Published in final edited form as:
*Nat Genet*. 2020 January ; 52(1): 48–55. doi:10.1038/s41588-019-0549-x.

# N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells

**Abdulkadir Abakir**[1], **Tom C. Giles**[2,3], **Agnese Cristini**[4], **Jeremy M. Foster**[5], **Nan Dai**[5], **Marta Starczak**[6], **Alejandro Rubio-Roldan**[7], **Miaomiao Li**[8,9], **Maria Eleftheriou**[1], **James Crutchley**[1], **Luke Flatt**[1], **Lorraine Young**[1], **Daniel J. Gaffney**[10], **Chris Denning**[1], **Bjørn Dalhus**[8,11], **Richard D. Emes**[2,12], **Daniel Gackowski**[6], **Ivan R. Corrêa Jr**[5], **Jose L. Garcia-Perez**[7,13], **Arne Klungland**[8,9,*], **Natalia Gromak**[4,*], **Alexey Ruzov**[1,*]

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

## Introductory



*Correspondence to: Alexey Ruzov, alexey.ruzov@nottingham.ac.uk, Natalia Gromak, natalia.gromak@path.ox.ac.uk and Arne Klungland, arne.klungland@medisin.uio.no.

**Ethics statement:** Use of the HUES7 hESC line was approved by the UK Medical Research Council Steering Committee, in association with the UK Stem Cell Bank. 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).

**Data and materials availability:** The confocal raw data that support the findings of this study are available from the corresponding author upon request due to size considerations. 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).

**Code availability:** The in-house scripts used for the analysis can be found in the following online repository (https://bitbucket.org/ADAC_UoN/adac0175-code/src).

**Author contributions:** A.A. performed immunostaining, microscopy, DRIP, DIP, RNAseq, ChIP, qPCR, FACS sorting, cell culture experiments and contributed to bioinformatics analysis and data interpretation. T.C.G., A.R.R., J.L.G.P. and R.D.E. performed bioinformatics analysis. A.C. and N.G. performed S9.6 IP and western blots. J.M.F., N.D. and I.R.C. performed LC-MS/MS. M.S. and D.G. performed SID-UPLC-MS/MS. A.R., A.A., 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 lines samples. J.C., L.F, L.Y., C.D. and D.J.G. provided wild type 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:** Authors declare no competing interests.


Europe PMC Funders Author Manuscripts

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 cells[1–4]. Here we show that N6-methyladenosine ($m^6A$) modification, contributing to different aspects of mRNA metabolism[5, 6], is detectable on the majority of RNA:DNA hybrids in human pluripotent stem cells (hPSCs). We demonstrate that $m^6A$-containing R-loops accumulate during $G_2$/M and are depleted at $G_0/G_1$ phases of the cell cycle and that the $m^6A$ reader promoting mRNA degradation, YTHDF2[7], interacts with R-loops-enriched loci in dividing cells. Consequently, *YTHDF2* knockout leads to increased R-loop levels, cell growth retardation and accumulation of γH2AX, a marker for DNA double-strand breaks, in mammalian cells. Our results suggest that $m^6A$ regulates accumulation of R-loops, implying a role for this modification in safeguarding genomic stability.

Dynamic methylation of adenosine in RNA (N6-methyladenosine, $m^6A$) has been implicated in regulation of different aspects of mRNA metabolism in mammals by numerous studies[5, 6]. Although $m^6A$ is abundant in eukaryotic transcriptomes, its DNA counterpart, N6-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 DNA[8–10]. Despite the fact 6mA is reportedly widespread in fungal genomes[11], its prevalence in mammalian systems is currently poorly understood. This modification accumulates in preimplantation pig embryos[12]; however, evidence for its presence in mouse tissues is contradictory[13, 14]. In this study, we initially aimed to examine if this mark is detectable in human cell lines using a sensitive immunostaining method that we have previously employed to detect modified forms of cytosine in vertebrate models[15].

To confirm that we can differentiate between $m^6A$-modified mRNAs and 6mA present on genomic DNA, we performed immunostaining of hPSCs using previously validated anti-$m^6A$/6mA antibody[11] without the DNA denaturation step which is required for the immunochemical detection of modified bases in genomic DNA[11, 15, 16]. In these experiments, we observed prominent $m^6A$ staining that disappeared upon pre-treatment 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 double stranded nucleic acids and is routinely used for detection of cytosine modifications and 6mA in genomic DNA[11, 15, 16]. In these conditions, we also detected strong $m^6A$ signal in both nuclei and the cytoplasm of hPSCs and cancer cell lines. Notably, high levels of $m^6A$ staining were still evident in the mitotic chromatin in all our samples processed after RNase A treatment (Fig. 1a; Supplementary Note). To examine if the mitotic staining we observed indicates the presence of 6mA in the human genome, we performed LC-MS/MS quantification of 6mA and modified forms of cytosine in the DNA of two hPSCs lines either cultured under standard conditions or after enrichment for mitotic cells using colcemid treatment[17]. 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 (ppm) levels, suggesting that this modification, if present in the hPSCs genomes, only occurs at levels substantially lower than that of 5-formylcytosine[18] (Fig. 1b). These results confirmed previously published LC-MS data indicating the absence of 6mA in the genome of mouse embryonic stem cells and tissues[14].

Attempting to explain the discrepancy between our LC-MS/MS data and immunostaining results, we hypothesized that the mitotic anti-m$^6$A/6mA antibody-specific signal was caused by the presence of this modification on the RNA component of R-loops[1]. 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 cells[1–4]. To test this hypothesis, we immunostained hPSCs using m$^6$A antibody after treatment of the samples with *E. coli* RNase H, an enzyme that specifically degrades RNA molecules present in RNA:DNA hybrids. Notably, mitotic m$^6$A staining significantly decreased or disappeared in the hPSCs pre-treated with RNase H, corroborating the presence of this modification on the RNA strand of RNA:DNA hybrids (Fig. 1c-d; Supplementary Note). Confirming our immunostaining results, we also detected a release of m$^6$A (but not of ribo-5-methylcytidine, ribo-m$^5$C) to filtrate by stable-isotope dilution ultra-performance liquid chromatography with tandem mass spectrometry (SID-UPLC-MS/MS) upon treatment of hPSCs-derived nucleic acids with RNase H (Fig. 1e, f; Supplementary Note). Overall, these results strongly suggested that m$^6$A modification is associated with the RNA components of RNA:DNA hybrids in hPSCs.

To examine the genomic distribution of m$^6$A-marked RNA:DNA hybrids, we modified a previously published DNA:RNA immunoprecipitation technique (DRIP, referred here as S9.6 DRIP)[2, 19] by replacing anti-RNA:DNA hybrid S9.6 antibody[20] with anti-m$^6$A antibody (designated here as m$^6$A DNA immunoprecipitation, m$^6$A 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$^6$A DIP in parallel with S9.6 DRIP coupled with high-throughput sequencing on hiPSCs (Fig. 2a; Extended Data Fig. 2). Although both types of IP resulted in generation of large peak datasets, the majority of m$^6$A DIP and S9.6 DRIP peaks were not detectable in the control samples pre-treated with RNase H, confirming that the presence of methylated adenosine is correlated to the RNA component of R-loops in hPSCs (Fig. 2b, Supplementary Note). Both m$^6$A- 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 (Fig. 2c, 3a; Supplementary Note). Despite the number of m$^6$A 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$^6$A- and S9.6 peaks was RNase H-dependent, and the density of S9.6 DRIP reads was noticeably increased across the m$^6$A peaks that do not overlap with S9.6 peaks (Fig. 2e, Extended Data Fig. 3a), we concluded that difference in the peak numbers we observed was likely due to different sensitivity of the corresponding antibodies and, therefore, our results imply that m$^6$A marks most of the RNA:DNA hybrids in hPSCs. In line with this explanation, m$^6$A DIP demonstrated approximately 3.6-fold more efficient enrichment for the synthetic m$^6$A-containing RNA:DNA hybrid compared with S9.6 DRIP in our spike-in experiments (Extended Data Fig. 1b). We also observed similar distribution of common m$^6$A/S9.6- and m$^6$A-only peaks amongst different genomic features (Extended Data Fig. 3b, c).

Since the RNase H-sensitive m$^6$A 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 m$^6$A-containing R-loops during cell cycle, we performed m$^6$A DIP and S9.6 DRIP on $G_0/G_1$, S and $G_2$/M flow cytometry-sorted hPSCs populations (Extended Data Fig. 4a), followed by quantitative PCR (qPCR) of LINE-1 repeats and individual intronic sequences enriched in both m$^6$A- and S9.6 peaks (Fig. 3a; Extended Data Fig. 4b). These experiments demonstrated that RNA:DNA hybrids accumulate on LINE-1 retrotransposons during S phase, max out at $G_2$/M and drastically decrease at $G_0/G_1$ phases of the cell cycle in hPSCs (Fig. 3b). Consistently, a recent study demonstrated that retrotransposition active LINE-1-derived mRNAs are enriched in cells exiting mitosis[21]. The intronic R-loops were found in high levels at both S and $G_2$/M phases, but were also significantly depleted at $G_0/G_1$ phase (Fig. 3c, d). Importantly, these cell cycle-specific changes were essentially equivalent in both m$^6$A DIP and S9.6 DRIP, suggesting that m$^6$A is present on RNA:DNA hybrids throughout all stages of the cell cycle (Fig. 3b-d). Notably, m$^6$A DIP qPCR enrichment substantially increased on the repetitive and intronic loci upon small interfering RNA (siRNA)-mediated knock down of RNase H1 in hPSCs (Extended Data Fig. 5a-d). Moreover, the intronic and repetitive m$^6$A DIP-containing sequences were also enriched in the two round IP (S9.6 DRIP followed by m$^6$A DIP or m$^6$A RNA IP) procedures, further confirming the presence of m$^6$A on the RNA components of R-loops (Extended Data Fig. 6a-e). In sum, these results suggested that the turn-over rates of m$^6$A-marked R-loops vary for cell cycle phases.

Given that deposition of m$^6$A is known to affect stability of mRNAs[5–7], we hypothesized that this mark may also modulate the stability of R-loops. Since siRNA-mediated knockdown of m$^6$A methyltransferase METTL3 led to accumulation of RNA:DNA hybrids in hPSCs (Extended Data Fig. 7a-b, 8a-f; Supplementary Note), we next enquired if any of the previously characterized m$^6$A reader proteins may interact with mitotic chromatin enriched in m$^6$A-containing R-loops. First, we examined for the presence of the m$^6$A readers in proteins interacting with RNA:DNA hybrids immuno-precipitated from HeLa cells using S9.6 antibody[22]. The analysis showed an enrichment of YTHDF1 – a protein promoting translation of m$^6$A-containing mRNAs[23], HNRNPA2B1 – a nuclear m$^6$A reader previously implicated in mRNA processing[24], and YTHDF2 – an m$^6$A-interacting protein that regulates degradation of cytoplasmic mRNAs[7] as well as METTL3 in the R-loop IP, 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 the chromatin during mitosis (Fig. 4d, e), YTHDF2 migrated to mitotic chromatin in dividing hiPSCs (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 m$^6$A-containing synthetic RNA:DNA substrates in electrophoretic mobility-shift assays (Extended Data Fig. 9f, g; Supplementary Note) and in MicroScale Thermophoresis (MST) analysis that demonstrated that YTHDF2 shows comparable dissociation constant values for its interaction with m$^6$A-marked single-stranded RNA and m$^6$A-RNA:DNA duplexes in this assay (Fig. 4h). Furthermore, YTHDF2 ChIP showed that this m$^6$A 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 HNRNPA2B1 with LINE-1 repeats but were able to detect binding of this protein to R-loops-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 upon METTL3 knock-down, confirming their interaction with m$^6$A 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 siRNaseH1 hPSCs, strongly suggesting the association of this m$^6$A reader with R-loops *in vivo* (Extended Data Fig. 10e, f). To assess the functional significance of YTHDF2 migration to mitotic chromatin, we performed its siRNA-mediated depletion (siYTHDF2) in hPSCs. S9.6 DRIP- and m$^6$A DIP qPCR showed a significant enrichment in both repetitive and individual intronic R-loops sequences in siYTHDF2 hPSCs relative to siCTL cells (Extended Data Fig. 10g-i). To further confirm these results, we next assessed the levels of R-loops in *YTHDF2* knockout (KO) HAP1[25] cells expressing a truncated version of this protein that does not co-localize with mitotic chromatin (Fig. 5a, 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 accretion of m$^6$A on RNA:DNA hybrids (Fig. 5d) and cell growth retardation (Fig. 5e). Subsequent analysis of recently published *Ythdf2* constitutive knockout[26] mice-derived neural stem cells (mNSCs) confirmed these results demonstrating increased levels of S9.6 immunostaining and accumulation of RNA:DNA hybrids in LINE-1open reading frames upon depletion of *Ythdf2* in this system (Fig. 5f, g). In line with these results, *YTHDF2* KO HAP1 cells displayed an increased accumulation of a marker for DNA double-strand breaks, phosphorylated (ser139) histone variant H2AX (γH2AX)[27] both at the nucleus-wide level and at R-loop-enriched loci (Fig. 6a, b). Correspondingly, we also observed elevated levels of γH2AX staining in the cortex of *Ythdf2* KO embryos and *Ythdf2* KO mNSCs (Fig. 6c) as well as, to a lesser extent, in hPSCs upon siRNA-mediated depletions of METTL3 and HNRNPA2B1 (Fig. 6d, Supplementary Note). Moreover, the γH2AX intensity significantly decreased upon overexpression of RNase H1 in *YTHDF2* KO HAP1 cells (Fig. 6e). Overall, these results suggest that YTHDF2 prevents accumulation of m$^6$A-containing RNA:DNA hybrids contributing to inhibition of R-loop-dependent DNA damage in mammalian cells. Correspondingly, YTHDF2 has been previously identified as one of the factors promoting genomic stability in a genome-wide siRNA screen[28].

The nature of the techniques we used for m$^6$A mapping is limited by the specificity and sensitivity of the available antibody. Even so, our results show that m$^6$A modification is present on the RNA within R-loops, potentially contributing to various aspects of their biology (Supplementary Note). In this context, the YTHDF2-mediated regulation of RNA:DNA hybrids may represent a specific mechanism of preventing accumulation of co-transcriptional R-loops during mitosis. Together with previously described factors suppressing formation of these structures[29–31], YTHDF2 plays a role in safeguarding genomic stability.

## Methods

### Cell culture, flow cytometry and RNA-interference-mediated knockdowns

REBL-PAT hiPSCs and HUES7 hESCs were maintained in Essential 8™ (E8) medium with supplement (#A1517001) on Matrigel™-coated tissue culture flasks at 37 °C with 5 % $CO_2$. Use of the HUES7 hESC line was approved by the UK Medical Research Council Steering Committee, in association with the UK Stem Cell Bank. Cells were passaged every 3–4 d using TrypLE™ Select Enzyme (#12563029). hiPSCs were treated with 1:100 dilution of KaryoMAX® Colcemid™ Solution (Thermo Fisher Scientific, catalogue number 15212012) for 3 h. HeLa, LN-18 and U87MG cells were maintained on DMEM (GIBCO) supplemented with 10 % bovine serum. $G_0/G_1$, S and $G_2$/M phases flow cytometry sorting was performed according to the previously described method[32]. 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 $G_0/G_1$, S and $G_2$/M cells using Beckman Coulter Astrios EQ and Beckman Coulter Kaluza 2.1 software. For *METTL3* and *YTHDF2* depletion, hiPSCs were transfected with 50 pmol of siRNA duplexes against human *METTL3* (Dharmacon™, catalogue number 56339), human *HNRNPA2B1* (Thermofisher, Catalogue number 4390824 siRNA, ID: s6714), human *YTHDF2* (Qiagen, catalogue number GS51441), human *RNase H1* (Dharmacon, Catalogue number M-012595-00-0010) and nontargeting siRNA #2 (Thermo Fisher Scientific, catalogue number D-001210-02) using DharmaFECT™ (GE Lifesciences) in antibiotic-free medium. Cells were collected for analysis 72 h after transfection. Expression of *METTL3, HNRNPA2B1, RNase H1* and *YTHDF2* was analysed by qPCR, according to standard procedures. Gene expression was normalized by comparison to levels 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 the exon 3 leading to frameshift and generating premature stop codon) HAP1 cells (Horizon Discovery, catalogue number HZGHC006678c001) and their isogenic wild type parental HAP1 cells (Horizon Discovery) were cultured on DMEM/F12 (Gibco Life Technologies, Catalog number 11320033) supplemented with 20 % heat-inactivated foetal bovine serum containing 1 % pen/strep at 37 °C in a humidified incubator with 5 % $CO_2$. Culture medium was changed daily and the cells were passaged using trypsin every 48 h. For determining the growth curve cells were counted using haemocytometer. Statistical significance was determined using 2-tailed t-test following assessment of the variance with F-test. The deletion in the 3$^{rd}$ exon of *YTHDF2* gene was validated by PCR (See Supplementary Table 1 for primer sequences) and by sequencing.

For overexpression of RNase H1 in mammalian cells we C-terminally eGFP-tagged human RNASEH1 (nuclear isoform) pEGFP-RNASEH1 plasmid[33]. This construct was a gift from Andrew Jackson & Martin Reijns (Addgene plasmid # 108699 ; http://n2t.net/addgene:108699 ; RRID:Addgene_108699). pmaxGFP™ (Lonza) was used as a control GFP-only plasmid.

### Animals and *Ythdf2* KO mouse model

Generation of the *Ythdf2* conditional knockout mice, followed by cre-mediated deletion and derivation of mNSCs from E14.5 embryonic forebrains were described previously[26]. 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).

### Immunocytochemistry, immunohistochemistry, confocal microscopy and image quantification

Immunocytochemistry and immunohistochemistry were performed as described[15, 34]. Sections of paraffin-embedded E14.5 wild type and *Ythdf2* KO mouse embryonic brain were used for γH2AX immunohistochemistry. The sections were dewaxed according to standard procedures. Cells were fixed in 4 % formaldehyde for 15 min. Cells and tissue sections were permeabilised with PBS containing 0.5 % Triton X-100 for 15 min. After permeabilisation, cells were treated with 25 mg/ml RNase A (Qiagen, catalogue number 19101) in PBS or with a mixture of 25 mg/ml RNase A and 10 U of RNase H in 1X RNase H buffer (NEB, catalogue number M0297S) overnight at 37 °C. DNase I (Qiagen, catalogue number 79254) treatment (20 U per sample) was carried out for 4 h at room temperature. The samples were incubated in 4N HCl for 1 h at 37 °C. Competition experiments were performed as described previously[15] using $N^6$-Methyl-2'-deoxyadenosine-5'-triphosphate (Trilink, catalogue number NU-949S) or unmodified dATP and dTTP from dNTP set (NEB, catalogue number N0446S). Immunostaining for RNA:DNA hybrids was performed according to previously published protocol[35] using S9.6 antibody (Merck Millipore, catalogue number MABE1095). The antibodies used for immunochemistry and their dilutions are provided in Supplementary Note. Control staining without primary antibodies produced no detectable signal. Images (500 nm optical sections) were acquired with a Zeiss LSM 700 AxioObserver confocal microscope using a Plan-Apochromat 63x/1.40 Oil DIC M27 objective and processed using Image J and Adobe Photoshop. 2.5XD signal intensity plots and intensity profiles were generated using ZEN Zeiss LSM 700 imaging software as described previously[15, 36]. Confocal raw data are available upon request. Co-localization coefficients were determined using the inbuilt analysis function of ZEN as described[15, 34]. Quantification of the $m^6A$, γH2AX and S9.6 signal intensities was performed according to the previously described method[34]. Mean values of the average of 18-60 nuclei signal intensities were calculated for each experimental point.

### Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

DNA and total RNA were isolated according to standard procedures. Up to 15 μg of purified DNA was digested to nucleosides for subsequent LC-MS analysis. Genomic DNAs and RNAs were digested to nucleosides by treatment with the Nucleoside Digestion Mix (NEB, M0649S) overnight at 37 °C. LC-MS/MS analysis was performed in duplicate by injecting digested DNAs and RNAs on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A Triple Quadrupole Mass Detector operating in the positive

electrospray ionization mode (+ESI). UHPLC was carried out using a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of methanol and 10 mM aqueous ammonium formate (pH 4.4). MS data acquisition was performed in the dynamic multiple reaction monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: dC at m/z 228→112, d5mC at m/z 242→126, d5hmC at m/z 258→142, d5fC at m/z 256→140, dA at 252→136, d6mA at 266→150, rC at m/z 244→112, $m^5C$ at 258→126, Cm at 258→112, rA at 268→136, $m^1A$ at 282→150, Am at 282→136, $m^6A$ at 282→150, and $m^6{}_2A$ at 296→164. External calibration curves with known amounts of the corresponding nucleosides were used to calculate the ratios within the samples analysed.

### Stable-isotope dilution ultra-performance liquid chromatography with tandem mass spectrometry (SID-UPLC-MS/MS)

The cells were resuspended in ice-cold buffer containing 10 mM Tris-HCl, 5 mM $Na_2EDTA$, 0.15 mM deferoxamine mesylate (pH 8.0) and 0.5 % SDS. The samples were incubated at 37 °C for 30 min followed by addition of 2.5 mg/ml Proteinase K and further incubation at 37 °C for 1.5 h. The nucleic acids were isolated using phenol/chloroform extraction and precipitated using ethanol. The precipitate was removed to another tube with plastic spatula, washed with 70 % ethanol and dissolved in MilliQ-grade deionized water. 5-10 μg of nucleic acids were treated with 5 U of RNase H (NEB) overnight in RNase H Reaction Buffer (NEB) at 37 °C. After incubation samples were ultrafiltered using Amicon Ultra-0.5 MWCO 3 kDa Centrifugal Filter (Merck) at 14000 g for 10 min. Subsequently, the samples were rinsed twice with MilliQ-grade deionized water (14000 g for 15 min). To recover the nucleic acids the filter was placed upside down in a clean microcentrifuge tube and centrifuged at 1000 g for 3 min. The ultrafiltrates containing released (oligo)nucleotides with molecular weight less than 3 kDa and the remaining nucleic acids were treated with 1U of nuclease P1 for 1 h in a buffer containing 200 mM ammonium acetate, 0.2 mM $ZnCl_2$ (pH 4.6) and 10 μg/sample tetrahydrouridine at 37°C, followed by addition of 10 % $NH_4OH$ and 1.3 U of alkaline phosphatase and subsequent additional 1 h incubation at 37 °C. Chromatographic analysis was performed using previously described method[37] adapted for determination of $m^6A$, mrC and adenosine (See details in Supplementary Note).

**RNA:DNA hybrids and protein co-immunoprecipitation with S9.6 antibody** was performed from non-cross-linked HeLa cells as described previously[22]. The genomic DNA was isolated from HeLa cells and sonicated as described[38]. The western blots of RNA:DNA hybrid IP samples were probed with the following antibodies: Top1 (Abcam, catalogue number ab109374, dilution 1:2000), YTHDF1 (ProteinTech, catalogue number 17479-1-AP, dilution 1:1000), YTHDF2 (ProteinTech, catalogue number 24744-1-AP, dilution 1:500), METTL3 (Bethyl Laboratories, catalogue number A301-567A, dilution 1:2000), HNRNPA2B1 (Novus, catalogue number NB120-6102SS, dilution 1:500) and Lamin B1 (Abcam, catalogue number ab16048, dilution 1:2000). Images were acquired by chemiluminescence using autoradiography.

### m$^6$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, catalogue number 19101) for 30 min in lysis buffer, phenol-chloroform extraction and ethanol precipitation. The DNA was fragmented to 300-600 bp using Covaris S2 ultrasonicator (Covaris Inc). Genomic DNA of the control samples was treated with 10 U of RNase H (NEB, catalogue number M0297S) in 1x RNase H buffer overnight at 37 °C before the immunoprecipitation. 10 μg of genomic DNA was used for immunoprecipitation. S9.6 DRIP was carried out essentially as described in the previously published protocol[39] using S9.6 antibody (Merck Millipore, catalogue number MABE1095) and anti-mouse magnetic Dynabeads (Invitrogen). m$^6$A DIP was performed using anti-m$^6$A rabbit polyclonal antibody (Synaptic systems, catalogue number 202003) and magnetic anti-rabbit Dynabeads (Invitrogen, M-280; polyclonal sheep anti-rabbit IgG; catalogue number 10716653) with denaturation step before the IP (10 min at 95º C) analogously to meDIP technique[40] (Supplementary Fig. 6). The corresponding primary IgG-only and secondary IgG only (Dynabeads only) DRIP reactions were used in control immunoprecipitations. For the two round (S9.6 DRIP followed by m$^6$A DIP) DRIP/DIP, approximately 500 ng of the nucleic acids recovered from multiple DRIP reactions performed in parallel were used for m$^6$A DIP followed by qPCR analysis. For S9.6 DRIP followed by m$^6$A RIP experiment, nucleic acids recovered from DRIP were denatured for 30 min at 95 ºC followed by digestion of the DNA components of RNA:DNA hybrids using Turbo DNase (Invitrogen, catalogue number: AM1907) for 30 min at 37 ºC. After inactivation of DNase, approximately 100 ng of the recovered RNA was used for the m$^6$A RIP performed using EpiMark® N6-Methyladenosine Enrichment Kit (NEB, catalogue number E1610S). The eluted RNA was reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen, catalogue number 2072529A) and random hexamers (Invitrogen, catalogue number 1831815) and analysed by qPCR. IgG-only reactions and reactions carried out without reverse transcription were used as controls.

For validation of m$^6$A-DIP using synthetic oligonucleotides, 0.1-1 pmol of m$^6$A-containing- and non-modified RNA:DNA hybrids and individual single stranded RNA or DNA oligonucleotides were spiked-in with 5 μg of mouse genomic DNA. The RNA and DNA oligonucleotides used for spike-in experiments and primers used to amplify spike-in controls are listed in Supplementary Table 1. The RNA oligonucleotides were synthesised by Dharmacon. To generate RNA:DNA hybrids, the RNA and DNA oligonucleotides were mixed in equimolar concentrations, incubated for 20 min at 98º C, slowly cooled down in a heating block and placed on ice. Quantitative PCR analysis of m$^6$A DIP and S9.6 DRIP samples was carried out with SYBR Green PCR Master Mix (Sigma) according to standard procedures. Fold enrichment was calculated as $2^{ddCt}$, where dCt=Ct(enriched)-Ct(input) and ddCt=dCt – Ct (IGG). The primers used for DRIP/DIP-qPCR and qPCR analysis of the corresponding transcripts are listed in Supplementary Table 1. The primers for α-satellites were obtained from Novus (catalogue number NBP1-71654SS). Generic primers amplifying Alu elements from the indicated families and evolutionarily young L1Hs were used. Human LINE-1 primers were designed to detect L1PA1 and L1PA2 classes of these retroelements. The primers for mouse LINE-1 ORF1 were previously published[41]. The primers used for

DRIP-qPCR of Alu-Y, Alu-S and LINE-1s were also employed for qPCR. Gene expression was normalized by comparison to levels of *GAPDH* gene expression.

**Chromatin immunoprecipitation (ChIP)** was performed using EZ-Magna ChIP™ A/G Chromatin Immunoprecipitation Kit (Merck, catalogue number 17-10086) according to manufacturer's instructions using anti-YTHDF2 rabbit polyclonal (ProteinTech, catalogue number 24744-1-AP), anti-HNRNPA2B1 mouse monoclonal (Novus, catalogue number NB120-6102SS) and anti-γH2AX mouse monoclonal (Merck, catalogue number 05-636, clone JBW301) primary antibodies. ChIP was analysed by quantitative PCR carried out with SYBR Green PCR Master Mix (Sigma) according to standard procedures. Fold enrichment was calculated as $2^{ddCt}$, where dCt=Ct(enriched)-Ct(input) and ddCt=dCt – Ct (IGG). The primers used for DRIP-qPCR were also used for ChIP-qPCR analysis.

## Purification of recombinant YTHDF2

Full-lengths YTHDF2 was cloned into pET-28b with N-terminal His-tag (Genescript). The plasmid was expressed in BL21(DE3) cells and incubated with LB-medium (Puls medical, 244610) with sorbitolat at 37° C until OD600 0.7. Expression was induced with 300 μM IPTG overnight at 18° C. Cells were pelleted by centrifugation at 3000 g for 10 min at 4° C, resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM ME, 10 mM imidazole) and sonicated. After sonication, the extract was centrifuged at 19000 g for 20 min at 4° C. The supernatant was loaded to Protino® 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 3000 g for 2 min, the Ni-agarose-bound YTHDF2 was washed with the buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM ME and 50 mM imidazole. Recombinant YTHDF2 was eluted with 5 washes (1.5 ml each) of elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM ME and 300mM imidazole).

**Microscale thermophoresis (MST)** was employed to study the interaction of YTHDF2 with modified or unmodified- RNA and RNA:DNA hybrid synthetic substrates used in EMSA experiments. The purified full-lengths YTHDF2 with N-terminal His-tag was labelled with the NT-647 RED-tris-NTA dye for His-tagged proteins following the recommended labelling protocol (Nanotemper Technologies). 100 μl of 200 nM protein in PBS buffer was mixed with 100 μl of 100 nM dye in the supplemented PBS-T buffer and incubated at room temperature for 30 minutes. The labelled protein was then added to a 1:1 dilution series of the respective substrate in UltraPure DNase/RNase-Free Distilled Water (FisherScientific) to a final concentration of 50 nM. The different protein-substrate samples were loaded into standard NT.115 MST capillaries (Nanotemper). Fluorescence profiles were measured at 25° C in a Monolith NT.115 instrument using the red channel. Data was collected at 20 % (single-strand substrates) or 40 % (hybrid substrates) MST power and 60 % excitation power. The changes in fluorescence ($\Delta F_{norm}$) due to thermophoresis were measured as the signal difference between time points 0 and 10 seconds. Data were normalized and plotted as a function of the ligand concentration, and a binding curve was fitted to the average of six independent dilution series of each substrate.

### Library preparation and high throughput sequencing

Sequencing libraries were prepared according to the NEB Next DNA Ultra Library Preparation Kit for Illumina (NEB, E7370). DNA was sonicated to 400-600 bp (Covaris S2) and adapters were ligated (NEB, E73355S) 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, KK4823). Sequencing was performed using the Illumina NextSeq500 platform to generate 2 x 150 bp reads. Primary IgG-only DRIP reactions resulted in the DNA amounts insufficient for successful library production even with 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-seq libraries were constructed using the Illumina TruSeq Stranded Total RNA sample preparation kits (Illumina, Inc., San Diego, CA), according to the manufacturers guidelines, and then sequenced on 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[42]. Reads that passed filtering were aligned to the human Ensembl genome (build hg38.89) using BWA with default parameters[43]. As each biological sample was split across multiple lanes of sequencing, the corresponding alignments were merged with Samtools[44] and de-duplicated to remove PCR artefacts with picard-tools MarkDuplicates[45]. The impact of each pulldown was assessed using Phantompeakqualtools[46] and the highly modified regions (HMRs, peaks) were identified using MACS2.1.1[46, 47]. As the exact mode of genomic distribution of $m^6A$-containing RNA:DNA hybrids was initially unknown, we performed detection of both narrow and broad peaks using -q 0.01 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[48]. We performed peak calling against input DNA and against secondary IgG-only control samples. More than 96 % of the $m^6A$ DRIP peaks called against input were also identified using IgG-only controls. Peaks called against input were used for further analysis. Consensus peaks were defined using the dba.peakset() function to select for peaks overlapping in both replicates. In each instance the replicate sample BAM/bed files along with the corresponding input samples were used as input. Additional details of bioinformatics analysis are provided in Supplementary Note. Details on software and data deposition are listed in the Life Sciences Reporting Summary.

### Statistics and reproducibility

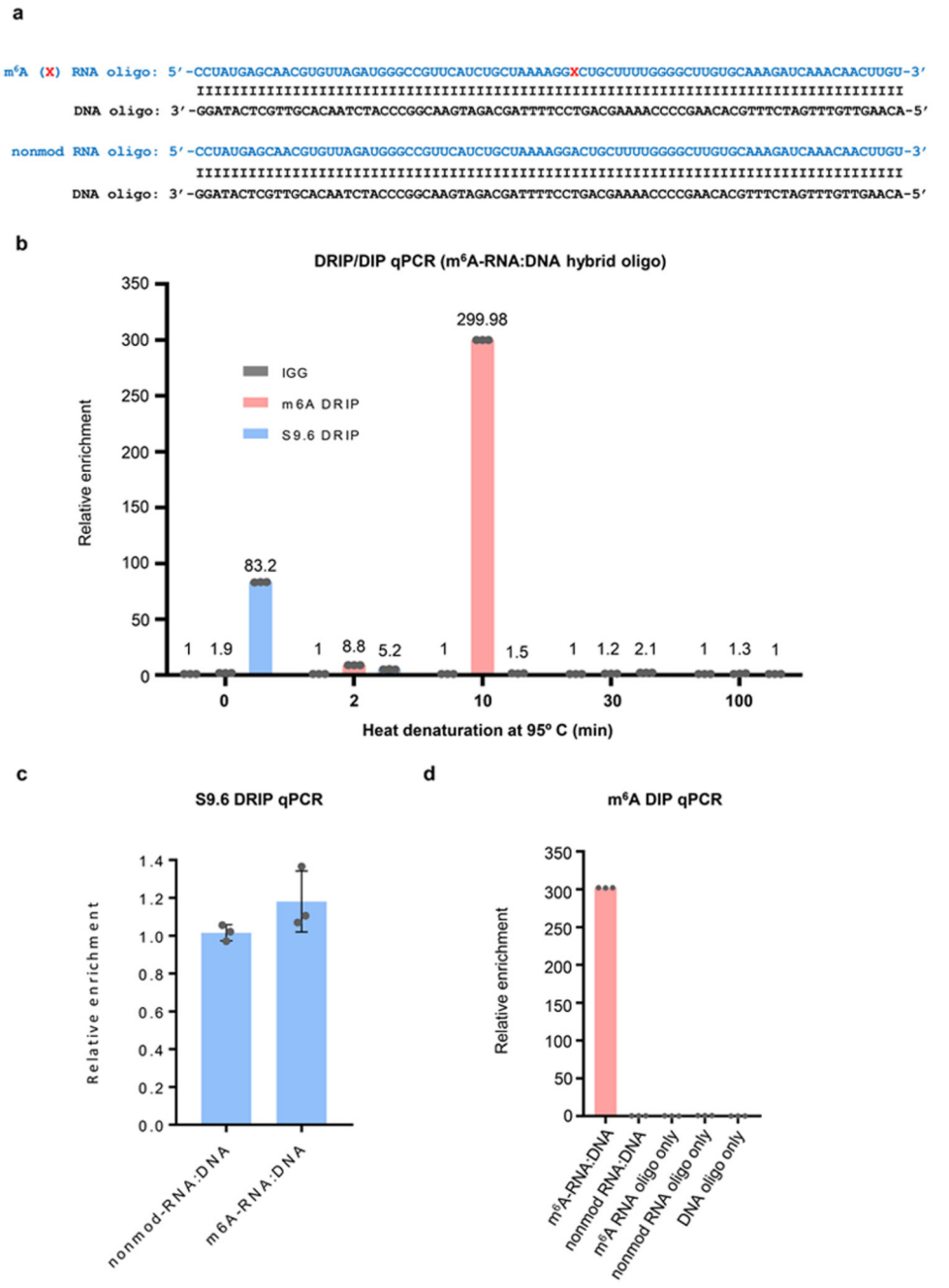
At least 2 and typically 3 independent experiments were carried out for most of the assays. DRIP and DIP were performed in two and RNaseq in three biologically independent experiments. All experiments were replicated independently. We observed generally good correlation between the replicates. Statistical tests used for individual experiments are described in corresponding figure legends. For quantification of the $m^6A$, $\gamma$H2AX and S9.6

signal intensities, statistical significance was determined using unpaired two-tailed Student’s t-test or unpaired two-tailed Welch's t-test. Signal intensity and qPCR data were plotted and analyzed in GraphPad Prism 7.04.

## Extended Data



Europe PMC Funders Author Manuscripts




**Extended Data Fig. 1. Validation of m$^6$A DIP using synthetic RNA and DNA substrates.**

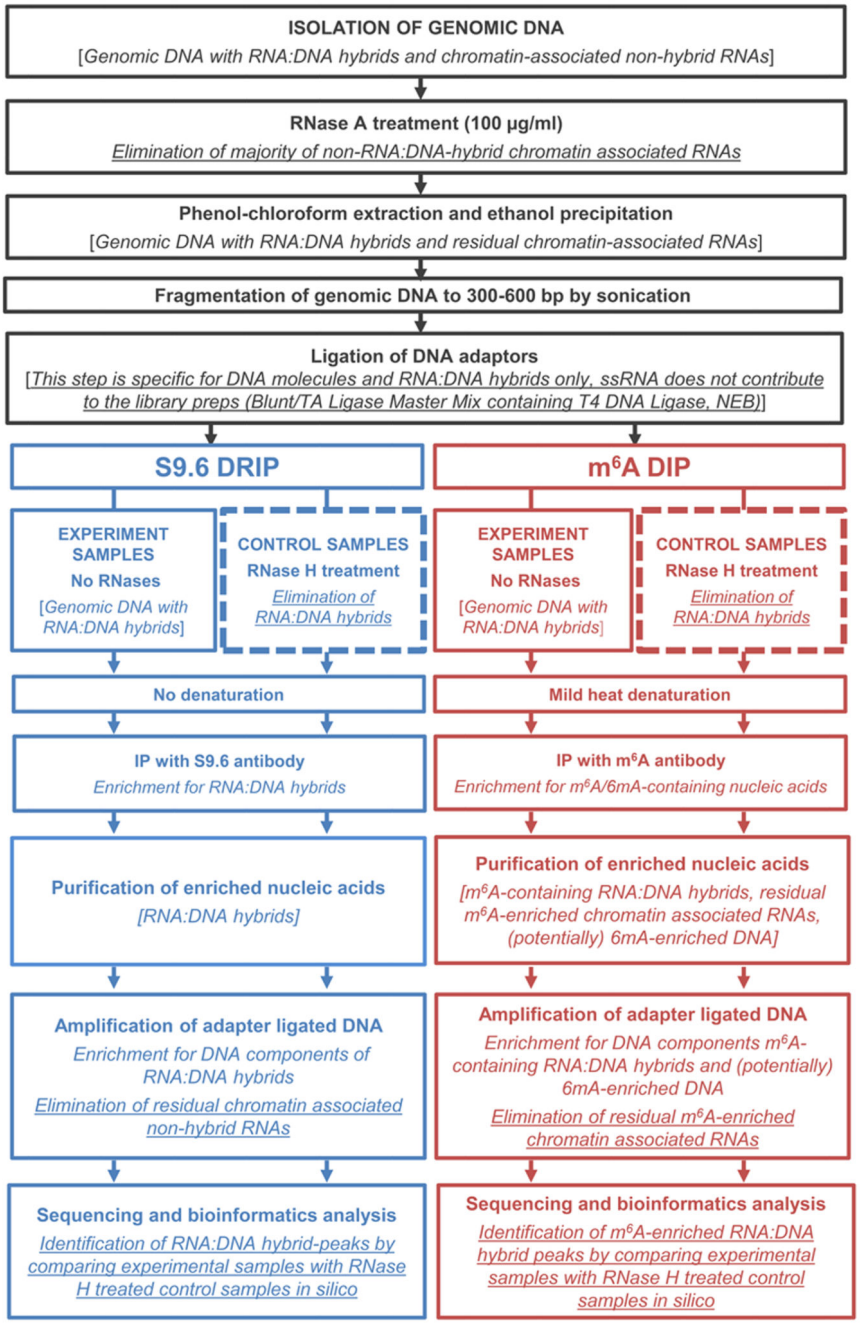
Validation of m$^6$A DIP using synthetic RNA and DNA substrates. See also Supplementary Note. (a) Schematic demonstrating the sequences of synthetic oligonucleotides and RNA:DNA hybrids used for spike in validation experiments. (b) The relative enrichment of DIP/DRIP performed on 0.1 pmol of spike in synthetic m$^6$A-containing RNA:DNA hybrid (81-mer with 46 % GC content) shown in (a) using anti-m$^6$A 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$^6$A-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.

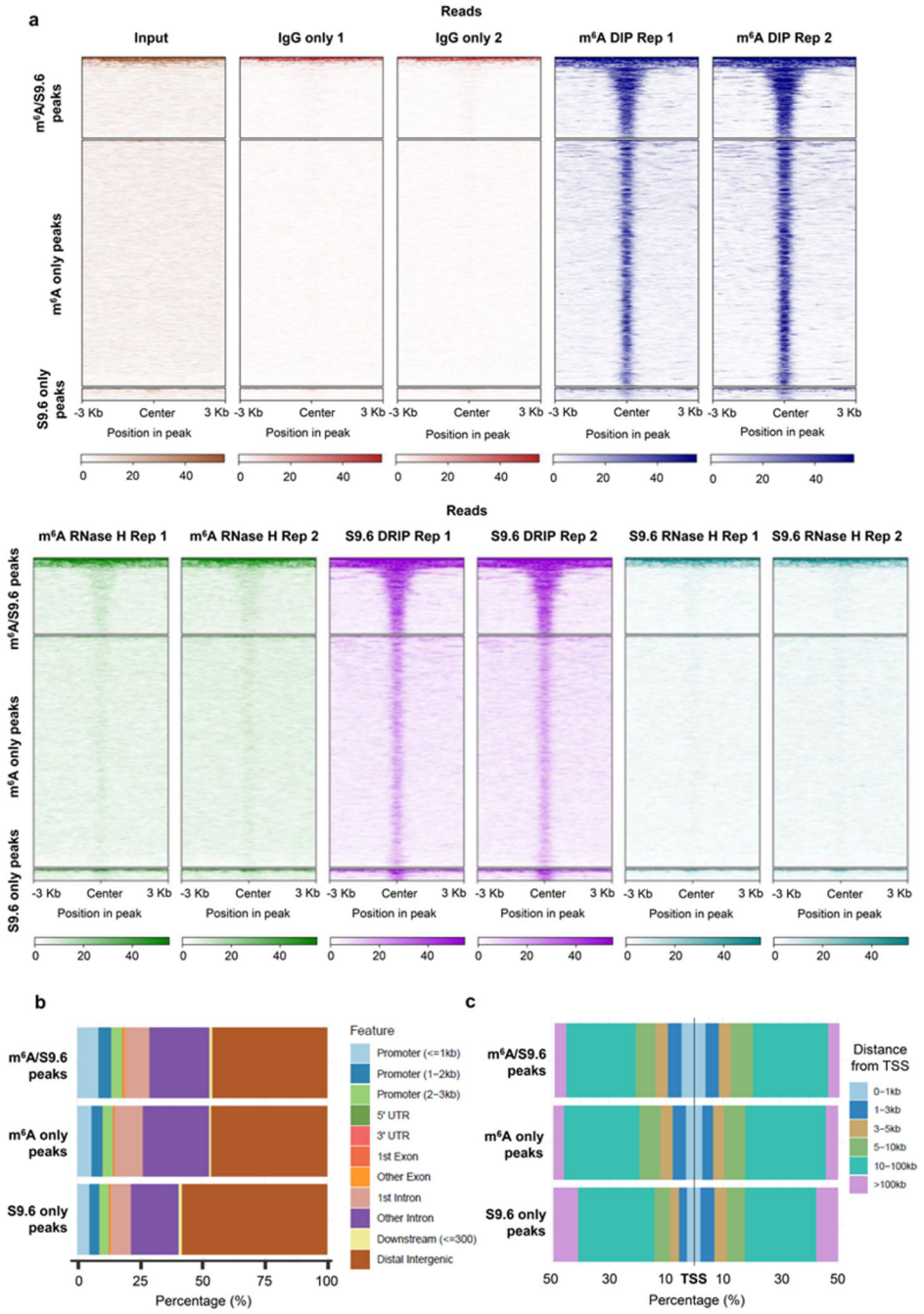


Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts




**Extended Data Fig. 2. Detailed schematic illustrating S9.6 DRIP and m$^6$A DIP techniques.**
Detailed schematic illustrating S9.6 DRIP and m$^6$A DIP techniques.




**Extended Data Fig. 3. m$^6$A is present on the majority of RNA:DNA hybrids in hiPSCs.**

m$^6$A 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$^6$A peaks overlapping with S9.6 peaks (m$^6$A/S9.6), m$^6$A peaks that do not overlap with S9.6 DRIP peaks (m$^6$A only) and S9.6 peaks that do not correspond to m$^6$A 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.

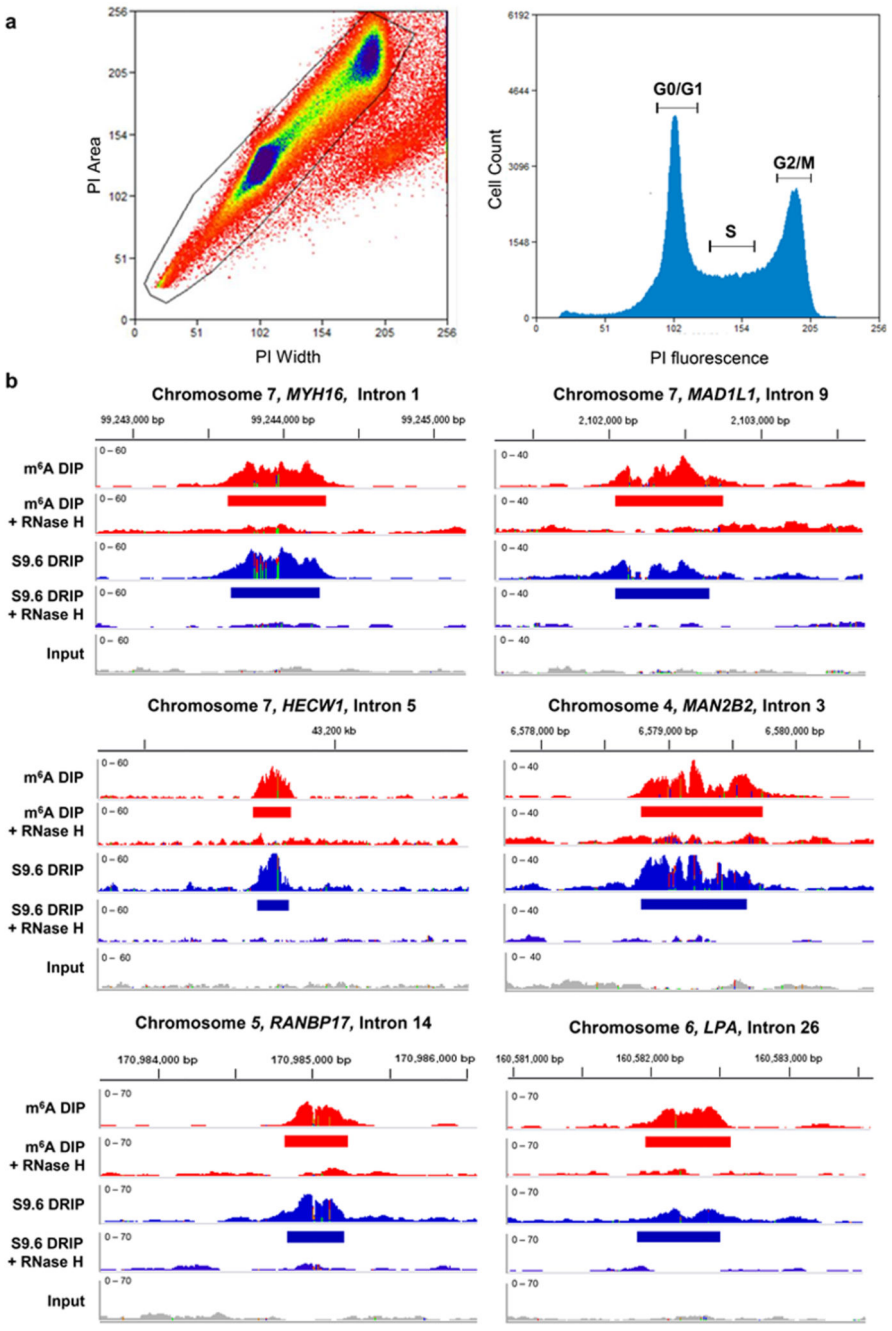



Europe PMC Funders Author Manuscripts

(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.



Europe PMC Funders Author Manuscripts

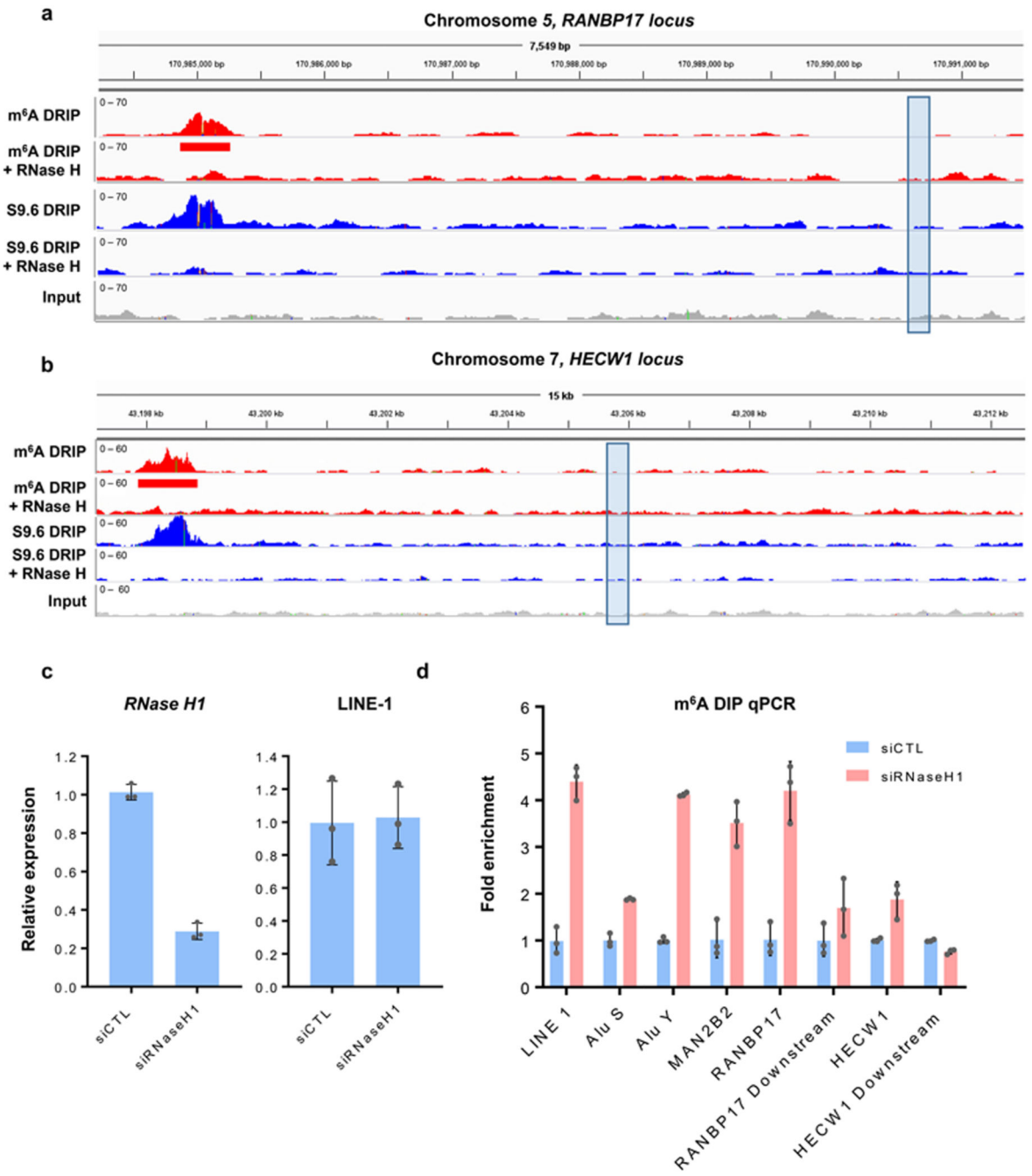



**Extended Data Fig. 4. RNA:DNA hybrids exhibit cell cycle-specific dynamics in hPSCs.**

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 $G_0/G_1$, S and $G_2$/M phases are marked. (b) The coverage plots of $m^6A$ DIP and S9.6 DRIP densities (CPK) in the intronic regions of the indicated genes. $m^6A$ and S9.6 peaks are marked with red and blue rectangles.




Europe PMC Funders Author Manuscripts



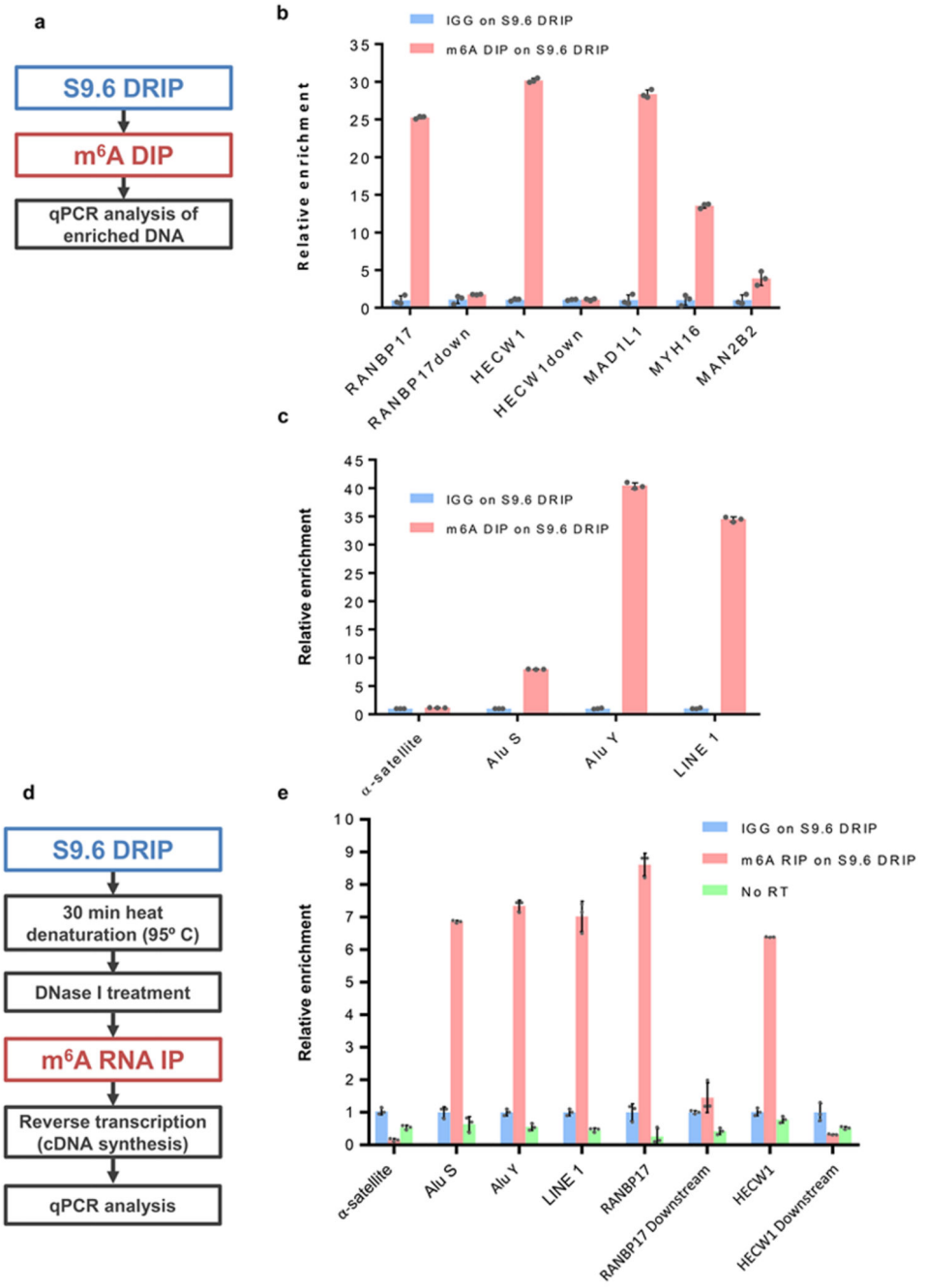

**Extended Data Fig. 5. m$^6$A DIP signal is increased upon RNase H1 knockdown in hPSCs.**

m$^6$A DIP signal is increased upon RNase H1 knockdown in hPSCs. (a, b) The coverage plots of m$^6$A and S9.6 DRIP/DIP densities (CPK) in the regions located downstream of the m$^6$A 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$^6$A DIP qPCR of the indicated repeats and intronic sequences performed on siCTL and siRNaseH1 hiPSCs. 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.



Europe PMC Funders Author Manuscripts



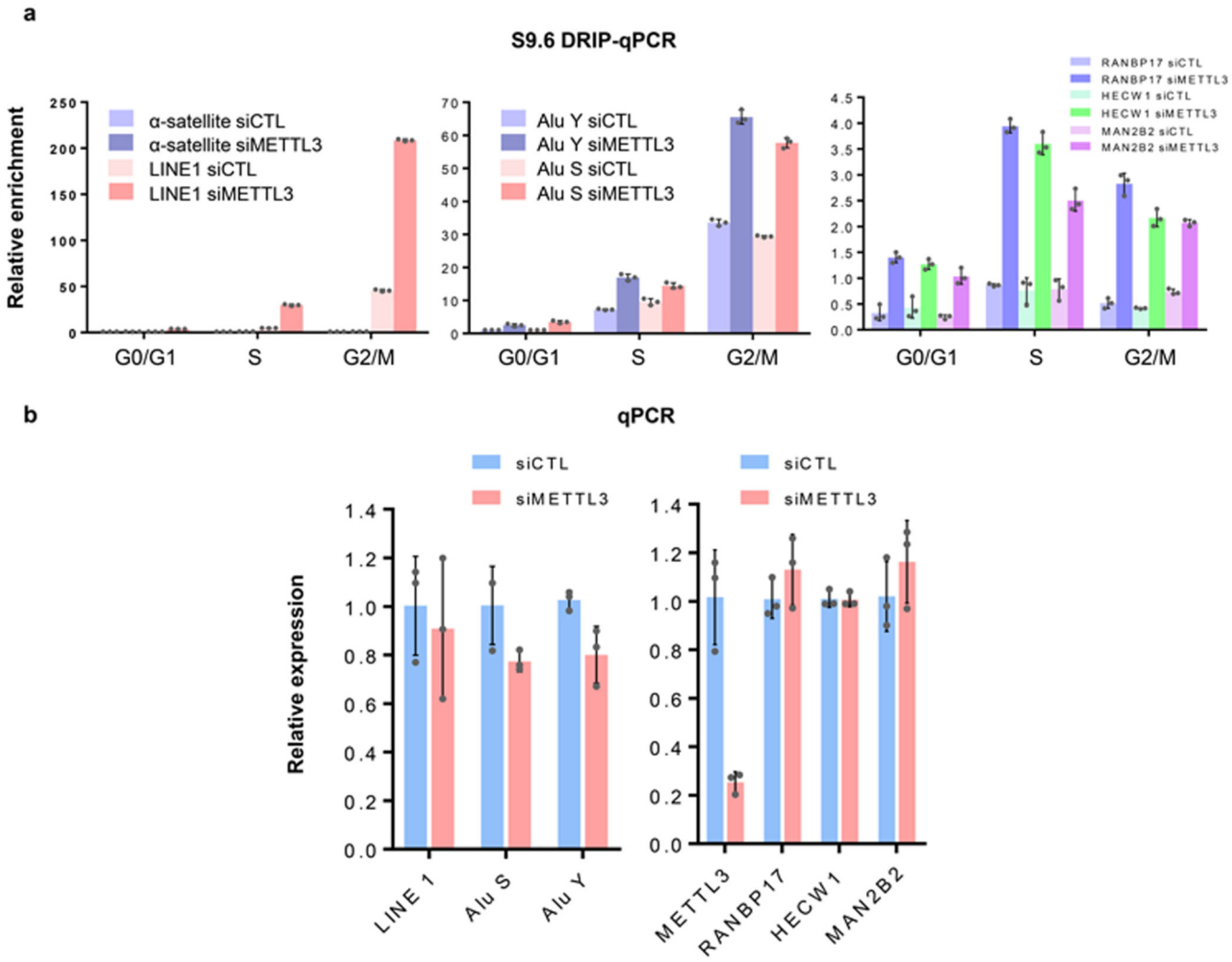

**Extended Data Fig. 6. m$^6$A is detectable on the RNA components of R-loops.**

m$^6$A is detectable on the RNA components of R-loops. (a) Schematic representation of the two round (S9.6 followed by m$^6$A) DRIP/DIP procedure. (b, c) The results of the two round (S9.6 DRIP followed by m$^6$A DIP) DRIP/DIP performed on individual intronic (b) and repetitive (c) m$^6$A/S9.6 peak containing sequences. (d) Schematic representation of the m$^6$A 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$^6$A RIP performed on the S9.6 DRIP of individual and repetitive DRIP/m$^6$A 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.



Europe PMC Funders Author Manuscripts

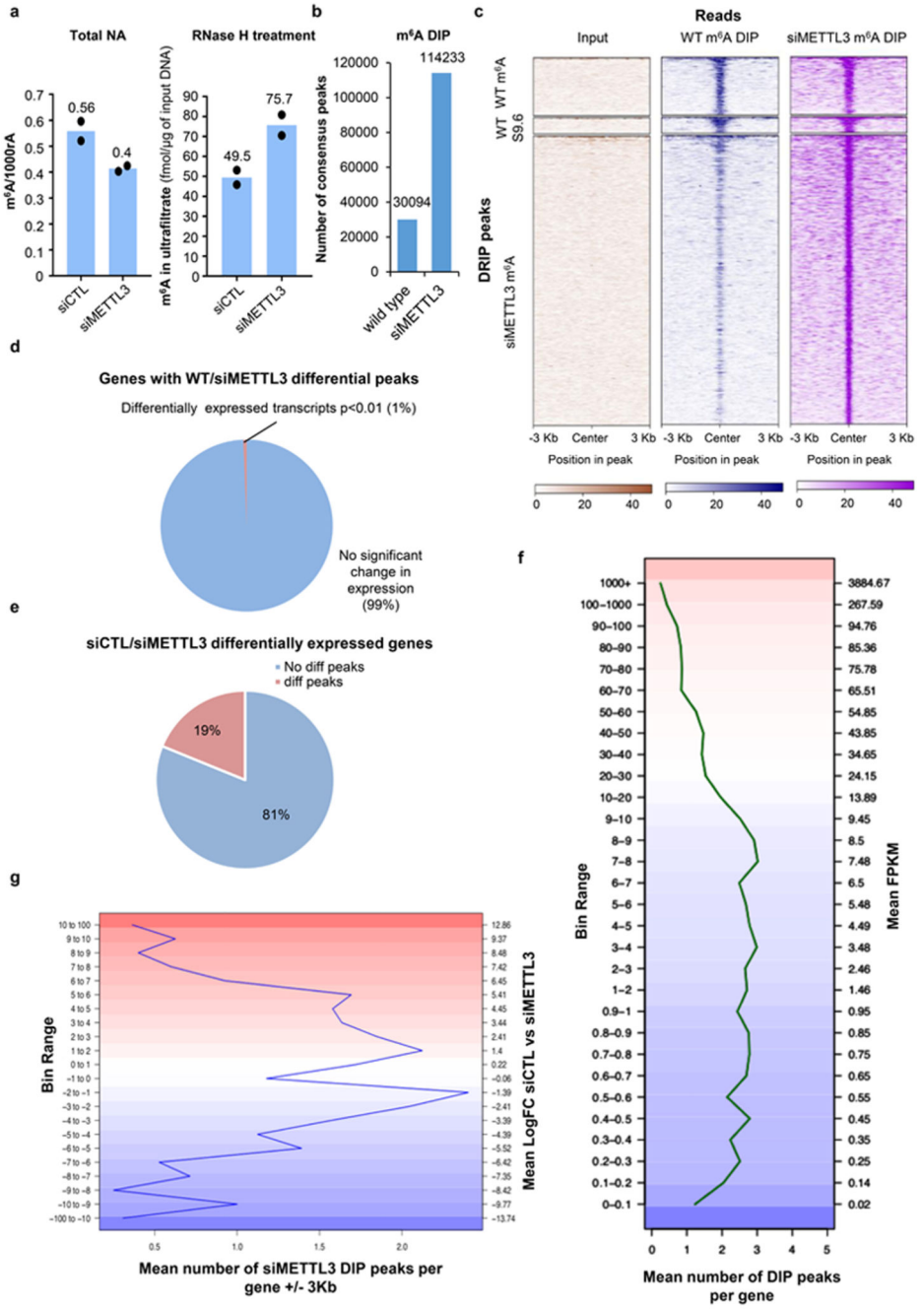



**Extended Data Fig. 7. METTL3 depletion leads to accumulation of RNA:DNA hybrids in hPSCs.**

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.




Europe PMC Funders Author Manuscripts




**Extended Data Fig. 8. METTL3 depletion leads to an increase in the m$^6$A DIP peaks in hPSCs.**

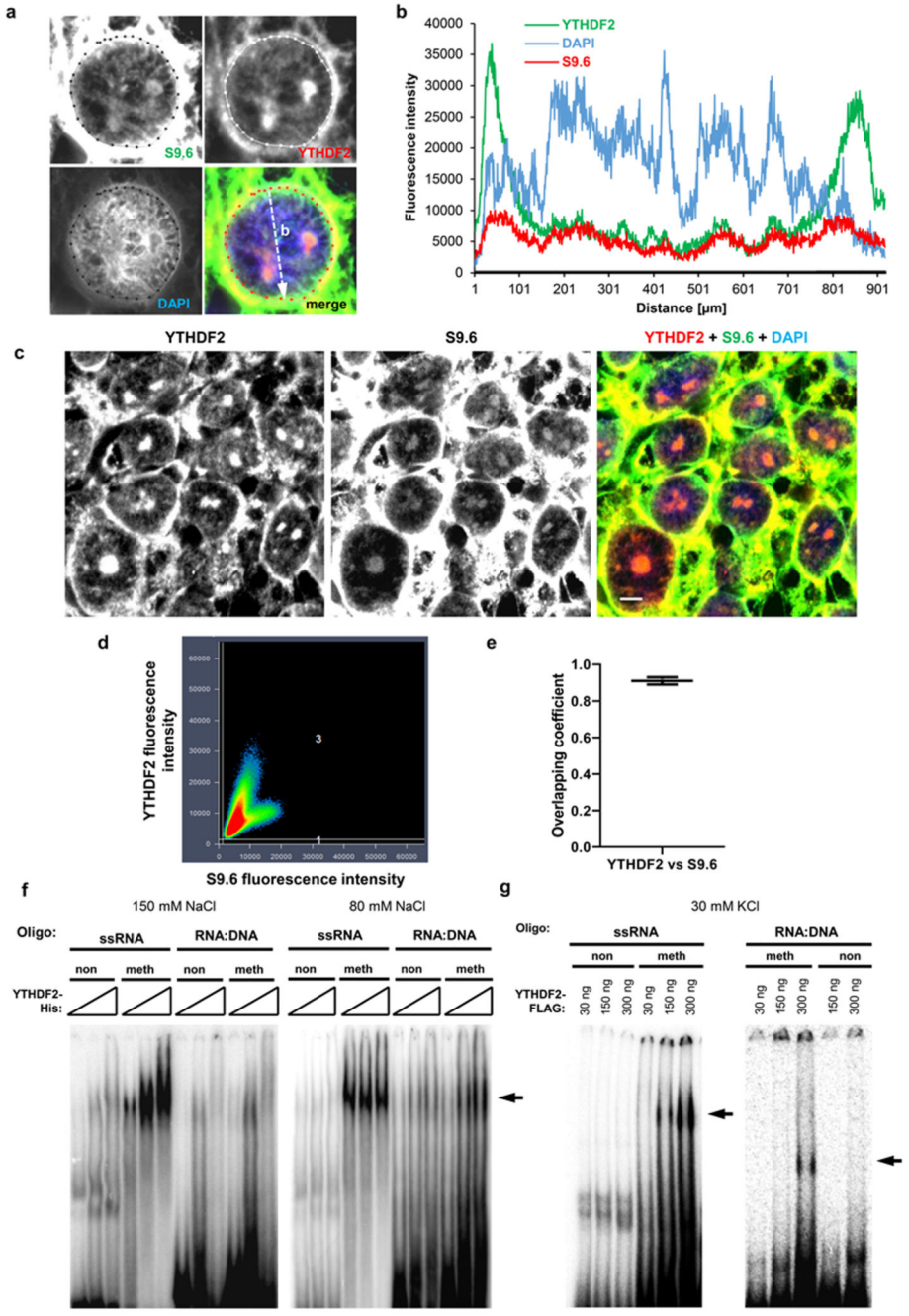
METTL3 depletion leads to an increase in the m$^6$A DIP peaks in hPSCs. (a) SID-UPLC-MS/MS quantification of m$^6$A in the total nucleic acids (left panel), and in ultrafiltrate fractions released upon RNase H treatment of siCTL and siMETTL3 hPSCs (right panel). The data are means of 2 independent experiments. (b) The total numbers of consensus m$^6$A peaks in WT and siMETTL3 hiPSCs. (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$^6$A DIP in WT and siMETTL3 hPSCs. The colour of each line represents




Europe PMC Funders Author Manuscripts

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.

Europe PMC Funders Author Manuscripts

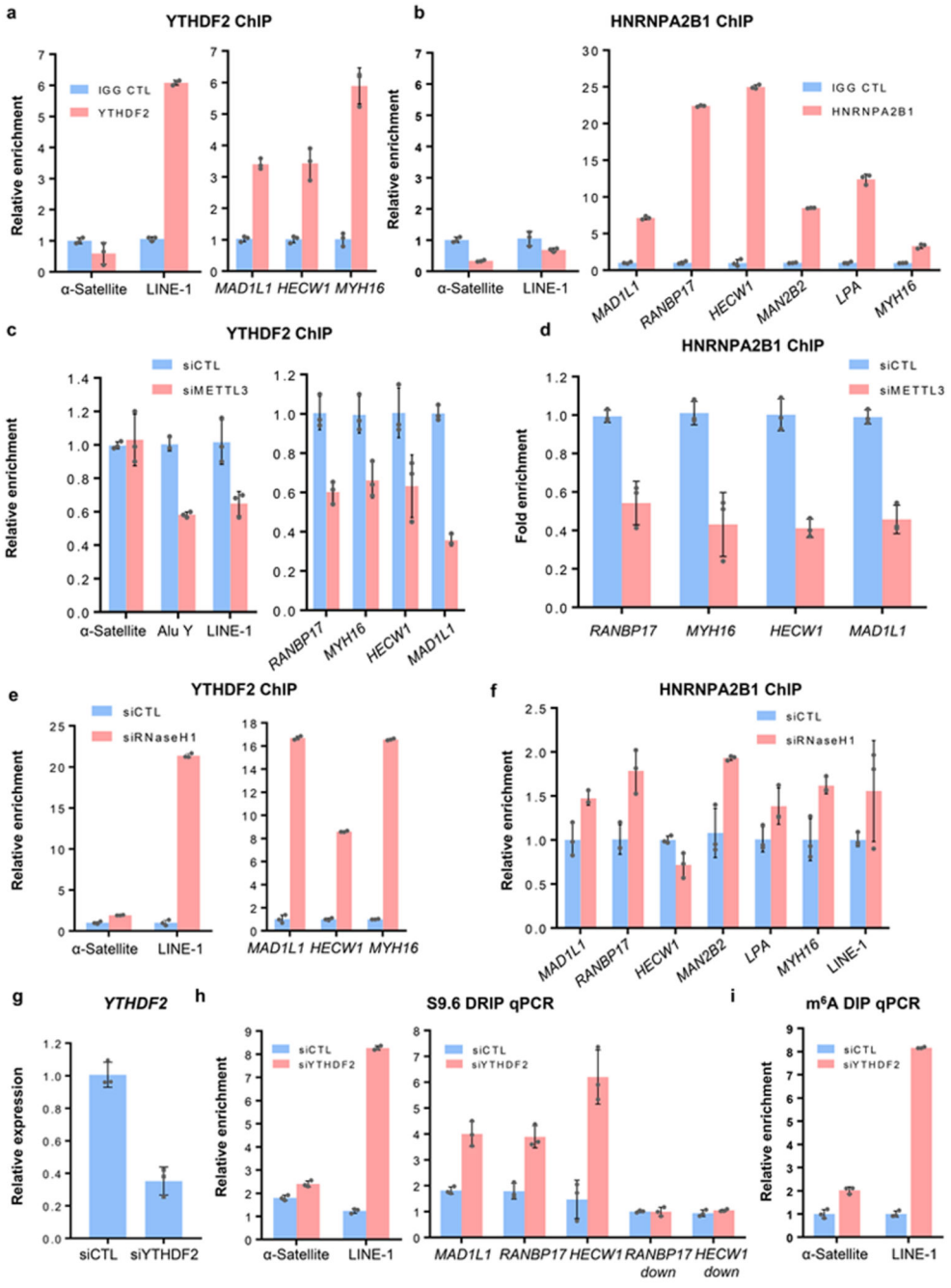



**Extended Data Fig. 9. YTHDF2 co-localizes with R-loops *in vivo* and interacts with synthetic m$^6$A-marked RNA:DNA hybrids *in vitro*.**

YTHDF2 co-localizes with R-loops *in vivo* and interacts with synthetic m$^6$A-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 DAPI signals across the nuclear region marked with an arrow in (a). (c) Representative image of hPSCs immunostained for YTHDF2 and

R-loops. Merged view and YTHDF2/S9.6 channels are shown. Scale bars are 20 μm 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 hPSCs 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 $m^6A$-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.

Europe PMC Funders Author Manuscripts




**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 m$^6$A 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank V. Wright, S. Malla, M. Loose, S. Rajani, D. Mosqueira, L. Lewis, R. Marcus, D. Onion, D. Bates and B. Coyle for technical help, D. Huangfu (Memorial Sloan Kettering Cancer Center) for hPSCs promoters dataset and S. Peña Perez for animal care. This work was supported by Medical Research Council IMPACT DTP PhD Studentship [grant number MR/N013913/1] to A.A and by Biotechnology and Biological Sciences Research Council [grant number BB/N005759/1] to A.R. N.G.'s lab is supported by a Royal Society University Research Fellowship and a John Fell award [grant number BVD07340] to N.G. J.L.G.P.'s lab is supported by MINECO-FEDER [grant number SAF2017-89745-R] to J.L.G.P., the European Research Council [grant number ERC-Consolidator ERC-STG-2012-233764] to J.L.G.P., and a private donation from Ms Francisca Serrano (Trading y Bolsa para Torpes, Granada, Spain). C.D. is supported by British Heart Foundation [SP/15/9/31605, PG/14/59/31000], Medical Research Council [MR/M017354/1], NC3Rs [CRACK-IT:35911-259146], and Heart Research UK [TRP01/12]. L.Y. was supported by Medical Research Council [grant number MR/M017354/1]. A.K. is supported by the Norwegian Research Council [grant number 275286] and the Health Authority South-East [grant number 2018086]. B.D. is supported by the Norwegian Health Authority South-East (Regional Core Facility for Structural Biology) [grant number 2015095] to B.D. D.J.G. is funded by Wellcome Trust [grant number 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.

## References

1. Chédin F. Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet. 2016; 32:828–838. [PubMed: 27793359]
2. Sanz LA, et al. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell. 2016; 63:167–178. [PubMed: 27373332]
3. Lu WT, et al. Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. Nat Commun. 2018; 9
4. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. Nat Rev Genet. 2015; 16:583–597. [PubMed: 26370899]
5. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev. 2015; 29:343–355.
6. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017; 169:1187–1200. [PubMed: 28622506]
7. Wang X, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014; 7481:117–120.
8. Zhang G, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015; 161:893–906. [PubMed: 25936838]
9. Greer EL, et al. DNA Methylation on N6-Adenine in C. elegans. Cell. 2015; 161:868–78. [PubMed: 25936839]
10. Fu Y, et al. N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell. 2015; 161:879–892. [PubMed: 25936837]
11. Mondo SJ, et al. Widespread adenine N6-methylation of active genes in fungi. Nat Genet. 2017; 49:964–968. [PubMed: 28481340]

12. Liu J, et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat Commun. 2016; 7
13. Yao B, et al. DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. Nat Commun. 2017; 8
14. Schiffers 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. 2017; 56:11268–11271. [PubMed: 28371147]
15. Wheldon LM, et al. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. Cell Rep. 2014; 7:1353–1361. [PubMed: 24882006]
16. Santos F, Dean W. Using immunofluorescence to observe methylation changes in mammalian preimplantation embryos. Methods Mol Biol. 2006; 325:129–137. [PubMed: 16761724]
17. Rieder CL, Palazzo RE. Colcemid and the mitotic cycle. J Cell Sci. 1992; 102:387–92. [PubMed: 1506421]
18. Pfaffeneder T, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl. 2011; 50:7008–7012. [PubMed: 21721093]
19. Chan YA, et al. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet. 2014; 10:e1004288. [PubMed: 24743342]
20. Boguslawski SJ, et al. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods. 1986; 89:123–130. [PubMed: 2422282]
21. Mita P, et al. LINE-1 protein localization and functional dynamics during the cell cycle. Elife. 2018; 7
22. Cristini A, Groh M, Kristiansen MS, Gromak N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep. 2018; 23:1891–1905. [PubMed: 29742442]
23. Wang X, et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell. 2015; 161:1388–1399. [PubMed: 26046440]
24. Alarcón CR, et al. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. Cell. 2015; 162:1299–12308. [PubMed: 26321680]
25. Carette JE, et al. Haploid genetic screens in human cells identify host factors used by pathogens. Science. 2009; 5957:1231–1235.
26. Li M, Zhao X, Wang W, Shi H, Pan Q, Lu Z, Perez SP, Suganthan R, He C, Bjørås M, Klungland A. Ythdf2-mediated m(6)A mRNA clearance modulates neural development in mice. Genome Biol. 2018; 19:69. [PubMed: 29855337]
27. Bonner WM, et al. GammaH2AX and cancer. Nat Rev Cancer. 2008; 8:957–967. [PubMed: 19005492]
28. Paulsen RD, et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol Cell. 2009; 35:228–239. [PubMed: 19647519]
29. Salas-Armenteros I, Pérez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J. 2017; 36:3532–3547. [PubMed: 29074626]
30. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nat Commun. 2017; 8
31. Benitez-Guijarro M, et al. RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. EMBO J. 2018; 37doi: 10.15252/embj.201798506
32. Pozarowski P, Darzynkiewicz Z. Analysis of cell cycle by flow cytometry. Methods Mol Biol. 2004; 281:301–311. [PubMed: 15220539]
33. Bubeck D, Reijns MA, Graham SC, Astell KR, Jones EY, Jackson AP. PCNA directs type 2 RNase H activity on DNA replication and repair substrates. Nucleic Acids Res. 2011; 39:3652–3266. [PubMed: 21245041]
34. Abakir A, Wheldon L, Johnson AD, Laurent P, Ruzov A. Detection of Modified Forms of Cytosine Using Sensitive Immunohistochemistry. J Vis Exp. 2016; 114

35. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell. 2017; 170:774–786. [PubMed: 28802045]
36. Ramsawhook AH, et al. Immunostaining for DNA Modifications: Computational Analysis of Confocal Images. J Vis Exp. 2017; 127
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. 2016; 88:12128–12136. [PubMed: 28193047]
38. Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014; 10:e1004318. [PubMed: 24787137]
39. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012; 45:814–825. [PubMed: 22387027]
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. 2005; 37:853–862. [PubMed: 16007088]
41. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017; 49:1502–1510. [PubMed: 28846101]
42. Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014; 15:182. [PubMed: 24925680]
43. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
44. Li H, et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
45. Chang VY, Federman N, Martinez-Agosto J, Tatishchev SF, Nelson SF. Whole exome sequencing of pediatric gastric adenocarcinoma reveals an atypical presentation of Li-Fraumeni syndrome. Pediatr Blood Cancer. 2013; 60:570–574. [PubMed: 23015295]
46. Marinov GK, Kundaje A, Park PJ, Wold BJ. Large-scale quality analysis of published ChIP-seq data. G3 (Bethesda). 2014; 4:209–223. [PubMed: 24347632]
47. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
48. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 7381:389–393.


Europe PMC Funders Author Manuscripts

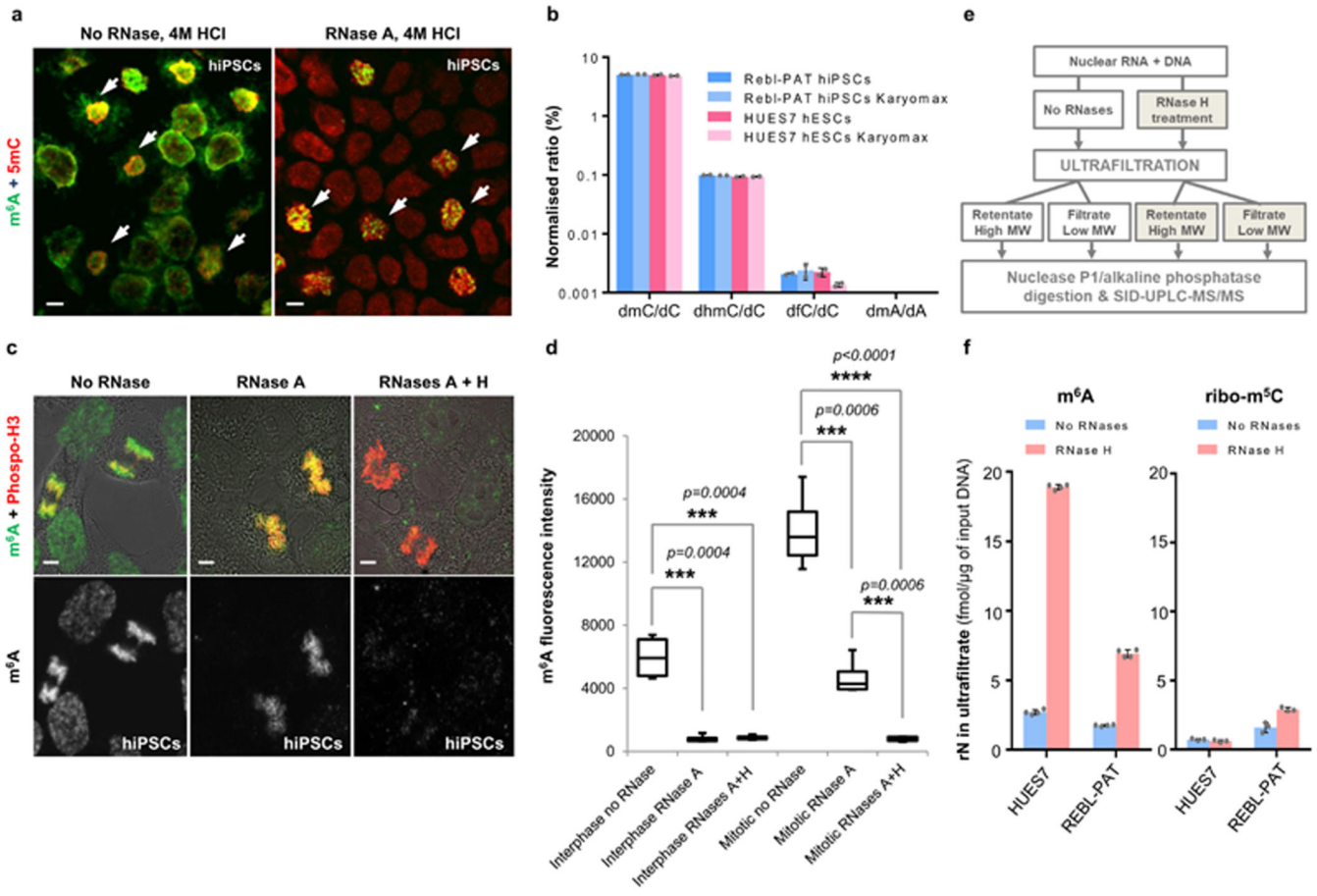



**Fig. 1.**

m$^6$A marks the RNA components of RNA:DNA hybrids in hPSCs. (a) m$^6$A and 5-methyldeoxycytosine (5mC) co-immunostaining of KaryoMAX-treated hiPSCs without RNases and after RNase A treatment. Merged images are shown. Mitotic cells are arrowed. (b) The ratios of the indicated deoxynucleotides obtained from the quantification of LC-MS/MS peaks in KaryoMAX-treated and untreated hiPSCs/hESCs DNA. Data are means ± SD, n=2 MS experiments. (c) Immunostaining of hiPSCs using anti-m$^6$A and anti-phospho-Histone H3 antibodies without RNases and after RNase A or combined RNases A/H treatments. Merged views are presented. (d) Box plots showing quantification of m$^6$A signal intensity in the interphase and mitotic hiPSCs at indicated immunostaining conditions. The elements of the box plots are: center line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all the data; n=20 nuclei for each condition. Significance was determined by unpaired two-tailed Student's t-test. No adjustments were made for multiple comparisons. (e) Schematic illustrating design of the experiment on SID-UPLC-MS/MS analysis of hPSCs-derived nucleic acids released and retained upon RNase H treatment. (f) SID-UPLC-MS/MS quantification of m$^6$A and ribo-m$^5$C in the fractions of hESCs- and hiPSCs-derived nucleic acids released upon RNase H treatment. Data are shown as means ± SD, n=4/n=3 MS experiments for m$^6$A/ribo-m$^5$C quantification. Scale bars are 10 µm in (a) and 5 µm in (c). KaryoMAX treatment was used to enrich hPSCs for mitotic




Europe PMC Funders Author Manuscripts

cells in (a, b). The experiments shown in (a, c) were repeated independently 6 times with similar results.



Europe PMC Funders Author Manuscripts

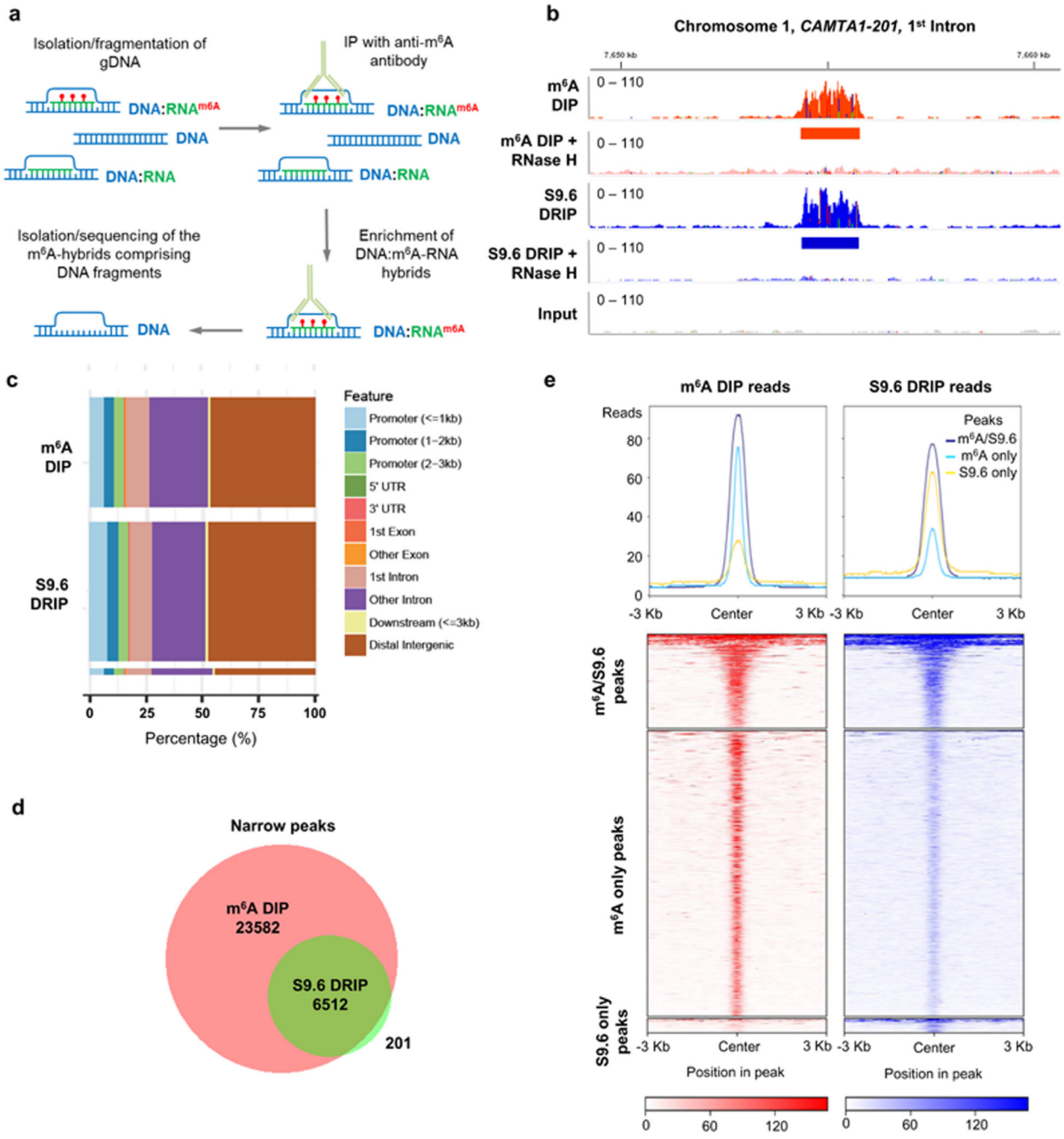



**Fig. 2.**

m$^6$A is present on the majority of the RNA:DNA hybrids in hPSCs. (a) Schematic illustrating the m$^6$A DIP technique. See also Extended Data Fig. 2. (b) The coverage plots of m$^6$A DIP and S9.6 DRIP densities (CPK) in the intron of *CAMTA1-201* gene. m$^6$A DIP and S9.6 DRIP peaks are marked with red and blue rectangles. (c) Distribution of m$^6$A and S9.6 peaks at the indicated genomic features in hiPSCs. (d) Venn diagram illustrating an overlap between m$^6$A DIP and S9.6 DRIP consensus peaks in REBL-PAT hiPSCs. (e) Heatmaps showing the distribution of density of m$^6$A DIP and S9.6 DRIP reads across genomic




Europe PMC Funders Author Manuscripts

regions containing the 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 $m^6A$ peaks (S9.6 only). The color of each line represents the density of reads for a given peak. The width of the heatmaps is normalized by peak length. Median numbers of reads per normalized region within each of the peak subsets are plotted over the top of the heatmaps. As the exact mode of genomic distribution of $m^6A$-containing RNA:DNA hybrids was initially unknown, we performed detection of both narrow and broad peaks in the datasets. The results shown were obtained from analyses of the narrow $m^6A$ and S9.6 peaks.



Europe PMC Funders Author Manuscripts

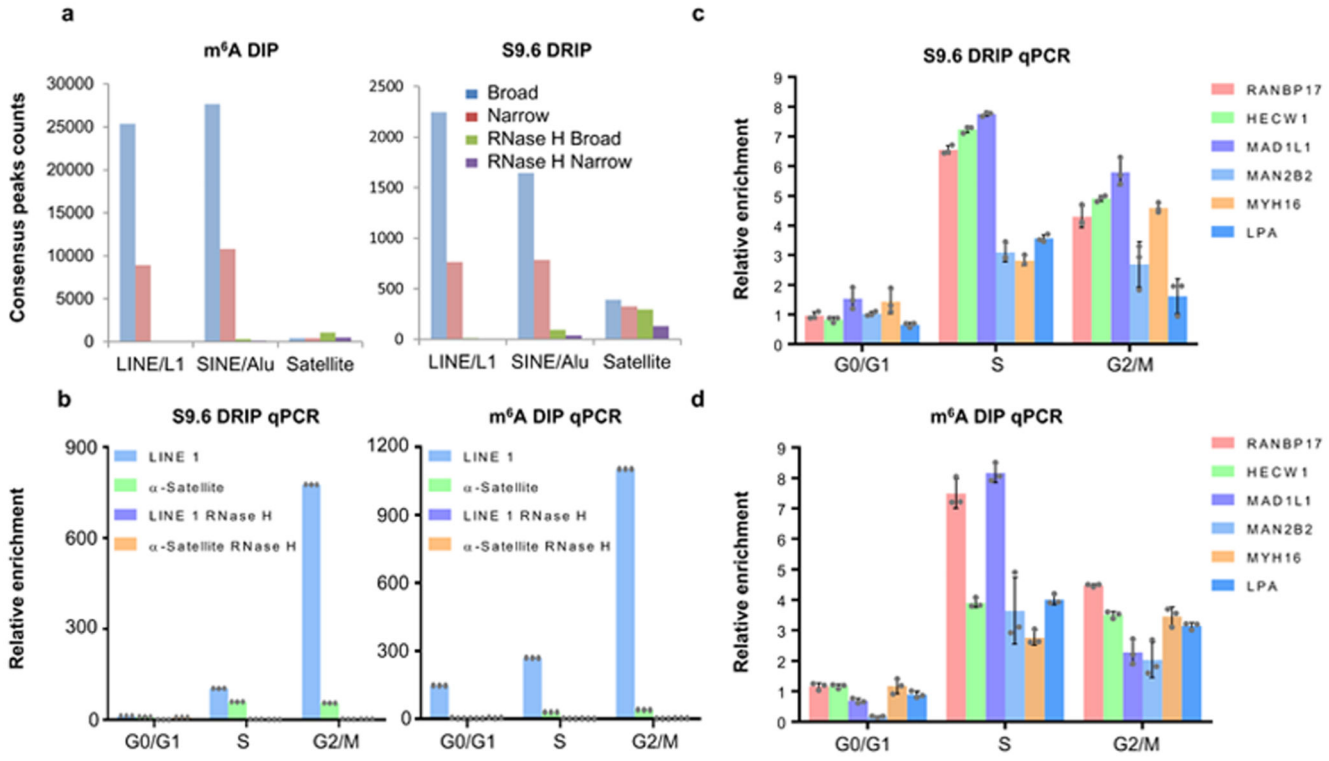



**Fig. 3.**

RNA:DNA hybrids exhibit cell cycle-specific dynamics in hPSCs. (a) The m$^6$A DIP and S9.6 DRIP consensus narrow peak counts of the indicated repetitive elements in hiPSCs. (b) The results of m$^6$A DIP and S9.6 DRIP qPCR of the indicated repeats performed on hiPSCs sorted at different cell cycle phases. Generic primers amplifying evolutionarily young L1Hs were used. (c, d) The results of S9.6 DRIP (c) and m$^6$A DIP qPCR (d) of the RNA:DNA peaks localized in the introns of the indicated genes (See Extended Data Fig. 9b) performed on hiPSCs sorted at different phases of the cell cycle. Data are means ± SD, n=3 independent experiments in (b-d).




Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

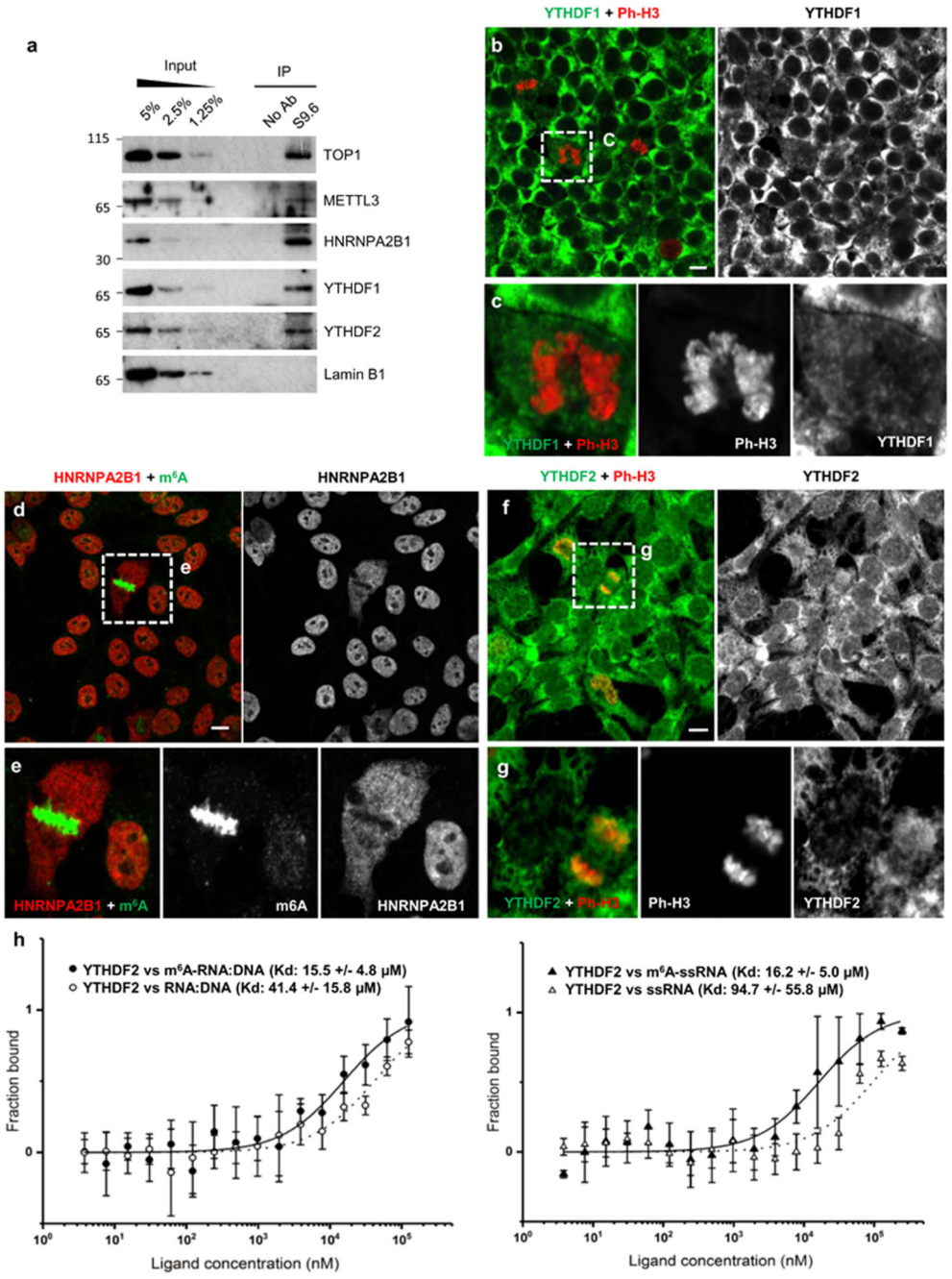



**Fig. 4.**

m$^6$A reader proteins interact with RNA:DNA hybrids. (a) Western blot of RNA:DNA hybrids protein co-IP probed with indicated antibodies. Top1 and Lamin B1 serve as positive and negative controls for R-loop IP, respectively. The experiments were repeated independently 2 times for METTL3 and 3 times for other proteins with similar results. The blots were cropped. The full scans of the blots are shown in Source Data 1. (b, c) Immunostaining of hiPSCs using anti-YTHDF1 and anti-phospho-Histone 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 phospho-Histone 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. The locations of the views shown in (c, e, g) are marked with dotted rectangles in (b, d, f). Scale bars are 10 μm. The experiments shown in (b-g) were repeated independently 4 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 The binding is shown as fraction 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 values are shown for each of the interactions. Error bars show SD, the centre values are means, n=6 independent series of experiments.



Europe PMC Funders Author Manuscripts

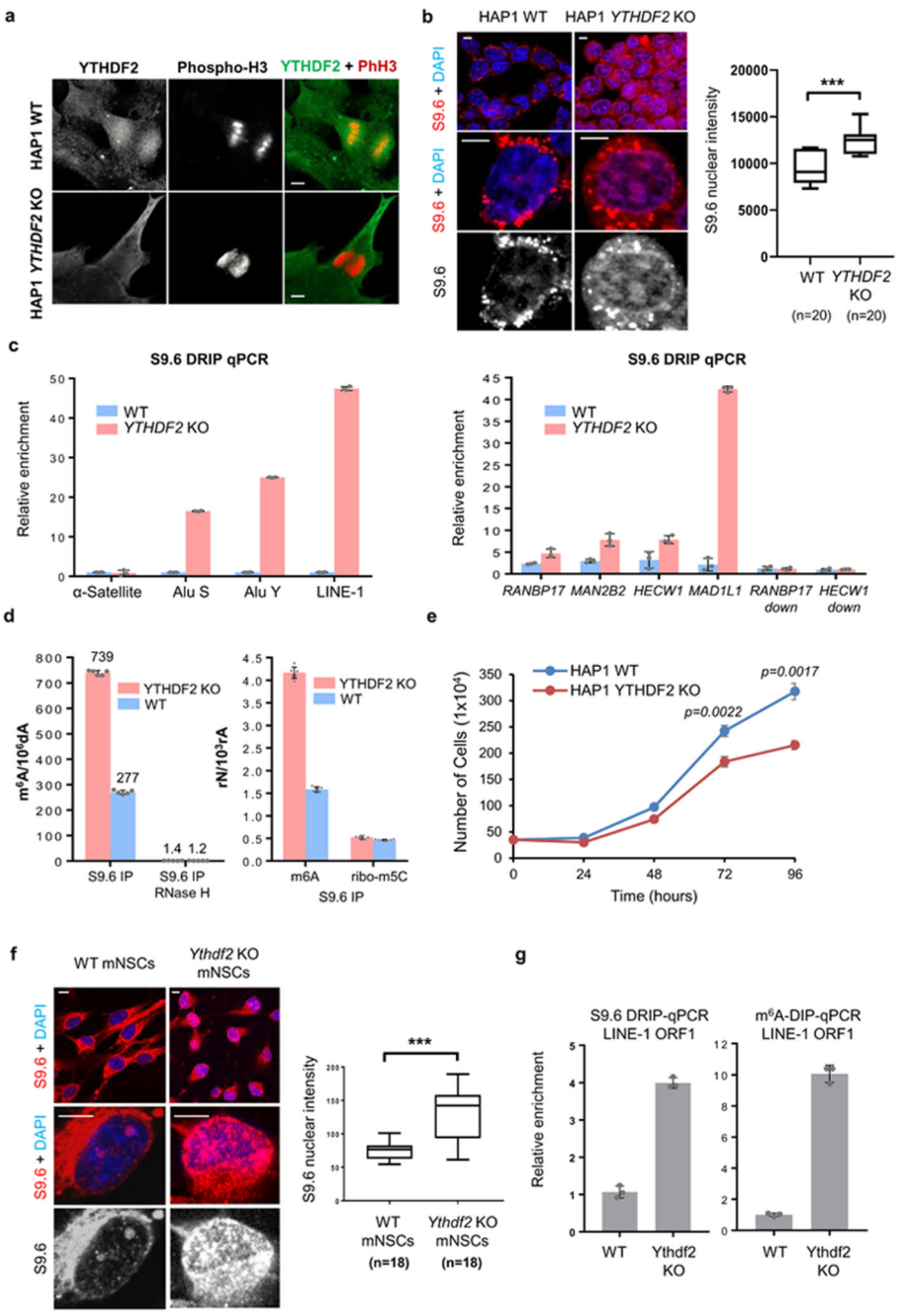



**Fig. 5.**

YTHDF2 depletion leads to accumulation of R-loops, increased accretion of m$^6$A on RNA:DNA hybrids and cell growth retardation. (a) Immunostaining of WT and *YTHDF2* KO HAP1 for YTHDF2 and phospho-Histone H3. The experiments were repeated independently 3 times with similar results. (b) Immunostaining of WT and *YTHDF2* KO HAP1 for R-loops alongside the 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




Europe PMC Funders Author Manuscripts

quantification of $m^6A$ and ribo-$m^5C$ in S9.6-IPs performed on WT and *YTHDF2* KO HAP1 and normalized for dA or rA. RNase H-pre-treated samples were used as controls. Data are means ± SD, n=5 (left) and n=7 (right panel) measurements of 4 independent samples. (e) The 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 $m^6A$ 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 are 10 μm. Data are means ± SD, n=3 independent experiments in (c, e, g). The elements of the box plots (b, f) are: centre line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all the data; n, sample size, ***$p < 0.0001$. Significance was determined by unpaired two-tailed Student's (b, f) or unpaired two-tailed Welch's (e) t-test.



Europe PMC Funders Author Manuscripts

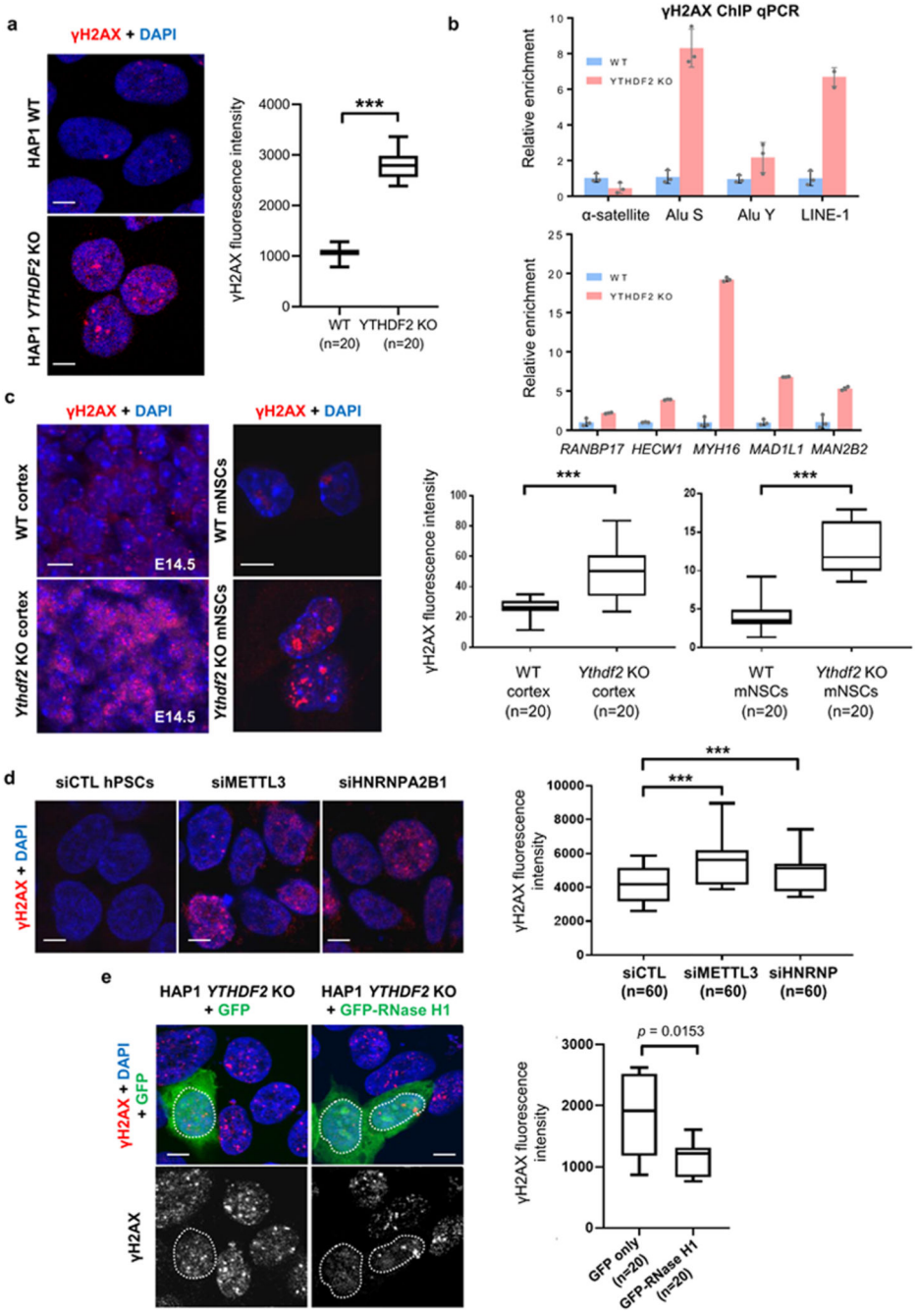



**Fig. 6.**

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. (b) The 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 ± SD, n=3 independent experiments. (c) Representative images of WT/ *Ythdf2* KO embryonic brain cortex and mNSCs

immunostained for γH2AX alongside quantification of γH2AX signal intensity in these tissues/cells. (d) Immunostaining of siCTL, siMETTL and siHNRNPA2B1 hPSCs for γH2AX and quantification of γH2AX signal intensity in these cells. (e) 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 the GFP-positive cells. P value is indicated. Examples of the nuclei used for signal quantification are marked with dotted shapes. Merged images and S9.6 channel views are shown in (a, c-e). Scale bars are 10 μm. The elements of the box plots shown in (a, c-e) are: centre line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all the data, n, sample size, ***$p < 0.0001$. Significance was determined by unpaired two-tailed Student's t-test, no adjustments were made for multiple comparisons.