESCs and impaired the upregulation of genes induced by inhibition of activin–NODAL signalling during neuroectoderm differentiation (Fig. 3b, Extended Data Fig. 10a–e, Supplementary Table 4 and Supplementary Discussion). Notably, the decrease in WTAP expression also led to upregulation of m$^6$A-marked mRNAs (Extended Data Fig. 10f), confirming that WTAP-dependent m$^6$A deposition destabilizes mRNAs[16,24,25]. Moreover, transcripts that are rapidly downregulated after inhibition of activin–NODAL signalling were enriched in m$^6$A-marked mRNAs (Extended Data Fig. 10f). Finally, simultaneous knockdown of METTL3, METTL14 and WTAP in hESCs resulted in an even stronger dysregulation of target transcripts of activin–NODAL signalling (Fig. 3c, d and Extended Data Fig. 8d) and

defective neuroectoderm differentiation (Fig. 3d and Extended Data Fig. 8e, f). Together, these results show that the interaction of SMAD2/3 with METTL3–METTL14–WTAP can promote m$^6$A deposition on a subset of transcripts, including a number of pluripotency regulators that are also transcriptionally activated by activin–NODAL signalling. The resulting negative feedback destabilizes these mRNAs and causes their rapid degradation following inhibition of activin–NODAL signalling. This mechanism allows timely exit from pluripotency and induction of neuroectoderm differentiation (Extended Data Fig. 9d).

In conclusion, this analysis of the SMAD2/3 interactome reveals interactions between TGFβ signalling and a wide variety of cellular processes. Our results suggest that SMAD2/3 could act as a hub, coordinating several proteins known to have a role in mRNA processing and modification, apoptosis, DNA repair and transcriptional regulation. This function is illustrated by our results that show activin–NODAL-sensitive regulation of m$^6$A. Activin–NODAL signalling connects transcriptional and epitranscriptional regulation through the interaction between SMAD2/3 and the METTL3–METTL14–WTAP complex, and primes its transcriptional targets for rapid degradation upon withdrawal of signalling (Extended Data Fig. 9d). This avoids overlaps between the pluripotency and neuroectoderm transcriptional programs, thereby facilitating changes in cell identity. We anticipate that further studies will clarify the other non-canonical functions of SMAD2/3, and will dissect how they are related to epigenetic, transcriptional and epitranscriptional regulation of gene expression.

Our findings also clarify and broaden our understanding of the function of m$^6$A in cell-fate decisions. They establish that depletion of m$^6$A in hPSCs does not lead to differentiation, contrary to predictions from studies in mouse-epiblast stem cells[22]. This could imply that there are important functional differences in epitranscriptional regulations between human and mouse pluripotent states. Moreover, widening the conclusions from previous reports[23], we demonstrate that deposition of m$^6$A is specifically necessary for neuroectoderm induction, but not for definitive endoderm differentiation. This can be explained by the fact that in contrast to its strong inhibitory effect on the neuroectoderm lineage[28], expression of NANOG is necessary for the early stages of mesendoderm specification[13,29]. Finally, our results establish that m$^6$A modification of RNA is a dynamic event that is directly modulated by extracellular cues such as TGFβ. Considering the many functions of TGFβ signalling, the regulation we describe here may have an essential function in many cellular contexts that require a rapid response or change in cell state, such as the inflammatory response or cellular proliferation.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 29 November 2016; accepted 22 January 2018.**
**Published online 28 February 2018.**

1. Wu, M. Y. & Hill, C. S. TGFβ superfamily signaling in embryonic development and homeostasis. *Dev. Cell* **16,** 329–343 (2009).
2. Oshimori, N. & Fuchs, E. The harmonies played by TGFβ in stem cell biology. *Cell Stem Cell* **11,** 751–764 (2012).
3. Gaarenstroom, T. & Hill, C. S. TGFβ signaling to chromatin: how SMADs regulate transcription during self-renewal and differentiation. *Semin. Cell Dev. Biol.* **32,** 107–118 (2014).
4. Heyn, H. & Esteller, M. An adenine code for DNA: a second life for N$^6$-methyladenine. *Cell* **161,** 710–713 (2015).
5. Pickup, M., Novitskiy, S. & Moses, H. L. The roles of TGFβ in the tumour microenvironment. *Nat. Rev. Cancer* **13,** 788–799 (2013).
6. Vallier, L., Reynolds, D. & Pedersen, R. A. NODAL inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275,** 403–421 (2004).
7. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/NODAL and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118,** 4495–4509 (2005).
8. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFβ/activin/NODAL signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132,** 1273–1282 (2005).
9. Smith, J. R. *et al.* Inhibition of activin/NODAL signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* **313,** 107–117 (2008).
10. Bertero, A. *et al.* Activin/NODAL signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* **29,** 702–717 (2015).
11. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23,** 1534–1541 (2005).
12. Vallier, L. *et al.* Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem Cells* **27,** 2655–2666 (2009).
13. Teo, A. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **25,** 238–250 (2011).
14. Kubo, A. *et al.* Development of definitive endoderm from embryonic stem cells in culture. *Development* **131,** 1651–1662 (2004).
15. Ke, S. *et al.* A majority of m$^6$A residues are in the last exons, allowing the potential for 3′ UTR regulation. *Genes Dev.* **29,** 2037–2053 (2015).
16. Ke, S. *et al.* m$^6$A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* **31,** 990–1006 (2017).
17. Dominissini, D. *et al.* Topology of the human and mouse m$^6$A RNA methylomes revealed by m$^6$A-seq. *Nature* **485,** 201–206 (2012).
18. Meyer, K. D. *et al.* 5′ UTR m$^6$A promotes cap-independent translation. *Cell* **163,** 999–1010 (2015).
19. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149,** 1635–1646 (2012).
20. Barbieri, I. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m$^6$A-dependent translation control. *Nature* **552,** 126–131 (2017).
21. Vu, L. P. *et al.* The N$^6$-methyladenosine (m$^6$A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23,** 1369–1376 (2017).
22. Geula, S. *et al.* Stem cells. m$^6$A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347,** 1002–1006 (2015).
23. Batista, P. J. *et al.* m$^6$A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15,** 707–719 (2014).
24. Schwartz, S. *et al.* Perturbation of m$^6$A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. *Cell Reports* **8,** 284–296 (2014).
25. Wang, X. *et al.* N$^6$-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).
26. Bartosovic, M. *et al.* N$^6$-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3′-end processing. *Nucleic Acids Res.* **45,** 11356–11370 (2017).
27. Bertero, A. *et al.* Optimized inducible shRNA and CRISPR/Cas9 platforms for *in vitro* studies of human development using hPSCs. *Development* **143,** 4405–4418 (2016).
28. Vallier, L. *et al.* Activin/NODAL signalling maintains pluripotency by controlling NANOG expression. *Development* **136,** 1339–1349 (2009).
29. Mendjan, S. *et al.* NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell* **15,** 310–325 (2014).
30. Brown, S. *et al.* Activin/NODAL signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* **29,** 1176–1185 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank Cambridge Genomic Services for help with next-generation sequencing. This work was supported by the European Research Council starting grant 'Relieve IMDs' (L.V., S.B., A.B., P.M.); the Cambridge University Hospitals National Institute for Health Research Biomedical Research Center (L.V., J.K., A.-S.L.); the Wellcome Trust PhD program (A.O., L.Y.); a British Heart Foundation PhD studentship (FS/11/77/39327 to A.B.); a Grant-in-Aid for JSPS Fellows (16J08005 to S.N.); and a core support grant from the Wellcome Trust and Medical Research Council to the Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute.

**Author Contributions** A.B. conceived the study, performed or contributed to most of the experiments, analysed data and wrote the manuscript with input from the other authors. S.B. contributed to study conception, performed co-immunoprecipitation, NeMeRIP and RNA-IP experiments, and analysed data. P.M., I.R.d.I.M, and C.S. analysed NeMeRIP–seq. A.O. performed PLA and co-immunoprecipitations and analysed RNA-seq data. D.O., L.Y. and J.K. assisted with hPSC gene editing and differentiation; N.C.H. performed quantitative proteomics and data analysis. A.-S.L., S.N. and R.G. assisted with hPSC culture. E.F. optimized NeMeRIP–seq libraries. J.U. contributed to study conception and supervision. H.G.S. supervised quantitative proteomics. S.M. contributed to study conception and supervision, and assisted with SMAD2/3 co-immunoprecipitation. L.V. conceived, supervised and supported the study, and wrote and provided final approval of the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.V. (lv225@cam.ac.uk).

**Reviewer Information** *Nature* thanks C. Mason and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**hPSC culture and differentiation.** Feeder- and serum-free culture of hESCs (H9/WA09 line; WiCell) and hiPSCs ($A1AT^{R/R}$; ref. 31) have been previously described[32]. In brief, cells were plated on gelatin and MEF medium-coated plates, and cultured in chemically defined medium (CDM) containing bovine serum albumin (BSA). CDM was supplemented with 10 ng/ml activin-A and 12 ng/ml FGF2 (both from M. Hyvonen, Dept. of Biochemistry, University of Cambridge). Cells were passaged every 5–6 days with collagenase IV and plated as clumps of 50–100 cells dispensed at a density of 100–150 clumps per $cm^2$. Differentiation was initiated in adherent hESC cultures 48 h after passaging. Definitive endoderm specification was induced for three days (unless stated otherwise) by culturing cells in CDM (without insulin) with 20 ng/ml FGF2, 10 μM LY294002 (PI3K inhibitor; Promega), 100 ng/ml activin-A and 10 ng/ml BMP4 (R&D), as previously described[33]. Neuroectoderm was induced for three days (unless stated otherwise) in CDM-BSA with 12 ng/ml FGF2 and 10 μM SB431542 (activin–NODAL–TGFβ signalling inhibitor; Tocris), as previously described[34]. These same culture conditions were used for activin–NODAL signalling inhibition experiments. hPSCs were routinely monitored for absence of karyotypic abnormalities and mycoplasma infection. As hESCs were obtained from a commercial supplier, cell line identification was not performed. hiPSCs were previously generated in house and genotyped by Sanger sequencing[31].

**Molecular cloning.** Plasmids carrying inducible shRNAs were generated by cloning annealed oligonucleotides into the pAAV-Puro_iKD or pAAV-Puro_siKD vectors as previously described[27]. All shRNA sequences were obtained from the RNAi Consortium TRC library[35] (https://www.broadinstitute.org/rnai/public/). Of the shRNAs that had been validated, the most powerful ones were chosen (the sequences are reported in Supplementary Table 5). Generation of a vector containing shRNAs against METTL3, METTL14 and WTAP (cloned in this order) was performed by Gibson assembly of PCR products containing individual shRNA cassettes, as previously described[27]. The resulting vector was named pAAV-Puro_MsiKD-M3M14W. Generation of the matched control vector containing three copies of the scrambled shRNA sequence (pAAV-Puro_MsiKD-SCR3×) has been described previously[27].

A targeting vector for the *AAVS1* locus carrying constitutively-expressed *NANOG* was generated starting from pAAV_TRE-eGFP[36]. First, the TRE-eGFP cassette was removed using PspXI and EcoRI, and substituted with the CAG promoter (cut from pR26-CAG_eGFP[27] using SpeI and BamHI) by ligating blunt-ended fragments. The resulting vector (pAAV-Puro_CAG) was then used to clone the full-length *NANOG* transcript, which includes its full 5′ and 3′ UTRs. The full-length *NANOG* transcript was constructed from three DNA fragments. The 5′ (bases 1–301) and 3′ (bases 1878–2105) ends were synthesized (IDT) with 40 bp overlaps corresponding to pGem3Z vector linearized with SmaI. The middle fragment was amplified from cDNA of H9 hESCs obtained by retrotranscription with poly-T primer using primers 5′-TTGTCCCCAAAGCTTGCCTTGCTTT-3′ and 5′-CAAAAACGGTAAGAAATCAATTAA-3′. The three fragments and the linearized vector were assembled using a Gibson reaction (NEB) and the sequence of the construct was confirmed by Sanger sequencing. The full length *NANOG* transcript was then subcloned into KpnI- and EcoRV-digested pAAV-Puro_CAG following KpnI and HincII digestion. The resulting vector was named pAAV-Puro_CAG-NANOG.

**Inducible gene knockdown.** Clonal inducible knockdown hESCs for METTL3, METTL14, WTAP or matched controls expressing a scrambled (SCR) shRNA were generated by gene targeting of the *AAVS1* locus with pAAV-Puro_siKD plasmids, which was verified by genomic PCR, all as previously described[27,36]. This same approach was followed to generate multiple inducible knockdown hESCs for METTL3, METTL14 and WTAP (plasmid pAAV-Puro_MsiKD-M3M14W) or matched controls expressing three copies of the SCR shRNA (plasmid pAAV-Puro_MsiKD-SCR3×). Inducible knockdown hESCs for SMAD2, FOXH1, SETDB1, EP300, CREBBP, B2M and matched controls expressing a scrambled shRNA were generated using pAAV-Puro_iKD vectors[27] in hESCs expressing a randomly integrated wild-type tetracycline resistance gene. Two wells were transfected for each shRNA in order to generate independent biological replicates. Following selection with puromycin, the resulting targeted cells in each well were pooled and expanded for further analysis. Given that 20 to 50 clones were obtained for each well, we refer to these lines as 'clonal pools'. Gene knockdown was induced by adding 1 μg/ml tetracycline hydrochloride (Sigma-Aldrich) to the culture medium. Unless indicated otherwise in the text or figure legends, inducible knockdown in undifferentiated hESCs was induced for five days, while differentiation assays were performed in hESCs in which knockdown had been induced for ten days.

**Generation of NANOG-overexpressing hESCs.** NANOG-overexpressing H9 hESCs were obtained by zinc-finger nuclease (ZFN)-facilitated gene targeting of the *AAVS1* locus with pAAV-Puro_CAG-NANOG. This was performed by lipofection of the targeting vector and zinc-finger plasmids followed by puromycin selection, clonal isolation and genotyping screening of targeted cells, all as previously described[27].

**SMAD2/3 co-immunoprecipitation.** Approximately $2 \times 10^7$ cells were used for each immunoprecipitation. Unless stated otherwise, all biochemical steps were performed on ice or at 4 °C, and ice-cold buffers were supplemented with complete protease inhibitors (Roche), PhosSTOP Phosphatase Inhibitor Cocktail (Roche), 1 mg/ml leupeptin, 0.2 mM DTT, 0.2 mM PMSF and 10 mM sodium butyrate (all from Sigma-Aldrich). Cells were fed with fresh medium for 2 h before being washed with PBS, scraped in cell dissociation buffer (CDB, Gibco) and pelleted at 250 g for 10 min. The cell pellet was then washed once with 10 volumes of PBS, and once with 10 volumes of hypotonic lysis buffer (HLB; 10 mM HEPES pH 7.6, 10 mM KCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.2 mM EGTA). The pellet was resuspended in 5 volumes of HLB and incubated for 5 min to induce cell swelling. The resulting cell suspension was homogenized using the 'loose' pestle of a Dounce homogenizer (Jencons Scientific) for 35–50 strokes until plasma membrane lysis was complete (as judged by microscopic inspection). The nuclei were pelleted at 800*g* for 5 min, washed once with ten volumes of HLB, and resuspended in 1.5 volumes of high-salt nuclear lysis buffer (HSNLB; 20 mM HEPES pH 7.6, 420 mM NaCl, 2 mM $MgCl_2$, 25% glycerol, 0.2 mM EDTA, 0.2 mM EGTA). High-salt nuclear extraction was performed by homogenizing the nuclei using the 'tight' pestle of a Dounce homogenizer for 70 strokes, followed by 45 min of incubation with rotation. The resulting lysate was clarified for 30 min at 16,000*g* and transferred to a dialysis cassette using a 19-gauge syringe. Dialysis was performed for 4 h in 1 l of dialysis buffer (20 mM HEPES pH 7.6, 50 mM KCl, 100 mM NaCl, 2 mM $MgCl_2$, 10% glycerol, 0.2 mM EDTA, 0.2 mM EGTA) with gentle stirring, and the buffer was changed once after 2 h. After dialysis, the sample was clarified from minor protein precipitates for 10 min at 17,000*g*, and the protein concentration was assessed. Immunoprecipitations were performed by incubating 0.5 mg of protein with 5 μg of goat polyclonal SMAD2/3 antibody (R&D systems, AF3797) or goat IgG negative control antibody (R&D systems, AB-108-C) for 3 h at 4 °C with rotation. This was followed by incubation with 10 μl of protein-G agarose for 1 h. Beads were washed three times with dialysis buffer and processed for western blot or mass spectrometry. This co-immunoprecipitation protocol is referred to as 'co-IP2' in the Supplementary Discussion and in Extended Data Fig. 1. The alternative SMAD2/3 co-immunoprecipitation protocol (co-IP1) has been previously described[10].

**Mass spectrometry.** Label-free quantitative mass-spectrometric analysis of proteins co-immunoprecipitated with SMAD2/3 or from control IgG co-immunoprecipitations was performed on three replicates for each condition. After immunoprecipitation, samples were prepared as previously described[37] with minor modifications. Proteins were eluted by incubation with 50 μl of 2 M urea and 10 mM DTT for 30 min at room temperature with agitation. Then, 55 mM chloroacetamide was added for 20 min to alkylate reduced disulfide bonds. Proteins were pre-digested on the beads with 0.4 μg of mass-spectrometry-quality trypsin (Promega) for 1 h at room temperature with agitation. The suspension was cleared from the beads by centrifugation. The beads were then washed with 50 μl of 2 M urea, and the combined supernatants were incubated overnight at room temperature with agitation to complete digestion. 0.1% trifluoroacetic acid was then added to inactivate trypsin, and peptides were loaded on $C_{18}$ StageTips[38]. Tips were prepared for binding by sequential equilibration for 2 min at 800*g* with 50 μl methanol, 50 μl Solvent B (0.5% acetic acid; 80% acetonitrile) and 50 μl Solvent A (0.5% acetic acid). Subsequently, peptides were loaded and washed twice with Solvent A. Tips were stored in dry conditions until analysis. Peptides were eluted from the StageTips and separated by reversed-phase liquid chromatography on a 2.5-h-long segmented gradient using EASY-nLC 1000 (ThermoFisher Scientific). Eluting peptides were ionized and injected directly into a Q Exactive mass spectrometer (ThermoFisher Scientific). The mass spectrometer was operated in TOP10 sequencing mode, meaning that one full mass-spectrometry scan was followed by higher energy collision induced dissociation (HCD) and subsequent detection of the fragmentation spectra of the 10 most abundant peptide ions (MS/MS). Collectively, ~160,000 isotype patterns were generated resulting from ~6,000 mass-spectrometry runs. Consequently, ~33,000 MS/MS spectra were measured.

Quantitative mass spectrometry based on dimethyl labelling of samples was performed as described for label-free quantitative mass spectrometry but with the following differences. Dimethyl labelling was performed as previously reported[39,40]. In brief, trypsin-digested protein samples were incubated with dimethyl labelling reagents (4 μl of 0.6 M $NaBH_3CN$ together with 4 μl of 4% $CH_2O$ or $CD_2O$ for light or heavy labelling, respectively) for 1 h at room temperature with agitation. The reaction was stopped by adding 16 μl of 1% $NH_3$. Samples were acidified with 0.1% trifluoroacetic acid, and finally loaded on StageTips. Each immunoprecipitation was performed twice, switching the labels.

**Analysis of mass-spectrometry data.** The raw label-free quantitative mass-spectrometry data were analysed using the MaxQuant software suite[41].

Peptide spectra were compared against the human database (Uniprot) using the integrated Andromeda search engine, and peptides were identified with FDR < 0.01, determined by false matches against a reverse decoy database. Peptides were assembled into protein groups with an FDR < 0.01. Protein quantification was performed using the MaxQuant label-free quantification algorithm requiring at least two ratio counts, in order to obtain label-free quantification (LFQ) intensities. Collectively, the MS/MS spectra were matched to ~20,000 known peptides, leading to the identification of 3,635 proteins in at least one of the conditions analysed. Statistical analysis of the data was performed using the Perseus software package (MaxQuant). First, common contaminants and reverse hits were removed, and only proteins identified by at least two peptides (one of those being unique to the respective protein group) were considered as high-confidence identifications. Proteins were then filtered for those identified in all replicates of at least one condition. LFQ intensities were converted to their log values, and missing intensity values were imputed by representative noise values[42]. One-tailed *t*-tests were then performed to determine the specific interactors in each condition by comparing the immunoprecipitations with the SMAD2/3 antibody to those with the IgG negative controls. Statistical significance was set with a permutation-based FDR < 0.05 (250 permutations). Fold-enrichment over IgG controls was calculated from LFQ intensities.

This same pipeline was used to analyse mass-spectrometry data based on dimethyl labelling, with the following two exceptions. First, an additional mass of 28.03 Da (light) or 32.06 Da (heavy) was specified as 'labels' at the N terminus and at lysines. Second, during statistical analysis of mass-spectrometry data, the outlier significance was calculated based on protein intensity (significance *B*[41]), and was required to be below 0.05 for both the forward and the reverse experiments.

**Biological interpretation of mass-spectrometry data.** The SMAD2/3 protein–protein interaction network was generated using Cytoscape v.2.8.3[43]. First, all the annotated interactions involving the SMAD2/3-binding proteins were inferred by interrogating protein–protein interaction databases through the PSIQUIC Universal Web Service Client. IMEx-complying interactions were retained and merged by union. Then, a subnetwork involving only the SMAD2/3 interactors was isolated. Finally, duplicate nodes and self-loops were removed to simplify visualization. Note that based on our results all the proteins shown would be connected to SMAD2/3, but such links were omitted to simplify visualization and highlight those interactions with SMAD2/3 that were already known. Proteins lacking any link and small complexes of less than three factors were not shown, in order to improve presentation clarity. Note that since the nodes representing SMAD2 and SMAD3 shared the same links, they were fused into a single node (SMAD2/3). Functional enrichment analysis was performed using Fisher's exact test implemented in Enrichr[44], and only enriched terms with a Benjamini–Hochberg adjusted *P* value < 0.05 were considered. For Gene Ontology (GO) enrichment analysis, the 2015 GO annotation was used. For mouse phenotype enrichment analysis, level 3 of the Mouse Genomic Informatics (MGI) annotation was used. To compare protein abundance in different conditions, a cut-off of absolute LFQ intensity $\log_2$ fold-change larger than 2 was chosen, as label-free mass spectrometry is currently not sensitive enough to detect smaller changes with confidence[37].

**Proximity ligation assay.** Proximity ligation assay (PLA) was performed using the Duolink *In situ* Red Starter Kit Goat/Rabbit (Sigma-Aldrich). Cells were cultured on glass coverslips and prepared by fixation in 4% paraformaldehyde (PFA) in PBS for 10 min at room temperature, followed by two gentle washes in PBS. All subsequent incubations were performed at room temperature unless otherwise stated. Samples were permeablilized in PBS containing 0.25% Triton X-100 for 20 min, blocked in PBS with 0.5% BSA for 30 min, and incubated with the two primary antibodies of interest (diluted in PBS with 0.5% BSA; see Supplementary Table 6) for 1 h at 37 °C in a humid chamber. The Duolink *In situ* PLA probes (anti-rabbit minus and anti-goat plus) were mixed and diluted 1:5 in PBS with 0.5% BSA, and pre-incubated for 20 min. Following two washes with PBS containing 0.5% BSA, the coverslips were incubated with the PLA probe solution for 1 h at 37 °C in a humidified chamber. Single-antibody and probes-only negative controls were performed for each antibody tested to confirm assay specificity. Coverslips were washed twice in wash buffer A for 5 min under gentle agitation, and incubated with 1× ligation solution supplemented with DNA ligase (1:40 dilution) for 30 min at 37 °C in a humidified chamber. After two more washes in wash buffer A for 2 min with gentle agitation, coverslips were incubated with 1× amplification solution supplemented with DNA polymerase (1:80 dilution) for 1 h 40 min at 37 °C in a humid chamber. Samples were protected from light from this step onwards. Following two washes in wash buffer B for 10 min, the coverslips were dried overnight, and finally mounted on a microscope slide using Duolink *In situ* Mounting Medium with DAPI. Images of random fields of view were acquired using a LSM 700 confocal microscope (Leica) using a Plan-Apochromat 40×/1.3 Oil DIC M27 objective, performing *z*-stack with optimal spacing (~0.36 μm). Images were analysed automatically using ImageJ. For this, nuclear (DAPI) and PLA *z*-stacks were first individually flattened (max intensity projection) and thresholded to remove background noise. Nuclear images were further segmented using the watershed function. Total nuclei and PLA spots were quantified using the 'analyse particle' function of ImageJ, and nuclear PLA spots were quantified using the 'speckle inspector' function of the ImageJ plugin BioVoxxel.

**RNA immunoprecipitation.** Approximately $2 \times 10^7$ cells were used for each RNA immunoprecipitation (RIP). Unless stated otherwise, all biochemical steps were performed on ice or at 4 °C, and ice-cold buffers were supplemented with cOmplete Protease Inhibitors (Roche) and PhosSTOP Phosphatase Inhibitor Cocktail (Roche). Cells were fed with fresh culture medium 2 h before being washed once with room-temperature PBS and UV crosslinked in PBS at room temperature using a Stratalinker 1800 at 254 nm (400 mJ/cm$^2$). Crosslinked cells were scraped off in cell-dissociation buffer (CDB, Gibco) and pelleted at 250*g* for 5 min. The cell pellet was incubated in five volumes of isotonic lysis buffer (ILB; 10 mM Tris-HCl pH 7.5, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.32 M sucrose) for 12 min to induce cell swelling. Then, Triton X-100 was added to a final concentration of 0.3%, and cells were incubated for 6 min to lyse the plasma membranes. Nuclei were pelleted at 600*g* for 5 min, washed once with ten volumes of ILB, and finally resuspended in two volumes of nuclear lysis buffer (NLB; 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 0.1% Tween) supplemented with 800 U/ml RNasin Ribonuclease Plus Inhibitor (Promega) and 1 μM DTT. The nuclear suspension was transferred to a Dounce homogenizer (Jencons Scientific) and homogenized by performing 70 strokes with a tight pestle. The nuclear lysate was incubated with rotation for 30 min, homogenized again by perfoming 30 additional strokes with the tight pestle, and incubated in rotation for 15 min at room temperature after addition of 12.5 μg/ml of DNase I (Sigma). The protein concentration was assessed, and approximately 1 mg of protein was used for overnight immunoprecipitation with rotation with the primary antibody of interest (Supplementary Table 6), or with equal amounts of non-immune species-matched IgG. Ten per cent of the protein lysate used for immunoprecipitation was saved as pre-immunoprecipitation input and stored at −80 °C for subsequent RNA extraction. Immunoprecipitation reactions were then incubated for 1 h with 30 μl of protein-G agarose, then washed twice with 1 ml of LiCl wash buffer (50 mM Tris-HCl pH 7.5, 250 mM LiCl, 0.1% Triton X-100, 1 mM DTT) and twice with 1 ml of NLB. Beads were resuspended in 90 μl of 30 mM Tris-HCl pH 9.0, and DNase-digested using the RNase-free DNase kit (QIAGEN) by adding 10 μl of RDD buffer and 2.5 μl of DNase. The pre-immunoprecipitation input samples were similarly treated in parallel, and samples were incubated for 10 min at room temperature. The reaction was stopped by adding 2 mM EDTA and by heating at 70 °C for 5 min. Proteins were digested by adding 2 μl of proteinase K (20 mg/ml; Sigma-Aldrich) and by incubating at 37 °C for 30 min. Finally, RNA was extracted with 1 ml of TriReagent (Sigma-Aldrich) according to the supplier's instructions. The RNA was resuspended in nuclease-free water, and half of the sample was used in a reverse-transcription reaction using SuperScript II (ThermoFisher) using the manufacturer's protocol. The other half was used in a control reaction with no reverse transcriptase to confirm successful removal of DNA contaminants. Samples were quantified by quantitative real-time PCR (qPCR), and normalized first to the pre-immunoprecipitation input and then to the IgG control using the $\Delta\Delta C_t$ approach (see below). Supplementary Table 5 shows the primers used.

**Chromatin immunoprecipitation.** Approximately $2 \times 10^7$ cells were used for each chromatin immunoprecipitation (ChIP), and cells were fed with fresh medium 2 h before collection. ChIP was performed using a previously described protocol[10,30]. Briefly, cells were crosslinked on plates, first with protein–protein crosslinkers (10 mM dimethyl 3,3′-dithiopropionimidate dihydrochloride and 2.5 mM 3,3′-dithiodipropionic acid di-*N*-hydroxysuccinimide ester; Sigma-Aldrich) for 15 min at room temperature, then with 1% formaldehyde for 15 min. Crosslinking was quenched with glycine, after which cells were collected, subjected to nuclear extraction, and sonicated to fragment the DNA. Following pre-clearing, the lysate was incubated overnight with the antibodies of interest (Supplementary Table 6) or non-immune IgG. ChIP was completed by incubation with protein-G agarose beads followed by subsequent washes with high salt and LiCl-containing buffers (all exactly as previously described[10,30]). Crosslinking was reverted, first by adding DTT (for disulfide bridge-containing protein–protein crosslinkers), then by incubating in high salt at high temperature. DNA was finally purified by sequential phenol–chloroform and chloroform extractions. Samples were analysed by qPCR using the $\Delta\Delta C_t$ approach (see Supplementary Table 5 for primer sequences). First, a region in the last exon of *SMAD7* was used as internal control to normalize for background binding. Second, the enrichment was normalized to the the enrichment observed in non-immune IgG ChIP controls.

**m$^6$A dot blot.** m$^6$A dot blots were performed as described with minor modifications[23]. Poly-A RNA was purified from total cellular RNA using the Dynabeads mRNA Purification Kit (ThermoFisher), diluted in 50 μl of RNA loading buffer (RLB; 2.2 M formaldehyde, 50% formamide, 0.5× MOPS buffer

(20 mM MOPS, 12.5 mM $CH_3COONa$, 1.25 mM EDTA, pH 7.0)), incubated at 55 °C for 15 min, and snap-cooled on ice. An Amersham Hybond-XL membrane was rehydrated in water for 3 min, then in 10× saline-sodium citrate buffer (SSC; 1.5 M NaCl, 150 mM $Na_3C_6H_5O_7$, pH 7.0) for 10 min, and finally 'sandwiched' in a 96-well dot blot hybridization manifold (ThermoFisher Scientific). Following two washes of the wells with 150 µl of 10× SSC, the RNA was spotted onto the membrane. After UV crosslinking for 2 min at 254 nm using a Stratalinker 1800 (Stratagene), the membrane was washed once with Tris-buffered saline Tween buffer (TBST; 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20), and blocked for 1 h at room temperature with TBST supplemented with 4% non-fat dry milk. Incubations with the anti-m$^6$A primary antibody (Synaptic Systems, 202-111; used at 1 µg/ml) and the mouse-HRP secondary antibody (Supplementary Table 6) were each performed in TBST with 4% milk for 1 h at room temerature, and were followed by three 10-min washes at room temperature in TBST. Finally, the membrane was incubated with ECL2 Western Blotting Substrate (Pierce), and exposed to X-Ray Super RX film.

**m$^6$A nuclear-enriched methylated RNA immunoprecipitation.** m$^6$A MeRIP of nuclear-enriched RNA for analysis by deep sequencing (NeMeRIP–seq) was performed using modifications of previously described methods[23,45]. For each sample, $7.5 \times 10^7$ hESCs were used, and three biological replicates were performed per condition. Cells were fed with fresh medium for 2 h before washing with PBS, scraping in cell dissociation buffer (CDB, Gibco), and pelleting at 250 *g* for 5 min. The cell pellet was incubated in five volumes of isotonic lysis buffer (ILB; 10 mM Tris-HCl pH 7.5, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.32 M sucrose, 1,000 U/ml RNAsin RNase inhibitor (Promega), 1 mM DTT) for 10 min to induce cell swelling. Then, Triton X-100 was added to a final concentration of 0.3% and cells were incubated for 6 min to lyse the plasma membranes. Nuclei were pelleted at 600 *g* for 5 min and washed once with ten volumes of ILB. RNA was extracted from the nuclear pellet using the RNeasy midi kit (QIAGEN) according to the manufacturer's instructions. Residual contaminating DNA was digested in solution using the RNase-free DNase set from QIAGEN, and RNA was re-purified by sequential acid phenol–chloroform and chloroform extractions followed by ethanol precipitation. At this stage, complete removal of DNA contamination was confirmed by qPCR of the resulting RNA without a reverse-transcription step. RNA was then chemically fragmented in 20 µl reactions each containing 20 µg of RNA in fragmentation buffer (10 mM $ZnCl_2$, 10 mM Tris-HCl pH 7.0). Such reactions were incubated at 95 °C for 5 min, followed by inactivation with 50 mM EDTA and storage on ice. The fragmented RNA was then cleaned up by ethanol precipitation. In preparation for MeRIP, 15 µg of anti-m$^6$A antibody (Synaptic Systems, 202-003) or equivalent amounts of rabbit non-immune IgG were crosslinked to 0.5 mg of magnetic beads using the Dynabeads Antibody Coupling Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Following equilibration of the magnetic beads by washing with 500 µl of binding buffer (50 mM Tris-HCl pH 7.5, 150 mM $NaCl_2$, 1% NP-40, 1 mM EDTA), MeRIP reactions were assembled with 300 µg of the fragmented RNA in 3 ml of binding buffer supplemented with 3,000 U of RNAsin RNase inhibitor. Samples were incubated with rotation at 7 r.p.m. for 1 h at room temperature. Fragmented RNA (5 µg, 10% of the amount used for MeRIP) was set aside as pre-MeRIP input control. MeRIP reactions were washed twice with binding buffer, once with low-salt buffer (LSB; 0.25× SSPE (saline-sodium phosphate-EDTA buffer; 150 mM NaCl, 10 mM $NaHPO_4$, 10 mM $Na_2$-EDTA, pH 7.4), 37.5 mM $NaCl_2$, 1 mM EDTA, 0.05% Tween-20), once with high-salt buffer (HSB; 0.25× SSPE, 137.5 mM $NaCl_2$, 1 mM EDTA, 0.05% Tween-20), and twice with TE–Tween buffer (TTB; 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.05% Tween-20). Each wash was performed by incubating the beads with 500 µl of buffer at 7 r.p.m. for 3 min at room temperature. Finally, RNA was eluted from the beads by four successive incubations with 75 µl of elution buffer (50 mM Tris-HCl pH 7.5, 150 mM $NaCl_2$, 20 mM DTT, 0.1% SDS, 1 mM EDTA) at 42 °C. Both the RNA from pooled MeRIP eluates and the pre-MeRIP input were purified and concentrated by sequential acid phenol–chloroform and chloroform extractions followed by ethanol precipitation. Glycogen (30 µg) was added as carrier during ethanol precipitation. RNA was resuspended in 15 µl of ultrapure RNase-free water. DNA libraries were prepared for deep sequencing using the TruSeq Stranded total RNA kit (Illumina) according to the manufacturer's instructions with the following exceptions: (1) Ribo-Zero treatment was performed only for pre-NeMeRIP samples, as there was minimal ribosomal RNA contamination in m$^6$A NeMeRIP samples; (2) since samples were pre-fragmented, the fragmentation step was bypassed and 30 ng of RNA from each sample was used directly for library prep; (3) owing to the small size of the library, a twofold excess of Ampure XP beads was used during all purification steps in order to retain small fragments; (4) owing to the presence of contaminating adapter dimers, the library was gel-extracted using gel-safe stain and a dark reader in order to remove fragments smaller than ~120 bp. Pooled libraries were diluted and denatured for sequencing on the NextSeq 500 (Illumina) according to the manufacturer's instructions. Samples were pooled to obtain >30 million unique clusters per sample. The PhiX control library (Illumina) was spiked into the main library pool at 1% vol/vol for quality control purposes. Sequencing was performed using a high output flow cell with 2 × 75 cycles of sequencing, which provided ~800 million paired-end reads from ~400 million unique clusters from each lane. Overall, an average of ~33 million and ~54 million paired-end reads were generated for m$^6$A MeRIP and pre-MeRIP samples, respectively.

m$^6$A MeRIP samples to be analysed by qPCR (NeMeRIP–qPCR) were processed as described for NeMeRIP–seq, but starting from $2.5 \times 10^7$ cells. MeRIP of cytoplasmic RNA was performed using RNA extracted from the cytoplasmic fraction of cells that were being processed for NeMeRIP. In both cases, MeRIP was performed as for NeMeRIP–seq, but using 2.5 µg of anti-m$^6$A antibody (or equivalent amounts of rabbit non-immune IgG) and 50 µg RNA in 500 µl binding buffer supplemented with 500 U RNAsin RNase inhibitor. At the end of the protocol, the RNA was resuspended in 15 µl ultrapure RNase-free water. For m$^6$A MeRIP on total RNA, the protocol just described was followed exactly, with the exception that the subcellular fractionation step was bypassed, and that total RNA was extracted from $5 \times 10^6$ cells. For m$^6$A MeRIP on mRNA, poly-A RNA was purified from 75 µg total RNA using the Dynabeads mRNA Purification Kit, and 2.5 µg of the resulting mRNA was used for chemical fragmentation and subsequent MeRIP with 1 µg anti-m$^6$A antibody. At the end of all these protocols, cDNA synthesis was performed using all of the MeRIP material in a 30 µl reaction containing 500 ng random primers, 0.5 mM dNTPs, 20 U RNaseOUT, and 200 U SuperScript II (all from Invitrogen) according to the manufacturer's instructions. cDNA was diluted tenfold, and 5 µl of this dilution was used for qPCR using KAPA Sybr Fast Low Rox (KAPA Biosystems). For each gene of interest, two primer pairs were designed against either the region containing the m$^6$A peak[23] or against a negative region (a portion of the same transcript lacking the m$^6$A peak; Supplementary Table 5). Results of MeRIP–qPCR for each gene were then calculated using the $\Delta\Delta C_t$ approach by using the negative region to normalize both for the expression level of the transcript of interest and for background binding.

**Analysis of NeMeRIP–seq data.** Quality of raw sequencing data was assessed using Trimmomatic v.0.35[46], with parameters 'LEADING:3 TRAILING:3 SLIDINGWINDOW:5:10 MINLEN:40'. Reads were aligned to the GRCh38 human genome assembly using TopHat 2.0.13[47] with parameters '–library-type fr-firststrand –transcriptome-index' and the Ensembl GRCh38.83 annotation. Identification of novel splice junctions was allowed. Paired-end and unpaired reads passing quality control were concatenated and mapped in 'single-end' mode in order to be used with MeTDiff[48], which supports only single-end reads. Reads with MAPQ < 20 were filtered out. m$^6$A peak calling and differential RNA methylation in the exome were assessed using MetDiff[48] with pooled inputs for each condition, 'GENE_ANNO_GTF = GRCh38.83, MINIMAL_MAPQ = 20', and of the remaining parameters as default (PEAK_CUTOFF_FDR = 0.05; DIFF_PEAK_CUTOFF_FDR = 0.05). MetDiff calculates *P* values using a likelihood ratio test, then adjusts them to FDR by Benjamini–Hochberg correction. An additional cut-off of absolute fold-change >1.5 (meaning an absolute $\log_2$ fold-change >0.585) was applied for certain analyses as specified in the figure legends or tables. Given known differences between epitranscriptome maps as a function of pipeline[49,50], we confirmed the site-specific and general trends in our data by using an additional pipeline[45]. For this, MACS2[51] was used with parameters '-q 0.05–nomodel–keep-dup all' in m$^6$A NeMeRIP–seq and paired inputs after read alignment with Bowtie 2.2.2.0 (reads with MAPQ < 20 were filtered out). Peaks found in at least two samples were kept for further processing, and a consensus MACS2 peak list was obtained, merging those located within a distance less than 100 bp. The MetDiff and MACS2 peak lists largely overlapped (Extended Data Fig. 5d), and differed primarily because MACS2 identifies peaks throughout the genome while MetDiff identifies only peaks found on the exome (Extended Data Fig. 5c). For the following analyses focused on exonic m$^6$A peaks, we considered a stringent consensus list of only those MetDiff peaks that overlapped with MACS2 peaks (Supplementary Table 2, 'exon m6a'). We assessed the reproducibility of m$^6$A NeMeRIP–seq triplicates in peak regions using the Bioconductor package fCCAC v1.0.0[52]. Hierarchical clustering (Euclidean distance, complete method) of *F* values corresponding to the first two canonical correlations divided the samples into activin and SB431542 clusters. Normalized read-coverage files were generated using the function 'normalize_bigwig' in RSeQC-2.6[53] with default parameters. The distribution of m$^6$A coverage across genomic features was plotted using the Bioconductor package RCAS[54] with sampleN = 0 (no downsampling) and flankSize = 2500. Motif finding on m$^6$A peaks was performed using DREME with default parameters[55]. For visualization purposes, the three biological replicates were combined. The Biodalliance genome viewer[56] was used to generate figures. Gene expression in this experiment was estimated from the pre-MeRIP input samples (which represent an RNA-seq sample on nuclear-enriched RNA species). Quantification, normalization of read counts and estimation of differential gene expression in pre-MeRIP input samples were performed using featureCounts[57]

and DESeq2[58]. For assessment of reproducibility, regularized log transformation of count data was computed, and biological replicates of input samples of the same condition were clustered together in the PC space[59]. Estimation of differential m$^6$A deposition onto each peak in NeMeRIP samples versus input controls was performed using an analogous approach. Functional-enrichment analysis of m$^6$A-marked transcripts was performed using Enrichr[44], as described above for mass-spectrometry data. The coordinates of SMAD2/3 ChIP–seq peaks in hESCs[30] were transferred from their original mappings on hg18 to hg38 using liftOver. Overlap of the resulting intervals with m$^6$A peaks significantly downregulated after 2 h of SB431542 treatment was determined using GAT[60] with default parameters. SMAD2/3-binding sites were assigned to the nearest gene using the annotatePeaks.pl function from the HOMER suite[61] with standard parameters. The significance in the overlap between the resulting gene list and that of genes encoding for transcripts with m$^6$A peaks that are significantly downregulated after 2 h of SB431542 treatment was calculated by a hypergeometric test where the population size corresponded to the number of genes in the standard Ensembl annotation (GRCh38.83).

m$^6$A peaks on introns were identified in three steps (Extended Data Fig. 6d). First, MetDiff was used to simultaneously perform peak calling and differential methylation analysis. Since MetDiff only accepts a transcriptome GTF annotation as an input to determine the genomic space onto which it identifies m$^6$A peaks, in order to determine peaks on introns, we followed the strategy recommended by the package developers of running the software using a custom transcriptome annotation that includes introns[48,62]. This 'extended' transcriptome annotation was built using Cufflinks 2.2.1[63] with parameters '–library-type = fr-firststrand -m 100 -s 50' and guided by the Ensembl annotation (GRCh38.83). This was assembled using all available pre-NeMeRIP input reads. The result was an extended transcriptome annotation including all of the transcribed genome that could be detected and reconstructed from our nuclear-enriched input RNA samples, thus including most expressed introns. Then, MetDiff was run using this extended annotation as input for GENE_ANNO_GTF, pooled inputs for each condition, WINDOW_WIDTH = 40, SLIDING_STEP = 20, FRAGMENT_LENGHT = 250, PEAK_CUTOFF_PVALUE = 1E-03, FOLD_ENRICHMENT = 2, MINIMAL_MAPQ = 20, and all other parameters as default. In a second step, the peaks identified by MetDiff were filtered for robustness by requiring that they overlapped with MACS2 peak calls, exactly as for exome-focused MetDiff peak calls (Extended Data Fig. 5d). Finally, only peaks that strictly did not overlap with any exon based on the Human Gencode annotation V.27 were retained to ensure specificity of mapping to introns (Supplementary Table 2; 'intron m6A'). MetDiff scores for the resulting peak list were used to assess differential m$^6$A deposition based on the cutoff of FDR < 0.05.

m$^6$A exon peaks spanning splice sites were selected from those identified by both the MetDiff analysis on the transcribed genome that was just described and by MACS2. Among these peaks, those presenting sequencing reads that overlap both an exon and an upstream or downstream intron were further selected (Supplementary Table 2; 'splice-site spanning m6A'). Peaks accomplishing MetDiff-calculated FDR < 0.05 and absolute fold-change >1.5 ($\log_2$ fold-change < −0.585) were used to create densities of RPKM-normalized reads inside exons and in the ±500 bp surrounding introns. Biological replicates were merged and depicted on 10 bp-binned heatmaps for visualization purposes. To study the covariation of m$^6$A peaks inside each transcriptional unit, the exonic peak with the greatest downregulated MetDiff fold-change was compared to the mean fold-change of the rest of the m$^6$A peaks found within the gene (both on exons and on introns). The resulting correlation was significant ($P < 2 \times 10^{-16}$; adjusted $R^2 = 0.2221$)

**RNA sequencing.** Poly-A purified opposing-strand-specific mRNA libraries were prepared from 200 ng of total RNA using the TruSeq Stranded mRNA HT sample preparation kit (Illumina). Samples were individually indexed for pooling using a dual-index strategy. Libraries were quantified both with a Qubit (ThermoFisher Scientific) and by qPCR using the NGS Library Quantification Kit (KAPA Biosystems). Libraries were then normalized and pooled. Pooled libraries were diluted and denatured for sequencing on the NextSeq 500 (Illumina) according to the manufacturer's instructions. Samples were pooled so as to obtain >30 million unique clusters per sample (18 samples were split in two runs and multiplexed across four lanes per run). The PhiX control library (Illumina) was spiked into the main library pool at 1% vol/vol for quality control purposes. Sequencing was performed using a high-output flow cell with 2 × 75 cycles of sequencing, which provided ~800 million paired-end reads from ~400 million unique clusters from each run. Overall, a total of ~80 million paired-end reads per sample were obtained.

**Analysis of RNA-seq data.** Reads were trimmed using Sickle[64] with '$q = 20$ and $l = 30$'. To prepare for reads alignment, the human transcriptome was built with TopHat2 v.2.1.0[44] based on Bowtie v.2.2.6[65] by using the human GRCh38.p6 as reference genome, and the Ensembl gene transfer format (GTF) as annotation (http://ftp.ensembl.org/pub/release-83/gtf/homo_sapiens/). All analyses were performed using this transcriptome assembly. Alignment was performed using TopHat2 with standard parameters. Using Samtools view[66], reads with MAPQ >10 were kept for further analyses. Subsequent quantitative data analysis was performed using SeqMonk[67]. The RNA-seq pipeline was used to quantify gene expression as reads per million mapped reads (RPM), and differential expression analysis for binary comparisons was performed using the R package DESeq2[58]. A combined cut-off of negative binomial test $P < 0.05$ and abs.FC > 2 was chosen. Analysis of differentially expressed transcripts across all samples was done using the Bioconductor timecourse package[68] in R. The Hotelling $T^2$ score for each transcript was calculated using the MB.2D function with all parameters set to their default value. Hotelling $T^2$ scores were used to rank probes according to differential expression across the time course, and the top 5% differentially expressed transcripts were selected for complete Euclidean hierarchical clustering ($k$-means preprocessing; max of 300 clusters) using Perseus software. $Z$-scores of $\log_2$ normalized expression values across the timecourse were calculated and used for this analysis. Eight gene clusters were defined, and gene-enrichment analysis for selected clusters was performed using the Fisher's exact test implemented in Enrichr[44]. Only enriched terms with a Benjamini–Hochberg adjusted $P$ value < 0.05 were considered. Principal component analysis was performed on the same list of top 5% differentially expressed transcripts using Perseus.

**Quantitative real-time PCR.** Cellular RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit and the On-Column DNase I Digestion Set (both from Sigma-Aldrich) following the manufacturer's instructions. 500 ng of RNA was used for complementary DNA (cDNA) synthesis using SuperScript II (Invitrogen) according to the manufacturer's instructions. cDNA was diluted 30-fold, and 5 µl was used for qPCR with SensiMix SYBR low-ROX (Bioline) and 150 nM forward and reverse primers (Sigma-Aldrich; see Supplementary Table 5 for primer sequences). Samples were run as technical duplicates in 96-well plates on a Stratagene Mx-3005P (Agilent), and results were analysed using the delta-delta cycle threshold ($\Delta\Delta C_t$) approach[69] using *RPLP0* as housekeeping gene. The reference sample used as control to calculate the relative gene expression is indicated in each figure or figure legend. In cases where multiple control samples were used as reference, the average $\Delta C_t$ from all controls was used when calculating the $\Delta\Delta C_t$. All primers were designed using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and were validated to have a qPCR efficiency >98% and to produce a single PCR product.

**mRNA-stability measurements.** RNA stability was measured by collecting RNA samples at different time points following transcriptional inhibition with 1 µg/ml actinomycin D (Sigma-Aldrich). Following qPCR analyses using equal amounts of mRNA, gene expression was expressed as relative to the beginning of the experiment (no actinomycin D treatment). The data were then fitted to a one-phase decay-regression model[70], and statistical differences in mRNA half-life were evaluated by comparing the model fits by extra sum-of-squares $F$ test.

**Western blots.** Samples were prepared by adding Laemmli buffer (final concentrations: 30 mM Tris-HCl pH 6.8, 6% glycerol, 2% SDS, 0.02% bromophenol blue and 0.25% β-mercaptoethanol), and were denatured at 95 °C for 5 min. Proteins were loaded and run on 4–12% NuPAGE Bis-Tris Precast Gels (Invitrogen), then transferred to polyvinylidene fluoride (PVDF) membranes by liquid transfer using NuPAGE Transfer buffer (Invitrogen). Membranes were blocked for 1 h at room temperature in PBS, 0.05% Tween-20 (PBST) supplemented with 4% non-fat dried milk, and incubated overnight at 4 °C with primary antibody diluted in the same blocking buffer (Supplementary Table 6). After three washes in PBST, membranes were incubated for 1 h at room temperature with horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in blocking buffer (Supplementary Table 6), then washed a further three times with PBST before incubation with Pierce ECL2 Western Blotting Substrate (Thermo) and exposure with X-Ray Super RX Films (Fujifilm).

**Immunofluorescence.** Cells were fixed for 20 min at 4 °C in PBS with 4% PFA, rinsed three times with PBS, and blocked and permeabilized for 30 min at room temperature using PBS with 10% donkey serum (Biorad) and 0.1% Triton X-100 (Sigma-Aldrich). Primary antibodies (Supplementary Table 6) were diluted in PBS with 1% donkey serum and 0.1% Triton X-100 and incubated overnight at 4 °C. This was followed by three washes with PBS and further incubation with AlexaFluor secondary antibodies (Supplementary Table 6) for 1 h at room temperature away from light. Cells were finally washed three times with PBS, and DAPI (Sigma-Aldrich) was added to the first wash to stain nuclei. Images were acquired using a LSM 700 confocal microscope (Leica).

**Flow cytometry.** Single-cell suspensions were prepared by incubation in cell–cell dissociation buffer (CDB; Gibco) for 10 min at 37 °C followed by extensive pipetting. Cells were washed twice with PBS and fixed for 20 min at 4 °C with PBS, 4% PFA. After three washes with PBS, cells were first permeabilized for 20 min at room temperature with PBS, 0.1% Triton X-100, then blocked for 30 min at room temperature with PBS containing 10% donkey serum. Primary and secondary antibody incubations (Supplementary Table 6) were performed for 1 h each at

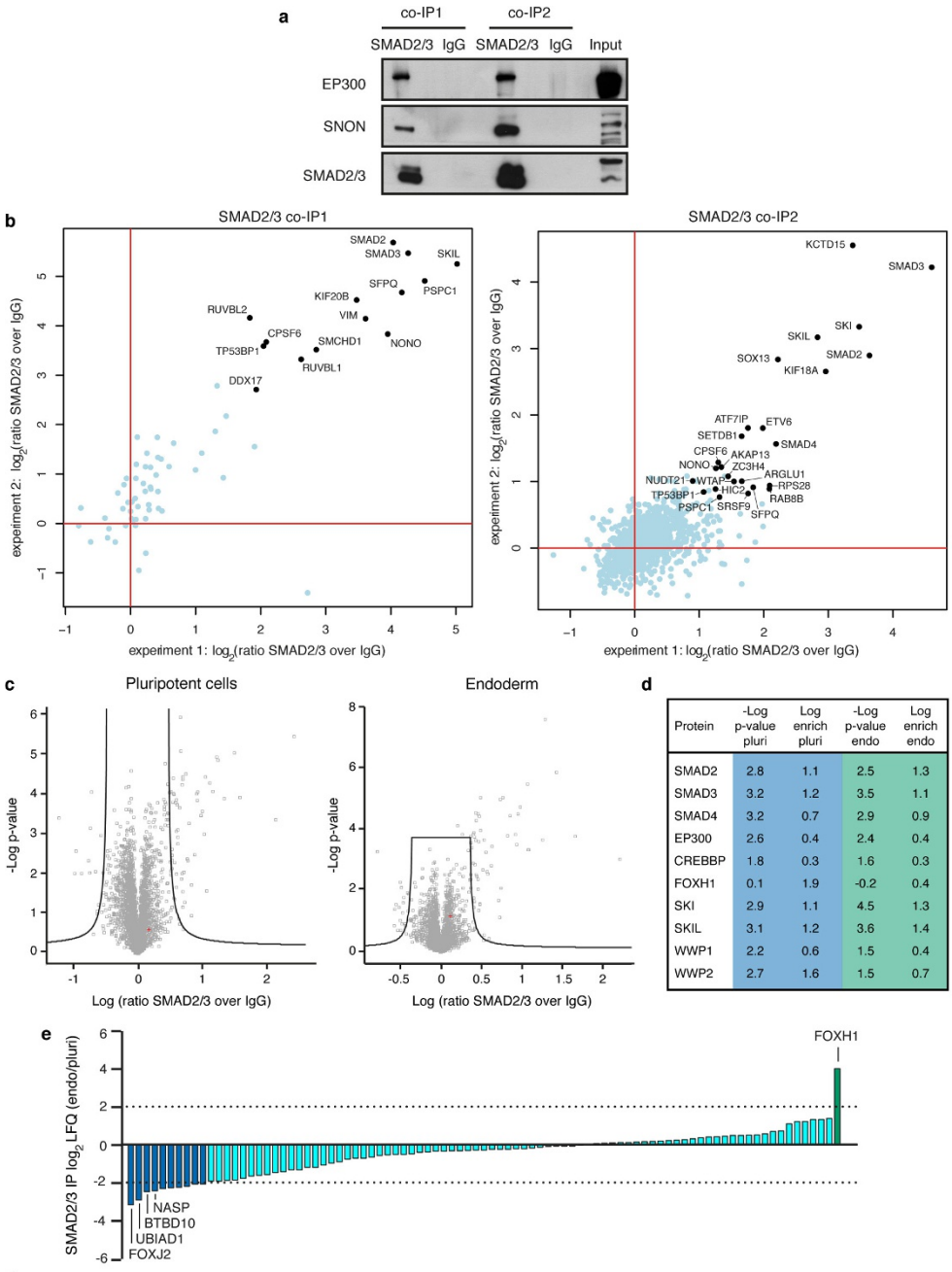
room temperature in PBS, 1% donkey serum, 0.1% Triton X-100, and cells were washed three times with this same buffer after each incubation. Flow cytometry was performed using a Cyan ADP flow cytometer, and at least 10,000 events were recorded. Data analysis was performed using FlowJo X.

**Statistics and reproducibility.** Unless described otherwise in a specific section of the Methods, standard statistical analyses were performed using GraphPad Prism 7 using default parameters. The type and number of replicates, the statistical test used, and the test results are described in the figure legends. The level of significance in all graphs is represented as follows: $*P<0.05$, $**P<0.01$ and $***P<0.001$. Test assumptions (for example, normal distribution) were confirmed where appropriate. For analyses with $n<10$, individual data points are shown, and the mean ± s.e.m. is reported for all analyses with $n>2$. The mean is reported when $n=2$, and no other statistics were calculated for these experiments owing to the small sample size. No experimental samples were excluded from the statistical analyses. Sample size was not pre-determined through power calculations, and no randomization or investigator blinding approaches were implemented during the experiments and data analyses. When representative results are presented, the experiments were reproduced in at least two independent cultures, and the exact number of such replications is detailed in the figure legend.

**Code availability.** Custom bioinformatics scripts used to analyse the data presented in the study have been deposited in GitHub (http://github.com/pmb59/neMeRIP-seq).

**Data availability.** The mass-spectrometry proteomics data that support the findings of this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the identifier PXD005285. Nucleotide sequencing data that support the findings of this study have been deposited to Array Express with identifiers E-MTAB-5229 and E-MTAB-5230. Source Data for the graphical representations found in all figures and Extended Data figures are provided in the Supplementary Information of this manuscript. Electrophoretic gel Source Data (uncropped scans with size marker indications) are presented in Supplementary Fig. 1. Supplementary Tables 1 to 4 provide the results of bioinformatics analyses described in the text and figure legends. All other data that support the findings of this study are available from the corresponding author upon reasonable request.


31. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–394 (2011).
32. Vallier, L. Serum-free and feeder-free culture conditions for human embryonic stem cells. *Methods Mol. Biol.* **690,** 57–66 (2011).
33. Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51,** 1754–1765 (2010).
34. Vallier, L. *et al.* Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS One* **4,** e6082 (2009).
35. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124,** 1283–1298 (2006).
36. Pawlowski, M. *et al.* Inducible and deterministic forward programming of human pluripotent stem cells into neurons, skeletal myocytes, and oligodendrocytes. *Stem Cell Reports* **8,** 803–812 (2017).
37. Hubner, N. C. & Mann, M. Extracting gene function from protein-protein interactions using quantitative BAC interactomics (QUBIC). *Methods* **53,** 453–459 (2011).
38. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–1906 (2007).
39. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4,** 484–494 (2009).
40. Hubner, N. C., Nguyen, L. N., Hornig, N. C. & Stunnenberg, H. G. A quantitative proteomics tool to identify DNA–protein interactions in primary cells or blood. *J. Proteome Res.* **14,** 1315–1329 (2015).
41. 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 (2008).
42. Hubner, N. C. *et al.* Quantitative proteomics combined with BAC transgeneomics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–754 (2010).
43. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).
44. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).
45. Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of $N^6$-methyladenosine by $m^6A$-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* **8,** 176–189 (2013).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
47. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
48. Cui, X. *et al.* MeTDiff: a novel differential RNA methylation analysis for MeRIP–seq data. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **PP,** 1 (2015).
49. Saletore, Y. *et al.* The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13,** 175 (2012).
50. Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14,** 23–31 (2017).
51. Zhang, Y. *et al.* Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9,** R137 (2008).
52. Madrigal, P. fCCAC: functional canonical correlation analysis to evaluate covariance between nucleic acid sequencing datasets. *Bioinformatics* **33,** 746–748 (2017).
53. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012).
54. Uyar, B. *et al.* RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic Acids Res.* **45,** e91 (2017).
55. Bailey, T. L. DREME: motif discovery in transcription factor ChIP–seq data. *Bioinformatics* **27,** 1653–1659 (2011).
56. Down, T. A., Piipari, M. & Hubbard, T. J. P. Dalliance: interactive genome viewing on the web. *Bioinformatics* **27,** 889–890 (2011).
57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
59. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17,** 13 (2016).
60. Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29,** 2046–2048 (2013).
61. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
62. Meng, J., Cui, X., Rao, M. K., Chen, Y. & Huang, Y. Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics* **29,** 1565–1567 (2013).
63. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).
64. Joshi, N. & Fass, J. *Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files* v.1.33 https://github.com/najoshi/sickle (2011).
65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
66. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
67. Andrews, S. *SeqMonk: A tool to visualise and analyse high throughput mapped sequence data* https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/ (2014).
68. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** e3 (2004).
69. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **25,** 402–408 (2001).
70. Harrold, S., Genovese, C., Kobrin, B., Morrison, S. L. & Milcarek, C. A comparison of apparent mRNA half-life using kinetic labeling techniques vs decay following administration of transcriptional inhibitors. *Anal. Biochem.* **198,** 19–29 (1991).




| Protein | -Log p-value pluri | Log enrich pluri | -Log p-value endo | Log enrich endo |
|---|---|---|---|---|
| SMAD2 | 2.8 | 1.1 | 2.5 | 1.3 |
| SMAD3 | 3.2 | 1.2 | 3.5 | 1.1 |
| SMAD4 | 3.2 | 0.7 | 2.9 | 0.9 |
| EP300 | 2.6 | 0.4 | 2.4 | 0.4 |
| CREBBP | 1.8 | 0.3 | 1.6 | 0.3 |
| FOXH1 | 0.1 | 1.9 | -0.2 | 0.4 |
| SKI | 2.9 | 1.1 | 4.5 | 1.3 |
| SKIL | 3.1 | 1.2 | 3.6 | 1.4 |
| WWP1 | 2.2 | 0.6 | 1.5 | 0.4 |
| WWP2 | 2.7 | 1.6 | 1.5 | 0.7 |

f

| GO molecular function term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| RNA Pol II transcription factor binding | 1 | 6.72E-06 | 8 |
| Poly(A)-specific ribonuclease activity | 3 | 1.97E-03 | 3 |
| SMAD binding | 4 | 1.97E-03 | 5 |
| Transcription corepressor activity | 8 | 2.57E-03 | 7 |
| Activating transcription factor binding | 9 | 4.42E-03 | 4 |

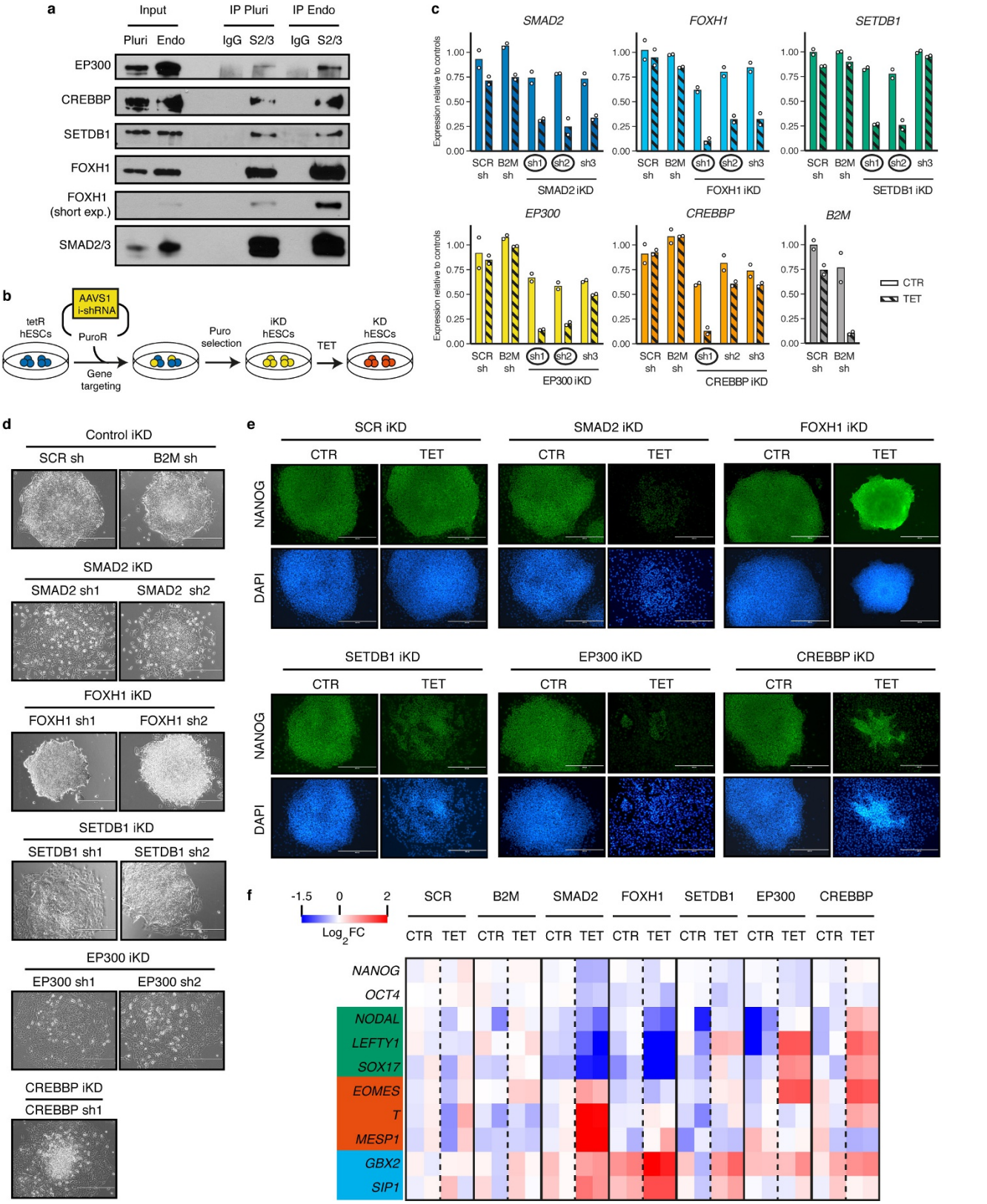
| GO biological process term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| mRNA processing | 1 | 1.53E-03 | 12 |
| mRNA modification | 2 | 2.60E-02 | 3 |
| Gene expression | 4 | 4.25E-02 | 12 |
| mRNA deadenylation-dependent decay | 5 | 4.25E-02 | 4 |
| Embryonic organ development | 11 | 4.78E-02 | 5 |

| GO cellular compartment term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| Nuclear body | 1 | 7.33E-07 | 11 |
| Transcription factor complex | 2 | 4.06E-06 | 11 |
| Nucleotide-excision repair complex | 7 | 2.84E-04 | 3 |
| Methyltransferase complex | 11 | 6.52E-03 | 4 |
| mRNA cleavage factor complex | 16 | 2.19E-02 | 3 |

| MGI mouse phenotype term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| Abnormal embryogenesis/development | 1 | 1.92E-08 | 19 |
| Abnormal survival | 2 | 7.16E-06 | 24 |
| Abnormal craniofacial morphololgy | 5 | 2.82E-02 | 8 |
| Abnormal respiratory system | 6 | 2.82E-02 | 6 |
| Abnormal digestive system | 7 | 2.82E-02 | 7 |

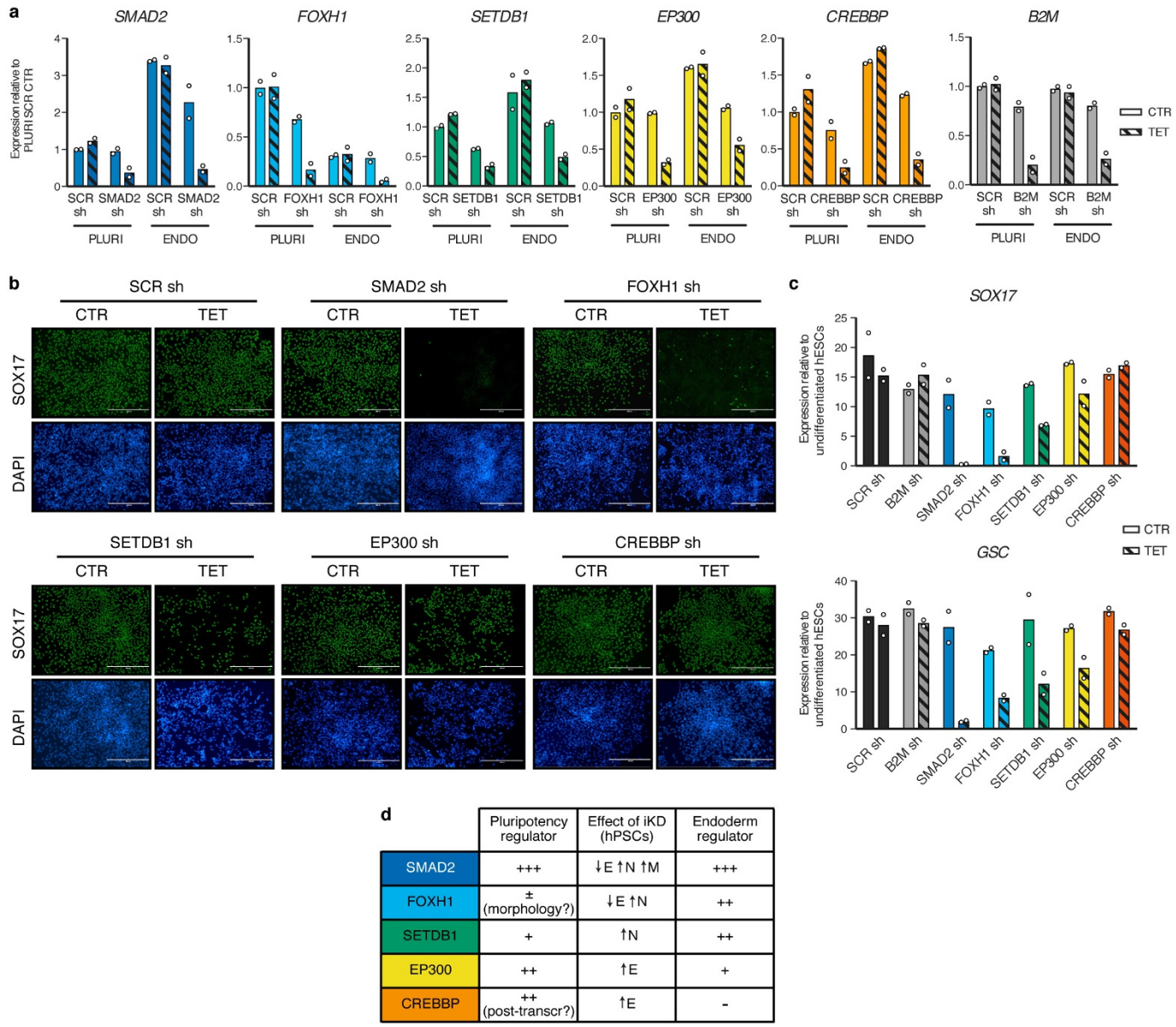
**Extended Data Figure 1** | See next page for caption.

**Extended Data Figure 1 | Optimized co-immunoprecipitation protocol to define the SMAD2/3 interactome in hPSCs and early endoderm cells.** **a**, Western blots of SMAD2/3 or control (IgG) immunoprecipitations from nuclear extracts of hESCs following the co-IP1 or co-IP2 protocols. Input is 5% of the material used for immunoprecipitations. Results are representative of two independent experiments. For gel Source Data, see Supplementary Fig. 1. **b**, Scatter plots of the $\log_2$ ratios of label-free quantification (LFQ) intensities for proteins identified by quantitative mass spectrometry in SMAD2/3 co-immunoprecipitations compared with IgG negative control co-immunoprecipitations. The experiments were performed from nuclear extracts of hESCs. The SMAD2/3 and IgG negative control co-immunoprecipitations were differentially labelled after immunoprecipitation using the dimethyl method, followed by a combined run of the two samples in order to compare the abundance of specific peptides and identify enriched peptides. The values for technical dye-swap duplicates are plotted on different axes, and proteins whose enrichment was significant (significance $B < 0.01$) are shown in black and named. As a result of this comparison between the two co-immunoprecipitation protocols, co-IP2 was selected for further experiments (see Supplementary Discussion). **c**, Volcano plots of statistical significance against fold-change for proteins identified by label-free quantitative mass spectrometry in SMAD2/3 or IgG negative control immunoprecipitations in pluripotent hESCs or early endoderm (see Fig. 1a). The black lines indicate the threshold used to determine specific SMAD2/3 interactors, which are located to the right ($n = 3$ co-immunoprecipitations; one-tailed $t$-test: permutation-based FDR $< 0.05$). **d**, Selected results of the analysis described in **c** for SMAD2, SMAD3 and selected known bona fide SMAD2/3-binding partners (full results can be found in Supplementary Table 1). **e**, Mean label-free quantification (LFQ) intensity $\log_2$ ratios in endoderm (endo) and pluripotent cells (pluri) for all SMAD2/3 interactors. Differentially enriched proteins are shown as green and blue bars. **f**, Selected results from gene ontology (GO) enrichment analysis, and enrichment analysis for mouse phenotypes annotated in the MGI database. All putative SMAD2/3-interacting proteins were considered for this analysis ($n = 89$ proteins; Fisher's exact test followed by Benjamini–Hochberg correction for multiple comparisons). For each term, its rank in the analysis, the adjusted $P$ value, and the number of associated genes are reported.




**Extended Data Figure 2** | See next page for caption.

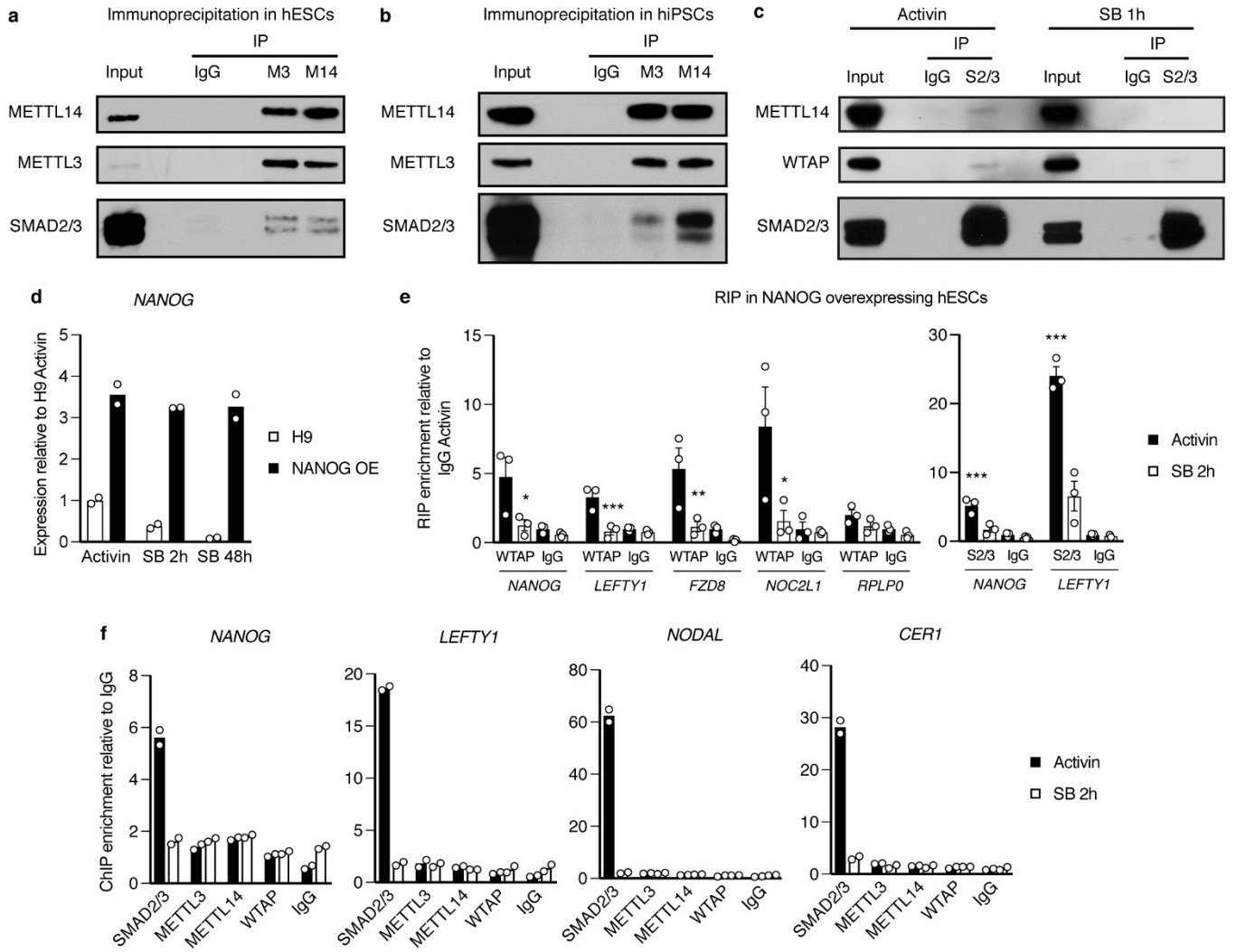
**Extended Data Figure 2 | Functional characterization of transcriptional and epigenetic cofactors of SMAD2/3 in hPSCs. a**, Western blots of SMAD2/3 or control (IgG) immunoprecipitations from nuclear extracts of pluripotent hESCs (pluri) or hESCs differentiated into endoderm for 36 h (endo). Input is 5% of the material used for immunoprecipitations. Results are representative of two independent experiments. **b**, Schematic of the experimental approach for the generation of iKD hESC lines for SMAD2/3 cofactors. **c**, qPCR screening of iKD hESCs cultured in the absence (CTR) or presence (TET) of tetracycline for three days. Three distinct shRNAs were tested for each gene. Expression is normalized to the mean level in hESCs carrying negative control shRNAs (scrambled (SCR) or against B2M) and cultured in the absence of tetracycline. The mean is indicated, $n = 2$ independent clonal pools. Note than for the B2M shRNA only the scrambled shRNA was used as negative control. shRNAs selected for further experiments are circled. **d**, Phase-contrast images of iKD hESCs expressing the indicated shRNAs (sh) and cultured in the presence of tetracycline for six days to induce knockdown. Scale bars, 400 μm. Results are representative of two independent experiments. **e**, Immunofluorescence for the pluripotency factor NANOG in iKD hESCs for the indicated genes cultured in the absence (CTR) or presence of tetracycline (TET) for six days. Scale bars, 400 μm. Results are representative of two independent experiments. **f**, Heat map summarizing qPCR analyses of iKD hESCs cultured as in **e**. $\log_2$ fold-changes (FC) are compared to scrambled control ($n = 2$ clonal pools). Germ-layer markers are grouped in boxes: green, endoderm; red, mesoderm; blue, neuroectoderm.



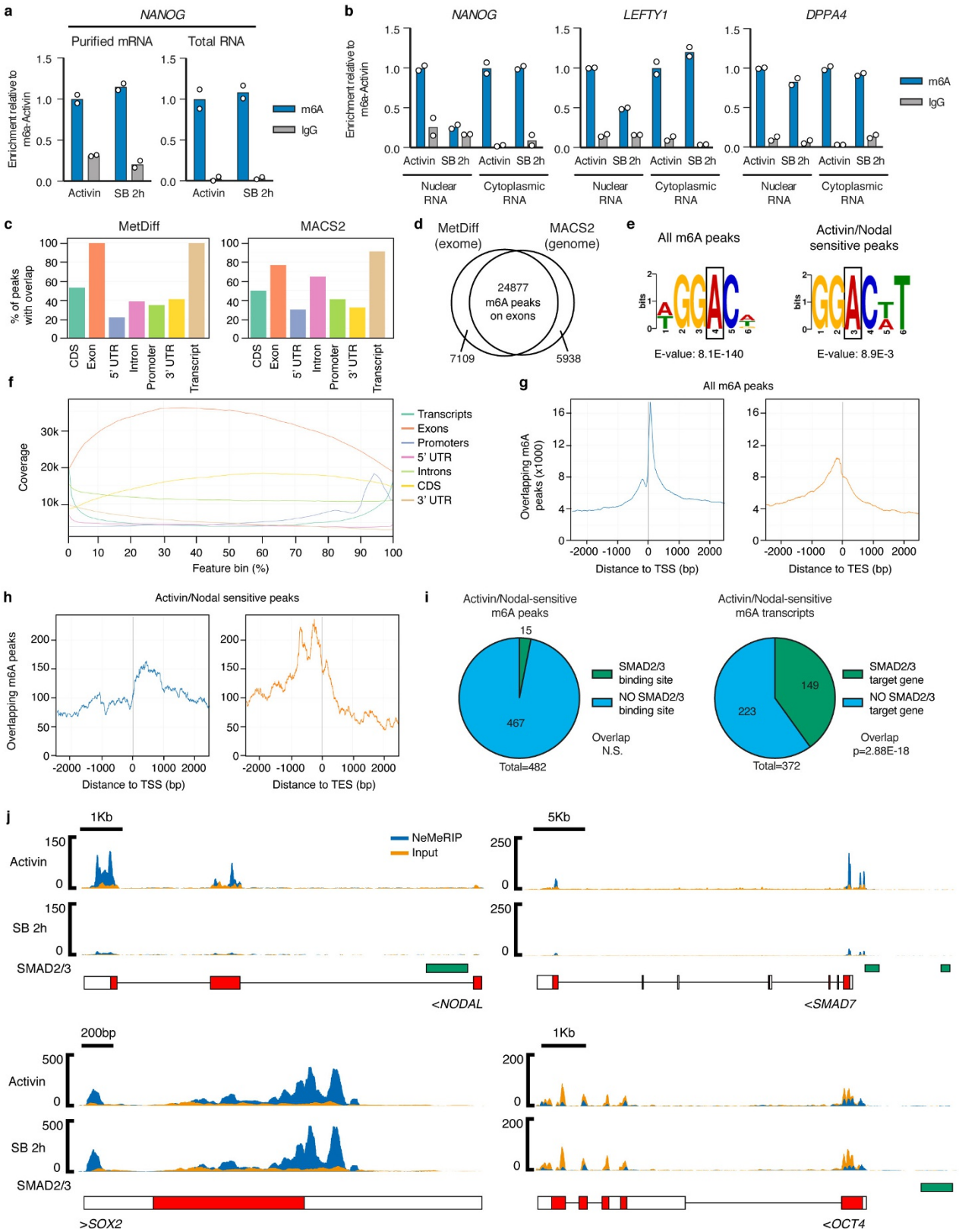

d

| | Pluripotency regulator | Effect of iKD (hPSCs) | Endoderm regulator |
|---|---|---|---|
| SMAD2 | +++ | ↓E ↑N ↑M | +++ |
| FOXH1 | ± (morphology?) | ↓E ↑N | ++ |
| SETDB1 | + | ↑N | ++ |
| EP300 | ++ | ↑E | + |
| CREBBP | ++ (post-transcr?) | ↑E | - |

**Extended Data Figure 3 | Functional characterization of transcriptional and epigenetic cofactors of SMAD2/3 during endoderm differentiation.** **a**, qPCR validation of iKD hESCs in pluripotent cells (PLURI) or following endoderm differentiation (ENDO). Pluripotent cells were cultured in the absence (CTR) or presence (TET) of tetracycline for six days. For endoderm differentiation, tetracycline treatment was initiated in undifferentiated hESCs for three days in order to ensure gene knockdown at the start of endoderm specification, and was then maintained during differentiation (three days). For each gene, the shRNA resulting in the strongest level of knockdown in hPSCs was selected (refer to Extended Data Fig. 2). Expression is normalized to the mean level in pluripotent hESCs carrying scrambled control shRNA and cultured in the absence of tetracycline. The mean is indicated, $n = 2$ independent clonal pools. **b**, Immunofluorescence of the endoderm marker SOX17 following endoderm differentiation of iKD hESCs expressing the indicated shRNAs and cultured as described in **a**. Scale bars, 400 μm. Results are representative of two independent experiments. **c**, qPCR following endoderm differentiation of iKD hESCs. The mean is indicated, $n = 2$ independent clonal pools. **d**, Table summarizing the phenotypic results presented in Extended Data Fig. 2 and in this figure. E, endoderm; N, neuroectoderm; M, mesoderm.

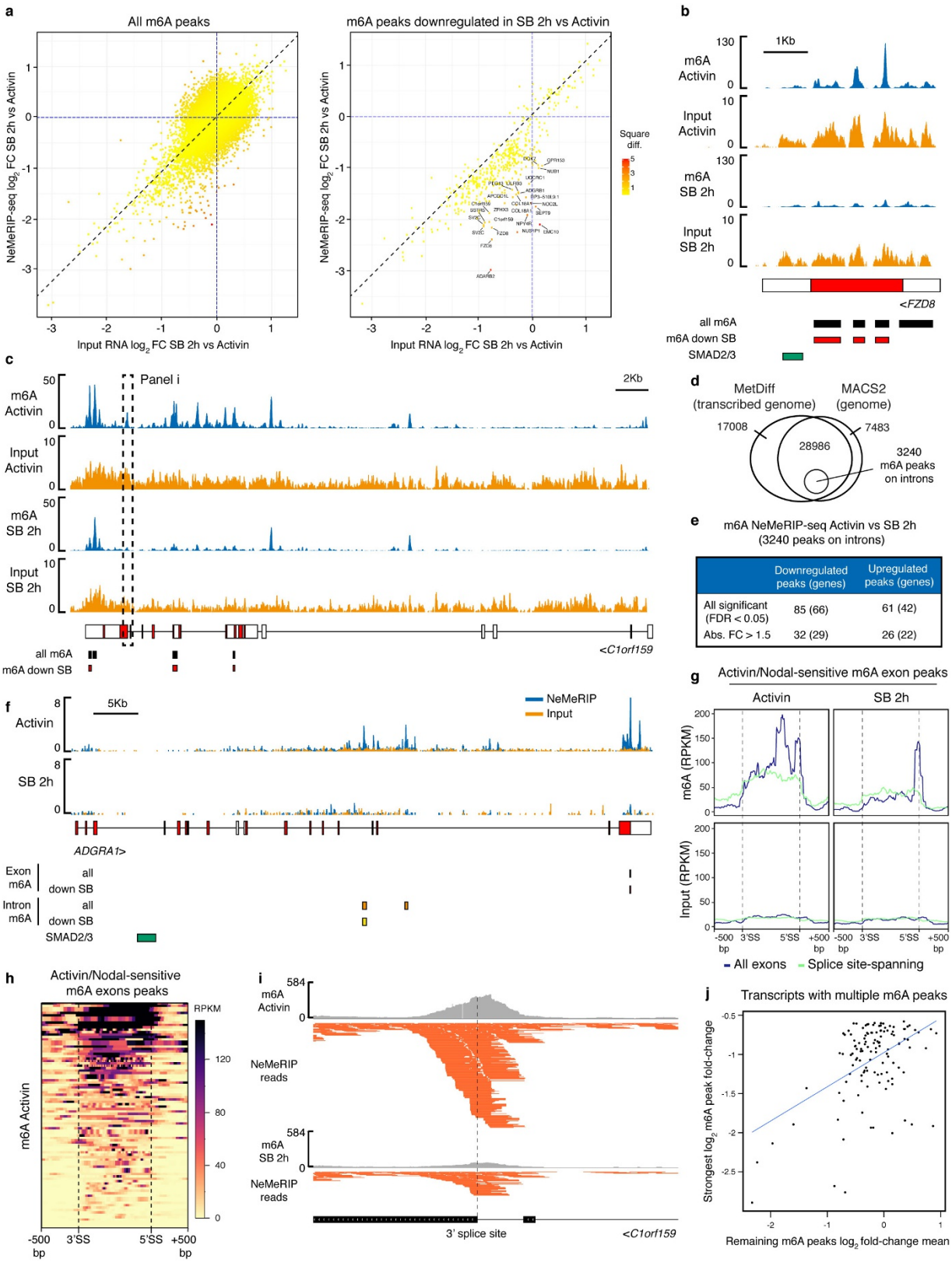



**Extended Data Figure 4 | Mechanistic insights into the functional interaction between SMAD2/3 and the m$^6$A methyltransferase complex.** **a**–**c**, Western blots of SMAD2/3, METTL3, METTL14 or control (IgG) immunoprecipitations from nuclear extracts of hESCs (**a**, **c**) or hiPSCs (**b**). Input is 5% of the material used for immunoprecipitations. In **c**, immunoprecipitations were performed from hPSCs maintained in the presence of activin or treated for 1 h with the activin–NODAL signalling inhibitor SB431542. Results are representative of three (**a**) or two (**b**, **c**) independent experiments. **d**, qPCR validation of hESCs constitutively overexpressing NANOG (NANOG OE) following gene targeting of the *AAVS1* locus with pAAV-Puro_CAG-NANOG. Parental wild-type H9 hESCs (H9) were analysed as negative controls. Cells were cultured in the presence of activin or treated with SB431542 for the indicated times. The mean is shown, $n = 2$ cultures. NANOG-overexpressing cells are resistant to downregulation of NANOG following inhibiton of activin–NODAL signalling. **e**, RNA immunoprecipitation experiments for WTAP, SMAD2/3 or IgG control in NANOG-overexpressing hESCs maintained in the presence of activin or treated for 2 h with SB431542. Enrichment of the indicated transcripts was measured by qPCR and expressed relative to background levels observed in control IgG RNA immunoprecipitations in the presence of activin. *RPLP0* was tested as a negative control transcript. Mean ± s.e.m., $n = 3$ cultures. Significance of differences from activin (left) or IgG (right) RIP was tested by two-way ANOVA with post hoc Holm–Sidak comparisons. **f**, ChIP–qPCR in hESCs for ChIP against the indicated proteins or the negative control ChIP (IgG). qPCR was performed for validated genomic SMAD2/3-binding sites associated with the indicated genes[10,30]. hESCs were cultured in the presence of activin or treated for 2 h with SB431542. Enrichment is normalized against background binding observed with IgG ChIP. The mean is shown, $n = 2$ technical replicates. Results are representative of three independent experiments.



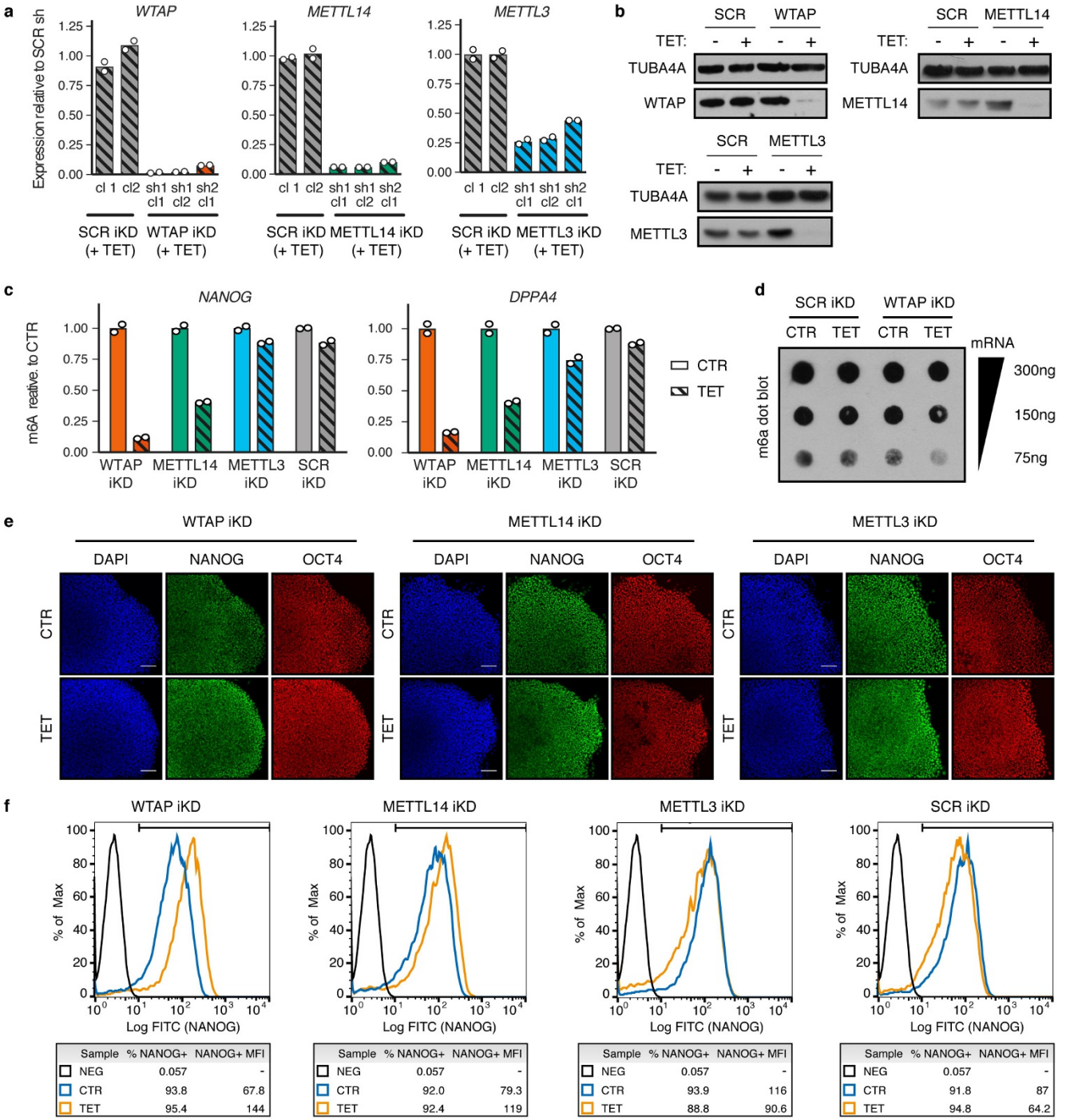

**Extended Data Figure 5** | See next page for caption.

**Extended Data Figure 5 | Monitoring changes in m$^6$A deposition that are rapidly induced by inhibition of activin–NODAL signalling.** **a**, **b**, MeRIP–qPCR results from purified mRNA, total cellular RNA or cellular RNA species separated by nuclear and cytoplasmic subcellular fractionation. hESCs were cultured in pluripotency-maintaining conditions containing activin or in conditions in which activin–NODAL signalling was inhibited for 2 h with SB431542. IgG MeRIP experiments were performed as negative controls. The mean is indicated, $n = 2$ technical replicates. Differences between activin- and SB431542-treated cells were observed only in the nuclear-enriched fraction. Therefore, the nuclear-enriched MeRIP protocol (NeMeRIP) was used for subsequent experiments (refer to the Supplementary Discussion). Results are representative of two independent experiments. **c**, Overlap with the indicated genomic features of m$^6$A peaks identified by NeMeRIP–seq using two different bioinformatics pipelines in which peak calling was performed using MetDiff or MACS2. For each pipeline, the analyses were performed on the union of peaks identified from data obtained in hESCs cultured in the presence of activin or with inhibition of activin–NODAL signalling for 2 h with SB431542 ($n = 3$ cultures). Note that the sum of the percentages within each graph is not 100% because some m$^6$A peaks overlap several feature types. MetDiff is an exome peak caller, and, accordingly, 100% of peaks map to exons. MACS2 identifies peaks throughout the genome. **d**, Venn diagrams showing the overlap of peaks identified by the two pipelines. Only MetDiff peaks that were also identified by MACS2 were considered for subsequent analyses focused on m$^6$A peaks on exons. **e**, Top sequence motifs identified *de novo* on all m$^6$A exon peaks, or on those that showed significant downregulation following inhibition of activin–NODAL signalling (activin–NODAL-sensitive m$^6$A peaks; Supplementary Table 2). The position of the methylated adenosine is indicated by a box. **f**, Coverage profiles for all m$^6$A exon peaks across the length of different genomic features. Each feature type is expressed as 100 bins of equal length with 5′ to 3′ directionality. **g**, **h**, Overlap of m$^6$A exon peaks and transcription start sites (TSS) or transcription end sites (TES). In **g**, the analysis was performed for all m$^6$A peaks. In **h**, only activin–NODAL-sensitive peaks were considered. **i**, Left, activin–NODAL-sensitive m$^6$A exon peaks were evaluated for direct overlap with SMAD2/3-binding sites as indicated by ChIP–seq[30]. $n = 482$ peaks; FDR = 0.41 as calculated by the permutation test implemented by the GAT python package; N.S., not significant based on 95% confidence interval. Right, overlap was calculated after the same features were mapped to their corresponding transcripts or genes, respectively. $n = 372$ genes; hypergeometric test $P = 2.88 \times 10^{-18}$, significant based on 95% confidence interval. **j**, m$^6$A NeMeRIP–seq results for selected transcripts ($n = 3$ cultures; replicates combined for visualization). Coverage tracks represent read enrichments normalized by million mapped reads and size of the library. Blue, sequencing results of m$^6$A NeMeRIP; orange, sequencing results of pre-NeMeRIP input RNA (negative control). GENCODE gene annotations are shown (red, protein coding exons; white, untranslated exons; all potential exons are shown and overlaid). The location of SMAD2/3 ChIP–seq-binding sites is also shown. Compared to the other genes shown, m$^6$A levels on *SOX2* were unaffected by inhibition of activin–NODAL signalling, showing specificity of action. *POU5F1* (also known as *OCT4*) is used as a negative control since it is known to not have a m$^6$A site[23], as confirmed by the lack of m$^6$A enrichment compared to the input.

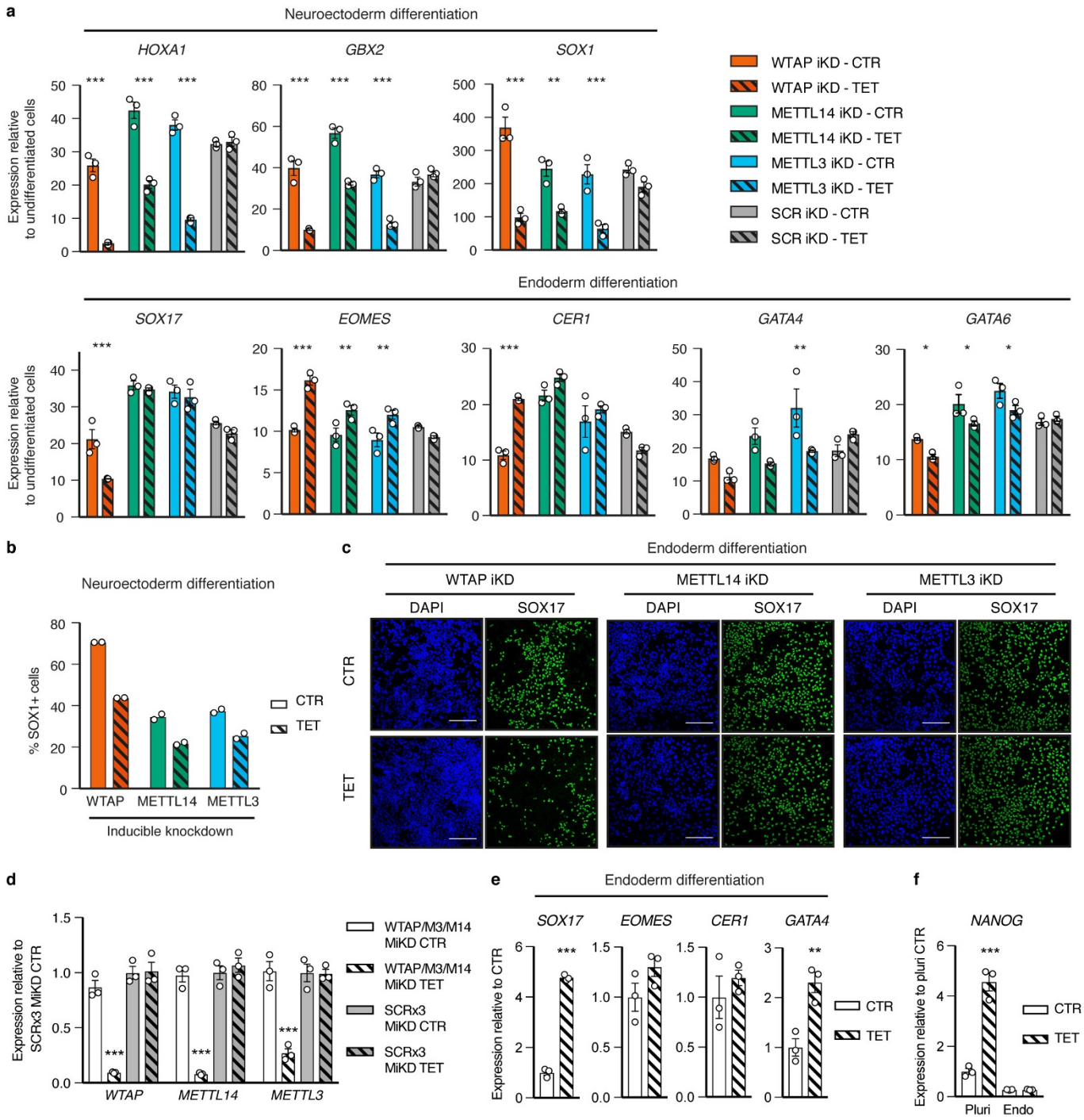



**Extended Data Figure 6** | See next page for caption.

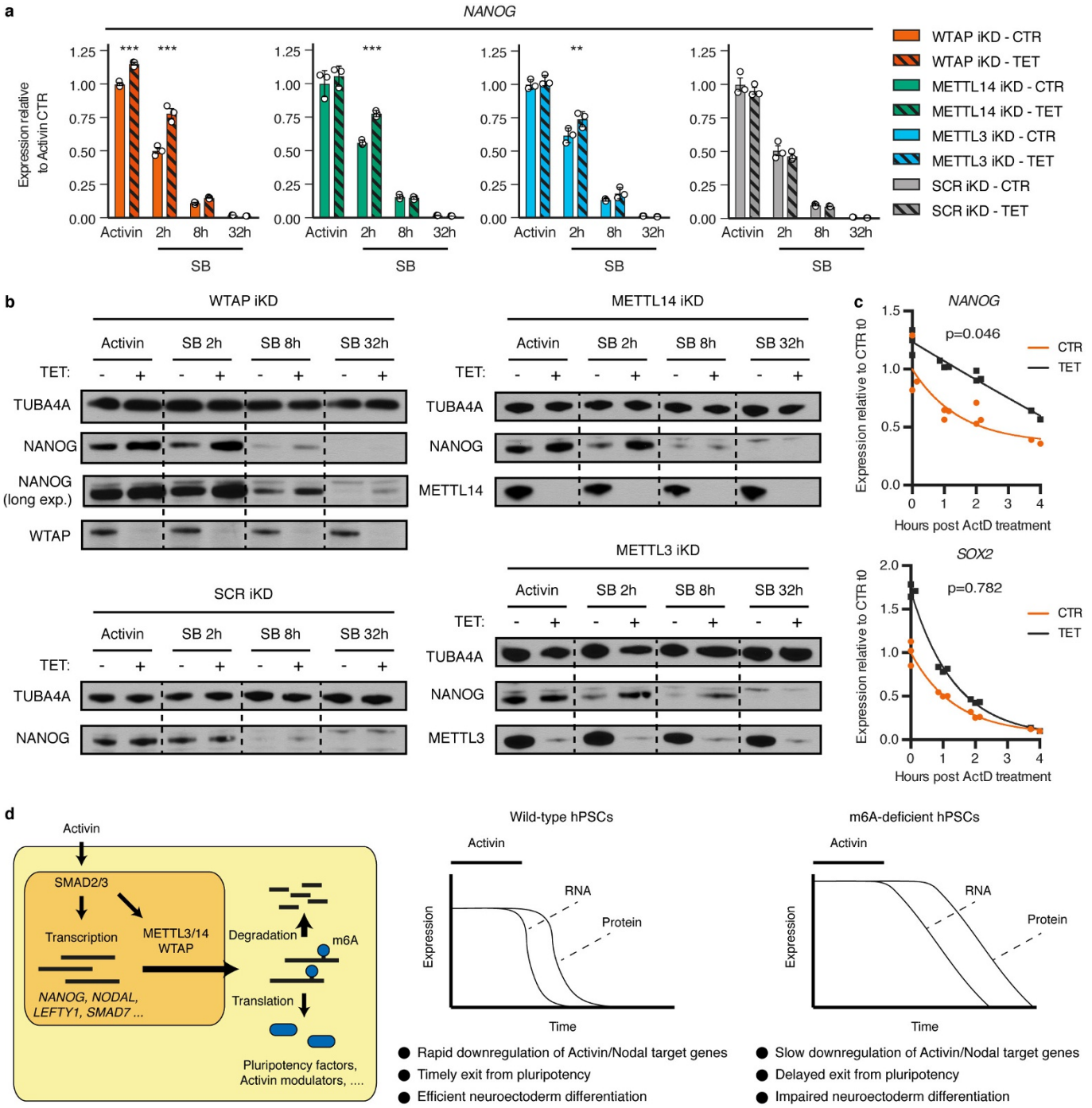
**Extended Data Figure 6 | Features of activin–NODAL-sensitive differential m$^6$A deposition. a**, Scatter plot of the average $\log_2$ fold-change in SB431542 versus activin-treated hESCs for m$^6$A NeMeRIP–seq and pre-NeMeRIP input RNA ($n = 3$ cultures). The analysis was performed for all m$^6$A exon peaks (left), or for those peaks that were significantly downregulated following inhibition of activin–NODAL signalling (right). Data are colour coded according to the square of the difference between the two values (square diff.). **b**, **c**, As Extended Data Fig. 5j, but for representative transcripts that are stably expressed following inhibition of activin–NODAL signalling for 2 h ($n = 3$ cultures; replicates combined for visualization). The m$^6$A NeMeRIP and input tracks were separated and are shown at different scales to facilitate comparison between the conditions. The m$^6$A peaks and those significantly downregulated after 2 h of SB431542 treatment are indicated. **d**, Venn diagram illustrating the strategy for identification of m$^6$A peaks on introns. Peaks mapping to the transcribed genome were obtained by running MetDiff using an extended transcriptome annotation based on the pre-NeMeRIP input RNA, which has a high abundance of introns. The resulting peaks were first filtered by overlap with genome-wide MACS2-identified peaks, and then by lack of overlap with annotated exons. **e**, Results of MetDiff differential methylation analysis in activin versus 2 h SB431542 treatment for m$^6$A peaks on introns. $n = 3$ cultures; $P$ value calculated by likelihood ratio test implemented in the MetDiff R package, and adjusted to FDR by Benjamini–Hochberg correction. See Supplementary Table 2 for the FDR of individual peaks. Abs.FC, absolute fold-change. **f**, As Extended Data Fig. 5j, but for a representative transcript that shows activin–NODAL-sensitive m$^6$A deposition in introns ($n = 3$ cultures; replicates combined for visualization). The m$^6$A peaks on exons, introns, and those significantly downregulated after SB431542 treatment within each subset are indicated. **g**, Plots of RPKM-normalized mean m$^6$A coverage for m$^6$A exon peaks significantly downregulated after SB431542 treatment (absolute fold-change $> 1.5$). Data for all such peaks is in blue, whereas green lines report coverage for only those peaks characterized by next generation sequencing reads that span exon–intron junctions. Exons were scaled proportionally, and the positions of the 3′ and 5′ splice sites (SS) are indicated. A window of 500 bp on either side of the splice sites is shown. m$^6$A, signal from m$^6$A NeMeRIP–seq; input, signal from pre-NeMeRIP input RNA. The results show that coverage of activin–NODAL-sensitive m$^6$A peaks often spans across splice sites (highlighted by the dotted lines). **h**, Heat map representing in an extended form the data shown in **g** for all activin–NODAL-sensitive m$^6$A exon peaks in hESCs cultured in the presence of activin. There are multiple regions in which sequencing coverage extends across exon–intron junctions (see Supplementary Table 2). **i**, Example of an activin–NODAL-sensitive peak located in the proximity of a 3′ splice site ($n = 3$ cultures; replicates combined for visualization). This peak is shown within its genomic context in **c**, where it is indicated by a dotted box. Top, m$^6$A NeMeRIP–seq coverage; bottom, individual next generation sequencing reads. Multiple reads span the exon–intron junction (indicated by the dashed line). **j**, Relationship between the decrease of m$^6$A on the most affected exonic peak located on a transcript ($y$ axis) and the mean change of all other peaks mapping to the same transcript ($x$ axis). The analysis considered transcripts with multiple m$^6$A peaks and with at least one peak significantly decreasing after inhibition of activin–NODAL signalling with SB431542 (absolute fold-change $> 1.5$). Sensitivity of m$^6$A deposition to activin–NODAL signalling across these transcripts is correlated.



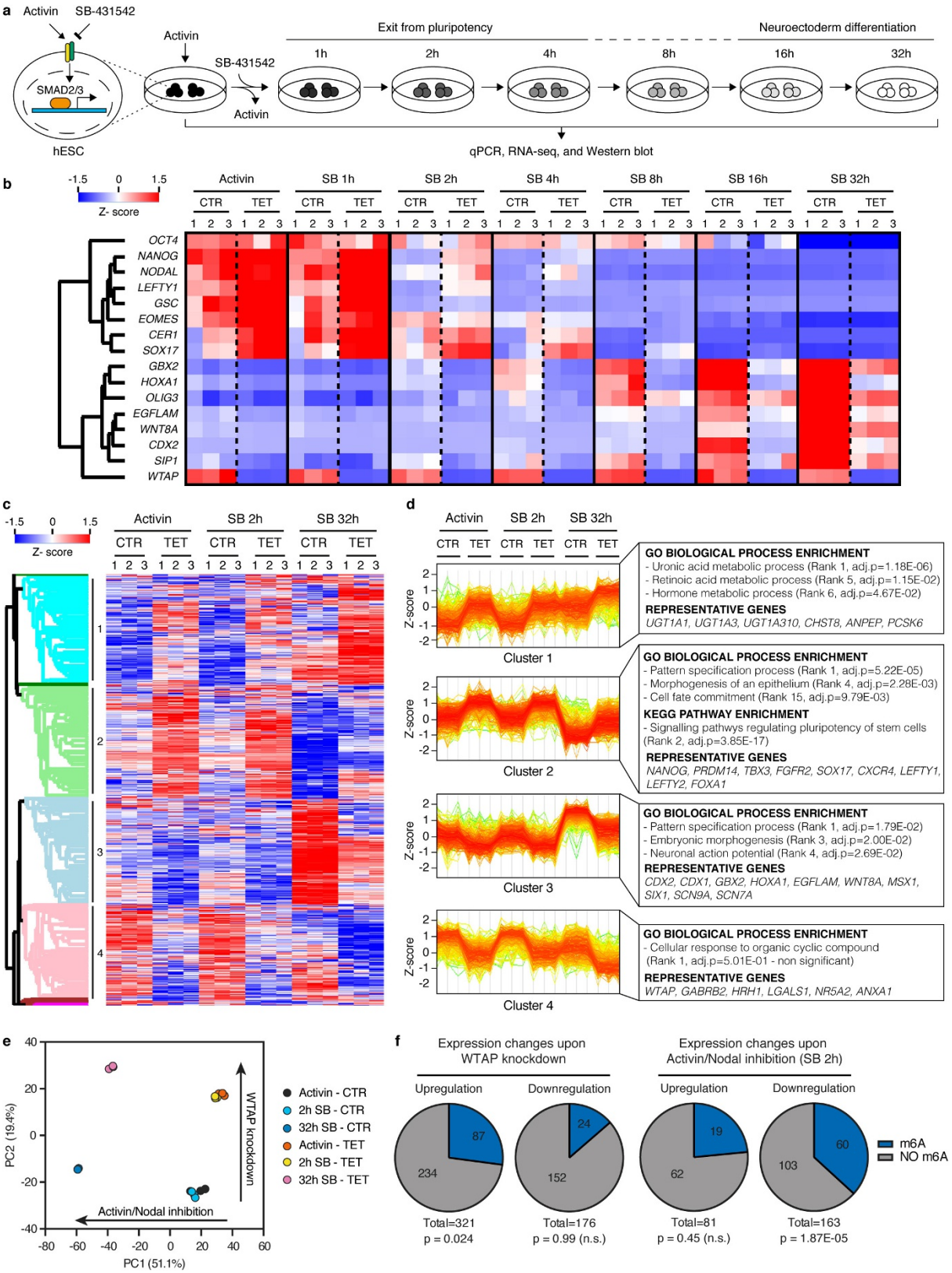

**Extended Data Figure 7 | Generation and functional characterization of hPSCs following iKD of the subunits of the m$^6$A methyltransferase complex. a**, qPCR validation of iKD hESCs cultured in the presence of tetracycline for five days (TET) to drive gene knockdown. Two distinct shRNAs and multiple clonal sublines (cl) were tested for each gene. Expression is normalized to the mean level in hESCs carrying a negative control scrambled (SCR) shRNA. For each gene, sh1 cl1 was selected for further analyses. The mean is indicated, $n = 2$ cultures. **b**, Western blot validation of selected iKD hESCs for the indicated genes. TUB4A4 (α-tubulin), loading control. Results are representative of three independent experiments. **c**, MeRIP–qPCR in iKD hESCs cultured for ten days in the absence (CTR) or presence of tetracycline (TET). m$^6$A abundance is shown relative to control conditions in the same hESC line. The mean is shown, $n = 2$ technical replicates. Results are representative of two independent experiments. **d**, m$^6$A dot blot in WTAP or scramble shRNA control iKD hESCs treated as described in **c**. Decreasing amounts of mRNA were spotted to allow semiquantitative comparisons, as indicated. Results are representative of two independent experiments. **e**, Immunofluorescence of the pluripotency markers NANOG and OCT4 in iKD hESCs cultured for three passages (15 days) in the absence (CTR) or presence of tetracycline (TET). Scale bars, 100 μm. Results are representative of two independent experiments. **f**, Flow cytometry showing NANOG expression in cells treated as in **e**. The percentage and median fluorescence intensity (MFI) of NANOG-positive cells (NANOG$^+$) are shown. The gates used for the analysis are indicated, and were determined on the basis of a secondary-antibody-only negative staining (NEG). Results are representative of two independent experiments.




**Extended Data Figure 8 | Function of the m$^6$A methyltransferase complex during germ-layer specification. a**, qPCR analysis following neuroectoderm or endoderm differentiation of iKD hESCs cultured in absence (CTR) or presence of tetracycline (TET). Tetracycline treatment was initiated in undifferentiated hESCs for ten days and was maintained during differentiation (three days). Expression was normalized against the mean level in undifferentiated hESCs. Mean ± s.e.m., $n = 3$ cultures. Significant differences versus the same iKD line in control conditions were calculated by two-way ANOVA with post hoc Holm–Sidak comparisons. **b**, Flow cytometry quantification of the percentage of SOX1-positive cells (SOX1$^+$) in cells treated as in **a**. Mean is shown, $n = 2$ cultures. **c**, Immunofluorescence of the lineage marker SOX17 in endoderm-differentiated hESCs treated as in **a**. Scale bars, 100μm. Results are representative of two independent experiments. **d**, qPCR validation of multiple inducible knockdown (MiKD) hESCs simultaneously expressing shRNAs against WTAP, METTL3 and METTL14. Cells expressing three copies of the scrambled shRNA (SCR3×) were used as negative control. Cells were cultured in the presence of tetracycline (TET) for five days to drive gene knockdown. Mean ± s.e.m., $n = 3$ cultures. Significant differences versus SCR3× hESCs in control conditions were calculated by two-way ANOVA with post hoc Holm–Sidak comparisons. **e**, **f**, qPCR analysis following endoderm differentiation of WTAP, METTL3 and METTL14-MiKD hESCs treated as described in **a**. Mean ± s.e.m., $n = 3$ cultures. Significant differences versus control conditions were calculated by two tailed $t$-test (**e**) or two-way ANOVA with post hoc Holm–Sidak comparisons (**f**).




**Extended Data Figure 9 | Function of the m$^6$A methyltransferase complex during exit from pluripotency induced by inhibition of activin–NODAL signalling. a**, qPCR analyses in iKD hESCs cultured in absence (CTR) or presence (TET) of tetracycline for ten days, then subjected to inhibition of activin–NODAL signalling with SB431542 (SB) for the indicated time (see Extended Data Fig. 10a). Activin, cells maintained in standard pluripotency-promoting culture conditions containing activin and collected at the beginning of the experiment. Mean ± s.e.m., $n = 3$ cultures. Significant differences versus same iKD line in control conditions were calculated by two-way ANOVA with post hoc Holm–Sidak comparisons. **b**, Western blots of cells treated as described in **a**. TUBA4A, loading control. Results are representative of two independent experiments. **c**, Measurement of mRNA stability in WTAP iKD hESCs cultured in absence (CTR) or presence (TET) of tetracycline for ten days. Samples were collected following transcriptional inhibition using actinomycin D (ActD) for the indicated time. The statistical significance of differences between the mRNA half lives in tetracycline versus control is shown ($n = 3$ cultures, comparison of fits to one-phase decay model by extra sum-of-squares $F$-test). The difference was significant for *NANOG* but not for *SOX2* (95% confidence interval). **d**, Model showing the interplays between activin–NODAL signalling and m$^6$A deposition in hPSCs (left), and the phenotype induced by impairment of the m$^6$A methyltransferase complex (right).




**Extended Data Figure 10** | See next page for caption.

**Extended Data Figure 10 | Genome-wide analysis of the relationship between WTAP and activin–NODAL signalling. a**, Schematic of the experimental approach to investigate the transcriptional changes induced by the knockdown of the m$^6$A methyltransferase complex subunits during neuroectoderm specification of hESCs. **b**, qPCR analyses of WTAP iKD hESCs subjected to the experiment outlined in **a** ($n = 3$ cultures). Activin, cells maintained in standard pluripotency-promoting culture conditions containing activin and collected at the beginning of the experiment. *Z*-scores indicate differential expression measured in number of standard deviations from the mean across all time points. **c**, RNA-seq analysis at selected time points from the samples shown in panel **b** ($n = 3$ cultures). The heat map shows *Z*-scores for the top 5% differentially expressed genes (1789 genes as ranked by the Hotelling $T^2$ statistic). Genes and samples were clustered based on their Euclidean distance, and the four major gene clusters are indicated (see Supplementary Discussion). **d**, Expression profiles of genes belonging to the clusters indicated in **c**. Selected results of gene-enrichment analysis and representative genes for each cluster are shown (cluster 1: $n = 456$ genes; cluster 2: $n = 471$ genes; cluster 3: $n = 442$ genes; cluster 4: $n = 392$ genes; Fisher's exact test followed by Benjamini–Hochberg correction for multiple comparisons). **e**, Principal component analysis of RNA-seq results in **c** ($n = 3$ cultures). The top 5% differentially expressed genes were considered for this analysis. For each of the two main principal components (PC1 and PC2), the fraction of inter-sample variance that they explain and their proposed biological meaning are reported. **f**, Proportion of transcripts marked by at least one high-confidence m$^6$A peak[23] in transcripts significantly up- or downregulated following WTAP iKD in hESCs maintained in the presence of activin (left), or following inhibition of activin–NODAL signalling for 2 h with SB431542 in control cells (right). Differential gene expression was calculated in three cultures using the negative binomial test implemented in DEseq2 with a cutoff of $P < 0.05$ and abs.FC $> 2$. The number of genes in each group and the hypergeometric probabilities of the observed overlaps with m$^6$A-marked transcripts are reported (n.s.: not significant at 95% confidence interval).

nature research

Corresponding author(s): Ludovic Vallier

☐ Initial submission ☐ Revised version ☒ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

1. Sample size

   Describe how sample size was determined.

   No power calculations were performed. Sample size was determined depending on the experiment type based on what is standard practice in the field of pluripotent stem cell biology to statically examine a large effect within an in vitro system which experiences only limited biological variability (n=2-4, see Figure Legends).

2. Data exclusions

   Describe any data exclusions.

   No data points were excluded from any of the analyses presented.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   All of the presented experiments were successfully reproduced.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   No randomization was performed since this was not relevant to the study: all treatment/control experiments were performed on the same starting cell population (no covariates).

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinding was performed. This was deemed unnecessary since none of the analyses reported involved procedures that could be influenced by investigator bias (such as manual counting/measuring and/or morphological assessments). Indeed, all analyses presented involved automated processing of data through experimental instrumentation and/or computing procedures (including counting of PLA signals presented in Fig. 2d, see the Methods)

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

All of the software used for the analyses presented is described in detail in the relevant Methods, which also describes the relevant parameters used when these were not the default. The software used was: MaxQuant, Perseus, Cytoscape, Enrichr, ImageJ, R/Bioconductor, Trimmomatic, TopHat 2.0.13, MetDiff, fCCAC v1.0.0, RSeQC-2.6, RCAS, DREME, DESeq2, Cufflinks, GraphPad Prism 6, Sickle, Samtools view, SeqMonk, timecourse (Bioconductor). All bioinformatic scripts have been deposited to GitHub (http://github.com/pmb59/neMeRIP-seq).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All unique materials (cell lines and plasmids) are readily available from the authors

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

All of the antibodies used are detailed in Supplementary Table 5, which also reports the application for which they were used and the amount/dilution. The list includes: anti-CREBBP (Cell Signalling 7389); anti-EP300 (Santa Cruz sc-584); anti-FOXH1 (R&D BAF4248); anti-METTL14 (Sigma-Aldrich HPA038002); anti-METTL3 (Proteintech 15073-1-AP, and Bethyl Lab A301-567A); anti-NANOG (R&D AF1997, and Abcam ab21624); anti-OCT4/POU5F1 (Santa Cruz sc-5279); anti-SETDB1 (Cell Signalling 2196); anti-SMAD2/3 (R&D AF3797, and Cell Signalling 12470S); anti-SNON/SKIL (Santa Cruz sc-9595x); anti-SOX1 (R&D AF3369); anti-SOX17 (R&D AF1924); anti-TUB4A4 (Sigma-Aldrich T6199); anti-WTAP (Bethyl Lab A301-436A). All of the antibodies were validated to recognize the relevant human protein, and most of them were specifically validated for the relevant application, as specified on the relevant catalog pages on the suppliers' websites. In the few cases in which the antibody was not previously validated for a specific application, extensive testing in house with the appropriate negative controls was performed to confirm the specificity (see Methods and Figure legends). Such applications were: anti-METTL14 (IP and ChIP); anti-METTL3 (ChIP); anti-NANOG (flow cytometry and PLA); anti-SMAD2/3 (RIP and PLA); anti-SOX1 (flow-cytometry); anti-WTAP (PLA, RIP, and ChIP).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

The H9 hESC line was obtained from WiCell (Madison, Wisconsin). The A1ATR/R hiPSC line was obtained in house and previously described in Yusa et al 2011.

Yusa, K. et al. Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature 478, 391-4 (2011).

b. Describe the method of cell line authentication used.

No authentication was performed on the H9 hESCs as they were used directly from the commercial supplier. The A1ATR/R was genotyped in house to confirm the presence of A1AT R/R allele. Both cell lines were routinely karyotyped by standard G-banding to confirm euploidy.

c. Report whether the cell lines were tested for mycoplasma contamination.

Yes, mycoplasma screening was performed every month.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.

nature research

Corresponding author(s): Ludovic Vallier

☐ Initial submission ☐ Revised version ☒ Final submission

# Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

## Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

## Methodological details

| | |
|---|---|
| 5. Describe the sample preparation. | Biological samples: H9 hESCs (Extended Data Figure 7f), or hESC-derived neuroectoderm (Extended Data Figure 8b). Single cell suspensions were prepared by incubation in cell cell dissociation buffer (CDB; Gibco) for 10' at 37° followed by extensive pipetting. Cells were washed twice with PBS and fixed for 20' at 4°C with PBS 4% PFA. After three washes with PBS, cells were first permeabilized for 20' at RT with PBS 0.1% Triton X-100, then blocked for 30' at RT with PBS 10% donkey serum. Primary and secondary antibodies incubations (Supplementary Table 5) were performed for 1h each at RT in PBS 1% donkey serum 0.1% Triton X-100, and cells were washed three times with this same buffer after each incubation. |
| 6. Identify the instrument used for data collection. | Cyan ADP flow cytometer (Beckman Coulter) |
| 7. Describe the software used to collect and analyze the flow cytometry data. | Data collection: Summit software (Beckman Coulter). Data analysis: FlowJo X |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | No cell sorting was performed in the study |
| 9. Describe the gating strategy used. | Cells were first gated on the basis of forward and side scatter properties, after which singlets were isolated on the basis of relationship between side scatter area peak area and width. A secondary-only negative control was used to determine the background fluorescence, and positive cells were quantified by setting a boundary so that less than 1% of the secondary-only control cells would be considered positive. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒