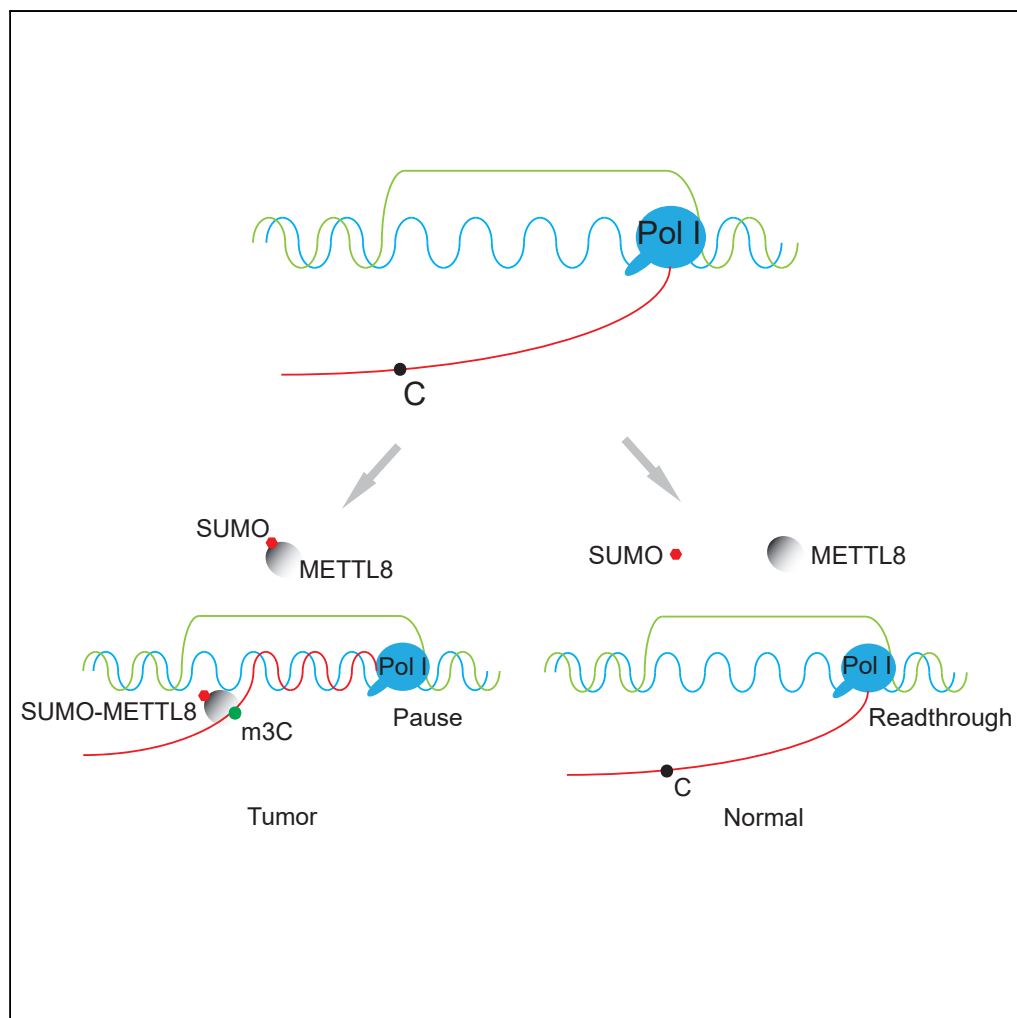**Article**

The SUMOylated METTL8 Induces R-loop and Tumorigenesis via m3C

Li-Hong Zhang,
Xue-Yun Zhang,
Tao Hu, ..., Ning
Sun, Yan Luo,
Xiang Gao

luoyan2011@zju.edu.cn (Y.L.)
shawngao@zju.edu.cn (X.G.)

HIGHLIGHTS

DNA:RNA hybrid
structures are regulated
by RNA methyltransferase
via 3-methylcytidine

SUMOylation stabilizes
the RNA
methyltransferase
complex in the nucleus

Dysregulation of
DNA:RNA hybrids may
induce tumorigenesis in
mammalian cells

Article

The SUMOylated METTL8 Induces
R-loop and Tumorigenesis via m3C

Li-Hong Zhang,[1,2,6] Xue-Yun Zhang,[1,2,6] Tao Hu,[4] Xin-Yun Chen,[3] Jing-Jia Li,[1,2] Manfred Raida,[5] Ning Sun,[3] Yan Luo,[1,2,*] and Xiang Gao[1,2,7,*]

SUMMARY

R-loops, three-stranded DNA-DNA:RNA hybrid structures, are best known for their deleterious effects on genome stability. The regulatory factors of this fundamental genetic structure remain unclear. Here, we reveal an epigenetic factor that controls R-loop stability. METTL8, a member of the methyltransferase-like protein family that methylates 3-methylcytidine (m3C), is a key factor in the R-loop regulating methyltransferase complex. Biochemical studies show that METTL8 forms a large SUMOylated nuclear RNA-binding protein complex (∼0.8 mega daltons) that contains well-reported R-loop related factors. Genetic ablation of METTL8 results in an overall reduction of R-loops in cells. Interaction assays indicated METTL8 binds to RNAs and is responsible for R-loop stability on selected gene regions. Our results demonstrate that the SUMOylated METTL8 promotes tumorigenesis by affecting genetic organization primarily in, or in close proximity to, the nucleolus and impacts the formation of regulatory R-loops through its methyltransferase activity on m3C.

INTRODUCTION

Epigenetic regulation is generally recognized to be related to covalent modifications of histones within chromatin or of genetic material such as DNA itself (Allfrey et al., 1964; Gold et al., 1963a, 1963b), which has powerful regulatory roles that impact the transcriptional landscape. Methyltransferase-like proteins (METTLs) share a well-conserved seven-beta strand s-adenosylmethionine (SAM)-binding site, which is by nature a common donor for the methyl group (Catoni, 1953; Schlenk and Depalma, 1957). In mammals, the function of several METTL proteins has been characterized, including the formation of N6-methyladenosine (m6A) in mRNA by a complex of METTL3 and METTL14 in human cells (Roundtree et al., 2017; Fu et al., 2014). METTL8, a member of a group of methyltransferase-like proteins, is originally identified as a tension induced protein (TIP) that is involved in mesenchymal to adipose tissue transition (Badri et al., 2008; Jakkaraju et al., 2005); it is also found to be mutated in colon cancers (Yeon et al., 2018) and regulates mouse embryo stem cell differentiation (Gu et al., 2018). Although biochemically METTL8 has been shown to modify m3C in RNAs (Xu et al., 2017), the molecular mechanism(s) regarding METTL8's role in epigenetic-related pathways remain poorly understood.

R-loops are naturally occurring genome-wide structures that are essential for diverse cellular functions (Ginno et al., 2012; Skourti-Stathaki et al., 2011; Yu et al., 2003). Deleterious effects of an excess number of R-loops have been linked to genome instability and DNA replication impairment (Aguilera and Garcia-Muse, 2012; Tuduri et al., 2009), and factors (SETX, TOPI, RNase H, Xrn2) that attenuate R-loop accumulation have been well reported (Aguilera and Garcia-Muse, 2012; Hatchi et al., 2015; Kogoma et al., 1993; Skourti-Stathaki et al., 2011; El Hage et al., 2010; Manzo et al., 2018). Owing to its DNA:RNA hybrid nature, R-loop formation usually occurs at the GC skew promoter regions of a genome (Chen et al., 2017; Ginno et al., 2012, 2013; Nadel et al., 2015; Sanz et al., 2016; Stork et al., 2016), and transcriptional supercoiling of the DNA also facilitates the formation of R-loops (Gottipati and Helleday, 2009; Brochu et al., 2018; Drolet, 2006; Drolet and Brochu, 2019). In addition to promoters, G-rich sequences have also been identified as transcription pausing elements downstream of poly(A) sites, where they promote poly(A)-dependent transcriptional termination (Gromak et al., 2006) that also involves R-loop formation. The intronic sequence also contributes to alternative splicing by promoting trans-splicing (Takahara et al., 2005). Additionally, computational prediction of the G-rich R-Loop Zooms in R-loop DB (Jenjaroenpun et al., 2017) revealed the genome-wide distribution of R-loops, which indicates the possible link between transcription and R-loops, suggesting that the latter may constitute a fundamental structure that impacts transcriptional elongation, termination, and overall outputs. To the best of our knowledge, although R-loops are emerging as "a double-edged sword" to genomic structure (Skourti-Stathaki and Proudfoot, 2014), the only factor that

[1]Department of Biochemistry and Cancer Institute of the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China

[2]Key Laboratory of Cancer Prevention and Intervention of China National Ministry of Education, Hangzhou 310009, China

[3]Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai 200032, China

[4]Department of Spine Surgery, Shanghai East Hospital, Tongji University, Shanghai 200092, China

[5]Life Sciences Institute, National University of Singapore, Singapore 117456, Singapore

[6]These authors contributed equally

[7]Lead Contact

*Correspondence:
luoyan2011@zju.edu.cn (Y.L.),
shawngao@zju.edu.cn (X.G.)

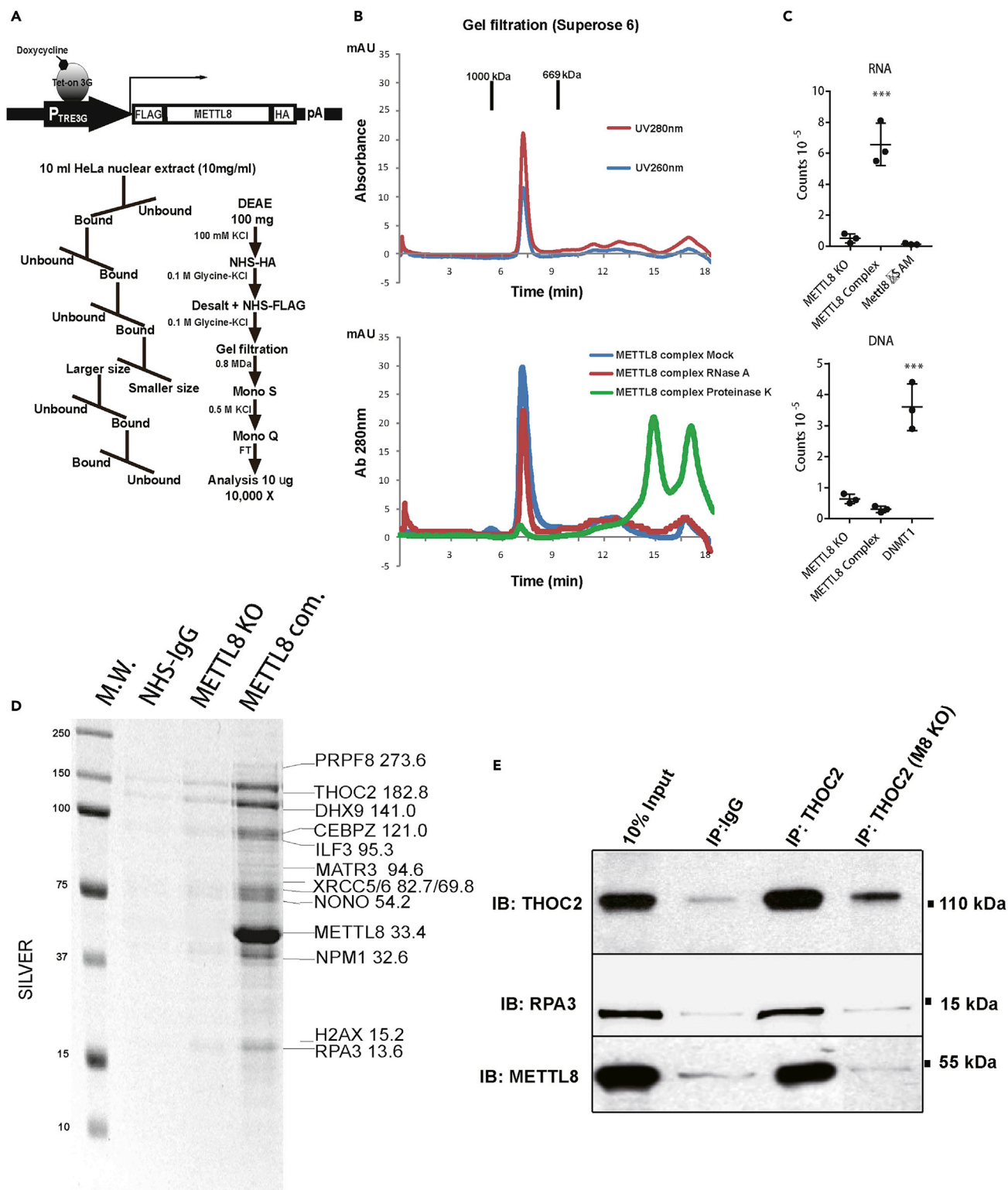https://doi.org/10.1016/j.isci.2020.100968

Figure 1. Biochemical Revelation of METTL8's Association with RNA Splicing and Export Factors

(A) Schematic representation of the doxycycline-inducible double-tagged (FLAG/HA) expression vector (top). Large-scale (>10 L) METTL8 complex purification steps (bottom).

Figure 1. *Continued*

(B) Top panel, gel filtration graph (Superose 6) showing the native size of the nuclear METTL8 complex using dual-channel UV (UV280nm, UV260nm). Bottom panel, gel filtration graph (Superose 6) detected by UV absorbance at 280nm of the METTL8 complex either treated with RNase A (Red) or Proteinase K (Green).

(C) Top panel, *in vitro* methylation of RNA by METTL8 complex purified either from wild-type (WT) or METTL8 KO cells; recombinant METTL8 mutant (METTL8ΔSAM, enzymatic dead) were included for negative control. Bottom panel, *in vitro* methylation of DNA by METTL8 complex purified either from WT or METTL8 KO cells; DNA methyltransferase DNMT1 was included for positive control.

(D) Silver staining of the purified nuclear METTL8 complex on 12% acrylamide gel; METTL8 KO cells and NHS column conjugated with IgG were included for control. Mass-spectrometry-determined peptide sequences and bands were labeled according to their molecular size.

(E) Endogenous coimmunoprecipitation using agarose beads conjugated with anti-THOC2 antibody; METTL8 and RPA3 were shown to interact with THOC2 *in vivo*.

positively regulates R-loops is METTL3 through its m6A activity (Yang et al., 2019) and the biological function of m3C remain unclear.

Here, we unravel the above questions through purification of the METTL8 complex. The overlapping nucleolus localization of METTL8 and R-loops (Hatchi et al., 2015; Salvi and Mekhail, 2015; Skourti-Stathaki et al., 2014), as well as the identification of the METTL8-THOC2-RPA3 association, indicates that regulatory R-loops are maintained in the nucleus, especially in the nucleolus, through METTL8 depending on its m3C activity. Intriguingly, SUMOylation guides METTL8 to the nucleolus, whereby it plays a stabilizing role for the complex via the SUMO-SIM interaction; both the m3C activity and SUMOylation are essential for METTL8 tumorigenicity in xenograft mice models. Taken together, our work establishes a novel R-loop pathway where SUMO-modulated nuclear methyltransferase (METTL8) orchestrates genetic structure R-loops through its enzymatic activity on m3C.

RESULTS

Biochemical Revelation of METTL8's Association with R-Loops and RNA Splicing and Export Factors

Increasing evidence indicated that nuclear factors play critical roles in epigenetic regulation; we decided to investigate the function of a previously enigmatic epigenetic factor family (METTL2/6/8, Figure S1A). Our approach for the initial study was to use a "guilty by association" strategy (Zheng et al., 2003) to unravel associated partners/potential functions of the larger family member, METTL8, through a series of large-scale conventional (>10 L of mammalian suspension cell line) biochemical purification steps. To ensure the purity of the final complex, a HeLaS3 cell line (ATCC, CCL-2.2) clone that stably expressed inducible FLAG/HA-double-tagged METTL8 (Figure 1A, top) was established. METTL8 expression was induced at the lowest level (Figure S1B, lane 1), and 100 mg of nuclear extract made from this clone was subjected to a scheme of chromatographic steps (Figure 1A, bottom). The complex native size was characterized with Superose 6 analytical gel filtration column to be ∼800 kDa (Figure 1B, top), and it was resistant to RNase A digestion (Figure 1B, bottom, red line), suggesting the METTL8 protein complex is held together through direct interaction. Through the purification process, the METTL8 complex was traced with commercially available monoclonal antibody and its ability to methylate RNAs (Figure 1C, upper) but not DNAs (Figure 1C, down). Preparative size exclusion chromatography (Sephacryl S-300) showed that the nuclear complex is slightly larger than its cytosolic counterpart (Figure S1C), which was suggested to be involved in protein translation regulation through affecting the formation of ribosome-mRNA polysome (Gu et al., 2018). The protein components were visualized on silver staining (Figure 1D, denoted as METTL8 Com.). Their sequence identities were then determined by FASP protein mass spectrometry (see Methods, Table S1). We repeated the same purification scheme (Figure 1A) using NHS columns (GE Healthcare) conjugated with either IgG as a blank control (Figure 1D, NHS-IgG) or METTL8 knockout cells (Figure 1D, METTL8 KO); mass spectrometry analysis of the IgG control was also performed (Table S1) to evaluate background noise.

After repeated mass spectrometry analysis (n = 3), METTL8 was confirmed to be in complex with several well-known protein complexes such as the nucleolar SWAP (NPM1-NCL-PARP1) complex (Borggrefe et al., 1998), the DNAPK (KU70/80, PRKDC, DHX9) complex (Gottlieb and Jackson, 1993), the mammalian homolog of the yeast Tho/Trex complex component THOC2 (a well-established R-loop regulator [Bhatia et al., 2014; Gomez-Gonzalez and Aguilera, 2007; Gomez-Gonzalez and Aguilera, 2009; Gomez-Gonzalez et al., 2011; Gonzalez-Aguilera et al., 2008; Hamperl and Cimprich, 2014; Salas-Armenteros et al., 2017]), and parts of the Spliceosome/U1, U5. 4/6 snRNP complex (summary in Table S1). Key interactions from
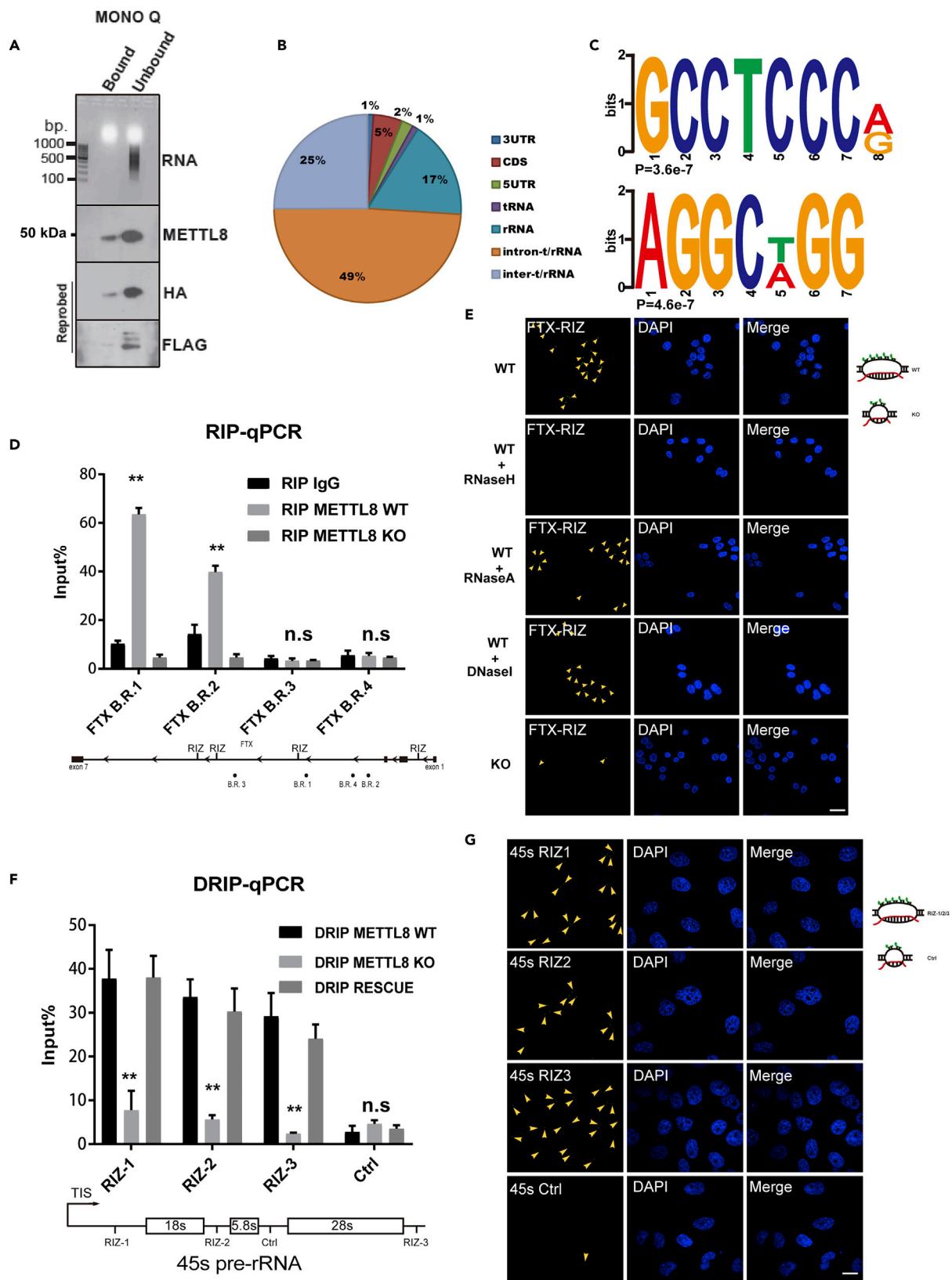
Figure 2. METTL8 Binds to RNA and Regulates R-Loop Formation on the Ribosomal RNA Gene

(A) Purified nuclear METTL8 complex from the final step (Mono Q column) (Figure 1A) was collected either for immunoblotting (bottom panels, the original probe is METTL8 [rabbit], then stripped (Figure S2A) and re-probed with HA [rabbit] and FLAG [mouse]) or for RNA extraction; RNA was visualized on an agarose gel (top panel).

(B) PAR-CLIP classifications of RNA species binding to the METTL8 complex (complete list in Table S2, control in Table S3).

(C) Motif analysis of METTL8-binding RNA that is rich in Gs and Cs (top two motifs, analysis was performed using the DREME online server).

(D) Validation of METTL8 interaction with RNA via immunoprecipitation followed by quantitative RT-PCR indicated as RIP-qPCR, the primers were designed around the locus of the METTL8-binding regions (B.R.) detected in PAR-CLIP (schematic at bottom), and amplification folds of the FTX (B)R. (1–4) were normalized to input RNAs; data are means $\pm$ SD (n = 3, **p < 0.01, Student's t test); METTL8 KO cell lysates were included for negative control.

(E) FISH confocal imaging indicating that RIZ-FTX (arrowheads) loses (upper right, schematic) its nuclear localization after METTL8 knockout (KO); scale bar, 10 µm; treatment with RNase H, RNase A, and DNase I was included as controls.

(F) DNA:RNA hybrid immunoprecipitation (DRIP) analysis of HeLa METTL8 WT and KO cells; the number of R-loop-binding sites (on 45s pre-rRNA) was calibrated against the input nucleic acids; data are means $\pm$ SD (n = 3); **p < 0.01, Student's t test. Bottom boxes, a schematic of 45s pre-rRNA with the location of RIZ-1/2/3 indicated; a gene site without RIZ was included as control (Ctrl).

(G) Fluorescence *In Situ* Hybridization (FISH) confocal microscope imaging using the RIZ probes (green) indicated in (F, RIZ-1/2/3). Arrowheads indicate R-loops detected by the probes and a schematic figure was shown on the upper right.

the Tho/Trex complex were further validated using a THOC2 antibody; endogenous THOC2 was shown to be associated with METTL8 and RPA3 but not with IgG control, and METTL8 knockout (KO) cell lysate was also included for specificity control (Figure 1E). Moreover, evidence from GST pull-down assay indicated purified METTL8 interacted directly with THOC2 and RPA3 (Figure S1E). Thus far, our purification efforts revealed the identity of the METTL8 nuclear complex and elucidated potential functions in which METTL8 may take part (RNA splicing and maturation). Owing to the fact that METTL8's binding partners, TREX (THOC1/2) and RPA family, have well-established roles in R-loop regulation, we are particularly interested in whether METTL8 is a novel factor for this pathway.

METTL8 Binds to RNA and Regulates R-Loop Formation on the Ribosomal RNA Gene

During biochemical analysis of the complex, we observed OD260/OD280 > 2 at the peak center of METTL8 complex (Figure 1B), which suggested that the complex is not entirely composed of protein and that the other components are likely to be RNAs (Glasel, 1995). Our first approach in confirming the binding of RNAs is Trizol (Invitrogen) phase separation of the purified complex. We followed the commercially available extraction protocol, and the RNAs were extracted and visualized on a gel in conjunction with the Trizol-denatured HA/FLAG-tagged METTL8 protein (Figure 2A); the HA/FLAG blot was stripped/reprobed from the original METTL8 blot (Figure S2A). Next, the HA/FLAG-METTL8 over-expressing cells were subjected to PAR-CLIP (see Methods) sequencing, and a total of 1,252 unique RNAs (with an enrichment factor >2) were identified (Table S2). The bound RNAs were mostly t/r-RNA introns (49%) and inter t/r-RNA (25%) in the nucleolus (Figure 2B). Moreover, the METTL8-binding RNA motif was determined to be GC rich using DREME (Bailey, 2011) (Figure 2C, top two hits with p = 3.6×10^{-7} and p = 4.6×10^{-7}). We also sequenced background-binding RNAs to the FLAG antibody in METTL8 KO cells; our data have indicated that they are mostly ribosomal and skeleton (microtubule) gene RNAs (Table S3). In addition to our earlier evidence regarding its binding with the R-loop regulator, TREX (THOC1/2) (Figures 1D, 1E, and S1E), we reason that METTL8 complex might be involved in R-loop regulation, as R-loops are reported to be GC skewed (Chen et al., 2017; Jenjaroenpun et al., 2017).

We validated one of the top PAR-CLIP hits using RNA immunoprecipitation (RIP-qPCR); the result indicated that five prime to Xist (FTX), which was reported to affect X-inactivation, binds exclusively to METTL8 complex in binding region 1 (B.R. 1) and B.R. 2 but weakly to B.R. 3/4;, METTL8 KO cells were included for negative control (Figure 2D, schematic of the B.R. genome location at the bottom). METTL8 was knocked out from HeLa cells using CRISPR-Cas9 targeting exon 3 (Figure S1D), and off-targets were checked *in silico*. FISH (fluorescence *in situ* hybridization) imaging (Napoli, 2013) using FTX probes for the R-loop-prone region (online-computed GC-rich probes [Jenjaroenpun et al., 2015], denoted as R-loop Initiation Zooms, RIZs) showed more displaced GC-rich single-stranded DNA in the nucleolus of WT HeLa cells than in KO METTL8 (Figure 2E top and bottom panels, Schematic Figure on right); treatment with an RNA:DNA hybrid-specific enzyme (RNase H) abolishes the FISH signal, whereas RNase A and DNase I have insignificant effect (Figure 2E middle panels, quantification in Figure S2E, n = 4, ***p < 0.005).

Regulation of R-loops by METTL8 was evaluated on the 45s pre-rRNA gene using DNA:RNA hybrid immunoprecipitation–quantitative PCR (see Methods, DRIP-qPCR). DRIP-qPCR indicated that the R-loops

from RIZ-1, RIZ-2, and RIZ-3 sites were significantly (n = 5, **p < 0.01) reduced in the METTL8 KO samples (Figure 2F, schematic on bottom). The control (Ctrl) site without RIZ sequence does not bind to the S9.6 antibody and remains unchanged after METTL8 KO, suggesting METTL8 directly affects R-loop formation on selected gene loci that are mostly (49%) t/r-RNAs (Figure 2F). FISH imaging (Napoli, 2013) indicated more R-loops at 45s RIZ-1, RIZ-2, and RIZ-3 than in the control region (Figure 2G), and quantification of R-loop positive cells was deemed significant (Figure S2F, n = 4, **p < 0.001, ***p < 0.005, ****p < 0.0001), suggesting that these R-loops are formed selectively. With evidence from both its R-loop-related co-factors (THOC2 and RPA3), together with indications from its binding to GC-skew RNAs that is sensitive to RNase H and the DRIP-qPCR data, and repeated validations on other models genes such as the *ACTB* gene (Figures S2B–S2D), we reasoned that METTL8's function is centered in R-loop regions. Therefore, we developed assays to investigate this hypothesis.

METTL8 Induces R-Loops in Cell Nucleolus

To quantify the number of R-loops *in vitro*, we customized a dot blot assay by serial dilution of equal amounts of starting nucleic acids on a nylon membrane (positively charged, Roche). The loading quantity was checked with EtBr (ethidium bromide). Initially, the detected number of R-loops from the total nucleic acid extract was similar between METTL8 WT and KO cell lines (Figure S3A); however, after repeated (n > 3) investigation supported by a recent publication (Zhang et al., 2015), we found that there are cross-reactions of the anti-DNA-RNA hybrid (S9.6) antibody with AU-rich RNA:RNA duplexes that are abundant in the cytosol. Therefore, the pure cell nucleus and/or nucleolus was isolated (Figure S3B), and a biochemical marker of each cellular compartment was checked with antibodies against UBF1 (a nuclear marker) NCL (nucleolus marker) or alpha-tubulin (a cytoplasm marker) (Figure S3C). The specificity of the S9.6 antibody was monitored by comparing the purified nuclei (fixed on a slide) R-loop signals before and after RNase H and RNase A treatment, which digest R-loops and single-stranded RNAs, respectively (Figure 3A). The results in Figure 3A clearly indicated that the S9.6 antibody specifically labels the nuclear R-loops and no cytoplasmic signals were detected after nuclear isolation.

After having developed a reliable R-loop quantification method using pure nucleoli, the difference in the number of R-loops between WT and KO cell nucleoli were quantified using dot blots (Figure 3B, lane 1–6). The dot intensity analysis (quantified by ImageJ) indicates that the number of R-loops from METTL8 KO cells was significantly lower (Figure 3B, rows 2 and 4) from repeated assays (Figure 3C, n = 3, *p < 0.05). The above finding was also confirmed multiple times by confocal immunofluorescence imaging (Figures 3D–3F and S3D). We showed that the nucleolus R-loop signal, in many cases, completely diminished after Mettl8 ablation (Figure 3D, enlarged inset), whereas the cytoplasmic signal remained detectable (cytoplasmic AU-rich RNAs). The line intensity graph (quantified by ImageJ line intensity plugin) indicated a nuclear localization of R-loops in WT but not KO cells (Figure 3E, n = 3). Up to 10 unbiased image fields (Figure S3D) were taken, and the R-loop difference between WT and KO HeLa S3 cells was determined to be significant (Figure 3F, n = 204, **p < 0.01).

Nuclear Localization of METTL8 Is Maintained by SUMOylation at Lysine 80

One interesting observation arising from complex purification is that extra bands appeared in the nuclear fraction of METTL8 but not from the cytosolic fraction (Figure S1C). This prompted us to test whether the METTL8 protein is post-translationally modified. We confirmed the molecular shift to be approximately 10 kDa by incubating cell lysates with three antibodies (anti-FLAG, anti-HA, and anti-METTL8) targeting the same double-tagged METTL8 protein (Figure 4A). Analysis of the amino acid motif suggested this modification is SUMO on a conserved lysine across species (Figure 4B, lysine 80 in humans). Next, *in vitro* SUMO assays using purified recombinant proteins (rMETTL8, rMETTL8-K80R, SUMO-1, and SUMO-1gg [lacking C terminus]) indicated that our prediction is correct. Point mutation of the SUMO receptor lysine (K) to arginine (R) completely prevented METTL8 SUMOylation (arrowhead) *in vitro* (Figure 4C, lanes 3 and 4). Additionally, an increased amount of Ginkgolic acid, a known SUMO ligase inhibitor, completely abolished the METTL8 SUMOylation (arrowhead) *in vitro* (Figure 4D, lanes 3 and 4). We also showed that addition of purified METTL8 complex promotes the *in vitro* SUMOylation reaction, suggesting SUMOylatin of METTL8 is independent from the complex but is enhanced when the complex is formed (Figure 4E). Computational analysis indicates that METTL8 is the only member of the METTL2/6/8 family that contains both a SUMO site and a SUMO interaction motif (SIM) (Table S4). YFP fusion protein (YFP-METTL8, YFP-METTL8-K80R) imaging suggests that lysine 80 is responsible for METTL8's nuclear enrichment in METTL8 KO HCT116 cells (Figure 4F, arrowheads; Figure S3E, showing the KO effect),
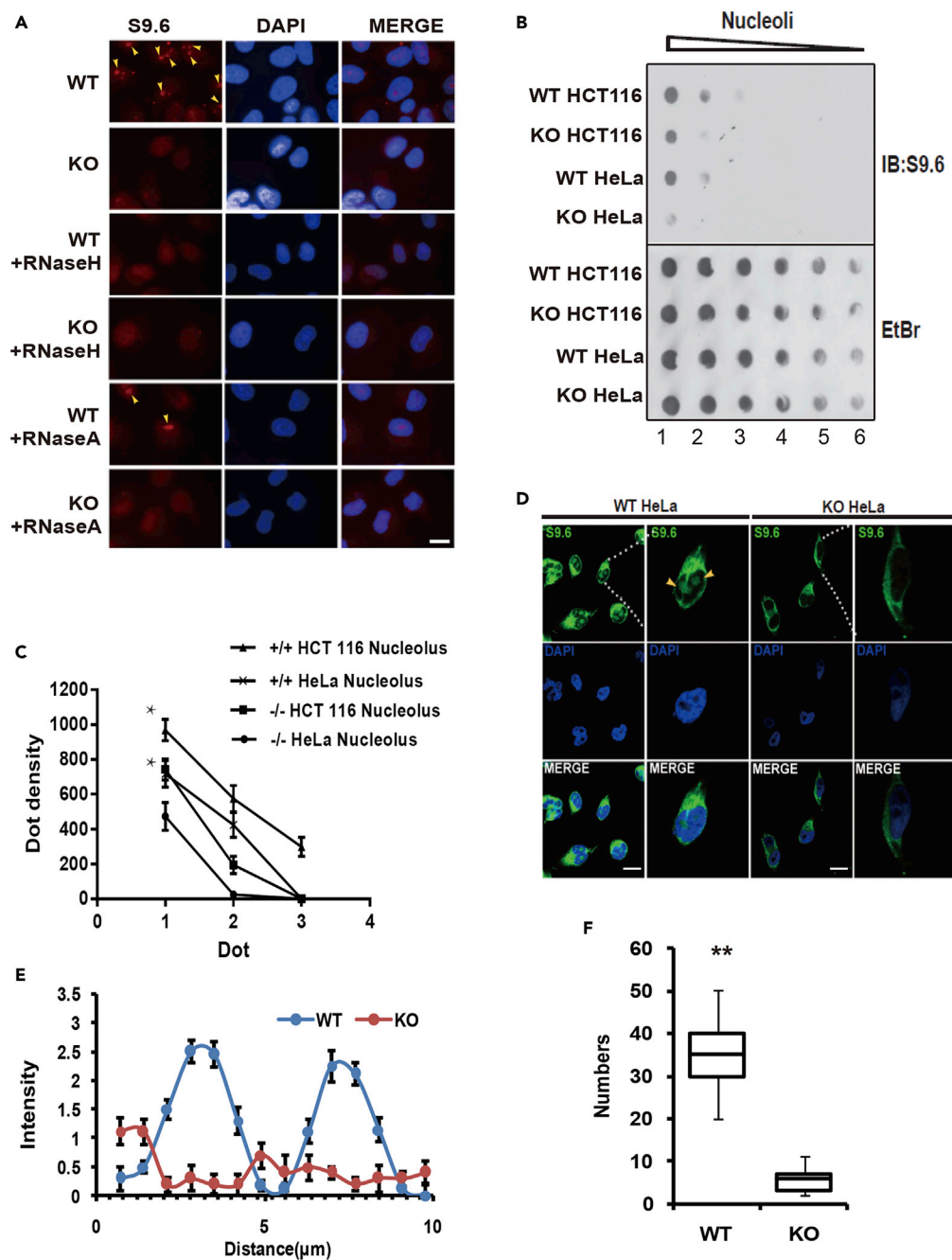
Figure 3. METTL8 Induces R-Loops in Cell Nucleolus

(A) R-loops (red) immunofluorescence imaging of WT and KO HeLa nuclei either mock treated or treated with RNaseH (1 μg/mL) and RNase A (1 μg/mL). Scale bar, 10 μm.

(B) Dot blot displays of R-loop structures from the nucleolar extract. Loading was visualized by ethidium bromide (EtBr) staining; the number of R-loops was visualized with the S9.6 antibody (IB: S9.6).

(C) Dot blot densitometry analysis of HeLa and HCT116 nucleolar R-loops from repeated assays (n = 3, *p < 0.05, Student's t test).

(D–F) The representative confocal image (D) of R-loops with an S9.6 antibody; R-loops (green) were stained with an S9.6 antibody, and DAPI (blue) was included to counter-stain the HeLa nucleus. Scale bar, 10 μm. Line scan analysis (E) of WT (blue) and KO (red) HeLa cells (means ± SD, n = 3). Statistical analysis (F) of R-loop-positive cells; data are means ± SD (n = 204). **p < 0.01, Student's t test.
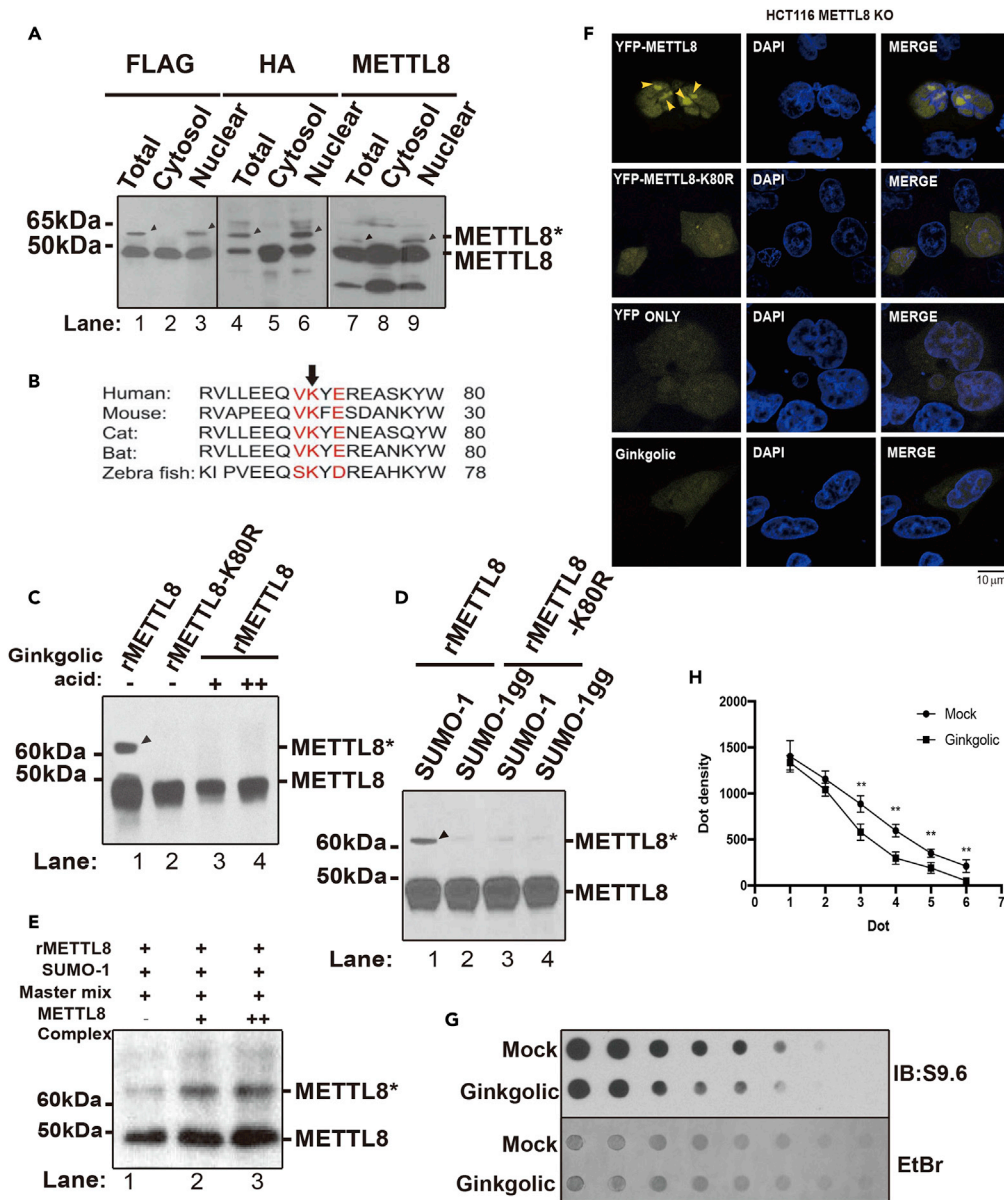
Figure 4. Nuclear Localization of METTL8 Is Maintained by SUMOylation at Lysine 80

(A) Western blot analysis of METTL8 protein band patterns (modifications) in fractioned or whole HeLa cell lysates with three different antibodies; bands were visualized separately in parallel using FLAG, HA, and METTL8 antibodies (left to right); the molecular weight is indicated on the left.

(B) Motif analysis of METTL8 protein at its lysine (K) SUMOylation site across species. (C–E) *In vitro* SUMOylation of recombinant METTL8 (rMETTL8) protein.

(C–E) (C) Together with the SUMOylation Master Mix, the bacterial purified rMETTL8 or SUMOylation (rMETTL8-K80R) mutant protein was either incubated with two different analogs of the SUMO protein (SUMO-1, SUMO-1gg [inactive]),

(D) an increasing dose of SUMO E3-ligase inhibitor (Ginkgolic acid), (E) or with an increasing quantity of purified METTL8 nuclear complex (from Figure 1D). The molecular weight of SUMOylated METTL8 protein is denoted on the left, and the modified band is denoted as METTL8* on the right.

(F) Confocal imaging of either YFP-METTL8 or YFP-METTL8-K80R plasmids transfected into METTL8 KO HCT116 cells (yellow arrowheads), YFP and YFP-METTL8 transfected cells treated with Ginkgolic acid were included as controls, nuclei were counter-stained with DAPI (blue); scale bar, 10 μm.

(G) Dot blot of R-loop structures from nucleolar extracts with or without Ginkgolic treatment (10 μg/mL); loading and R-loop structure was visualized with EtBr and the S9.6 antibody (IB: S9.6), respectively.

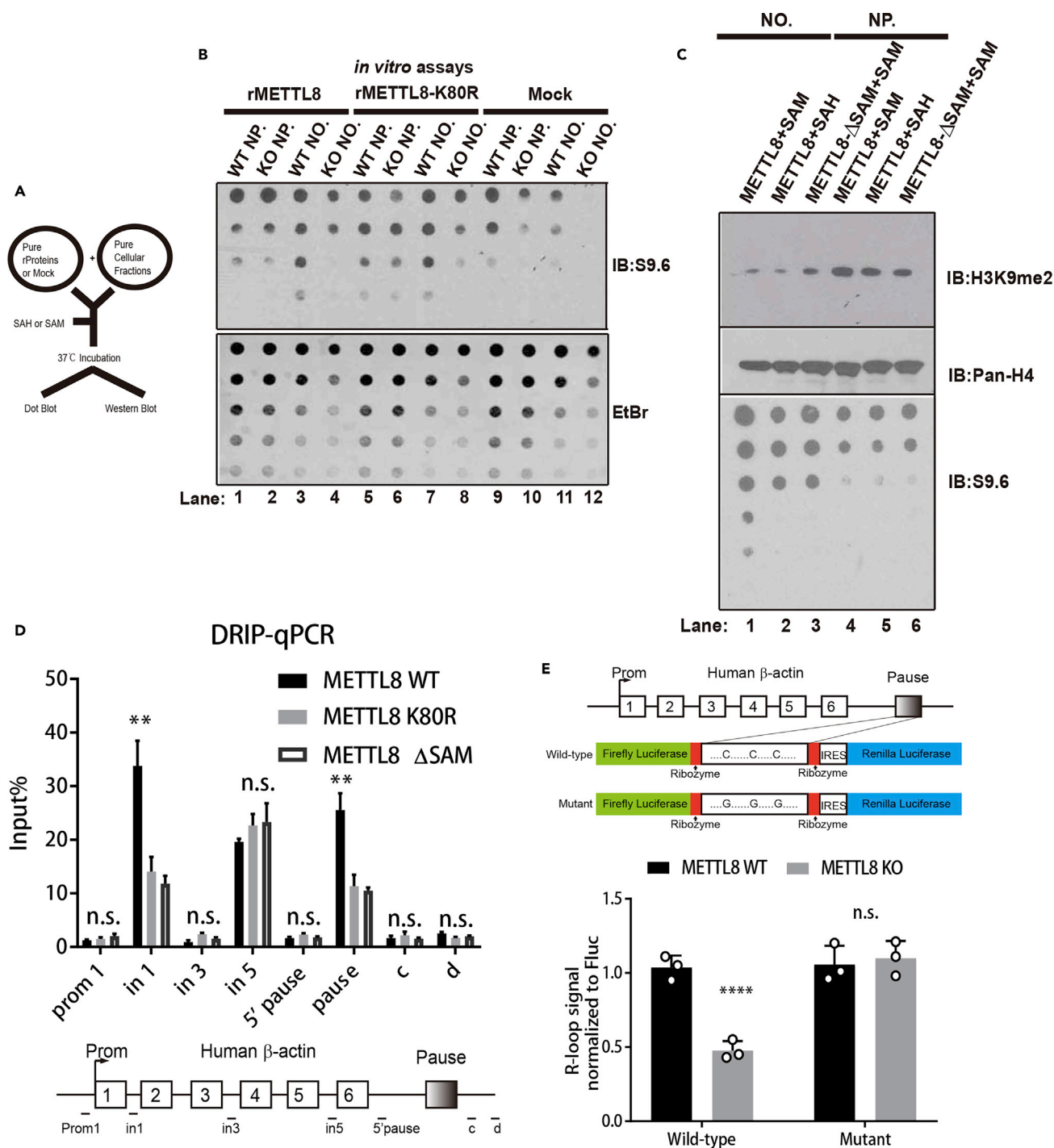(H) Dot blot densitometry analysis of (G), n = 3, **$p < 0.01$, Student's t test.

Figure 5. SAM Domain Is Essential for METTL8 R-Loop-Inducing Ability via m3C

(A) Schematic of the *in vitro* methylation assay; components were mixed as indicated and incubated at 37°C followed by either a dot blot or western blot.

(B) *In vitro* methylation of genomic nucleic acid with purified recombinant proteins, compared with a control (lane 12), rMETTL8-K80R retained its R-loop-inducing ability in KO nucleolar (NO.) (lane 8).

(C) *In vitro* methylation of nucleolar (NO.) and nucleoplasm (NP.) nucleic acids with recombinant proteins; wild-type METTL8 is able to increase R-loops in NO. (lane 1) but not in NP. (lane 4); the methyl-donor mutant (METTL8-ΔSAM) also fails to increase R-loops in NO.(lane 3).

Figure 5. *Continued*

(D) DNA:RNA hybrid immunoprecipition (DRIP) analysis of METTL8 WT, METTL8 K80R, and METTL8 ΔSAM inducing R-loops in HeLa cells; the number of R-loop-binding sites (on beta-actin) was calibrated against the input nucleic acids; data are means ± SD (n = 3); **p < 0.01, Student's t test. Bottom boxes, a schematic of human beta-actin gene dash lines indicating Prom1, in1, in3, in5, 5'pause, and the pause site.

(E) Top, schematic of the luciferase reporter system for R-loop-induced transcription pause. The *ACTB* gene pause region, containing m3C sites (Cs) or mutated sites (Gs), was inserted into a dual luciferase reporter vector. Bottom, quantification of R-loops and readthrough products of the wild-type and mutant reporters in METTL8 WT and METTL8 KO cells. Data are means ± SD (n = 3, ****p < 0.0001, Student's t test).

and Ginkgolic acid inhibits not only the nuclear enrichment of METTL8 (Figure 4F, bottom panels) but also its downstream genomic structure: R-loops (Figures 4G and 4H). We conclude from the above evidences that METTL8 is SUMOylated at a conserved site (lysine 80 in human), which is essential for its function on R-loops in the nuclear compartment. Next, we proceed to explore the biological connections between these motifs (SAM and SUMO) with R-loops.

SAM Domain Is Essential for METTL8 R-Loop-Inducing Ability via m3C

In vitro assays (Figure 5A, schematic) indicated that, although the K80R METTL8 mutant loses its nuclear localization, it retained its ability to induce R-loops *in vitro*, suggesting that SUMO only affects its cellular localization but not its enzyme activity toward R-loops (Figure 5B, lane 8 versus lane 12). A possible explanation for this observation is that SUMOylation frequently targets entire groups of physically interacting proteins and serves as a distinguishing marker for functionally engaged protein complexes (Jentsch and Psakhye, 2013). Additionally, METTL8 is the only member of its family that contains both a SUMO and a SIM motif (Table S4); therefore, we conclude that the METTL8 complex is an active, functional nuclear complex cross-linked by SUMO-SIM interaction. The enzymatically dead counterpart (rMETTL8-ΔSAM, lacking the canonical S-Adenosyl Methionine donor domain) fails to induce R-loops (Figure 5C, lane 3 versus lane 1, Pan-H4 for loading control), which indicates the SAM domain (methyl donor) is responsible for its R-loop stabilizing activity. We also tested the R-loop inducing ability of these METTL8 mutants; DRIP-qPCR has indicated that both the K80R and ΔSAM mutant failed in inducing R-loops on "in 1" and "pause" sites, whereas WT METTL8 could (Figure 5D), suggesting that, although K80R could induce R-loops *in vitro* (Figure 5B), the dislocation of the K80R mutant abolishes its ability to induce R-loops in cells (also supported by its inhibitor in Figure 4G). To investigate the effect of m3C on R-loops stability, we constructed a minigene reporter system in which two luciferases (Firefly and Renilla, Figure 5E, top) are separated by the *ACTB* pause element that contains C-rich RIZ. R-loop formation would induce transcription pause that only produces the Firefly transcripts, whereas reduction of R-loop would generate readthrough, which in turn produces both the Firefly and the Renilla transcripts. Our data showed that METTL8 KO causes a reduction in R-loops only in the wild-type luciferase reporter but not in the Mutant reporter where Cs has been mutated to Gs (Figure 5E, bottom), suggesting METTL8 regulates R-loop-induced transcription pause via its ability to modify m3C. Having validated various mechanistic aspects of METTL8 on R-loop biogenesis from Figure 1–5, taking into consideration that earlier report indicated R-loops affect genome stability and are potentially tumorigenic (Hamperl et al., 2017). We proceeded to test the tumorigenicity of METTL8 using xenograft mice models.

The SAM and SUMO Domains Are Tumorigenic in Mice Model

In Figure 3 we have showed METTL8 promotes R-loops formation in cell lines, and since R-loops are reported to be tumorigenic, we proceeded to test if METTL8 could promote tumorigenesis via regulating R-loops. First, we showed in soft agar colony assay, METTL8 promotes cell proliferation and both its m3C donor domain and SUMOylation domain are essential for this function (Figure 6A). Next we showed in immunodeficient mice xenograft models that METTL8 KO HCT116 cells (Figure S3E) and their mutants, METTL8ΔSAM and METTL8K80R, are less tumorigenic than their wild-type counterpart (Figure 6B) and their difference in tumor volume is deemed to be significant (Figure 6C, n = 4, **p < 0.01, ***p < 0.005). Xenograft data suggested that both the SAM and SUMO domains, which were shown to affect METTL8 R-loop-inducing ability (Figure 5), are essential for the tumorigenicity of METTL8. Clinically, based on published datasets in human patients with colorectal cancer (Smith et al., 2010), high level of METTL8 significantly has a lower survival rate than the low-level cohort (Figure 6D, n = 504, p = 0.0024), indicating METTL8 may contribute to therapeutic response in patients with cancer.
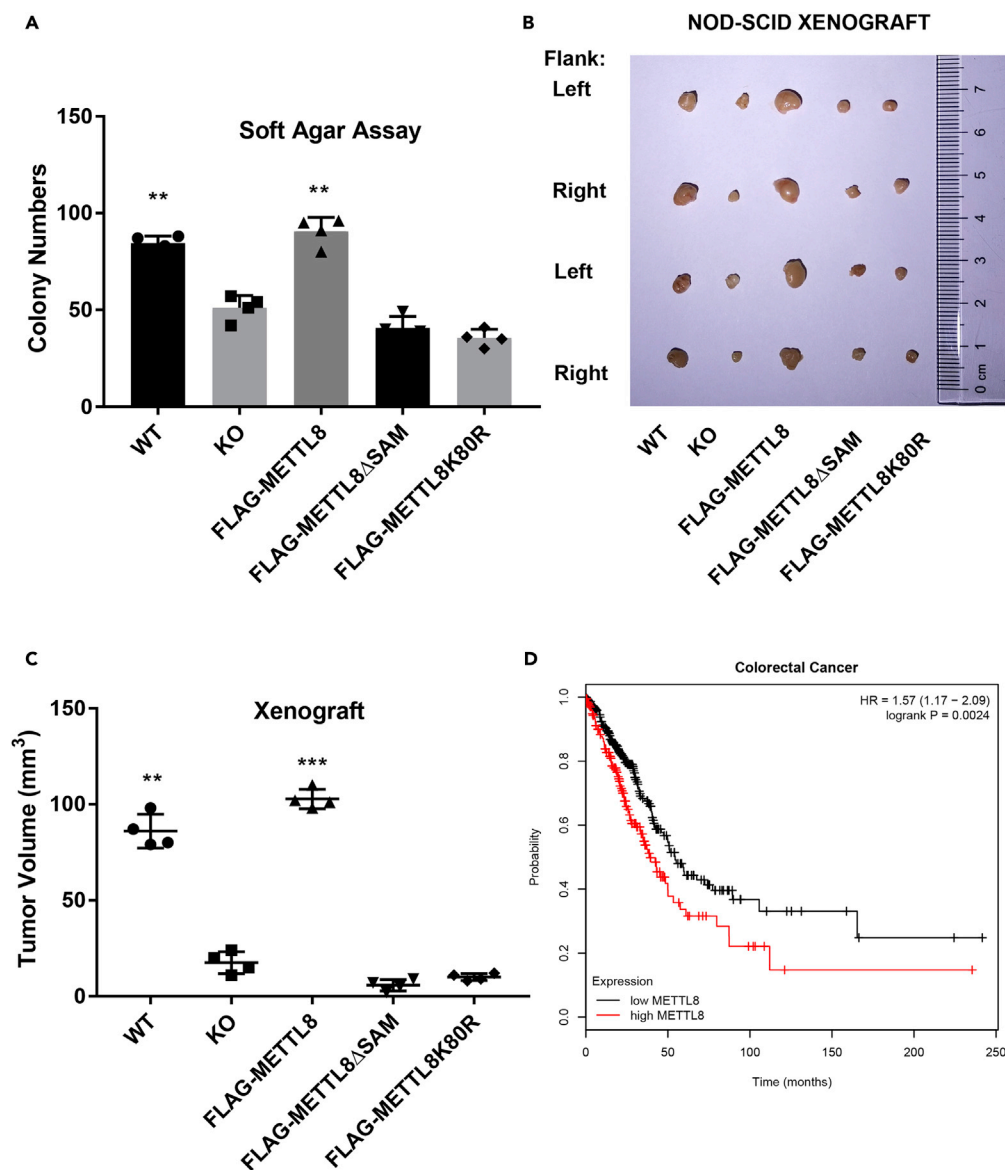
Figure 6. The SAM and SUMO Domains Are Tumorigenic in Mice Model

(A) HCT116 WT or KO cells overexpressing METTL8, ΔSAM, and K80R were seeded at 2,000 cells/well in quadruples, and after 2 weeks, colonies grown up in soft agar were apparent. Colonies with more than 50 cells were counted, data are mean ± SD (n = 4, **p < 0.01, Student's t test).

(B and C) 1×10^5 HCT116 WT or KO cells overexpressing METTL8, ΔSAM, and K80R were mixed with matrigel and injected subcutaneously into the left or right flank on the back of 6-week-old female NOD-SCID mice; after 4 weeks, tumor tissue was dissected, photographed (B), and weighted (C, n = 4, **p < 0.01, ***p < 0.005, Student's t test).

(D) Patients with colorectal cancer as enlisted in the study (GSE17536&37) were stratified according to METTL8 expression level, and survival rate was monitored in a cohort of patients (n = 504, p = 0.0024, log rank test) with the indicated Mettl8 gene level.

Taken together, our data showed METTL8 regulates R-loops through its methyltransferase activity, which was identified to be m3C (Xu et al., 2017), and its SUMOylation on Lysine 80 contributes to its nuclear enrichment, which is essential for its R-loop inducing activity. Xenograft mice models also suggested that both the SAM and the SUMO domains are responsible for METTL8 tumorigenicity. The METTL family is emerging to be an interesting class of proteins that regulate various aspects of RNA metabolism such as circRNA (Zhou et al., 2017) and R-loop (Yang et al., 2019), which are both related to tumorigenesis and genome stability.

DISCUSSION

METTL8 Complex and R-Loops

Epigenetic regulation is generally recognized to be related to covalent modifications of histones within chromatin or of genetic material such as DNA itself (Allfrey et al., 1964; Gold et al., 1963a, 1963b), which essentially impacts the transcriptional landscape. Recent research has found the prevalent, non-canonical genomic structure of R-loops, which are associated with epigenetic signatures in mammals (Sanz et al., 2016); however, the mechanisms by which R-loops regulate genetic readout (transcription or RNA processing) and detail information regarding the role of R-loops in RNA splicing remain elusive. Here, through biochemical approaches, we first show that the METTL8 complex is an RNA processing complex containing well-characterized proteins involved in RNA splicing, mRNA export, and spliceosome assembly (Figure 1). We show that its key component, METTL8, controls R-loop production in the cell nucleus and, specifically, in the nucleolus (Figure 3), thus physically linking these biological processes together (Aguilera and Gomez-Gonzalez, 2008; Li and Manley, 2006). Interestingly, in addition to the TREX (THOC1/THOC2) complex, the NFAT (IL2/IL3) complex, another R-loop-related METTL8 complex component, Replication Protein A (RPA), has also been identified (Figure 1). RPA is a DNA-binding protein that binds to displaced single-stranded DNA in long and persistent R-loops (>1 kb) in the regions involving class switching (Nguyen et al., 2017; Yu et al., 2003); its co-appearance suggests that METTL8's main function is related to regulatory R-loops in the genome.

SUMOylation as R-Loop Regulator

We have shown that the SUMOylation signaling pathway regulates METTL8 nucleus-cytoplasm shuttling through an evolutionary well-conserved site (K80 in humans) (Figure 5B). This modification is important for METTL8's enrichment in the nucleus (Figure 4), and it may be critical for the formation of the 800 kDa METTL8 complex; studies have shown that SUMO-SIM (SUMO Interacting Motif) fosters both intra- and inter-molecular interactions in protein complexes (Jentsch and Psakhye, 2013), and METTL8 is the only member of its family that contains both a SUMO and a SIM site (Table S4), which suggests that METTL8 may interact with itself and acts as glue to hold the complex together. In addition to METTL8, other R-loop regulators, such as DNA topoisomerase I (TOP1) and RPA, are also SUMOylated (Li et al., 2015; Jentsch and Psakhye, 2013), indicating that SUMOylation could be an important layer of regulation for R-loop biogenesis. Moreover, SUMO ligases have been shown to commonly target assembled complexes that are functionally engaged (Psakhye and Jentsch, 2012; Dou et al., 2010), and most of these SUMO complexes are found in the nucleus; the nucleolar-enriched METTL8 complex regulating R-loops characterized in this study is another good example of SUMO signaling target.

Non-Canonical Genetic Structures and Their Regulators

GC-rich non-canonical genetic structures, such as R-loops, circRNAs, G-quadruplex (guanidine-rich) (Parkinson et al., 2002), and I-motifs (cytosine-rich) (Phan et al., 2000), were previously considered mistakes during cell biogenesis. However, more evidences indicate that the GC-rich region of the genome is in fact an important part of the genetic code. Here, we show that the METTL8 complex influences R-loops formation in the nucleus (Figure 3) and it affects R-loop regulated transcription through m3C (Figure 5). It is known that R-loops influence transcription events such as pausing and splicing, together with recent work showing METTL3 could promote R-loop formation and facilitate transcription termination via m6A (Yang et al., 2019), we believe it is likely that nucleic acid base modifications such as methylation could be a generic biologic mechanism in which R-loops are regulated. In addition, our unpublished data have suggested METTL8 could perturb circular RNA production, which is likely due to the regulatory upstream factors such as genomic R-loops.

Limitations of the Study

Although all data are safely reproduced in this study, some limitations should be noted. In current field, studies on R-loops heavily rely on the S9.6 antibody; assays such as the commonly used DRIP, cell imaging, and quantitative assays (dot blots) are influenced, sometimes distorted, by nonspecific binding to the AU-rich RNAs. We are able to circumvent interference from cytosolic AU-rich RNA by cellular fractioning; we are unable, unfortunately, to assess the influence of nuclear AU-rich RNAs on R-loops detection sensitivity with the current measure.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

All data are available in the main text or the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100968.

AUTHOR CONTRIBUTIONS

X.G. and Y.L. conceived the project; X.G. and L.-H.Z. performed the experiments and wrote the paper; L.-H.Z., X.-Y.Z., and X.-Y.C. repeated the data; J.-J.L and T.H. performed animal works; M.R. performed the FASP mass spectrometry; and N.S. provided the Cas9-CRISPR reagents.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription by-products to threats to genome stability. Mol. Cell *46*, 115–124.

Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. *9*, 204–217.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U S A *51*, 786–794.

Badri, K.R., Zhou, Y., Dhru, U., Aramgam, S., and Schuger, L. (2008). Effects of the SANT domain of tension-induced/inhibited proteins (TIPs), novel partners of the histone acetyltransferase p300, on p300 activity and TIP-6-induced adipogenesis. Mol. Cell Biol. *28*, 6358–6372.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Bhatia, V., Barroso, S.I., Garcia-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature *511*, 362–365.

Borggrefe, T., Wabl, M., Akhmedov, A.T., and Jessberger, R. (1998). A B-cell-specific DNA recombination complex. J. Biol. Chem. *273*, 17025–17035.

Brochu, J., Vlachos-Breton, E., Sutherland, S., Martel, M., and Drolet, M. (2018). Topoisomerases I and III inhibit R-loop formation to prevent unregulated replication in the chromosomal Ter region of *Escherichia coli*. PLoS Genet. *14*, e1007668.

Catoni, G.L. (1953). S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. J. Biol. Chem. *204*, 403–416.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol. Cell *68*, 745–757 e5.

Dou, H., Huang, C., Singh, M., Carpenter, P.B., and Yeh, E.T. (2010). Regulation of DNA repair through deSUMOylation and SUMOylation of replication protein A complex. Mol. Cell *39*, 333–345.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol. Microbiol. *59*, 723–730.

Drolet, M., and Brochu, J. (2019). R-loop-dependent replication and genomic instability in bacteria. DNA Repair (Amst). *84*, 102693.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. *24*, 1546–1558.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chedin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. *23*, 1590–1600.

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

Glasel, J.A. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques *18*, 62–63.

Gold, M., Hurwitz, J., and Anders, M. (1963a). The enzymatic methylation of RNA and DNA. Biochem. Biophys. Res. Commun. *11*, 107–114.

Gold, M., Hurwitz, J., and Anders, M. (1963b). The enzymatic methylation of rna and DNA, Ii. On the species specificity of the methylation enzymes. Proc. Natl. Acad. Sci. U S A *50*, 164–169.

Gomez-Gonzalez, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc. Natl. Acad. Sci. U S A *104*, 8409–8414.

Gomez-Gonzalez, B., and Aguilera, A. (2009). R-loops do not accumulate in transcription-defective hpr1-101 mutants: implications for the functional role of THO/TREX. Nucleic Acids Res. *37*, 4315–4321.

Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J. *30*, 3106–3119.

Gonzalez-Aguilera, C., Tous, C., Gomez-Gonzalez, B., Huertas, P., Luna, R., and Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. Mol. Biol. Cell *19*, 4310–4318.

Gottipati, P., and Helleday, T. (2009). Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. Mutagenesis *24*, 203–210.

Gottlieb, T.M., and Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. Cell *72*, 131–142.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell Biol. *26*, 3986–3996.

Gu, H., Do, D.V., Liu, X., Xu, L., su, Y., Nah, J.M., Wong, Y., Li, Y., Sheng, N., Tilaye, G.A., et al. (2018). The STAT3 target Mettl8 regulates mouse ESC differentiation via inhibiting the JNK pathway. Stem Cell Rep. *10*, 1807–1820.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. Cell *170*, 774–786 e19.

Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. DNA Repair (Amst). *19*, 84–94.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., Mckinney, K.M., Eaton, M.L., et al. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol. Cell *57*, 636–647.

Jakkaraju, S., Zhe, X., Pan, D., Choudhury, R., and Schuger, L. (2005). TIPs are tension-responsive proteins involved in myogenic versus adipogenic differentiation. Dev. Cell *9*, 39–49.

Jenjaroenpun, P., Wongsurawat, T., Sutheeworapong, S., and Kuznetsov, V.A. (2017). R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. Nucleic Acids Res. *45*, D119–D127.

Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S.P., and Kuznetsov, V.A. (2015). QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. Nucleic Acids Res. *43*, 10081.

Jentsch, S., and Psakhye, I. (2013). Control of nuclear activities by substrate-selective and protein-group SUMOylation. Annu. Rev. Genet. *47*, 167–186.

Kogoma, T., Hong, X., Cadwell, G.W., Barnard, K.G., and Asai, T. (1993). Requirement of homologous recombination functions for viability of the Escherichia coli cell that lacks RNase HI and exonuclease V activities. Biochimie *75*, 89–99.

Li, M., Pokharel, S., Wang, J.T., Xu, X., and Liu, Y. (2015). RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. Nat. Commun. *6*, 6720.

Li, X., and Manley, J.L. (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev. *20*, 1838–1847.

Manzo, S.G., Hartono, S.R., Sanz, L.A., Marinello, J., de Biasi, S., Cossarizza, A., Capranico, G., and Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. Genome Biol. *19*, 100.

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., et al. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics Chromatin *8*, 46.

Napoli, S. (2013). Detailed analysis of promoter-associated RNA. Methods Mol. Biol. *1049*, 199–213.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. Mol. Cell *65*, 832–847 e4.

Parkinson, G.N., Lee, M.P., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. Nature *417*, 876–880.

Phan, A.T., Gueron, M., and Leroy, J.L. (2000). The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. J. Mol. Biol. *299*, 123–144.

Psakhye, I., and Jentsch, S. (2012). Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. Cell *151*, 807–820.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. Cell *169*, 1187–1200.

Salas-Armenteros, I., Perez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., and Aguilera, A. (2017). Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J. *36*, 3532–3547.

Salvi, J.S., and Mekhail, K. (2015). R-loops highlight the nucleus in ALS. Nucleus 6, 23–29.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016). Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell *63*, 167–178.

Schlenk, F., and Depalma, R.E. (1957). The formation of S-adenosylmethionine in yeast. J. Biol. Chem. *229*, 1037–1050.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. *28*, 1384–1396.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol. Cell *42*, 794–805.

Smith, J.J., Deane, N.G., Wu, F., Merchant, N.B., Zhang, B., Jiang, A., Lu, P., Johnson, J.C., Schmidt, C., Bailey, C.E., et al. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology *138*, 958–968.

Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chedin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife *5*, 17548–17558.

Takahara, T., Tasic, B., Maniatis, T., Akanuma, H., and Yanagisawa, S. (2005). Delay in synthesis of the 3' splice site promotes trans-splicing of the preceding 5' splice site. Mol. Cell *18*, 245–251.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., de Vos, J., Thomas, A., Theillet, C., et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat. Cell Biol. *11*, 1315–1324.

Xu, L., Liu, X., Sheng, N., Oo, K.S., Liang, J., Chionh, Y.H., Xu, J., Ye, F., Gao, Y.G., Dedon, P.C., and Fu, X.Y. (2017). Three distinct 3-methylcytidine (m(3)C) methyltransferases modify tRNA and mRNA in mice and humans. J. Biol. Chem. *292*, 14695–14703.

Yang, X., Liu, Q.L., Xu, W., Zhang, Y.C., Yang, Y., Ju, L.F., Chen, J., Chen, Y.S., Li, K., Ren, J., et al. (2019). m(6)A promotes R-loop formation to facilitate transcription termination. Cell Res. *29*, 1035–1038.

Yeon, S.Y., Jo, Y.S., Choi, E.J., Kim, M.S., Yoo, N.J., and Lee, S.H. (2018). Frameshift mutations in repeat sequences of ANK3, HACD4, TCP10L, TP53BP1, MFN1, LCMT2, RNMT, TRMT6, METTL8 and METTL16 genes in colon cancers. Pathol. Oncol. Res. *24*, 617–622.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol. *4*, 442–451.

Zhang, Z.Z., Pannunzio, N.R., Hsieh, C.L., Yu, K., and Lieber, M.R. (2015). Complexities due to single-stranded RNA during antibody detection of genomic RNA:DNA hybrids. BMC Res. Notes *8*, 127.

Zheng, L., Roeder, R.G., and Luo, Y. (2003). S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. Cell *114*, 255–266.

Zhou, C., Molinie, B., Daneshvar, K., Pondick, J.V., Wang, J., van Wittenberghe, N., Xing, Y., Giallourakis, C.C., and Mullen, A.C. (2017). Genome-wide maps of m6A circRNAs identify widespread and cell-type-specific methylation patterns that are distinct from mRNAs. Cell Rep. *20*, 2262–2276.

Supplemental Information

The SUMOylated METTL8 Induces

R-loop and Tumorigenesis via m3C

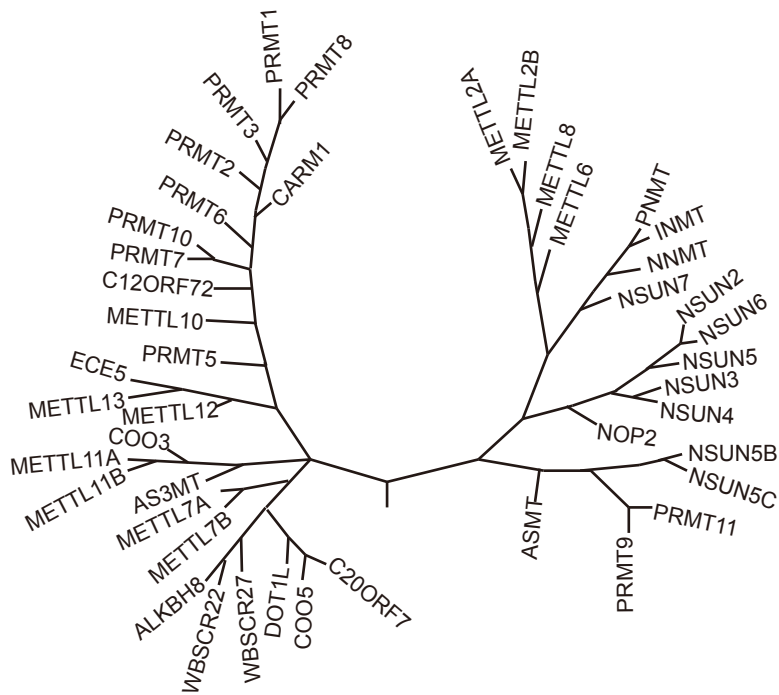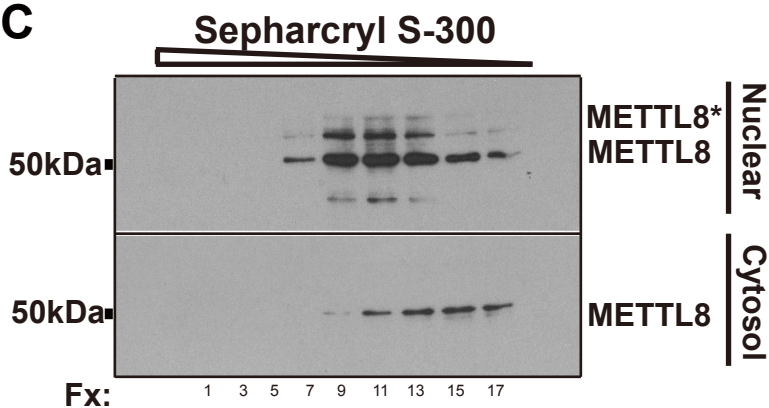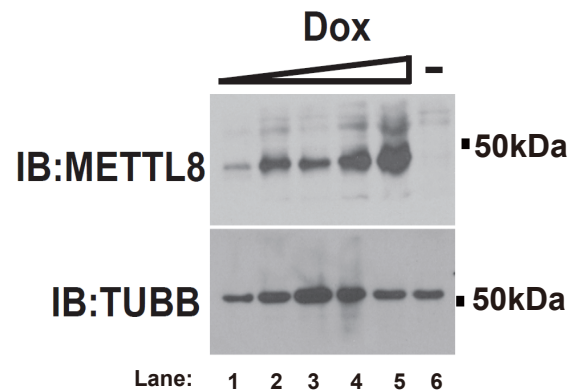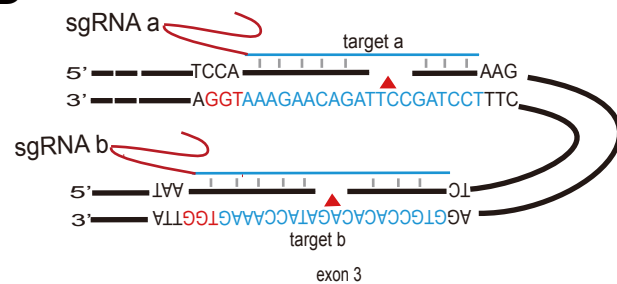Li-Hong Zhang, Xue-Yun Zhang, Tao Hu, Xin-Yun Chen, Jing-Jia Li, Manfred Raida, Ning Sun, Yan Luo, and Xiang Gao
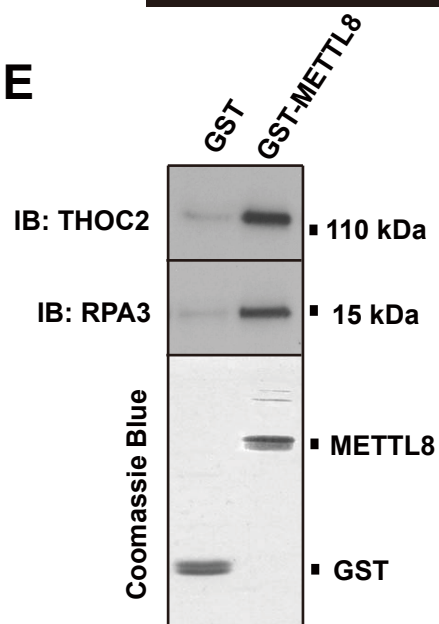
Figure S1. METTL8 protein family and its co-purification of binding partners, Related to Figure 1 (A) The distance between each methyltransferase was computed based on their primary protein sequence and plots were generated using MEGA 7.0.14 software. (B) The concentration of Doxycycline was titrated at 0.125 µg/ml so that ectopic induced METTL8 is close to endogenous protein level (lane 1). (C) Gel filtration (Sepharcryl S-300) separation of nuclear and cytosol METTL8 complex, alternative fractions were loaded onto gel and visualized with anti-METTL8 monoclonal antibody (Sigma, Atlas antibodies). (D) Schematic representation of the genomic locus of the double-nickase CRISPR/Cas9-mediated METTL8 knock out (top, with 2 sgRNAs targeting opposite DNA strands on the 3rd exon). (E) GST-pulldown assay with purified GST-METTL8 fusion protein and the GST protein; direct interaction with THOC2 and RPA3 was visualized with their respective antibody, protein molecular weight was indicated on right.
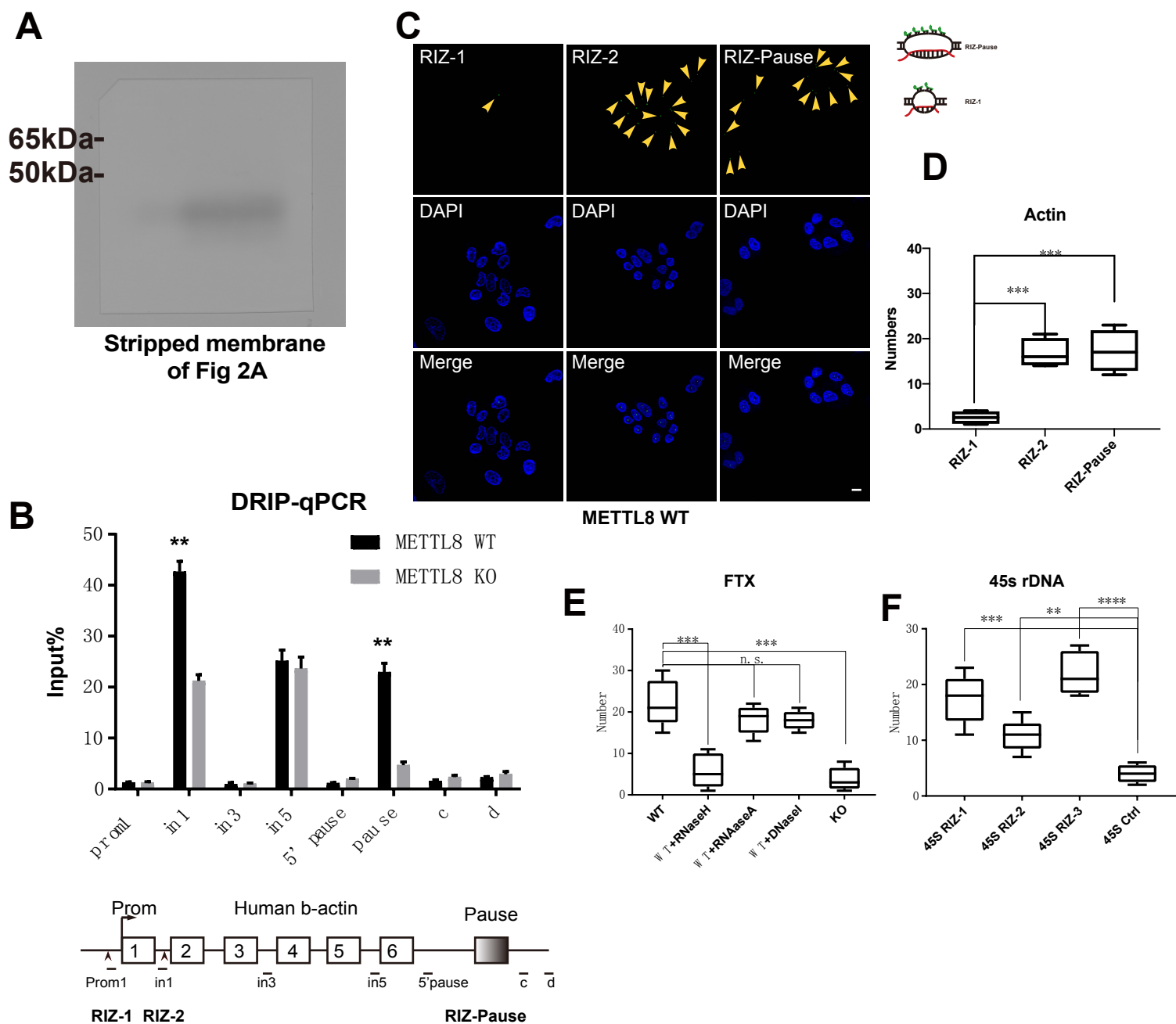
Figure S2. METTL8 regulates R-loop formation on the beta-actin gene, Related to Figure 2 (A) The stripped membrane of Figure 2A. (B) DRIP-qPCR analysis of HeLa METTL8 WT (black bars) and KO (gray bars); the number of R-loop binding sites (on beta-actin) was calibrated against the input nucleic acids; data are means ± SD (n=3, **P<0.01, Student's t-test). Bottom boxes, a schematic of human beta-actin gene dash lines indicating Prom 1, in 1, in3, in5, 5'pause and the pause sites. (C) FISH confocal microscopy imaging using the RIZ probes (green) indicated in (B, schematic on bottom); nuclei were counterstained using DAPI (blue). Scale bar, 10 μm. Arrowheads indicate R-loops detected by the FISH probes. (D) Statistic analysis of (C), stained dots (RIZ-1/2/Pause) were counted and data are mean ± SD, n=4, ****P<0.0001, Student's t-test. (E-F) Statistic analysis of Figure 2E and Figure 2G, respectively. Stained dots (R-loops) were counted and data are mean ± SD, n=4, **P<0.01, ***P<0.005, ****P<0.0001, Student's t-test.
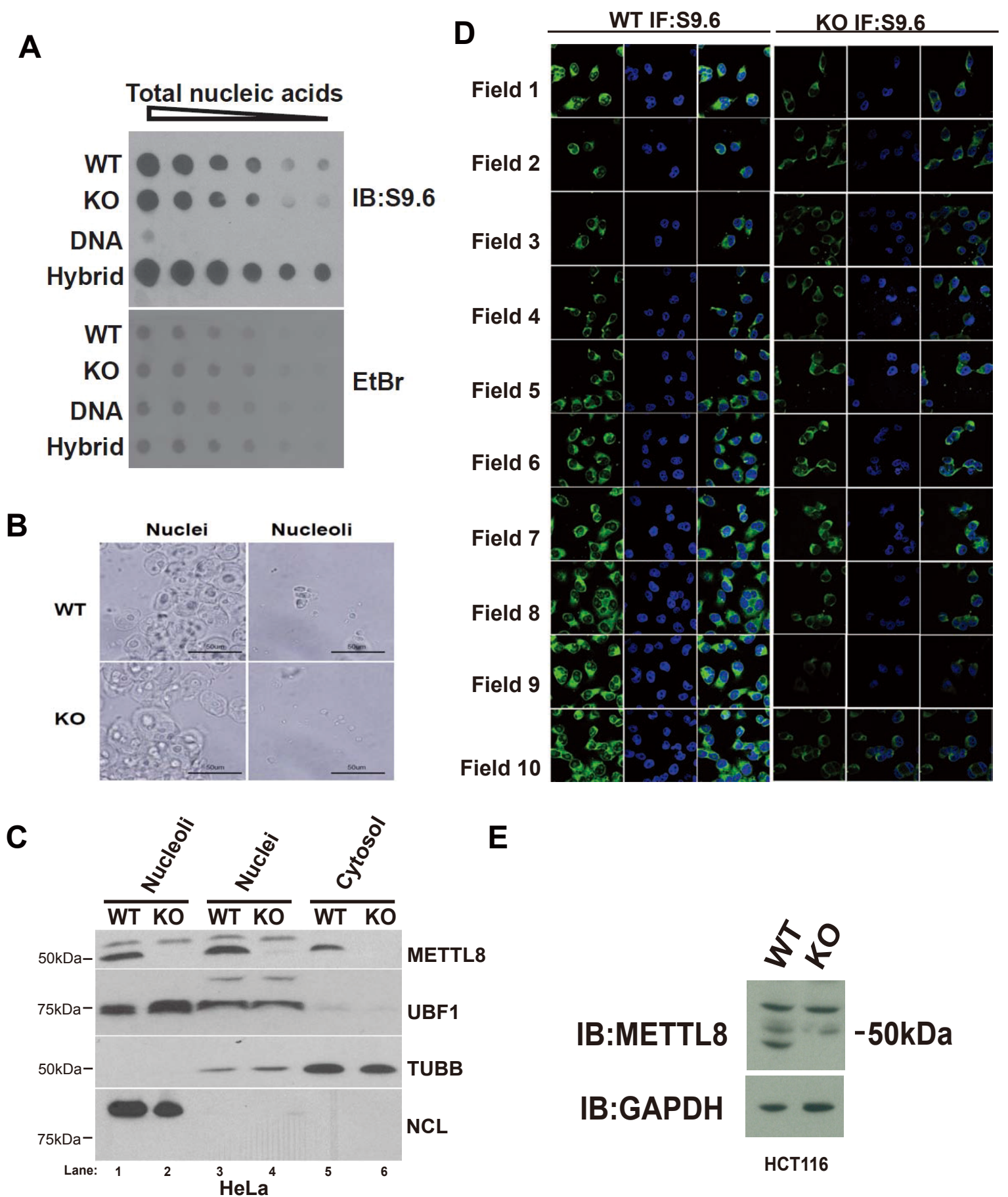
Figure S3. METTL8 regulates R-loops in the nucleolus, Related to Figure 3 (A) Dot blots display of R-loop structures from total HeLa cell nucleic acids. DNA (S9.6 negative, row 3) and DNA : RNA hybrid (S9.6 positive, row 4) were included for control. (B) Phase contrast microscopy showing purified HeLa nulei and nucleolus. Scale bar, 50 μm. (C) Western blots showing the fractioned HeLa METTL8 proteins from WT and KO cells, including cytosol (TUBB), nucleus (UBF1) and nucleolus (NCL) markers as controls, the band for METTL8 is below the 50 kDa marker. (D) Confocal imaging of R-loops (green) with S9.6 antibody, DAPI were included to counter-stain HeLa nucleus. 10 different view fields were taken and statistical analysis was performed on R-loops positive and negative cells (Figure 3F). Scale bar, 10 μm. (E) Western blots of HCT116 METTL8 proteins from WT and KO cells, GAPDH was included for loading control.

Transparent Methods

CRISPR/Cas9 directed knock out

Double-nikase CRISPR-Cas9 mediated knockout cells was done following the instruction of published protocols (Ran et al., 2013). Briefly, guiding RNA targeting to human METTL8 coding region was designed based on instructions on Feng Zhang's Target Finder (http://crispr.mit.edu). The two single guide RNAs with sequence: AAAGAACAGATTCCGATCCT, GTGCCACACAGATACCAAAG were cloned into PX462 vector based on the probability of the off-targets and their predicted on-site activity. Cas9 GFP expression was monitored 24h post transfection and positive clones were handpicked after 20 days of selection under puromycin (8 μg/mL). Cells were then expanded in the absence of puromycin to avoid the generation of any heterogeneity due to additional cutting, and clones screened by western blotting and immunofluorescence. Off-target site analysis of sgRNAs to identify potential off-target sites in the human genome, each sgRNA was BLASTed against the human genome (hg19). All sgRNAs of specified length to the sgRNA 3' end that conform to the NGG PAM used in experiments contains zero possible off-target.

Dot blot analysis

Dot blot was performed according to BioRad Bio-Dot Module (Product No.: 1706547 BioRad) instruction. Briefly, purified nucleic acids (DNA, RNA, R-loops) were treated accordingly (keep on ice or at 37-50℃ water bath) was loaded onto a nylon membrane (Product No.: 11209272001 Roche) using BioRad Bio-Dot Module connected to a vacuum, and then the DNA was crossed-linked using UV crosslinker (Stratalinker 2400, Stratagene). The DNA/RNA hybrid content was visualized using anti-DNA/RNA hybrid antibody (S9.6), while loading control was visualized under UV after ethidium bromide (10 μg/ml) staining.

Dot densitometry and Line scan analysis

Image pixel quantification was performed using ImageJ particle analysis function (area) and Line Profile Tool plugin (intensity), each dot/line (respectively) were scanned at least 3 times and image area/intensity values were averaged and standard deviations were calculated, graphs were drawn using Microsoft Excel.

Fluorescence *In situ* Hybridization (FISH)

FITC tagged DNA probes are synthesized commercially (Generay Biotechnology). R-loops Initiation Zoom (RIZ) sequences (complete list in Table S5) were predicted by the QmRLFS algorithm in R-loop DB (http://rloop.bii.a-star.edu.sg/). FITC tagged RIZ probes were reconstituted in hybridization buffer (10% Dextran Sulfate, 50% Formamide, 2×SSC, 1 mM EDTA, 50 mM Sodium Phosphate buffer pH 7.0). Cells were fixed in 4% PFA　(25℃, 10 min) and permeabilized at least one hour or overnight in 70% ethanol (4℃) before incubating at 37℃ (overnight, in a humidified chamber) with 5 μl 50 mM pre-denatured (85℃, 10 min) RIZ probes. Then the cells were washed 3 times with 50% Formamide, 2×SSC for 5 min each at room temperature to remove all unhybridized or partially hybridized probes. Mount the coverslip on a slide with mounting media containing DAPI (Product No.: H-1200, Vector Laboratories) and fix it in place with a minimal amount of nail varnish and proceed to fluorescence imaging immediately.

Laser confocal microscopy

All confocal imagining were performed on Nikon A1R confocal microscope with ×60 object oil lens. A combination laser wavelength of 405/488/514 nm was used to detect DAPI/FITC/YFP

respectively. The image was analyzed on NIS-Elements Viewer 4.20 and exported to Adobe Illustrator for figure arrangement.

Cell nucleolar isolation

Nucleoli were prepared from HeLa S3, HCT116 cells following the LamondLab.com Nucleolar Isolation Protocol with some modification. Briefly, 10^7 cells were harvest by trypsinization followed by 3× wash with cold PBS. Then cells were resuspended in 3 mL Hypotonic Buffer A (10 mM Hepes, pH 7.9, 10 mM KCl, 1.5 mM MgCl$_2$, 0.5 mM DTT) and swelled on ice for 5 minutes. Swelled cells were checked under a phase contrast microscope and subjected to homogenization (10 strokes) using glass Douncer (Pestle "B", Wheaton Scientific). Homogenized cells were centrifuged at 2000 g for 10 min at 4℃, and supernatants (cytoplasm) and pellets (nuclei) were saved for analysis. 90% of the pellets containing nuclei were subject to nucleolar purification. Nuclei were resuspended in 1 mL S1 (0.25 M sucrose, 10 mM MgCl$_2$) and sonicated on ice for 6 × 5 seconds bursts with 5 seconds intervals between each burst (power setting 5, Misonix XL 2020). Sonicated nuclei were layered over 2 mL of S3 (0.88 M sucrose, 0.5 mM MgCl$_2$) and centrifuged at 3000 g for 10 min. The pellet contains nucleoli while the supernatant is retained as nucleoplasm.

Cell culture and generation of inducible protein complex system

HeLa S3 (ATCC CCL-2.2) cell line is maintained at 37℃, 5% CO$_2$ until 80% confluence before each passage. HeLa S3 suspension culture is achieved by inoculating exponential growth HeLa S3 cell into 10 liters of JMEM (Product No.: M0518 SIGMA) containing 10% FBS and 1% Pen-strep in a Corning (Product No.: CLS4500125 SIGMA) spinner flask. Inducible EF1a-FLAG-METTL8-HA stable cell line is generated by transfecting Clonetech (Cat. No.: 631337) TET-ON 3G vectors containing the EF1a-FLAG-METTL8-HA cassette with Blasticidine selection marker into HeLa S3 cell line, then the transfected HeLa S3 cells were maintained under 100 μM Blasticidine (Cat No.: A1113903 ThermoFisher Scientific) selection medium for two weeks or until colonies appeared. Positive colonies were handpicked and subjected to inducible expression profiling. Final Doxycyclin (Product No.: D3447 SIGMA) concentration for the endogenious level of METTL8 protein expression is deciphered at 0.125 μg/mL.

Liquid chromatography and purification scheme

1. DEAE anion column

DEAE Sepharose resin (Product No.: 17-0709-01) is purchased from GE Healthcare and poured into GE Healthcare xk 16/20 column, the resin bed is packed with a packing reservoir at a flow rate of 0.5 mL/min. After packing, the column is equilibrated with 2 times column volume of binding buffer. 10 mL of 10 mg/mL nuclear extract was loaded onto the DEAE column in low ionic strength buffer (10 mM Tris-HCl, 50 mM NaCl, pH 8.0) and the bounded proteins were step-wise eluted with buffer containing 100 mM, 300 mM, 500 mM and 1 M of NaCl. Each step-fraction was pooled and subjected to western blot analysis using antibody specified in the result section. The peak contains the highest amount of target protein was dialyzed with low ionic buffer overnight at 4℃ and subjected to the next step of purification.

2. Antibody NHS column generation

NHS-agarose-antibody column (Product No.:17-0716-01 GE Healthcare) is generated according to manufacturer's recommendation, anti-FLAG (Product No.: F1804, SIGMA) and anti-HA (Product No.: sc-7392, Santa Cruz) antibodies were purchased commercially.

3. NHS affinity column

Dialyzed DEAE fractions from previous steps were centrifuged to remove traces of precipitation before loading onto NHS-agarose-antibody column using a peristaltic pump. To ensure maximum antibody-antigen binding capacity, the sample was repeatedly passed through the antibody column at a flow rate of 0.2 mL/min for up to 16 hrs at 4℃. After applying the sample with the peristaltic pump, the charged column is linked to an ÄkTA FPLC system for wash and elution steps of the immune-captured protein complex. The wash steps include 2 CV of binding buffer (10 mM Tris-HCl, 50 mM NaCl, pH 8.0) equilibration and 2 CV of low ionic strength buffer (10 mM Tris-HCl, 100 mM NaCl, pH 8.0), after the wash, the complex is eluted with buffer containing 500 mM NaCl.

4. Gel filtration column

Gel filtration (Superose 6 10/300 GL GE Healthcare) column was used to characterize the native size of the immune-purified protein complex, and Sephacryl S-300 16/60 (GE Healthcare) was used to prepare protein fractions from the nuclear and cytosolic compartments. Sample was loaded onto the column and eluted with 1.5 CV isotonic buffers (10 mM Tris-HCl, 150 mM NaCl, pH 8.0).

5. Mono S strong cation column

Mono S strong cation exchange column was chosen for the final polishing step. The column is first equilibrated with binding buffer (10 mM Phosphate, 10 mM NaCl, pH 6.0), and sample dialyzed in the same buffer was loaded on to the column with ÄkTA FPLC system. The complex was eluted using a stepwise method. A salt gradient ladder of 30 mM, 100 mM, 300 mM, and 500 mM was pumped into the column and the complex is collected at 500 mM.

Western blotting and silver staining

All western blots were done according to BioRad Mini-PROTEAN and Mini Trans-Blot electrophoresis Transfer Cell protocol. Silver staining was also done with BioRad Silver Stain Plus kit.

Mass-spectrometry

A FASP (filter-aided sample preparation) aided mass-spectrometry was incorporated downstream of the complex purification. Peptides from the protein complex were extracted and solubilized and loaded onto FASP columns, samples were then alkylated, washed and digested on FASP Protein Digestion kit (Product No.: 44250 EXPEDEON). Digested peptides were then eluted, and dissolved in 0.1% formic acid in water. The eluted peptides were next submitted to nano-flow HPLC coupled to a QTOF mass spectrometer (1260 nano HPLC Agilent Technologies).

Raw Mass-spectrometry data were directly loaded onto PEAKS software and database search was performed against human IPI database (version 3.65), Carboxymethylated Cys was set as fixed modification, oxidized Met, deamidation of Asn and Gln, pyroGlu formation of the N terminus and acetylation of the N terminus as variable modification.

Expression and purification of recombinant proteins

To obtain purified, active proteins for *in vitro* enzyme assays, 6*HIS-SUMO1, 6*HIS-METTL8, 6*HIS-METTL8-K80R and 6*HIS-SUMO1gg were expressed in BL21 E. coli. at 20℃ under 100 μM IPTG induction for a period of 10~16 hours. The cells were then harvested by centrifugation and resuspended in cold PBS and sonicated twice at 10 Watts 5 mins to dissolve the recombinant proteins. The proteins were pre-cleared by centrifugation and

passing through empty agarose column which helps to remove background contaminants. The pre-cleared protein lysates were loaded either onto Gstrap or Histrap (GE healthcare) columns and the bounded recombinant proteins were eluted with 10 mM glutathione and 0.2 M imidazole respectively. The purified proteins were dialyzed against 10 mM Tris-HCl (pH 7.9) + 3 mM MgCl$_2$ to remove Imidazole and Glutathione. Protein concentration is determined by UV absorbance and snaps frozen in liquid nitrogen for storage up to 2 months.

In vitro sumoylation and methylation assay

In vitro sumoylation is carried out using SUMOLink SUMO-1 Kit (Product No.: 40120 ACTIVE MOTIF). Briefly, a master mix of 2 µg GST-UBA2, 1 µg GST-AOS1 and 0.5 µg 6*HIS-UBC9, is mixed with 3 µg of 6*HIS-SUMO1/1aa (non-conjugatable form) and 1~10 µg of METTL8, METTL8 complex or its K80R mutant. These enzymes are then incubated in a buffer containing a final concentration of 10 mM ATP at 37℃ for 1 hour. The reaction is stopped by adding ½ volume of SDS-PAGE sample buffers and boiled for western blots or further analysis.

In vitro methylation is carried out with a master mix of 2 µl S-adenosyl-L-[methyl-3H]-methionine, 2 g of METTL8, METTL8 complex, METTL8-ΔSAM, METTL8-K80R or DNMT1 protein, in combination with either METTL8 WT and KO HeLa nucleoli, or with purified DNA and KO from HeLa cells. Components are incubated in Histone methyltransferase buffer (25 mM Tris-Cl at pH 8.0, 20% glycerol) at 37℃ for 1 hour, and the reaction is either stopped by adding ½ volume of 4× SDS-PAGE sample buffer and boiled for SDS-PAGE separation, or proceed to Dot Blot analysis and liquid scintillation counter (Tri-Carb 2810TR, PerkinElmer).

PAR-CLIP and RNA immunoprecipitation

Photoactivated-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) is used for identifying METTL8 binding RNAs, the detailed procedure is published earlier (Hafner et al., 2010). Briefly, we incorporated photoreactive ribonucleoside analogs into nascent RNA transcripts in living cells, and irradiated the cells with UV at 365 nm inducing efficient crosslinking of the photoreactive nucleoside-labeled RNAs to interact with RNA binding proteins (RBPs), then the FLAG-tagged METTL8 protein complex is immunoprecipitated from the photo-activated cells and the crosslinked and co-immunoprecipitated RNAs were isolated. The isolated RNAs were then reverse-transcribed into cDNA library and subjected to Illumina Hiseq 2000 deep sequencing, raw sequencing data were mapped to the human genome version hg19 by Bowtie 1.0.0, METTL8 binding RNAs were listed in Table S2. For experimental control, the same procedure was repeated in METTL8 KO cells with FLAG antibody expressing FLAG vector, background binding RNAs were mapped to the hg19 and listed in Table S3.

DRIP-qPCR

DNA:RNA (S9.6) immunoprecipitation were performed followed by qPCR. Briefly, 10 µg S9.6 anti-R-loops antibody (Kerafast 1 µg/µl ENH001) were incubated with 5 µg of fragmented DNA in a 1.5 ml tube containing 40 µl of prewashed protein A beads, mix overnight in 4℃. Beads were washed the next day with TE buffer supplemented with 0.05% Tween, purify DNA using Qiagen MinElute kit. qPCR were performed using 2x SYBR premix (Takara RR820L), 0.2 l of 10 µM primers (Primer list) were mixed with 2 µl of S9.6-immunoprecipitate or 2 µl of 1:10 input chromatin.

Dual Luciferase assay

Either Wild-type or Mutant (C to G in m3C RIZ motif) *ACTB* pause regions were cloned into dual luciferase reporter vector with pGL4 backbone (Promega). The DNA sequence for the Wild-type *ACTB* pause region is GGGGAGTTCCTGGCCTGGCCCTTCTATGTCTCCCCAGGTACCCCAGTTTTTCTGGGTTCACCCAGAGTGCAG ATGCTTGAGGAGGTGGGAAGGGACTATTTGGGGGTGTCTGGCTCAGGTGCCATGCCTCACTGGG, whereas the Mutant pause region is GGGGAGTTCCTGGCCTGGCCCTTCTATGTCTCCCCAGGTACCCCAGTTTTTCTGGGTTCA**GGG**AGAGTG**G**A GATG**G**TTGAGGAGGTGGGAAGGGA**G**TATTTGGGGGTGTCTGGCTCAGGTGCCATGCCTCACTGGG, C to G mutations are underlined in bold. The self-cleaving ribozyme gene is subcloned from p370HU (Addgene 27386), and the IRES is subcloned from pIRES2-EGFP (Clontech). These reporter minigenes are then transfected into WT and METTL8 KO HCT116 cells in 24-wells at 90% confluence. Luciferase measurement was performed with standard protocol (Promega), briefly, the transfected cells were lysed in Passive Lysis buffer 24 hours after transfection, 20 µl of cell lysate was transferred to silicon-coated tube containing 100 µl LAR II buffer, mixed by pipetting 2 or 3 times, and Firefly luciferase activity was recorded with a luminometer. Next, 100 µl of Stop & Glo Reagent was added and vortex briefly, Renilla luciferase activity was recorded with the luminometer. All measurements are repeated in triplicates, background luminescence were measured and subtracted from all samples with nontransfected control cells.

ImageJ analysis

Dotblots and microscope images were loaded onto ImageJ (64-bit Java 1.8.0) free software, images were adjusted to gray scale and threshold were adjusted to auto. For dotblots, the dot intensities were measured using "measure" under the "analyze" menu; for line intensity scan, a line were drawn on the image at the sites of interest and line intensity were measured using "line scan" plugin. All measurements were performed in triplicates.

Soft agar colony formation assay

SeaPlaque™ Agarose with low melting temperature (Lonza, 50100) was used in this assay. In triplicate manner, 1.5ml culture medium with 0.6% agarose was first plated into each well of a 6-well plate. After base agarose solidified, another 1.5ml of 0.4% agarose was plated on top in culture medium containing 2000 cells per well. After 20 days colonies grown in soft agar were stained by 1 mg/ml Thiazolyl Blue Tetrazolium Bromide (Sigma, M5655) dissolve in water and then scanned in Bio-Rad Gel Doc EZ system. A population with more than 50 cells was counted as one surviving colony using Quantity One software.

Xenograft assay

1×10^5 METTL8 WT and KO HCT116 cells overexpressing different mutants were mixed with 1/3 matrigel and injected subcutaneously into the left or right flank on the back of 6-week old female NOD-SCID mice. After 4 weeks, tumor tissue was dissected and photographed.

Supplemental References

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P.,

Rothballer, A., Ascano, M., Jr., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M. & Tuschl, T. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell, 141, 129-41.

Ran, F. A., Hsu, P. D., Wright, J., Agawala, V., Scott, D. A. & Zhang, F. 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 8, 2281-2308.