# 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 Sadee[4], An-Sofie Lenaerts[1], Shota Nakanoh[1], Rodrigo Grandy[1], Edward Farnell[5], Jernej Ule[4], Hendrik G. Stunnenberg[3], Sasha Mendjan[1,‡], and Ludovic Vallier[1,2,#].

[1] Wellcome Trust - MRC Cambridge Stem Cell Institute Anne McLaren Laboratory and Department of Surgery, University of Cambridge, UK.

[2] Wellcome Trust Sanger Institute, Hinxton UK.

[3] Department of Molecular Biology, Radboud University Nijmegen, The Netherlands.

[4] Francis Crick Institute and Department of Molecular Neuroscience, University College London, UK.

[5] Department of Pathology, University of Cambridge, UK.

[*] These authors contributed equally to this work.

[†] Current address: Department of Pathology, University of Washington, Seattle, WA, USA.

[‡] Current address: Institute of Molecular Biotechnology, Vienna, Austria.

[#] Corresponding author (lv225@cam.ac.uk).

**The TGFβ pathway plays an essential role in embryonic development, organ homeostasis, tissue repair, and disease[1,2]. This diversity of tasks is achieved through the intracellular effector SMAD2/3, whose canonical function is to control activity of target genes by interacting with transcriptional regulators[3]. Nevertheless, a complete description of the factors interacting with SMAD2/3 in any given cell type is still lacking. Here we address this limitation by describing the interactome of SMAD2/3 in human pluripotent stem cells (hPSCs). 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 deposits $N^6$-methyladenosine (m6A)[4]. We uncover that SMAD2/3 promotes binding of the m6A methyltransferase complex onto a subset of transcripts involved in early cell fate decisions. This mechanism destabilizes specific SMAD2/3 transcriptional targets, including the pluripotency factor *NANOG,* thereby poising 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 regulations of the epitranscriptome. These novel aspects of TGFβ signalling could have far-reaching implications in many other cell types and in diseases such as cancer[5].**

## Main

Activin and Nodal, two members of the TGFβ superfamily, play essential roles in cell fate decision in hPSCs[6–8]. Activin/Nodal signalling is necessary to maintain pluripotency, and its inhibition drives differentiation toward the neuroectoderm lineage[6,9,10]. Activin/Nodal also cooperates with BMP and WNT to drive mesendoderm specification[11–14]. Thus, we used hPSC differentiation into definitive endoderm as a model system to interrogate the SMAD2/3 interactome during a dynamic cellular process. For that we developed an optimized SMAD2/3 co-immunoprecipitation (co-IP) protocol compatible with mass-spectrometry analyses (Extended Data Fig. 1a-b and Supplementary Discussion). This method allowed a comprehensive and unbiased examination of the proteins interacting with SMAD2/3 for the first time in any given cell type. By examining human embryonic stem cells (hESCs) and hESCs induced to differentiate towards endoderm (Fig. 1a), we identified 89 SMAD2/3 partners (Fig. 1b, Extended Data Fig. 1c-d, and Supplementary Table 1). Of these, only 11 factors were not shared between hESCs and endoderm differentiating cells (Extended Data Fig. 1e), suggesting that the SMAD2/3 interactome is largely conserved across these two lineages (Supplementary Discussion). Importantly, this list included known SMAD2/3 transcriptional and epigenetic cofactors (including FOXH1, SMAD4, SNON, SKI, EP300, SETDB1, and CREBBP[3]), which validated our method. Furthermore, we performed functional experiments on FOXH1, EP300, CREBBP, and SETDB1, which uncovered the essential function of these SMAD2/3 transcriptional and epigenetic cofactors in hPSC fate decisions (Extended Data Fig. 2 and 3, and Supplementary Discussion).

Interestingly, our proteomic experiments also revealed that SMAD2/3 interacts with complexes involved in functions that have never been associated with TGFβ signalling (Fig.

1b and Extended Data Fig. 1f), such as ERCC1-XPF (DNA repair) and DAPK3-PAWR (apoptosis). Most notably, we identified several factors involved in mRNA processing, modification, and degradation (Fig. 1b), such as the METTL3-METTL14-WTAP complex (deposition of $N^6$-methyladenosine, or m6A), the PABP-dependent poly(A) nuclease complex hPAN (mRNA decay), the cleavage factor complex CFIm (pre-mRNA 3' end processing), and the NONO-SFPQ-PSPC1 factors (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, which include not only transcriptional and epigenetic regulations, but also novel "non-canonical" molecular functions.

To further explore this hypothesis, we investigated the interplays between Activin/Nodal and m6A deposition. m6A is the most common RNA modification, regulating multiple aspects of mRNA biology including decay and translation[4,15–19]. However, whether this is a dynamic event that can be modulated by extracellular cues remains to be established. Furthermore, while m6A is known to regulate hematopoietic stem cells[20,21] and the transition between the naïve and primed pluripotency states[22,23], its function in hPSCs and during germ layer specification is unclear. We first validated the interaction of SMAD2/3 with METTL3-METTL14-WTAP using co-IP followed by Western Blot in both hESCs and human induced pluripotent stem cells (hiPSCs; Fig. 2a and Extended Data Fig. 4a-b). Interestingly, 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 at the nuclear level (Fig. 2c-d). These observations suggest that SMAD2/3 and the m6A methyltransferase complex interact in an Activin/Nodal signalling-dependent fashion.

To investigate the functional relevance of this interaction, we assessed the transcriptome-wide effects of Activin/Nodal inhibition on the deposition of m6A by performing nuclear-enriched m6A methylated RNA immunoprecipitation followed by deep sequencing (NeMeRIP-seq; Extended Data Fig. 5a-d, and Supplementary Discussion). In agreement with previous reports[17,19,24], deposition of m6A onto exons was enriched around stop codons and transcription start sites, and occurred at a motif corresponding to the m6A consensus sequence (Extended Data Fig. 5e-g). Assessment of differential m6A deposition revealed that Activin/Nodal inhibition predominantly resulted in reduced m6A levels in selected transcripts (Supplementary Table 2; average absolute $\log_2$ fold-change of 0.56 and 0.35 for m6A decrease and increase, respectively). Decrease in m6A deposition was predominantly observed on peaks located near to stop codons (Extended Data Fig. 5h), a location which has been reported to decrease the stability of mRNAs[16,24,25]. Interestingly, transcripts showing reduced m6A levels after Activin/Nodal inhibition largely and significantly overlapped with genes bound by SMAD2/3 (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 m6A deposition was largely associated with transcripts rapidly decreasing during the exit from pluripotency triggered by Activin/Nodal inhibition (Extended Data Fig. 6a). Transcripts behaving in this fashion were enriched in pluripotency regulators and in 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 to Activin/Nodal-sensitive m6A deposition remained unchanged following Activin/Nodal inhibition (Extended Data Fig. 6a-c and Supplementary Table 3). Considered together, these findings establish that Activin/Nodal signalling can regulate m6A deposition on a number of specific transcripts.

We then examined the underlying molecular mechanisms. RNA immunoprecipitation experiments on nuclear RNAs showed that inhibition of Activin/Nodal signalling impaired binding of WTAP to multiple m6A-marked transcripts including *NANOG* and *LEFTY1* (Fig. 2f and Extended Data Fig. 4d-e), while SMAD2/3 itself interacted with such 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 m6A methyltransferase complex onto nuclear RNAs. Interestingly, recent reports have established that m6A deposition occurs co-transcriptionally and involves nascent pre-RNAs[16,26,20]. Considering the broad overlap between SMAD2/3 transcriptional targets and transcripts showing Activin/Nodal-sensitive m6A deposition (Extended Data Fig 5i), we therefore hypothesized that SMAD2/3 could facilitate co-transcriptional recruitment of the m6A methyltransferase complex onto nascent transcripts. Supporting this notion, inhibition of Activin/Nodal signalling mainly resulted in downregulation of m6A not only on exons, but also onto 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 for m6A peaks within the same transcript (Extended Data Fig. 6j), suggesting that SMAD2/3 regulates m6A deposition at the level of a genomic locus rather than on a specific mRNA peak. Nevertheless, a stable and direct binding of the m6A methyltransferase complex to the DNA could not be detected (Extended Data Fig. 4f). Thus, co-transcriptional recruitment might rely on indirect and dynamic interactions with the chromatin. Considering all these results, we propose a model in which Activin/Nodal signalling promotes co-transcriptional m6A deposition by facilitating the recruitment of the m6A methyltransferase complex onto nascent mRNAs (Fig. 2h).

To understand the functional relevance of these regulations in the context of hPSC cell fate decisions, we performed inducible knockdown experiments for the various subunits of the

m6A methyltransferase complex[27] (Extended Data Fig. 7a-b). As expected, decrease in WTAP, METTL14, or METTL3 expression reduced the deposition of m6A (Extended Data Fig 7c-d). Interestingly, prolonged knockdown did not affect pluripotency (Extended Data Fig. 7e-f). However, expression of m6A methyltransferase complex subunits was necessary for neuroectoderm differentiation induced by the inhibition of Activin/Nodal signalling, while it was dispensable for Activin-driven endoderm specification (Fig. 3a and Extended Data Fig. 8a-c). Activin/Nodal is known to block neuroectoderm induction by promoting NANOG expression[28], while NANOG is required for the early stages of endoderm specification[13]. Therefore, we monitored the levels of this factor during neuroectoderm differentiation. We observed that both transcript and protein were upregulated following impairment of m6A methyltransferase activity (Fig. 3b and Extended Data Fig. 9a-b), while mRNA stability was increased (Extended Data Fig. 9c). These results show that m6A deposition decreases the stability of the *NANOG* mRNA to facilitate its downregulation upon loss of Activin/Nodal signalling, thus facilitating exit from pluripotency and neuroectoderm specification (Extended Data Fig. 9d). Additional transcriptomic analyses showed that WTAP knockdown resulted in a global upregulation of genes transcriptionally activated by SMAD2/3 in hESCs, while it impaired the upregulation of genes induced by Activin/Nodal inhibition during neuroectoderm differentiation (Fig. 3b, Extended Data Fig. 10a-e, Supplementary Table 4, and Supplementary Discussion). Importantly, the decrease in WTAP expression also led to the upregulation of mRNAs marked by m6A (Extended Data Fig. 10f), confirming that WTAP-dependent m6A deposition destabilises mRNAs[16,24,25]. Moreover, transcripts rapidly downregulated after Activin/Nodal inhibition were enriched in m6A-marked mRNAs (Extended Data Fig. 10f). Finally, simultaneous knockdown of METTL3, METTL14, and WTAP in hESCs resulted in an even stronger dysregulation of Activin/Nodal target transcripts (Fig. 3c-d and Extended Data Fig. 8d) and defective neuroectoderm differentiation (Fig. 3d

and Extended Data Fig. 8e-f). Taken together, these results indicate that the interaction of SMAD2/3 with METTL3-METTL14-WTAP can promote m6a 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).

To conclude, this first analysis of the SMAD2/3 interactome reveals novel interplays between TGFβ signalling and a diversity 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 possibility is illustrated by our results regarding Activin/Nodal-sensitive regulation of m6A. Indeed, through the interaction between SMAD2/3 and the METTL3-METTL14-WTAP complex, Activin/Nodal signalling connects transcriptional and epitranscriptional regulations to "poise" several of its transcriptional targets for rapid degradation upon signalling withdrawal (Extended Data Fig. 9d). As a result, 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 these are interrelated with chromatin epigenetic, transcriptional, and epitranscriptional regulations.

Our findings also clarify and substantially broaden our understanding of the function of m6A in cell fate decisions. They establish that depletion of m6A in hPSCs does not lead to differentiation, contrary to predictions from studies in mouse epiblast stem cells[22]. This could

imply important functional differences in epitranscriptional regulations between the human and murine pluripotent state. Moreover, widening the conclusions from previous reports[23], we demonstrate that deposition of m6A 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 actually necessary for the early stages of mesendoderm specification[13,29]. Finally, our results establish that m6A is a dynamic event directly modulated by extracellular clues such as TGFβ. Considering the broad importance of TGFβ signalling, the regulation we describe here might have an essential function in many cellular contexts requiring a rapid response or change in cell state, such as the inflammatory response or cellular proliferation.

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

**Acknowledgments**

We thank Cambridge Genomic Services for help in next generation sequencing. The 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-IP, NeMeRIP, and RNA-IP experiments, and analysed data. P.M., I.R.d.l.M, and C.S. analysed NeMeRIP-seq. A.O. performed PLA and co-immunoprecipitations, and analyzed RNA-seq. D.O., L.Y., and J.K. assisted hPSC gene editing and differentiation; N.C.H. performed quantitative proteomics and data analysis. A.L., S.N., and R.G. assisted hPSC culture. E.F. optimized NeMeRIP-seq sequencing 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 SMAD2/3 co-IP. L.V. conceived, supervised, and supported the study, wrote and provided final approval of the manuscript.

**Author information**

Reprints and permissions information is available at www.nature.com/reprints. No competing financial interests are declared. Correspondence and requests for materials should be addressed to Ludovic Vallier (lv225@cam.ac.uk).

## References

1. Wu, M. Y. & Hill, C. S. Tgf-beta superfamily signaling in embryonic development and homeostasis. *Dev. Cell* **16,** 329–43 (2009).
2. Oshimori, N. & Fuchs, E. The harmonies played by TGF-β in stem cell biology. *Cell Stem Cell* **11,** 751–64 (2012).
3. Gaarenstroom, T. & Hill, C. S. TGF-β signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation. *Semin. Cell Dev. Biol.* **32,** 107–8 (2014).
4. Heyn, H. & Esteller, M. An Adenine Code for DNA: A Second Life for N6-

Methyladenine. *Cell* **7,** 710–3 (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,** 403–21 (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,** 4495–509 (2005).

8. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132,** 1273–82 (2005).

9. Smith, J. R. *et al.* Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* **313,** 107–17 (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,** 702–17 (2015).

11. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23,** 1534–41 (2005).

12. Vallier, L. *et al.* Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem Cells* **27,** 2655–66 (2009).

13. Teo, A. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **2,** 238–250 (2011).

14. Kubo, A. *et al.* Development of definitive endoderm from embryonic stem cells in culture. *Development* **131,** 1651–62 (2004).

15. Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3′ UTR regulation. *Genes Dev.* **29,** 2037–53 (2015).

16. Ke, S. *et al.* m6A 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 m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–6 (2012).

18. Meyer, K. D. *et al.* 5' UTR m6A 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,** 1635–46 (2012).

20. Barbieri, I. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* **552,** 126-131 (2017).

21. Vu, L. P. *et al.* The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23,** 1369–1376 (2017).

22. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347,** 1002–6 (2015).

23. Batista, P. J. *et al.* m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* **15,** 707–19 (2014).

24. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **8,** 284–96 (2014).

25. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–20 (2014).

26. Bartosovic, M. *et al.* N6-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,** 4405–18 (2016).

28. Vallier, L. *et al.* Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136,** 1339–49 (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,** 1176–85 (2011).

## Figure legends

**Figure 1. Identification of the SMAD2/3 interactome.**

(**a**) Experimental approach. (**b**) Interaction network from all known protein-protein interactions between selected SMAD2/3 partners identified in pluripotent and endoderm cells (n=3 co-IPs; one-tailed t-test: permutation-based FDR<0.05). Nodes describe: (1) the lineage in which the proteins were significantly enriched (shape); (2) significance of the enrichment (size is proportional to the maximum -log p-value); (3) function of the factors (colour). Complexes of interest are marked.

**Figure 2. Activin/Nodal signalling promotes m6A deposition on specific regulators of pluripotency and differentiation.**

(**a-b**) Western blots of SMAD2/3 (S2/3), METTL3 (M3), or control (IgG) immunoprecipitations (IPs) from nuclear extracts of hESCs (representative of three experiments). Input is 5% of the material used for IP. In b, IPs were performed from hESCs maintained in presence of Activin or treated for 1h with SB-431542 (SB; Activin/Nodal inhibitor). For gel source data, see Supplementary Figure 1. (**c**) Proximity ligation assays (PLA) for SMAD2/3 and WTAP in hESCs maintained in presence of Activin or SB (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 ± SEM, n=4 PLA. 2-way ANOVA with post-hoc Holm-Sidak comparisons: **=$p<0.01$, and ***=$p<0.001$. (**e**) Representative results of nuclear-enriched m6A methylated RNA immunoprecipitation followed by deep-sequencing (m6A NeMeRIP-seq; n=3 cultures, replicates combined for visualization). Signal represents read enrichment normalized by 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 ChIP-seq data[30] are shown. (**f-g**) RNA immunoprecipitation (RIP) experiments for WTAP, SMAD2/3, or IgG control in hESCs maintained in presence of Activin or treated with SB. *RPLP0* and *PBGD* were used as negative controls as they present no m6A. f: mean ± SEM, n=3 cultures. 2-way ANOVA with post-hoc Holm-Sidak comparisons: *=$p<0.05$, and **=$p<0.01$. g: mean, n=2 cultures. (**h**) Model for the mechanism by which SMAD2/3 promotes m6A deposition. P: phosphorylation; W: WTAP; M14: METTL14.

**Figure 3. The m6A methyltransferase complex antagonizes Activin/Nodal signalling in hPSCs to promote timely exit from pluripotency.**

(**a**) Immunofluorescence for neural marker SOX1 following neuroectoderm differentiation of tetracycline (TET)-inducible knockdown (iKD) hESCs (representative of two experiments). CTR: no TET; DAPI: nuclei. Scale bars: 100μm. (**b**) qPCR analyses in WTAP iKD hESCs

subjected to Activin/Nodal signalling inhibition with SB for the indicated time. Act: Activin. Mean ± SEM, n=3 cultures. 2-way ANOVA with post-hoc Holm-Sidak comparisons: **=p<0.01, and ***=p<0.001. (**c**) Western blot validation of multiple inducible knockdown (MiKD) hESCs for WTAP, METTL3 (M3), and METTL14 (M14). Cells expressing three copies of the scrambled shRNA (SCR3x) were used as negative control. (**d**) qPCR analyses in undifferentiated MiKD hESCs, or following their neuroectoderm differentiation. Mean ± SEM, n=3 cultures. Two-tailed t-test: **=p<0.01, and ***=p<0.001.

# Methods

### hPSC culture and differentiation

Feeder- and serum-free culture of hESCs (H9/WA09 line; WiCell) and hiPSCs ($A1AT^{R/R}$;[31]) was previously described[32]. Briefly, 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 10ng/ml Activin-A and 12ng/ml FGF2 (both from Dr Marko 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/$cm^2$. Differentiation was initiated in adherent hESC cultures 48h following passaging. Definitive endoderm specification was induced for 3 days (unless stated otherwise) by culturing cells in CDM (without insulin) with 20ng/ml FGF2, 10μM LY294002 (PI3K inhibitor; Promega), 100ng/ml Activin-A, and 10ng/ml BMP4 (R&D), as previously described[33]. Neuroectoderm was induced for 3 days (unless stated otherwise) in CDM-BSA with 12ng/ml FGF2 and 10μM SB-431542 (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

363 abnormalities and mycoplasma infection. Since hESCs were obtained by a commercial
364 supplier cell line identification was not performed. hiPSCs were previously generated *in house*
365 and genotyped by Sanger sequencing[31].
366
367 **Molecular cloning**
368 Plasmids carrying inducible shRNAs were generated by cloning annealed oligonucleotides
369 into the pAAV-Puro_iKD or pAAV-Puro_siKD vectors as previously described[27]. All shRNA
370 sequences were obtained from the RNAi Consortium TRC library[35]
371 (https://www.broadinstitute.org/rnai/public/). Whenever shRNAs had been validated, the most
372 powerful ones were chosen (the sequences are reported in Supplementary Table 5). Generation
373 of a vector containing shRNAs against METTL3, METTL14, and WTAP (cloned in in this
374 order) was performed by Gibson assembly of PCR products containing individual shRNA
375 cassettes, as previously described[27]. The resulting was named pAAV-Puro_MsiKD-
376 M3M14W. Generation of the matched control vector containing three copies of the scrambled
377 shRNA sequence (pAAV-Puro_MsiKD-SCR3x) was previously described[27].
378
379 A targeting vector for the AAVS1 locus carrying constitutively-expressed *NANOG* was
380 generated starting from pAAV_TRE-EGFP[36]. First, the TRE-EGFP cassette was removed
381 using PspXI and EcoRI, and substituted with the CAG promoter (cut from pR26-
382 CAG_EGFP[27] using SpeI and BamHI) by ligating blunt-ended fragments. The resulting vector
383 (pAAV-Puro_CAG) was then used to clone full-length the *NANOG* transcript, which includes
384 its full 5' and 3' UTR. The full-length *NANOG* transcript was constructed from 3 DNA
385 fragments. The 5' (1–301bp) and 3' (1878–2105bp) ends were synthesised (IDT) with 40bp
386 overlaps corresponding to pGem3Z vector linearised with SmaI. The middle fragment was
387 amplified from cDNA of H9 hESCs obtained by retrotranscription with poly-T primer using

primers 5'-TTGTCCCCAAAGCTTGCCTTGCTTT-3'and 5'–CAAAAACGGTAAGAAA-TCAATTAA-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-SCR3x). 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 tetR. Two wells were transfected for each shRNA in order to generate independent biological replicates. Following selection with puromycin, all 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 tetracycline hydrochloride (Sigma-Aldrich) to the culture medium at the concentration of 1μg/ml. Unless indicated owtherwise in the text or figure legends, inducible knockdown in undifferentiated hESCs was induced for 5 days, while differentiation assays were performed in hESCs in which knockdown had been induced for 10 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 $2x10^7$ cells were used for each immunoprecipitation (IP). 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), 1mg/ml Leupeptin, 0.2mM DTT, 0.2mM PMSF, and 10mM sodium butyrate (all from Sigma-Aldrich). Cells were fed with fresh medium for 2h before being washed with PBS, scraped in cell dissociation buffer (CDB, Gibco), and pelleted at 250g for 10'. The cell pellet was then washed once with 10 volumes of PBS, and once with 10 volumes of hypotonic lysis buffer (HLB: 10mM HEPES pH 7.6; 10mM KCl; 2mM $MgCl_2$; 0.2mM EDTA; 0.2mM EGTA). The pellet was resuspended in 5 volumes of HLB and incubated for 5' 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 800g for 5', washed once with 10 volumes of HLB, and resuspended in 1.5 volumes of high-salt nuclear lysis buffer (HSNLB: 20mM HEPES pH 7.6; 420mM NaCl; 2mM $MgCl_2$; 25% glycerol; 0.2mM EDTA; 0.2mM 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' of incubation in rotation. The resulting lysate was clarified for 30' at 16,000g,

and transferred to a dialysis cassette using a 19-gauge syringe. Dialysis was performed for 4h in 1l of dialysis buffer (DB: 20mM HEPES pH 7.6; 50mM KCl; 100mM NaCl; 2mM $MgCl_2$; 10% glycerol; 0.2mM EDTA; 0.2mM EGTA) under gentle stirring, and the buffer was changed once after 2h. After the dialysis, the sample was clarified from minor protein precipitates for 10' at 17,000g, and the protein concentration was assessed. Immunoprecipitations were performed by incubating 0.5mg of protein with 5µg of goat polyclonal SMAD2/3 antibody (R&D systems, catalogue number: AF3797) or goat IgG negative control antibody (R&D systems, catalogue number: AB-108-C) for 3h at 4°C in rotation. This was followed by incubation with 10µl of Protein G-Agarose for 1h. Beads were finally washed three times with DB, and finally 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) was 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 50µl of 2M urea and 10mM DTT for 30' at RT in agitation. Then, 55mM chloroacetamide was added for 20' to alkylate reduced disulphide bonds. Proteins were pre-digested on the beads with 0.4µg of mass spectrometry-quality trypsin (Promega) for 1h at RT in agitation. The suspension was cleared from the beads by centrifugation. The beads were then washed with 50ul of 2M Urea, and the merged supernatants were incubated overnight at RT in 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' at 800g with 50µl methanol, 50µl Solvent B (0.5% acetic acid; 80% acetonitrile), and 50µl Solvent A (0.5% acetic acid). Subsequently, peptides were loaded and washed twice with Solvent A. Tips were dry-stored until analysis. Peptides were eluted from the StageTips and separated by reversed-phase liquid chromatography on a 2.5h 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 a TOP10 sequencing mode, meaning that one full mass spectrometry (MS) scan was followed by higher energy collision induced dissociation (HCD) and subsequent detection of the fragmentation spectra of the 10 most abundant peptide ions (tandem mass spectrometry; MS/MS). Collectively, ~160000 isotype patterns were generated resulting from ~6000 mass spectrometry (MS) runs. Consequently, ~33000 tandem mass spectrometry (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]. Briefly, trypsin digested protein samples were incubated with dimethyl labelling reagents (4µl of 0.6M $NaBH_3CN$ together with 4µl of 4% $CH_2O$ or $CD_2O$ for light or heavy labelling, respectively) for 1h at RT in agitation. The reaction was stopped by adding 16µl of 1% $NH_3$. Samples were acidified with 0.1% trifluoroacetic acid, and finally loaded on stage-tips. Each immunoprecipitation was performed twice, switching the labels.

**Analysis of mass spectrometry data**

The raw label-free quantitative mass spectrometric data was analysed using the MaxQuant software suite[41]. Peptide spectra were searched against the human database (Uniprot) using the integrated Andromeda search engine, and peptides were identified with an 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 2 ratio counts, in order to obtain label free quantification (LFQ) intensities. Collectively, the MS/MS spectra were matched to ~20000 known peptides, leading to the identification of 3635 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 having been identified in all replicates of at least one condition. LFQ intensities were logarithmized, and missing intensity values were imputed by representing 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 against the IgG negative controls. Statistical significance was set with a permutation-based FDR<0.05 (250 permutations). Fold-enrichment over IgG controls were calculated from LFQ intensities.

This same pipeline was used to analyze mass spectrometry data based on dimethyl labelling, with the following two exceptions. First, an additional mass of 28.03Da (light) or 32.06Da (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 experiment.

512

513 **Biological interpretation of mass spectrometry data**

514 The SMAD2/3 protein-protein interaction network was generated using Cytoscape v2.8.3[43].
515 First, all the annotated interactions involving the SMAD2/3 binding proteins were inferred by
516 interrogating protein-protein interaction databases through the PSIQUIC Universal Web
517 Service Client. IMEx-complying interactions were retained and merged by union. Then, a
518 subnetwork involving only the SMAD2/3 interactors was isolated. Finally, duplicate nodes
519 and self-loops were removed to simplify visualization. Note that based on our results all the
520 proteins shown would be connected to SMAD2/3, but such links were omitted to simplify
521 visualization and highlight those interactions with SMAD2/3 that were already known.
522 Proteins lacking any link and small complexes of less than three factors were not shown to
523 improve presentation clarity. Note that since the nodes representing SMAD2 and SMAD3
524 shared the very same links, they were fused into a single node (SMAD2/3). Functional
525 enrichment analysis was performed using the Fisher's exact test implemented in Enrichr[44], and
526 only enriched terms with a Benjamini-Hochberg adjusted p-value<0.05 were considered. For
527 Gene Ontology (GO) enrichment analysis, the 2015 GO annotation was used. For mouse
528 phenotype enrichment analysis, the level 3 of the Mouse Genomic Informatics (MGI)
529 annotation was used. To compare protein abundance in different conditions, a cut-off of
530 absolute LFQ intensity $\log_2$ fold-change larger than 2 was chosen, as label-free mass
531 spectrometry is at present not sensitive enough to detect smaller changes with confidence[37].

532

533 **Proximity ligation assay (PLA)**

534 PLA was performed using the Duolink In Situ Red Starter Kit Goat/Rabbit (Sigma-Aldrich).
535 Cells were cultured on glass coverslips and prepared by fixation in PBS 4% PFA for 10' at
536 RT, followed by two gentle washes in PBS. All subsequent incubations were performed at RT

unless otherwise stated. Samples were permeablilized in PBS 0.25% Triton X-100 for 20', blocked in PBS 0.5% BSA for 30', and incubated with the two primary antibodies of interest (diluted in PBS 0.5% BSA; see Supplementary Table 6) for 1h 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 0.5% BSA, and pre-incubated for 20'. Following two washes with PBS 0.5% BSA, the coverslips were incubated with the PLA probe solution for 1h at 37°C in a humid 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' under gentle agitation, and incubated with 1x ligation solution supplemented with DNA ligase (1:40 dilution) for 30' at 37°C in a humid chamber. After two more washes in Wash Buffer A for 2' under gentle agitation, coverslips were incubated with 1x amplification solution supplemented with DNA polymerase (1:80 dilution) for 1h 40' 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', 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 40x/1.3 Oil DIC M27 objective, performing z-stack with optimal spacing (~0.36µm). Images were automatically analysed 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 (RIP)**

Approximately $2x10^7$ cells were used for each 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 2h before being washed once with RT PBS and UV cross-linked in PBS at RT using a Stratalinker 1800 at 254nm wavelength (irradiation of $400mJ/cm^2$). Crosslinked cells were scraped in cell dissociation buffer (CDB, Gibco) and pelleted at 250g for 5'. The cell pellet was incubated in five volumes of isotonic lysis buffer (ILB: 10mM Tris-HCl pH 7.5; 3mM $CaCl_2$; 2mM $MgCl_2$; 0.32M sucrose) for 12' to induce cell swelling. Then, Triton X-100 was added to a final concentration of 0.3%, and cells were incubated for 6' to lyse the plasma membranes. Nuclei were pelleted at 600g for 5', washed once with ten volumes of ILB, and finally resuspended in two volumes of nuclear lysis buffer (NLB: 50mM Tris-HCl pH 7.5; 100mM NaCl; 50mM KCl; 3mM $MgCl_2$; 1mM EDTA; 10% glycerol; 0.1% Tween) supplemented with 800U/ml RNasin Ribonuclease Plus Inhibitor (Promega) and 1µM 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 in rotation for 30', homogenized again by perfoming 30 additional strokes with the tight pestle, and incubated in rotation for 15' more minutes at RT after addition of 12.5µg/ml of DNase I (Sigma). The protein concentration was assessed, and approximately 1mg of protein was used for overnight IP in rotation with the primary antibody of interest (Supplementary Table 6), or with equal amounts of non-immune species-matched IgG. 10% of the protein lysate used for IP was saved as pre-IP input and stored at -80°C for subsequent RNA extraction. IPs were incubated for 1h with 30µl of Protein G-Agarose, then washed twice with 1ml of LiCl wash buffer (50mM Tris-HCl pH 7.5; 250mM LiCl; 0.1% Triton X-100; 1mM DTT) and twice with 1ml of NLB. Beads were resuspended in 90µl of 30mM Tris-HCl pH 9.0, and DNase-digested using the RNase-free DNase kit (QIAGEN) by

adding 10µl of RDD buffer and 2.5µl of DNase. The pre-IP input samples were similarly treated in parallel, and samples were incubated for 10' at RT. The reaction was stopped by adding 2mM EDTA and by heating at 70°C for 5'. Proteins were digested by adding 2µl of Proteinase K (20mg/ml; Sigma-Aldrich) and by incubating at 37°C for 30'. Finally, RNA was extracted by using 1ml of TriReagent (Sigma-Aldrich) according to the supplier's instructions. The RNA was resuspended in nuclease-free water, and half of the sample was subjected to retrotranscription using SuperScript II (ThermoFisher) using the manufacturer's protocol. The other half was subjected to 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-IP input and then to the IgG control using the ΔΔCt approach (see below). Supplementary Table 5 reports all the primers used.

**Chromatin immunoprecipitation (ChIP)**

Approximately $2x10^7$ cells were used for each ChIP, and cells were fed with fresh media 2h before collection. ChIP was performed using a previously described protocol[10,30]. Briefly, cells were cross-linked on plates first with protein-protein crosslinkers (10mM dimethyl 3,3'-dithiopropionimidate dihydrochloride and 2.5mM 3,3'-dithiodipropionic acid di-N-hydroxysuccinimide ester; Sigma-Aldrich) for 15' at RT, then with 1% formaldehyde for 15'. Cross-linking 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]). Cross-linking was reverted first by adding DTT (for disulphide bridge-containing protein-protein cross-linkers), then by incubating in high salt at high temperatures. DNA was finally

purified by sequential phenol-chloroform and chloroform extractions. Samples were analysed by qPCR using the ΔΔCt 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. Secondly, the enrichment was normalized to the one observed in non-immune IgG ChIP controls.

**m6A dot blot**

m6A dot blot was performed with minor modifications to what previously described[23]. poly-A RNA was purified from total cellular RNA using the Dynabeads mRNA Purification Kit (ThermoFisher), diluted in 50µl of RNA loading buffer [RLB: 2.2M formaldehyde; 50% formamide; 0.5x MOPS buffer (20mM MOPS; 12.5mM $CH_3COONa$; 1.25mM EDTA; pH 7.0)], incubated at 55°C for 15', and snap cooled on ice. An Amersham Hybond-XL membrane was rehydrated in water for 3', then in 10x saline-sodium citrate buffer (SSC: 1.5M NaCl 150mM $Na_3C_6H_5O_7$; pH 7.0) for 10', and finally "sandwiched" in a 96-well dot blot hybridization manifold (ThermoFisher Scientific). Following two washes of the wells with 150µl of 10x SSC, the RNA was spotted on the membrane. After ultraviolet light (UV) cross-linking for 2' at 254nm using a Stratalinker 1800 (Stratagene), the membrane was washed once with TBST buffer, and blocked for 1h at RT with Tris-buffered saline Tween buffer (TBST: 20mM Tris-HCl pH 7.5; 150mM NaCl; 0.1% Tween-20) supplemented with 4% non-fat dry milk. Incubations with the anti-m6A primary antibody (Synaptic System, catalogue number: 202-111; used at 1µg/ml) and the mouse-HRP secondary antibody (Supplementary Table 6) were each performed in TBST 4% milk for 1h at RT, and were followed by three 10' washes at RT in TBST. Finally, the membrane was incubated with Pierce ECL2 Western Blotting Substrate, and exposed to X-Ray Super RX Films.

**m6A nuclear-enriched methylated RNA immunoprecipitation**

m6A MeRIP on nuclear-enriched RNA to be analysed by deep sequencing (NeMeRIP-seq) was performed following modifications of previously described methods[23,45]. $7.5x10^7$ hESCs were used for each sample, and three biological replicates per condition were generated. Cells were fed with fresh medium for 2h before being washed with PBS, scraped in cell dissociation buffer (CDB, Gibco), and pelleted at 250g for 5'. The cell pellet was incubated in five volumes of isotonic lysis buffer (ILB: 10mM Tris-HCl pH 7.5; 3mM $CaCl_2$; 2mM $MgCl_2$; 0.32M sucrose; 1,000U/ml RNAsin ribonuclease inhibitor, Promega; and 1mM DTT) for 10' to induce cell swelling. Then, Triton X-100 was added to a final concentration of 0.3% and cells were incubated for 6' to lyse the plasma membranes. Nuclei were pelleted at 600g for 5', washed once with ten volumes of ILB. RNA was extracted from the nuclear pellet using the RNeasy midi kit (QIAGEN) according to 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 retrotranscription step. RNA was then chemically fragmented in 20µl reactions each containing 20µg of RNA in fragmentation buffer (FB: 10mM $ZnCl_2$; 10mM Tris-HCl pH 7.0). Such reactions were incubated at 95°C for 5', followed by inactivation with 50mM EDTA and storage on ice. The fragmented RNA was then cleaned up by ethanol precipitation. In preparation to the MeRIP, 15µg of anti m6A-antibody (Synaptic Systems, catalogue number: 202-003) or equivalent amounts of rabbit non-immune IgG were cross-linked to 0.5mg of magnetic beads by using the Dynabeads Antibody Coupling Kit (ThermoFisher Scientific) according to manufacturer's instructions. Following equilibration of the magnetic beads by washing with 500µl of binding buffer (BB: 50mM Tris-HCl pH 7.5; 150mM $NaCl_2$; 1% NP-40; 1mM EDTA), MeRIP reactions were assembled with

300µg of the fragmented RNA in 3ml of BB supplemented with 3000U of RNAsin ribonuclease inhibitor. Samples were incubated at 7rpm for 1h at RT. 5µg of fragmented RNA (10% of the amount used for MeRIP) were set aside as pre-MeRIP input control. MeRIP reactions were washed twice with BB, once with low–salt buffer [LSB: 0.25x SSPE (saline-sodium phosphate-EDTA buffer: 150mM NaCl; 10mM $NaHPO_4$-$H_2O$; 10mM $Na_2$-EDTA; pH 7.4); 37.5mM $NaCl_2$; 1mM EDTA; 0.05% Tween-20), once with high-salt buffer (HSB: 0.25x SSPE; 137.5mM $NaCl_2$; 1mM EDTA; 0.05% Tween-20), and twice with TE-Tween buffer (TTB: 10mM Tris-HCl pH 7.4; 1mM EDTA; 0.05% Tween-20). Each wash was performed by incubating the beads with 500µl of buffer at 7rpm for 3' at RT. Finally, RNA was eluted from the beads by four successive incubations with 75µl of elution buffer (EB: 50mM Tris-HCl pH 7.5; 150mM $NaCl_2$; 20mM DTT; 0.1% SDS; 1mM 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. 30µg of glycogen were added as carrier during ethanol precipitation. RNA was resuspended in 15µl of ultrapure RNAse-free water. Preparation of DNA libraries for deep sequencing was performed using the TruSeq Stranded total RNA kit (Illumina) according to manufacturer's instructions with the following exceptions: (1) Ribo-Zero treatment was performed only for pre-NeMeRIP samples, as ribosomal RNA contamination in m6A NeMeRIP samples was minimal; (1) since samples were pre-fragmented, the fragmentation step was bypassed and 30ng of RNA for each sample were used directly for library prep; (3) due to the small size of the library, a 2-fold excess of Ampure XP beads was used during all purification steps in order to retain small fragments; (4) due 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 ~120bp. 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

686 >30M unique clusters per sample. The PhiX control library (Illumina) was spiked into the
687 main library pool at 1% vol/vol for quality control purposes. Sequencing was performed using
688 a high output flow cell with 2x75 cycles of sequencing, which provided ~800M paired end
689 reads from ~400M unique clusters from each lane. Overall, an average of ~33M and ~54M
690 paired-end reads were generated for m6A MeRIP and pre-MeRIP samples, respectively.
691
692 Samples for m6A MeRIP to be analysed by qPCR (NeMeRIP-qPCR) were processed as just
693 described for NeMeRIP-seq, but starting from $2.5x10^7$ cells. MeRIP from cytoplasmic RNA
694 was performed from RNA extracted from the cytoplasmic fraction of cells being processed for
695 NeMeRIP. In both cases, MeRIP was performed as for NeMeRIP-seq, but using 2.5µg of anti
696 m6A-antibody (or equivalent amounts of rabbit non-immune IgG) and 50µg of RNA in 500µl
697 of BB supplemented with 500U of RNAsin ribonuclease inhibitor. At the end of the protocol,
698 RNA was resuspended in 15µl of ultrapure RNAse-free water. For m6A MeRIP on total RNA,
699 the protocol just described was followed exactly, with the exception that the subcellular
700 fractionation step was bypassed, and that total RNA was extracted from $5x10^6$ cells. For m6A
701 MeRIP on mRNA, poly-A RNA was purified from 75µg of total RNA using the Dynabeads
702 mRNA Purification Kit, and 2.5µg of the resulting mRNA were used for chemical
703 fragmentation and subsequent MeRIP with 1µg of anti-m6A antibody. At the end of all these
704 protocols, cDNA synthesis was performed using all of the MeRIP material in a 30µl reaction
705 containing 500ng random primers, 0.5mM dNTPs, 20U RNaseOUT, and 200U of SuperScript
706 II (all from Invitrogen), all according to manufacturer's instructions. cDNA was diluted 10-
707 fold, and 5µl were used for qPCR using KAPA Sybr Fast Low Rox (KAPA Biosystems). For
708 each gene of interest, two primer pairs were designed either against the region containing the
709 m6A peak[23], or against a negative region (portion of the same transcript lacking the m6A
710 peak; Supplementary Table 5). Results of MeRIP-qPCR for each gene were then calculated

using the ΔΔCt 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**

QC of raw sequencing data was assessed using Trimmomatic v0.35[46], with parameters 'LEADING:3 TRAILING:3 SLIDINGWINDOW:5:10 MINLEN:40'. Reads were aligned to 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 QC were concatenated and mapped in 'single-end' mode in order to be used with MeTDiff[48], which only supports single-end reads. Reads with MAPQ<20 were filtered out. m6A peak calling and differential RNA methylation in the exome was assessed using MetDiff[48] with pooled inputs for each conditions, GENE_ANNO_GTF=GRCh38.83, MINIMAL_MAPQ=20, and rest of parameters as default (PEAK_CUTOFF_FDR=0.05; DIFF_PEAK_CUTOFF_FDR=0.05). MetDiff calculates p-values by a likelihood ratio test, then adjust them to FDR by Benjamini-Hochberg correction. An additional cut-off of absolute fold-change>1.5 (meaning an absolute log2 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 m6A 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 in a distance closer than 100bp. The MetDiff and MACS2 peak lists largely overlapped (Extended Data Fig. 5d), and differed primarily because MACS2 identifies peaks throughout the genome

736 while MetDiff only identifies peaks found on the exome (Extended Data Fig. 5c). For the
737 following analyses focused on exonic m6A peaks we considered a stringent consensus list of
738 only those MetDiff peaks overlapping with MACS2 peaks (Supplementary Table 2, "exon
739 m6a"). We assessed the reproducibility of m6A NeMeRIP-seq triplicates in peak regions using
740 the Bioconductor package fCCAC v1.0.0[52]. Hierarchical clustering (euclidean distance,
741 complete method) of F values corresponding to first two canonical correlations divided the
742 samples in Activin and SB clusters. Normalized read coverage files were generated using the
743 function 'normalise_bigwig' in RSeQC-2.6[53] with default parameters. The distribution of m6A
744 coverage across genomic features was plotted using the Bioconductor package RCAS[54]
745 with sampleN=0 (no downsampling) and flankSize=2500. Motif finding on m6A peaks was
746 performed using DREME with default parameters[55]. For visualization purposes, the three
747 biological replicates were combined. The Biodalliance genome viewer[56] was used to generate
748 figures. Gene expression in this experiment was estimated from the pre-MeRIP input samples
749 (which represent an RNA-seq sample on nuclear-enriched RNA species). Quantification,
750 normalisation of read counts, and estimation of differential gene expression in pre-MeRIP
751 input samples were performed using featureCounts[57] and DESeq2[58]. For assessment
752 of reproducibility regularised log transformation of count data was computed, and biological
753 replicates of input samples of the same condition clustered together in the PC
754 space[59]. Estimation of differential m6A deposition onto each peak in NeMeRIP samples
755 versus input controls was performed using an analogous approach. Functional enrichment
756 analysis of m6A-marked transcripts was performed using Enrichr[44], as described above for
757 mass-spectrometry data. The coordinates of SMAD2/3 ChIP-seq peaks in hESCs[30] were
758 transferred from their original mapping on hg18 to hg38 using liftOver. Overlap of the
759 resulting intervals with m6A peaks significantly downregulated after 2h of SB was determined
760 using GAT[60] with default parameters. SMAD2/3 binding sites were assigned to the closest

761 gene using the annotatePeaks.pl function from the HOMER suite[61] with standard parameters.
762 The significance in the overlap between the resulting gene list and that of genes encoding for
763 transcripts with m6A peaks significantly downregulated after 2h of SB was calculated by a
764 hypergeometric test where the population size corresponded to the number of genes in the
765 standard Ensemble annotation (GRCh38.83).
766
767 m6A peaks on introns were identified in three steps (Extended Data Fig 6d). First, MetDiff
768 was used to simultaneously perform peak calling and differential methylation analysis. Since
769 MetDiff only accepts a transcriptome GTF annotation as an input to determine the genomic
770 space onto which it identifies m6A peaks, in order to determine peaks onto introns we
771 followed the strategy recommended by the package developers of running the software using a
772 custom transcriptome annotation that includes introns[48,62]. This "extended" transcriptome
773 annotation was built using Cufflinks 2.2.1[63] with parameters '--library-type=fr-firststrand -m
774 100 -s 50' and guided by the Ensemble annotation (GRCh38.83). This was assembled using all
775 pre-NeMeRIP input reads available. The result was an extended transcriptome annotation
776 including all of the transcribed genome that could be detected and reconstructed from our
777 nuclear-enriched input RNA samples, thus including most expressed introns. Then, MetDiff
778 was run using this extended annotation as input for GENE_ANNO_GTF, pooled inputs for
779 each conditions, WINDOW_WIDTH=40, SLIDING_STEP=20,
780 FRAGMENT_LENGHT=250, PEAK_CUTOFF_PVALUE=1E-03,
781 FOLD_ENRICHMENT=2, MINIMAL_MAPQ=20, and all other parameters as default). In a
782 second step, the peaks identified by MetDiff were filtered for robustness by requiring that they
783 overlapped with MACS2 peak calls, exactly as for exome-focused MetDiff peak calls
784 (Extended Data Fig. 5d). Finally, only peaks that strictly did not overlap with any exon based
785 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 m6A deposition based on the cutoff of FDR<0.05.

m6A exon peaks spanning splice sites were selected from those identified both by the MetDiff analysis on the transcribed genome that was just described and by MACS2. Among these peaks, those presenting sequencing reads overlapping to both an exon and upstream/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 ± 500bp surrounding introns. Biological replicates were merged and depicted on 10bp-binned heatmaps for visualization purposes. To study the covariation of m6A peaks inside each transcriptional unit, the exonic peak with the greatest down regulated MetDiff fold-change was compared to the mean fold-change of the rest of m6A peaks found within the gene (both on exons and on introns). The resulting correlation was significant ($p<2E-16$; adjusted $R^2=0.2221$)

**RNA sequencing (RNA-seq)**

Polyadenylated (poly-A) purified opposing strand-specific mRNA library libraries were prepared from 200ng 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 >30M unique clusters per sample (18 samples were split in two runs and multiplexed across 4

811 lanes per run). The PhiX control library (Illumina) was spiked into the main library pool at 1%
812 vol/vol for quality control purposes. Sequencing was performed using a high output flow cell
813 with 2x75 cycles of sequencing, which provided ~800M paired end reads from ~400M unique
814 clusters from each run. Overall, a total of ~80M paired end reads per sample were obtained.
815
816 **Analysis of RNA-seq data**
817 Reads were trimmed using Sickle[64] with 'q=20 and l=30'. To prepare for reads alignment, the
818 human transcriptome was built with TopHat2 v2.1.0[44] based on Bowtie v2.2.6[65] by using the
819 human GRCh38.p6 as reference genome, and the Ensembl gene transfer format (GTF) as
820 annotation (http://ftp.ensembl.org/pub/release-83/gtf/homo_sapiens/). All analyses were
821 performed using this transcriptome assembly. Alignment was performed using TopHat2 with
822 standard parameters. Using Samtools view[66], reads with MAPQ>10 were kept for further
823 analyses. Subsequent quantitative data analysis was performed using SeqMonk[67]. The RNA-
824 seq pipeline was used to quantify gene expression as reads per million mapped reads (RPM),
825 and differential expression analysis for binary comparisons was performed using the R
826 package DESeq2[58]. A combined cut-off of negative binomial test $p<0.05$ and abs.FC>2 was
827 chosen. Analysis of differentially expressed transcripts across all samples was done using the
828 R/Bioconductor timecourse package[68]. The Hotelling $T^2$ score for each transcript was
829 calculated using the MB.2D function with all parameters set to their default value. Hotelling
830 $T^2$ scores were used to rank probes according to differential expression across the time-course,
831 and the top 5% differentially expressed transcripts were selected for complete Euclidean
832 hierarchical clustering (k-means preprocessing; max of 300 clusters) using Perseus software.
833 Z-scores of $\log_2$ normalized expression values across the timecourse were calculated and used
834 for this analysis. 8 gene clusters were defined, and gene enrichment analysis for selected
835 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 (PCA) was performed on the same list of top 5% differentially expressed transcripts using Perseus.

**Quantitative real-time PCR (qPCR)**

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 manufacturer's instructions. 500ng of RNA was used for complementary DNA (cDNA) synthesis using SuperScript II (Invitrogen) according to manufacturer's instructions. cDNA was diluted 30-fold, and 5µl were used for qPCR using SensiMix SYBR low-ROX (Bioline) and 150nM forward and reverse primers (Sigma-Aldrich; see Supplementary Table 5 for primer sequences). Samples were run in technical duplicates on 96-well plates on a Stratagene Mx-3005P (Agilent), and results were analysed using the delta-delta cycle threshold (ΔΔCt) 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 ΔCt from all controls was used when calculating the ΔΔCt. 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 was then fit to a one-

861 phase decay regression model[70], and statistical differences in mRNA half-live were evaluated
862 by comparing the model fits by extra sum-of-squares F test.
863
864 **Western blot**
865 Samples were prepared by adding Laemmli buffer (final concentration of 30mM Tris-HCl pH
866 6.8, 6% glycerol, 2% sodium dodecyl sulphate/SDS, 0.02% bromophenol blue, and 0.25% β-
867 mercaptoethanol), and were denatured at 95°C for 5'. Proteins were loaded and run on 4-12%
868 NuPAGE Bis-Tris Precast Gels (Invitrogen), then transferred to polyvinylidene fluoride
869 (PVDF) membranes by liquid transfer using NuPAGE Transfer buffer (Invitrogen).
870 Membranes were blocked for 1h at RT in PBS 0.05% Tween-20 (PBST) supplemented with
871 4% non-fat dried milk, and incubated overnight at 4°C with the primary antibody diluted in the
872 same blocking buffer (Supplementary Table 6). After three washes in PBST, membranes were
873 incubated for 1h at RT with horseradish peroxidase (HRP)-conjugated secondary antibodies
874 diluted in blocking buffer (Supplementary Table 6), then further washed three times with
875 PBST before being incubated with Pierce ECL2 Western Blotting Substrate (Thermo) and
876 exposed to X-Ray Super RX Films (Fujifilm).
877
878 **Immunofluorescence**
879 Cells were fixed for 20' at 4°C in PBS 4% PFA, rinsed three times with PBS, and blocked and
880 permeabilized for 30' at RT using PBS with 10% donkey serum (Biorad) and 0.1% Triton X-
881 100 (Sigma-Aldrich). Primary antibodies (Supplementary Table 6) were diluted in PBS 1%
882 donkey serum 0.1% Triton X-100 and incubated overnight at 4°C. This was followed by three
883 washes with PBS and by further incubation with AlexaFluor secondary antibodies
884 (Supplementary Table 6) for 1h at RT protected from light. Cells were finally washed three
885 times with PBS, and 4′,6-Diamidine-2′-phenylindole dihydrochloride (DAPI; Sigma-Aldrich)

886 was added to the first wash to stain nuclei. Images were acquired using a LSM 700 confocal
887 microscope (Leica).
888
889 **Flow cytometry**
890 Single cell suspensions were prepared by incubation in cell cell dissociation buffer (CDB;
891 Gibco) for 10' at 37° followed by extensive pipetting. Cells were washed twice with PBS and
892 fixed for 20' at 4°C with PBS 4% PFA. After three washes with PBS, cells were first
893 permeabilized for 20' at RT with PBS 0.1% Triton X-100, then blocked for 30' at RT with
894 PBS 10% donkey serum. Primary and secondary antibodies incubations (Supplementary Table
895 6) were performed for 1h each at RT in PBS 1% donkey serum 0.1% Triton X-100, and cells
896 were washed three times with this same buffer after each incubation. Flow cytometry was
897 performed using a Cyan ADP flow-cytometer, and at least 10,000 events were recorded. Data
898 analysis was performed using FlowJo X.
899
900 **Statistics and reproducibility**
901 Unless described otherwise in a specific section of the Methods, standard statistical analyses
902 were performed using GraphPad Prism 7 using default parameters. The type and number of
903 replicates, the statistical test used, and the test results are described in the figure legends. The
904 level of significance in all graphs is represented as it follows (p denotes the p-value):
905 *=$p<0.05$, **=$p<0.01$, and ***=$p<0.001$. Test assumptions (e.g. normal distribution) were
906 confirmed where appropriate. For analyses with $n<10$ individual data points are shown, and
907 the mean ± SEM is reported for all analyses with $n>2$. The mean is reported when $n=2$, and no
908 other statistics were calculated for these experiments due to the small sample size. No
909 experimental samples were excluded from the statistical analyses. Sample size was not pre-
910 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 to 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 (Source Data Table Figure 1 and 3, and Source Data Extended Data Figure 1 to 10). Electrophoretic gel source data (uncropped scans with size marker indications) are presented in Supplementary Figure 1. Supplementary Tables 1 to 4 provide the results of bioinformatics analyses described in the text and figure legends. All other data that supports the findings of this study are available from the corresponding author upon reasonable request.

## Methods specific references

31. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–4 (2011).

936 32. Vallier, L. Serum-Free and Feeder-Free Culture Conditions for Human Embryonic
937 Stem Cells. *Springer Protoc.* **690,** 57–66 (2011).
938 33. Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem
939 cells under chemically defined conditions that recapitulate liver development.
940 *Hepatology* **51,** 1754–1765 (2010).
941 34. Vallier, L. *et al.* Early cell fate decisions of human embryonic stem cells and mouse
942 epiblast stem cells are controlled by the same signalling pathways. *PLoS One* **4,** e6082
943 (2009).
944 35. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an
945 arrayed viral high-content screen. *Cell* **124,** 1283–98 (2006).
946 36. Pawlowski, M. *et al.* Inducible and Deterministic Forward Programming of Human
947 Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. *Stem*
948 *Cell Reports* **8,** 803–812 (2017).
949 37. Hubner, N. C. & Mann, M. Extracting gene function from protein-protein interactions
950 using Quantitative BAC InteraCtomics (QUBIC). *Methods* **53,** 453–9 (2011).
951 38. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment,
952 pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*
953 **2,** 1896–906 (2007).
954 39. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex
955 peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4,**
956 484–494 (2009).
957 40. Hubner, N. C., Nguyen, L. N., Hornig, N. C. & Stunnenberg, H. G. A quantitative
958 proteomics tool to identify DNA-protein interactions in primary cells or blood. *J.*
959 *Proteome Res.* **14,** 1315–29 (2015).
960 41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized

961 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–72 (2008).
963 42. Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–54 (2010).
965 43. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).
967 44. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).
969 45. Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* **8,** 176–89 (2013).
973 46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–20 (2014).
975 47. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
977 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).
979 49. Saletore, Y. *et al.* The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13,** 175 (2012).
981 50. Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14,** 23–31 (2016).
983 51. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
985 52. Madrigal, P. fCCAC: functional canonical correlation analysis to evaluate covariance

between nucleic acid sequencing datasets. *Bioinformatics* **33,** 746–8 (2017).

53. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–5 (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,** 1653–9 (2011).

56. Down, T. A., Piipari, M. & Hubbard, T. J. P. Dalliance: interactive genome viewing on the web. *Bioinformatics* **27,** 889–90 (2011).

57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–30 (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,** 2046–8 (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,** 576–89 (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 (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle.* (2011).

65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).

66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).

67. Andrews, S. SeqMonk: A tool to visualise and analyse high throughput mapped sequence data. *Available at http://www.bioinformatics.bbsrc.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,** 1–25 (2004).

69. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) 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).

## Extended Data Figure legends

**Extended Data Figure 1. Optimized SMAD2/3 co-immunoprecipitation protocol to define its interactome in hPSCs and early endoderm cells.**

(**a**) Western blots of SMAD2/3 or control (IgG) immunoprecipitations (IPs) from nuclear extracts of hESCs following the co-IP1 or co-IP2 protocols. Input is 5% of the material used for IP. Results are representative of two independent experiments. For gel source data, see Supplementary Figure 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-IPs compared with IgG negative control co-IPs. The experiments were performed from nuclear extracts of hESCs. The SMAD2/3 and IgG negative control co-IPs were differentially labelled post-IP 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 ones. 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-IP 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 IPs 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-IPs; one-tailed t-test: permutation-based FDR<0.05). (**d**) Selected results of the analysis described in panel c for SMAD2, SMAD3, and selected known *bona fide* SMAD2/3 binding partners (full results can be found in Supplementary Table 1). (**e**) Average label free quantification (LFQ) intensity $\log_2$ ratios in endoderm (Endo) and pluripotency (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 Mouse Genomics Informatics (MGI) database. All SMAD2/3 putative interacting proteins were considered for this analysis (n=89 proteins; Fisher's exact test followed by Benjamini-Hochberg correction for multiple comparisions).

For each term, its rank in the analysis, the adjusted p-value, and the number of associated genes are reported.

**Extended Data Figure 2. Functional characterization of SMAD2/3 transcriptional and epigenetic cofactors in hPSCs.**

(**a**) Western blots of SMAD2/3 or control (IgG) immunoprecipitations (IPs) from nuclear extracts of pluripotent hESCs (Pluri), or hESCs differentiated into endoderm for 36h (Endo). Input is 5% of the material used for IP. Results are representative of two independent experiments. (**b**) Schematic of the experimental approach for the generation of tetracycline-inducible knockdown (iKD) hESC lines for SMAD2/3 cofactors. (**c**) qPCR screening of iKD hESCs cultured in absence (CTR) or presence of tetracycline for 3 days (TET). Three distinct shRNAs were tested for each gene. Expression is shown as normalized on the average level in hESCs carrying negative control shRNAs (scrambled, SCR, or against B2M) and cultured in absence of tetracycline. The mean is indicated, n=2 independent clonal pools. Note than for the B2M shRNA only the SCR 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 presence of tetracycline for 6 days to induce knockdown. Scale bars: 400μm. Results are representative of two independent experiments. (**e**) Immunofluorescence for the pluripotency factor NANOG in iKD hESCs for the indicated genes cultured in absence (CTR) or presence of tetracycline (TET) for 6 days. DAPI: nuclear staining; scale bars: 400μm. Results are representative of two independent experiments. (**f**) Heatmap summarizing qPCR analyses of iKD hESCs cultured as in panel e. log2 fold-changes (FC) are compared to SCR CTR (n=2 clonal pools). Germ layer markers are grouped in boxes (green: endoderm; red: mesoderm; blue: neuroectoderm).

**Extended Data Figure 3. Functional characterization of SMAD2/3 transcriptional and epigenetic cofactors during endoderm differentiation.**

(**a**) qPCR validation of inducible knockdown (iKD) hESCs in pluripotency (PLURI) and following endoderm differentiation (ENDO). Pluripotent cells were cultured in absence (CTR) or presence of tetracycline (TET) for 6 days. For endoderm differentiation, tetracycline treatment was initiated in undifferentiated hESCs for 3 days in order to ensure gene knockdown at the start of endoderm specification, and was then maintained during differentiation (3 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 shown as normalized to the average level in pluripotent hESCs carrying a scrambled (SCR) control shRNAs and cultured in absence of tetracycline. The mean is indicated, n=2 independent clonal pools. (**b**) Immunofluorescence for the endoderm marker SOX17 following endoderm differentiation of iKD hESCs expressing the indicated shRNAs (sh) and cultured as described in panel a. DAPI shows nuclear staining. Scale bars: 400μm. 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 m6A methyltransferase complex.**

**(a-c)** Western blots of SMAD2/3 (S2/3), METTL3 (M3), METTL14 (M14), or control (IgG) immunoprecipitations (IPs) from nuclear extracts of hPSCs (hESCs for panels a and c, and hiPSCs for panel b). Input is 5% of the material used for IP. In c, IPs were performed from hPSCs maintained in presence of Activin or treated for 1h with the Activin/Nodal inhibitor SB-431542 (SB). Results are representative of three (panel a) or two (panels 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 control. Cells were cultured in presence of Activin or treated with SB for the indicated time points. The mean is indicated, n=2 cultures. NANOG OE cells are resistant to downregulation of NANOG following Activin/Nodal inhibiton. (**e**) RNA immunoprecipitation (RIP) experiments for WTAP, SMAD2/3 (S2/3), or IgG control in NANOG overexpressing hESCs maintained in presence of Activin or treated for 2 hours with SB. Enrichment of the indicated transcripts was measured by qPCR and expressed over background levels observed in IgG RIP in presence of Activin. *RPLP0* was tested as a negative control transcript. Mean ± SEM, n=3 cultures. Significance was tested for differences versus Activin (left panel) or versus IgG (right panel) by 2-way ANOVA with post-hoc Holm-Sidak comparisons: *=$p<0.05$, **=$p<0.01$, and ***=$p<0.001$. (**f**) Chromatin immunoprecipitation (ChIP) qPCR in hESCs for the indicated proteins or for the negative control ChIP (IgG). qPCR was performed for validated genomic SMAD2/3 binding sites associated to the indicated genes[10,30]. hESCs were cultured in presence of Activin or treated for 2h with SB. The enrichment is expressed as normalized levels to background binding observed in IgG ChIP. The mean is indicated, n=2 technical replicates. Results are representative of three independent experiments.

**Extended Data Figure 5. Monitoring the changes in m6A deposition rapidly induced by Activin/Nodal inhibition.**

(**a-b**) m6A methylated RNA immunoprecipitation (MeRIP) qPCR results from purified mRNA, total cellular RNA, or cellular RNA species separated following nuclear/cytoplasmic subcellular fractionation. hESCs were cultured in pluripotency-maintaining conditions containing Activin, or subjected to Activin/Nodal inhibition for 2h with SB-431542 (SB). IgG

MeRIP experiments were performed as negative controls. The mean is indicated, n=2 technical replicates. Differences between Activin and SB-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 m6A 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 presence of Activin or subjected to Activin/Nodal inhibition for 2h with SB (n=3 cultures). Note that the sum of the percentages within each graph does not add to 100% because some m6A 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 MACS2 were considered for subsequent analyses focused on m6A peaks on exons. (**e**) Top sequence motifs identified *de novo* on all m6A exon peaks, or on such peaks that showed significant downregulation following Activin/Nodal inhibition (Activin/Nodal-sensitive m6A peaks; Supplementary Table 2). The position of the methylated adenosine is indicated by a box. (**f**) Coverage profiles for all m6A 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 m6A exon peaks to transcription start sites (TSS) or transcription end sites (TES). In g, the analysis was performed for all m6A peaks. In h, only Activin/Nodal-sensitive peaks were considered. (**i**) On the left, Activin/Nodal-sensitive m6A exon peaks were evaluated for direct overlap with SMAD2/3 binding sites measured by ChIP-seq[30]. n=482 peaks; FDR=0.41 (non-significant at 95% confidence interval, N.S.) as calculated by the permutation test implemented by the GAT python package. On the right,

overlap was calculated after the same features were mapped to their corresponding transcripts or genes, respectively. A significant overlap was observed for the transcript-gene overlap. n=372 genes; hypergeometric test p-value (p) of 2.88E-18, significant at 95% confidence interval. (**j**) m6A 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 m6A NeMeRIP. Orange: sequencing results of pre-NeMeRIP input RNA (negative control). GENCODE gene annotations are shown (red: protein coding exons; white: untranslated exons; note that all potential exons are shown and overlaid). The location of SMAD2/3 ChIP-seq binding sites is also reported. Compared to the other genes shown, the m6A levels on *SOX2* were unaffected by Activin/Nodal inhibition, showing specificity of action. *OCT4/POU5F1* is reported as negative control since it is known not to have any m6A site[23], as confirmed by the lack of m6A enrichment compared to the input.

**Extended Data Figure 6. Features of Activin/Nodal-sensitive differential m6A deposition.** (**a**) Scatter plot of the average $\log_2$ fold-change (FC) in SB-431542 (SB) versus Activin-treated hESCs for m6A NeMeRIP-seq and pre-NeMeRIP input RNA (n=3 cultures). The analysis was performed for all m6A exon peaks (left), or for such peaks significantly downregulated following Activin/Nodal inhibition (right). Data was colour coded according to the square of the difference between the two values (square diff.). (**b-c**) As in Extended Data Fig. 5j, but for representative transcripts whose expression is stable following Activin/Nodal inhibition for 2 hours (n=3 cultures; replicates combined for visualization). The m6A NeMeRIP and input tracks were separated and have a different scale in order to facilitate visual comparison between the conditions. The m6A peaks and those significantly downregulated after SB treatment for 2h are indicated. (**d**) Venn diagram illustrating the

strategy for the identification of m6A 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 is abundant with 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 vs SB 2h for m6A peaks on introns. n=3 cultures; p-value calculated by likelihood ratio test implemented in the MetDiff R package, and adjusted to False Discovery Rate (FDR) by Benjamini-Hochberg correction. See Supplementary Table 2 for the FDR of individual peaks. abs. FC: absolute fold-change. (**f**) As in Extended Data Fig. 5j, but for a representative transcript that shows Activin/Nodal-sensitive m6A deposition in introns (n=3 cultures; replicates combined for visualization). The m6A peaks on exons, introns, and those significantly downregulated after SB treatment within each subset are indicated. (**g**) Plots of RPKM-normalized mean m6A coverage for m6A exon peaks significantly downregulated after SB treatment (absolute fold-change>1.5). Data for all such peaks is in blue, while 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 position of the 3' and 5' splice sites (SS) is indicated. A window of 500 base pairs (bp) on either side of the splice sites is shown. m6A: signal from m6A NeMeRIP-seq; input: signal from pre-NeMeRIP input RNA. The results show that coverage of Activin/Nodal-sensitive m6A peaks often spans across splice sites (highlighted by the dotted lines). (**h**) Heatmap representing in an extended form the data shown in panel g for all Activin/Nodal-sensitive m6A exon peaks in hESCs cultured in presence of Activin. Multiple regions where sequencing coverage extends across exon-intron junctions can be observed (see Supplementary Table 2). (**i**) Example of an Activin/Nodal-sensitive peaks located in the proximity of a 3' splice site (n=3 cultures; replicates combined for visualization). This peak can be visualized within its genomic context

in panel c, where it is indicated by a dotted box. Data plotted on top is m6A NeMeRIP-seq coverage, while individual next generation sequencing reads are shown on the bottom. Multiple reads spanning the exon-intron junction (indicated by the dashed line) can be observed. (**j**) Relationship between the decrease of m6A on the most strongly 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 m6A peaks and with at least one peak significantly decreasing after Activin/Nodal inhibition with SB (absolute fold-change>1.5). Sensitivity of m6A deposition to Activin/Nodal signalling across these transcripts correlated.

**Extended Data Figure 7. Generation and functional characterization of inducible knockdown hPSCs for the subunits of the m6A methyltransferase complex.**

(**a**) qPCR validation of tetracycline-inducible knockdown (iKD) hESCs cultured in presence of tetracycline (TET) for 5 days to drive gene knockdown. Two distinct shRNAs (sh) and multiple clonal sublines (cl) were tested for each gene. Expression is shown as normalized on the average level in hESCs carrying a negative control scrambled (SCR) shRNA. For each gene, sh1 cl1 was chosen 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**) m6A methylated RNA immunoprecipitation (MeRIP)-qPCR in iKD hESCs cultured for 10 days in absence (CTR) or presence of tetracycline (TET). m6A abundance is reported relative to control conditions in the same hESC line. The mean is indicated, n=2 technical replicates. Results are representative of two independent experiments. (**d**) m6A dot blot in WTAP or SCR iKD hESCs treated as described in panel c. Decreasing amounts of mRNA were spotted to facilitate semi-quantitative comparisons, as indicated. Results are representative of two independent experiments. (**e**) Immunofluorescence for the pluripotency markers NANOG and

OCT4 in iKD hESCs cultured for three passages (15 days) in absence (CTR) or presence of tetracycline (TET). DAPI shows nuclear staining. Scale bars: 100μm. Results are representative of two independent experiments. (**f**) Flow cytometry quantifications for NANOG in cells treated as described for panel e. The percentage and median fluorescence intensity (MFI) of NANOG positive cells (NANOG+) are reported. The gates used for the analysis are shown, and were determined based on a secondary antibody only negative staining (NEG). Results are representative of two independent experiments.

**Extended Data Figure 8. Function of the m6A methyltransferase complex during germ layer specification.**

(**a**) qPCR analysis following neuroectoderm or endoderm differentiation of inducible knockdown (iKD) hESCs cultured in absence (CTR) or presence of tetracycline (TET). Tetracycline treatment was initiated in undifferentiated hESCs for 10 days and was maintained during differentiation (3 days). Expression is shown as normalized on the average level in undifferentiated hESCs. Mean ± SEM, n=3 cultures. Significant differences vs same iKD line in control conditions were calculated by 2-way ANOVA with post-hoc Holm-Sidak comparisons: *=$p<0.05$, **=$p<0.01$, and ***=$p<0.001$. (**b**) Flow cytometry quantification of the percentage of SOX1 positive cells (SOX1+) in cells treated as described for panel a. Mean is indicated, n=2 cultures. (**c**) Immunofluorescent stainings for the lineage marker SOX17 in endoderm-differentiated hESCs treated as described for panel a. DAPI shows nuclear staining. 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 (M3), and METTL14 (M14). Cells expressing three copies of the scrambled shRNA (SCR3x) were used as negative control. Cells were cultured in presence of tetracycline (TET) for 5 days to drive gene knockdown. Mean ± SEM, n=3 cultures. Significant differences vs SCR3x hESCs in control conditions were calculated by 2-

way ANOVA with post-hoc Holm-Sidak comparisons: ***=p<0.001. (**e-f**) qPCR analysis following endoderm differentiation of WTAP, METTL3, and METTL14 MiKD hESCs treated as described for panel a. Mean ± SEM, n=3 cultures. Significant differences versus control conditions were calculated by two tailed t-test (panel e) or 2-way ANOVA with post-hoc Holm-Sidak comparisons (panel f): **=p<0.01, and ***=p<0.001.

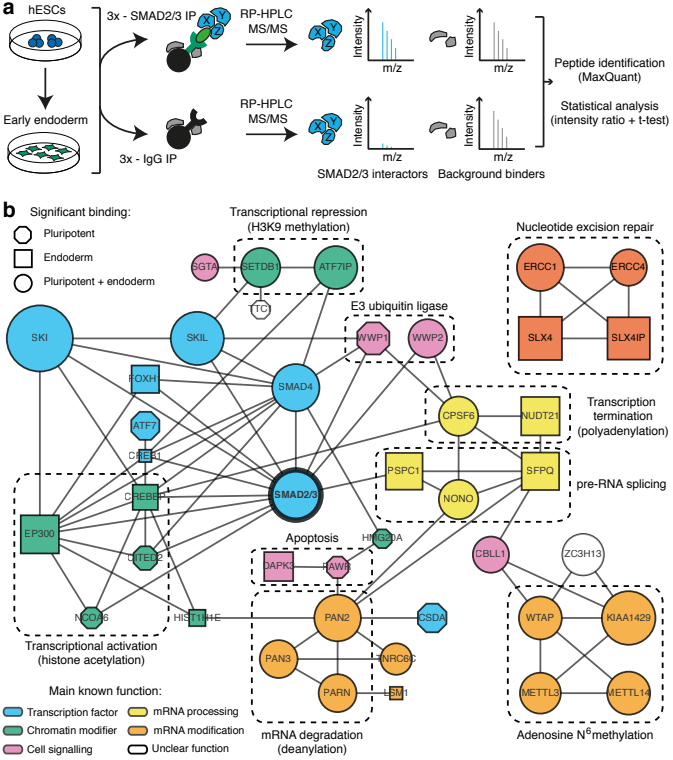
**Extended Data Figure 9. Function of the m6A methyltransferase complex during pluripotency exit induced by Activin/Nodal inhibition.**

(**a**) qPCR analyses in inducible knockdown (iKD) hESCs cultured in absence (CTR) or presence of tetracycline (TET) for 10 days, then subjected to Activin/Nodal signalling inhibition with SB-431542 (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 ± SEM, n=3 cultures. Significant differences vs same iKD line in control conditions were calculated by 2-way ANOVA with post-hoc Holm-Sidak comparisons: **=p<0.01, and ***=p<0.001. (**b**) Western blots of cells treated as described in panel a. TUBA4A (α-tubulin): loading control. Results are representative of two independent experiments. (**c**) Measurement of mRNA stability in WTAP iKD hESCs cultured in absence (CTR) or presence of tetracycline (TET) for 10 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 TET vs CTR is reported (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 *SOX2* (95% confidence interval). (**d**) Model showing the interplays between Activin/Nodal signalling and m6A deposition in hPSCs (left), and the phenotype induced by impairment of the m6A methyltransferase complex (right).

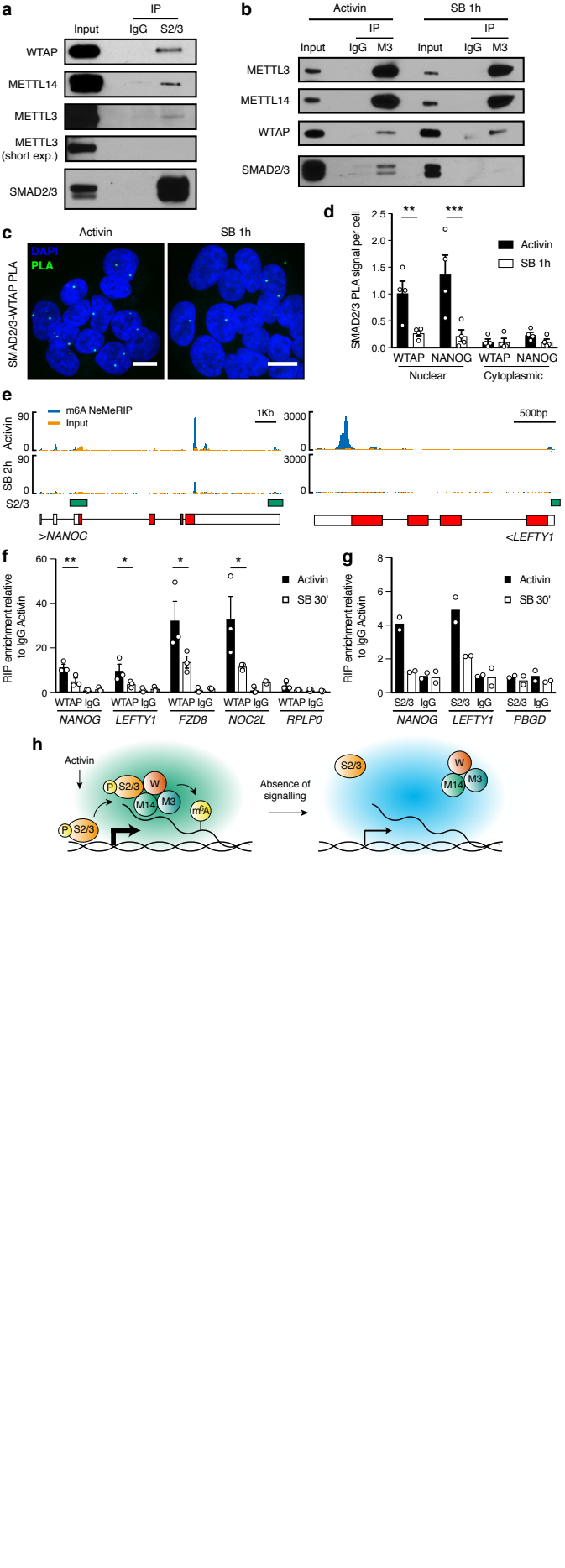
**Extended Data Figure 10. Genome wide analysis of the relationship between WTAP and Activin/Nodal signalling.**

(**a**) Schematic of the experimental approach to investigate the transcriptional changes induced by the knockdown of the m6A methyltransferase complex subunits during neuroectoderm specification of hESCs. (**b**) qPCR analyses of WTAP inducible knockdown (iKD) hESCs subjected to the experiment illustrated in panel a (n=3 cultures). Activin: cells maintained in standard pluripotency-promoting culture conditions containing Activin and collected at the beginning of the experiment. SB: SB-431542. Z-scores indicate differential expression measured in number of standard deviations from the average across all time points. (**c**) RNA-seq analysis at selected time points from the samples shown in panel b (n=3 cultures). The heatmap depicts 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 the Supplementary Discussion). (**d**) Expression profiles of genes belonging to the clusters indicated in panel c. Selected results of gene enrichment analysis and representative genes for each cluster are reported (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 comparisions). (**e**) Principal component analysis (PCA) of RNA-seq results described in panel 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 m6A peak[23] in transcripts significantly up- or downregulated following WTAP inducible knockdown in hESCs maintained in presence of Activin (left), or following Activin/Nodal inhibition for 2 hours with SB in control cells

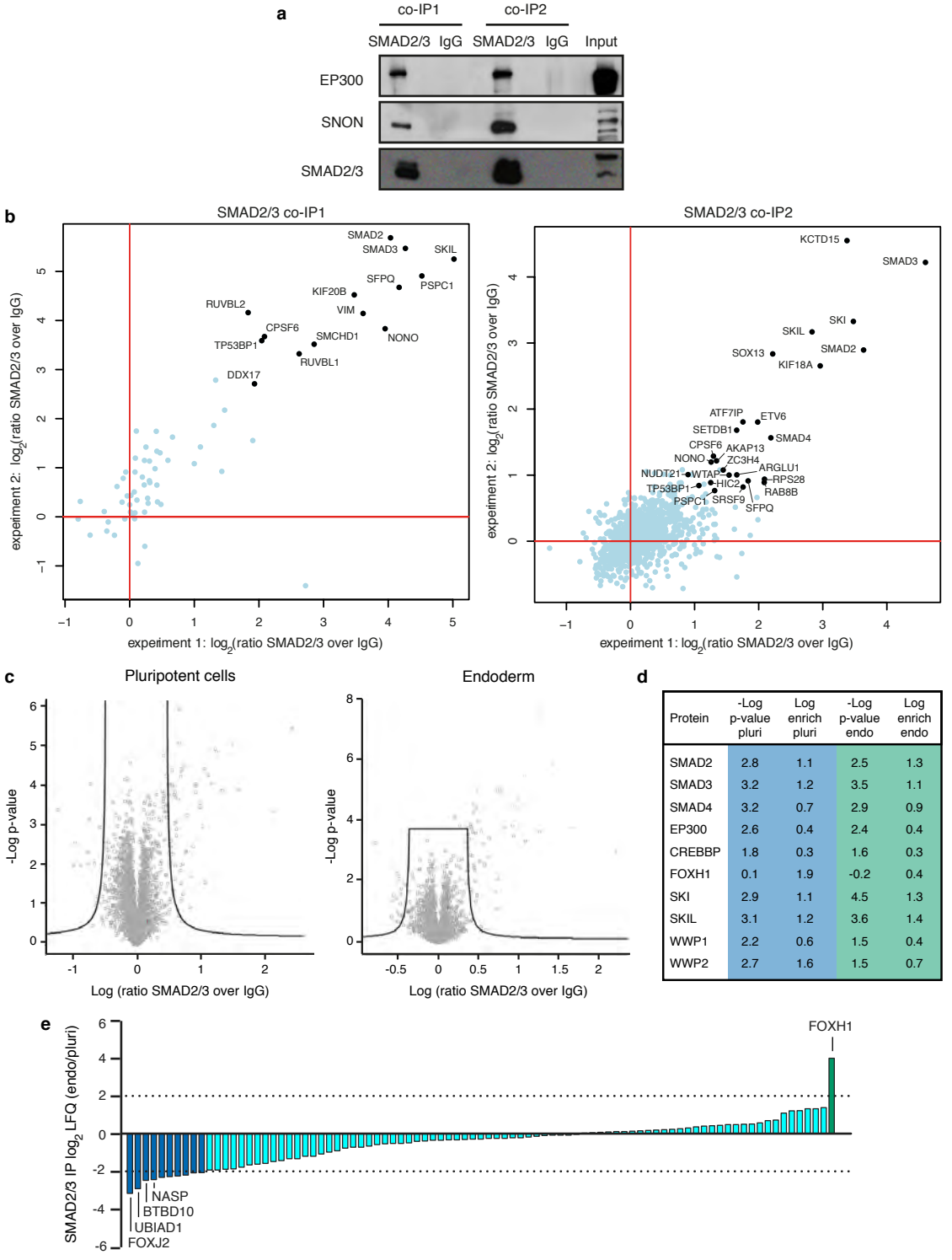
(right). Differential gene expression was calculated on n=3 cultures using the negative binomial test implemented in DEseq2 with a cutoff of $p<0.05$ and abs.FC>2. The number of genes in each group and the hypergeometric probabilities of the observed overlaps with m6A-marked transcripts are reported (n.s.: non-significant at 95% confidence interval).


a
hESCs
Early endoderm
3x - SMAD2/3 IP
3x - IgG IP
RP-HPLC MS/MS
RP-HPLC MS/MS
Intensity
m/z
Peptide identification (MaxQuant)
Statistical analysis (intensity ratio + t-test)
SMAD2/3 interactors
Background binders


b
Significant binding:
Pluripotent
Endoderm
Pluripotent + endoderm
Transcriptional repression (H3K9 methylation)
SGTA
SETDB1
ATF7IP
TCF
E3 ubiquitin ligase
WWP1
WWP2
Nucleotide excision repair
ERCC1
ERCC4
SLX4
SLX4IP
SKI
SKIL
FOXH1
SMAD4
ATF7
CREB1
CREBBP
SMAD2/3
EP300
CITED2
NCOR2
HIST1H1E
CPSF6
NUDT21
Transcription termination (polyadenylation)
PSPC1
NONO
SFPQ
pre-RNA splicing
Apoptosis
HIPK2A
DAPK3
PAWR
CBLL1
ZC3H13
PAN2
PAN3
PARN
CSDA
YNRC6B
LSM1
WTAP
KIAA1429
METTL3
METTL14
Transcriptional activation (histone acetylation)
mRNA degradation (deanylation)
Adenosine N6-methylation
Main known function:
Transcription factor
Chromatin modifier
Cell signalling
mRNA processing
mRNA modification
Unclear function


a
Input
IP
IgG
S2/3
WTAP
METTL14
METTL3
METTL3 (short exp.)
SMAD2/3
b
Activin
SB 1h
IP
Input
IgG
M3
METTL3
METTL14
WTAP
SMAD2/3
c
Activin
SB 1h
DAPI
PLA
SMAD2/3-WTAP PLA
d
SMAD2/3 PLA signal per cell
**
***
Activin
SB 1h
WTAP
NANOG
Nuclear
Cytoplasmic
e
m6A NeMeRIP
Input
Activin
SB 2h
S2/3
1Kb
500bp
>NANOG
<LEFTY1
f
RIP enrichment relative to IgG Activin
Activin
SB 30'
WTAP IgG
NANOG
LEFTY1
FZD8
NOC2L
RPLP0
g
RIP enrichment relative to IgG Activin
S2/3 IgG
NANOG
LEFTY1
PBGD
h
Activin
P
S2/3
W
M14
M3
mA
Absence of signalling


a
WTAP iKD
METTL14 iKD
METTL3 iKD
DAPI
SOX1
CTR
TET
b
CDX2
NANOG
LEFTY1
Expression relative to Actvin CTR
WTAP iKD CTR
WTAP iKD TET
Act
1h
2h
4h
8h
16h
32h
SB
c
WTAP/M3/ M14/ iKD
SCR3x iKD
TET:
METTL14
METTL3
WTAP
TUBA4A
d
Undifferentiated hESCs
Neuroectoderm
NANOG
LEFTY1
CER1
CDX2
HOXA1
SOX1
HOXA1
Expression relative to CTR
CTR
TET



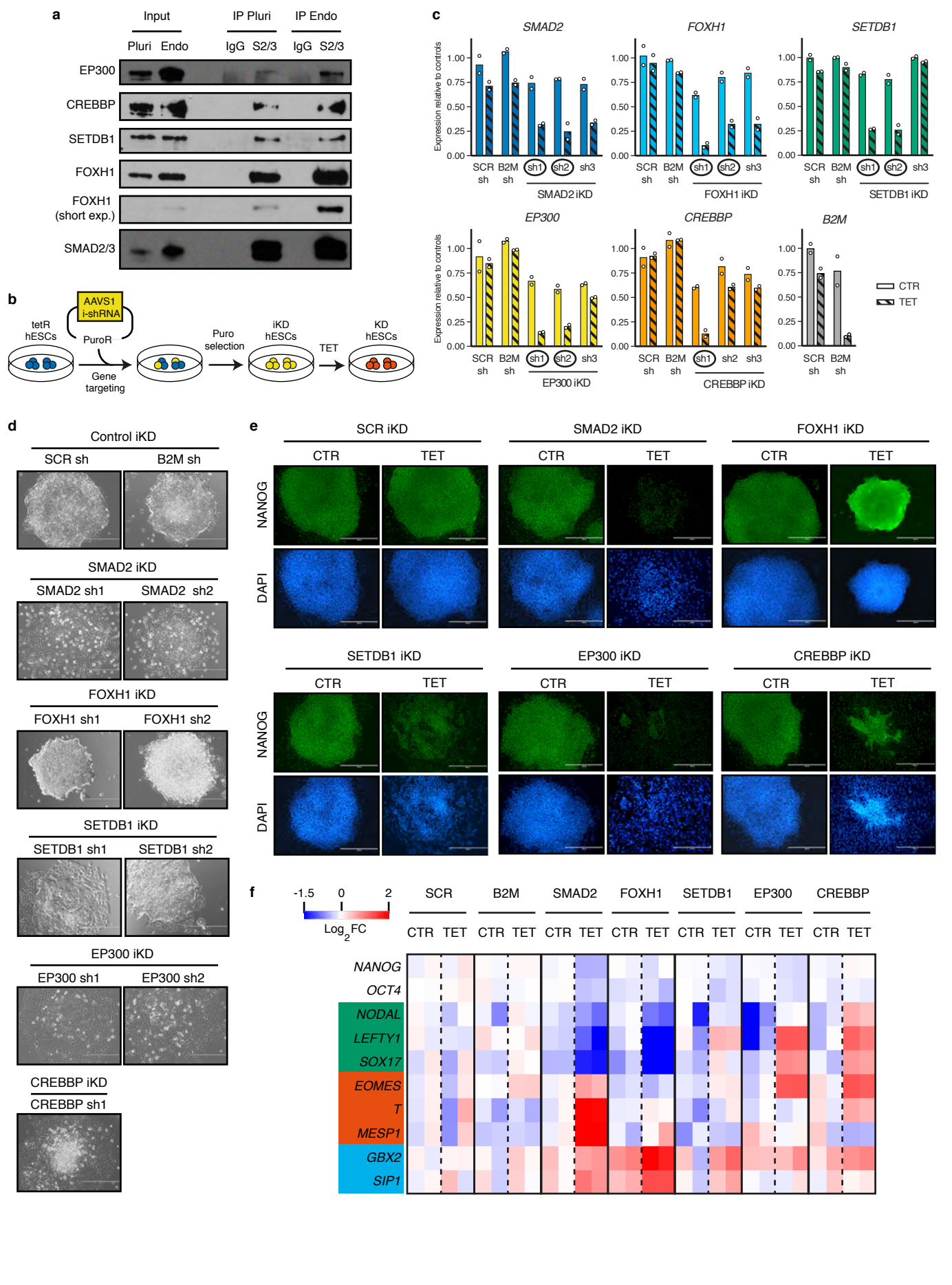

| Protein | -Log p-value pluri | Log enrich pluri | -Log p-value endo | Log enrich endo |
|---|---|---|---|---|
| SMAD2 | 2.8 | 1.1 | 2.5 | 1.3 |
| SMAD3 | 3.2 | 1.2 | 3.5 | 1.1 |
| SMAD4 | 3.2 | 0.7 | 2.9 | 0.9 |
| EP300 | 2.6 | 0.4 | 2.4 | 0.4 |
| CREBBP | 1.8 | 0.3 | 1.6 | 0.3 |
| FOXH1 | 0.1 | 1.9 | -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 |

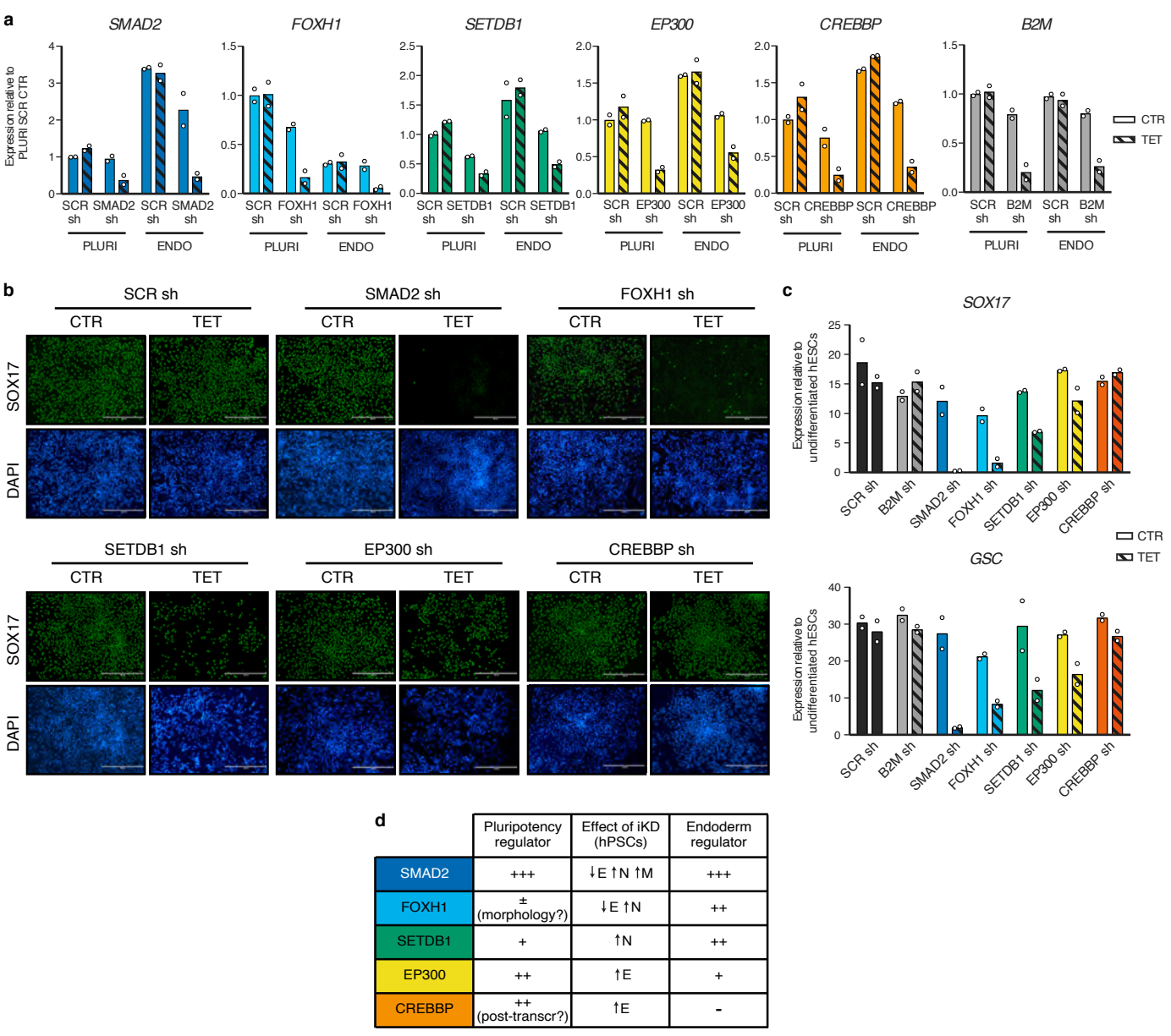
| GO molecular function term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| RNA Pol II transcription factor binding | 1 | 6.72E-06 | 8 |
| Poly(A)-specific ribonuclease activity | 3 | 1.97E-03 | 3 |
| SMAD binding | 4 | 1.97E-03 | 5 |
| Transcription corepressor activity | 8 | 2.57E-03 | 7 |
| Activating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

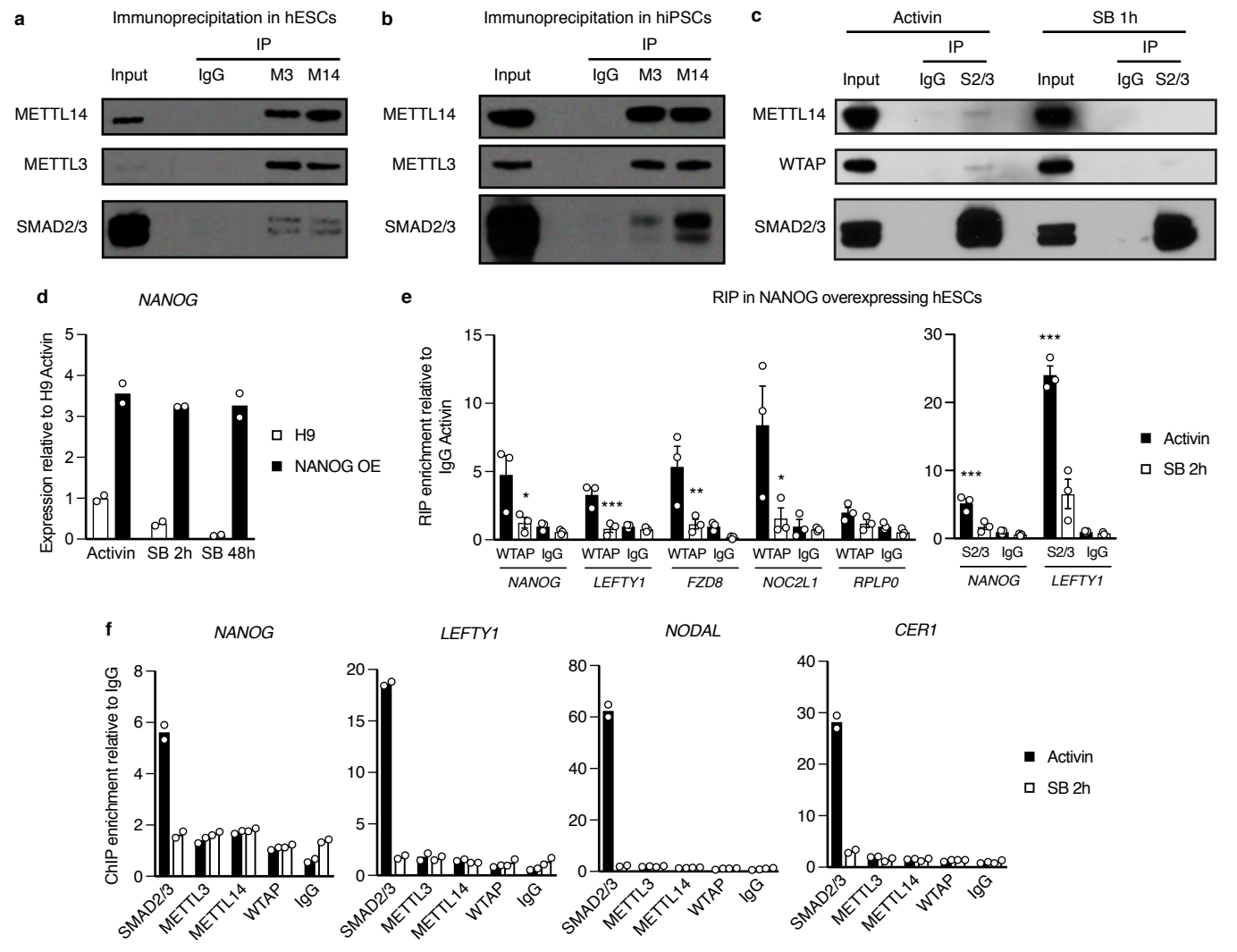
| MGI mouse phenotype term | Rank | Adj. p-value | Genes |
|---|---|---|---|
| Abnormal embryogenesis/development | 1 | 1.92E-08 | 19 |
| Abnormal survival | 2 | 7.16E-06 | 24 |
| Abnormal craniofacial morphololgy | 5 | 2.82E-02 | 8 |
| Abnormal respiratory system | 6 | 2.82E-02 | 6 |
| Abnormal digestive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

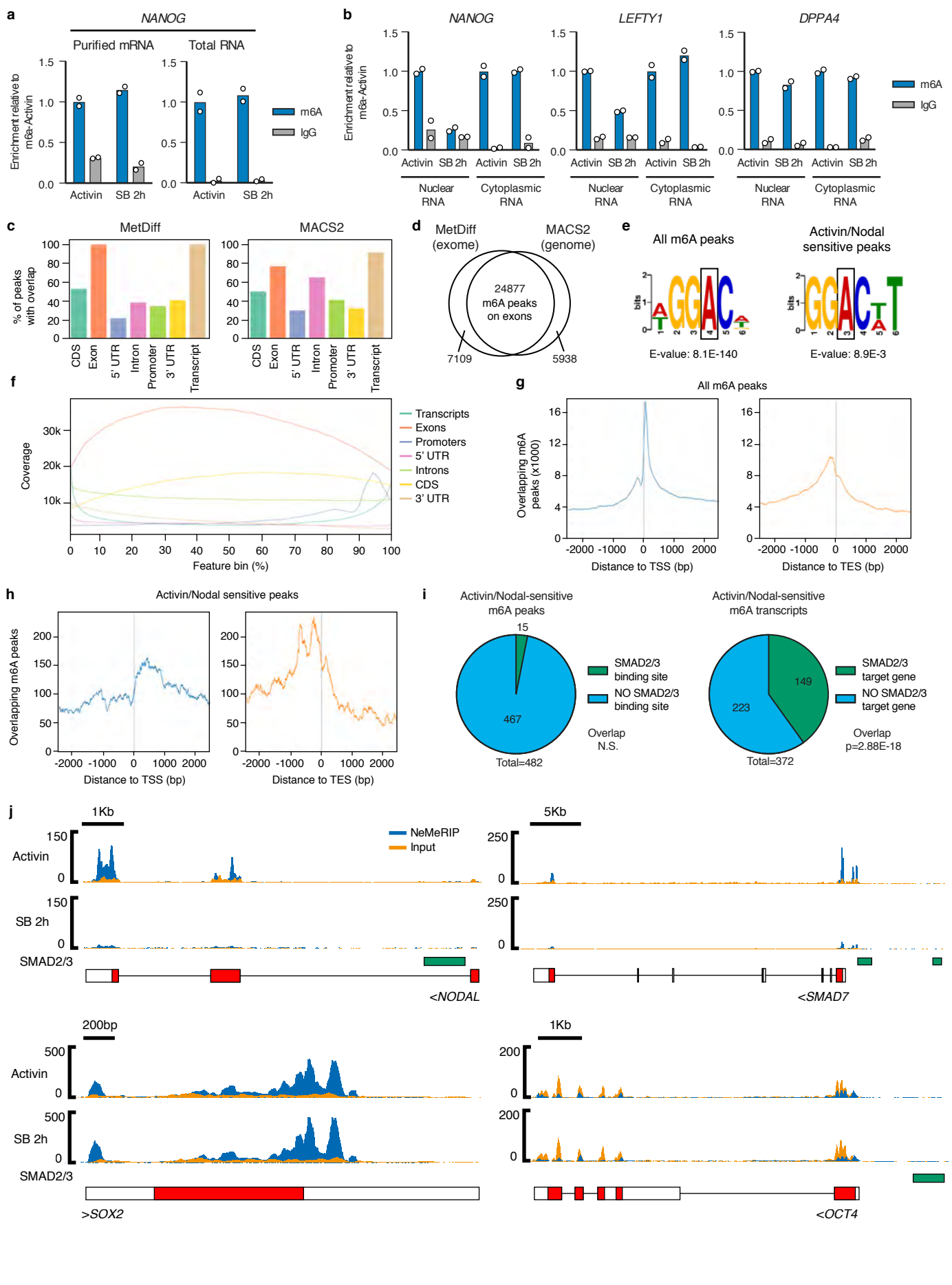

a
Input
IP Pluri
IP Endo
Pluri Endo
IgG S2/3
IgG S2/3
EP300
CREBBP
SETDB1
FOXH1
FOXH1 (short exp.)
SMAD2/3
b
AAVS1 i-shRNA
tetR hESCs
PuroR
Gene targeting
Puro selection
iKD hESCs
TET
KD hESCs
c
SMAD2
FOXH1
SETDB1
EP300
CREBBP
B2M
Expression relative to controls
SCR sh
B2M sh
sh1
sh2
sh3
SMAD2 iKD
FOXH1 iKD
SETDB1 iKD
EP300 iKD
CREBBP iKD
CTR
TET
d
Control iKD
SCR sh
B2M sh
SMAD2 iKD
SMAD2 sh1
SMAD2 sh2
FOXH1 iKD
FOXH1 sh1
FOXH1 sh2
SETDB1 iKD
SETDB1 sh1
SETDB1 sh2
EP300 iKD
EP300 sh1
EP300 sh2
CREBBP iKD
CREBBP sh1
e
SCR iKD
SMAD2 iKD
FOXH1 iKD
SETDB1 iKD
EP300 iKD
CREBBP iKD
CTR
TET
NANOG
DAPI
f
-1.5
0
2
Log2FC
SCR
B2M
SMAD2
FOXH1
SETDB1
EP300
CREBBP
NANOG
OCT4
NODAL
LEFTY1
SOX17
EOMES
T
MESP1
GBX2
SIP1

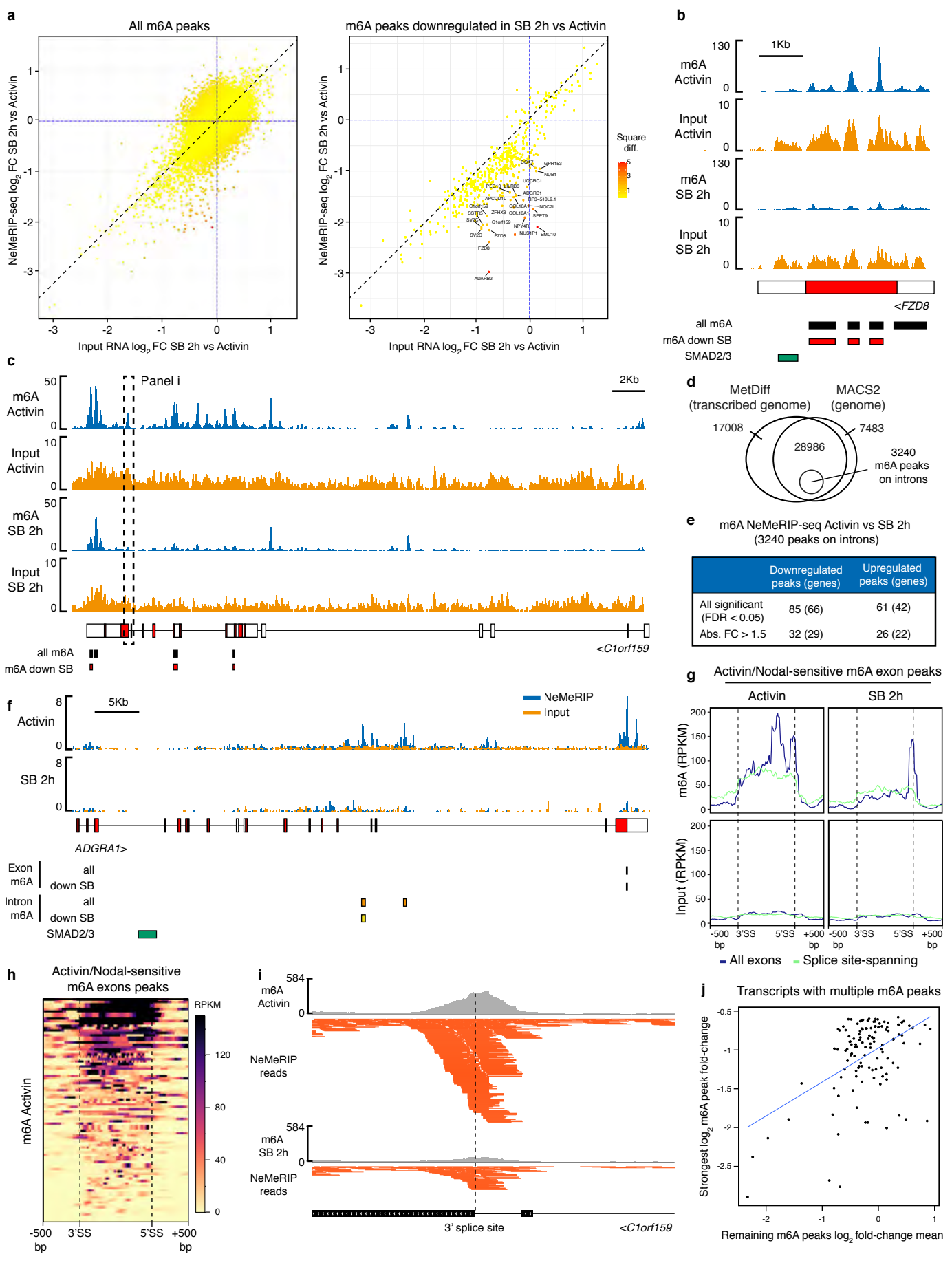


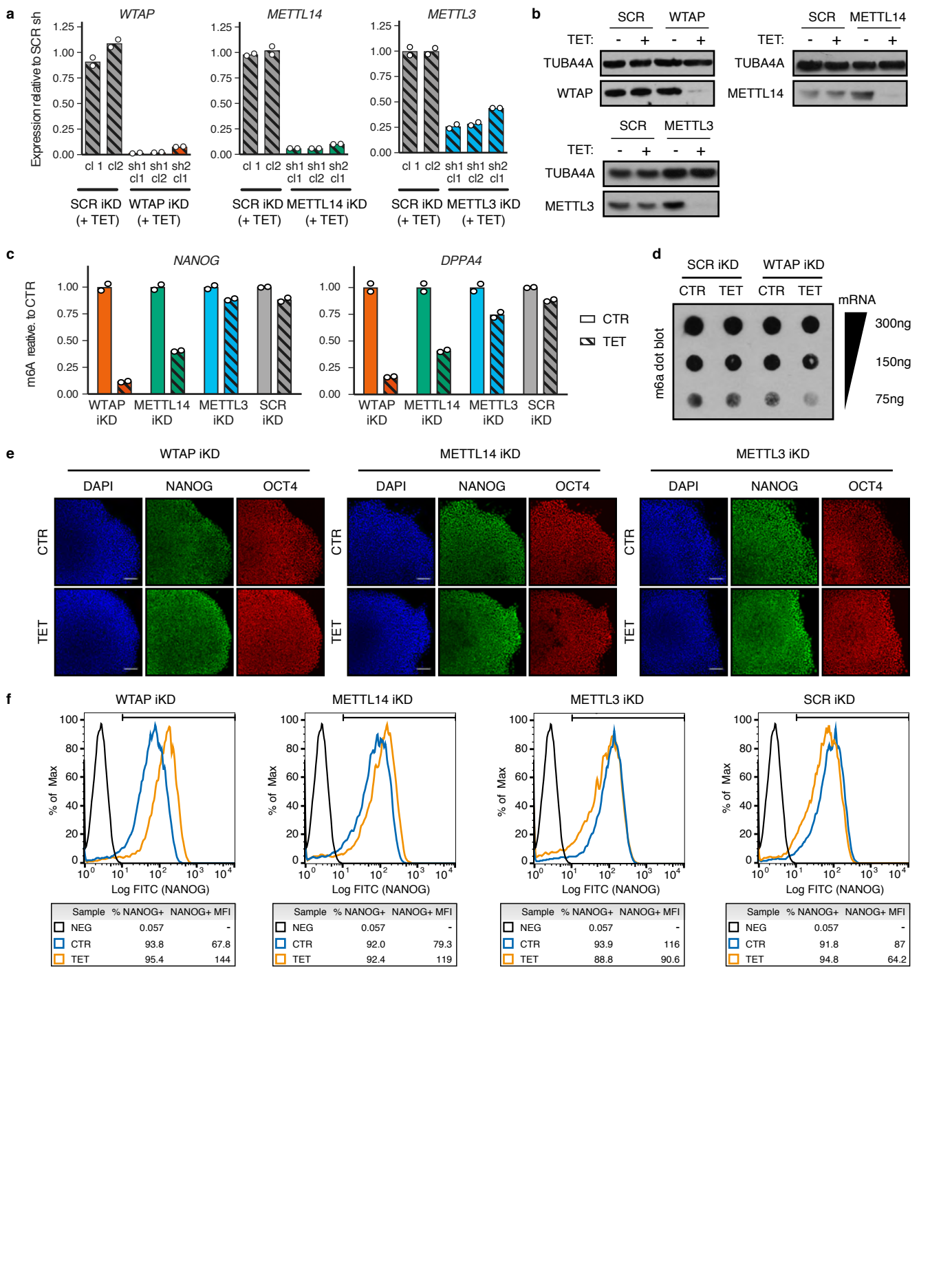

d

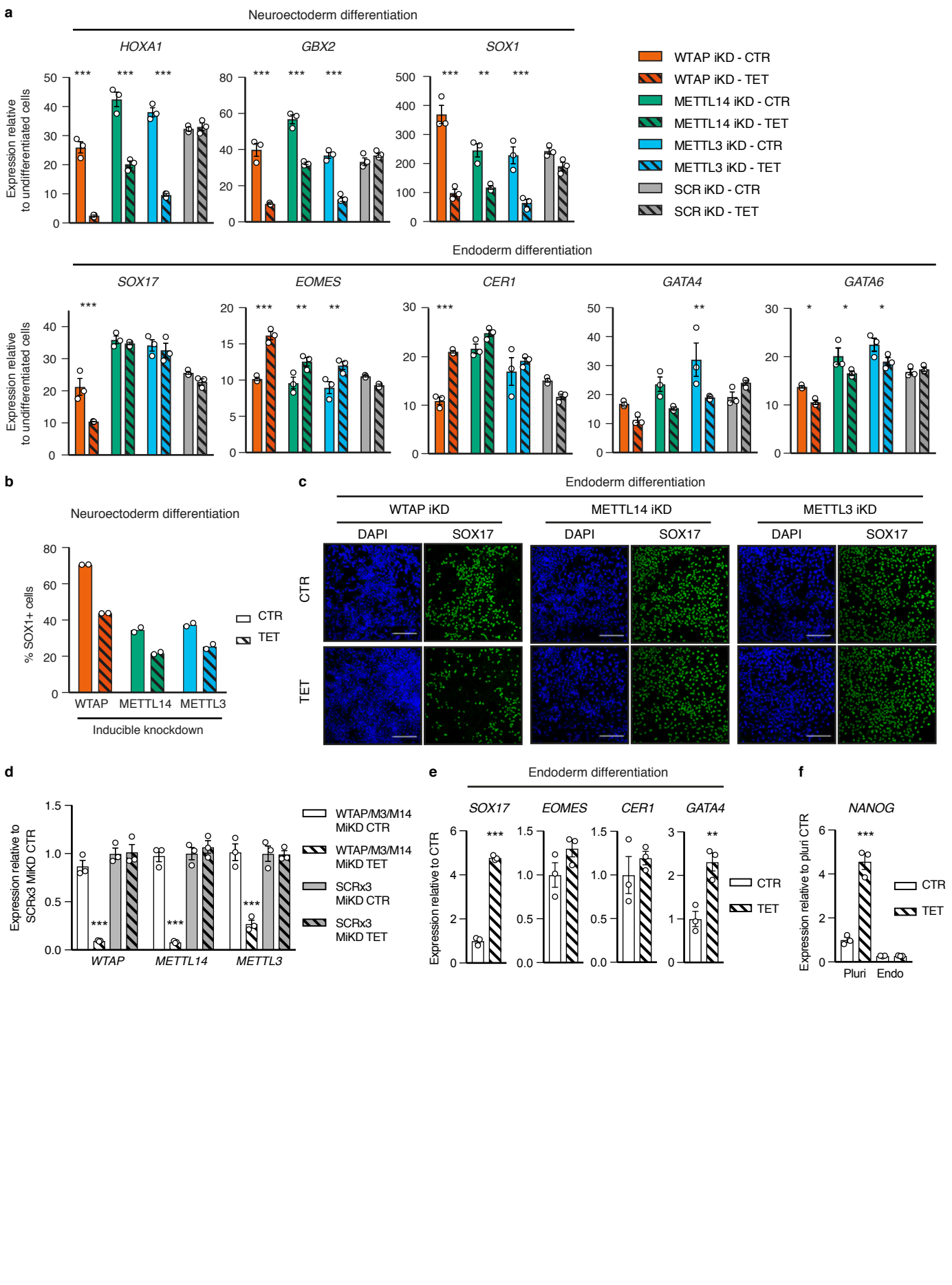
| | Pluripotency regulator | Effect of iKD (hPSCs) | Endoderm regulator |
|---|---|---|---|
| SMAD2 | +++ | ↓E ↑N ↑M | +++ |
| FOXH1 | ± (morphology?) | ↓E ↑N | ++ |
| SETDB1 | + | ↑N | ++ |
| EP300 | ++ | ↑E | + |
| CREBBP | ++ (post-transcr?) | ↑E | - |

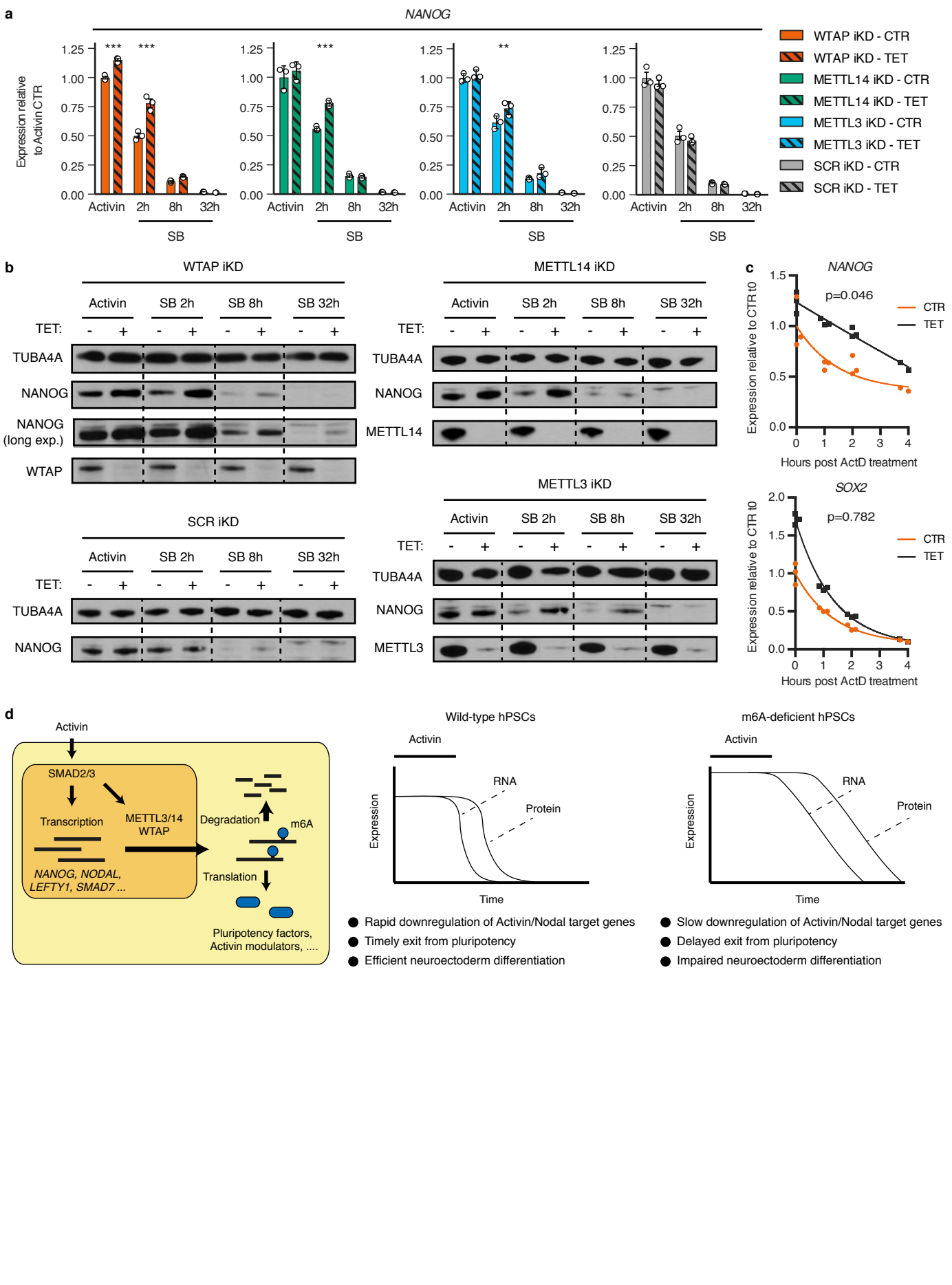

a
Immunoprecipitation in hESCs
IP
Input
IgG
M3
M14
METTL14
METTL3
SMAD2/3
b
Immunoprecipitation in hiPSCs
IP
Input
IgG
M3
M14
METTL14
METTL3
SMAD2/3
c
Activin
SB 1h
IP
IP
Input
IgG
S2/3
Input
IgG
S2/3
METTL14
WTAP
SMAD2/3
d
NANOG
Expression relative to H9 Activin
H9
NANOG OE
Activin
SB 2h
SB 48h
e
RIP in NANOG overexpressing hESCs
RIP enrichment relative to IgG Activin
WTAP
IgG
NANOG
LEFTY1
FZD8
NOC2L1
RPLP0
S2/3
Activin
SB 2h
f
NANOG
LEFTY1
NODAL
CER1
ChIP enrichment relative to IgG
SMAD2/3
METTL3
METTL14
WTAP
IgG
Activin
SB 2h

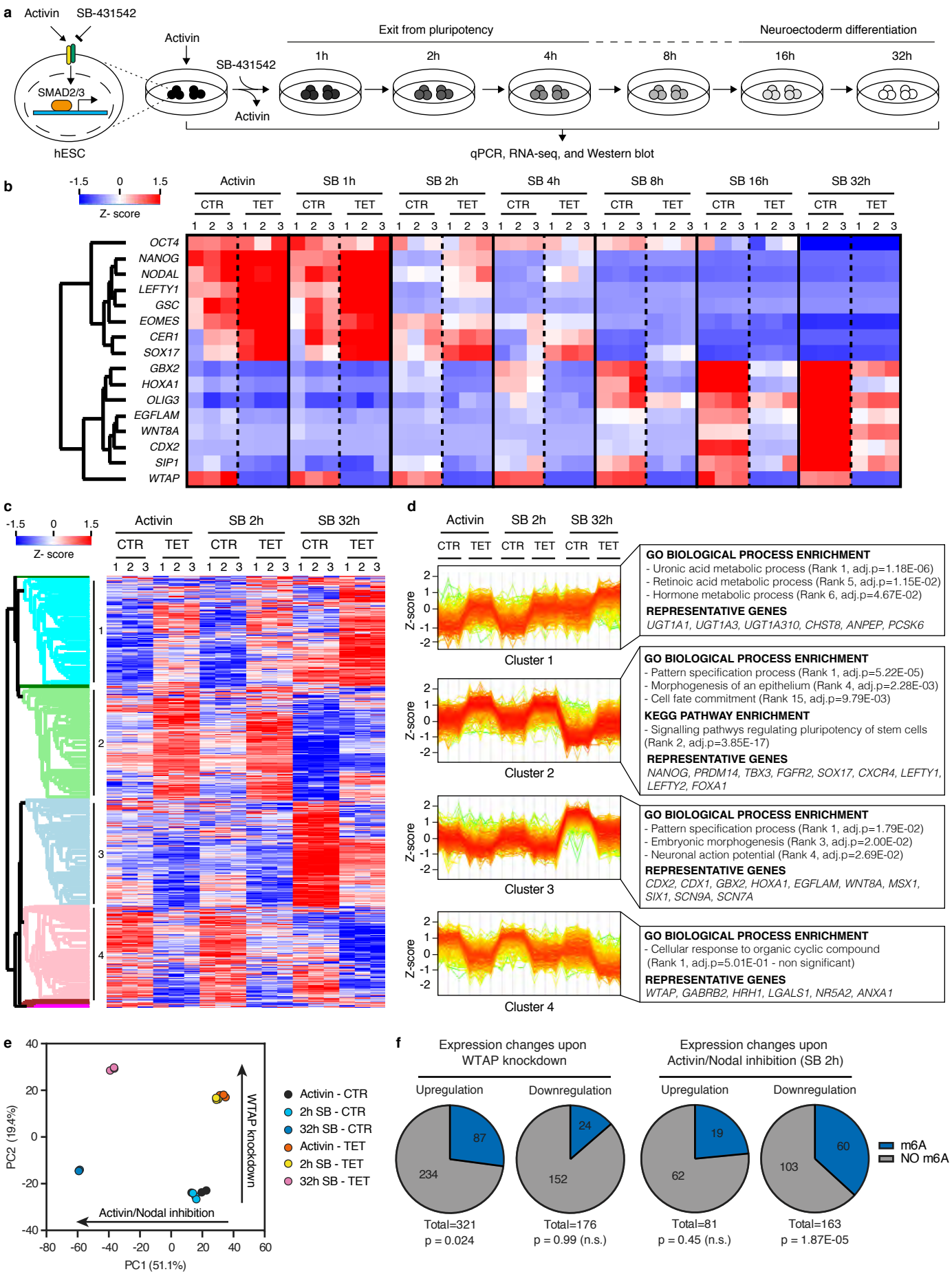

a
NANOG
Purified mRNA
Total RNA
Enrichment relative to m6a-Activin
m6A
IgG
Activin
SB 2h
b
NANOG
LEFTY1
DPPA4
Nuclear RNA
Cytoplasmic RNA
c
MetDiff
MACS2
% of peaks with overlap
CDS
Exon
5' UTR
Intron
Promoter
3' UTR
Transcript
d
MetDiff (exome)
MACS2 (genome)
24877 m6A peaks on exons
7109
5938
e
All m6A peaks
Activin/Nodal sensitive peaks
E-value: 8.1E-140
E-value: 8.9E-3
f
Coverage
Feature bin (%)
Transcripts
Exons
Promoters
5' UTR
Introns
CDS
3' UTR
g
All m6A peaks
Overlapping m6A peaks (x1000)
Distance to TSS (bp)
Distance to TES (bp)
h
Activin/Nodal sensitive peaks
Overlapping m6A peaks
Distance to TSS (bp)
Distance to TES (bp)
i
Activin/Nodal-sensitive m6A peaks
15
467
Total=482
SMAD2/3 binding site
NO SMAD2/3 binding site
Overlap N.S.
Activin/Nodal-sensitive m6A transcripts
149
223
Total=372
SMAD2/3 target gene
NO SMAD2/3 target gene
Overlap p=2.88E-18
j
1Kb
5Kb
NeMeRIP
Input
Activin
SB 2h
SMAD2/3
<NODAL
<SMAD7
200bp
1Kb
>SOX2
<OCT4


a
All m6A peaks
m6A peaks downregulated in SB 2h vs Activin
NeMeRIP-seq log2 FC SB 2h vs Activin
Input RNA log2 FC SB 2h vs Activin
Square diff.
b
1Kb
m6A Activin
Input Activin
m6A SB 2h
Input SB 2h
<FZD8
all m6A
m6A down SB
SMAD2/3
c
Panel i
2Kb
m6A Activin
Input Activin
m6A SB 2h
Input SB 2h
<C1orf159
all m6A
m6A down SB
d
MetDiff (transcribed genome)
MACS2 (genome)
17008
28986
7483
3240 m6A peaks on introns
e
m6A NeMeRIP-seq Activin vs SB 2h (3240 peaks on introns)
Downregulated peaks (genes)
Upregulated peaks (genes)
All significant (FDR < 0.05)
85 (66)
61 (42)
Abs. FC > 1.5
32 (29)
26 (22)
f
5Kb
NeMeRIP
Input
Activin
SB 2h
<ADGRA1
Exon m6A
Intron m6A
all
down SB
SMAD2/3
g
Activin/Nodal-sensitive m6A exon peaks
Activin
SB 2h
m6A (RPKM)
Input (RPKM)
-500 bp
3'SS
5'SS
+500 bp
All exons
Splice site-spanning
h
Activin/Nodal-sensitive m6A exons peaks
m6A Activin
RPKM
-500 bp
3'SS
5'SS
+500 bp
i
m6A Activin
NeMeRIP reads
m6A SB 2h
NeMeRIP reads
3' splice site
<C1orf159
j
Transcripts with multiple m6A peaks
Strongest log2 m6A peak fold-change
Remaining m6A peaks log2 fold-change mean


a
WTAP
METTL14
METTL3
Expression relative to SCR sh
1.25
1.00
0.75
0.50
0.25
0.00
cl 1 cl2 sh1 cl1 sh1 cl2 sh2 cl1
SCR iKD (+ TET)
WTAP iKD (+ TET)
METTL14 iKD (+ TET)
METTL3 iKD (+ TET)
b
SCR WTAP
TET: - + - +
TUBA4A
WTAP
SCR METTL14
TET: - + - +
TUBA4A
METTL14
SCR METTL3
TET: - + - +
TUBA4A
METTL3
c
NANOG
DPPA4
m6A reative. to CTR
1.00
0.75
0.50
0.25
0.00
WTAP iKD
METTL14 iKD
METTL3 iKD
SCR iKD
CTR
TET
d
SCR iKD
WTAP iKD
CTR TET CTR TET
m6a dot blot
mRNA
300ng
150ng
75ng
e
WTAP iKD
METTL14 iKD
METTL3 iKD
DAPI NANOG OCT4
CTR
TET
f
WTAP iKD
METTL14 iKD
METTL3 iKD
SCR iKD
% of Max
Log FITC (NANOG)
Sample % NANOG+ NANOG+ MFI
NEG 0.057 -
CTR 93.8 67.8
TET 95.4 144
NEG 0.057 -
CTR 92.0 79.3
TET 92.4 119
NEG 0.057 -
CTR 93.9 116
TET 88.8 90.6
NEG 0.057 -
CTR 91.8 87
TET 94.8 64.2


a
Neuroectoderm differentiation
HOXA1
GBX2
SOX1
Expression relative to undifferentiated cells
WTAP iKD - CTR
WTAP iKD - TET
METTL14 iKD - CTR
METTL14 iKD - TET
METTL3 iKD - CTR
METTL3 iKD - TET
SCR iKD - CTR
SCR iKD - TET
Endoderm differentiation
SOX17
EOMES
CER1
GATA4
GATA6
b
Neuroectoderm differentiation
% SOX1+ cells
CTR
TET
WTAP
METTL14
METTL3
Inducible knockdown
c
Endoderm differentiation
WTAP iKD
METTL14 iKD
METTL3 iKD
DAPI
SOX17
CTR
TET
d
Expression relative to SCRx3 MiKD CTR
WTAP/M3/M14 MiKD CTR
WTAP/M3/M14 MiKD TET
SCRx3 MiKD CTR
SCRx3 MiKD TET
WTAP
METTL14
METTL3
e
Endoderm differentiation
SOX17
EOMES
CER1
GATA4
Expression relative to CTR
CTR
TET
f
NANOG
Expression relative to pluri CTR
CTR
TET
Pluri
Endo


a
NANOG
Expression relative to Activin CTR
1.25
1.00
0.75
0.50
0.25
0.00
***
***
***
**
Activin
2h
8h
32h
SB
WTAP iKD - CTR
WTAP iKD - TET
METTL14 iKD - CTR
METTL14 iKD - TET
METTL3 iKD - CTR
METTL3 iKD - TET
SCR iKD - CTR
SCR iKD - TET
b
WTAP iKD
METTL14 iKD
METTL3 iKD
SCR iKD
Activin
SB 2h
SB 8h
SB 32h
TET:
- + - + - + - +
TUBA4A
NANOG
NANOG (long exp.)
WTAP
METTL14
METTL3
c
NANOG
p=0.046
SOX2
p=0.782
Expression relative to CTR t0
Hours post ActD treatment
CTR
TET
d
Activin
SMAD2/3
Transcription
METTL3/14
WTAP
NANOG, NODAL, LEFTY1, SMAD7 ...
Degradation
m6A
Translation
Pluripotency factors, Activin modulators, ....
Wild-type hPSCs
m6A-deficient hPSCs
RNA
Protein
Expression
Time
Rapid downregulation of Activin/Nodal target genes
Timely exit from pluripotency
Efficient neuroectoderm differentiation
Slow downregulation of Activin/Nodal target genes
Delayed exit from pluripotency
Impaired neuroectoderm differentiation


a
Activin
SB-431542
SMAD2/3
hESC
Activin
SB-431542
Activin
Exit from pluripotency
Neuroectoderm differentiation
1h
2h
4h
8h
16h
32h
qPCR, RNA-seq, and Western blot
b
-1.5
0
1.5
Z- score
Activin
SB 1h
SB 2h
SB 4h
SB 8h
SB 16h
SB 32h
CTR
TET
1 2 3
OCT4
NANOG
NODAL
LEFTY1
GSC
EOMES
CER1
SOX17
GBX2
HOXA1
OLIG3
EGFLAM
WNT8A
CDX2
SIP1
WTAP
c
-1.5
0
1.5
Z- score
Activin
SB 2h
SB 32h
CTR
TET
1 2 3
1
2
3
4
d
Activin
SB 2h
SB 32h
CTR
TET
Z-score
Cluster 1
GO BIOLOGICAL PROCESS ENRICHMENT
- Uronic acid metabolic process (Rank 1, adj.p=1.18E-06)
- Retinoic acid metabolic process (Rank 5, adj.p=1.15E-02)
- Hormone metabolic process (Rank 6, adj.p=4.67E-02)
REPRESENTATIVE GENES
UGT1A1, UGT1A3, UGT1A310, CHST8, ANPEP, PCSK6
Cluster 2
GO BIOLOGICAL PROCESS ENRICHMENT
- Pattern specification process (Rank 1, adj.p=5.22E-05)
- Morphogenesis of an epithelium (Rank 4, adj.p=2.28E-03)
- Cell fate commitment (Rank 15, adj.p=9.79E-03)
KEGG PATHWAY ENRICHMENT
- Signalling pathwys regulating pluripotency of stem cells (Rank 2, adj.p=3.85E-17)
REPRESENTATIVE GENES
NANOG, PRDM14, TBX3, FGFR2, SOX17, CXCR4, LEFTY1, LEFTY2, FOXA1
Cluster 3
GO BIOLOGICAL PROCESS ENRICHMENT
- Pattern specification process (Rank 1, adj.p=1.79E-02)
- Embryonic morphogenesis (Rank 3, adj.p=2.00E-02)
- Neuronal action potential (Rank 4, adj.p=2.69E-02)
REPRESENTATIVE GENES
CDX2, CDX1, GBX2, HOXA1, EGFLAM, WNT8A, MSX1, SIX1, SCN9A, SCN7A
Cluster 4
GO BIOLOGICAL PROCESS ENRICHMENT
- Cellular response to organic cyclic compound (Rank 1, adj.p=5.01E-01 - non significant)
REPRESENTATIVE GENES
WTAP, GABRB2, HRH1, LGALS1, NR5A2, ANXA1
e
PC2 (19.4%)
PC1 (51.1%)
WTAP knockdown
Activin/Nodal inhibition
Activin - CTR
2h SB - CTR
32h SB - CTR
Activin - TET
2h SB - TET
32h SB - TET
f
Expression changes upon WTAP knockdown
Expression changes upon Activin/Nodal inhibition (SB 2h)
Upregulation
Downregulation
87
234
Total=321
p = 0.024
24
152
Total=176
p = 0.99 (n.s.)
19
62
Total=81
p = 0.45 (n.s.)
60
103
Total=163
p = 1.87E-05
m6A
NO m6A

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

Bertero A. et al.,

# Supplementary Information

## Supplementary Discussion

### Optimization of the SMAD2/3 co-immunoprecipitation protocol

The first step towards defining the SMAD2/3 interactome was to identify a co-immunoprecipitation protocol allowing specific identification of the greatest number of interactors following mass spectrometry analysis. Indeed, the biochemical conditions used for protein extraction and immunoprecipitation have a profound effect on the stability of various protein-protein interactions[71]. Of note, SMAD2 and SMAD3 interact with several of their known binding partners through a set of contiguous hydrophobic patches, referred to as the "hydrophobic corridor", which is located on the surface of the MH2 domain[72]. Therefore, biochemical conditions that stabilize hydrophobic interactions might be preferable. To test this, we compared two SMAD2/3 co-immunoprecipitation (co-IP) methods that are expected to preferentially preserve different types of protein-protein interactions.

First, we tested the protocol that we had recently used to demonstrate the interaction of SMAD2/3 with COMPASS complexes[10]. This method, which we named co-IP1, relies on an isotonic buffer with low concentration of a mild detergent (0.1% Tween-20) both to solubilize nuclear proteins and to minimize background binding during immunoprecipitation. As such, these biochemical conditions are likely to preserve salt-sensitive hydrophilic bonds, while the presence of detergent might partially interfere with hydrophobic interactions. Secondly, we developed a different protocol that avoids the use of detergent and in which nuclear proteins are extracted using a high-salt buffer followed by dialysis of the lysate to re-adjust the salt content to physiological levels (co-IP2). In this case, hydrophobic interactions should be better preserved due to the lack of detergent. However, the high-salt conditions used for the nuclear extraction could disrupt certain hydrophilic bonds, which might only be partially re-established following dialysis.

Initial comparisons in hESCs demonstrated that both conditions allowed the detection of well-characterized SMAD2/3 binding factors by Western blot, with co-IP2 being slightly more efficient than co-IP1 (Extended Data Fig. 1a). We therefore performed small-scale pilot mass spectrometry analyses of SMAD2/3 co-IPs from hESCs using both methods (Extended Data Fig. 1b). Remarkably, co-IP2 allowed the identification of a larger number of SMAD2/3 interacting proteins following statistical analysis (23, compared to 12 for co-IP1; Extended Data Fig. 1b). Interestingly, roughly half of the proteins significantly enriched in co-IP1 experiments were also identified as specific binders in co-IP2 ones. On the other side, co-IP2-specific interactors included important transcription factors (such as SOX13, ETV6, and SMAD4), epigenetic regulators (like SETDB1 and ATF7IP), and RNA-binding proteins (for instance WTAP and CPSF6). Overall, these results showed that the co-IP2 protocol is more suitable for the large-scale analysis of SMAD2/3 interacting proteins. We therefore chose this method for subsequent experiments.

**Functional roles of the SMAD2/3 interactome**

Having identified the SMAD2/3 interactome in hPSCs, we decided to validate its functional role in pluripotency and endoderm differentiation. First, we focused on selected transcriptional and epigenetic cofactors (the transcription factor FOXH1, the histone acetyltransferases EP300 and CREBBP, and the histone methyltransferase SETDB1), as the function of these proteins in hPSCs is not fully understood. We confirmed their interactions with SMAD2/3 by co-IPs followed by Western Blot (Extended Data Fig. 2a). We then took advantage of our recently established OPTimized inducible gene KnockDown system (OPTiKD[27]; Extended Data Fig. 2b) to decrease the expression of these factors in hESCs and during differentiation (Extended Data Fig. 2c and 3a). Knockdown of SMAD2 was used as a positive control in these experiments as this factor is necessary for both pluripotency and endoderm specification[30,73].

Interestingly, knockdown of FOXH1 did not result in overt hESC differentiation, while this gene was required for endoderm differentiation (Extended Data Fig. 2d-f and 3b-d). This indicates that this well-known SMAD2/3 co-factor regulates only a specific subset of the transcriptional responses to Activin/Nodal signalling and it is predominantly involved in the expression of endoderm genes. This is in agreement with previous ChIP-seq results, which showed that FOXH1 and SMAD2/3 only weakly colocalize in pluripotent cells, while their genomic binding largely overlaps during endoderm differentiation[74]. Finally, loss of Foxh1 in the mouse embryo does not cause overt defects in the post-implantation epiblast, while it specifically impairs

patterning of the anterior primitive streak and formation of the node, prechordal mesoderm, notochord, and definitive endoderm[75–77].

In contrast to this, decreased expression of SETDB1, EP300, or CREBBP induced hESC differentiation, while having only moderate (SETDB1), little (EP300), or no effect (CREBBP) on endoderm specification (Extended Data Fig. 2d-f and 3b-d). Of note, gastrulation is not affected in mice knockout for Ep300 or Crebbp, which only show later embryonic defects such as heart malformations, defective neurulation, and impaired haematopoiesis and vasculogenesis[78–81]. This suggests that EP300 and CREBBP might be redundant during Activin/Nodal-induced endoderm specification, either because they compensate for each other or because other epigenetic regulators play a more significant role (for instance the COMPASS complexes[10]). Further studies involving conditional single and double conditional knockout for EP300 and CREBBP during hPSC differentiation will be required to clarify their role in early cell-fate choices.

Moving beyond the functional validation presented here, our data show that SMAD2/3 interacts with more than a dozen of different transcription factors and a similar number of epigenetic modifiers. Aside from well-known SMAD2/3 cofactors (such as SMAD4, SKI, and SNON), most of these proteins have never been previously reported to interact with SMAD2/3. These include multiple transcription factors (for instance ETV6, NFAT5, and SOX13) whose role in hPSCs is unknown despite being crucial for other developmental processes[82–86]. We anticipate that future studies will take advantage of the dataset we present to further dissect the transcriptional and epigenetic regulations involving SMAD2/3 in hPSCs.

To our surprise, the interactome of SMAD2/3 proved remarkably similar in undifferentiated hPSCs and hPSCs differentiating into endoderm (Extended Data Fig. 1e). Nevertheless, a few factors appear to differentially bind to SMAD2/3 in each condition. The most remarkable example is FOXH1, which as discussed above functionally interacts with SMAD2/3 to regulate expression of endoderm. Overall, only limited differences in the SMAD2/3 interactome could be sufficient to substantially modify the outcome of Activin/Nodal signalling in hPSCs. On the other hand, it is possible that at later stages of hPSC differentiation the changes in the SMAD2/3 interactome might become more significant, with novel partners such as EOMES driving yet other SMAD2/3-dependent transcriptional responses[13].

**Optimization of the m6A methylated RNA immunoprecipitation (MeRIP) experiments**

Since our data showed that SMAD2/3 physically interacts with the m6A methyltransferase complex (Figs. 1 and 2a-d), we hypothesized that Activin/Nodal signalling might regulate m6A deposition. To test this notion, we performed m6A methylated RNA immunoprecipitation (MeRIP) followed by qPCR to monitor the level of m6A onto Activin/Nodal-regulated transcripts following short-term signalling inhibition. Interestingly, treatment of hESCs for 2h with the Activin/Nodal inhibitor SB-431542 (SB) decreased m6A levels of specific nuclear transcripts (such as *NANOG* and *LEFTY1*, but not *DPPA4*), while cytoplasmic mRNAs were unaffected (Extended Data Fig. 5a-b). This suggested that SMAD2/3 might promote m6A deposition onto certain transcripts at the nuclear level, in agreement with its known localization. Therefore, in our following experiments we decided to focus on nuclear transcript by performing nuclear-enriched MeRIP, a method which we named NeMeRIP (Fig. 2e and Extended Data Fig. 5c-j).

**RNA-seq experiments in WTAP inducible knockdown cells.**

Having shown that WTAP regulates expression of several Activin/Nodal target genes (Fig. 3 and Extended Data Figs. 9 and 10a-b), we validated these findings at a genome-wide level. Therefore, we performed RNA-seq in WTAP inducible knockdown cells cultured in presence of Activin or following Activin/Nodal signalling inhibition (Extended Data Fig. 10c-e) These results confirmed that knockdown of WTAP globally alters the response to Activin/Nodal signalling by: (1) upregulating a large cohort of developmental regulators whose expression is maintained by Activin/Nodal in the pluripotent state, and by delaying the downregulation of such genes upon Activin/Nodal inhibition (cluster 2); (2) impairing the upregulation of neuroectoderm genes induced following inhibition of Activin/Nodal (cluster 3). Additionally, WTAP knockdown resulted in up- and downregulation of additional factors whose expression is largely independent from Activin/Nodal signalling, and which are not associated to developmental regulations (cluster 1 and 4, respectively). This showed that WTAP has additional functions other than modulating the response to Activin/Nodal signalling, in agreement with its role as a general regulator of the epitranscriptome. Overall, these findings demonstrated that WTAP has an important role in modulating the gene expression network controlled by Activin/Nodal signalling.

## Supplementary Information specific references


71. Sambrook, J. & Russell, D. W. Identification of associated proteins by coimmunoprecipitation. *CSH Protoc.* **2006,** (2006).
72. Massagué, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev.* **19,** 2783–810 (2005).
73. Sakaki-Yumoto, M., Liu, J., Ramalho-Santos, M., Yoshida, N. & Derynck, R. Smad2 Is essential for maintenance of the human and mouse primed pluripotent stem cell state. *J. Biol. Chem.* **288,** 18546–18560 (2013).
74. Kim, S. W. *et al.* Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* **357,** 492–504 (2011).
75. von Both, I. *et al.* Foxh1 is essential for development of the anterior heart field. *Dev. Cell* **7,** 331–45 (2004).
76. Yamamoto, M. *et al.* The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev.* **15,** 1242–56 (2001).
77. Hoodless, P. a *et al.* FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev.* **15,** 1257–71 (2001).
78. Oike, Y. *et al.* Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. *Blood* **93,** 2771–9 (1999).
79. Kung, A. L. *et al.* Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev.* **14,** 272–7 (2000).
80. Yao, T. P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93,** 361–72 (1998).
81. Shikama, N. *et al.* Essential function of p300 acetyltransferase activity in heart, lung and small intestine formation. *EMBO J.* **22,** 5175–85 (2003).
82. Tsuzuki, S. & Seto, M. TEL (ETV6)-AML1 (RUNX1) initiates self-renewing fetal pro-B cells in association with a transcriptional program shared with embryonic stem cells in mice. *Stem Cells* **31,** 236–47 (2013).
83. Ford, A. M. *et al.* The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. *J. Clin. Invest.* **119,** 826–36 (2009).
84. Melichar, H. J. *et al.* Regulation of gammadelta versus alphabeta T lymphocyte differentiation by the transcription factor SOX13. *Science* **315,** 230–3 (2007).
85. Wang, Y., Bagheri-Fam, S. & Harley, V. R. SOX13 is up-regulated in the developing mouse neuroepithelium and identifies a sub-population of differentiating neurons. *Dev. brain Res.* **157,** 201–8 (2005).
86. Berga-Bolaños, R., Alberdi, M., Buxadé, M., Aramburu, J. & López-Rodríguez, C. NFAT5 induction by the pre-T-cell receptor serves as a selective survival signal in T-lymphocyte development. *PNAS* **110,** 16091–6 (2013).