

# Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs

David Zünd<sup>1,2</sup>, Andreas R Gruber<sup>3,4</sup>, Mihaela Zavolan<sup>3,4</sup> & Oliver Mühlemann<sup>1</sup>

**Recruitment of the UPF1 nonsense-mediated mRNA decay (NMD) factor to target mRNAs was initially proposed to occur through interaction with release factors at terminating ribosomes. However, recently emerging evidence points toward translation-independent interaction with the 3' untranslated region (UTR) of mRNAs. We mapped transcriptome-wide UPF1-binding sites by individual-nucleotide-resolution UV cross-linking and immunoprecipitation in human cells and found that UPF1 preferentially associated with 3' UTRs in translationally active cells but underwent significant redistribution toward coding regions (CDS) upon translation inhibition, thus indicating that UPF1 binds RNA before translation and gets displaced from the CDS by translating ribosomes. Corroborated by RNA immunoprecipitation and by UPF1 cross-linking to long noncoding RNAs, our evidence for translation-independent UPF1-RNA interaction suggests that the triggering of NMD occurs after UPF1 binding to mRNA, presumably through activation of RNA-bound UPF1 by aberrant translation termination.**

The RNA helicase UPF1 (for up-frameshift; also called RENT1) is a multitasking enzyme implicated in DNA-damage response<sup>1</sup>, telomere maintenance<sup>2</sup>, Staufen-1-mediated mRNA decay<sup>3</sup> and histone-mRNA turnover<sup>4</sup>, but it is best known for its crucial role in NMD (reviewed in refs. 5,6). UPF1 is essential in plants<sup>7</sup>, mammals<sup>8</sup>, zebrafish<sup>9</sup> and fruit flies<sup>10</sup>, and RNA interference-mediated depletion of UPF1 in human cells alters the levels of about 10% of all mRNAs, with most of those being upregulated. Although this suggests that UPF1-bound mRNAs are generally marked for degradation<sup>11–14</sup>, distinguishing the direct effects of UPF1 binding from indirect effects has so far not been possible.

UPF1 belongs to the SF1 superfamily of helicases and contains the typical helicase core formed by two RecA-like domains oriented face to face, with the ATP-binding site located in the cleft in between<sup>15–18</sup>. *In vitro*, UPF1 binds RNA and DNA and has the capacity to unwind double-stranded nucleic acids in a 5'-to-3' direction<sup>19–21</sup>. The helicase domain of UPF1 is flanked by a conserved N-terminal cysteine- and histidine-rich (CH) domain and a C-terminal serine and glutamine (SQ)-rich region conserved only in higher eukaryotes; both domains independently inhibit UPF1's helicase and ATPase activities<sup>16,20,22</sup>. These recent studies suggest that in the absence of interacting proteins, the CH and SQ domains interact with the helicase core domain, thereby locking UPF1 on the RNA<sup>16,22</sup>. UPF1 helicase activity, which is essential for NMD<sup>17,19,23</sup>, thus requires abrogation of the inhibitory function of both the N- and C-terminal domains simultaneously<sup>22</sup>. Whereas the factors responsible for relieving the SQ domain-mediated inhibition are unknown, binding of the NMD cofactor UPF2 to UPF1 has been shown to move the CH domain to the opposite side of the helicase core domain, causing a large conformational change

that alleviates the inhibitory effect of the CH domain and stimulates ATP hydrolysis and helicase activity<sup>16,20</sup>.

NMD targets for rapid degradation mRNAs on which the stop codon is located in an unfavorable environment for proper translation termination<sup>5,6</sup>. NMD is thereby an efficient quality-control mechanism that eliminates mRNAs with CDS-interrupting premature-termination codons (PTCs). At the same time, NMD regulates the half-lives of many mRNAs with intact CDS, and this indicates a more general function for NMD in post-transcriptional gene regulation<sup>11–14</sup>. The exact mechanism of NMD substrate recognition is not yet known, but the prevailing models posit that UPF1 is recruited to mRNA by ribosomes stalling at stop codons where signals for proper translation termination are absent, through interaction with the eukaryotic release factors eRF1 and eRF3 (ref. 24). The UPF1-eRF3 interaction is thought to be mutually exclusive with the interaction between eRF3 and poly(A)-binding protein C1 (PABPC1)<sup>25</sup>, which is required for proper termination<sup>26–28</sup>. Corroborating this UPF1-versus-PABP competition model, PABPC1 efficiently suppresses NMD when brought near a PTC by artificial tethering<sup>25,28–32</sup> or by folding back of the poly(A) tail<sup>31</sup>. The model predicts a subsequent handover of the ribosome-associated UPF1 to a 3' UTR-located exon junction complex (EJC) through interaction with UPF2 and UPF3 (ref. 24). A structural model of this complex consisting of UPF1, UPF2, UPF3B and the four EJC core factors eIF4AIII, Y14, Magoh and the SELOR domain of MLN51 was recently generated<sup>33</sup>. In this UPF-EJC complex model, UPF1 interacts with the RNA downstream of the EJC<sup>33</sup>, but it remains to be clarified how a UPF1 molecule coming from a stalling ribosome upstream of the EJC could reach this position. Interaction of UPF1 with UPF2 and UPF3 promotes SMG1-mediated phosphorylation of

<sup>1</sup>Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland. <sup>2</sup>Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland. <sup>3</sup>Biozentrum, University of Basel, Basel, Switzerland. <sup>4</sup>Swiss Institute of Bioinformatics, Basel, Switzerland. Correspondence should be addressed to O.M. (oliver.muehlemann@dcb.unibe.ch).

Received 29 January; accepted 20 June; published online 7 July 2013; doi:10.1038/nsmb.2635

UPF1 (refs. 24,34,35), and phosphorylated UPF1 then interacts with the three 14-3-3-like domain-containing proteins SMG6, SMG5 and SMG7 (refs. 36–38). Whereas SMG6 is an endonuclease that cleaves the target mRNA near the PTC<sup>39,40</sup>, SMG5 and SMG7 in conjunction with PNRC2 activate exonucleolytic RNA degradation by promoting decapping and deadenylation<sup>41–43</sup>.

In contrast to the proposed translation termination-dependent recruitment of UPF1 to mRNA, described above, there is evidence for translation-independent UPF1-RNA interaction<sup>44</sup>. UPF1 was reported to associate with mRNA in a 3' UTR length-dependent manner and to be highly enriched on transcripts that contain NMD-inducing 3' UTRs<sup>44</sup>. This preferential binding of Upf1 to mRNAs containing NMD-sensitive 3' UTRs was unaffected by translation inhibition, thus suggesting that the translation dependence of NMD is conferred after UPF1 association with RNA. Contradicting these data<sup>44</sup>, a very recent study reported evidence for translation-dependent recruitment of UPF1 to 3' UTRs of NMD-targeted mRNAs<sup>45</sup>.

To map the transcriptome-wide RNA-binding sites of UPF1 in human cells, we performed individual-nucleotide-resolution UV cross-linking and immunoprecipitation (iCLIP)<sup>46</sup> experiments in HeLa cells. Owing to the seemingly contradictory evidence with regard to the translation dependence of UPF1-RNA interaction, iCLIP was conducted with both untreated cells and cells in which translation was inhibited for 6 h before cross-linking, and the UPF1 binding patterns under both conditions were compared. Although UPF1 bound preferentially to the 3' UTR of mRNAs under both conditions, translation inhibition led to increased UPF1 binding in CDS. We obtained a similar result with RNA immunoprecipitations (RIPs) in which CDS and 3' UTRs were separated by oligonucleotide-mediated RNase-H cleavage, and this suggests that in the absence of translation (and probably before translation initiation), UPF1 associates with

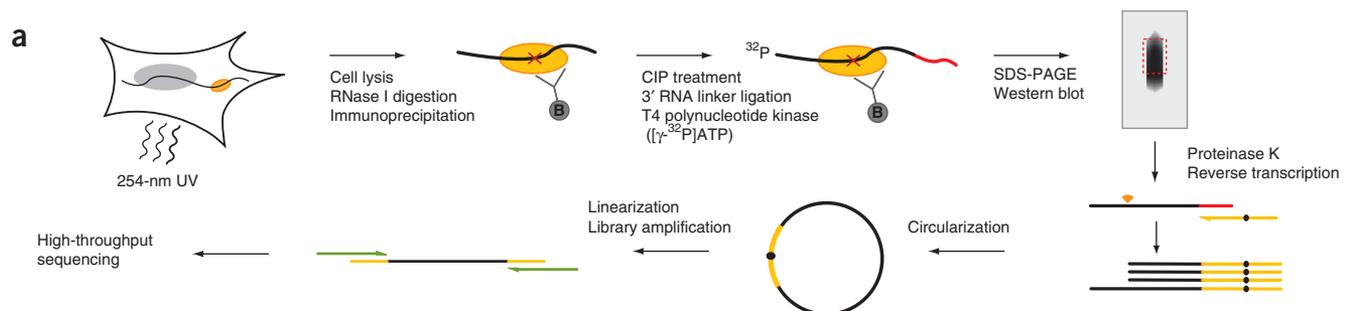
transcripts across their entire length and that elongating ribosomes displace UPF1 from CDS during translation.

## RESULTS

### iCLIP sequencing of endogenous UPF1 in human cells

Apart from minor modifications (described in Online Methods), we prepared iCLIP libraries as previously described<sup>46</sup>. During CLIP library preparation, reverse transcriptase aborts at >80% of the RNA-protein cross-link sites that it encounters, thus resulting in truncated cDNAs that are lost during cDNA library amplification in classical CLIP protocols<sup>47</sup>. In contrast, the iCLIP method preserves these truncated cDNAs by introducing the upstream adaptor on the DNA level through circularization of the cDNA, which is subsequently relinearized by restriction digestion (Fig. 1a). Hence, most of the sequences in the iCLIP library end at the reverse transcription-abortion site, thus allowing identification of the RNA-protein cross-link sites with individual-nucleotide resolution<sup>46</sup>.

To determine *in vivo* UPF1-RNA interaction sites transcriptome wide, we irradiated HeLa cells with UV-C light ( $\lambda = 254$  nm) before lysis. We subjected cell lysates to limited RNase I digestion to obtain RNA fragments, which were then coimmunopurified with endogenous UPF1 with a highly specific anti-UPF1 antibody and subsequently used for cDNA library preparation. We monitored the immunopurified and radioactively labeled RNA-protein adducts by western blotting (Fig. 1b). The smear representing RNA-protein adducts extended toward higher molecular weight with decreasing RNase I concentrations, and this signal was strongly reduced when the cells were depleted of UPF1 or not subjected to UV-C irradiation, a result indicating that most of the signal in the >150-kDa range represented UPF1 cross-linked to labeled RNA. HeLa cell extract treated with 1:1,000 diluted RNase I resulted in suitable RNA fragments of



**Figure 1** iCLIP of endogenous UPF1 in HeLa cells.

(a) Schematic illustration of the iCLIP procedure (details in Online Methods). CIP, calf intestinal alkaline phosphatase.

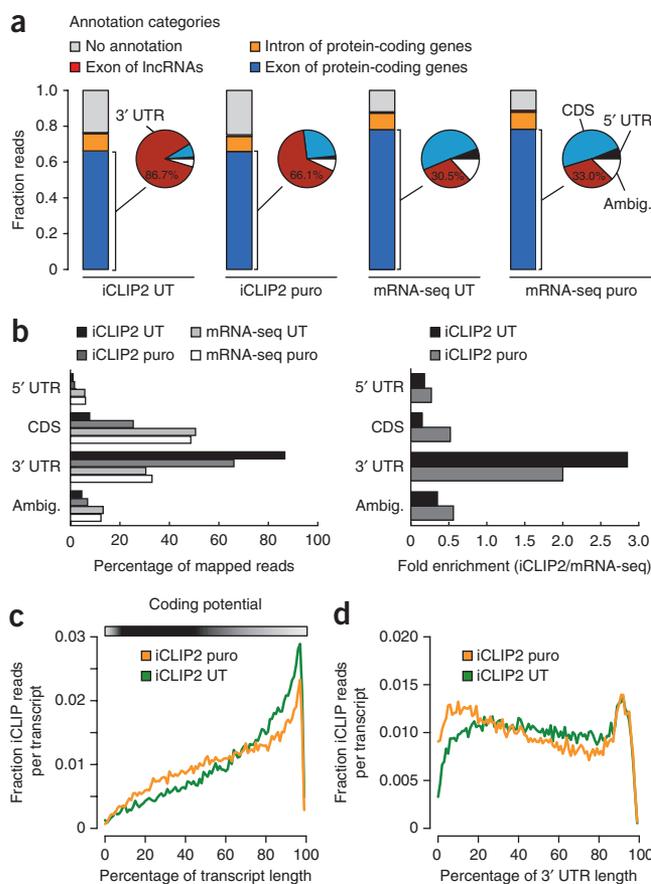
(b) Autoradiograph of immunopurified and <sup>32</sup>P-labeled UPF1-RNA complexes transferred to a nitrocellulose membrane. Results of RNase I digestion at 1:50 and 1:1,000 dilution are shown. Cells depleted of UPF1 (UPF1 KD) and cells that were not irradiated with UV-C light (no x-link) are controls for the specificity of the UPF1-RNA signal. The red arrow denotes the electrophoretic mobility of free UPF1 protein.

(c) Left, autoradiograph of <sup>32</sup>P-labeled UPF1-RNA adducts ([<sup>32</sup>P]RNA-x-UPF1) on a nitrocellulose membrane. Right, RNA from the indicated area (red square), isolated by proteinase-K treatment and analyzed on a 15% denaturing polyacrylamide (PAA) gel. The size range of the RNA fragments is indicated.

(d) Percentage of UPF1-associated reads uniquely mapping to exons, introns and lncRNAs from the iCLIP1 UT (untreated) experiment. (e) Left, exonic reads of the first iCLIP experiment with untreated cells (iCLIP1 UT) and an mRNA-seq experiment of untreated HeLa cells (mRNA-seq UT, used as reference), divided into four categories (5' UTR, CDS, 3' UTR and ambiguous), are shown as percentages of all exonic reads. Ambiguous (ambig.) refers to reads with different annotations in different transcripts of the same gene locus. Right,

enrichment factors calculated by division of the proportion of iCLIP1 UT reads by the proportion of reads of the mRNA-seq experiment for each category.

**Figure 2** Comparison of iCLIP from untreated and puromycin-treated cells. **(a)** Relative distribution of reads obtained in the iCLIP experiments with untreated (iCLIP2 UT) and puromycin-treated cells (iCLIP2 puro) and in the corresponding mRNA-seq data (mRNA-seq UT and mRNA-seq puro, respectively). The fraction of reads mapping to introns (orange) and exons (blue) of protein-coding genes and to lncRNAs (red), as well as reads with no annotation (gray), are represented as bars. The exonic subcategories 5' UTR (black), CDS (light blue), 3' UTR (dark red) and ambiguous (white) are represented as pie charts. **(b)** Representation of the distribution and enrichment of exonic reads in the categories 5' UTR, CDS, 3' UTR and ambiguous, as in **Figure 1e**. **(c)** Distribution of iCLIP2 UT and iCLIP2 puro reads along transcripts. Shown is the average proportion of reads as a function of the relative position in the transcript for the 5,000 most expressed genes (on the basis of the corresponding mRNA-seq data). For each gene, coordinates are relative to the start of the longest transcript variant annotated in RefSeq. The CDS density bar shows the average coding potential at each relative position among all transcripts in a white-to-black gradient (black, 100%; white, 0% CDS). **(d)** Similar to **c**, but showing distribution of reads along the 3' UTRs of transcripts. 3' UTRs <500 nt were excluded from this analysis.



approximately 100 nucleotides (nt) in length, which were isolated from the UPF1-RNA complex by proteinase-K digestion and separated on a denaturing polyacrylamide gel for visualization (**Fig. 1c**).

### UPF1 preferentially cross-linked to 3' UTRs

From the cDNA library of a first iCLIP experiment (iCLIP1 UT), we obtained  $2.9 \times 10^6$  single-end reads on an Illumina HiSeq 2000 sequencer. Of these,  $1.06 \times 10^6$  could be uniquely mapped to the human genome with the CLIPZ server<sup>48</sup> (**Supplementary Table 1**). The majority (76%) of the uniquely mapped reads mapped to exons and only 8% to introns, a result indicating that UPF1 mainly binds mRNA after splicing (**Fig. 1d**). Notably, a small fraction (0.57%) of the UPF1 iCLIP reads mapped to long noncoding RNAs (lncRNAs), thus providing a first hint that at least some UPF1 may bind RNA independently of translation (described below). Further classification of the exonic reads into 5' UTR-, CDS- and 3' UTR-derived and ambiguous regions (that had different annotation in different transcripts) revealed a pronounced enrichment of UPF1 binding to 3' UTRs, relative to the density of reads obtained from these regions in mRNA sequencing (mRNA-seq; **Fig. 1e**). Predominant interaction of UPF1 with 3' UTRs is consistent with the previously reported 3' UTR length-dependent association of UPF1 with mRNAs<sup>44,45</sup>; however, it did not allow us to distinguish whether the 3' UTR-specific cross-linking of UPF1 resulted from specific UPF1 recruitment to the 3' UTR (for example, by terminating ribosomes as described above) or whether UPF1 bound all over mRNAs before translation and subsequently was displaced from CDS by translating ribosomes.

To distinguish between these two possibilities, we performed an iCLIP experiment with HeLa cells treated with the translation inhibitor puromycin for 6 h before UV cross-linking (iCLIP2 puro) as well as an iCLIP experiment with untreated cells (iCLIP2 UT) for comparison. We obtained  $20.7 \times 10^6$  reads for iCLIP2 puro and  $15.4 \times 10^6$  reads for iCLIP2 UT. When aligned to the human genome,  $4.66 \times 10^6$  uniquely mapping reads could be identified for iCLIP2 puro and  $3.69 \times 10^6$  reads for iCLIP2 UT (**Supplementary Table 1**). In addition, mRNA-seq experiments were also performed for both conditions, and these data sets were used as a reference and to select highly expressed genes for subsequent analysis. Comparing the data obtained for the 5,000 most expressed genes from iCLIP1 UT and iCLIP2 UT, we found a strong correlation indicating a high degree of reproducibility (**Supplementary Fig. 1a**). As expected, we obtained

a slightly lower correlation between the data from iCLIP2 UT and from the puromycin-treated sample (iCLIP2 puro; **Supplementary Fig. 1b**). We also observed a high correlation between the mRNA-seq data from untreated and puromycin-treated cells, demonstrating that the puromycin treatment did not induce major changes in the transcriptome (**Supplementary Fig. 1c**).

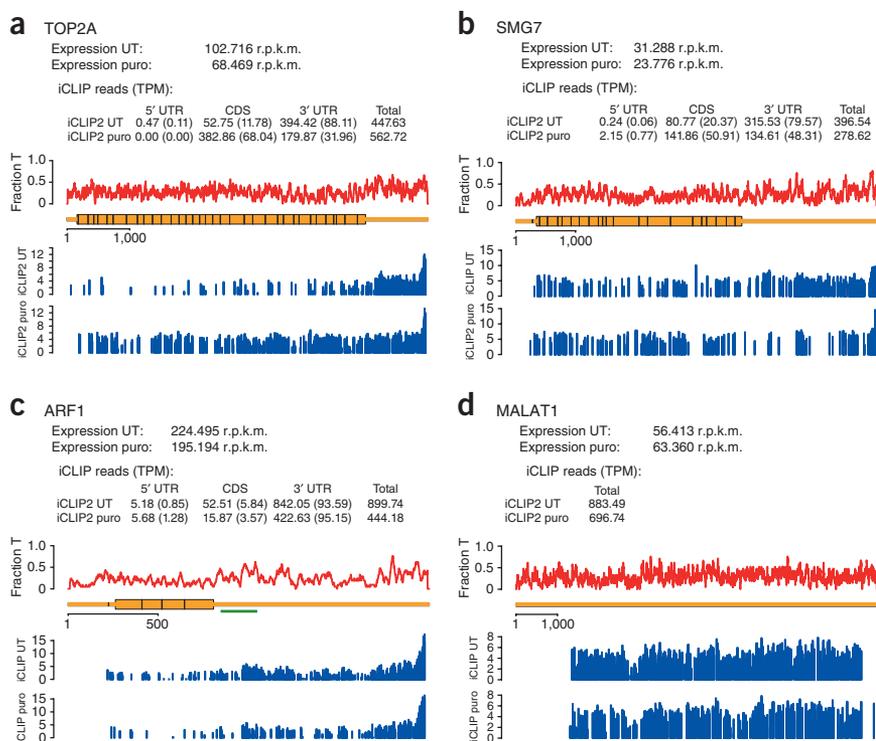
### Increased UPF1 occupancy in CDS upon translation inhibition

We aligned the unique UPF1-cross-linked reads to the human genome and analyzed their distribution between exons and introns of protein-coding genes as well as their distribution in lncRNAs. Again, the majority of UPF1-associated reads (~65%) mapped to exonic sequences and 8–10% to intronic sequences, irrespective of whether translation was inhibited (**Fig. 2a**). For ~24% of the reads, there was no annotation in the database, and a small fraction (0.7–0.9%) mapped to lncRNAs (discussed below).

The relative distribution of reads from iCLIP2 UT along mRNAs was almost identical to that observed in the initial iCLIP1 UT (comparison of **Fig. 2a** and **Fig. 1e**), again demonstrating the high reproducibility of our iCLIP data. In untreated cells, 86.7% of the mRNA-bound UPF1 was detected in 3' UTRs, corresponding to a nearly three-fold enrichment relative to the proportion of 3' UTR sequence present in the HeLa mRNAome (mRNA-seq). Notably, the puromycin-induced translation inhibition led to a marked relative redistribution of UPF1 binding from 3' UTRs toward the CDS: UPF1-cross-linked reads mapping to the CDS (relative to the total of uniquely mapped reads) were 3.3-fold increased at the expense of reads mapping to the 3' UTRs, which decreased by 25% (**Fig. 2** and **Supplementary Table 1**). This is consistent with a model in which UPF1 binds mRNAs over their entire length before translation, then is

**Figure 3** Examples of UPF1 association with individual transcripts. The number of reads located in 5' UTR, CDS and 3' UTR are shown for both iCLIP2 UT and iCLIP2 puro.

The expression level of an individual transcript is indicated by the number of reads in both the untreated and puromycin-treated mRNA-seq experiments, normalized for the transcript length and the total number of sequencing reads per kilobase per million (r.p.k.m.). The proportion of T nucleotides (fraction T; red) and the UPF1-binding sites (log<sub>2</sub> fold enrichment relative to a smoothed profile of mRNA-seq reads of the corresponding mRNA-seq sample, blue) are depicted along the schematically illustrated transcripts. Individual panels show the data for the DNA topoisomerase 2-alpha (TOP2A) (a), the NMD factor SMG7-encoding mRNA (b), the Staufen-mediated decay target ADP ribosylation factor 1 (ARF1) (c) and the long non-coding metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) (d). Green bar, Staufen-1-binding site in the ARF1 3' UTR. TPM, tags per million.



subsequently displaced from coding regions by the elongating ribosomes. This hypothesis is further supported by the profile of UPF1 iCLIP read densities along the 3' UTRs (Fig. 2d). Whereas in untreated cells the 3' UTRs were relatively uniformly covered by UPF1 (apart from a slight increase in read density toward the poly(A) tail that was also detected in the puromycin-treated sample), binding of UPF1 immediately downstream of the stop codon increased when translation was inhibited. This suggests that ribosomes hinder the UPF1 interaction with the RNA, opposite to what would be expected if terminating ribosomes recruited UPF1 to the 3' UTRs<sup>24,45</sup>. Collectively, our iCLIP data revealed a translation-dependent UPF1 redistribution from CDS to the 3' UTR.

#### Slight preference for UPF1 cross-linking to U nucleotides

To investigate whether UPF1 has a preferred binding sequence, we analyzed the nucleotide composition of the 1,000 most abundant UPF1 cross-link sites in CDS or 3' UTRs by calculating the relative frequencies of the four nucleotides at each position in a 21-nt window centered on the cross-linked nucleotide (Supplementary Fig. 2a,b). Consistent with the known U bias of UV-C cross-linking methods<sup>47</sup>, the reads showed a clear U enrichment at the cross-link sites. Additionally, a slightly increased U nucleotide frequency was apparent at positions neighboring the cross-link (from -2 to +2), but beyond this, the sequence composition around the cross-linked sites appears to reflect the overall nucleotide composition of the region (CDS or 3' UTRs). A search for over-represented sequence motifs with MEME in the region of the 1,000 most abundant UPF1 cross-linked sites did not identify a preferred binding motif for UPF1 (A.R.G., unpublished data). Nevertheless, UPF1 cross-links are not uniformly distributed along RNAs; some clustering is observed in both the untreated and the puromycin-treated samples. The percentage of UPF1 cross-link sites with another cross-link site within a short distance is much larger than would be expected if cross-links were distributed randomly within the respective region (CDS or 3' UTRs) (Supplementary Fig. 2c). For example, 60% of the iCLIP2 UT cross-links have another cross-link within the next 3 nt, whereas only 35% would be expected by chance.

#### UPF1 distribution on selected individual transcripts

After our global assessment of UPF1 interaction with the transcriptome, we inspected the UPF1 occupancy on individual transcripts (a few of which are depicted in Fig. 3). Confirming the mild preference for UV cross-linking of U nucleotides<sup>47</sup> or perhaps some sequence or structure constraints on UPF1 binding to RNAs, regions of high UPF1 occupancy tend to coincide with regions of high frequency of U nucleotides (Fig. 3, comparison of CLIP-read distribution with fraction-T trace). The mRNA encoding DNA topoisomerase 2-alpha (TOP2A) displays the frequently observed redistribution of UPF1 from mainly 3' UTRs in untreated cells toward CDS in puromycin-treated cells (Fig. 3a). In actively translating cells, the TOP2A CDS were essentially devoid of UPF1, which was detected mainly in the 3' UTRs. We observed a more even distribution of UPF1 throughout the entire mRNA in puromycin-treated cells, a result suggesting that translation inhibition prevents UPF1 from being stripped off the CDS by elongating ribosomes. We also observed a similar pattern, albeit less pronounced, with most mRNAs coding for NMD factors, which themselves are NMD targets<sup>14,49</sup>, as exemplified by SMG7 mRNA (Fig. 3b). These results imply that at least a fraction of UPF1 binds RNA independently of translation.

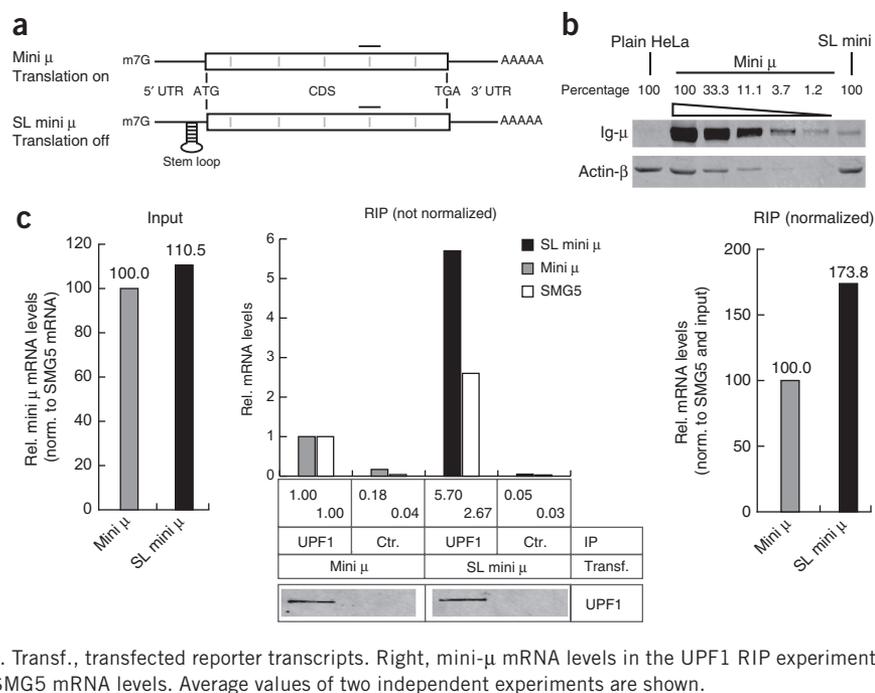
The mRNA encoding ADP ribosylation factor 1 (ARF1) is a well-characterized target for Staufen-mediated mRNA decay, and UPF1 was previously described to be recruited to ARF1 mRNA by interacting with STAU1 (ref. 50). The binding site for STAU1 was mapped to a branched hairpin structure in the first third of the 3' UTR of ARF1 (ref. 50), a region in which we also detected increased UPF1 cross-linking (Fig. 3c).

#### UPF1 associates with lncRNAs

Association of UPF1 with lncRNAs (Fig. 2a) and its similar distribution along the lncRNAs in puromycin-treated and control cells (Fig. 3d and Supplementary Fig. 3) further support the hypothesis that UPF1 binds RNA independently of translation. For example, in the highly abundant metastasis-associated lung adenocarcinoma

**Figure 4** Transcript-specific translation inhibition promotes UPF1 association.

(a) Schematic representation of the Ig- $\mu$  reporter transcripts mini  $\mu$  and SL mini  $\mu$ . The translation-inhibiting stable stem loop in the 5' UTR of SL mini  $\mu$  is depicted. The CDS is represented as a white box flanked by the 5' and 3' UTRs. Gray bars, exon-exon junctions; black line, position of the TaqMan assay used for real-time qPCR with reverse transcription (RT-qPCR). (b) Western blot analysis of Ig- $\mu$  and actin- $\beta$  protein in a serially diluted cell lysate of HeLa mini  $\mu$ , undiluted lysate of HeLa SL mini  $\mu$  for comparison and untransfected (plain) HeLa cells as a control, showing the extent of translation inhibition caused by the stem loop. (c) RIPs with anti-UPF1 antibody, from cells expressing either the mini- $\mu$  or the SL-mini- $\mu$  reporter construct. Left, relative (rel.) expression levels of mini- $\mu$  and SL-mini- $\mu$  mRNAs, determined by RT-qPCR and normalized (norm.) to endogenous SMG5 mRNA, in the cell lysates before the RIP. Middle, relative mRNA levels of mini  $\mu$  (with or without SL) and SMG5, co-purified in either the UPF1 or the control (ctr.) immunoprecipitations (IP). Transf., transfected reporter transcripts. Right, mini- $\mu$  mRNA levels in the UPF1 RIP experiments, normalized to the input and the coimmunopurified SMG5 mRNA levels. Average values of two independent experiments are shown.



transcript 1 (MALAT1, which in the cells that we studied appears to be a short isoform; **Fig. 3d**), UPF1 iCLIP reads are distributed relatively uniformly along the entire transcript. Furthermore, the overall density of UPF1 iCLIP reads is similar in untreated and puromycin-treated cells, consistent with the expectation that no ribosomes engage with noncoding MALAT1 RNA even in untreated cells. Four other inspected lncRNAs gave a similar picture (TUG1, GAS5, RP11-22011.1 and FGDS-AS1; **Supplementary Fig. 3**). Notably, the majority of MALAT1 transcripts reside in the nucleus<sup>51</sup>, thus suggesting that UPF1 might interact with mRNA even before its export to the cytoplasm.

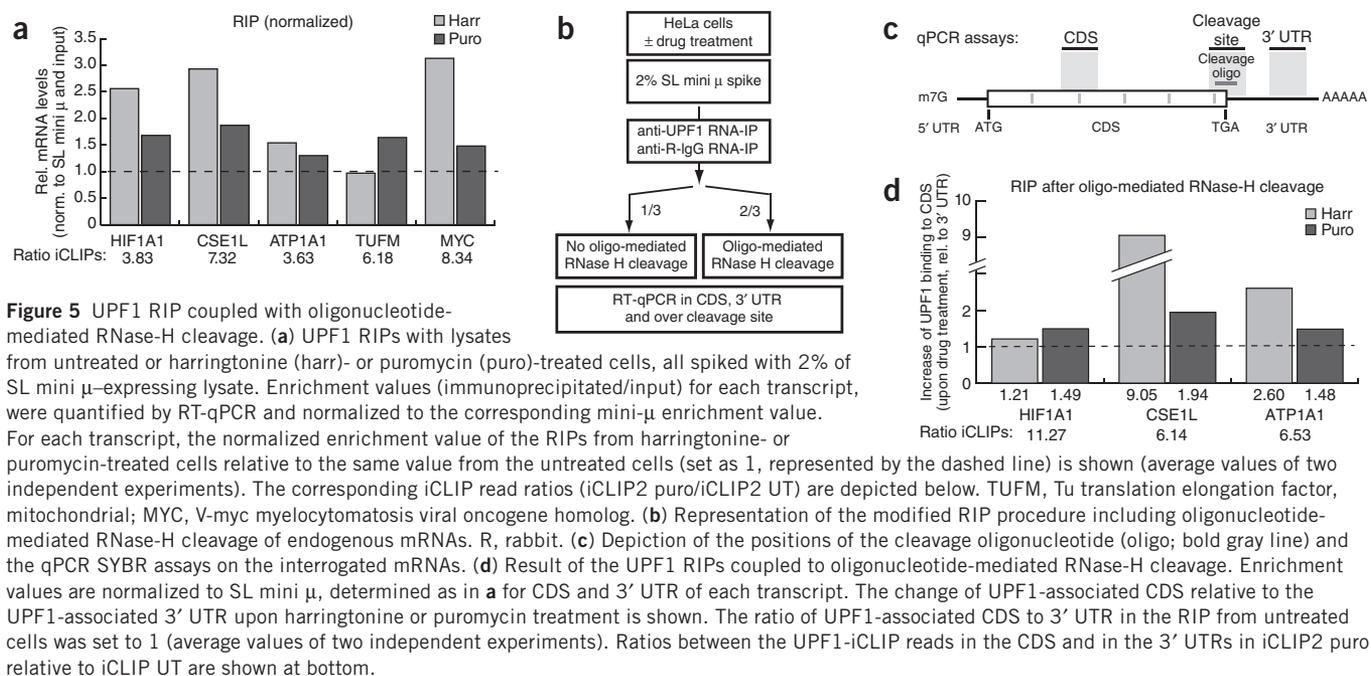
### Translationally blocked reporter transcript binds more UPF1

To rule out that the relative shift of UPF1 from 3' UTRs to CDS observed in our iCLIP data was a pleiotropic effect of puromycin or an artifact of the iCLIP technique, we carried out RIPs from HeLa cells that either expressed a normally translated immunoglobulin- $\mu$  (Ig- $\mu$ ) reporter (called mini  $\mu$  (ref. 52)) or a version with a stable stem loop in the 5' UTR (SL mini  $\mu$ ; **Fig. 4a**) that inhibits the 48S preinitiation-complex formation at the cap. The stem loop reduced translation of SL-mini- $\mu$  mRNA by 98% relative to mini  $\mu$  (**Fig. 4b**) while the mRNA levels of the two constructs were similar (**Fig. 4c**). Mini- $\mu$  and SL-mini- $\mu$  mRNAs were readily detected in the anti-UPF1 RIPs, whereas the control RIP performed with an anti-rabbit IgG precipitated very little mRNA, thus confirming specific association of the transcripts with UPF1 (**Fig. 4c**). The relative abundance of the two mini- $\mu$  reporter mRNAs in the UPF1 RIPs was determined by normalization of their mRNA levels in the immunoprecipitations to their input levels and to the amount of coimmunopurified endogenous SMG5 mRNA, which was known to efficiently associate with UPF1 (ref. 14). Compared to the normally translated mini- $\mu$  transcript, 1.7-fold more of the barely translated SL-mini- $\mu$  reporter transcript co-purified with UPF1, thus supporting our conclusion that the observed increased UPF1 association with CDS in our iCLIP2 puro experiment was caused by the blocked translation of the CDS rather than by a secondary effect of the puromycin treatment.

### RIP of selected endogenous mRNAs confirms the iCLIP data

To further validate the iCLIP data by another technique, we performed RIP from untreated or puromycin- and harringtonine-treated cells, using the same goat anti-UPF1 antibody as in the iCLIP experiments or a goat anti-rabbit IgG serum as a control (**Supplementary Fig. 4a**). To normalize for different immunoprecipitation efficiencies, we spiked all lysates with 2% of an SL mini  $\mu$ -containing lysate. First, we determined the enrichment ratios in the translation-inhibited cells relative to the untreated cells for five endogenous mRNAs that had more iCLIP reads in the iCLIP2 puro than in the iCLIP2 UT (**Fig. 5a**). Consistent with the iCLIP result, all five transcripts were also more enriched in RIPs of puromycin-treated cells than of untreated cells. As expected, the difference in RNA-associated UPF1 between the puromycin-treated and the untreated samples was less pronounced in RIP than in iCLIP because cross-linking captures even weak interactions that are lost in the RIP procedure. Harringtonine, which inhibits translation initiation<sup>53</sup>, also led to increased enrichment for four of the five tested mRNAs, a result indicating that the effect is independent of the drug used to inhibit translation.

To test whether the increased UPF1-mRNA association that occurred in iCLIP upon translation inhibition originated from an increased interaction with CDS, we modified the RIP protocol as follows (**Fig. 5b**). After immunoprecipitation, when the RNA was still bound to the beads, we added RNase H and DNA oligonucleotides complementary to the sequence around the stop codons of HIF1A1, CSE1L or ATP1A1 mRNAs to separate the CDS (and 5' UTR) from the 3' UTR (**Fig. 5b,c**). The average cleavage efficiencies varied from 86% (HIF1A1 oligonucleotide) to 57% (CSE1L oligonucleotide; **Supplementary Fig. 4b**). We determined the relative change of enrichment values upon translation inhibition (by puromycin or harringtonine), normalized for the spiked-in SL mini  $\mu$ , separately with quantitative PCR (qPCR) assays for CDS and for 3' UTRs (CDS/3' UTR ratios in **Fig. 5d**). Consistent with the iCLIP data, though less pronounced, the association of CDS with UPF1 increased for each of the three tested mRNAs relative to the association of 3' UTR with UPF1 when translation was inhibited. Collectively, our RIP data corroborate the findings from the iCLIP experiments and show that

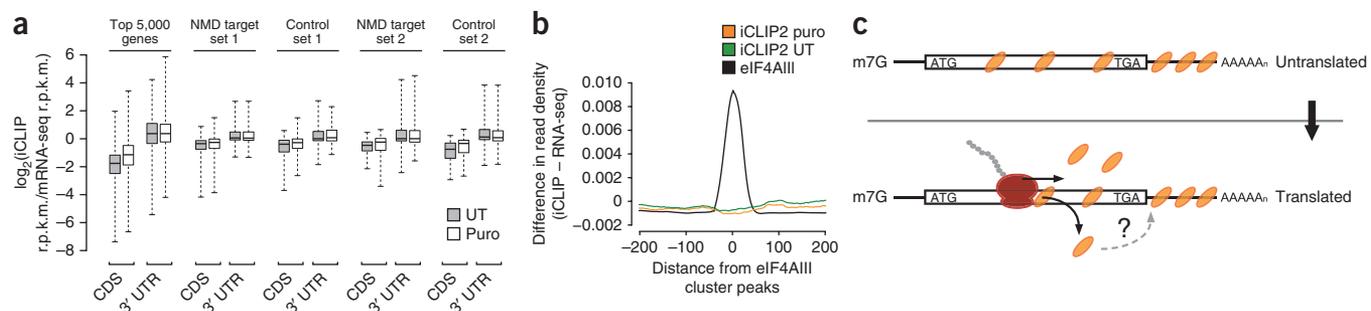


translation inhibition leads to more UPF1 binding in the CDS, indicating that UPF1 does engage with virgin messenger ribonucleoprotein particles (i.e. that are not yet translated) and gets displaced from the CDS by elongating ribosomes during translation, as has been shown for other mRNA-binding proteins, in particular for components of the EJC<sup>54,55</sup>.

### Similar UPF1 densities on NMD-targeted and non-NMD mRNAs

Several lines of evidence suggest that UPF1 specifically binds NMD-targeted transcripts (reviewed in refs. 5,6), and we therefore interrogated our data for differential UPF1 densities in CDS and 3' UTRs between two groups of experimentally determined NMD target genes<sup>12,14</sup> and randomly chosen control groups consisting of the same number of equally expressed non-NMD target genes (Fig. 6a). NMD target set 1 consisted of 144 genes that in a previous transcriptome profiling were significantly upregulated in both UPF1- and SMG6-depleted HeLa cells<sup>14</sup>. NMD target set 2 comprised 75 genes with an increase in both half-life and steady-state mRNA level upon UPF1

knockdown<sup>12</sup>. The 5,000 most expressed genes served as a reference population in this analysis. Overall, the median UPF1 densities for the corresponding categories were similar in all three groups and confirmed the previously discussed puromycin-induced UPF1 redistribution. In the top 5,000 genes, the increase in CDS density upon puromycin treatment was highly significant ( $P = 2.2 \times 10^{-16}$  by two-sided  $t$  test,  $n = 5,000$ ), and the decrease in 3' UTR density was also significant ( $P = 0.032$ ). However, there was no statistically significant difference ( $P \leq 0.05$ ) between the NMD target gene sets and the corresponding control sets in the categories '3' UTR UT', 'CDS puro' and '3' UTR puro'. Only the UPF1 density in CDS (but not 3' UTRs) of both NMD target sets were marginally but significantly changed in the UT data set ( $P = 0.039$  for set 1 by two-sided  $t$  test,  $n = 144$ ;  $P = 0.0015$  for set 2,  $n = 75$ ). Thus, a pronounced preference of UPF1 binding to NMD-targeted transcript clearly does not seem to exist, and this implies that UPF1 interaction with RNA precedes the decision of whether NMD ensues on a given mRNA.



**Figure 6** Neither NMD substrate mRNAs nor EJC-proximal regions are enriched with UPF1. **(a)** Comparison of UPF1-associated read density in CDS and 3' UTRs between untreated (UT) and puromycin-treated (puro) samples on four gene groups. NMD target set 1 (ref. 14) and NMD target set 2 (ref. 12) were compared to control groups consisting of the same number of randomly selected genes with similar expression levels (control set 1 and 2) and to the 5,000 most expressed genes. Enrichment factors for the iCLIP reads, normalized to the mRNA-seq reads are depicted as box plots for each category. Black lines, median values; boxes, upper and lower quartiles. **(b)** Comparison of UPF1 iCLIP sites with eIF4AIII-binding sites identified by CLIP-seq<sup>56</sup>. For iCLIP2 UT and iCLIP2 puro, the differences in read densities of the iCLIP experiments and the corresponding mRNA-seq experiments at various positions relative to the 5,000 most prominent eIF4AIII clusters (anchored at 0) are shown. **(c)** Model illustrating our conclusion of translation-independent binding of UPF along the entire mRNA and subsequent displacement from the CDS by elongating ribosomes upon translation. Displaced UPF1 proteins can reassociate with RNA, further contributing to an increased UPF1 occupancy in 3' UTRs.

### UPF1-binding sites and EJC positions do not correlate

There is compelling evidence for the formation of a transient multi-meric NMD complex consisting of UPF1, UPF2, UPF3 and the EJC core factors eIF4AIII, Y14, Magoh and MLN51 at some unknown point in the NMD pathway<sup>33</sup>. We therefore tested whether EJC- and UPF1-cross-linked sites clustered together on mRNA. Recently, the eIF4AIII- (and by inference the EJC-) binding sites in the HeLa cell transcriptome were identified with CLIP-seq<sup>56</sup>. We used these data to map the UPF1-associated read densities of iCLIP2 UT and iCLIP2 puro relative to the positions of the top 5,000 EJC clusters inferred from the previous study<sup>56</sup> (Fig. 6b). Within the 200 nt on either side of the aligned EJC-binding sites, we could not detect any enrichment of UPF1 reads. The lack of an obvious correlation between EJC and UPF1 positions could indicate that UPF1 interacts with mRNA independently of EJCs, but we cannot exclude that a possible correlation might be obscured by the comparison of two data sets obtained with different CLIP protocols.

### DISCUSSION

Transcriptome-wide profiling of UPF1-interaction sites with the iCLIP technique revealed a pronounced enrichment of UPF1-mRNA interaction on 3' UTRs and a concomitant under-representation on CDS and 5' UTRs (Figs. 1 and 2). We also observed UPF1 cross-linking to many lncRNAs, arguing for the association of at least some UPF1 with RNA independently of translation (Fig. 3d and Supplementary Fig. 3). That UPF1 even associated with the mainly nuclear lncRNA MALAT1 suggests that UPF1 may bind RNA inside the nucleus, even though the majority of the shuttling protein UPF1 resides in the cytoplasm at steady state<sup>57,58</sup>. The bias toward exonic rather than intronic sequence of the iCLIP reads relative to the mRNA-seq data identifies spliced mRNAs as the main substrate for UPF1 interaction.

Most notably, inhibition of translation elongation with puromycin before iCLIP resulted in a relative redistribution of UPF1: the UPF1 density increased in CDS and decreased in 3' UTRs (Fig. 2). We confirmed this finding by RIP experiments with a reporter transcript in which translation was inhibited *in cis* by the presence of a stem loop in the 5' UTR (Fig. 4) and for selected endogenous mRNAs (Fig. 5), demonstrating that the relative shift of UPF1 from 3' UTRs to CDS occurs irrespectively of the method used to inhibit translation. The low CDS occupancy of UPF1 in translation-competent cells implies that the elongating ribosomes strip off UPF1 when they translocate through the CDS (Fig. 6c).

The preference of UPF1 for 3' UTRs is in principle consistent with the proposed recruitment of UPF1 to mRNAs through eRF-bound ribosomes stalling at stop codons for a prolonged time and therefore triggering NMD (reviewed in ref. 59), but we consider this scenario unlikely because we did not detect enhanced UPF1 association with known NMD substrate mRNAs (Fig. 6a) and found that translation decreased rather than increased UPF1 density immediately downstream of stop codons (Fig. 2d). The simplest explanation for enhanced UPF1 occupancy of 3' UTRs in untreated cells is that the ongoing displacement of UPF1 from CDS by passing ribosomes leads to an apparent increase in binding to the 3' UTRs, where the interactions are not destabilized by ribosomes.

Given that NMD is a translation-dependent process, our finding of translation-independent formation of UPF1-RNA complexes implies that the decision of whether NMD ensues must occur at a step after UPF1 binding. That we detected no significant difference in UPF1 binding to CDS and 3' UTRs between experimentally determined endogenous NMD substrate mRNAs and a control group of non-NMD targets further supports this scenario (Fig. 6a). On the basis of previous work, it seems likely that the NMD-triggering step might

be the interaction of mRNA-bound UPF1 with eRFs bound to an aberrantly terminating ribosome. This UPF1-eRF interaction occurs in a complex that contains the UPF1-phosphorylating kinase SMG1 (ref. 24), but for the phosphorylation of its Ser/Thr-Gln motifs, UPF1 has to interact additionally with UPF2 and UPF3 (ref. 24). Because UPF1 has been shown to form a complex with UPF2, UPF3 and the EJC core<sup>20,33</sup>—and because UPF1 does not bind a specific sequence motif yet shows a strong preference for mRNA, and its distribution on mRNA is not random (Supplementary Fig. 2)—we hypothesized that UPF1- and EJC-binding sites might cluster on mRNA. However, a comparison between the positions of the top 5,000 eIF4AIII clusters identified in a previous study<sup>56</sup> relative to our UPF1-associated RNA sequences showed no correlation between the positions of EJC- and UPF1-interaction sites along mRNAs (Fig. 6b). These results might indicate that UPF1 binds mRNA independently of EJCs. However, it is also possible that *in vivo* the UPF1 association with the EJC either happens in a configuration in which UPF1 does not touch the RNA in a cross-linkable way or that it is very transient. Notably, besides its phosphorylation-dephosphorylation cycle, UPF1's ATPase and helicase activities are also essential for NMD<sup>24,60</sup>. Because we analyzed RNA cross-linked to the endogenous UPF1, we cannot distinguish between initial UPF1-binding sites and positions to which activated UPF1 has translocated by using its helicase activity. This could be addressed in the future by comparison of cross-link patterns of exogenously expressed recombinant wild-type UPF1 and different mutants, foremost a helicase-inactive UPF1 mutant.

In summary, we report here the first transcriptome-wide identification, to our knowledge, of human UPF1-interaction sites with RNA *in vivo*. We provide evidence that UPF1, like other RNA-binding proteins, interacts with mRNA rather promiscuously before translation and that the observed preferential occupancy of UPF1 in 3' UTRs at steady state results from the selective displacement of UPF1 from CDS by translating ribosomes. Our findings are thus inconsistent with the suggested specific recruitment of UPF1 to NMD-targeted transcripts. Instead, they indicate that NMD substrate selection occurs after UPF1 association with mRNA.

### METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession code.** The high-throughput sequencing data have been deposited in the Gene Expression Omnibus database under accession number [GSE47976](#).

*Note: Supplementary information is available in the [online version of the paper](#).*

### ACKNOWLEDGMENTS

We thank R. Bruggmann and S. Neuenschwander (Institute of Cell Biology, University of Bern, Bern, Switzerland) for initial bioinformatics support, H. le Hir (Institut de Biologie, École Normale Supérieure, Paris, France) for technical hints with the CLIP and for providing their eIF4AIII CLIP data, L. Maquat (University of Rochester, Rochester, New York, USA) for providing their protocol for RIP coupled with RNase-H cleavage, S. Rufener (Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland) for help with the RIPs and S. Kishore and N. Mittal (Biozentrum, University of Basel, Basel, Switzerland) for sharing the HeLa mRNA-seq data before publication. This work was supported by grants to O.M. from the European Research Council (StG 207419), the Swiss National Science Foundation (31003A-127614 and 31003A-143717) and the canton of Bern. The work of A.R.G. and M.Z. was supported by the University of Basel.

### AUTHOR CONTRIBUTIONS

D.Z. and O.M. conceived of the project. D.Z. performed all experiments. A.R.G. performed most of the bioinformatics analysis. D.Z., O.M., A.R.G. and M.Z. wrote the paper. O.M. provided resources.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Azzalin, C.M. & Lingner, J. The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. *Curr. Biol.* **16**, 433–439 (2006).
- Chawla, R. *et al.* Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication. *EMBO J.* **30**, 4047–4058 (2011).
- Kim, Y.K., Furic, L., Desgroseillers, L. & Maquat, L.E. Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell* **120**, 195–208 (2005).
- Kaygun, H. & Marzluff, W.F. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat. Struct. Mol. Biol.* **12**, 794–800 (2005).
- Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* **13**, 700–712 (2012).
- Nicholson, P. *et al.* Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell Mol. Life Sci.* **67**, 677–700 (2010).
- Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C. & Riha, K. Aberrant growth and lethality of *Arabidopsis* deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Res.* **40**, 5615–5624 (2012).
- Medghalchi, S.M. *et al.* Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* **10**, 99–105 (2001).
- Wittkopp, N. *et al.* Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. *Mol. Cell Biol.* **29**, 3517–3528 (2009).
- Avery, P. *et al.* *Drosophila* Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner. *RNA* **17**, 624–638 (2011).
- Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F. & Dietz, H.C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **36**, 1073–1078 (2004).
- Tani, H. *et al.* Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol.* **9**, 1370–1379 (2012).
- Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W. & Kulozik, A.E. The abundance of RNP51, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. *Nucleic Acids Res.* **35**, 4542–4551 (2007).
- Yepiskoposyan, H., Aeschmann, F., Nilsson, D., Okoniewski, M. & Muhlemann, O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17**, 2108–2118 (2011).
- Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).
- Chakrabarti, S. *et al.* Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell* **41**, 693–703 (2011).
- Cheng, Z., Muhrad, D., Lim, M.K., Parker, R. & Song, H. Structural and functional insights into the human Upf1 helicase core. *EMBO J.* **26**, 253–264 (2007).
- Clerici, M. *et al.* Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* **28**, 2293–2306 (2009).
- Bhattacharya, A. *et al.* Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6**, 1226–1235 (2000).
- Chamieh, H., Ballut, L., Bonneau, F. & Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**, 85–93 (2008).
- Czaplinski, K. *et al.* The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12**, 1665–1677 (1998).
- Fiorini, F., Boudvillain, M. & Le Hir, H. Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res.* **41**, 2404–2415 (2013).
- Franks, T.M., Singh, G. & Lykke-Andersen, J. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell* **143**, 938–950 (2010).
- Kashima, I. *et al.* Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* **20**, 355–367 (2006).
- Singh, G., Rebbapragada, I. & Lykke-Andersen, J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* **6**, e111 (2008).
- Cosson, B. *et al.* Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. *Mol. Cell Biol.* **22**, 3301–3315 (2002).
- Hoshino, S., Imai, M., Kobayashi, T., Uchida, N. & Katada, T. The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-poly(A) tail of mRNA: direct association of eRF3/GSPT with polyadenylate-binding protein. *J. Biol. Chem.* **274**, 16677–16680 (1999).
- Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W. & Kulozik, A.E. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* **27**, 736–747 (2008).
- Amrani, N. *et al.* A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112–118 (2004).
- Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V. & Izaurralde, E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* **26**, 1591–1601 (2007).
- Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z. & Muhlemann, O. Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol.* **6**, e92 (2008).
- Silva, A.L., Ribeiro, P., Inacio, A., Liebhaber, S.A. & Romao, L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA* **14**, 563–576 (2008).
- Melero, R. *et al.* The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3' end. *Nat. Struct. Mol. Biol.* **19**, 498–505 (2012).
- Arias-Palomo, E. *et al.* The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev.* **25**, 153–164 (2011).
- Yamashita, A., Kashima, I. & Ohno, S. The role of SMG-1 in nonsense-mediated mRNA decay. *Biochim. Biophys. Acta* **1754**, 305–315 (2005).
- Fukuhara, N. *et al.* SMG7 is a 14–3–3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell* **17**, 537–547 (2005).
- Okada-Katsuhata, Y. *et al.* N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res.* **40**, 1251–1266 (2012).
- Jonas, S., Weichenrieder, O. & Izaurralde, E. An unusual arrangement of two 14–3–3-like domains in the SMG5–SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev.* **27**, 211–225 (2013).
- Eberle, A.B., Lykke-Andersen, S., Muhlemann, O. & Jensen, T.H. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* **16**, 49–55 (2009).
- Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J. & Izaurralde, E. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**, 2609–2617 (2008).
- Cho, H., Kim, K.M. & Kim, Y.K. Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. *Mol. Cell* **33**, 75–86 (2009).
- Cho, H. *et al.* SMG5-PNRC2 is functionally dominant compared with SMG5–SMG7 in mammalian nonsense-mediated mRNA decay. *Nucleic Acids Res.* **41**, 1319–1328 (2013).
- Unterholzner, L. & Izaurralde, E. SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell* **16**, 587–596 (2004).
- Hogg, J.R. & Goff, S.P. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **143**, 379–389 (2010).
- Kurosaki, T. & Maquat, L.E. Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc. Natl. Acad. Sci. USA* **110**, 3357–3362 (2013).
- König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
- Sugimoto, Y. *et al.* Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* **13**, R67 (2012).
- Khorshid, M., Rodak, C. & Zavolan, M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res.* **39**, D245–D252 (2011).
- Huang, L. *et al.* RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell* **43**, 950–961 (2011).
- Kim, Y.K. *et al.* Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J.* **26**, 2670–2681 (2007).
- Hutchinson, J.N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).
- Bühler, M., Paillusson, A. & Muhlemann, O. Efficient downregulation of immunoglobulin  $\mu$  mRNA with premature translation-termination codons requires the 5'-half of the VDJ exon. *Nucleic Acids Res.* **32**, 3304–3315 (2004).
- Fresno, M., Jimenez, A. & Vazquez, D. Inhibition of translation in eukaryotic systems by harringtonine. *Eur. J. Biochem.* **72**, 323–330 (1977).
- Dostie, J. & Dreyfuss, G. Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.* **12**, 1060–1067 (2002).
- Ishigaki, Y., Li, X., Serin, G. & Maquat, L.E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**, 607–617 (2001).
- Saulière, J. *et al.* CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* **19**, 1124–1131 (2012).
- Lykke-Andersen, J., Shu, M.D. & Steitz, J.A. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**, 1121–1131 (2000).
- Mendell, J.T., ap Rhys, C.M. & Dietz, H.C. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* **298**, 419–422 (2002).
- Stalder, L. & Muhlemann, O. The meaning of nonsense. *Trends Cell Biol.* **18**, 315–321 (2008).
- Page, M.F., Carr, B., Anders, K.R., Grimson, A. & Anderson, P. SMG-2 is a phosphorylated protein required for mRNA surveillance in *Caenorhabditis elegans* and related to Upf1p of yeast. *Mol. Cell Biol.* **19**, 5943–5951 (1999).

## ONLINE METHODS

**UPF1 iCLIP protocol.** HeLa cells were cultivated in DMEM supplemented with 100 U/ $\mu$ L penicillin, 100  $\mu$ g/mL streptomycin and 10% FCS in ten 150 cm<sup>2</sup> dishes until 80% confluency (corresponding to a total of  $8 \times 10^7$  cells). For the iCLIP2 puro experiment, puromycin (Calbiochem) was added to a final concentration of 60.6  $\mu$ M 6 h before cross-linking. After medium removal, cells were covered with 12 mL ice-cold PBS and subjected to 150 mJ/cm<sup>2</sup> UV-C light (Stratagene Stratalinker 2400, 254 nm). After irradiation, cells were scraped off the dishes in 10 mL fresh PBS. Cell pellets were shock frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  until use.

After being thawed on ice, cells were lysed in 7 mL hypotonic gentle lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2 mM EDTA and 0.5% Triton X-100, supplemented with 1 $\times$  Halt Protease Inhibitor (Thermo Scientific)) for 20 min. The cell lysate was cleared by centrifugation at 4  $^\circ\text{C}$  and 13,000g for 15 min. For partial RNA digestion, RNase I (Ambion cat. no. AM2294) was used at 1:1,000 dilution, the optimal concentration determined in preliminary experiments. The cell lysate was incubated with 70  $\mu$ L of the diluted RNase I and 35  $\mu$ L Turbo DNase (Ambion) at 37  $^\circ\text{C}$  for 3 min, followed by a 3-min incubation on ice. Then 28  $\mu$ L of anti-UPF1 antibody (Bethyl, goat anti-RENT1, A300-038A) was added and rotated at 4  $^\circ\text{C}$  for 1 h. Meanwhile, 240  $\mu$ L of Dynabeads Protein G (Life Technologies) was washed twice with 1 mL PBS and once with 1 mL hypotonic gentle lysis buffer. After 1-h preincubation of the lysate with the anti-UPF1 antibody, the washed beads were added and incubated at 4  $^\circ\text{C}$  for 1 h. After immunoprecipitation, the beads were washed three times with IP wash buffer (50 mM HEPES-NaOH, pH 7.5, 300 mM KCl, 0.05% NP-40, 0.5 mM DTT and 1 $\times$  Halt Protease Inhibitor).

The RNA moiety of the coimmunopurified complex was subjected to cDNA library preparation as previously described<sup>46</sup>. The RNA was dephosphorylated, and a 5'-phosphorylated 3' RNA linker (p-UGAGAUCGGAAGAGCGGUUCAG-puromycin) was ligated to the 3' end of the purified RNA fragments. To detect the UPF1 RNA adducts by western blotting, the RNA was radiolabeled with [ $\gamma$ -<sup>32</sup>P]ATP, and the protein-mRNA complex was loaded onto a 4–12% gradient Novex NuPAGE gel (Invitrogen) and transferred with the iBlot system (Invitrogen) to a nitrocellulose membrane. The area of the nitrocellulose membrane containing the [<sup>32</sup>P]RNA-UPF1 complex was excised, and the RNA was released by Proteinase-K digestion. Reverse transcription of the RNA was performed with a DNA primer complementary to the 3' RNA linker (p-NNNNAACAGATCGG AAGAGCGTCTGGATCCTGAACCGC). The cDNA was separated on a 6% TBE-urea gel, and a gel piece covering the size range of 300–150 nt was excised. The cDNA was then extracted from the gel piece, and the single-stranded cDNA was circularized and hybridized to a DNA cut oligo (5'-GTTTCAGGATCCACGA CGCTCTTCAAAA-3') to generate a double-stranded DNA stretch containing a BamHI site. After linearization by BamHI digestion, the adaptors needed for high-throughput Illumina sequencing were introduced by PCR with the following primers: P5\_Solexa (AATGATACGCGCACCACCGAGATCTACTCTTTC CCTACAGCAGCTCTTCCGATCT) and P3\_Solexa (CAAGCAGAAGACGG CATACGAGATCGTCTCGGCATCTCTGCTGAACCGCTCTTCCGATCT). After PCR amplification (24–28 cycles), the cDNA libraries were separated on a 2.5% agarose gel, excised and isolated with the Wizard SV Gel and PCR Clean-Up System (Promega).

**Sample preparation for mRNA-seq UT and mRNA-seq puro.** Untreated and puromycin-treated HeLa cells were cultivated as described above. Total RNA was isolated with the GenElute Mammalian Total RNA Mini-prep Kit (Sigma) according to the manufacturer's protocol. Library preparation for high-throughput sequencing was performed with the TruSeq RNA and DNA Sample Preparation Kits v2 (Illumina).

**High-throughput sequencing.** Prior to high-throughput sequencing, the quality of the iCLIP cDNA libraries was assessed on an Agilent 2100 Bioanalyzer with the 2100 expert High Sensitivity DNA Assay. Fifty-nucleotide-long single-end reads were obtained on an Illumina HiSeq 2000 sequencer.

**Mapping and annotation of reads to the human genome.** Prior to mapping of reads from the iCLIP samples to the genome, the first 14 nt were trimmed from obtained raw reads, owing to low sequence-quality scores. Mapping to the human genome (hg19, Feb. 2009) and annotation was done with the CLIPZ server<sup>48</sup>.

For subsequent analyses on the genome level, only reads that mapped uniquely to the genome were considered.

To infer iCLIP cross-link sites, that is, the sites of polymerase abortion, 8 nt were subtracted (+ strand) and added (– strand) to start and end coordinates of the genomic positions of mapped reads, respectively. Reads for which this led to inconsistency, with the iCLIP truncation site being placed on an exon and the mapped read being in an intron or vice versa, were excluded from subsequent analyses.

Annotation of protein-coding transcripts, 5' UTRs, CDS and 3' UTRs is based on the transcript data available at the CLIPZ server, whereas annotation of noncoding RNAs is based on GENCODE noncoding RNA genes v12 obtained from the UCSC genome browser (<http://genome.ucsc.edu/ENCODE/downloads.html>).

As for the iCLIP samples, reads obtained from mRNA-seq experiments were analyzed with the CLIPZ server. Because the applied mRNA-seq protocol did not preserve strand information, we did not consider orientation of the reads when assigning reads to genes and annotation categories.

To rank genes on the basis of their expression in HeLa cells, from the mRNA-seq data for each gene we calculated normalized expression values in the form of r.p.k.m. values<sup>61</sup>. The length of a gene was calculated as the sum of unique genomic exon positions of transcripts associated with the gene. The top 5,000 genes were selected on the basis of their averaged expression of the two mRNA-seq samples.

**Mapping of reads to representative transcripts.** For each gene, a representative transcript was selected, which was the longest RefSeq transcript or otherwise the longest non-RefSeq transcript. Reads that were marked by the CLIPZ server as bacterial, fungal, vector, viral or rRNA, or that mapped to multiple loci in the genome, were not considered for mapping to the transcript database. Mapping was done with Segemehl<sup>62</sup> with parameters -D 1 -A 90 -E 10000. The same procedure was also applied to the mRNA-seq and the eIF4AIII CLIP-seq data<sup>56</sup> (GEO GSM1001331).

**Generation of iCLIP site profiles along transcripts and 3' UTRs.** The top 5,000 expressed genes were selected for this analysis. For each gene, the representative transcript was divided into 100 bins. We computed the fraction of iCLIP reads associated with each bin for each transcript and then visualized the average fraction across transcripts as a function of the bin number. For the iCLIP profiles on 3' UTRs, we used only 3' UTRs with a length of at least 500 nt.

**Calculation of enrichment scores for pentamers.** iCLIP sites were ranked by abundance of reads, and the top 1,000 sites located in the CDS and the top 1,000 sites located in the 3' UTR of representative transcripts were chosen for further analysis. Next, the sequences of the regions of  $-25$  to  $+25$  nucleotides around the iCLIP sites were extracted. For the background set, we selected for each iCLIP site a randomly chosen region of the same length in the exon where the iCLIP site was placed. For each pentamer, we calculated the  $z$  score quantifying its enrichment in iCLIP regions compared to the control regions as  $(n-\mu)/\sigma$ , where  $n$  was the average count of the pentamer in windows around the cross-link sites, and  $\mu$  and  $\sigma$  were the average and s.d. of the count in randomized windows.

**Generation of sequence logos centered on iCLIP truncation sites.** iCLIP sites were ranked by abundance of reads, and the top 1,000 sites located in the CDS and the top 1,000 sites located in the 3' UTRs of representative transcripts were chosen for analysis. The sequences of the regions of  $-10$  to  $+10$  nucleotides centered on the iCLIP sites were extracted, and sequence logos were generated with WebLogo<sup>63</sup>. To infer background nucleotide frequencies, we selected for each iCLIP site a randomly chosen region of the same length in the exon where the iCLIP site was located. From these windows, we calculated the frequencies of each individual nucleotide, which were again visualized with WebLogo.

**Calculation of the distance to the closest neighboring iCLIP sites.** Analysis was performed for the representative transcripts of the top 5,000 expressed genes. For each iCLIP site inferred on these genes, we determined the distance to the closest iCLIP site located in the same region (CDS or 3' UTR) of the corresponding transcript. We then shuffled the cross-link sites in the corresponding transcript

regions, preserving the identity of the cross-linked nucleotide, and calculated similarly the distances for the randomized sites.

**Selection of an NMD target-gene control group.** A list of 172 NMD target genes that showed significant upregulation in cells depleted for UPF1 and in cells depleted for SMG6 was obtained<sup>14</sup>. Of these, 144 genes could be linked through a matching gene symbol or accession number to genes in our database. For the control group, we randomly selected 144 genes with approximately the same expression levels as in the NMD target genes group (at most 10% deviation from r.p.k.m. values of the NMD target genes). The same procedure was applied to a second set of NMD target genes from another study<sup>12</sup>. For this set, we were able to retrieve 75 genes by matching gene symbols or accession numbers. Statistical analysis was done in R (version 2.13.0).

**Generation of read profiles centered on eIF4AIII HITS CLIP peaks.** eIF4AIII CLIP-seq data were mapped to the transcriptome as described above. Peaks of eIF4AIII CLIP were inferred with FindPeaks<sup>64</sup> with the parameter settings described in the initial analysis of this data set<sup>62</sup>. Peaks were ranked by abundance of CLIP reads, and the top 5,000 peaks were selected for further analysis according to the following criteria. To avoid over-representation of specific transcripts, at most three peaks were selected per transcript. In addition, peaks had to be separated from each other by at least 250 nt.

Profiles of UPF1 iCLIP reads, RNA-seq reads and eIF4AIII CLIP reads were then generated by averaging the fraction of reads in a window  $\pm 250$  nt around the peak site, computed for each individual peak. To adjust for the read trimming (described above), iCLIP reads were extended by 7 nt at the start coordinate of the read. To rule out a bias by preferential amplification or mappability of certain reads, we used reads from the mRNA-seq UT sample as background and subtracted mRNA-seq read fractions from the CLIP-seq fractions for each position.

**Visualization of iCLIP reads along selected transcripts.** CLIP-seq and mRNA-seq data were mapped to the transcriptome as described above. Read counts were normalized by library size and scaled to a total of 1,000,000. For each transcript, we first generated a smoothed profile of 5' start sites of mRNA-seq reads along the transcript (window length  $\pm 25$  nt), then calculated the  $\log_2$  fold enrichment of iCLIP reads compared to the corresponding mRNA-seq profile. Plots were generated with R.

**Plasmids.** The p-SL mini  $\mu$  WT uA1 plasmid was generated by insertion of a double-stranded oligonucleotide encoding a stable stem-loop structure into the KpnI site of p-mini  $\mu$  WT uA1 (ref. 14). The inserted sequence was GGGTTCC GTCCAAGCACTGTTGAAGCAGGAAACCCGGGTTGCTAGTCGATCGAC TAGCAACCCGGGTTTCTCTGCTTCAACAGTGCTTGACGGAACCC.

**UPF1 RNA immunoprecipitation from mini- $\mu$  construct-expressing cells.** HeLa cells were grown to ~80% confluency in 150 cm<sup>2</sup> dishes and transfected with 5  $\mu$ g p- $\beta$ -mini  $\mu$  WT uA1 or p- $\beta$  SL mini  $\mu$  WT uA1 and 20  $\mu$ L DreamFect (OZ Biosciences) according to the manufacturer's protocol. Two days after transfection,  $1.2 \times 10^7$  cells were lysed on ice for 20 min in 1.2 mL hypotonic gentle lysis buffer supplemented with 40 U/mL RiboLock RI RNase inhibitor (Fermentas). Lysates were cleared by centrifugation at 16,000g in a microcentrifuge for 15 min. From 100  $\mu$ L lysate previously incubated with 20 U/mL TURBO DNase (Ambion) at 37 °C for 5 min, input RNA samples were extracted with TRI reagent according to a standard TRIzol protocol. Protein input samples were prepared by mixture of 50  $\mu$ L lysate with 50  $\mu$ L 2 $\times$  SDS loading buffer (200 mM DTT, 120 mM Tris-HCl, pH 6.8, 0.44% SDS, 20% glycerol and 0.25% bromophenol blue) for subsequent western blot analysis. Immunoprecipitation was performed by head-over-tail incubation of 0.5 mL lysate with 3.7  $\mu$ g of either goat anti-RENT1 (Bethyl, A300-038A) or goat anti-rabbit IgG (Bio-Rad, 172-1053)

antibody at 4 °C for 90 min. After preincubation of the lysates with the antibodies, 32  $\mu$ L of Dynabeads Protein G equilibrated in hypotonic gentle lysis buffer was added and incubated for additional 90 min. Precipitates were washed six times with 1 mL Net-2 buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, and 0.1% Triton X-100, supplemented with 1 $\times$  Halt Protease Inhibitor). After the washes, one-third of the beads were incubated with 40  $\mu$ L 2 $\times$  SDS loading buffer at 90 °C for 5 min to elute the protein for subsequent western blot analysis. After the remaining two-thirds of the beads were DNase treated (1 $\times$  TURBO DNase buffer, 2.5  $\mu$ L Turbo DNase and 2.5 U RiboLock RI RNase inhibitor in a final volume of 50  $\mu$ L) at 37 °C for 10 min, the coimmunopurified RNA was isolated with TRI reagent.

**UPF1 RIP with endogenous transcripts.** Lysates of untreated HeLa cells or cells treated for 6 h before lysis with 60.6  $\mu$ M puromycin or 2  $\mu$ g/mL harringtonine (Santa Cruz Biotechnology) were spiked with 2% of SL mini  $\mu$ -containing cell lysate, and UPF1 RIP was performed as described above.

**UPF1 RIP coupled with oligonucleotide-mediated RNase-H cleavage.** The RIP protocol was performed as published elsewhere<sup>45</sup>. Oligonucleotides for RNA cleavage and SYBR Green qPCR assays are listed in **Supplementary Table 2**.

**Quantitative real-time reverse-transcription PCR.** The entire recovered RNA of the RIP RNA samples, or 1  $\mu$ g of total RNA of the input RNA samples, was reverse transcribed in 50  $\mu$ L of StrataScript RT buffer containing 0.1 mM DTT, 0.4 mM dNTPs, 300 ng random hexamers and 1  $\mu$ L of StrataScript Multi-Temp reverse transcriptase (Agilent) according to the manufacturer's protocol. Twelve microliters of the reverse-transcription reactions were amplified with Brilliant III Ultra-Fast qPCR Master Mix (Agilent) and corresponding TaqMan assays in a total volume of 15  $\mu$ L with the Rotor-Gene 6000 rotary analyzer (Corbett). The primers and TaqMan probes to measure SMG5 and mini- $\mu$  mRNA levels are described elsewhere<sup>14</sup>. To measure endogenous mRNA levels, the RT-qPCR protocol was adapted for the Brilliant II Fast SYBR Green QPCR Master Mix (Agilent) according to the manufacturer's manual.

**Immunoblotting.** Cell lysates corresponding to  $2 \times 10^5$  cells and one-third of the RIP precipitates were separated by 10% SDS-PAGE. Proteins were transferred to Optitran BA-S 85 reinforced nitrocellulose membrane (Schleicher and Schuell) and probed with 1:3,000-diluted polyclonal goat anti-RENT1 antibody (Bethyl, A300-038A), 1:5,000-diluted polyclonal rabbit anti-actin antibody (A5060, Sigma-Aldrich), 1:500-diluted AffiniPure goat anti-mouse IgM,  $\mu$  chain (Jackson ImmunoResearch) or 1:5,000-diluted polyclonal rabbit anti-CPSF73 antibody<sup>65</sup>. Donkey anti-rabbit IRDye 800CW and donkey anti-goat IRDye 800CW (LI-COR, 926-32213 and 926-32214, respectively), diluted 1:10,000, were used as secondary antibodies, and membranes were scanned on an Odyssey Infrared Imager.

Original images of autoradiographs and blots used in this study can be found in **Supplementary Fig. 5**.

61. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
62. Hoffmann, S. *et al.* Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* **5**, e1000502 (2009).
63. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
64. Fejes, A.P. *et al.* FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**, 1729–1730 (2008).
65. Jenny, A., Minvielle-Sebastia, L., Preker, P.J. & Keller, W. Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* **274**, 1514–1517 (1996).