$N^\circ$-methyladenosine regulates the stability of RNA:DNA hybrids in human cells

Abdulkadir Abakir1, Tom C. Giles2,3, Agnese Cristini4, Jeremy M. Foster5, Nan Dai5, Marta Starczak6, Alejandro Rubio-Roldan7, Miaomiao Li8,9, Maria Eleftheriou1, James Crutchley1, Luke Flatt1, Lorraine Young1, Daniel J. Gaffney10, Chris Denning1, Bjørn Dalhus8,11, Richard D. Emes12, Daniel Gackowski5, Jose L. Garcia-Perez7,13, Arne Klungland8,9, Natalia Gromak* and Alexey Ruzov*14

R-loops are nucleic acid structures formed by an RNA:DNA hybrid and unpaired single-stranded DNA that represent a source of genomic instability in mammalian cells14. Here we show that $N^\circ$-methyladenosine (m$^\circ$A) modification, contributing to different aspects of messenger RNA metabolism5,6, is detectable on the majority of RNA:DNA hybrids in human pluripotent stem cells. We demonstrate that m$^\circ$A-containing R-loops accumulate during G2/M and are depleted at G0/G1 phases of the cell cycle, and that the m$^\circ$A reader promoting mRNA degradation, YTHDF2 (ref. 7), interacts with R-loop-enriched loci in dividing cells. Consequently, YTHDF2 knock-out leads to increased R-loop levels, cell growth retardation and accumulation of $\gamma$H2AX, a marker for DNA double-strand breaks, in mammalian cells. Our results suggest that m$^\circ$A regulates accumulation of R-loops, implying a role for this modification in safeguarding genomic stability.

Dynamic methylation of adenosine in RNA ($N^\circ$-methyladenosine, m$^\circ$A) has been implicated in the regulation of different aspects of messenger RNA metabolism in mammals by numerous studies8–10. Although m$^\circ$A is abundant in eukaryotic transciptomes, its DNA counterpart, $N^\circ$-methyldeoxyadenosine (6mA) was previously thought to be restricted to unicellular organisms and only recently has been shown to exist in non-negligible quantities in metazoan DNA11–13. Despite the fact that 6mA is reportedly widespread in fungal genomes1, its prevalence in mammalian systems is currently poorly understood. This modification accumulates in pre-implantation pig embryos14; however, evidence for its presence in mouse tissues is contradictory15,16. In this study, we initially aimed to examine whether this mark is detectable in human cell lines by using a sensitive immunostaining method that we have previously employed to detect modified forms of cytosine in vertebrate models15.

To confirm that we can differentiate between m$^\circ$A-modified miRNAs and 6mA present on genomic DNA, we performed immunostaining of human pluripotent stem cells (hPSCs) by using previously validated anti-m$^\circ$A/6mA antibody15 without the DNA denaturation step required for the immunohistochemical detection of modified bases in gDNA1,13,14. In these experiments, we observed prominent m$^\circ$A staining that disappeared following pretreatment of the samples with RNase A (Supplementary Note). Next, we immunostained several human cell lines with the same antibody but after treatment of the samples with 4 M HCl, which allows denaturing of double-stranded nucleic acids and is routinely used for the detection of cytosine modifications and 6mA in gDNA1,13,14. Under these conditions, we also detected strong m$^\circ$A signal in both the nuclei and cytoplasm of hPSCs and cancer cell lines. Notably, high levels of m$^\circ$A staining were still evident in the mitotic chromatin in all our samples processed after RNase A treatment (Fig. 1a and Supplementary Note). To examine whether the mitotic staining we observed indicates the presence of 6mA in the human genome, we performed liquid chromatography–tandem mass spectrometry (LC–MS/MS) quantification of 6mA and modified forms of cytosine in the DNA of two hPSC lines either cultured under standard conditions or after enrichment for mitotic cells by using colcemid treatment15. Unlike the species of modified cytosine, 6mA was not detectable by LC–MS/MS in hPSCs under both experimental conditions even at low parts-per-million levels, suggesting that this modification, if present in hPSC genomes, occurs only at levels substantially lower than that of 5-formylcytosine (Fig. 1b). These results confirmed previously published liquid chromatography–mass spectrometry (LC–MS) data indicating the absence of 6mA in the genome of mouse embryonic stem cells and tissues11.

Attempting to explain the discrepancy between our LC–MS/MS data and immunostaining results, we hypothesized that the mitotic anti-m$^\circ$A/6mA antibody-specific signal was caused by the presence of this modification on the RNA component of R-loops1. R-loops are specific nucleic acid structures formed by an RNA:DNA hybrid and an unpaired single-stranded DNA that contribute to a number of important biological processes, ranging from transcriptional regulation to DNA repair, and represent a source of genomic instability in mammalian cells1–4. To test this hypothesis, we immunostained hPSCs by using m$^\circ$A antibody after treatment of the samples with Escherichia coli RNase H, an enzyme that specifically degrades RNA molecules present in RNA:DNA hybrids. Notably, mitotic m$^\circ$A staining significantly decreased or disappeared in the hPSCs

1Department of Stem Cell Biology, University of Nottingham, Nottingham, UK. 2Advanced Data Analysis Centre, University of Nottingham, Sutton Bonington, UK. 3Digital Research Service, University of Nottingham, Sutton Bonington, UK. 4Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. 5New England Biolabs, Inc., Ipswich, MA, USA. 6Department of Clinical Biochemistry, Nicolaus Copernicus University in Toruń, Bydgoszcz, Poland. 7GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain. 8Department of Microbiology, Oslo University Hospital, Oslo, Norway. 9Department of Molecular Medicine, University of Oslo, Oslo, Norway. 10Wellcome Sanger Institute, Cambridge, UK. 11Department of Medical Biochemistry, University of Oslo, Oslo, Norway. 12School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonnington, UK. 13MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK.

e-mail: arne.klungland@medisin.uio.no; natalia.gromak@path.ox.ac.uk; alexey.ruzov@nottingham.ac.uk
pretreated with RNase H, corroborating the presence of this modification on the RNA strand of RNA:DNA hybrids (Fig. 1c,d and Supplementary Note). Confirming our immunostaining results, we also detected release of m^6A (but not of ribo-5-methylcytidine, ribo-m^5C) to filtrate by stable-isotope dilution ultra-performance liquid chromatography with tandem mass spectrometry (SID–UPLC–MS/MS) following treatment of hPSC-derived nucleic acids with RNase H (Fig. 1e,f and Supplementary Note). Overall, these results strongly suggested that m^6A modification is associated with the RNA components of RNA:DNA hybrids in hPSCs.

To examine the genomic distribution of m^6A-marked RNA:DNA hybrids, we modified a previously published DNA:RNA immunoprecipitation technique (DRIP, referred to here as S9.6 DRIP) by replacing anti-RNA:DNA hybrid S9.6 antibody with anti-m^6A antibody (designated here as m^6A DNA immunoprecipitation, m^6A DIP). After validation of this technique using synthetic spike-in RNA:DNA hybrids and single-stranded oligonucleotides (Extended Data Fig. 1a–d), we performed m^6A DIP in parallel with S9.6 DRIP coupled with high-throughput sequencing on human induced pluripotent stem cells (hiPSCs) (Fig. 2a and Extended Data Fig. 2). Although both types of immunoprecipitation resulted in the generation of large peak datasets, the majority of m^6A DIP and S9.6 DRIP peaks were not detectable in the control samples pretreated with RNase H, confirming that the presence of methylated adenosine is correlated with the RNA component of R-loops in hPSCs (Fig. 2b and Supplementary Note). Both m^6A and S9.6 peaks exhibited virtually identical distribution across various genomic features, and repetitive elements and were enriched in transcribed regions of the human genome (Figs. 2c and 3a and Supplementary Note). Despite the number of m^6A DIP peaks being approximately fourfold greater relative to S9.6 DRIP, both sets of peaks displayed an essentially complete overlap at the sequence level (Fig. 2b,d). Since the presence of both m^6A and S9.6 peaks was RNase H dependent, and the density of S9.6 DRIP reads was noticeably increased across the m^6A peaks that did not overlap with S9.6 peaks (Fig. 2c and Extended Data Fig. 3a), we concluded that the difference in peak numbers we observed was probably due to variation in sensitivity of the corresponding
antibodies and, therefore, our results imply that m6A marks most of the RNA:DNA hybrids in hPSCs. In line with this explanation, m6A DIP demonstrated approximately 3.6-fold more efficient enrichment for the synthetic m6A-containing RNA:DNA hybrid compared with S9.6 DRIP in our spike-in experiments (Extended Data Fig. 1b). We also observed a similar distribution of common m6A/S9.6- and m6A-only peaks amongst different genomic features (Extended Data Fig. 3b,c).

Since the RNase H-sensitive m6A immunostaining signal was particularly high in mitotic chromatin (Fig. 1a,c), we hypothesized...
that this modification may accumulate on RNA:DNA hybrids in a cell-cycle-specific manner. To examine the dynamics of m6A-containing R-loops during the cell cycle, we performed m6A DIP and S9.6 DRIP on G0/G1, S and G2/M flow cytometry-sorted hPSC populations (Extended Data Fig. 4a), followed by quantitative PCR (qPCR) of LINE-1 repeats and individual intronic sequences enriched in both m6A and S9.6 peaks (Fig. 3a and Extended Data Fig. 4b). These experiments demonstrated that RNA:DNA hybrids accumulate on LINE-1 retrotransposons during S phase, max out at G2/M and drastically decrease at G0/G1 phases of the cell cycle in hPSCs (Fig. 3b). Consistently, a recent study demonstrated that retrotransposition-active, LINE-1-derived messenger RNAs are enriched in cells exiting mitosis21. Intronic R-loops were found at high levels at both S and G2/M phases, but were also significantly depleted at G0/G1 phase (Fig. 3c,d). Importantly, these cell-cycle-specific changes were essentially equivalent in both m6A DIP and S9.6 DRIP, suggesting that m6A is present on RNA:DNA hybrids throughout all stages of the cell cycle (Fig. 3b–d). Notably, m6A DIP–qPCR enrichment substantially increased on the repetitive and intronic loci following small interfering RNA-mediated knockdown of RNase H1 in hPSCs (Extended Data Fig. 5a–d). Moreover, intronic and repetitive m6A DIP-containing sequences were also enriched in the two-round immunoprecipitation (S9.6 DRIP followed by m6A DIP or m6A RNA immunoprecipitation) procedures, further confirming the presence of m6A on the RNA components of R-loops (Extended Data Fig. 6a–e). In sum, these results suggested that the turnover rates of m6A-marked R-loops vary for cell cycle phases.

Given that deposition of m6A is known to affect the stability of mRNAs2,23, we hypothesized that this mark may also modulate the stability of R-loops. Since siRNA-mediated knockdown of m6A methyltransferase METTL3 led to accumulation of RNA:DNA hybrids in hPSCs (Extended Data Figs. 7a,b and 8a–f and Supplementary Note), we next inquired whether any of the previously characterized m6A reader proteins might interact with mitotic chromatin enriched in m6A-containing R-loops. First, we examined for the presence of m6A readers in proteins interacting with RNA:DNA hybrids immunoprecipitated from HeLa cells by using S9.6 antibody24. The analysis showed an enrichment of YTHDF1—a protein promoting translation of m6A-containing mRNAs25, HNRNPA2B1—a nuclear m6A reader previously implicated in mRNA processing26—and YTHDF2—an m6A-interacting protein that regulates degradation of cytoplasmic mRNAs27 as well as METTL3 in R-loop immunoprecipitation, suggesting that these proteins interact with RNA:DNA hybrids (Fig. 4a). Our subsequent immunostaining experiments showed that, while YTHDF1 exhibited predominantly cytoplasmic localization in both interphase and mitotic hPSCs (Fig. 4b,c) and HNRNPA2B1 was specifically excluded from chromatin during mitosis (Fig. 4d,e), YTHDF2 migrated to mitotic chromatin in dividing hPSCs (Fig. 4f,g). Moreover, the nuclear fraction of YTHDF2 exhibited a high degree of co-localization with RNA:DNA hybrids in interphase cells (Extended Data Fig. 9a–e). In line with this, we also observed preferential interaction of YTHDF2 with m6A-containing synthetic RNA:DNA substrates in electrophoretic mobility-shift assay (EMSA; Extended Data Fig. 9f,g and Supplementary Note), and in microscale thermophoresis (MST) analysis, which demonstrated that YTHDF2 shows comparable dissociation constant values for its interaction with m6A-marked single-stranded RNA and m6A-RNA:DNA duplexes in this assay (Fig. 4h). Furthermore, YTHDF2 chromatin immunoprecipitation (ChIP) showed that this m6A reader interacts with both LINE-1s and intronic genomic regions enriched in RNA:DNA hybrids in these cells (Extended Data Fig. 10a). In contrast, we did not observe any interaction of
**Fig. 4 | m^6A reader proteins interact with RNA:DNA hybrids.** a, Immunoblot of RNA:DNA hybrid protein co-immunoprecipitation (IP)-probed with indicated antibodies. Top1 and Lamin B1 serve as positive and negative controls, respectively, for R-loop immunoprecipitation. The experiments were repeated independently two times for METTL3 and three times for other proteins, with similar results. The blots were cropped. Full scans of the blots are shown in Source Data 1. b–c, Immunostaining of hiPSCs by using anti-YTHDF1 and anti-pH-H3 antibodies imaged at two different magnifications. d–e, Immunostaining of hiPSCs for HNRNPA2B1 and m^6A imaged at two different magnifications. f–g, Immunostaining of hiPSCs for YTHDF2 and pH-H3 imaged at two different magnifications. Merged views and YTHDF1/HNRNPA2B1/YTHDF2 channels (b,d,f) or merged views and individual channels (c,e,g) are shown. Locations of the views shown in c,e,g are marked with dashed rectangles in b,d,f. All scale bars, 10 μm. Experiments shown in b–g were repeated independently four times with similar results. h, Microscale thermophoresis binding curves for YTHDF2 interaction with m^6A-containing/non-modified RNA:DNA hybrid and m^6A-marked/non-modified ssRNA synthetic substrates. Binding is shown as fraction of protein bound as a function of substrate concentration. Binding curves are fitted to the data points from experiments for m^6A-containing (filled circles/triangles) and unmodified (open circles/triangles) substrates. Dissociation constant (K_d) values are shown for each of the interactions. Error bars show s.d., the center values are means, n=6 independent series of experiments.
Fig. 5 | YTHDF2 depletion leads to accumulation of R-loops, increased accretion of m^A on RNA:DNA hybrids and cell growth retardation.

a, Immunostaining of WT and YTHDF2 KO HAP1 for YTHDF2 and pH-H3. The experiments were repeated independently three times with similar results.

b, Immunostaining of WT and YTHDF2 KO HAP1 for R-loops alongside quantification of S9.6 nuclear signal. c, DRIP-qPCR of the indicated sequences performed on WT and YTHDF2 KO HAP1. RANBP17 and HECW1 downstream regions lacking DRIP peaks were used as controls. d, SID-UPLC-MS/MS quantification of m^A and ribo-m^C in S9.6 immunoprecipitation performed on WT and YTHDF2 KO HAP1 and normalized for dA or rA. RNase H-pretreated samples were used as controls. Data are means ± s.d., n = 5 (left) and n = 7 (right) measurements of four independent samples.

e, Growth curves of WT and YTHDF2 KO HAP1. f, Immunostaining of WT/Ythdf2 KO mNSCs for R-loops, alongside the quantification of S9.6 nuclear signal. g, DRIP and m^A DIP-qPCR of mouse LINE-1 ORF1 performed on WT and Ythdf2 KO mNSCs. Individual channels (a) or S9.6 channel (b,f) with merged views are shown. Scale bars, 10 μm. Data are means ± s.d., n = 3 independent experiments. The elements of the box plots (b,f) are: center line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all data; n, sample size, ***P < 0.0001. Significance was determined by unpaired two-tailed Student’s t-test.
**Fig. 6 | YTHDF2 depletion leads to elevated levels of H2AX phosphorylation in human and mouse cells.**

**a.** Representative images of WT and YTHDF2 KO HAP1 cells immunostained for γH2AX, and quantification of γH2AX signal intensity in these cells. Results of γH2AX ChIP-qPCR of the indicated sequences performed on WT and YTHDF2 KO HAP1. Generic primers amplifying Alu elements from the indicated families and evolutionarily young L1Hs were used. Data are means ± s.d., n = 3 independent experiments.

**b.** Representative images of WT/Ythdf2 KO embryonic brain cortex and mNSCs immunostained for γH2AX, alongside quantification of γH2AX signal intensity in these tissues/cells. Immunostaining of siCTL, siMETTL3 and siHNRNPA2B1 hPSCs for γH2AX, and quantification of γH2AX signal intensity in these cells. Representative images of YTHDF2 KO HAP1 cells transfected with GFP-RNase H1 and GFP-only expression constructs immunostained for γH2AX, alongside the quantification of γH2AX signal intensity in GFP-positive cells. P value is indicated. Examples of nuclei used for signal quantification are outlined by dotted shapes. Merged images and S9.6 channel views are shown in **a,c–e.** Scale bars, 10 μm. **a,c–e.** The elements of the box plots are: center line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all data; n, sample size, ***P < 0.0001. Significance was determined by unpaired two-tailed Student’s t-test, and no adjustments were made for multiple comparisons.
HNRNPA2B1 with LINE-1 repeats but were able to detect binding of this protein to R-loop-containing intronic regions in ChIP experiments (Extended Data Fig. 10b). Interestingly, although the recruitment of both these proteins to R-loop-containing loci was reduced following METTL3 knockdown, confirming their interaction with m6A in chromatin-bound RNAs (Extended Data Fig. 10c,d), the accumulation of YTHDF2 (but not of HNRNPA2B1) at LINE-1s and intronic loci was dramatically increased in siRNA-HF1 hPSCs, strongly suggesting the association of this m6A reader with R-loops in vivo (Extended Data Fig. 10e,f). To assess the functional importance of YTHDF2 migration to mitotic chromatin, we performed its siRNA-mediated depletion (siYTHDF2) in hPSCs. S9.6 DRIP and m6A DIP–qPCR showed a significant enrichment in both repetitive and individual intronic R-loop sequences in siYTHDF2 hPSCs relative to siCTRL cells (Extended Data Fig. 10g–i). To further confirm these results, we next assessed the levels of R-loops in YTHDF2 knockout (KO) HAP1 cells expressing a truncated version of this protein that does not co-localize with mitotic chromatin (Fig. 5a and Supplementary Note). These experiments showed both the elevated levels of S9.6 immunostaining and dramatic 5–50-fold increase in R-loops at Alu-S, Alu-Y, LINE-1s and intronic sequences in YTHDF2 KO compared with isogenic wild type (WT) parental HAP1 cells (Fig. 5b,c). Moreover, YTHDF2 depletion in HAP1 cells also resulted in increased acetylation of m6A on RNA:DNA hybrids (Fig. 5d) and cell growth retardation (Fig. 5e). Subsequent analysis of recently published Yhdf2 constitutive KO mouse-derived neural stem cells (mNSCs) confirmed these results, demonstrating increased levels of S9.6 immunostaining and accumulation of RNA:DNA hybrids in LINE-1 open reading frames following depletion of Yhdf2 in this system (Fig. 5f,g). In line with these results, YTHDF2 KO HAP1 cells displayed increased accumulation of a marker for DNA double-strand breaks, phosphorylated (ser139) histone variant H2AX (γH2AX)35, at both the nucleus-wide level and R-loop-enriched loci (Fig. 6a,b). Correspondingly, we also observed elevated levels of γH2AX staining in the cortex of Yhdf2 KO embryos and Yhdf2 KO mNSCs (Fig. 6c) as well as, to a lesser extent, in hPSCs following siRNA-mediated depletions of METTL3 and HNRNPA2B1 (Fig. 6d and Supplementary Note). Moreover, γH2AX intensity significantly decreased following overexpression of RNase H1 in YTHDF2 KO HAP1 cells (Fig. 6e). Overall, these results suggest that YTHDF2 prevents accumulation of m6A-containing RNA:DNA hybrids, contributing to inhibition of R-loop-dependent DNA damage in mammalian cells. Correspondingly, YTHDF2 has previously been identified as one of the factors promoting genomic stability in a genome-wide siRNA screen36.

The nature of the techniques we used for m6A mapping is limited by the specificity and sensitivity of the available antibody. Even so, our results show that m6A modification is present on the RNA within R-loops, potentially contributing to various aspects of their biology (Supplementary Note). In this context, YTHDF2-mediated regulation of RNA:DNA hybrids may represent a specific mechanism in the prevention of accumulation of co-transcriptional R-loops during mitosis. Together with previously described factors suppressing the formation of these structures37–39, YTHDF2 plays a role in safeguarding genomic stability.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0549-x.

Received: 23 January 2019; Accepted: 13 November 2019; Published online: 16 December 2019

References
1. Chédin, F. Nascent connections: R-loops and chromatin patterning. Trends Genet. 32, 828–838 (2016).
2. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell 63, 167–178 (2016).
3. Lu, W. T. et al. Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. Nat. Commun. 9, 532 (2018).
4. Santos-Pereira, J. M. & Aguilera, A. R-loops: new modulators of genome dynamics and function. Nat. Rev. Genet. 16, 583–597 (2015).
5. Yue, Y., Liu, J. & He, C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Nature 7481, 117–120 (2014).
6. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. Cell 169, 1187–1206 (2017).
7. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 7481, 134–155 (2015).
8. Zhang, G. et al. N6-methyladenine DNA modification in Drosophila. Cell 161, 893–906 (2015).
9. Greer, E. L. et al. DNA methylation on N6-adenine in C. elegans. Cell 161, 868–878 (2015).
10. Yu, F. et al. N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell 161, 879–892 (2015).
11. Mondo, S. J. et al. Widespread adenine N6-methylation of active genes in fungi. Nat. Genet. 49, 964–968 (2017).
12. Lee, M. et al. Abundant DNA N6-methylation during early embryogenesis of zebrafish and pig. Nat. Commun. 7, 13052 (2016).
13. Yao, B. et al. DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. Nat. Commun. 8, 1122 (2017).
14. Schiffer, S. et al. Quantitative LC–MS provides no evidence for m(6)dA or m(4)dC in the genome of mouse embryonic stem cells and tissues. Angew. Chem. Int. Ed. Engl. 56, 11268–11271 (2017).
15. Wheldon, L. M. et al. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. Cell Rep. 7, 1353–1361 (2014).
16. Santos, F. & Dean, W. Using immunofluorescence to observe methylation changes in mammalian preimplantation embryos. Methods Mol. Biol. 325, 129–137 (2006).
17. Rieder, C. L. & Palazzo, R. E. Colcemid and the mitotic cycle. J. Cell Sci. 102, 387–392 (1992).
18. Pfaffeneder, T. et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl. 50, 7008–7012 (2011).
19. Chan, Y. A. et al. Genome-wide profiling of yeast DNA:RNA hybrid probe sites with DRIP-chip. PLoS Genet. 10, e1004288 (2014).
20. Boguslavski, S. J. et al. Characterization of monoclonal antibody to DNA. RNA and its application to immunodetection of hybrids. J. Immunol. Methods 51, 123–130 (1986).
21. Mita, P. et al. LINE-1 protein localization and functional dynamics during the cell cycle. elife 7, e30058 (2018).
22. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA Hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. Cell Rep. 23, 1891–1905 (2018).
23. Wang, X. et al. N(6)-methyladenosine modulates messenger RNA translation efficiency. Cell 161, 1388–1399 (2015).
24. Alarcón, C. R. et al. HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. Cell 162, 1299–1308 (2015).
25. Carette, J. E. et al. Haplodiploid genetic screens in human cells identify host factors used by pathogens. Science 3595, 1231–1235 (2019).
26. Li, M. et al. Yhdf2-mediated m(6)A mRNA clearance modulates neural development in mice. Genome Biol. 19, 69 (2018).
27. Bonner, W. M. et al. GammaH2AX and cancer. Nat. Rev. Cancer. 8, 957–967 (2008).
28. Paulsen, R. D. et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol. Cell 35, 228–239 (2009).
29. Salas-Armenteros, I. et al. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J. 36, 3532–3547 (2017).
30. Bayona-Felu, A., Casas-Lamesa, A., Reina, O., Bernués, J. & Azorin, F. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nat. Commun. 8, 283 (2017).
31. Benitez-Guijarro, M. et al. RNAse H2, mutated in Aicardi–Goutières syndrome, promotes LINE-1 retrotransposition. EMBO J. 37, e98506 (2018).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
DharmaFECT (GE Lifesciences) in antibiotic-free medium. Cells were collected and nontargeting siRNA no. 2 (Thermo Fisher Scientific, no. D-001210-02) using RNase H1 (Qiagen, no. GS51441), human (Dharmacon, no. M-012595-00-0010) HNRNPA2B1 (Dharmacon, no. 56339), human and depletion, hiPSCs were transfected with 50 pmol of siRNA YTHDF2 on DMEM (GIBCO) supplemented with 10% bovine serum. G0/G1, S and G2/M treated with a 1:100 dilution of KaryoMAX Colcemid Solution (Thermo Fisher Scientific, no. C05228) and fixed in 70% ethanol for 2 h, washed again with PBS and stained with 10 µg/mL propidium iodide (Sigma–Aldrich, no. P3566) in PBS supplemented with 0.1% Triton X-100 and 100 µg/mL RNase A (Qiagen, no. 19101). Propidium iodide–treated cells were spread on DNA containing slides by using Beckman Coulter Astros EQ and Beckman Coulter Kaluza 2.1 software. For MTTL3 and YTHDF2 depletion, hiPSCs were transfected with 50 pmol of siRNA duplexes against human MTTL3 (Dharmacon, no. 56339), human RNase H1 (Dharmacon, no. M-012595-00-0010) and non-targeting siRNA no. 2 (Thermo Fisher Scientific, no. D-001210-02) using DharmaFECT (GE Life sciences) in antibiotic-free medium. Cells were collected for analysis after 72 h after transfection. Expression of MTTL3, RNH2PA2B1, RNase H1 and YTHDF2 was analyzed by qPCR, according to standard procedures. Gene expression was normalized by comparison to expression of GAPDH gene expression. The primers used for qPCR are listed in Supplementary Table 1. YTHDF2 KO (CRISPR/Cas9-mediated deletion of 140 bp in exon 3, leading to frameshift and generation of premature stop codon) HAP1 cells (Horizon Discovery, no. HGZHC006678c001) and their isogenic WT parental HAP1 cells (Horizon Discovery) were cultured on DMEM/F12 (Gibco Life Technologies, no. 11320213) supplemented with 20% heat-inactivated fetal bovine serum, containing 1% pen/strep at 37 °C in a humidified incubator with 5% CO2. Culture medium was changed daily and the cells were passaged using trypsin every 48 h. For determination of growth curves, cells were counted using a hemocytometer. Statistical significance was determined using two-tailed t-test following assessment of variance with F-test. The deletion in the third exon of the YTHDF2 gene was validated by both PCR (see Supplementary Table 1 for primer sequences) and sequencing.

For overexpression of RNase H1 in mammalian cells, we used C-terminally eGFP-tagged human RNASEH1 (nuclear isoform) pEGFP-RNASEH1 plasmid. This construct was a gift from A. Jackson and M. Reijns (Addgene plasmid no. 108699; http://n2t.net/addgene:108699; RRID:Addgene_108699). pmaxGFP-S9.6 was co-transfected with either 25 µg/mL RNase A (Qiagen, no. 19101) in PBS or a mixture of 25 µg/mL RNase A and 10 U/mL of RNase H in 1 x RNase H buffer (NEB, no. R0587S), or 1 x RNase H buffer (NEB, no. R0587S) supplemented with 70% ethanol and dissolved in MilliQ-grade deionized water, and then 5–10 µg of the cDNA was treated with 5 U of RNase H (NEB) overnight in RNase H Reaction Buffer (NEB) at 37 °C. After incubation, samples were ultrafiltered using an Amicon Ultra 0.5 MWCO 3-kDa centrifugal filter (Merck) at 14,000 rpm for 10 min. Subsequently, the samples were resuspended in a mixture of 70% ethanol and 30% deionized water (14,000 rpm for 15 min). To recover nucleic acids, the filter was placed upside down in a clean microcentrifuge tube and centrifuged at 1,000 rpm for 3 min. Ultrafiltrates containing released (oligo)nucleotides of molecular weight <3 kDa and the remaining nucleic acids were treated with 1 U of nuclease P1 for 1 h in a buffer containing 200 mM of ammonium acetate, 0.1 M of ZnCl2 (pH 4.6) and 10 µg per sample tetrathydrodrazine at 37 °C, followed by the addition of 10% NH4OH and 1.3 U of alkaline phosphatase and subsequent additional 1 h incubation at 37 °C. Chromatographic analysis was performed using a method previously described4 adapted for the determination of m6A, ribo-m6C and adenosine (see details in Supplementary Note).

Animals and Yhdh2 KO mouse model. Generation of Yhdh2 conditional KO mice, followed by cre-mediated deletion and derivation of mScNS from E14.5 embryonic forebrains, was described previously. All mouse experiments were approved by the Norwegian Animal Research Authority, Norwegian Food Safety Authority and performed in accordance with institutional guidelines at the Center for Comparative Medicine at Oslo University Hospital. Animal work was conducted in accordance with the rules and regulations of the Federation of European Laboratory Animal Science Association (FELASA).

Immunocytochemistry, immunohistochemistry, confocal microscopy and image quantification. Immunocytochemistry and immunohistochemistry were performed as previously described. Sections of paraffin embedded E14.5 WT and Yhdh2 KO mouse embryonic brain were used for γHAX immunohistochemistry. The sections were dewaxed according to standard procedures. Cells were fixed in 4% formaldehyde for 15 min. Cells and tissue sections were permeabilized with PBS containing 0.5% Triton X-100 for 15 min. After permeabilization, cells were treated with either 25 mM γHAX (Qiagen, no. 19101) in PBS or a mixture of 25 mg/mL RNase A and 10 U/mL of RNase H in 1 x RNase H buffer (NEB, no. R0587S), or 1 x RNase H buffer (NEB, no. R0587S) supplemented with 70% ethanol and dissolved in MilliQ-grade deionized water, and then 5–10 µg of the cDNA was treated with 5 U of RNase H (NEB) overnight in RNase H Reaction Buffer (NEB) at 37 °C. After incubation, samples were ultrafiltered using an Amicon Ultra 0.5 MWCO 3-kDa centrifugal filter (Merck) at 14,000 rpm for 10 min. Subsequently, the samples were resuspended in a mixture of 70% ethanol and 30% deionized water (14,000 rpm for 15 min). To recover nucleic acids, the filter was placed upside down in a clean microcentrifuge tube and centrifuged at 1,000 rpm for 3 min. Ultrafiltrates containing released (oligo)nucleotides of molecular weight <3 kDa and the remaining nucleic acids were treated with 1 U of nuclease P1 for 1 h in a buffer containing 200 mM of ammonium acetate, 0.1 M of ZnCl2 (pH 4.6) and 10 µg per sample tetrathydrodrazine at 37 °C, followed by the addition of 10% NH4OH and 1.3 U of alkaline phosphatase and subsequent additional 1 h incubation at 37 °C. Chromatographic analysis was performed using a method previously described4 adapted for the determination of m6A, ribo-m6C and adenosine (see details in Supplementary Note).

RNA/DNA hybrids and protein co-immunoprecipitation with S9.6 antibody. RNA/DNA hybrids and protein co-immunoprecipitation with S9.6 antibody was performed from noncrosslinked HeLa cells as described previously. gDNA was isolated from HeLa cells and sonicated as previously described. Immunoblots of RNA/DNA hybrids immunoprecipitation samples were probed with the following antibodies: Top1 (Abcam, no. ab109374, dilution 1:2,000), YTHDF1 (ProteinTech, no. 715179-1-A, dilution 1:1,000), YTHDF2 (ProteinTech, no. 24744-1-A, dilution 1:500), MTTL3 (Bethyl Laboratories, no. A301-567A, dilution 1:2,000), RNHP2A2B1 (Novus, no. NB12-6102SS, dilution 1:500) and Lamin B1 (Abcam, no. ab16048, dilution 1:2,000). Images were acquired by chemiluminescence using autoradiography.

m’A DIP and S9.6 DRIP. Genomic DNA was isolated from REBL-PAT hiPSCs by SDS/Proteinase K treatment at 37 °C followed by incubation with 100 µg/mL RNase A (Qiagen, no. 19101) for 30 min in lysis buffer, phenol–chloroform extraction and ethanol precipitation. The DNA was fragmented to 300–600 bp using Covaris S2 ultrasound processor (Covaris). gDNA of 10 ng/µl was treated with 10 U of RNase H (NEB, no. M0297S) in 1 x RNase H buffer overnight at 37 °C before immunoprecipitation. gDNA (10 µg) was used for immunoprecipitation. S9.6 DRIP was carried out essentially as described in a previously published protocol9 using S9.6 antibody (Merck Millipore, no. MABE1095) and anti-mouse magnetic Dynabeads (Invitrogen). m’A DIP was performed using anti-m’A rabbit polyclonal antibody (Synaptic, no. 202003) and magnetic anti-rabbit Dynabeads (Invitrogen, M-280; polyclonal sheep anti-rabbit IgG, no. 10716653) with a denaturation step before immunoprecipitation (10 min at 95 °C) analogously to the meDIP technique1 (Extended Data Fig. 2). The corresponding primary IgG-only and secondary IgG-only (Dynabeads-only) DIP reactions were conducted in control immune complexes. For m’A DIP (S9.6 DRIP followed by m’A DIP) DIP/DRIP, approximately 500 ng of the nucleic acids recovered from multiple DRIP reactions, performed in parallel, was used for m’A DIP followed by qPCR analysis. For S9.6 DRIP followed by the m’A RIP experiment, nucleic acids recovered from DRIP were denatured for 30 min at
with sorbitol at 37 °C until OD600 in BL21(DE3) cells and incubated with LB-medium (Puls Medical, 244610) pET-28b with N-terminal His-tag (Genescript). The plasmid was expressed

Purification of recombinant YTHDF2. Full-length YTHDF2 was cloned into

Chromatin immunoprecipitation (ChIP). Chromatin immunoprecipitation (ChIP) was performed using the EZ-Magna ChIP A/G Chromatin Immunoprecipitation Kit (Merck, no. 17-10086) according to the manufacturer’s instructions, using anti-YTHDF2 rabbit polyclonal (ProteinTech, no. 24744-1-AP), anti-HNRNPA2B1 mouse monoclonal (Novus, no. NB120-6102SS) and anti-H2AX mouse monoclonal (Merck, no. 05-636, clone JBW301) primary antibodies. ChIP was analyzed by quantitative PCR carried out with the SYBR Green PCR Master Mix (Sigma) according to standard procedures. Fold enrichment was calculated as above. The primers used for DRIP–qPCR were also

Microscale thermophoresis. Microscale thermophoresis was employed to study the interaction of YTHDF2 with modified or unmodified RNA and the RNA:DNA hybrid synthetic substrates used in EMSA experiments. Purified full-length YTHDF2 with N-terminal His-tag (Genescript), the plasmid was expressed in BL21(DE3) cells and incubated with LB-medium (Puls Medical, 244610) with sorbitol at 37 °C until OD600 = 0.7. Expression was induced with 300 µM of isopropl-b-d-thiogalactoside overnight at 18 °C. Cells were pelleted by centrifugation at 3,000 g for 10 min at 4 °C, resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM b-mercaptoethanol (ME), 10 mM imidazole) and sonicated. After sonication, the extract was centrifuged at 19,000 g for 20 min at 4 °C. The supernatant was loaded to Proteo Ni-NTA agarose prepared as described by the producer (Macherey-Nagel, 745400.100) in a 50-ml tube and incubated at 4 °C with rotation for 30 min. After centrifugation at 3,000 g for 2 min, the Ni-agarose-bound YTHDF2 was washed with buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM ME and 50 mM imidazole. Recombinant YTHDF2 was eluted with five washes (1.5 ml each) of elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM ME and 300 mM imidazole).

Library preparation and high-throughput sequencing. Sequencing libraries were prepared according to the NEB Next DNA Ultra Library Preparation Kit for Illumina (NEB, E73570). DNA was sonicated to 400–600 bp (Covaris S2) and adapters were ligated (NEB, no. E73555S) according to the protocol. Adapter-ligated DNA was digested with USER enzyme as stated in the protocol. Following immunoprecipitation, the enriched adapter-ligated DNA was amplified for 15 cycles and libraries were quantified using the Kapa Library Quantification Kit (Kapa Biosystems, no. KK4823). Sequencing was performed using the Illumina NextSeq500 platform to generate 2 × 150-bp reads. Primary IgG-only DRIP reactions resulted in low levels of DNA insufficient for successful library production, even with the maximum number (15) of amplification cycles recommended by NEB.

Whole-transcriptome sequencing. Total RNA was isolated from REBL-PAT hiPSCs according to standard procedures. RNA-sequence libraries were constructed using the Illumina TruSeq Stranded Total RNA sample preparation kit according to the manufacturer’s guidelines, and then sequenced on an Illumina HiSeq 4000 generating 20–50 million 75-bp paired-end reads per sample.

Bioinformatics analysis. The 150-bp Illumina paired-end reads were trimmed using Skewer to remove low-quality sequences[44]. Reads that passed filtering were aligned to the human Ensembl genome (build hg38.89) using the Burrows–Wheeler aligner with default parameters[47]. As each biological sample was split across multiple lanes of sequencing, the corresponding alignments were merged with Samtools[55] and de-duplicated to remove PCR artifacts with picard-tools MarkDuplicates[61]. The impact of each pulldown was assessed using Phantomeakqualtools[63], and highly modified regions (peaks) were identified using MACS2.1.1 (refs. [64, 65]). Because the exact mode of genomic distribution of m6A-containing RNA:DNA hybrids was initially unknown, we performed downstream peaks calling with CLASH+ software using q0.01 settings for narrow peaks and -broad-cutoff 0.1 (q 0.01) for broad peaks. High-confidence peaks and consensus peaks were identified using the bioconductor package DiffBind[64]. We performed peak calling against input DNA and against secondary IgG-only control samples. More than 96% of the m6A DRIP peaks called against input were also identified using IgG-only controls. Peaks called against input were used for further analysis. Conserved peaks across the broad and narrow peaks were selected with the following criteria. The broad and narrow peaks were merged with Samtools[55] and de-duplicated to remove PCR artifacts with picard-tools MarkDuplicates[61]. The impact of each pulldown was assessed using Phantomeakqualtools[63], and highly modified regions (peaks) were identified using MACS2.1.1 (refs. [64, 65]). Because the exact mode of genomic distribution of m6A-containing RNA:DNA hybrids was initially unknown, we performed downstream peaks calling with CLASH+ software using q0.01 settings for narrow peaks and -broad-cutoff 0.1 (q 0.01) for broad peaks. High-confidence peaks and consensus peaks were identified using the bioconductor package DiffBind[64]. The in-house scripts used for the analysis can be found in the following online repository: https://bitbucket.org/ADAC_UoN/adac0175-code/src. Code availability

Data availability

The confocal raw data that support the findings of this study are available from the corresponding author on request, due to size considerations. The deep-sequencing data have been deposited in the NCBI Sequence Read Archive with the Bioproject ID: PRJNA474076. The annotated bed files have been deposited in the following online repository: https://bitbucket.org/ADAC_U0a/adac1075-bed-files/src. Source data for Fig. 4, Extended Data Fig. 9 and Supplementary Fig. 9 are provided with the paper.

References

32. Pozarowski, P. & Darzykiewicz, Z. Analysis of cell cycle by flow cytometry. Methods Mol. Biol. 281, 301–311 (2004).
33. Rubeck, D. et al. PCNA directs type 2 Rnasin H activity on DNA replication and repair substrates. Nucleic Acids Res. 39, 3652–3666 (2011).
34. Abakir, A., Wheldon, L., Johnson, A. D., Laurent, P. & Ruzov, A. Detection of modified forms of cytosine using sensitive immunohistochemistry. J. Vis. Exp. https://doi.org/10.3791/54416 (2016).
35. Hamperl, S., Bocek, M. J., Saldívar, J. C., Swigut, T. & Cimprich, K. A. Transcription–replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786 (2017).

36. Ramsawhok, A. H. et al. Immunostaining for DNA modifications: computational analysis of confocal images. *J. Vis. Exp.* https://doi.org/10.3791/5631 (2017).

37. Gackowski, D. et al. Accurate, direct, and high-throughput analyses of a broad spectrum of endogenously generated DNA base modifications with isotope-dilution two-dimensional ultraperformance liquid chromatography with tandem mass spectrometry: possible clinical implication. *Anal. Chem.* **88**, 12128–12136 (2016).

38. Groh, M., Lufrino, M. M., Wade-Martins, R. & Gronak, N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10**, e1004318 (2014).

39. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).

40. Weber, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).

41. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).

42. Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).

43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

44. Li, H. et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

45. Chang, V. Y., Federman, N., Martinez-Agosto, J., Tatishchev, S. F. & Nelson, S. F. Whole exome sequencing of pediatric gastric adenocarcinoma reveals an atypical presentation of Li–Fraumeni syndrome. *Pediatr. Blood Cancer* **60**, 570–574 (2013).

46. Marinov, G. K., Kundaje, A., Park, P. J. & Wold, B. J. Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)* **4**, 209–223 (2014).

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

48. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **7381**, 389–393 (2012).

Acknowledgements

We thank Y. Wright, S. Malla, M. Loose, S. Rajani, D. Mosqueira, L. Lewis, R. Marcus, D. Onison, D. Bates and B. Coyle for technical help; D. Huangfu (Memorial Sloan Kettering Cancer Center) for the hPSC promoters dataset; and S. Peña Perez for animal care. This work was supported by a Medical Research Council IMPACT DTP PhD Studentship (grant no. MR/N013913/1) to A.A and by the Biotechnology and Biological Sciences Research Council (grant no. BB/N005759/1) to A.R. N.G.'s laboratory is supported by a Royal Society University Research Fellowship and a John Fell award (grant no. BVD07340) to N.G. I.L.G.-P.'s laboratory is supported by MINECO-FEDER (grant no. SAF2017–89745–R) to J.L.G.-P., the European Research Council (grant no. ERC-Consolidator ERC-STG-2012-233764) to J.L.G.-P. and a private donation from F. Serrano (Trading y Bolsa para Torpes, Granada, Spain). C.D. is supported by the British Heart Foundation (nos. SP/15/9/31605 and PG/14/59/31000), the Medical Research Council (no. MR/M017354/1), NC3Rs (no. CRACK-IT:35911–259146) and Heart Research UK (no. TRP0112). L.Y. was supported by the Medical Research Council (grant no. MR/M017354/1). A.K. is supported by the Norwegian Research Council (grant no. 275286) and Health Authority South-East (grant no. 20180806). B.D. is supported by Norwegian Health Authority South-East (Regional Core Facility for Structural Biology) (grant no. 2015095). D.I.G. is funded by the Wellcome Trust (grant no. WT206194). This work was partially supported by Health Data Research UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities.

Author contributions

A.A. performed immunostaining, microscopy, DRIP, DIP, RNA-seq, ChIP, qPCR, FACS sorting, cell culture experiments and contributed to bioinformatics analysis and data interpretation. T.G.C., A.R.R., J.L.G.-P. and R.D.E. performed bioinformatics analysis. A.C. and N.G. performed S9.6 immunoprecipitation and immunoblots. J.M.F., N.D. and I.R.C. performed LC–MS/MS. M.L. and A.K. contributed to EMSA and mouse KO experiments. M.L., B.D. and A.K. generated His-fused YTHDF2 and performed MST. M.E. provided cell line samples. M.S. and D.G. performed SID–UPLC–MS/MS. A.R., A.A., A.C. and N.G. performed S9.6 immunoprecipitation and immunoblots. J.M.F., N.D. and I.R.C. performed LC–MS/MS. M.E. contributed to EMSA and mouse KO experiments. M.L., B.D. and A.K. generated His-fused YTHDF2 and performed MST. M.E. provided cell line samples. J.C., L.F. Y., C.D. and D.I.G. provided the WT REBL-PAT transcriptome dataset. A.R. conceived, designed and coordinated the study and drafted the manuscript together with A.A., N.G., J.L.G.-P. and I.R.C. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-019-0549-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0549-x.

Correspondence and requests for materials should be addressed to A.K., N.G. or A.R.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Validation of m^6A DIP using synthetic RNA and DNA substrates. Validation of m^6A DIP using synthetic RNA and DNA substrates. See also Supplementary Note. (a) Schematic demonstrating the sequences of synthetic oligonucleotides and RNA:DNA hybrids used for spike in validation experiments. (b) The relative enrichment of DIP/DRIP performed on 0.1 pmol of spike in synthetic m^6A-containing RNA:DNA hybrid (81-mer with 46 % GC content) shown in (a) using anti-m^6A and S9.6 antibodies along with IGG for IP after the indicated time intervals of heat denaturation. (c) S9.6 DRIP exhibits comparable efficiencies in precipitating non-modified and m^6A-containing RNA:DNA hybrids. Relative enrichment of DRIP performed on the m^6A-containing RNA:DNA synthetic substrate normalized against that of DRIP done on equivalent amount (0.1 pmol) of non-modified spike in synthetic RNA:DNA hybrid. (d) The results of m^6A DIP on 0.1 pmol of the indicated spike in synthetic oligonucleotides and RNA:DNA hybrids. Unlike non-modified RNA:DNA hybrid substrate or single stranded m^6A-containing RNA oligonucleotide, m^6A-containing RNA:DNA hybrid is efficiently detected by m^6A DIP technique. Data are means ± SD, n = 3 independent experiments.
Extended Data Fig. 2 | Detailed schematic illustrating S9.6 DRIP and m^6A DIP techniques. Detailed schematic illustrating S9.6 DRIP and m^6A DIP techniques.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | m^6A is present on the majority of RNA-DNA hybrids in hiPSCs. m^6A is present on the majority of RNA-DNA hybrids in hiPSCs. (a) Heatmaps showing the distribution of density of indicated reads across genomic regions containing peaks (3 kb around peak center) of the three categories: m^6A peaks overlapping with S9.6 peaks (m^6A/S9.6), m^6A peaks that do not overlap with S9.6 DRIP peaks (m^6A only) and S9.6 peaks that do not correspond to m^6A DIP peaks (S9.6 only). The colour of each line represents the density of reads for a given peak. The width of the heatmaps is normalized by peak length. (b, c) Distribution of the m^6A/S9.6-, m^6A only- and S9.6 only peaks at the indicated genomic features (b) and relative to transcription start site (TSS) (c) in hiPSCs.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | RNA:DNA hybrids exhibit cell cycle-specific dynamics in hPSCs. RNA:DNA hybrids exhibit cell cycle-specific dynamics in hPSCs. (a) The diagram illustrating gating of single hiPSCs using PI-Area and PI-Width signals (left panel) and DNA content frequency histogram (right panel) of a representative hiPSCs population used for cell cycle analysis. hPSCs at G0/G1, S and G2/M phases are marked. (b) The coverage plots of m6A DIP and S9.6 DRIP densities (CPK) in the intronic regions of the indicated genes. m6A and S9.6 peaks are marked with red and blue rectangles.
Extended Data Fig. 5 | m^6A DIP signal is increased upon RNase H1 knockdown in hPSCs. m^6A DIP signal is increased upon RNase H1 knockdown in hPSCs. (a, b) The coverage plots of m^6A and S9.6 DRIP/DIP densities (CPK) in the regions located downstream of the m^6A and S9.6 peaks (marked with red rectangles). The location of regions that were used as controls in (d) is designated by blue rectangles. (c) Relative expression of RNase H1 and LINE1 transcripts in hPSCs transfected with control non-targeting (siCTL) and RNase H1 (siRNaseH1) siRNAs. (d) The results of m^6A DIP qPCR of the indicated repeats and intronic sequences performed on siCTL and siRNaseH1 hPSCs. The regions without peaks (RANBP17 Downstream and HECW1 Downstream) were used as controls. Generic primers amplifying Alu elements from the indicated families and evolutionarily young L1Hs were used for DRIP qPCRs and qPCR. Data are means ± SD, n = 3 independent experiments.
Extended Data Fig. 6 | m^6A is detectable on the RNA components of R-loops. m^6A is detectable on the RNA components of R-loops. (a) Schematic representation of the two round (S9.6 followed by m^6A) DRIP/DIP procedure. (b, c) The results of the two round (S9.6 DRIP followed by m^6A DIP) DRIP/DIP performed on individual intronic (b) and repetitive (c) m^6A/S9.6 peak containing sequences. (d) Schematic representation of the m^6A RIP performed on the RNA isolated from S9.6 IP-ed nucleic acids, followed by RT-qPCR of the candidate sequences. (e) The results of the m^6A RIP performed on the S9.6 DRIP of individual and repetitive DRIP/m^6A DIP-peak containing sequences. No RT represents control samples processed without reverse transcription. The regions without peaks (RANBP17 Downstream and HECW1 Downstream) were used as controls in (b and e). Representative results of analysis of one of 3 independent biological samples are presented. Data are means ± SD, n = 3 technical repeats.
Extended Data Fig. 7 | METTL3 depletion leads to accumulation of RNA:DNA hybrids in hPSCs. METTL3 depletion leads to accumulation of RNA:DNA hybrids in hPSCs. (a) The results of S9.6 DRIP qPCR of the indicated repeats and intronic sequences performed on siCTL and siMETTL3 hiPSCs sorted at different cell cycle phases. (b) Relative expression of the indicated transcripts in siCTL and siMETTL3 hiPSCs. The difference between the values of Y-axes in (a) and in Fig. 3b is due to incorporation of RNase H control samples in the analysis shown in Fig. 3b. Data are means ± SD, n = 3 independent experiments.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | METTL3 depletion leads to an increase in the m^6A DIP peaks in hPSCs. METTL3 depletion leads to an increase in the m^6A DIP peaks in hPSCs. (a) SID-UPLC-MS/MS quantification of m^6A in the total nucleic acids (left panel), and in ultrafiltrate fractions released upon RNase H treatment of siCTL and siMETTL3 hPSCs (right panel). The data are means of 2 independent experiments. (b) The total numbers of consensus m^6A peaks in WT and siMETTL3 hPSCs. (c) Heatmaps showing the distribution of density of indicated reads across peak-containing genomic regions (3 kb around peak center) for S9.6 DRIP in WT and m^6A DIP in WT and siMETTL3 hPSCs. The colour of each line represents the density of reads for a given peak. The width of the heatmaps is normalized by peak length. (d) Pie chart showing the percentages of transcripts differentially expressed between siCTL and siMETTL3 hPSCs (p < 0.01) and without significant changes in expression amongst genes containing siMETTL3-specific m^6A peaks. P values were calculated using the Standard ballgown parametric F-test. (e) Pie chart demonstrating the percentages of genes differentially expressed between siCTL and siMETTL3 hPSCs containing siMETTL3-specific m^6A peaks (diff peaks) and without such peaks (No diff peaks). (f) Average profile of m^6A peak densities for all genes sorted based on levels of their expression in siMETTL3 hPSCs. The colour gradient represents log_{10} of mean RPKM per bin. (g) Average profile of m^6A peak densities for all genes sorted based on the fold change of their expression between siCTL and siMETTL3 hPSCs. The colour gradient represents log_{10} of mean fold change (FC) per bin.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | YTHDF2 co-localizes with R-loops in vivo and interacts with synthetic m\(^\text{tA}\)-marked RNA:DNA hybrids in vitro. YTHDF2 co-localizes with R-loops in vivo and interacts with synthetic m\(^\text{tA}\)-marked RNA:DNA hybrids in vitro. (a) Co-immunostaining of YTHDF2 with R-loops (S9.6) in a representative hiPSCs interphase nucleus. Merged view and individual channels are shown. The area used for S9.6/YTHDF2 signals quantification is indicated. The arrow designates the region used for generation of signal intensity profile shown in (b). (b) The profile showing intensities of YTHDF2, S9.6 and 4,6-diamidino-2-phenylindole (DAPI) signals across the nuclear region marked with an arrow in (a). (c) Representative image of hiPSCs immunostained for YTHDF2 and R-loops. Merged view and YTHDF2/S9.6 channels are shown. Scale bars are 20 \(\mu\)m. The experiments shown in (a, c) were repeated independently 3 times with similar results. (d) Scatter diagram for YTHDF2 and S9.6 fluorescence intensities in an individual hiPSCs nucleus. (e) The value of overlapping coefficient of YTHDF2 vs S9.6 intensities quantified for 20 cells immunostained for YTHDF2 and S9.6. The centre value is median, error bars show minimal and maximal values. (f, g) The results of EMSA using recombinant YTHDF2-His fusion (f) or YTHDF2-FLAG (g) and 0.15 pmol of unmodified (non) or m\(^\text{tA}\)-containing (meth) RNA oligonucleotides (ssRNA) or corresponding synthetic RNA:DNA hybrids (RNA:DNA). Triangles indicate increasing concentrations of the protein (10, 100, 300 ng). Concentrations of the recombinant protein (g) and NaCl/KCl in the binding buffer are indicated. The RNA and RNA:DNA-protein complexes are arrowed. See also Supplementary Note. The gel images were cropped. The full scans of the gels are shown in Source Data 2.
Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | YTHDF2 interacts with R-loop-containing loci in vivo and its depletion leads to accumulation of RNA:DNA hybrids in hPSCs.

YTHDF2 interacts with R-loop-containing loci in vivo and its depletion leads to accumulation of RNA:DNA hybrids in hPSCs. (a-f) The results of YTHDF2 (a, c, e) and HNRNPA2B1 (b, d, f) ChIP qPCR of the indicated repeats and intronic sequences performed on REBL-PAT hiPSCs (a, b), siCTL and siMETTL3 hPSCs (c, d) as well as on siCNTL and siRNase H1 hPSCs (e, f). (g) Relative expression of YTHDF2 transcript in hiPSCs transfected with siYTHDF2 siRNAs compared with that in non-targeting control siRNA (siCTL) transfected cells. (h, i) The results of S9.6 DRIP qPCR (h) and m6A DIP qPCR (i) of the indicated repeats and intronic sequences performed on siCTL and siYTHDF2 hiPSCs. Generic primers amplifying Alu elements from the indicated families and evolutionarily young L1Hs were used for DRIP qPCRs and qPCR. The regions without peaks (RANBP17 Downstream and HECW1 Downstream) were used as controls. Representative results of analysis of one of 3 independent biological samples are presented. Data are means ± SD, n = 3 technical repeats.
Corresponding author(s): Arne Klungland, Natalia Gromak, Alexey Ruzov
Last updated by author(s): Nov 6, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Deep sequencing data were collected using Illumina NextSeq Control Software 2.1.0.31 and Illumina RTA 2.4.11 & bcl2fastq 2.18 software.
Confocal images were obtained using Zeiss LSM 710 Zen 2012 Black Edition Service pack 5 imaging software.
hPSCs were sorted using Weasel version 3.0.2 (Walter and Eliza Hall Institute for Medical Research) and Beckman Coulter Kaluza Analyzis 2.1 software.

Data analysis

Deep sequencing data were processed using open source software tools. The following packages were used: Samtools version 1.3.1, Picard Tools Dedup version 2.5.0, MACS2 1.1, PhantomPeakQualTools version - Feb 13, 2012, Bedtools version 2.27.1, tximport, R 3.6, Bowtie2, IGV 2.4.0, Skewer version 0.2.2, BWA version 0.7.15, Deeptools version 2.5.7, awk version 4.0.2, Hisat2 version 2.1.0, StringTie version 1.2.3, Ballgown version 2.18.0, ChiPQC version 1.18.2, ggplot2 version 3.2.0, GenomicFeatures version 1.34.8, DESeq version 1.34.1, AnnotationDbi version 1.44.0, DiffBind version 2.10.0, Biobase version 2.42.0, GenomicRanges version 1.34.0, GenomelInfoDb version 1.18.2, BioVenn.

The full pipeline including in-house scripts used for the analysis can be found in the following online repository (https://bitbucket.org/ADAC_UoN/adac0175-code/)
Confocal Images were processed using Fiji Image J, Adobe Photoshop CS6 Version 13.0 x32, and Zeiss LSM 710 Zen 2012 Black Edition Service pack 5 imaging software.
FACS data were analysed using Weasel V3.0.2 (Walter and Eliza Hall Institute for Medical Research) and Beckman Coulter Kaluza Analyzis 2.1 software. Signal intensity and qPCR data were plotted using GraphPad Prism 7.04.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The deep sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) with the Bioproject ID: PRJNA474076 (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB4074125). The annotated bed files have been deposited to the following online repository (https://bitbucket.org/ADAC_UoN/adac1075-bed-files/src). The in-house scripts used for the analysis can be found in the following online repository (https://bitbucket.org/ADAC_UoN/adac0175-code/src).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

- Samples size was chosen to ensure reproducibility of the results at affordable costs. Generally, sample size was determined based on previously published studies and on experience and was equivalent to that is routinely used for any particular assay. Sample sizes are indicated separately for different experiments. At least 2 and typically 3 independent experiments were carried out for most of the assays. DRIP and DIP were performed in two and RNaseq in three biologically independent experiments. We observed generally good correlation between the replicates for different experiments suggesting that chosen samples size was sufficient. No statistical methods were used to predetermine sample sizes.

Data exclusions

- In case of experimental mistakes or occasional loss of samples/reagents, the experiment was discarded and repeated. Otherwise, no data were excluded from the analysis.

Replication

- All experiments were replicated independently. All attempts at replication were successful.

Randomization

- No randomization was required as the study was based on molecular and cellular biology techniques and did not involve allocation of experimental units across different treatment groups, no human subjects were involved.

Blinding

- Blinding was not relevant to the study as the study was based on molecular and cellular biology techniques, the results were derived from objective quantitative methods. No subjective measurements were taken.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a
- Involved in the study
  - Antibodies
  - Eukaryotic cell lines
  - Palaeontology
  - Animals and other organisms
  - Human research participants
  - Clinical data

Methods

- n/a
- Involved in the study
  - ChIP-seq
  - Flow cytometry
  - MRI-based neuroimaging

Antibodies

- Antibodies used
  - Anti-m6A rabbit polyclonal (Synaptic systems, catalogue number 202003, Lot number 2-92, 1:200 dilution), anti-m1A mouse monoclonal (Diagenode catalogue number C15200235, 1:500 dilution), anti-5-methylcytosine (5mC) mouse monoclonal (clone 33D3, Diagenode, catalogue number C15200081-100, Lot number RD-002, 1:100 dilution), anti-phospho-Histone H3 (Ser10) mouse monoclonal (6G3, Cell Signalling, catalogue number 9706, Lot number 05636 iBW301, 1:200 dilution), anti-YTHDF1 rabbit polyclonal (ProteinTech, catalogue number 17479-1-AP, Lot number 00040713, 1:500 dilution), anti-YTHDF2 rabbit monoclonal (6G3, Cell Signalling, catalogue number 9706, Lot number 05636 iBW301, 1:200 dilution)
Authentication REBL-PAT hiPSCs and HUES7 hESC were authenticated by karyotyping, gene expression profiling (using qPCR and immunocytochemistry), phenotypic assays and differentiation into 3 germ layers. Wild type parental HAP1 cells, source: Horizon Discovery.

Validation Anti-γH2AX antibody (Merck, catalogue number 05-636, clone JBW301) has been validated for immunostaining and ChIP on human cells and mouse tissues by the vendor and in a number of studies (http://www.merckmillipore.com/GB/en/product/Anti-phospho-Histone-H2AX-Ser139-Antibody-clone-JBW301, MF_NB-05-636#documentation; Sato S, et al. (2016) Nat Commun. 2015 6:7035. doi: 10.1038/ncomms8035; Kanu N, et al. (2015) Oncogene. 34(46):5699-708; Abdou I, Poirier GG, Hendzel MJ, Weinfeld M. (2015) Nucleic Acids Res. 2015 43(2):875-92).

S9.6 antibody (Merck Millipore, catalogue number MABE1095) is widely used for DRIP (Boguslawski, S.J., et al. (1986). J. Immunol Methods. 89(1):123-130, El Hage, A., et al. (2014). PLoS Genet. 10(10):e1004716). It was previously validated for DRIP and DRIPseq on a number of species including human (Skourti-Stathaki, K., et al. (2011). Mol. Cell. 42(6):794-805; Ginno, P.A., et al. (2012). Mol. Cell. 45(6):814-825; Bhata, V., et al. (2014). Nature. 511(7509):362-365; Cristini A, et al. (2019) Cell Rep. 23, 1891-1905) and for immunocytochemistry on human and mouse cells (Bhata, V., et al. (2014). Nature. 511(7509):362-365; Ginno, P.A., et al. (2012). Mol. Cell. 45(6):814-825).

Top1 (Abcam, catalogue number ab109374), METTL3 (Bethyl Laboratories, catalogue number A301-567A) and Lamin B1 (Abcam, catalogue number ab16048) antibodies have been validated for Western blot on human and mouse cells previously (https://www.abcam.com/topoisomerase-i-antibody-epr5375-ab109374-references.html#top-200; Cristini A, et al. (2016) Nucleic Acids Res. 44(3):1161-78; e6; Cristini, A., et al. (2018) Cell Rep. 23, 1891-1905; https://www.bethyl.com/product/A301-567A/METTL3+MT-A70+Antibody; Takata MA, et al. (2017) Nature. 550(7674):124-127; https://www.abcam.com/lamin-b1-antibody-nuclear-envelope-marker-ab16048-references.html#top-500; Du Q, et al. (2019) Nat Commun. 10(1):416; Vadnais C, et al. (2018) Nat Commun. 9(1):1418; Dingar D, et al. (2018) Nat Commun. 9(1):3502) and used in a number of studies.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

REBL-PAT hiPSCs, source: Chris Denning’s lab, University of Nottingham, described in Mosqueira D, et al. Eur Heart J. 43, 3879-3892 (2018).

HUES7 hESC, source: Chris Denning’s lab, University of Nottingham, derived in DA Melton’s lab, described in Cowan CA, et al. N Engl J Med. 350, 1353-1356 (2004).

U87MG, LN-18 and HeLa cell lines, source: ATCC.

YTHDF2 KO HAP1 cells, source: Horizon Discovery (catalogue number H2GHCO06678c001).

Wild type parental HAP1 cells, source: Horizon Discovery.

Authentication

REBL-PAT hiPSCs and HUES7 hESC were authenticated by karyotyping, gene expression profiling (using qPCR and immunocytochemistry), phenotypic assays and differentiation into 3 germ layers.

U87MG, LN-18 and HeLa cell lines were authenticated by vendor.

HeLa cells were authenticated by short tandem repeat (STR) analysis by the vendor.
U87MG and LN-18 were authenticated by gene expression (using qPCR and immunostaining) and phenotypic profiling. YTHDF2 KO HAP1 cells were authenticated by immunostaining, PCR and sequencing. Both YTHDF2 KO and parental HAP1 cells were also authenticated by vendor.

Mycoplasma contamination
All cell lines tested negative for mycoplasma contamination

Commonly misidentified lines
(See ICLAC register)
No commonly misidentified cell lines were used.

Animals and other organisms
Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals
C57BL/6 wild-type and YTHDF2 targeted mouse embryos at embryonic day 14.5 were used. Sex of the embryos was not determined.

Wild animals
The study did not involve wild animals.

Field-collected samples
The study did not involve samples collected from the field.

Ethics oversight
All mouse experiments were approved by the Norwegian Animal Research Authority by Norwegian Food Safety Authority and done in accordance with institutional guidelines at the Centre for Comparative Medicine at Oslo University Hospital. Animal work was conducted in accordance with the rules and regulations of the Federation of European Laboratory Animal Science Association’s (FELASA).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Reads are submitted to SRA (SUB4074125): Bioproject PRJNA474076
Bed files are available at: https://bitbucket.org/ADAC_UoN/adac1075-bed-files/src
The datasets can be accessed at:
https://www.ncbi.nlm.nih.gov/account/?back_url=https%3A%2F%2Fdataview.ncbi.nlm.nih.gov
with user = thomgiles; pass= n0078022

Files in database submission

DRIP/DIP:
K1_6mA_R2.fastq.gz (fastq)
K1_6mA_R1.fastq.gz (fastq)
K2_6mA_R2.fastq.gz (fastq)
K2_6mA_R1.fastq.gz (fastq)
K3_6mA_R2.fastq.gz (fastq)
K3_6mA_R1.fastq.gz (fastq)
K4_6mA_R2.fastq.gz (fastq)
K4_6mA_R1.fastq.gz (fastq)
K3_DNA-RNA_R2.fastq.gz (fastq)
K3_DNA-RNA_R1.fastq.gz (fastq)
K4_DNA-RNA_R2.fastq.gz (fastq)
K4_DNA-RNA_R1.fastq.gz (fastq)
K1_DNA-RNA_R2.fastq.gz (fastq)
K1_DNA-RNA_R1.fastq.gz (fastq)
K2_DNA-RNA_R2.fastq.gz (fastq)
K2_DNA-RNA_R1.fastq.gz (fastq)
Input_R2.fastq.gz (fastq)
Input_R1.fastq.gz (fastq)
SAMN12283116 / IGG1 IGG1.R1.fastq.gz IGG1.R2.fastq.gz
SAMN12283117 / IGG2 IGG2.R1.fastq.gz IGG2.R2.fastq.gz
SAMN12283118 / SiM3-S1 SiM3-S1.R1.fastq.gz SiM3-S1.R2.fastq.gz
SAMN12283119 / SiM3-S2 SiM3-S2.R1.fastq.gz SiM3-S2.R2.fastq.gz

RNAseq:
E8_48.R2.fastq.gz (fastq)
E8_48.R1.fastq.gz (fastq)
E8_48_2.R2.fastq.gz (fastq)
E8_48_2.R1.fastq.gz (fastq)
E8_72.R2.fastq.gz (fastq)
E8_72.R1.fastq.gz (fastq)
E8_72_2.R2.fastq.gz (fastq)
E8_72_2.R1.fastq.gz (fastq)
**Methodology**

**Replicates**

2 Biological replicates were used for each DRIP experiment. 4 or 3 Biological replicates were used for RNAseq experiments. Replicates agreement is:

**DRIPseq:**
- K2 S9.6 DRIP RNase H, Replicate 1
- K1 S9.6 DRIP RNase H, Replicate 2
- K3 S9.6 DRIP, Replicate 1
- K4 S9.6 DRIP, Replicate 2
- K3 m6A DIP RNase H, Replicate 1
- K4 m6A DIP RNase H, Replicate 2
- K1 m6A DIP, Replicate 1
- K2 m6A DIP, Replicate 2
- SiM3-S1 m6A DIP Replicate 1
- SiM3-S2 m6A DIP Replicate 2

**RNAseq:**

- Wild type hPSCs:
  - Replicate 1 E8_48.
  - Replicate 2 E8_48_2.
  - Replicate 3 E8_72.
  - Replicate 4 E8_72_2.

- siCTL hPSCs:
  - SAMN12283120 / siCTL-S1-RNA siCTL-S1-RNA.R1.fastq.gz siCTL-S1-RNA.R2.fastq.gz
  - SAMN12283121 / siCTL-S2-RNA siCTL-S2-RNA.R1.fastq.gz siCTL-S2-RNA.R2.fastq.gz
  - SAMN12283122 / siCTL-S3-RNA siCTL-S3-RNA.R1.fastq.gz siCTL-S3-RNA.R2.fastq.gz
  - SAMN12283123 / SiM3-S1-RNA SiM3-S1-RNA.R1.fastq.gz SiM3-S1-RNA.R2.fastq.gz
  - SAMN12283124 / SiM3-S2-RNA SiM3-S2-RNA.R1.fastq.gz SiM3-S2-RNA.R2.fastq.gz
  - SAMN12283125 / SiM3-S3-RNA SiM3-S3-RNA.R1.fastq.gz SiM3-S3-RNA.R2.fastq.gz

- siMETTL3 hPSCs:
  - SAMN12283123 / SiM3-S1-RNA SiM3-S1-RNA.R1.fastq.gz SiM3-S1-RNA.R2.fastq.gz
  - SAMN12283124 / SiM3-S2-RNA SiM3-S2-RNA.R1.fastq.gz SiM3-S2-RNA.R2.fastq.gz
  - SAMN12283125 / SiM3-S3-RNA SiM3-S3-RNA.R1.fastq.gz SiM3-S3-RNA.R2.fastq.gz

**Sequencing depth**

Paired end 150 bp Illumina sequencing, >30x depth in all DRIP-seq samples.

| Sample | Total reads | Uniquely mapped reads |
|--------|-------------|-----------------------|
| m6A DIP, Replicate 1 | 49842240 | 48015356 |
| m6A DIP, Replicate 2 | 56332128 | 54533737 |
| S9.6 DRIP, Replicate 1 | 72691406 | 41780898 |
| S9.6 DRIP, Replicate 2 | 77961427 | 4933291 |
| m6A DIP RNase H, Replicate 1 | 79112558 | 76866727 |
| m6A DIP RNase H, Replicate 2 | 87247890 | 82861485 |
| S9.6 DRIP RNase H, Replicate 1 | 89788100 | 82922838 |

Genome browser session (e.g. UCSC)

As Ensembl (GRCh38.90) genome version was used for mapping and the downstream visualization analysis was performed using IGV, no web genome browser is available but an IGV session can be made available upon request. The bed files have been deposited to the following online repository (https://bitbucket.org/ADAC_UoN/adac1075-bed-files/src).
No BAM alignments are present for the RNAseq because a pseudo-alignment method (Salmon) was used to quantify transcripts.

### Antibodies

S9.6 antibody (Merck Millipore, catalogue number MABE1095) and anti-m6A rabbit polyclonal antibody (Synaptic systems, catalogue number 202003) were used for DRIP/DIP-seq and DRIP/DIP-qPCR.

### Peak calling parameters

Peaks were called according to the recommended default MACS2 parameters described here: https://github.com/taoliu/MACS/blob/master/README.md

The code is available online at https://bitbucket.org/ADAC_UoN/adac0175-code/src

### Data quality

Reads quality was assessed directly by the sequencing center using fastQC, the impact of each of the pull-downs on read distribution was assessed using Phantompeakqualtools and ChipQC was employed for assessing the peaks’ genomic distribution and the quality of the data. Consensus peaks were defined using the dba.peakset() function to select for peaks overlapping in both replicates. Peaks were called with a q value of 0.01. The numbers of narrow peaks were as follows:

- m6A DIP Replicate 1, 58120;
- m6A DIP Replicate 2, 26699;
- m6A DIP RNase H Replicate 1, 1742;
- m6A DIP RNase H Replicate 2, 1103;
- S9.6 DRIP Replicate 1, 11236;
- S9.6 DRIP Replicate 2, 8691;
- S9.6 DRIP RNase H Replicate 1, 506;
- S9.6 DRIP RNase H Replicate 2, 483;
- m6A DIP siMETTL3 Replicate 1, 381150
- m6A DIP siMETTL3 Replicate 2, 133698

### Software

Deep sequencing data were collected using Illumina NextSeq Control Software 2.1.0.31 and Illumina RTA 2.4.11 & bcl2fastq 2.18 software. The 150 bp Illumina paired end reads were trimmed using Skewer to remove low quality sequences. Reads that passed filtering were aligned to the human Ensembl genome (build hg38.89) using BWA with default parameters. As each biological sample was split across multiple lanes of sequencing, the corresponding alignments were merged with Samtools and de-duplicated to remove PCR artefacts with picard-tools MarkDuplicates. The highly modified regions (HMRs, peaks) were identified using MACS2.1.1. High confidence consensus peaks were identified using the bioconductor package DiffBind. In each instance the replicate sample BAM/bed files along with the corresponding input samples were used as input. The Ensembl Genome GTF file (Homo_sapiens.GRCh38.89.gtf) was downloaded and processed to extract gene biotypes, annotated gene features, introns, intergenic regions and sequences 3kb up/downstream of genes. The Repeatmasker Library (db20140131) was downloaded from UCSC, converted to the ensemble gene build and processed to identify repeat masked regions. The Ensembl regulatory build (HgGRCh38.Regulatory_Build.regulatory_features.20161111) was downloaded and processed to identify regulatory features. The consensus and DMR bedfiles were annotated with this list of gene features, gene biotypes, repeat-masked sequences and regulatory regions using BedTools intersect.

The code is available Online at https://bitbucket.org/ADAC_UoN/adac0175-code/src
Flow Cytometry

Plots
Confirm that:
☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
☒ All plots are contour plots with outliers or pseudocolor plots.
☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
G0/G1, S and G2/M phases flow cytometry sorting was performed according to the previously described method (Pozarowski, P., Darzynkiewicz, Z. Analysis of cell cycle by flow cytometry. Methods Mol Biol. 281, 301-311 (2004)). Briefly, enzymatically dissociated hPSCs were washed in PBS and fixed in 70 % ethanol for 2 h, washed with PBS again and stained with 10 μg/ml propidium iodide (PI) (Sigma-Aldrich, catalogue number P3566) in PBS supplemented with 0.1 % Triton X-100 and 100 μg/ml RNase A (Qiagen, catalogue number 19101). PI treated hPSCs were sorted based on the DNA content into G0/G1, S and G2/M cells using Beckman Coulter Astrios EQ and Beckman Coulter Kaluza Analysis 2.1 software.

Instrument
Beckman Coulter Astrios EQ, serial number AW24025

Software
Weasel V3.0.2 (Walter and Eliza Hall Institute for Medical Research), Beckman Coulter Kaluza Analysis 2.1

Cell population abundance
2,000,000 sorted cells were used for each experimental point. The FACS-sorted cell populations were 98 % pure. The purity was confirmed by re-sorting of the obtained cell populations following PI treatment.

Gating strategy
Cells were gated using a forward scatter vs side scatter plot to exclude any debris and apoptotic cells. Single cells were then gated for plotting the side scatter against area. From there G0, S phase and G2/M cell populations were gated on the PI (488-620/29) plot, with G2 showing a peak at double the fluorescence intensity of G0. Please see Methods and Supplementary Figures.

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