Supplementary Information:

Nonsense-mediated mRNA Decay Substrates Involve Distinct and Successive Upf1-bound Complexes

Marine Dehecq, Laurence Decourty, Abdelkader Namane, Caroline Proux, Joanne Kanaan, Hervé Le Hir, Alain Jacquier, Cosmin Saveanu

Access to primary data (mass-spectrometry and RNASeq):

RNASeq data. GEO accession GSE102099

MS data. MS data are available via ProteomeXchange with identifier **PXD007159**.

<u>Contents :</u>

Supplementary Figures :

Fig. S1 - Workflow for quantitative analysis of MS/MS results from affinity purified complexes.
Fig. S2 - N-terminal and C-terminal tagged Upf1 enrich similar sets of specific proteins.
Fig. S3 - Controls of total tagged protein levels in the presence or absence of other NMD components.
Fig. S4 - Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences.

Fig. S5 - Deletion of NMD4 and EBS1 stabilize a set of transcripts that is also stabilized in the absence of UPF1.

Fig. S6 - The helicase domain of Upf1 alone can destabilize RPL28 pre-mRNA, an NMD reporter. Fig. S7 - Comparison between the canonical SURF/DECID model and our extended Detector/Effector model.

Strains, oligonucleotides and plasmid tables.

Supplementary tables:

Table **S1** - LTOP2 values for the proteins identified in affinity purified samples Table **S2** - Average LTOP2 values and standard deviation of the results from independent

Table **S2** - Average LTOP2 values and standard deviation of the results from independent replicate experiments.

Table **S3** - Enrichment values in purification over total protein content (log2 transformed values). Table **S4** - Average enrichment values and standard deviation of the results from independent replicate experiments.

Table **S5** - Summary of the number of replicates, the number of proteins robustly quantified and the efficiency of RNase treatment for each purification type.

Table **S6** - RNASeq raw data analysis results summary.

Note: tables **S1** to **S4** are provided as a single separate Excel Office Open XML file, and **S5** and **S6** as individual spreadsheet files.



Fig. S1. Workflow for quantitative analysis of MS/MS results from affinity purified complexes.

Notes about the depicted procedure: 1. MaxQuant output for peptide intensities and their association with a given protein was used as the main input for computing LTOP2 scores and enrichment. To calculate the false discovery rate (FDR) of the MS/MS analysis, MaxQuant builds reverse sequence « artificial » proteins that serve as negative controls for the identification procedure. Reversed sequences and common contaminants (trypsin, keratins) were removed from further analyses in the early steps of the analysis.

2. A protein group corresponds to a single protein or several proteins with very high sequence similarity that cannot be discriminated by peptide analysis. For further analyses, we used the identity of the protein of the group with most coverage.

3. The TEV protease was added in each purification experiments with the same relative amount to elute complexes from beads. This step is important to be able to compare replicates between them and the different purification types.

4. The comparison of our input LTOP2 with the abundance data from Ho et al. 2017, allowed to calculate a factor that served to adjust LTOP2 values and make them compatible with the dynamic range and scale of published abundance values.



Fig. S2. N-terminal and C-terminal tagged Upf1 enrich similar sets of specific proteins. **a)** Estimation of the levels of overexpression for N-terminal tagged Upf1 fragments, in comparison with chromosomally C-terminal tagged protein. G6PDH was used as a loading control. Serial dilutions were used to test the ability of the immunoblot signal to estimate protein levels. **b)** Enrichment values for purifications done with chromosomally C-terminal tagged Upf1 (x axis) and N-terminally TAP tagged Upf1.



Fig. S3. Controls of total tagged protein levels in the presence or absence of other NMD components. Total protein extracts from the described strains were tested by immunoblot to detect the protein A part of the TAP tag. G6PDH was used as a loading control.

| а | | | |
|---|--------------------------|---|------------------|
| hSmg6 hSmg5 yNmd4 | 1239 849 1 | 1 10 20 30 40 50 50 70 80 MELEIRPLFLVPD- NGFIDHLASLARLLESRKYILVVPLIVINELDGLAKGQETDHRAGGYARVVQEKARKSIEFL 13 AQSAMSP-YLVPDTQALCHHLPVIRQLATSGRFIVIIPRTVIDGLDLLKKEHPGARDGIRYL 90 MTQYNFIIDASAFEKGLGNIKRWCSDCTEAVTLNFYIPTFTLNELDFLQQRRKSFAARESLKFI 65 | 13 9 |
| hSmg6 hSmg5 yNmd4 | 1314 910 66 | 81 90 100 110 120 130 140 150 160 EQRFESR DSCLRALTSR GNELESIAFRSEDITGQLGNNDDLILSCCLHYC SCCLHYC 130 EAEFKKG NRYIRCOKEV GKSFERHKLK RO-DADAWTLYKILDSC KQL 95 DRLDDSKFANLKVFIEFPEVLDIILWSDVMEH NDSSG KI N IAKLYKLKNLKSCIYKCYLEGNE 129 | 66 5 9 |
| hSmg6 hSmg5 yNmd4 | 1367 956 130 | 161 170 180 190 200 210 220 230 240 AKDFMPASKEEPIRLLREVVLLTDRNLR VKALTRNVPVRD 141 TLAQGAG-EE-DPSGMVTIITGPLDNPSVL SGPMQAALQAAAHAS 99 GLH-WILSEDPQIREMAMQCNIPSCSIVDVDSILSKDMNDKSFRESEKFNNMMLKNG-TKEESEN 193 | 07 9 3 |
| hSmg6 hSmg5 yNmd4 | 1408 1000 194 | 241 250 260 268 VDIKNVLDFYKQWKEIG 1418 GREIIRTNFNKTVYASRGT-GELWSP 218 | |
| | | | |
| D hSmg6 [14-3-3] hSmg5 [14-3-3] hSmg7 [14-3-3] yEbs1 [14-3-3] | 576 19 1 23 | 10 20 30 40 50 60 70 80 QQQELHRLLRVADNQELQLSNLLSRDRI-SPEGLEKMAQLRAELLQLYERCILL-DIEFSDNQNVDQIL- 64 TKRLYRAVVEAVHRLDLILCNKTAYQEVFKPENISLRNKLRELCVKLMFLHPVDYGRKAEELL-81 MSLQSAQYLRQAEVLKADMTDSKLGPAEVWTSRQALQDLYQKMLVT-DLEYALDKKVEQDL-60 SILKSNQLFQDYALLNGFLAFVHSKLNAAILTSIESQCGKSFAA-DLDSFDQSSISSILD 59 | 2 |
| hSmg6 [14-3-3] hSmg5 [14-3-3] hSmg7 [14-3-3] yEbs1 [14-3-3] | 643 82 61 60 | 90 100 110 120 130 140 150 160 WKNAFYQVI EKFR | 7 4 5 9 |
| hSmg6 [14-3-3] hSmg5 [14-3-3] hSmg7 [14-3-3] yEbs1 [14-3-3] | 698 135 116 140 | 170180190200210220230240KLEDYMDGLAIRSKPUSASGKEMDWAQMACHRCLVYLGDLSRYQEQASDTANYGKARSWY75ELOCCIDWTHVTDPLIGCK | 4059 |
| hSmg6 [14-3-3] hSmg5 [14-3-3] hSmg7 [14-3-3] yEbs1 [14-3-3] | 755 201 176 210 | 250 260 270 280 290 300 LKAQHIAPKNGRPYNQLALLAVYTRRKLDAVYYYMRSLAASNPILTAKESLMSLFEETKRK815 YQALSVAPQIGMPFNQLGTLAGSKYYNVEAMYCYLRCIQSEVSFEGAYGNLKRLYDKAAKMYHQL265 RHAAQLVPSNGQPYNQLAILASSKGDHLTTIFYYCRSIAVKFPFPAASTNLQKALSKALES236 RLASLVLPSAGETYSQAGAIFLQTGNLGIAVFNFVKGMMTKMPSPVSIKNFGALMVDNK291 | |
| • | | | |
| hSmg5 [HHR] hSmg7 [HRR] yEbs1 [292-585] | 267 237 292 | 10 20 30 40 50 60 70 80 KKCETRKLSPGKKRCKDIKRLLVNFMYLQSLLQPKSSSVDSELTSLCQSVLEDFNLCLFYLPSSPN 33 RDEVKTKWGVSDFIKAFIKFHGHVYLSKSLEKLSPLREKLEEQFKRLLFQKAFNSQQLVHVTVINL- 30 RDEVKTKWGVSDFIKAFIKFHGHVYLSKSLEKLSPLREKLEEQFKRLLFQKAFNSQQLVHVTVINL- 30 | 3 2 8 |
| hSmg5 [HHR] hSmg7 [HRR] yEbs1 [292-585] | 334 303 349 | 90 100 110 120 130 140 150 160 LSLASEDEEEYESGYA-FLPDL-LIFOMVIICLMCVHSLERAGSKQYSAAIAFTLALFSHLVNHVN 39 FQLHHLRDFSNETEOHTYSQDEQLCWTQLLALFMSFLGILCKCPLQNES-35 NNGIKLRHLENALYETMSARYLNNIKTIFHNLIITIGGFHLLKRRSDVSAKTLKDLRSNELDYLNFAFKYIAHILNDI-42 | 727 |
| hSmg5 [HHR] hSmg7 [HRR] yEbs1 [292-585] | 398 353 428 | 170 * 180 190 200 210 220 230 240 IRLQAELEEGENPVPVLMAEGLLPAVKVFLDWLRTNPDLIIVCAQSSQSLWNRLSVLLNLLPAAG 69 QEESYNAYPLPAVKVSMDWLRLRP-RVFQEAVVDERQYIWPWLISLLNSF | 9 |
| hSmg5 [HHR] hSmg7 [HRR] yEbs1 [292-585] | 700 402 497 | 250 260 270 280 290 300 310 320 | 543 |
| hSmg5 [HHR] hSmg7 [HRR] yEbs1 [292-585] | 766 909 65 | 330 340 350 360 VRICCIRSFGHFIARLQGSILOFNPEVGIFVSIAQSEQESLLQQA 809 IRQQRLISIGKWIADNQPRLIQCENEVGKLLFITEIPELILEDPS 499 * (410-647) for hSmg5 LRLQAVVNISSQLLQ-NNNCGVEWSDNKSRYIFN 585 | |

Fig. S4: Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences. Alignment of the PIN domains of hSmg6, hSmg5 and ScNmd4 (a), the 14-3-3 domains of hSmg6, hSmg5, hSmg7 and Ebs1 (b) and of the helical hairpin region (HHR) of hSmg5, hSmg7 and Ebs1 (292-585) (c). These alignments were obtained with the algorithm Mafft with default parameters; colour represents percentage of identity or similarity (BLOSUM62).



Fig. S5. Deletion of *NMD4* and *EBS1* stabilize a set of transcripts that is also stabilized in the absence of *UPF1*. **a**) Workflow used for RNAseq experiments and analyses. (B) to (G) Examples of NMD substrates sequencing profiles in WT, *upf1* Δ , *nmd4* Δ and *ebs1* Δ experiments. NMD substrates belong to different classes, intron containing (RPL28; **b** and **c**), uORF (DAL7, DAL2; **d** and **e**), non-coding RNA (SUT439; **f**) and long 3'UTR (YOR304C-A; **g**). Profiles were normalized using the samples median counts. For RPL28 (**b** and **c**), we represented the profile of the entire transcript showing that the signal in the exon is similar in each strain (**b**). A zoom of the intron region (**c**) shows a higher signal in this specific region of unspliced RPL28 in mutants by comparison of WT. H. Scatter plot of transcript log2 fold change in *upf1* Δ against *ebs1* Δ . The dashed line represents the limit over which RNA are considered as stabilized, a value of 0.5 in log2 (1.4 fold change).



Fig. S6. The helicase domain of Upf1 alone can destabilize RPL28 pre-mRNA, an NMD reporter. Total RNA from wild-type or $upf1\Delta$ strain transformed with an empty plasmid (pControl) or plasmids expressing various Upf1 fragments (see Figure 3A) was tested by reverse transcription and quantitative PCR. The levels of RPL28 pre-mRNA were normalized using an NMD-insensitive transcript (RIM1) and an NMD efficiency score was calculated based on the difference between a wild type and a $upf1\Delta$ strain.

a SURF/DECID NMD model

b Extended Detector/Effector model



Fig. S7. Comparison between the canonical SURF/DECID model (a) and our extended Detector/Effector model (b) for NMD. Orange and blue squares mark equivalent steps in both models. Light grey elements in the revised model represent optional steps that can further enhance the NMD process under certain conditions and in specific organisms.

Strains used in this study

Dehecq et al., Detection and degradation of nonsense-mediated mRNA decay substrates involve two distinct Upf1-bound complexes

| strain | original strain | Genotype | Reference |
|-------------------------------|-----------------|---|---------------------------|
| LMA2154 (BY4741) | | MATa ura3∆0 his3∆1 leu2∆0 met15∆0 | Brachmann et al. 1998 |
| LMA2155 (BY4742) | | MATa $ura3\Delta 0$ his $3\Delta 1$ leu $2\Delta 0$ lys $2\Delta 0$ | Brachmann et al. 1998 |
| LMA2194 (UPF1-TAP) | BY4741 | NAM7-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| LMA3730 (UPF1-TAP) | BY4741 | NAM7-CRAP::URA3 | This study |
| LMA2192 (UPF2-TAP) | BY4741 | NMD2-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| LMA3731 (UPF2-TAP) | BY4741 | NMD2-CRAP::URA3 | This study |
| LMA2193 (UPF3-TAP) | BY4741 | UPF3-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| LMA4263 (UPF3-TAP) | BY4741 | UPF3-CRAP::URA3 | This study |
| LMA3317 (NMD4-TAP) | BY4741 | NMD4-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| LMA3728/3729 (NMD4-TAP) | BY4741 | NMD4-CRAP::URA3 | This study |
| LMA4264 (DCP1-TAP) | BY4741 | DCP1-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| (EBS1-TAP) | BY4741 | EBS1-CRAP::URA3 | This study |
| LMA3312 (HRR25-TAP) | BY4741 | HRR25-TAP::HIS3MX | Ghaemmaghami et al., 2003 |
| LMA3849 (UPF1-TAP/NMD4-HA) | BY4741 | NAM7-CRAP::URA3 / NMD4-HA::KANMX6 | This study |
| LMA3851 (UPF1-TAP/EBS1-HA) | BY4741 | NAM7-CRAP::URA3 / EBS1-HA::KANMX6 | This study |
| LMA3852 (UPF1-TAP/EDC3) | BY4741 | NAM7-CRAP::URA3 / EDC3-HA::KANMX6 | This study |
| LMA1667 (upf1∆) | BY4741 | NAM7::KANMX6 | Giaever et al., 2002 |
| LMA1669 (upf2∆) | BY4741 | NMD2::KANMX6 | Giaever et al., 2002 |
| LMA1671 (upf3∆) | BY4741 | UPF3::KANMX6 | Giaever et al., 2002 |
| LMA3732 (nmd4∆) | BY4741 | NMD4::KANMX6 | Giaever et al., 2002 |
| LMA4112 (UPF2-TAP/upf1∆) | BY4741 | NAM7::KANMX6 / NMD2-CRAP::URA3 | This study |
| LMA4113 (UPF3-TAP/upf1∆) | BY4741 | NAM7::KANMX6 / UPF3-CRAP::URA3 | This study |
| LMA4114 (NMD4-TAP/upf1∆) | BY4741 | NAM7::KANMX6 / NMD4-CRAP::URA3 | This study |
| LMA3739 (UPF1-TAP/upf2∆) | BY4741 | NMD2::KANMX6 / NAM7-CRAP::URA3 | This study |
| LMA4701/4702 (NMD4-TAP/upf2∆) | BY4741 | NMD2::KANMX6 / NMD4-CRAP::URA3 | This study |
| LMA3735 (UPF1-TAP/upf3∆) | BY4741 | UPF3::KANMX6 / NAM7-CRAP::URA3 | This study |
| LMA3736 (UPF1-TAP/nmd4∆) | BY4741 | NMD4::KANMX6 / NAM7-CRAP::URA3 | This study |
| LMA4523 (upf1∆/upf2∆) | BY4741 | NAM7::HIS3MX / NMD2::KANMX6 | This study |
| LMA4524 (upf1∆/upf3∆) | BY4741 | NAM7::HIS3MX / UPF3::KANMX6 | This study |
| LMA4525 (upf1∆/ebs1∆) | BY4741 | NAM7::HIS3MX / EBS1::KANMX6 | This study |
| LMA4678 (upf1∆/nmd4∆) | BY4741 | NAM7::HIS3MX / NMD4::KANMX6 | This study |
| LMA3853/3854 (nmd4∆/ebs1∆) | BY4742 | NMD4::ProMFalpha2NAT / EBS1::KANMX6 | This study |

Plasmids used in this study

Dehecq et al., Detection and degradation of nonsense-mediated mRNA decay substrates involve two distinct Upf1-bound complexes

| Plasmid pRS316 | Description and ID (yeast marker) pRS316 (URA3) | Reference Sikorski & Hieter, 1989 |
|--------------------------|--|--------------------------------------|
| pCM189-NTAP | pl.1233 (H1) (URA3) | This study |
| pCM189-NTAP-NAM7-FL | pl.1442 (TAP-UPF1-FL) (URA3) | This study |
| pCM189-NTAP-NAM7-CH | pl.1443 (TAP-UPF1-CH) (URA3) | This study |
| pCM189-NTAP-NAM7-NoCH | pl.1444 (TAP-UPF1-Cter) (URA3) | This study |
| pCM189-NTAP-Upf1 2-853 | pl.1521 (TAP-UPF1-NoSQ) (URA3) | This study |
| pCM189-NTAP-Upf1 208-853 | pl.1522 (TAP-UPF1-Cter-NoSQ) (URA3) | This study |
| pDEST14-NAM7 | pl.1350 – Gateway destination vector | This study |
| pDONR201-NAM7 | pl.1330 – Gateway source vector | This study |
| pCRBlunt-CRAP(6-HisTAP) | pl.1287 – TAP to CRAP cassette vector | This study |

Oligonucleotides used in this study

| Oligonucleotide | Usage | Sequence |
|--|------------------------------|--|
| CS887_fw_RPL28intron | real time PCR forward primer | CCATCTCACTGTTGAGACGG |
| CS888_rv_RPL28intron | real time PCR reverse primer | CTCAGTTTGCGATGGAAGAG |
| CS889_rv_RPL28exon2 | real time PCR reverse primer | ATGTTGACCACCGGCCATAC |
| CS946_RPL28_fw_exonoverlapQPCR | real time PCR forward primer | TCACGTCTCAGCCGGTAAAG |
| CS1076_fw_RIM1Qex1ex2 | real time PCR forward primer | GTTAGAAAAGGCGCTTTGGTATATG |
| CS1077_rv_RIM1QRTex2 | real time PCR reverse primer | AACCGTCGTCTCTCTCGAAG |
| CS1429_DAL7_fwQ | real time PCR forward primer | TGAAACTTTGCCAGCGGCCTTC |
| CS1430_DAL7_rvQ | real time PCR reverse primer | TCCCAACGACCACAGTTCAAACC |
| CS1359_fw_NAM7_2_pTM189Not | construction oligonucleotide | ttaagaaaatctcatcctccggggcacttGATgcgG TCGGTTCCGGTTCTCACAC |
| CS1361_rv_NAM7_208_pTM189Notv1 | construction oligonucleotide | ATAACTAATTACATGATGCGGCCCTC CTGCAGGGCTTAATTGGATCTCCATT TTGCCTC |
| CS1362_fw_NAM7_s208_pTM189Not | construction oligonucleotide | ttaagaaaatctcatcctccggggcacttGATgcgA ATAAAGACGCTACAATTAATGATATT GACG |
| CS1364_rv_NAM7_971_pTM189Notv2 | construction oligonucleotide | ATAACTAATTACATGATGCGGCCCTC CTGCAGGGCTTATATTCCCAAATTGC TGAAGTC |
| CS1393_rv_UPF1_853STOP_pTM189 Notv2 | construction oligonucleotide | ATAACTAATTACATGATGCGGCCCTC CTGCAGGGCTTActgaggacgaactaattga ac |