Regulation of SETD2 stability by its intrinsically disordered regions maintains the fidelity of H3K36me3 deposition

Saikat Bhattacharya[1], Ning Zhang[1], Hua Li[1] and Jerry L. Workman[1*]

[1]Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA

ABSTRACT

The SET domain-containing protein SETD2 is the sole methyltransferase in mammals that can tri-methylate histone H3 at lysine 36. H3K36me3 is known to be involved in transcription elongation, pre-mRNA splicing, DNA methylation, and DNA damage repair. However, knowledge of the regulation of the SETD2 enzyme itself is limited. Here we show that the poorly characterized N-terminal region of SETD2 plays a determining role in regulating the stability of SETD2. This stretch of 1-1403 amino acid residues which contains disordered regions, is targeted for degradation by the proteasome. In addition, the SETD2 protein is aggregate-prone and forms insoluble inclusion bodies in nuclei especially upon proteasome inhibition. Removal of the N-terminal segment results in the stabilization of SETD2 and leads to a marked increase in global H3K36me3 which, uncharacteristically, can happen in an RNA Pol II-independent manner. The spurious H3K36me3 is deposited in a non-canonical distribution including reduced enrichment over gene bodies and exons. An increased SETD2 abundance leads to widespread changes in transcription and alternative splicing. Thus, the regulation of SETD2 levels through intrinsically disordered region-facilitated proteolysis is important to maintain the fidelity of transcription and splicing related processes.

INTRODUCTION

The N-terminal tails of histones protrude from the nucleosome and are hotspots for the occurrence of a variety of post-translational modifications (PTMs) that play key roles in regulating epigenetic processes. H3K36me3 is one such important functionally characterized PTM. In yeast, this mark suppresses cryptic transcription from within the coding region of genes by preventing histone exchange (Venkatesh *et al*, 2012). In mammalian cells, it is involved in the recruitment of DNA repair machinery, in splicing and also, in establishing DNA methylation patterns by acting as a binding site for the enzyme DNMT3a (Qin *et al*, 2017)(Kolasinska-zwierz *et al*, 2009)(Dhayalan *et al*, 2010)(Li *et al*, 2013)(Pfister *et al*, 2014). H3K36me3 marks actively transcribed genes and is important in preventing the spread of heterochromatin boundaries by antagonizing PRC2 activity and the H3K27me3 repressive mark (Yuan *et al*, 2011). Recent reports have emphasized the tumor-suppressive role of H3K36me3 in renal cancers especially, where the

gene coding for the SETD2 histone methyltransferase is often deleted or mutated (Chiang *et al*, 2018)(Su *et al*, 2017)(Li *et al*, 2016).

In yeast, the SET domain-containing protein Set2 (ySet2) is the sole H3K36 methyltransferase (Strahl *et al*, 2002). ySet2 interacts with the large subunit of the RNA polymerase II, Rpb1, through its SRI domain, and co-transcriptionally deposits H3K36me3 (Xiao *et al*, 2003). The deletion of the SRI domain from ySet2 abolishes both the Set2-RNA Pol II interaction and H3K36me3 methylation in yeast (Suzuki *et al*, 2016). H3K36 methylation is a highly conserved histone mark and Set2 homologs are found in more complex eukaryotes (McDaniel & Strahl, 2017). These homologs share the conserved features like the AWS (**a**ssociated **w**ith **S**ET), SET [**S**u(var)3-9, **E**nhancer-of-zeste and **T**rithorax] and Post-SET domains that are required for the catalytic activity of the enzyme, and also, the protein-protein interaction regions such as the WW and SRI (**S**et2-**R**pb1 **I**nteraction) domains. However, there are differences in both the manner of H3K36me3 deposition in mammals as well as in the enzyme itself as compared to yeast. For instance, although SETD2 is the sole methyltransferase in mammals that can deposit the histone H3K36me3 mark, there are additional enzymes such as NSD1, NSD2, and ASH1L that can deposit H3K36me1 and me2 (Lucio-Eterovic *et al*, 2010)(Kuo *et al*, 2011)(Tanaka *et al*, 2007). Notably, SETD2 has a long N-terminal segment that is not present in ySet2. The function of this region has remained obscure (McDaniel & Strahl, 2017)(Hacker *et al*, 2016).

Here we show that SETD2 is an inherently aggregate-prone protein and its N-terminal region regulates its half-life. This, in turn, is important for the fidelity of H3K36me3 deposition and the processes of transcription and splicing. Although the regulation of SETD2 stability by the proteasome has been reported earlier, our findings reveal that SETD2 is regulated by its intrinsically disordered regions and highlights the importance of such sequences in governing appropriate protein function and activity.

RESULTS

Deletion of the N-terminal region of SETD2 leads to changes in transcription and splicing

SETD2 is a large protein with 2564 amino acids. The C-terminal region of the protein (1404-2564 residues) has characterized domains that are conserved with ySet2. However, SETD2 has an N-terminal segment (1-1403 residues) not found in ySet2, the function of which is not clear [Figure 1a]. Notably, the

NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) revealed that the SETD2 N-terminus lacks any known domains or motifs and did not yield any clue about its possible function. We hypothesized that if this region is functionally important, then the rescue of *setd2Δ* cells with a truncated form of SETD2, that lacks the N-terminal region (1-1403 amino acids), will have dissimilar outcomes on the transcriptome as compared to rescue with full-length SETD2. To test this hypothesis, constructs were made to express full-length (SETD2 FL) or the C-terminal segment of SETD2 (1404-2564, SETD2 N3) under the control of a CMV promoter. Recombinant expression of SETD2 has been previously used to investigate the function of the protein (Zhang *et al*, 2020)(Chiang *et al*, 2018)(Zhu *et al*, 2016)(Carvalho *et al*, 2014)(Chen *et al*, 2017a). The constructs were introduced in *setd2Δ* 293T (KO) cells in which exon 3 of both the alleles of the endogenous SETD2 gene were disrupted using TALEN (Hacker *et al*, 2016). Subsequently, high-throughput RNA-seq was performed.

Expression of SETD2 FL and SETD2 N3 led to transcriptome changes in the KO cells [supplementary information S1]. Strikingly, a comparison of SETD2 N3 versus SETD2 FL-expressing cells revealed that a total of 1234 genes exhibited significant differential expression [Fold change (FC) ≥2, false discovery rate (FDR) ≤0.05] [Figure 1b]. The majority of these genes (1118 genes) were upregulated on SETD2 N3 expression and only 116 were downregulated as compared to SETD2 FL. Noticeably, Gene Ontology Enrichment Analysis revealed that the upregulated genes are not enriched in any pathway. This suggests that the removal of the N-terminal region of SETD2 leads to widespread changes in transcription without affecting any specific pathway.

Analysis of our transcriptome data to look for possible differences in splicing also revealed widespread changes in cells expressing SETD2 N3 versus SETD2 FL. A total of 2726 differential alternative splicing events (FDR ≤0.05) were identified. Out of these, 1924 events were upregulated and 802 were downregulated upon SETD2 N3 expression [Figure 1c]. The changes belonged to the different kinds of splicing events, such as skipped exon, alternative 5’ splice site, alternative 3’ splice site, mutually exclusive exons, and retained intron. The skipped exon and retained intron events were the most common.

Collectively, our data suggest that removal of the N-terminal region of SETD2 leads to widespread changes in the transcriptome of mammalian cells and indicates a functional role of this region.

Removal of the N-terminal region stabilizes the SETD2 protein and leads to a marked increase in global H3K36me3 levels

To gain insights into how the N-terminal region might affect the function of SETD2, we first wanted to test whether removal of this stretch affects the catalytic activity of SETD2. To check the activity of the exogenously introduced SETD2 constructs, KO 293T cells were used. Consistent with the role of SETD2 as the sole H3K36me3 depositor in humans, in the KO cells the H3K36me3 mark was not detected in the whole-cell lysates by immunoblotting [supplementary information S2a]. Next, constructs to express Halo-tagged SETD2 FL or SETD2 N3 were introduced in KO cells by transfection. 72 hours post-transfection, whole-cell lysates were prepared and analyzed by western blotting. The expression of the empty vector (VC) did not rescue H3K36me3 as expected [Figure 1d]. Strikingly, the expression of SETD2 N3 in KO cells led to a marked increase in the H3K36me3 level as compared to the rescue with SETD2 FL [Figure 1d]. The other two H3K36 methyl marks, H3K36me1 and H3K36me2, largely remained unchanged [Figure 1d].

Some methyltransferases are known to harbor autoinhibitory domains (Wang *et al*, 2015)(Hou *et al*, 2019)(Qiao *et al*, 2011). It is unknown whether SETD2 contains such a domain, and if its absence in SETD2 N3 affects the H3K36me3 levels. To discern between this possibility and whether the difference in H3K36me3 is simply due to altered abundance of the truncated (N3) vs the full-length (FL) protein, western blotting of the whole-cell lysates was performed with an anti-Halo antibody. Strikingly, although Halo-SETD2 FL could rescue H3K36me3 in KO cells, it could not be detected using an anti-Halo antibody, suggesting that the expression level was very low [Figure 1e](discussed in the Discussion section). A robust signal was obtained for both Halo-Vector control (VC) and Halo-SETD2 N3 lanes demonstrating that the expression of the smaller SETD2 fragment was considerably higher than SETD2 FL [Figure 1e]. RT-PCR using primers specific for SETD2 FL and SETD2 N3 transcripts confirmed that the transcripts were produced robustly [supplementary information S2b]. Also, the transcript levels of

SETD2 FL and SETD2 N3 were comparable, which was also observed in our RNA-seq data, suggesting that the differences observed in protein abundance are not due to the differences in the transcription of the two constructs but rather, might be due to instability of the full-length SETD2 protein. Similar results were obtained with the GFP-tagged constructs of SETD2. H3K36me3 level was much higher on SETD2 N3 expression and the fluorescence intensity of GFP-SETD2 FL was considerably weaker than GFP-Vector control and GFP-SETD2 N3 [supplementary information S2c, d].

Thus, full-length SETD2 does not accumulate in human cells. The removal of the N-terminal segment leads to a marked increase in its concentration that is also manifested in the noticeable increase in the global H3K36me3 level. Also, these experiments demonstrate that in the absence of the N-terminal region, SETD2 retains its histone methyltransferase activity.

SETD2 is robustly degraded by the ubiquitin-proteasome pathway

Autophagy and the ubiquitin-proteasome system (UPS) are the major pathways for protein degradation in mammalian cells (Cooper, 2000). To investigate whether autophagy plays a role in SETD2 turn-over, 293T cells expressing GFP-SETD2 FL were treated with increasing concentration of the lysosome inhibitor chloroquine. Chloroquine treatment did not have an apparent effect on the GFP-SETD2 FL level [supplementary information S2e].

To test whether SETD2 is targeted for degradation by the UPS, SETD2 FL expression was checked after treating the cells with the proteasome inhibitor MG132. Proteasome inhibition led to an increase in the accumulation of SETD2. This was confirmed by western blotting of Halo-SETD2 FL expressing cells as well as microscopy of GFP-SETD2 FL [supplementary information S3a, Figure 1f]. The addition of MG132 did not have a prominent effect on Halo or GFP expression [supplementary information S3b, c]. Also, RT-PCR did not reveal any change in the transcript abundance of both endogenous as well as exogenous SETD2 on MG132 treatment, suggesting that the increased SETD2 protein abundance observed is due to protein stabilization [Figure 1g]. Importantly, an increase in the accumulation of endogenous SETD2 was also observed by western blotting upon MG132 treatment of WT 293T cells, implying that degradation by the proteasome is not limited to the recombinant protein [Figure 1h].

To check whether the robust proteasomal degradation of SETD2 is unique to 293T cells, the expression of GFP-SETD2 FL was tested in HEPG2 and HELA cells. SETD2 behaved similarly in these cell lines. Very weak expression of GFP-SETD2 FL was observed which increased upon proteasome inhibition by MG132 treatment [supplementary information S3d]. Our results suggest that the short half-life of SETD2 due to UPS mediated decay is not cell line-specific and is a characteristic of the SETD2 protein.

Next, we turned our attention to the question of how the SETD2 protein is targeted for degradation. We checked the effect of the E3 ubiquitin ligase SPOP and anaphase-promoting complex (APC) on SETD2 degradation as these have been reported to promote degradation of the SETD2 protein (Zhu *et al*, 2016)(Dronamraju *et al*, 2018) [please see supplementary information S4 and the associated results for details]. Our results agreed with the previously published data that SPOP regulates SETD2 stability, but also indicated that the post-translational control of SETD2 abundance is more complicated than previously thought and may contain several redundancies. We also investigated the effect of PEST motifs on SETD2 degradation as it is known to lead to robust degradation (37). However, no apparent effect on SETD2 stability was observed on the deletion of a PEST motif [please see supplementary information S5 and the associated results for details].

SETD2 has long intrinsically disordered regions

Our experiments by mutating SPOP and APC binding motifs in SETD2 hinted that the control of SETD2 expression may be more complex than would be the case if it were solely mediated by a distinct domain(s) or motif(s). Besides the fact that the SETD2 N-terminus lacks any known domains or motifs, this segment has very little sequence similarity with other known protein sequences. Consequently, homology modeling of the structure using SWISS-MODEL (https://swissmodel.expasy.org/) failed. Strikingly, the sequence is also 81% disordered based on Robetta prediction (http://robetta.bakerlab.org/). This was of interest because intrinsically disordered regions (IDRs) are known to regulate protein half-life (Tompa *et al*, 2008)(van der Lee *et al*, 2014). A study with 3,273 proteins in yeast, 4,502 in mouse, and 3,971 in humans revealed strong correlations between IDRs and protein half-life. The study revealed that

long N-terminal and internal disordered segments contribute to short protein half-life *in vivo* (van der Lee *et al*, 2014). This prompted us to perform a deeper analysis of the disordered regions in SETD2 protein.

Prediction of the disordered region in SETD2 was performed using IUPRED2 (https://iupred2a.elte.hu/plot) (Mészáros *et al*, 2018). IUPred provides a score between 0 and 1 for every residue, corresponding to the likelihood of the given residue being a part of a disordered region. Hence, a score of greater than 0.5 means that it is more likely that the residue is a part of a disordered region. Strikingly, 1546 of the 2564 (60.29%) residues of SETD2 protein returned a score of >0.5 as compared to a Halo control (0%) [supplementary information S6]. Furthermore, 67.6% of residues in the N-terminus scored >0.5 as compared to 51.4% of residues of C-terminus.

Next, we looked for long disordered regions (>30 residues) in SETD2 in accordance with a previously published report (van der Lee *et al*, 2014). The length was based on findings that a critical minimum length of ~30 residues allows a disordered terminus of a ubiquitinated substrate to efficiently initiate proteasomal degradation (Da Fonseca *et al*, 2012)(Lasker *et al*, 2012b). Strikingly, 17 such stretches were found in SETD2 out of which 11 were in the N-terminal region, 5 in C-terminal fragment and 1 overlapping between and the N and C-terminal segments [supplementary information S6].

Our findings regarding stability and disordered region prediction in SETD2 suggest the possibility that the disordered regions of SETD2 may govern its half-life.

Multiple disordered segments of SETD2 have a combined effect on its half-life

Interestingly, proteins with several internal disordered segments have shorter half-lives than proteins with only one such segment (van der Lee *et al*, 2014). SETD2 undergoes robust degradation and is predicted to have numerous disordered segments throughout. We speculated that if the multiple disordered segments of SETD2 collectively enhance its proteolysis, then the expression of the shorter fragments will be higher than the longer ones. To test this, a series of constructs were made to express Halo-tagged N- or C-terminal truncations of SETD2 in 293T cells [Figure 2a, b, c, d]. The use of truncation mutants to determine the region of protease sensitivity in a protein is a commonly used approach and has been successfully used (Sano *et al*, 2007)(Zhu *et al*, 2016). Western blotting of whole-cell extracts with an anti-Halo revealed that for both the N and C-terminal truncations, SETD2 expression anti-correlated with the

length of the fragment [Figure 2c, d]. Also consistent with our hypothesis, three non-overlapping fragments of Halo-SETD2; C4, 504-1403 and N3 expressed robustly in 293T cells whereas SETD2 FL could not be detected by anti-Halo western blotting [Figure 2e, f]. Notably, microscopy analyses revealed that fragment 504-1403 was cytoplasmic, unlike C4 and N3 which were nuclear [supplementary information S7a]. To test whether the localization of this fragment alters its stability, the C-myc nuclear localization signal (NLS) was added to GFP-504-1403 (504-1403') and the expression was checked. Microscopy revealed that the addition of NLS resulted in the nuclear translocation of 504-1403 and also, reduced its expression level [supplementary information S7b]. This was also confirmed by western blotting of the whole-cell lysates [supplementary information S7c]. Nevertheless, the expression of all the SETD2 fragments was very robust compared to SETD2 FL and continued to display sensitivity to MG132.

To confirm that the differences observed in the expression are indeed due to the shorter half-life of larger SETD2 fragments, a time-chase experiment was performed to monitor the expression of Halo-N1 and N4 on the treatment of cells with cycloheximide, a translation inhibitor. Post 2 hours cycloheximide treatment, Halo-N1 could no longer be detected by western blotting whereas no appreciable decrease was observed for Halo-N4 even after 4 hours [Figure 2g, h]. Cycloheximide chase experiment for Halo-N4 was performed to include later time points. Halo-N4 could be clearly detected even after 48 hours of treatment [Figure 2i] demonstrating that the shorter fragment N4 has a longer half-life than N1. Importantly, the observed differences in the half-life of the proteins along with our RT-PCR and RNA-seq data also validate that the dissimilarities in the expression of the SETD2 constructs are indeed due to the differences in protein stability and is not due to experimental variations like transfection efficiency.

From these experiments, no specific region emerged in the SETD2 protein that is particularly targeted for UPS-mediated decay and confirmed that the turnover of SETD2 is regulated by factors besides those that were previously reported. Furthermore, consistent with the report that IDRs have a combined effect on protein half-life, multiple regions of SETD2 were found to co-operatively regulate its proteolysis.

IDR-rich N-terminus can reduce the half-life of ySet2

In yeast, ySet2 is a well-characterized protein that is degraded by the proteasome (Fuchs *et al*, 2012). Our data revealed that ySet2 is expressed much more than SETD2 in 293T cells [supplementary information S8b]. We speculated that the disparity in expression between ySet2 and SETD2 could be due to differences in disordered region abundance between the proteins. Disordered region prediction revealed that overall ySet2 is a well-ordered protein with a much lower proportion of residues (24.69%) predicted to be disordered as compared to its homolog SETD2 [supplementary information S6d]. The gain or loss of disordered segments may be an important contributor to the degradation rate of proteins during evolution (van der Lee *et al*, 2014). It is possible that similar evolutionary factors favored the development of the IDR rich SETD2 to promote its degradation. We wanted to test whether an N-terminal segment of SETD2 can destabilize the ySet2 protein when fused to its N-terminus [Figure 3a].

First, the expression of GFP- or Halo-SETD2 FL/N3/ySet2 was compared in 293T cells by microscopy and western blotting of whole-cell extracts with an anti-Halo antibody. The results showed that the expression of ySet2 was significantly higher than SETD2 FL and is comparable to SETD2 N3 with which it shares conservation [supplementary information S8b, c]. Also, ySet2 localized to the nucleus of mammalian cells [supplementary information S8c]. Next, to test whether the N-terminal segment of SETD2 can destabilize ySet2, constructs were made to express Halo-tagged chimeric proteins that have increasing stretches of the N-terminal segment of SETD2 (967-1403, 504-1403 and 1-1403) fused upstream of ySet2 [Figure 3b]. The expression of these constructs was tested in 293T cells by probing whole-cell extracts with an anti-Halo antibody. A progressive decrease in the accumulation of chimeric ySet2 was observed with the increasing length of the N-terminal segment and all the chimeras displayed sensitivity to MG132 treatment [Figure 3c].

Next, we wanted to test whether the destabilization brought about by the SETD2 N-terminus is influenced by the order in which the fragments are added upstream of ySet2. For this, chimeric constructs were prepared in which the segments α, β or α + β were fused upstream of ySet2 while the fragment γ was not included [Figure 3d]. The expression of these chimeric constructs was tested in 293T cells by probing whole-cell extracts with an anti-Halo antibody [Figure 3e]. Addition of α and β destabilized ySet2

and fusing α + β destabilized it further, even in the absence of fragment γ. The data also suggests that the SETD2 fragments co-operatively regulate stability consistent with the idea that IDRs have a combined effect on protein half-life.

The UPS and the architecture of the proteasome itself are conserved from yeast to mammals (Da Fonseca *et al*, 2012)(Lasker *et al*, 2012a). Consistent with that, our data shows that ySet2 responds to MG132 treatment in 293T cells, suggesting, that it is targeted for degradation through UPS in human cells too [supplementary information S8d and Figure 3c, e]. It is speculated that the observed relationships between IDRs and protein-half life may be evolutionarily conserved (van der Lee *et al*, 2014). Hence, we wondered whether the IDR-rich SETD2 N-terminal segment can enhance ySet2 proteolysis in yeast. To test this, the expression of FLAG-tagged WT and chimeric ySet2 proteins was checked in the yeast strain BY4741. The expression was scored by probing the whole-cell extracts with an anti-FLAG antibody. Although the addition of SETD2 fragment γ (967-1403) did not have a large impact, the addition of β and α + β had a drastic destabilization effect on ySet2 protein [Figure 3f].

Collectively, our results demonstrate that the IDR rich N-terminus of SETD2 protein has a destabilization effect on both SETD2 and its yeast homolog ySet2. Importantly, they also suggest evolutionary conservation of the role of the N-terminal region of SETD2 in bringing about the degradation of a protein that is key evidence to support IDR-mediated degradation

SETD2 forms puncta that are characteristic of intranuclear aggregates

Microscopy revealed that GFP-C4 is nuclear and formed puncta [supplementary information S7a]. This was surprising because this segment doesn't have an NLS with a significant score as per NLS mapper prediction (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) (Kosugi *et al*, 2009). We decided to characterize the NLS of SETD2 to better understand the unexpected localization of the SETD2 fragments. To this end, the localization of a series of GFP-SETD2 fragments was checked using fluorescence microscopy (data not shown). Of all the SETD2 fragments tested, the data revealed the presence of putative NLS in three fragments of SETD2: 967-1690, 1964-2263 and 2423-2564 [Figure 4a]. This was consistent with NLS mapper prediction that revealed the presence of NLS in each of these segments [supplementary information S9a]. To validate these NLS, site-directed mutagenesis was

performed to mutate the K and R residues to A and the disruption of the nuclear localization of the mutated GFP-SETD2 fragments was confirmed by microscopy [Figure 4a, supplementary information 9b]. Further, to validate that full-length SETD2 has only these three NLS, site-directed mutagenesis was performed to mutate these NLS one by one [Figure 4a]. The cytoplasmic localization of the GFP-SETD2 FL mutant, in which all the three NLS were disrupted, confirmed that SETD2 has three NLS [Figure 4a].

Importantly, this also shows that the SETD2 fragment C4 forms nuclear puncta without an NLS. Also, treatment with MG132 caused nuclear translocation and formation of puncta by SETD2 504-1403 [supplementary information S7b] suggesting that SETD2 might be an aggregate-prone protein. To test whether the full-length SETD2 protein also behaves similarly, the localization of NLS mutants of SETD2 was tested on MG132 treatment. Strikingly, all the mutants exhibited the formation of nuclear puncta [Figure 4a]. Interestingly, very similar to our observations with SETD2, truncated N-terminal fragments of huntingtin with expanded glutamine repeats are known to form nuclear aggregates in cell culture even without an NLS (Cooper *et al*, 1998).

To characterize the aggregate-prone tendency of SETD2 further, 293T cells expressing GFP-SETD2 truncations (described in Figure 2) were observed under the microscope with or without MG132 treatment. The microscopic approach corroborated the expression level data obtained from immunoblotting; confirming that smaller SETD2 fragments express more robustly than larger ones (data not shown). Remarkably, treatment with MG132 led to a marked increase in the formation of puncta by all the SETD2 fragments [Figure 4b]. Notably, the completely cytoplasmic fragment C3 showed a pan-cellular distribution with a tendency to form puncta upon MG132 treatment similar to the segment 504-1403 and SETD2 FL NLS Mutant 3. Furthermore, puncta were also formed by WT GFP-SETD2 FL protein suggesting that it might not be an artifact caused due to protein misfolding resulting from the truncations or mutations [Figure 4b]. To test whether endogenous SETD2 behaves similarly, immunofluorescence of 293T cells was performed with an anti-SETD2 antibody. Importantly, proteasomal inhibition caused speckle-like staining, revealing that endogenous SETD2 also forms puncta [Figure 4b]. This observation, together with the fact that despite very weak expression, SETD2 FL is aggregate prone,

suggests that aggregation is an intrinsic property of the SETD2 protein that is exacerbated by increased protein abundance.

SETD2 forms ubiquitinated insoluble aggregates

Exogenously expressed as well as endogenous SETD2 formed puncta, especially on MG132 treatment, that are reminiscent of inclusion bodies comprising of aggregated proteins. The most striking pattern was observed with fragment C4 that formed distinct puncta even in the absence of proteasome inhibition. As aggregates are often polyubiquitinated, to confirm that the puncta formed by SETD2 are aggregated structures, the colocalization of RFP-ubiquitin with GFP-C4 was tested. Clear colocalization was observed suggesting that C4 puncta are indeed ubiquitinated aggregate structures [Figure 5a]. GFP-C4 did not colocalize with RFP-Fibrillarin indicating that the puncta are not nucleolar [Figure 5a]. To biochemically substantiate that SETD2 is ubiquitinated, Halo-N3 was affinity-purified from HEK293T cells co-expressing HA-ubiquitin with or without MG132 treatment. The purified proteins were then resolved on a gel and analyzed for the presence of ubiquitination. Western blotting with an anti-HA antibody revealed that SETD2 N3 was indeed ubiquitinated [Figure 5b].

To confirm further that SETD2 forms aggregates, we checked the solubility of Halo-SETD2 fragments C4, 504-1403 and N3. 293T cells expressing Halo-SETD2 fragments were lysed, their soluble and insoluble fractions separated and afterward analyzed by western blotting with an anti-Halo antibody. Correlating with the microscopy observations, C4 that formed spontaneous puncta was highly insoluble, as to a lesser extent was N3 [Figure 5c]. The segment 504-1403 was soluble [Figure 5c]. Interestingly, the data shows that the different regions of the same protein have different aggregation propensity under similar experimental conditions and are not merely a consequence of expression from a strong CMV promoter.

Collectively, our data show that SETD2 forms aggregated ubiquitinated puncta (discussed in the Discussion section).

At high cellular levels, SETD2 has a reduced RNA pol II dependency for H3K36me3 deposition

We found that the removal of the N-terminal region of SETD2 leads to the stabilization of the remaining portion that shares conserved domains with ySet2. This leads to a marked increase in global H3K36me3

levels as scored by western blotting [Figure 1d, supplementary information 2c]. Studies conducted in yeast have revealed that the deposition of the H3K36me3 mark is strictly dependent on the ySet2-Pol II association (Suzuki *et al*, 2016). We were curious whether the marked increase in the global H3K36me3 level that occurs on SETD2 N3 expression happens in an RNA Pol II-dependent manner.

To test this, Halo-SETD2 constructs without the SRI domain were introduced in *setd2Δ* 293T cells. Similar to the findings for ySet2 in yeast, removal of the SRI domain from full-length SETD2 protein (FLΔSRI) led to a marked decrease in H3K36me3 levels as compared to the FL [Figure 6a, supplementary information 10a]. Strikingly though, removal of the SRI domain from SETD2 N3 (N3ΔSRI) had a very marginal effect on the H3K36me3 levels [Figure 6a]. To confirm that the removal of the SRI domain leads to the abolishment of SETD2-Pol II interaction, Halo-FLAG-SETD2 N3 and Halo-FLAG-SETD2 N3ΔSRI were affinity purified from 293T extracts using Halo ligand-conjugated magnetic resin. Elution of proteins purified using this technique involves cleaving off the Halo tag with TEV protease, leaving the FLAG epitope which can be detected from the eluted bait by immunoblotting [Figure 6b]. Immunoblotting with an anti-Pol II antibody confirmed that the deletion of the SRI domain from SETD2 leads to the abolishment of SETD2-Pol II interaction [Figure 6b].

We wondered whether the decreased dependency on Pol II interaction for the H3K36me3 activity of SETD2 N3 is due to the loss of a possible autoinhibition by the N-terminal region of SETD2 or is due to the increased expression of SETD2 N3 fragment as compared to the full-length protein. To address these possibilities, Halo-SETD2 N3 constructs under the control of the CMVD2 promoter were introduced in *setd2Δ* cells. CMVD2 promoter is a truncated form of CMV and exhibits a much weaker activity. The weaker activity of the CMVD2 promoter was confirmed by RT-PCR [supplementary information S10b]. The reduced expression of SETD2 N3 under the regulation of CMVD2 promoter was verified by analyzing whole-cell extracts with an anti-Halo antibody [Figure 6a]. Notably, analysis of H3K36me3 revealed that the RNA Pol II dependency of SETD2 N3 was restored at reduced expression level as SETD2 N3ΔSRI did not exhibit much activity when expressed using the CMVD2 promoter [Figure 6a, supplementary information S10a].

We conclude that at high cellular levels, SETD2 has a reduced RNA pol II dependency for H3K36me3 deposition.

High levels of SETD2 leads to non-canonical H3K36me3 distribution

We wondered whether the increased SETD2 accumulation also leads to an altered distribution of the H3K36me3 mark. For this, first, spike-in normalized H3K36me3 ChIP-Seq of WT and KO cells was performed. Metagene analysis revealed clear enrichment of H3K36me3 within the coding region of the genes in the WT cells [supplementary information S10c]. As expected, a similar pattern was not observed in KO cells that lack SETD2 and hence, H3K36me3 [supplementary information S10c]. Additionally, metagene analysis of H3 normalized H3K36me3 of protein-coding genes highlighted that H3K36me3 is greatly enriched on highly expressed genes as compared to the lowly expressed ones, consistent with the idea that the mark is associated with transcriptionally active genes [supplementary information S10d]. Furthermore, a closer inspection of the H3 normalized H3K36me3 distribution within the coding region revealed that H3K36me3 is more enriched over exons than introns [supplementary information S10e].

Next, ChIP-Seq of H3K36me3 was performed post introduction of GFP-SETD2 constructs in KO cells and their distribution was compared. On rescue of KO cells with SETD2 constructs, the H3K36me3 mark continued to be enriched within the gene bodies. Notably, this was true even for SETD2 N3ΔSRI and is consistent with the reports that association with RNA Pol II is not required for Set2 association with highly transcribed genes (discussed in the Discussion section). Also, the H3K36me3 level correlated with gene expression [supplementary information S10f]. However, a closer inspection of the distribution reveals that it is skewed towards the 5’ end of the genes in cells rescued with N3 as compared to FL [Figure 6c]. Furthermore, analysis of global H3K36me3 peaks revealed that there is more deposition of H3K36me3 at intergenic regions post-rescue with SETD2 N3 as compared to FL [Figure 6d]. In addition, there was an aberrant deposition of H3K36me3 on genes such as CLIC4 although no change in its expression was observed in our RNA-seq data [Figure 6e].

The ratio of average exon signal divided by the average intron signal was lower in N3 rescued cells than FL rescued cells [Figure 6f]. This suggests that on an increased accumulation of SETD2, the enrichment of H3K36me3 over exons decreases. The 5’ skewing and loss of enrichment over exons are

exemplified in genes such as ARID1A [Figure 6g]. Furthermore, analysis of the genes showing differential alternative splicing events in N3 as compared to FL display opposite trends of exon enrichment [supplementary information S10g]. This kind of opposite trend was not observed in FL expressing cells in which the exon enrichment values were ~1.5 [supplementary information S10g]. Also, the H3K36me3 levels were higher on the genes that showed increased splicing in N3 versus FL [supplementary information S10h].

Thus, in the absence of N-terminal rich IDRs, there is reduced initiation of UPS-mediated decay of SETD2. This results in its high abundance, reduced Pol II dependency and leads to global changes in the canonical distribution pattern of H3K36me3 [Figure 7].

DISCUSSION

In recent years many reports have highlighted the adverse effect of the loss of H3K36me3 that occurs due to SETD2 deletion. Here we show that the other end of the spectrum can also be detrimental as an excess of SETD2 might have inadvertent outcomes. In such a scenario, its IDRs play an important role in tuning and maintaining the requisite intracellular amount of the protein.

IDRs are common in proteins important in transcription pathways

Regulating protein half-life is critical for the wellbeing of cells. Altered protein half-life can lead to abnormal development and diseases such as cancer and neurodegeneration (Ciechanover, 2007). Considering the role of H3K36me3 in a variety of important cellular processes, it is reasonable that regulating the activity of the methyltransferase responsible for the deposition of this mark is important. Although the role of the proteasome in regulating SETD2 stability has been reported earlier, how SETD2 degradation occurs so robustly was not clear. Furthermore, it was not clear why the cellular half-life of SETD2 is so tightly regulated. We not only show that multiple regions of SETD2 are targeted by the proteasome, but also found that the disordered segments of SETD2 are important in this process. Importantly, our work clearly illustrates the importance of the N-terminal segment of SETD2 in governing its function, which has been a mystery. Strikingly, an analysis of the functional annotations of proteins with long disordered segments revealed enrichment for associations with regulatory and transcription functions (van der Lee *et al*, 2014). This suggests that the IDR-mediated regulation of protein half-life that

we discovered for SETD2 might be a prevalent mechanism employed by cells to govern essential processes.

SETD2 IDRs possibly act as efficient sites for initiating proteasomal degradation

Disordered segments act directly to regulate protein half-lives by forming initiation sites for degradation by the proteasome (van der Lee *et al*, 2014)(Zhao *et al*, 2010)(Prakash *et al*, 2004)(Verhoef *et al*, 2009). Consistent with this, we did not find any distinct segment of SETD2 that is degradation prone. We also did not see any apparent effect of the presence of D and KEN box motifs or PEST sequences on the half-life of SETD2. The catalytic residues are situated deep within the proteasome core particle and are only accessible through a long narrow channel. A terminal disordered segment of 30 residues or an internal disordered segment of at least 40 residues can span twice this distance and thus, could be cleaved by the core particle (van der Lee *et al*, 2014). Thus, proteins such as SETD2 that have 17 long disordered segments are expected to be processed quickly due to the efficient initiation of degradation. In addition to that, the disordered segments are spread across the length of the protein and co-operatively leads to the robust degradation of SETD2.

Variations in the N-terminal region might tune SETD2 half-life

As IDRs have a role in protein stability, variation in IDRs might be a mechanism for the divergence of half-life among orthologous proteins. IDRs are not required to attain a specified three-dimensional conformation in order to exert their functional effect and hence, can undergo mutations without greatly affecting their functionality. Such variation in disordered segments may provide an evolutionary mechanism for fine-tuning protein turnover rates. These forces might lead to inter- as well as intra-species divergence of protein half-life. Analysis of Drosophila SETD2 revealed large and abundant disordered regions [supplementary information S6e]. The differences in the degree of disorder between Drosophila and human SETD2 proteins might lead to differences in their half-lives much like what we observed between SETD2 and ySet2. Such differences in half-lives between homologs might be needed to adjust for the differences in the mechanism of deposition of H3K36 methylation. For instance, unlike in yeast where ySet2 performs all three states of methylation of H3K36, SETD2 does not appear to be majorly responsible for me1 and me2 deposition. Therefore, even with a shorter half-life, SETD2 might be able to

do the required H3K36me3 deposition. Some evidence for this is provided by studies on human cancers that show that total H3K36me3 levels are not significantly impacted by a monoallelic loss of SETD2 (Ho *et al*, 2016)(Roberti *et al*, 2016).

In addition to inter-species differences, intra-species variation in the length of IDRs may arise through mechanisms such as repeat expansion, alternative splicing, and alternative transcription start sites. Interestingly, in the ENSEMBL database, SETD2 has 9 splice variants out of which three are protein-coding. One of those with transcript ID ENST00000638947.1 codes for only a 591 residue segment of SETD2 that, importantly, contains the catalytic SET domain. This segment is trimmed off the IDRs present in full-length SETD2 and might be expected to accumulate to a higher level with inadvertent consequences as discussed below. Furthermore, differences in protein turnover may disturb protein abundance and could lead to disease (Ciechanover, 2007)(Yang *et al*, 2012). Strikingly, such missense and truncation mutations that can potentially alter half-life, are found in SETD2 in cancers [supplementary information S11].

SETD2 is aggregate-prone

We found that SETD2 is an inherently aggregate-prone protein. Although aggregation and intrinsic disorder are considered independent properties, a positive correlation has been observed between the two (Zhang *et al*, 2019)(Carvalho *et al*, 2013)(Uemura *et al*, 2018)(Kirilyuk *et al*, 2012). Overaccumulation of SETD2 presents cells with challenges because of the aggregate-prone nature of the protein. The insolubility of proteins leads to their sequestration in inclusion bodies and inactivation. Besides its own inactivation, aggregation of SETD2 might lead to the co-aggregation of other proteins leading to their inactivation and proteotoxic stress. Ubiquitinated aggregates like the ones formed by SETD2 can directly inhibit or clog proteasomes (Hipp *et al*, 2012)(Nonaka & Hasegawa, 2009). As aggregation is a concentration-dependent phenomenon, possibly the robust degradation mechanism in cells ensures that SETD2 is kept at levels that maintain its solubility and activity. This mechanism is often employed by cells for IDR-containing proteins (Babu, 2016). Interestingly, SETD2, aka, HYPB was initially identified in a screen to find interactors of the aggregate prone protein Huntingtin (Faber *et al*, 1998). It is possible that the reported Huntingtin-SETD2 interaction was due to the aggregation propensities of these proteins. In

fact, the N-terminus of SETD2 (SETD2 C4) behaves very similarly to the polyQ containing N-terminus of Huntingtin protein. Like mutant Htt, SETD2 C4 localizes to the nucleus in absence of an NLS and forms spontaneous puncta characteristic of aggregated proteins. The nucleoplasm promotes the formation of such aberrant and insoluble protein aggregates due to the strong crowding forces from highly concentrated macromolecules (approximately 100 mg/ml) (Mikecz, 2009). It will be interesting to investigate in the future whether the aggregate prone tendency of SETD2 enables it to form a part of RNA-granules and regulate transcription.

SETD2 over-abundance might have inadvertent consequences

Besides the challenge of dealing with undesired protein aggregation that a high level of SETD2 may pose to the cells, our data show that overexpression of SETD2 leads to widespread changes in the transcriptome which is consistent with the observed global increase in H3K36me3 and its altered distribution profile observed in metagene analysis. Strikingly almost 50 percent of the differentially expressed genes (551/1234) were long non-coding RNA (lncRNA) almost all of which are significantly upregulated (537/551). The list includes genes like NEAT1, MALAT1, MMP10, MMP12 and MMP13 that not only are functionally important but also of disease relevance (West *et al*, 2014)(Hirata *et al*, 2015)(Chen *et al*, 2017b)(Vaalamo *et al*, 1998). Furthermore, genes like the H3K9 demethylase KDM3A (FC 2.005, FDR 8.28e-14), the splicing factor U2AF1 (FC 2.009, FDR 1.6924e-5), 40 genes from the ZNF family, etc. were upregulated amongst others that have well established functional importance.

Besides affecting the transcriptome, the aberrant H3K36me3 deposition upon SETD2 overexpression can result in adverse consequences as the H3K36me3 mark is required for proper recruitment of important epigenetic regulators such as the PWWP domain-containing proteins DNMT3a, MutSα and MORF (Dhayalan *et al*, 2010)(Li *et al*, 2013)(Luco *et al*, 2010)(Huang *et al*, 2018). Aberrant H3K36me3 deposition might result in the mistargeting of such regulators. It's possible that the expression of SETD2 shorter splice variant with transcript ID ENST00000638947.1 will lead to similar perturbations in the H3K36me3 profile to what was observed upon expression of SETD2 N3. Also, the missense and truncation mutations found in SETD2 in cancers can potentially alter half-life, leading to aberrant H3K36me3 deposition [supplementary information S11]. We focused on the changes in the H3K36me3

mark upon SETD2 overexpression. However, SETD2 also has non-histone targets like tubulin, the methylation of which is important for metaphase transition (Park *et al*, 2016). Hence, the consequences of an increase in SETD2 abundance might not be limited to changes in histone methylation.

Pol II association is required for enhancing SETD2 activity but not for its recruitment

Our data shows that when the expression level of SETD2 is high, it has a reduced dependency on RNA Pol II for H3K36me3 deposition. In fact, from the ChiP-Seq data of H3K36me3, it appears that RNA Pol II association is not required for chromatin recruitment of SETD2 as previously assumed, but rather this interaction is required for the activation of SETD2 enzymatic activity. When present in its normal cellular amounts, SETD2 is not active enough to lead to robust H3K36me3 deposition without the RNA Pol II association. Pol II association enhances SETD2 activity and hence, although SETD2 FL protein is barely detectable, the rescue of H3K36me3 can be readily seen in the *setd2Δ* cells. At high abundance, despite its low activity, a robust H3K36me3 deposition is seen likely due to the increased copy number of SETD2 protein in cells. Recent studies in yeast have also challenged the notion that the Pol II association is required for chromatin recruitment of Set2 (Wang *et al*, 2015). The study found that the engagement of Set2 and Pol II through the SRI domain is rather required for the activation of Set2. In fact, the Set2ΔSRI ChiP profile shows an enrichment on the coding sequence of genes that is skewed towards the 5' end much like what we found for H3K36me3 when SETD2 N3ΔSRI is introduced in *setd2Δ* cells (Suzuki *et al*, 2016). Possibly, Set2/SETD2 can continue to engage with the transcription elongation complex even in the absence of its interaction with RNA Pol II. In future studies, it will be interesting to determine how the gene body enrichment of H3K36me3 and Set2/SETD2 occurs even without the Pol II association.

MATERIALS AND METHODS

Plasmids- SETD2-HaloTag® human ORF in pFN21A was procured from Promega. Deletion mutants of SETD2 were constructed by PCR (Phusion polymerase, NEB) using full-length SETD2 as a template and individual fragments were cloned. All constructs generated were confirmed by sequencing. SETD2-GFP, mRFP-Ub, pLenti puro HA-Ubiquitin, pTagRFP-C1-Fibrillarin and pCDNA3-ySet2 were procured from Addgene.

Cell line maintenance and drug treatment- The cell lines used in this study (HEK293T, HEPG2 and HELA) were procured from ATCC. Cells were maintained in DMEM supplemented with 10% FBS and 2 mM L-glutamine at 37 °C with 5% CO_2. MG132 (Sigma) was added at a final concentration of 10 µM for 12 hours. Chloroquine (Sigma) treatment was done as indicated in the text. Cycloheximide (Sigma) was added at a final concentration of 10 µM. Transfections were performed at cell confluency of 40% using Fugene HD (Promega) using a ratio of 1:4 of the plasmid (µg) to transfection reagent (µl).

Isolation of total RNA and PCR- Total RNA was extracted from cells as per the manufacturer's (Qiagen) instructions. It was further treated with DNaseI (NEB) for 30 min at 72 °C to degrade any possible DNA contamination. RNA (2 µg) was subjected to reverse transcription using the QScript cDNA synthesis mix according to the manufacturer's instructions. cDNAs were then amplified with the corresponding gene-specific primer sets. For RTPCR, PCR was conducted for 24 cycles using the condition of 30 s at 94 °C, 30 s at 60 °C and 30 s at 72 °C. The PCR products were analyzed on a 1% agarose gels containing 0.5 µg/ml ethidium bromide. The sequence of oligos is in supplementary information 12.

Histone isolation and immunoblot analysis- Histones were isolated and analyzed as described previously (Bhattacharya *et al*, 2017). For immunoblotting, histones were resolved on 15% SDS–polyacrylamide gel, transferred to PVDF membrane and probed with antibodies. Signals were detected by using the ECL plus detection kit (ThermoFisher).

Antibodies- H3K36me3 (CST, 4909S), H3K36me2 (Active Motif, 39255), H3K36me1 (Abcam, ab9048), H3 (CST, 9715S), Halo (Promega, G9211), β-actin (Abcam, ab8224), SETD2 (Abclonal, A3194), FLAG (Sigma-Aldrich, A8592) , Pol II (Abcam, ab5095).

Cell Fractionation- To prepare soluble and insoluble extracts, 293T cells were washed with 1xPBS, collected by centrifugation, and resuspended in lysis buffer (50 mM Tris, pH 7.5, 350 mM NaCl, 1% Triton-X 100, 0.1% Na-deoxycholate and a protease inhibitor mix). The lysed cells were centrifuged at 13,000 rpm for 20 min. The supernatant was collected as the soluble fraction. The pellet was washed with lysis buffer containing 600 mM NaCl). The remaining insoluble pellet following another centrifugation was resuspended in Laemmli buffer (Biorad) and solubilized by sonication on ice.

Affinity purification- 293T cells expressing the protein of interest were harvested in 1xPBS and collected by centrifugation. The cells were lysed by resuspending in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton-X 100, 0.1% Na-deoxycholate and a protease inhibitor cocktail). The lysed cells were centrifuged at 13,000 rpm for 20 min. The supernatant was collected and diluted 1:3 by adding dilution buffer (1x PBS, pH 7.5 with 1mM DTT and 0.005% NP-40). The diluted lysate was added to 100 µl of pre-equilibrated Magne® HaloTag® Beads (Promega, G7282) and incubated overnight on a rotator at 4 °C. The beads were then washed with wash buffer (50 mM Tris-HCL, pH 7.5, 300 mM NaCl, 0.005% NP40 and 1mM DTT. 2 µl AcTEV (ThermoFisher, 12575015) protease was used per 100 µl of elution buffer.

Immunofluorescence- 293T cells were plated onto glass coverslips in a 6-well plate. Cells were washed with 1x PBS and fixed in 4% paraformaldehyde for 20 min at 37 °C. Cells were then washed three times with cold 1x PBS and permeabilized for 5 mins with 1x PBS containing 0.2% Triton X-100. Permeabilized cells were then blocked for 30 min with blocking buffer (3% BSA and 0.1% Triton-X in 1x PBS). Cells were stained with primary antibodies against SETD2 (1:1,000; Abclonal) for 1 hr at room temperature. A secondary antibody conjugated with AlexaFluor 568 was applied for 1 hr at room temperature.

ChIP- Cells were cross-linked by 1% formaldehyde for 10 mins, and then quenched in 125 mM glycine for 5 mins. After washing with cold 1x PBS thrice, cells were harvested by scraping and pelleted down by centrifugation. The cell pellet was resuspended in swelling buffer (25 mM HEPES pH 8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP40, 1 mM DTT, protease inhibitor cocktail), kept in ice for 10 mins and then dounced. The nuclear pellet was obtained by centrifugation and resuspended in sonication buffer (50 mM HEPES pH 8, 140 mM NaCl, 1 mM EDTA, 1% Triton X 100, 0.1% Na-deoxycholate, 0.1% SDS, protease inhibitor cocktail), followed by sonication on ice for 12 cycles (30% amplitude, 10 secs on / 60 secs off) using a Branson Sonicator. For spike-in normalization, the spike-in chromatin and antibody were added in the reaction as per the manufacturer's recommendation (Active Motif). The chromatin was incubated with antibodies at 4 °C overnight and then added to 30 µl of protein G-Dyna beads (Thermo Fisher Scientific) for an additional 2 hours with constant rotation. The beads were extensively washed, and bound DNA was eluted with elution buffer (50 mM Tris-HCl pH 8, 5 mM EDTA, 50 mM NaCl, 1% SDS) and reverse-

crosslinked at 65 °C overnight. DNAs were purified using the QIAquick PCR purification kit (Qiagen) after the treatment of proteinase K and RNase A.

High throughput sequencing- Sequencing libraries were prepared using High Throughput Library Prep Kit (KAPA Biosystems) following the manufacturer's instructions. The library was sequenced on an Illumina HiSeq platform with paired reads of 75 bp for RNA-seq and single reads of 50 bp for ChIP-seq.

ChIP-seq analysis- Raw reads were demultiplexed into FASTQ format allowing up to one mismatch using Illumina bcl2fastq2 v2.18. Reads were aligned to the human genome (hg38) using Bowtie2 (version 2.3.4.1) with default parameters (Langmead & Salzberg, 2012). For samples with fly spike-in, reads were first mapped to the Drosophila melanogaster genome (dm6), and unmapped reads were then aligned to the human genome (hg38). Reads per million (RPM) normalized bigWig tracks were generated by extending reads to 150bp. For spike-in ChIP-seq data, we also generated spike-in normalized bigWig tracks (RPM normalization factor = 1E6 / number of reads aligned to hg38, and spike-in normalization factor = 1E6 / number of reads aligned to dm6).

Peak calling- epic2 (with options: -gn hg38 -fs 200 -fdr 0.05) was used to call wide peaks for H3K36me3 ChIP-seq data for FL, N3, and N3ΔSRI (Stovner *et al*, 2019). Next, ChIPseeker was applied (with options: genomicAnnotationPriority=c('Intergenic', '5UTR', '3UTR', 'Exon', 'Intron')) to obtain the genomic feature distribution (Ensembl 96 release) under peaks (Yu *et al*, 2015).

Metagene Plots- 14533 Protein-coding genes (Ensembl 96 release) were selected with length ≥ 600bp and no other genes within -2Kb TSS and +2Kb TES regions. Metagene regions were from -2Kb TSS to +2Kb TES. In addition, 2Kb upstream TSS and downstream TES regions are grouped into 100 bins (20bp per bin), respectively. The gene body region is grouped into 300 bins (at least 2bp per bin since the minimum gene length is 600bp). In total, each gene is grouped into 500 bins. The average normalized (RPM or spike-in) H3K36me3 signals in each bin were plotted using R package EnrichedHeatmap (Gu *et al*, 2018).

H3K36me3 on exons and introns- Protein-coding genes were selected, and the longest transcript for each gene was chosen. Also, we removed any overlapping transcripts (ignore strand). As a result, 15311

transcripts were used to calculate the H3K36me3 signal (RPM normalized) distribution on exons/introns. The average exon/intron signal is defined as the total H3K36me3 signals on all exons/introns divided by the total exon/intron length.

RNA-seq analysis- Raw reads were demultiplexed into FASTQ format allowing up to one mismatch using Illumina bcl2fastq2 v2.18. Reads were aligned to the human genome (hg38 and Ensembl 96 gene models) using STAR (version STAR_2.6.1c) (Dobin *et al*, 2013). TPM expression values were generated using RSEM (version v1.3.0) [5]. edgeR (version 3.24.3 with R 3.5.2) was applied to perform differential expression analysis, using only protein-coding and lncRNA genes (Robinson *et al*, 2009). To perform differential splicing analysis, we used rMATs (version 4.0.2) with default parameters starting from FASTQ files (Shen *et al*, 2014). FDR cutoff of 0.05 was used to determine statistical significance.

ACCESSION NUMBERS

The data sets are available in the Gene Expression Omnibus (GEO) database under the accession number GSE147752.

ACKNOWLEDGMENT

The authors are grateful to Dr. Kimryn Rathmell, Vanderbilt Institute for Infection, Immunology and Inflammation from providing SETD2 knock out 293T cells. The authors would like to thank the members of the Workman lab for their critical suggestions to improve the manuscript.

FUNDING

This work was supported by funding from the National Institute of General Medical Sciences (grant no. R35GM118068) and the Stowers Institute for Medical Research to Jerry L Workman.

AUTHOR INFORMATION

*Corresponding author: jlw@stowers.org

AUTHOR CONTRIBUTION

S.B. conceptualized the work, designed and performed the experiments. S.B. wrote the manuscript. N.Z. and H.L. analyzed the high-throughput sequencing data and made the related plots. J.L.W conceived the idea of the work, provided supervision, acquired funding and revised the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Babu MM (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44:** 1185–1200

Bhattacharya S, Reddy D, Jani V, Gadewal N, Shah S, Reddy R, Bose K, Sonavane U, Joshi R & Gupta S (2017) Histone isoform H2A1H promotes attainment of distinct physiological states by altering chromatin dynamics. *Epigenetics and Chromatin* **10:** 13072 Available at: https://doi.org/10.1186/s13072-018-0238-5

Carvalho S, Vítor AC, Sridhara SC, Filipa BM, Ana CR, Desterro JMP, Ferreira J & de Almeida SF (2014) SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *Elife* **2014:** 1–19

Carvalho SB, Botelho HM, Leal SS, Cardoso I, Fritz G & Gomes CM (2013) Intrinsically Disordered and Aggregation Prone Regions Underlie β-Aggregation in S100 Proteins. *PLoS One* **8:** 1–11

Chen K, Liu J, Liu S, Xia M, Zhang X, Han D, Jiang Y, Wang C & Cao X (2017a) Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. *Cell* **170:** 492-506.e14 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867417307614

Chen T, Wang H, Yang P & He ZY (2017b) Prognostic role of long noncoding RNA NEAT I in various carcinomas: A meta-analysis. *Onco. Targets. Ther.*

Chiang YC, Park IY, Terzo EA, Tripathi DN, Mason FM, Fahey CC, Karki M, Shuster CB, Sohn BH, Chowdhury P, Powell RT, Ohi R, Tsai YS, De Cubas AA, Khan A, Davis IJ, Strahl BD, Parker JS, Dere R, Walke CL, et al (2018) SETD2 haploinsufficiency for microtubule methylation is an early

driver of genomic instability in renal cell carcinoma. *Cancer Res.* **78:** 3135–3146

Ciechanover A (2007) Intracellular protein degradation from a vague idea through the lysosome and the ubiquitin-proteasome system and on to human diseases and drug targeting: Nobel lecture, December 8, 2004. *Ann. N. Y. Acad. Sci.* **1116:** 1–28

Cooper GM (2000) Protein Degradation. In *The Cell: A Molecular Approach* Sunderland, MA: Sinauer Associates

Cooper JK, Schilling G, Peters MF, Herring WJ, Sharp AH, Kaminsky Z, Masone J, Khan FA, Delanoy M, Borchelt DR, Dawson VL, Dawson TM & Ross CA (1998) Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum. Mol. Genet.* **7:** 783–790

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S & Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285:** 26114–26120

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21

Dronamraju R, Jha DK, Eser U, Adams AT, Dominguez D, Choudhury R, Chiang YC, Rathmell WK, Emanuele MJ, Churchman LS & Strahl BD (2018) Set2 methyltransferase facilitates cell cycle progression by maintaining transcriptional fidelity. *Nucleic Acids Res.* **46:** 1331–1344

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF & MacDonald ME (1998) Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* **7:** 1463–1474

Da Fonseca PCA, He J & Morris EP (2012) Molecular Model of the Human 26S Proteasome. *Mol. Cell* **46:** 54–66 Available at: http://dx.doi.org/10.1016/j.molcel.2012.03.026

Fuchs SM, Kizer KO, Braberg H, Krogan NJ & Strahl BD (2012) RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the set2 methyltransferase and histone H3 di- and trimethylation at lysine 36. *J. Biol. Chem.* **287:** 3249–3256

Gu Z, Eils R, Schlesner M & Ishaque N (2018) EnrichedHeatmap: An R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics*

Hacker KE, Fahey CC, Shinsky SA, Chiang Y-CCJCJ, DiFiore J V., Jha DK, Vo AH, Shavit JA, Davis IJ, Strahl BD & Rathmell WK (2016) Structure/Function Analysis of Recurrent Mutations in SETD2 Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and H3K36 Trimethylation. *J. Biol. Chem.* **291:** jbc.M116.739375 Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M116.739375

Hipp MS, Patel CN, Bersuker K, Riley BE, Kaiser SE, Shaler TA, Brandeis M & Kopito RR (2012) Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease. *J. Cell Biol.*

Hirata H, Hinoda Y, Shahryari V, Deng G, Nakajima K, Tabatabai ZL, Ishii N & Dahiya R (2015) Long noncoding RNA MALAT1 promotes aggressive renal cell carcinoma through Ezh2 and interacts with miR-205. *Cancer Res.*

Ho TH, Park IY, Zhao H, Tong P, Champion MD, Yan H, Monzon FA, Hoang A, Tamboli P, Parker AS, Joseph RW, Qiao W, Dykema K, Tannir NM, Castle EP, Nunez-Nateras R, Teh BT, Wang J, Walker CL, Hung M-C, et al (2016) High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. *Oncogene* **35:** 1565–1574 Available at: http://www.nature.com/doifinder/10.1038/onc.2015.221

Hou P, Huang C, Liu C-P, Yang N, Yu T, Yin Y, Zhu B & Xu R-M (2019) Structural Insights into Stimulation of Ash1L's H3K36 Methyltransferase Activity through Mrg15 Binding. *Structure* **27:** 837-845.e3 Available at: http://www.sciencedirect.com/science/article/pii/S0969212619300152

Huang Y, Gu L & Li GM (2018) H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J. Biol. Chem.* **293:** 7811–7823

Kirilyuk A, Shimoji M, Catania J, Sahu G, Pattabiraman N, Giordano A, Albanese C, Mocchetti I, Toretsky JA, Uversky VN & Avantaggiati ML (2012) An Intrinsically Disordered Region of the Acetyltransferase p300 with Similarity to Prion-Like Domains Plays a Role in Aggregation. *PLoS One* **7:**

Kolasinska-zwierz P, Down T, Latorre I, Liu T, Liu XS & Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41:** 376–381

Kosugi S, Hasebe M, Tomita M & Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 10171–10176

Kuo AJ, Cheung P, Chen K, Zee BM, Kioi M, Lauring J, Xi Y, Park BH, Shi X, Garcia BA, Li W & Gozani O (2011) NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Mol. Cell* **44:** 609–620 Available at: http://dx.doi.org/10.1016/j.molcel.2011.08.042

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*

Lasker K, Forster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A & Baumeister W (2012a) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci.* **109:** 1380–1387

Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A & Baumeister W (2012b) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 1380–1387

van der Lee R, Lang B, Kruse K, Gsponer J, de Groot NS, Huynen MA, Matouschek A, Fuxreiter M & Babu MM (2014) Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.* **8:** 1832–1844

Li F, Mao G, Tong D, Huang J, Gu L, Yang W & Li GM (2013) The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutS?? *Cell* **153:** 590–600 Available at: http://dx.doi.org/10.1016/j.cell.2013.03.025

Li J, Kluiver J, Osinga J, Westers H, van Werkhoven MB, Seelen MA, Sijmons RH, van den Berg A & Kok K (2016) Functional Studies on Primary Tubular Epithelial Cells Indicate a Tumor Suppressor Role of SETD2 in Clear Cell Renal Cell Carcinoma. *Neoplasia* **18:** 339–46 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27292023%5Cnhttp://www.pubmedcentral.nih.gov/articlerende

r.fcgi?artid=PMC4916950

Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC & Carpenter PB (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl. Acad. Sci.* **107:** 16952–16957

Luco RRF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM & Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* **327:** 996–1000 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2913848&tool=pmcentrez&rendertype=abstract

McDaniel SL & Strahl BD (2017) Shaping the cellular landscape with Set2/SETD2 methylation. *Cell. Mol. Life Sci.* **0:** 1–18 Available at: http://link.springer.com/10.1007/s00018-017-2517-x

Mészáros B, Erdös G & Dosztányi Z (2018) IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46:** W329–W337

Mikecz A Von (2009) Protein Aggregation in the Cell Nucleus□: Structure , Function and Topology. *Open Biol. J.* **2:** 193–199

Nonaka T & Hasegawa M (2009) A cellular model to monitor proteasome dysfunction by α-synuclein. *Biochemistry*

Park IY, Powell RT, Tripathi DN, Dere R, Ho TH, Blasius TL, Chiang YC, Davis IJ, Fahey CC, Hacker KE, Verhey KJ, Bedford MT, Jonasch E, Rathmell WK & Walker CL (2016) Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell* **166:** 950–962 Available at: http://dx.doi.org/10.1016/j.cell.2016.07.005

Pfister SX, Ahrabi S, Zalmas LP, Sarkar S, Aymard F, Bachrati CZ, Helleday T, Legube G, LaThangue NB, Porter ACGG & Humphrey TC (2014) SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Rep.* **7:** 2006–2018

Prakash S, Tian L, Ratliff KS, Lehotzky RE & Matouschek A (2004) An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11:** 830–837

Qiao Q, Li Y, Chen Z, Wang M, Reinberg D & Xu RM (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* **286:** 8361–8368

Qin J, Robine S, Louvard D, Huelsken J, Kelley R, Salzberg S, Liu Y, Qiu T, Jiang M, Pan Q, Han Y, Wang X, Li Q & Qin J (2017) Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. *Nature* **493:** 236–240 Available at: https://www.jci.org/articles/view/94292

Roberti A, Dobay MP, Bisig B, Vallois D, Boéchat C, Lanitis E, Bouchindhomme B, Parrens M-C, Bossard C, Quintanilla-Martinez L, Missiaglia E, Gaulard P & De Leval L (2016) Type II enteropathy-associated T-cell lymphoma features a unique genomic profile with highly recurrent SETD2 alterations. *Nat. Commun.* **7:** 12602 Available at: http://www.nature.com/doifinder/10.1038/ncomms12602

Robinson MD, McCarthy DJ & Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*

Sano M, Tokudome S, Shimizu N, Yoshikawa N, Ogawa C, Shirakawa K, Endo J, Katayama T, Yuasa S, Ieda M, Makino S, Hattori F, Tanaka H & Fukuda K (2007) Intramolecular control of protein stability, subnuclear compartmentalization, and coactivator function of peroxisome proliferator-activated receptor γ coactivator 1α. *J. Biol. Chem.* **282:** 25970–25980

Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q & Xing Y (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.*

Stovner EB, Sætrom P & Hancock J (2019) Epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics*

Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF & Allis CD (2002) Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Mol. Cell. Biol.* **22:** 1298–1306 Available at: http://mcb.asm.org/content/22/5/1298.abstract?ijkey=b692685301336f146af9f090f053cbc71e5aa6f9

&keytype2=tf_ipsecsha

Su X, Zhang J, Mouawad R, Compérat E, Rouprêt M, Allanic F, Parra J, Bitker MO, Thompson EJ, Gowrishankar B, Houldsworth J, Weinstein JN, Tost J, Broom BM, Khayat D, Spano JP, Tannir NM & Malouf GG (2017) NSD1 inactivation and SETD2 mutation drive a convergence toward loss of function of H3K36 writers in clear cell renal cell carcinomas. *Cancer Res.* **77:** 4835–4845

Suzuki S, Kato H, Suzuki Y, Chikashige Y, Hiraoka Y, Kimura H, Nagao K, Obuse C, Takahata S & Murakami Y (2016) Histone H3K36 trimethylation is essential for multiple silencing mechanisms in fission yeast. *Nucleic Acids Res.* **44:** 4147–4162

Tanaka Y, Katagiri Z ichiro, Kawahashi K, Kioussis D & Kitajima S (2007) Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397:** 161–168

Tompa P, Prilusky J, Silman I & Sussman JL (2008) Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins Struct. Funct. Bioinforma.* **71:** 903–909 Available at: https://doi.org/10.1002/prot.21773

Uemura E, Niwa T, Minami S, Takemoto K, Fukuchi S, Machida K, Imataka H, Ueda T, Ota M & Taguchi H (2018) Large-scale aggregation analysis of eukaryotic proteins reveals an involvement of intrinsically disordered regions in protein folding. *Sci. Rep.* **8:** 1–11 Available at: http://dx.doi.org/10.1038/s41598-017-18977-5

Vaalamo M, Karjalainen-Lindsberg ML, Puolakkainen P, Kere J & Saarialho-Kere U (1998) Distinct expression profiles of stromelysin-2 (MMP-10), collagenase-3 (MMP-13), macrophage metalloelastase (MMP-12), and tissue inhibitor of metalloproteinases-3 (TIMP-3) in intestinal ulcerations. *Am. J. Pathol.*

Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K & Workman JL (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489:** 452–455 Available at: http://dx.doi.org/10.1038/nature11326

Verhoef LGGC, Heinen C, Selivanova A, Halff EF, Salomons FA & Dantuma NP (2009) Minimal length

requirement for proteasomal degradation of ubiquitin-dependent substrates. *FASEB J.* **23:** 123–133

Wang Y, Niu Y & Li B (2015) Balancing acts of SRI and an auto-inhibitory domain specify Set2 function at transcribed chromatin. *Nucleic Acids Res.* **43:** 4881–4892

West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY & Kingston RE (2014) The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites. *Mol. Cell* **55:** 791–802 Available at: http://dx.doi.org/10.1016/j.molcel.2014.07.012

Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH & Strahl BD (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* **17:** 654–663

Yang C, Matro JC, Huntoon KM, Ye DY, Huynh TT, Fliedner SMJ, Breza J, Zhuang Z & Pacak K (2012) Missense mutations in the human SDHB gene increase protein degradation without altering intrinsic enzymatic function. *FASEB J.* **26:** 4506–4516

Yu G, Wang LG & He QY (2015) ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*

Yuan W, Xu M, Huang C, Liu N, Chen S & Zhu B (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286:** 7983–7989

Zhang J, Han C, Song K, Chen W, Ungerleider N, Yao L, Ma W & Wu T (2020) The long-noncoding RNA MALAT1 regulates TGF- β / Smad signaling through formation of a lncRNA-protein complex with Smads , SETD2 and PPM1A in hepatic cells. **:** 1–16

Zhang W, Zhang J, MacRaild CA, Norton RS, Anders RF & Zhang X (2019) Modulation of the aggregation of an amyloidogenic sequence by flanking-disordered region in the intrinsically disordered antigen merozoite surface protein 2. *Eur. Biophys. J.* **48:** 99–110 Available at: https://doi.org/10.1007/s00249-018-1337-8

Zhao M, Zhang NY, Zurawel A, Hansen KC & Liu CW (2010) Degradation of some polyubiquitinated proteins requires an intrinsic proteasomal binding element in the substrates. *J. Biol. Chem.* **285:** 4771–4780

Zhu K, Lei P-JJ, Ju L-GG, Wang X, Huang K, Yang B, Shao C, Zhu Y, Wei G, Fu X-DD, Li L & Wu M (2016) SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. *Nucleic Acids Res.* **45:** 1–14 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27614073

FIGURE LEGENDS

Figure 1: Removal of the SETD2 N-terminus leads to its increased abundance and changes in the transcriptome. (a) Cartoon illustrating the length and domains of ySet2/SETD2. (b) Volcano plot showing the expression changes occurring on SETD2 N3 expression as compared to SETD2 FL. Each dot is a gene. RNA was isolated 72 hours post-transfection. (c) RMATS plot depicting the differential alternative splicing events in SETD2 N3 rescued KO cells as compared to SETD2 FL rescued ones. The cartoon was taken from http://rnaseq-mats.sourceforge.net/ (d, e) Western blot of whole-cell lysates probed with the depicted antibodies. setd2Δ 293T (KO) cells were transfected with Halo-vector control (VC), SETD2 full-length (FL) or SETD2 N3 (N3) and lysates were prepared 72 hrs post-transfection. * denotes non-specific band. The expected bands for the target are depicted by arrows. (f) Microscopy images showing the effect of MG132 treatment on expression of GFP-SETD2 FL in 293T cells. (g) RNA was isolated from transfected cells described in (f), supplementary information S3c and RT-PCR was performed to check transcript levels. GAPDH was used as a normalization control. (h) Whole-cell lysates of 293T cells were prepared after 12 hrs of MG132 (10 µM) treatment and probed with the antibodies depicted.

Figure 2: Deleting SETD2 fragments leads to its stabilization. (a, b, e) Cartoon illustrating the truncations of SETD2 along with the known domains. (c, d) Western blot of whole-cell lysates probed with the depicted antibodies. Lysates of 293T cells expressing Halo-SETD2 constructs were prepared after 12 hrs of MG132 (10 µM) treatment. The expected band for the target is depicted by arrows. (f) Western blot of whole-cell lysates of 293T cells expressing Halo-SETD2 constructs probed with the depicted antibodies. (g, h, i) Western blot of whole-cell lysates probed with the depicted antibodies. Lysates of 293T cells expressing Halo-SETD2 constructs were prepared after cycloheximide (10 µM) treatment. The duration of the treatment is shown in the figure. HE- High Exposure, LE-Low Exposure.

Figure 3: The SETD2 N-terminus can destabilize ySet2. (a, b, d) Cartoon illustrating the chimeric ySet2 constructs made by fusing with the segments of N-terminus of SETD2. (c, e) Western blot of whole-cell

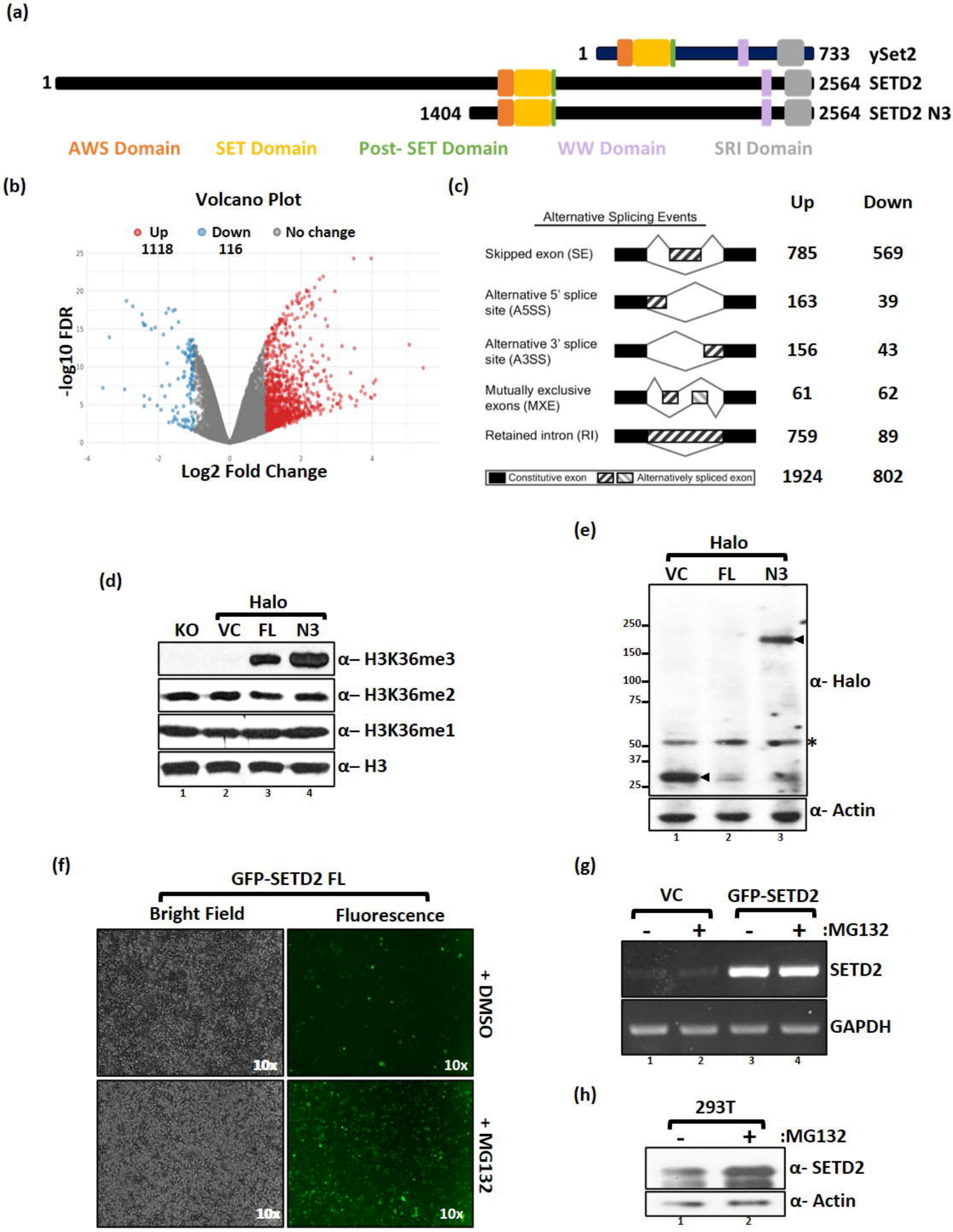
lysates probed with the depicted antibodies. Lysates of 293T cells expressing Halo-ySet2 constructs were prepared after 12 hrs of MG132 (10 µM) treatment. The expected band for the target is depicted by arrows. (f) Western blot of whole-cell lysates of yeast BY4741 expressing FLAG-ySet2 constructs (using pRS416) probed with the depicted antibodies. HE- High Exposure, LE-Low Exposure, *- non-specific.

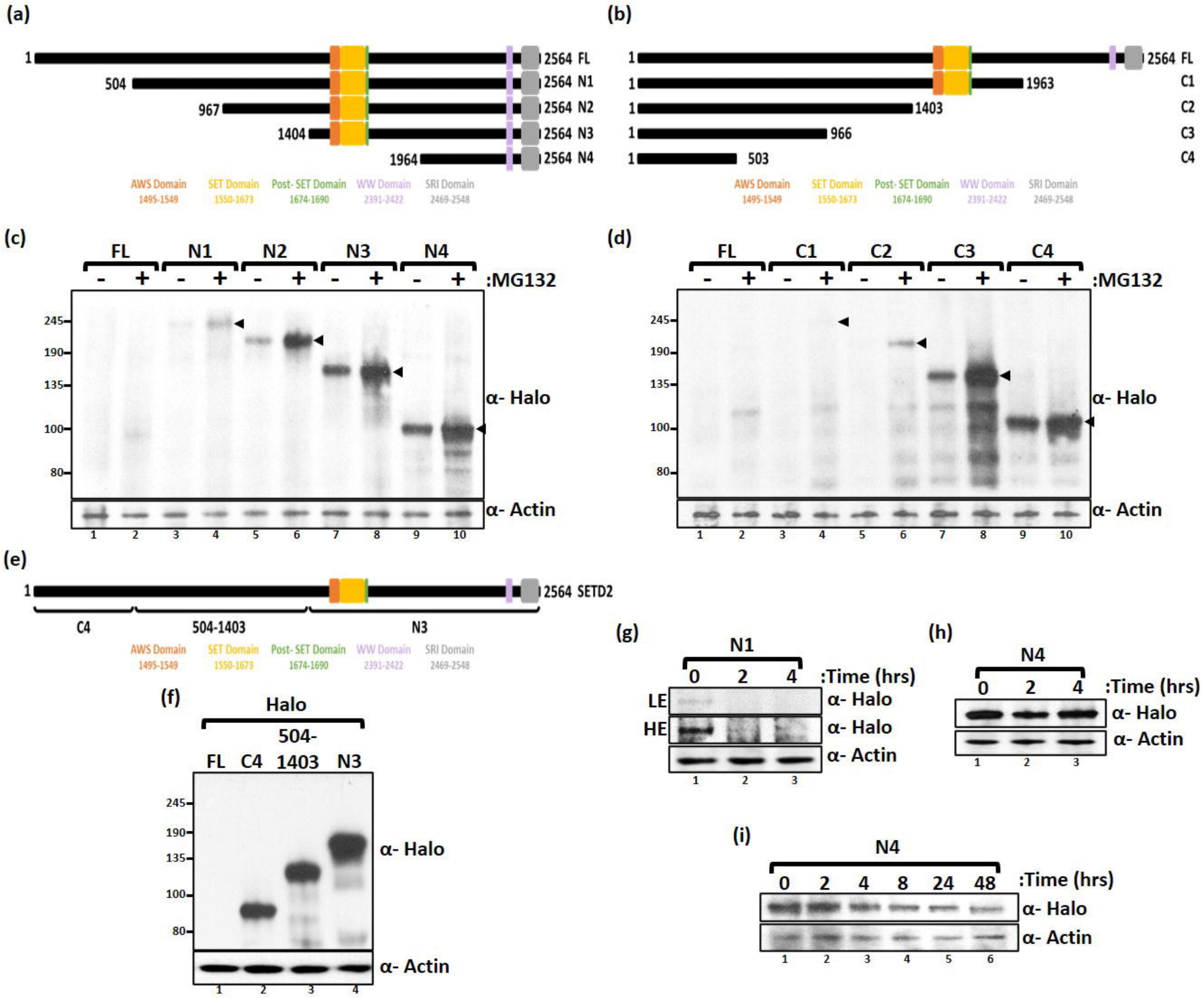
Figure 4: SETD2 has three nuclear localization signals. (a) Microscopy images showing localization of GFP-SETD2 fragments that have putative NLS and the NLS mutants. See the text for more details. (b) Microscopy images showing the localization of GFP-SETD2 truncations. The scale bar is 10 um.

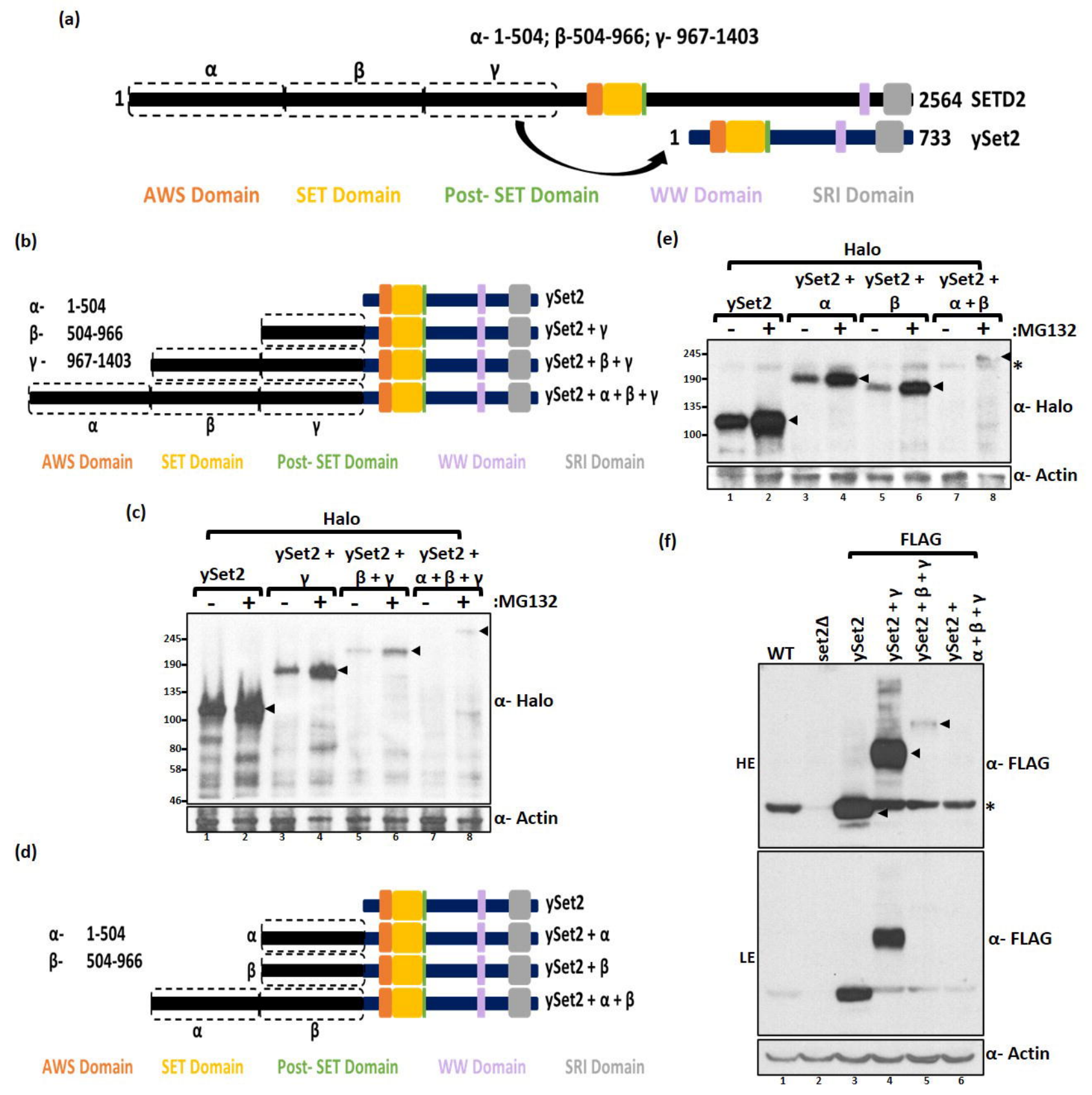
Figure 5: SETD2 forms ubiquitinated aggregates. (a) Microscopy images showing colocalization of GFP-SETD2 C4 fragment with RFP-Fibrillarin or RFP-Ub. The scale bar is 10 um. (b) Halo purification was performed from extracts of 293T cells co-expressing Halo-N3 and HA-Ub. Eluted and 0.5% of Input samples were resolved on a gel and probed with the depicted antibodies. (c) Soluble and insoluble fractions were separated of 293T cells expressing Halo-SETD2 fragments, resolved on a gel and probed with the depicted antibodies. *- non-specific.

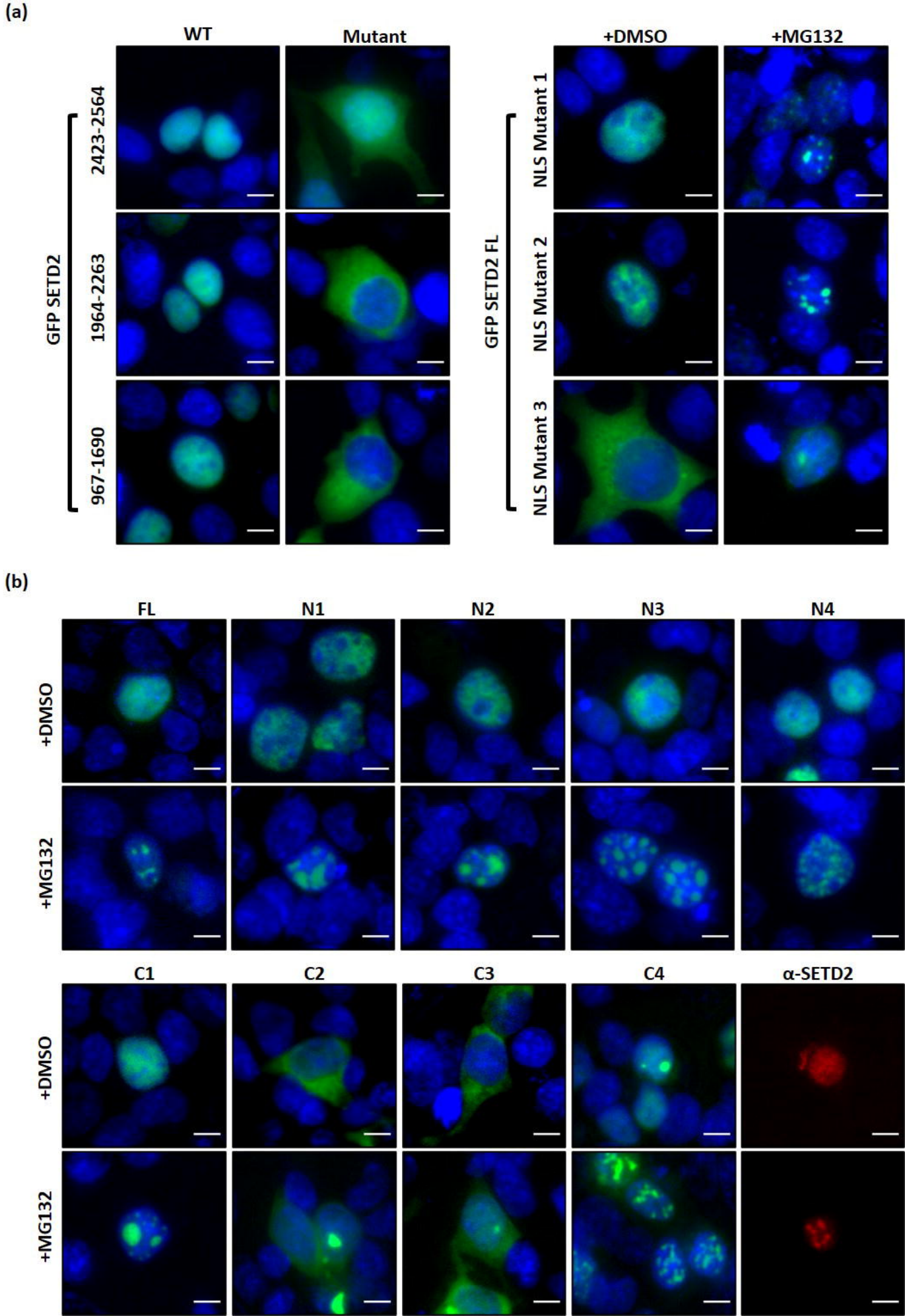
Figure 6: Increased SETD2 abundance leads to a non-canonical H3K36me3 deposition. (a) Western blot of whole-cell lysates of 293T cells expressing Halo-SETD2 constructs, under the control of either CMV or CMVD2, probed with the depicted antibodies (b) Halo purification was performed from extracts of 293T cells expressing Halo-N3. Input and eluted samples were resolved on a gel and probed with the depicted antibodies. (c, d) Metagene plot and histogram depicting the distribution of H3K36me3. (e, g) Genome browser view depicting the peaks of H3K36me3. The blue solid blocks on the X-axis denotes the exons. (f) Bar graph showing the H3K36me3 enrichment over exons as compared to introns. *- non-specific.

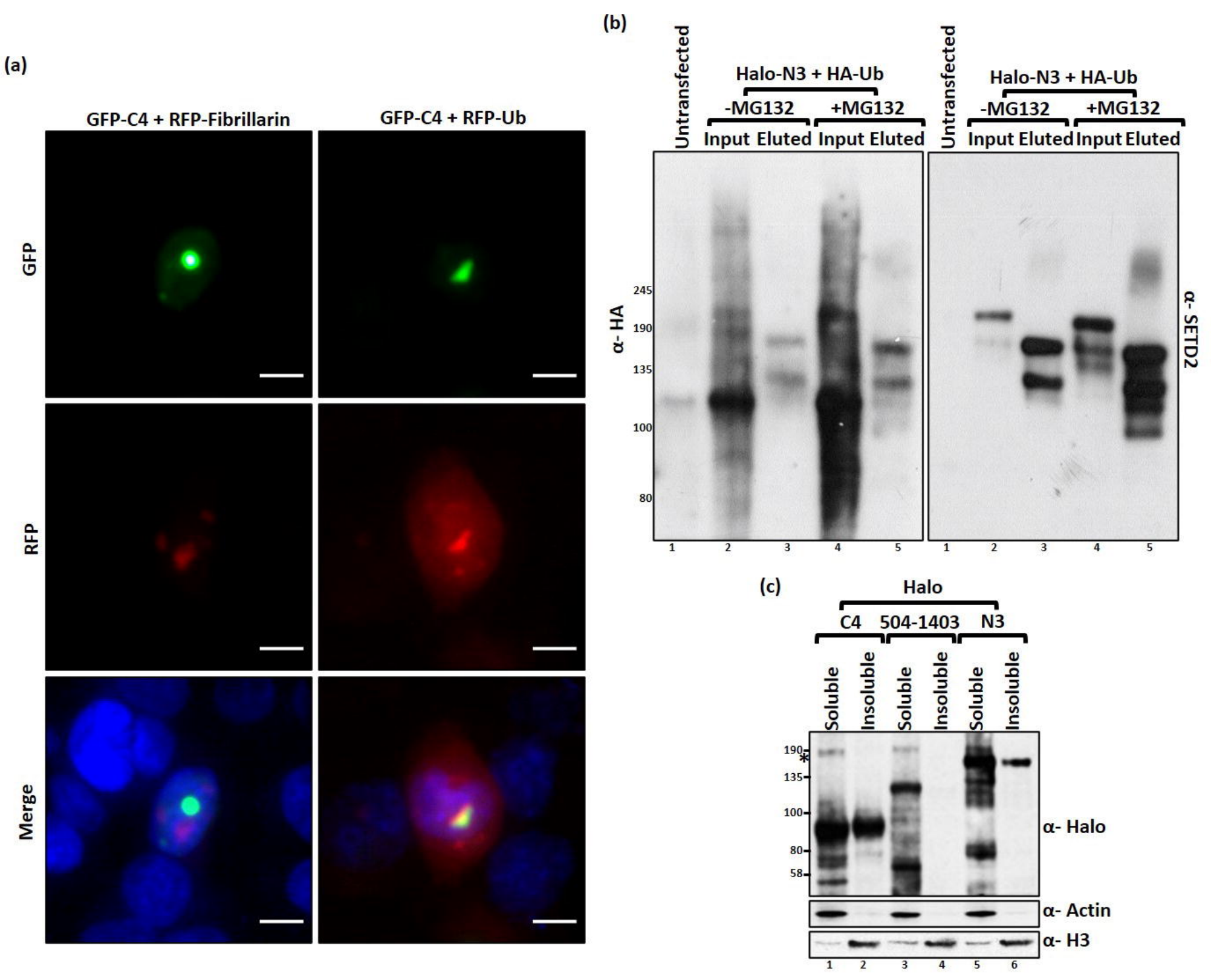
Figure 7: Removal of the N-terminal region from SETD2 stabilizes the protein and results in a non-canonical H3K36me3 distribution.

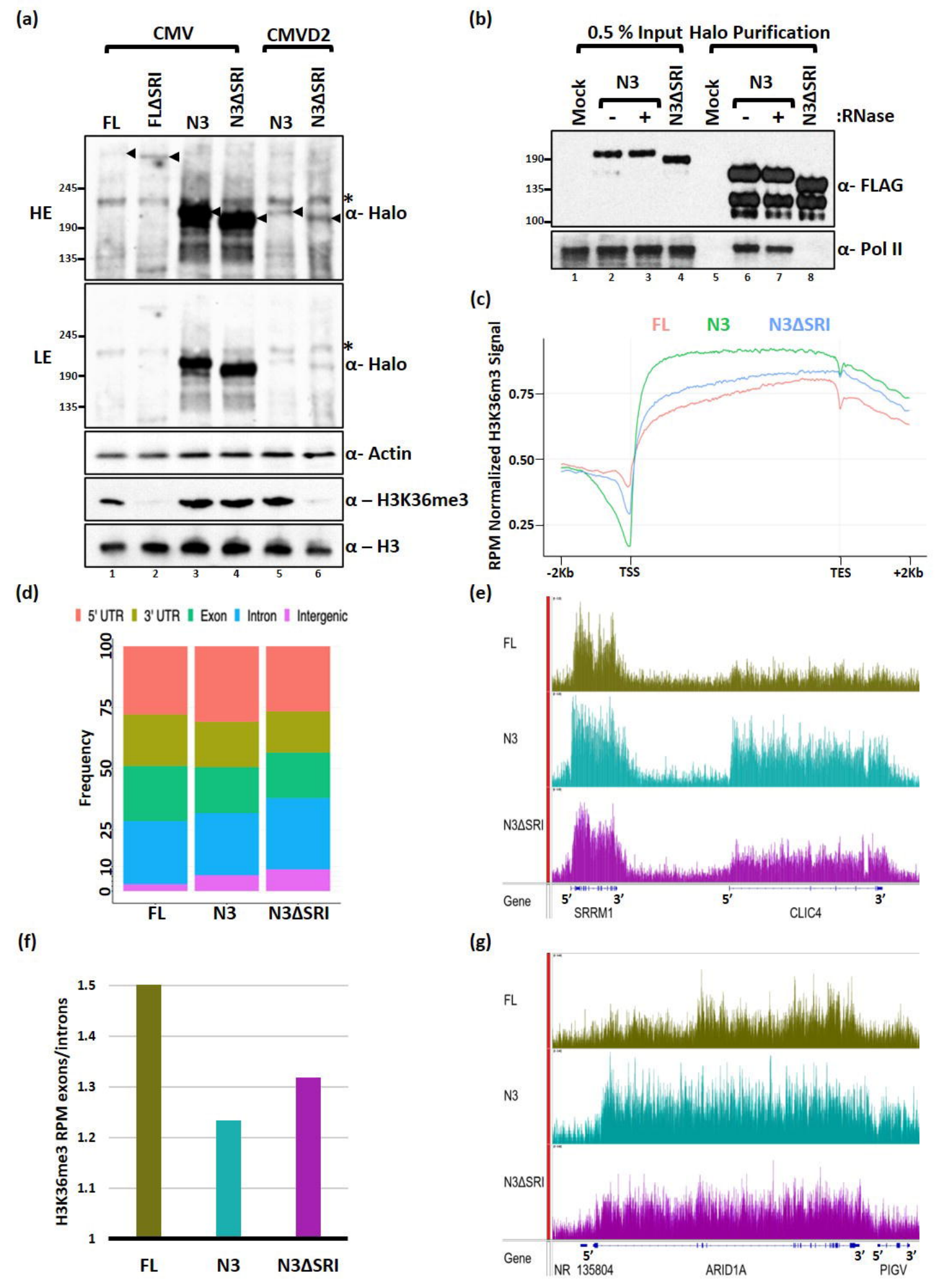

(a)
1
733 ySet2
1
2564 SETD2
1404
2564 SETD2 N3
AWS Domain
SET Domain
Post- SET Domain
WW Domain
SRI Domain
(b)
Volcano Plot
Up 1118
Down 116
No change
-log10 FDR
Log2 Fold Change
(c)
Up
Down
Alternative Splicing Events
Skipped exon (SE) 785 569
Alternative 5' splice site (A5SS) 163 39
Alternative 3' splice site (A3SS) 156 43
Mutually exclusive exons (MXE) 61 62
Retained intron (RI) 759 89
Constitutive exon
Alternatively spliced exon 1924 802
(d)
Halo
KO VC FL N3
α– H3K36me3
α– H3K36me2
α– H3K36me1
α– H3
1 2 3 4
(e)
Halo
VC FL N3
250
150
100
75
50
37
25
α- Halo
*
α- Actin
1 2 3
(f)
GFP-SETD2 FL
Bright Field
Fluorescence
+ DMSO
+ MG132
10x
(g)
VC
GFP-SETD2
- + - + :MG132
SETD2
GAPDH
1 2 3 4
(h)
293T
- + :MG132
α- SETD2
α- Actin
1 2

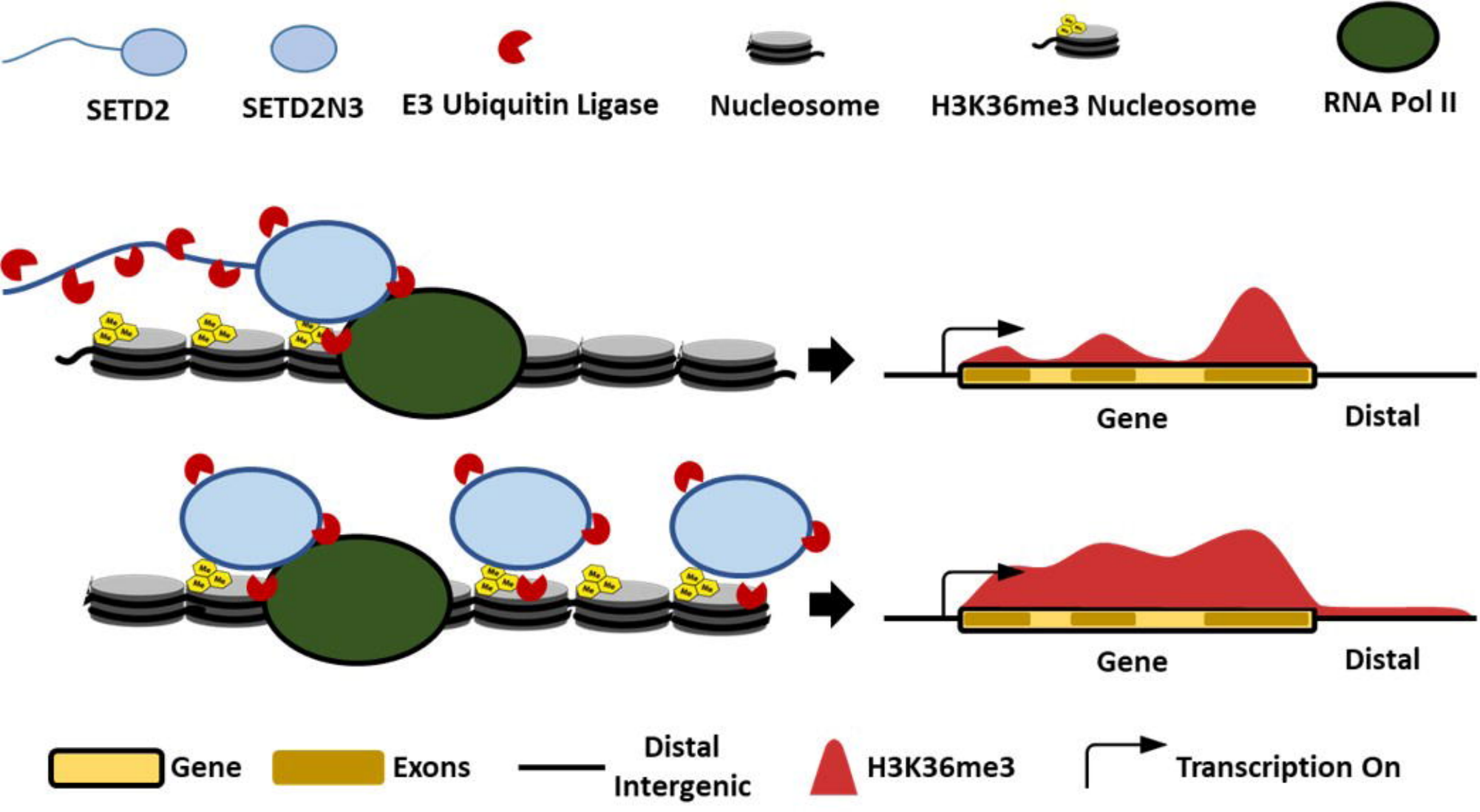

(a)
1
2564 FL
504
2564 N1
967
2564 N2
1404
2564 N3
1964
2564 N4
AWS Domain 1495-1549
SET Domain 1550-1673
Post- SET Domain 1674-1690
WW Domain 2391-2422
SRI Domain 2469-2548
(b)
1
2564 FL
1963
C1
1403
C2
966
C3
503
C4
(c)
FL N1 N2 N3 N4
- + - + - + - + - + :MG132
245
190
135
100
80
α- Halo
α- Actin
1 2 3 4 5 6 7 8 9 10
(d)
FL C1 C2 C3 C4
- + - + - + - + - + :MG132
245
190
135
100
80
α- Halo
α- Actin
1 2 3 4 5 6 7 8 9 10
(e)
1
2564 SETD2
C4
504-1403
N3
AWS Domain 1495-1549
SET Domain 1550-1673
Post- SET Domain 1674-1690
WW Domain 2391-2422
SRI Domain 2469-2548
(f)
Halo
FL C4 504-1403 N3
245
190
135
100
80
α- Halo
α- Actin
1 2 3 4
(g)
N1
0 2 4 :Time (hrs)
LE
HE
α- Halo
α- Halo
α- Actin
1 2 3
(h)
N4
0 2 4 :Time (hrs)
α- Halo
α- Actin
1 2 3
(i)
N4
0 2 4 8 24 48 :Time (hrs)
α- Halo
α- Actin
1 2 3 4 5 6


(a)
α- 1-504; β-504-966; γ- 967-1403
α
β
γ
1
2564 SETD2
1
733 ySet2
AWS Domain
SET Domain
Post- SET Domain
WW Domain
SRI Domain
(b)
α- 1-504
β- 504-966
γ - 967-1403
ySet2
ySet2 + γ
ySet2 + β + γ
ySet2 + α + β + γ
α
β
γ
AWS Domain
SET Domain
Post- SET Domain
WW Domain
SRI Domain
(c)
Halo
ySet2
ySet2 + γ
ySet2 + β + γ
ySet2 + α + β + γ
- + - + - + - +
:MG132
245
190
135
100
80
58
46
α- Halo
α- Actin
1 2 3 4 5 6 7 8
(d)
α- 1-504
β- 504-966
ySet2
ySet2 + α
ySet2 + β
ySet2 + α + β
α
β
α
β
AWS Domain
SET Domain
Post- SET Domain
WW Domain
SRI Domain
(e)
Halo
ySet2
ySet2 + α
ySet2 + β
ySet2 + α + β
- + - + - + - +
:MG132
245
190
135
100
*
α- Halo
α- Actin
1 2 3 4 5 6 7 8
(f)
FLAG
WT
set2Δ
ySet2
ySet2 + γ
ySet2 + β + γ
ySet2 + α + β + γ
HE
LE
α- FLAG
α- FLAG
*
α- Actin
1 2 3 4 5 6


(a)
WT
Mutant
+DMSO
+MG132
GFP SETD2
2423-2564
1964-2263
967-1690
GFP SETD2 FL
NLS Mutant 1
NLS Mutant 2
NLS Mutant 3
(b)
FL
N1
N2
N3
N4
+DMSO
+MG132
C1
C2
C3
C4
α-SETD2
+DMSO
+MG132


(a)
GFP-C4 + RFP-Fibrillarin
GFP-C4 + RFP-Ub
GFP
RFP
Merge
(b)
Untransfected
Halo-N3 + HA-Ub
-MG132
+MG132
Input
Eluted
α- HA
245
190
135
100
80
α- SETD2
1
2
3
4
5
(c)
Halo
C4
504-1403
N3
Soluble
Insoluble
190
135
100
80
58
*
α- Halo
α- Actin
α- H3
1
2
3
4
5
6


(a)
CMV
CMVD2
FL
FLΔSRI
N3
N3ΔSRI
N3
N3ΔSRI
HE
LE
245
190
135
α- Halo
α- Actin
α – H3K36me3
α – H3
1 2 3 4 5 6
(b)
0.5 % Input Halo Purification
Mock
N3
N3ΔSRI
Mock
N3
N3ΔSRI
:RNase
190
135
100
α- FLAG
α- Pol II
1 2 3 4 5 6 7 8
(c)
FL
N3
N3ΔSRI
RPM Normalized H3K36m3 Signal
0.75
0.50
0.25
-2Kb
TSS
TES
+2Kb
(d)
5' UTR
3' UTR
Exon
Intron
Intergenic
Frequency
100
75
50
25
10
0
FL
N3
N3ΔSRI
(e)
FL
N3
N3ΔSRI
Gene
5'
3'
SRRM1
CLIC4
(f)
H3K36me3 RPM exons/introns
1.5
1.4
1.3
1.2
1.1
1
FL
N3
N3ΔSRI
(g)
FL
N3
N3ΔSRI
Gene
NR 135804
ARID1A
PIGV


SETD2
SETD2N3
E3 Ubiquitin Ligase
Nucleosome
H3K36me3 Nucleosome
RNA Pol II
Me
Gene
Distal
Gene
Distal
Gene
Exons
Distal Intergenic
H3K36me3
Transcription On