





# Wobble tRNA modification and hydrophilic amino acid patterns dictate protein fate

Francesca Rapino<sup>1,2,3,4,14</sup>✉, Zhaoli Zhou<sup>1,2,3,5,14</sup>, Ana Maria Roncero Sanchez<sup>1,2,3,14</sup>, Marc Joiret<sup>1,2,3,6</sup>, Christian Seca<sup>1,2,3</sup>, Najla El Hachem<sup>1,2,3</sup> , Gianluca Valenti<sup>3,7</sup>, Sara Latini<sup>1,2,3</sup> , Kateryna Shostak<sup>2,3,8</sup>, Liesbet Geris<sup>2,3,6</sup> , Ping Li<sup>5</sup>, Gang Huang<sup>5</sup>, Gabriel Mazzucchelli<sup>2,3,9</sup>, Dominique Baiwir<sup>2,3,9</sup> , Christophe J. Desmet<sup>2,3,10,11</sup>, Alain Chariot<sup>2,3,8,12</sup>, Michel Georges<sup>2,3,11,13</sup> & Pierre Close<sup>1,2,3,12</sup>✉

Regulation of mRNA translation elongation impacts nascent protein synthesis and integrity and plays a critical role in disease establishment. Here, we investigate features linking regulation of codon-dependent translation elongation to protein expression and homeostasis. Using knockdown models of enzymes that catalyze the mcm<sup>5</sup>s<sup>2</sup> wobble uridine tRNA modification (U<sub>34</sub>-enzymes), we show that gene codon content is necessary but not sufficient to predict protein fate. While translation defects upon perturbation of U<sub>34</sub>-enzymes are strictly dependent on codon content, the consequences on protein output are determined by other features. Specific hydrophilic motifs cause protein aggregation and degradation upon codon-dependent translation elongation defects. Accordingly, the combination of codon content and the presence of hydrophilic motifs define the proteome whose maintenance relies on U<sub>34</sub>-tRNA modification. Together, these results uncover the mechanism linking wobble tRNA modification to mRNA translation and aggregation to maintain proteome homeostasis.

<sup>1</sup>Laboratory of Cancer Signaling, University of Liège, Liège, Belgium. <sup>2</sup>GIGA-Institute, University of Liège, Liège, Belgium. <sup>3</sup>University of Liège, Liège, Belgium. <sup>4</sup>Department of Radiation Oncology (Maastricht), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands. <sup>5</sup>Shanghai Key Laboratory of Molecular Imaging, Shanghai University of Medicine and Health Sciences, Shanghai, China. <sup>6</sup>Biomechanics Research Unit, University of Liège, Liège, Belgium. <sup>7</sup>Unité de Recherche Transitions, University of Liège, Liège, Belgium. <sup>8</sup>Laboratory of Medical Chemistry, University of Liège, Liège, Belgium. <sup>9</sup>Mass Spectrometry Laboratory, System Biology and Chemical Biology, University of Liège, Liège, Belgium. <sup>10</sup>Laboratory of Cellular and Molecular Immunology, University of Liège, Liège, Belgium. <sup>11</sup>Faculty of Veterinary Medicine, University of Liège, Liège, Belgium. <sup>12</sup>WELBIO, University of Liège, Liège, Belgium. <sup>13</sup>Unit of Animal Genomics, University of Liège, Liège, Belgium. <sup>14</sup>These authors contributed equally: Francesca Rapino, Zhaoli Zhou, Ana Maria Roncero Sanchez. ✉email: [Francesca.Rapino@uliege.be](mailto:Francesca.Rapino@uliege.be); [Pierre.Close@uliege.be](mailto:Pierre.Close@uliege.be)

Adequate regulation of mRNA translation rate and speed is essential for cellular homeostasis. mRNA translation impacts protein output, but it also regulates co-translational processes that assist nascent polypeptides. The speed of ribosomes on mRNA during translation is not uniform and an increasing amount of evidence has demonstrated that translation kinetics play a prominent role in proteostasis by impacting co-translational protein folding, translation fidelity, or protein quality control pathways<sup>1–9</sup>. In addition, cells must not only promote accurate folding but also have to prevent accumulation of misfolded proteins that may arise from errors in translation, aberrant mRNAs, or defects in the chaperone machinery<sup>10,11</sup>.

Local translation elongation rate is tightly regulated at the level of individual codons through the integration of different layers, including tRNA abundance and modification, tRNA selection, and codon usage. In this context, wobble uridine ( $U_{34}$ ) tRNA modifications, including the addition of both methoxycarbonylmethyl ( $mcm^5$ ) and thiol ( $s^2$ ), are crucial for the decoding of specific codons, mainly AA-ending codons (i.e., AAA, GAA, CAA)<sup>12–19</sup>. Consequently, loss of these modifications was shown to lower expression of functionally important proteins whose mRNA is enriched in such codons, a process that can be rescued by the replacement of AA-ending codons by synonymous,  $U_{34}$ -tRNA modification independent, AG-ending codons<sup>19,20</sup>. The enzymatic pathway leading to the addition of the  $mcm^5s^2$  modification at  $U_{34}$ -tRNAs ( $U_{34}$ -enzymes) encloses the acetyltransferase complex Elongator, a six subunit complex (ELP1–6) where ELP3 harbors the catalytic activity, the methyltransferase ALKBH8 (Alkylation repair homolog 8) and the thiouridylase complex composed of CTU1/2 proteins (Cytosolic thiouridylase proteins 1/2)<sup>12,14,15,21,22</sup>. Ribosome profiling experiments demonstrate that loss of  $U_{34}$ -enzymes leads to ribosome accumulation (i.e., translation elongation defects) mainly at AA-ending codons (i.e., AAA, GAA, CAA)<sup>16–19</sup>. However, the exact consequences of this codon-specific translation pausing on subsequent protein expression and homeostasis still remain elusive. Recent studies showed that loss of  $U_{34}$ -enzymes leads to increase protein aggregation<sup>17,19</sup>. However, the mechanisms linking codon-dependent translation dysregulation to protein aggregation remain poorly understood.

tRNA modifications emerge as key players in development and cancer. In particular, loss of the  $U_{34}$ -enzyme *Elp3* in murin cortical progenitors results in neurogenesis defects and microcephaly through induction of an unfolded protein response (UPR)<sup>18</sup>. Also, altered proteostasis consequent to *Elp3* loss affects the development of the cochlea and results in deafness<sup>23</sup>. Mutations in Elongator genes are associated with neurodegenerative disorders<sup>24</sup>. In cancer, we and others have recently uncovered a key role of  $U_{34}$ -enzymes in tumor development, metastasis and resistance to therapy by promoting efficient codon-specific translation. ELP3 was shown to drive Wnt-dependent intestinal tumor initiation by maintaining a subpopulation of *Lgr5*<sup>+</sup>/*Dclk1*<sup>+</sup>/*Sox9*<sup>+</sup> cancer stem cells<sup>25</sup>. Also,  $U_{34}$ -enzymes regulate breast cancer metastasis through the translation of the pro-invasive DEK oncoprotein<sup>20</sup>. Furthermore,  $U_{34}$ -enzymes promote resistance to targeted therapy in melanoma, through regulation of HIF1A mRNA translation in a codon-dependent manner<sup>19</sup>. Finally, Elongator and other  $U_{34}$ -enzymes were identified as a key determinant of gemcitabine sensitivity in gallbladder cancer by regulating the mRNA translation of hnRNPQ<sup>26</sup>. Together, these evidences highlight the role of  $U_{34}$ -enzymes in cancer by promoting the mRNA translation of key oncoproteins<sup>22</sup>. However, the extent of the proteome subset whose expression relies on the  $U_{34}$ -tRNA modification pathway remains to be determined.

Here, we took advantage of loss-of-function models of  $U_{34}$ -enzymes to investigate the impact of codon-dependent

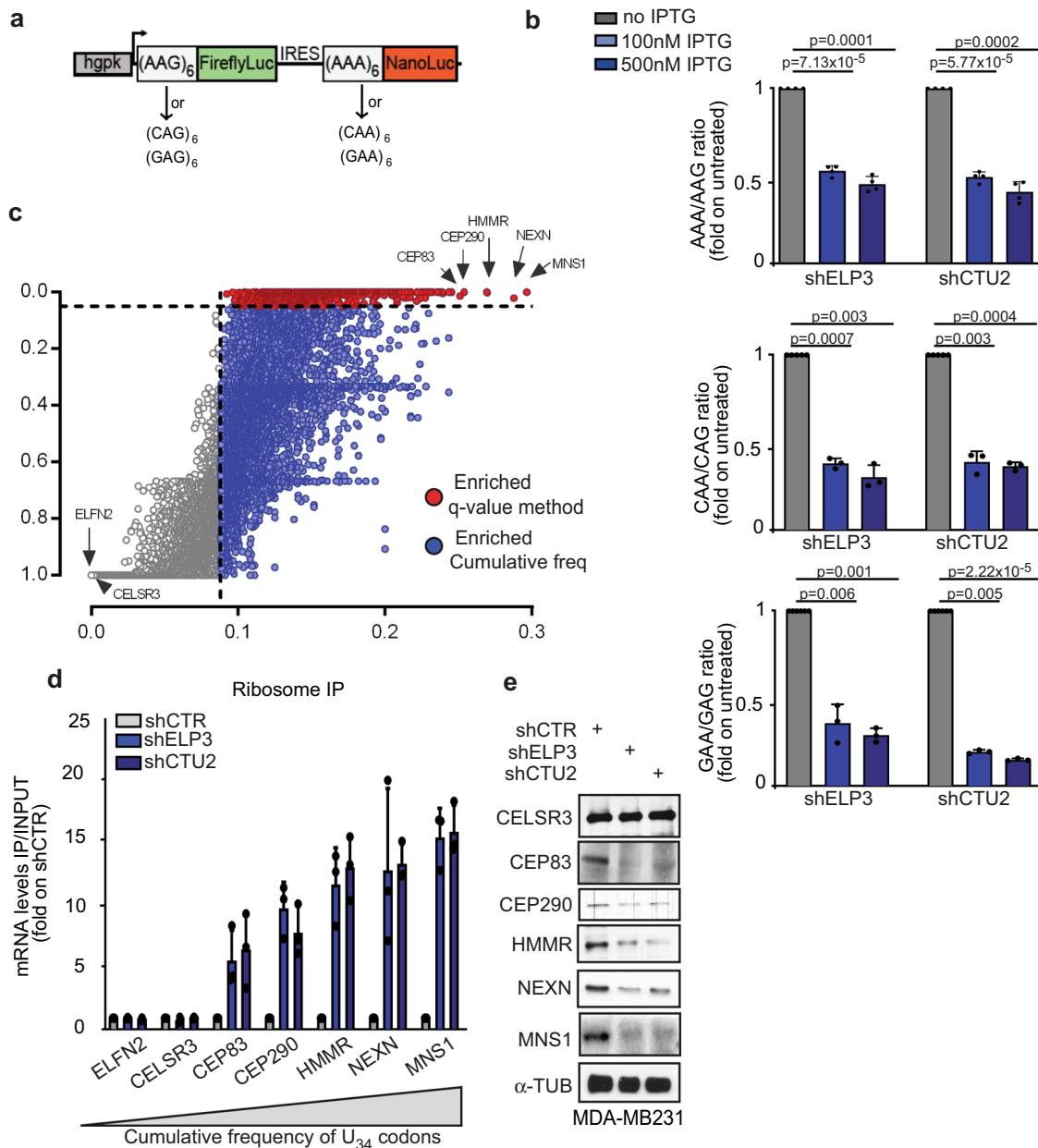
regulation of translation elongation on proteome expression and maintenance. Surprisingly, we find that gene codon content is necessary but not sufficient to predict protein fate upon defective  $U_{34}$ -tRNA modification. Patterns of hydrophilic pentasequences link defects in codon-dependent translation elongation to protein aggregation and subsequent degradation.

## Results

**Analysis of  $U_{34}$ -codons frequency across the human orfeome uncovers translational targets of the  $U_{34}$ -tRNA modification pathway.** Biological phenotypes associated with modulation of the  $mcm^5s^2$  wobble uridine tRNA modification have been associated with decoding of specific codon during translation elongation, mainly of three codons AAA, GAA, and CAA ( $U_{34}$ -codons)<sup>16–19</sup>. We designed dual-luciferase constructs encoding six in frame repetitions of  $U_{34}$ -codons (xAA) upstream of NanoLuc and six in frame repetitions of the corresponding synonymous codon (xAG), whose decoding does not require  $U_{34}$ -tRNA modification, upstream of FireflyLuc (Fig. 1a and Supplementary Fig. 1a). Depletion of the  $U_{34}$ -enzymes, ELP3, or CTU2, strongly decreased NanoLuc activity but did not impact FireflyLuc activity (Fig. 1b and Supplementary Fig. 1b–d), and no effect was seen using a construct lacking the codon repetitions (Supplementary Fig. 1e–f). Together with previous evidences<sup>19,20,27–29</sup>, these data indicate that decoding the -AA ending codons requires  $U_{34}$ -enzymes.

In order to identify the mechanisms linking codon-dependent translational optimization and protein fate, we computed the frequency bias across the human genome of the three  $U_{34}$ -codons (i.e., AAA, GAA, and CAA). To this end, we calculated a  $q$  value for each  $U_{34}$ -codon in every gene of the human orfeome in order to define codon enrichment in a statistical manner ( $q < 0.05$ , see the “Methods” section). Interestingly, 4331 genes display a high frequency of at least one of the three codons ( $q$  value  $< 0.05$ ; Supplementary Fig. 1g), among which 608 genes are strongly enriched in all three codons (Fig. 1c and Supplementary Fig. 1h). In line with previously published evidence in *S. cerevisiae*, genes enriched in these codons significantly cluster in biological pathways related to the cell cycle<sup>30,31</sup> (Supplementary Fig. 1i, Supplementary Data1) as well as in specific protein families, including kinesins (Supplementary Fig. 1j). As expected, depletion of ELP3 and CTU2, two of the  $U_{34}$ -enzymes<sup>12,13,22</sup>, leads to a strong increase in ribosome occupancy (i.e., defect in translation elongation) at the top five transcripts enriched in  $U_{34}$ -codons (i.e., CEP83, CEP290, HMMR, NEXN, and MNS1; Fig. 1d and Supplementary Fig. 1k). This correlates with a strong decrease in their protein expression levels (Fig. 1e). No difference in the expression levels of the corresponding mRNAs was detected (Supplementary Fig. 1l). These data indicate that high abundance of  $U_{34}$ -codons in transcripts predicts dependency toward  $U_{34}$ -enzymes for correct mRNA translation and corresponding protein expression.

**mRNA codon content is not sufficient to predict protein abundance upon loss of the  $U_{34}$ -enzymes.** In order to assess if the enrichment of  $U_{34}$ -codons in mRNAs is sufficient to predict their protein expression levels upon loss of  $U_{34}$ -enzymes, we performed RNA-seq and label-free quantitative proteomics upon ELP3 depletion (Fig. 2a and Supplementary Fig. 2a–b). Surprisingly, even though downregulated proteins were significantly enriched in  $U_{34}$ -codons in their corresponding mRNA as compared to the genome ORFs content ( $\chi^2$  test,  $p = 3.79 \times 10^{-19}$ ; Fig. 2a and Supplementary Data 2), a large number of genes enriched in  $U_{34}$ -codons remained unchanged in their protein expression (Fig. 2b and Supplementary Fig. 2c). Of note, no major

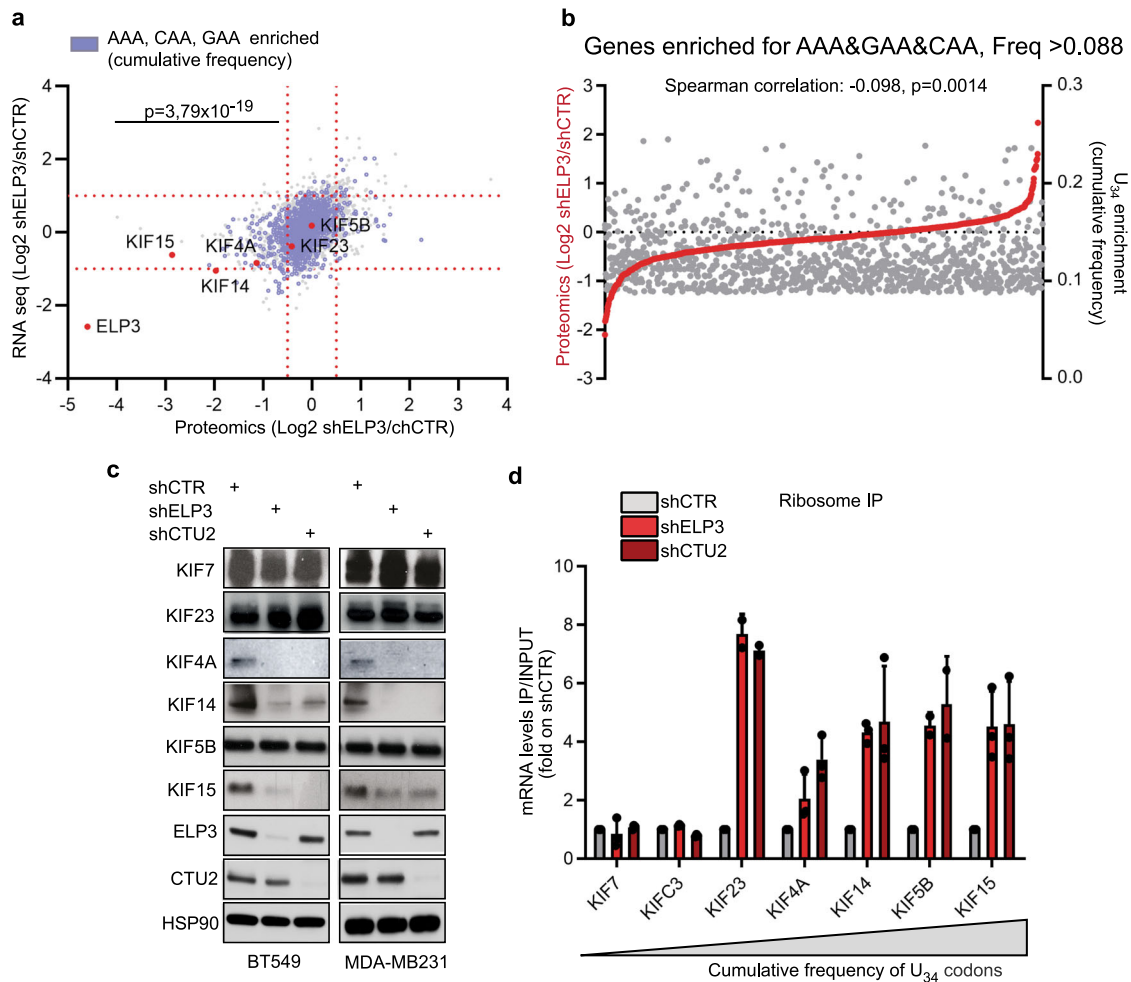


**Fig. 1 mRNA codon content predicts translation dependency on U<sub>34</sub>-enzymes.** **a** Schematic representation of the dual-luciferase construct. **b** NanoLuc/Firefly luminescence ratio of MDA-MB231 cells stably expressing the indicated constructs and depleted or not of ELP3 or CTU2 upon IPTG treatment. Data are normalized on non-treated conditions (IPTG 0 μM, n = 3-4 independent experiments, two-sided t-test, data are mean + s.d.). **c** AAA, CAA, and GAA codon frequency and q value were computed for each human gene, red and blue dots represent genes enriched in the three U<sub>34</sub>-codons (AAA&GAA&CAA q value <0.05 = red dots; AAA&GAA&CAA cumulative frequency >0.088 = blue dots). **d** qRT-PCR after ribosome immunoprecipitation in MDA-MB231 cells. n = 2 independent experiments, two-sided t-test, data are mean + s.d. **e** Levels of indicated proteins were determined in cells by western blot (n = 1 replicate).

changes at the mRNA steady-state levels were detected for U<sub>34</sub>-enriched genes upon ELP3 depletion (Supplementary Fig. 2b and Supplementary Data 2). In agreement with our codon-content analysis, many members of the kinesin family were found downregulated at the protein, but not the mRNA, level in the absence of ELP3 (i.e., KIF4A, KIF14, KIF15, KIF13B, KIF5A, KIF5C, and KIF1B; gene family enrichment analysis,  $p = 1.10 \times 10^{-13}$ ) (Fig. 2c and Supplementary Fig. 2d-e). Surprisingly, other kinesins, such as KIF5B or KIF23, even though strongly enriched in U<sub>34</sub>-codons (i.e., KIF5B: 17.11% and KIF23: 13.4%), were not downregulated in ELP3-depleted cells (Fig. 2a, c and Supplementary Data 2). Strikingly, ribosomes invariably accumulate at

kinesin transcripts enriched in U<sub>34</sub>-codons (i.e., KIF4A, KIF14, KIF15, KIF5B, and KIF23), independently of their protein expression levels, and with no change in the corresponding mRNA levels (Fig. 2d and Supplementary Fig. 2f). Therefore, the U<sub>34</sub>-codon content of mRNAs is predictive of changes in ribosome translational elongation, but is not sufficient to predict protein expression output in the absence of U<sub>34</sub>-enzymes.

**Loss of U<sub>34</sub>-enzymes leads to protein aggregation in a codon-dependent manner.** Defective translation elongation has been associated with increased protein misfolding and aggregation<sup>32</sup>. Accordingly, loss of U<sub>34</sub>-enzymes leads to increased protein

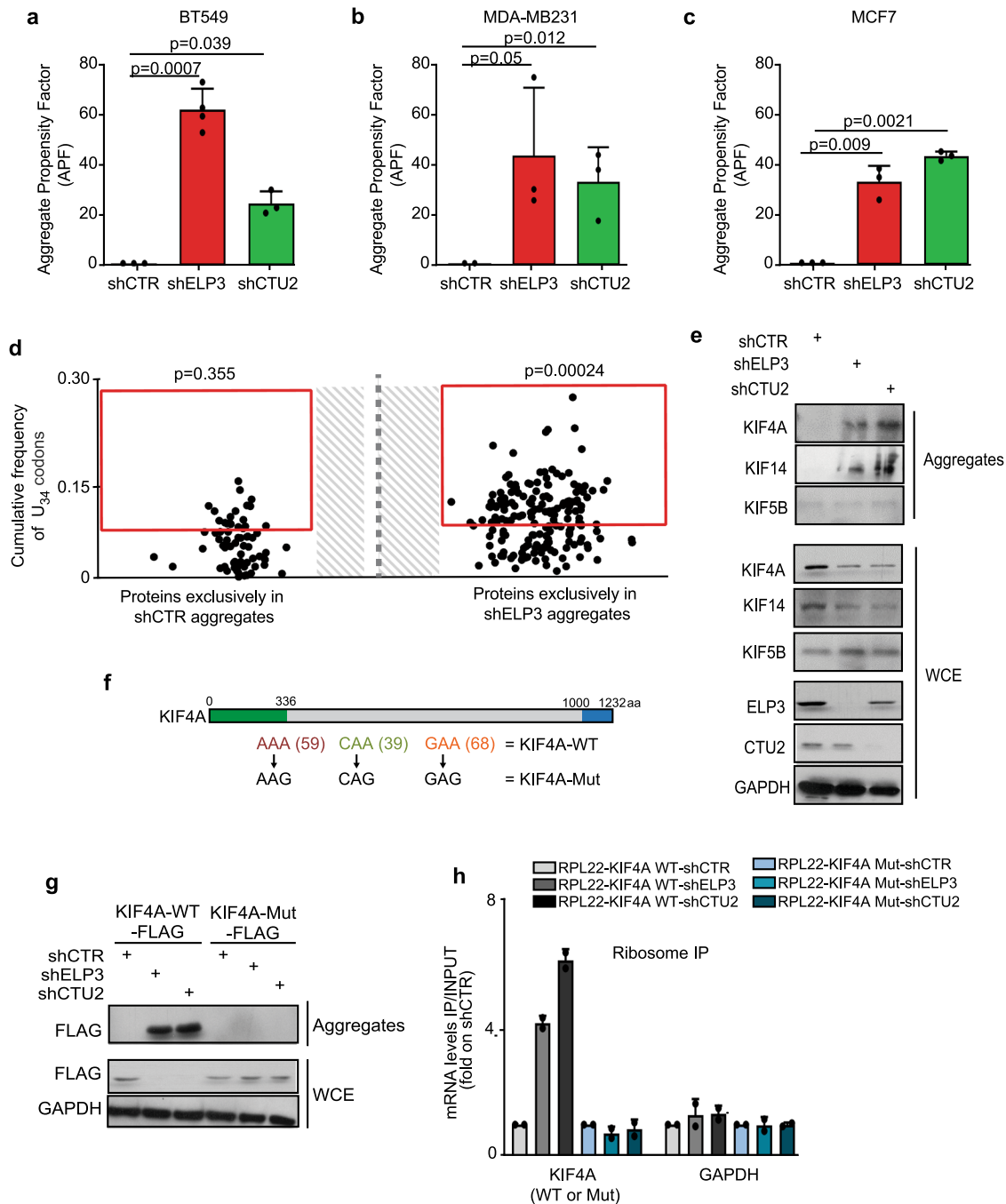


**Fig. 2 mRNA codon content poorly predicts protein expression fate upon loss of U<sub>34</sub>-enzymes.** **a** Dot plot of RNA-seq and proteomics of BT549 cells depleted or not of ELP3 ( $n = 3$  independent experiments); genes enriched in U<sub>34</sub>-codons are shown in blue. U<sub>34</sub>-codons enrichment in downregulated proteins was calculated by  $\chi^2$  test (two-sided). **b**, Frequency of U<sub>34</sub>-codons and protein expression in BT549 cells depleted of ELP3 was plotted, correlation was calculated by Spearman test (two-sided). **c** Levels of indicated proteins were determined in cells depleted of ELP3 or CTU2 by western blot ( $n = 4$  replicates). **d** qRT-PCR after ribosome immunoprecipitation in MDA-MB231 cells.  $n = 2$  independent experiments, data are mean + s.d.

aggregation in a variety of cell lines (Fig. 3a–c)<sup>19,33</sup>. We surmised that defects in ribosome kinetics upon loss of U<sub>34</sub>-enzymes might result in defective protein expression only for proteins that are integrated into aggregates for clearance. To test this hypothesis, we purified the aggregates formed in ELP3-depleted cells and we assessed their protein content by quantitative proteomics (Fig. 3d). Strikingly, proteins exclusively detected in aggregates upon ELP3 depletion were strongly biased toward U<sub>34</sub>-codons in their corresponding mRNAs ( $p = 0.00024$  vs  $p = 0.355$  in control aggregates compared to human orfome). Also, previously identified targets of U<sub>34</sub>-enzymes, KIF4A and KIF14, but not KIF5B, were detected by western blot in protein aggregates upon loss of U<sub>34</sub>-enzymes (Fig. 3e). Of note, KIF4A and KIF14 protein levels were partially rescued after blocking the autophagic flux with chloroquine, but not after proteasome inhibition (Supplementary Fig. 3a). Importantly, the systematic recoding of KIF4A cDNA (i.e., all AAA, CAA, GAA codons were mutated into their synonymous, U<sub>34</sub>-modification independent, codons AAG, CAG, GAG—KIF4A-Mut, as in refs. <sup>19,20</sup>; Fig. 3f) abolished the translation defects observed upon U<sub>34</sub>-enzymes depletion (Fig. 3g), prevented KIF4A aggregation and restored KIF4A expression levels (Fig. 3h). Of note, no changes in mRNA levels nor in global aggregation propensity were observed upon ELP3 depletion

among cells overexpressing the KIF4A-WT or -Mut constructs (Supplementary Fig. 3b–c). This shows that aggregation of protein targets of the U<sub>34</sub>-tRNA modification pathway (U<sub>34</sub>-targets) is caused by codon-dependent translational defects upon loss of U<sub>34</sub>-enzymes. Taken together, these data indicate that U<sub>34</sub>-enzymes prevent the aggregation and degradation of U<sub>34</sub>-targets through codon-dependent regulation of mRNA translation.

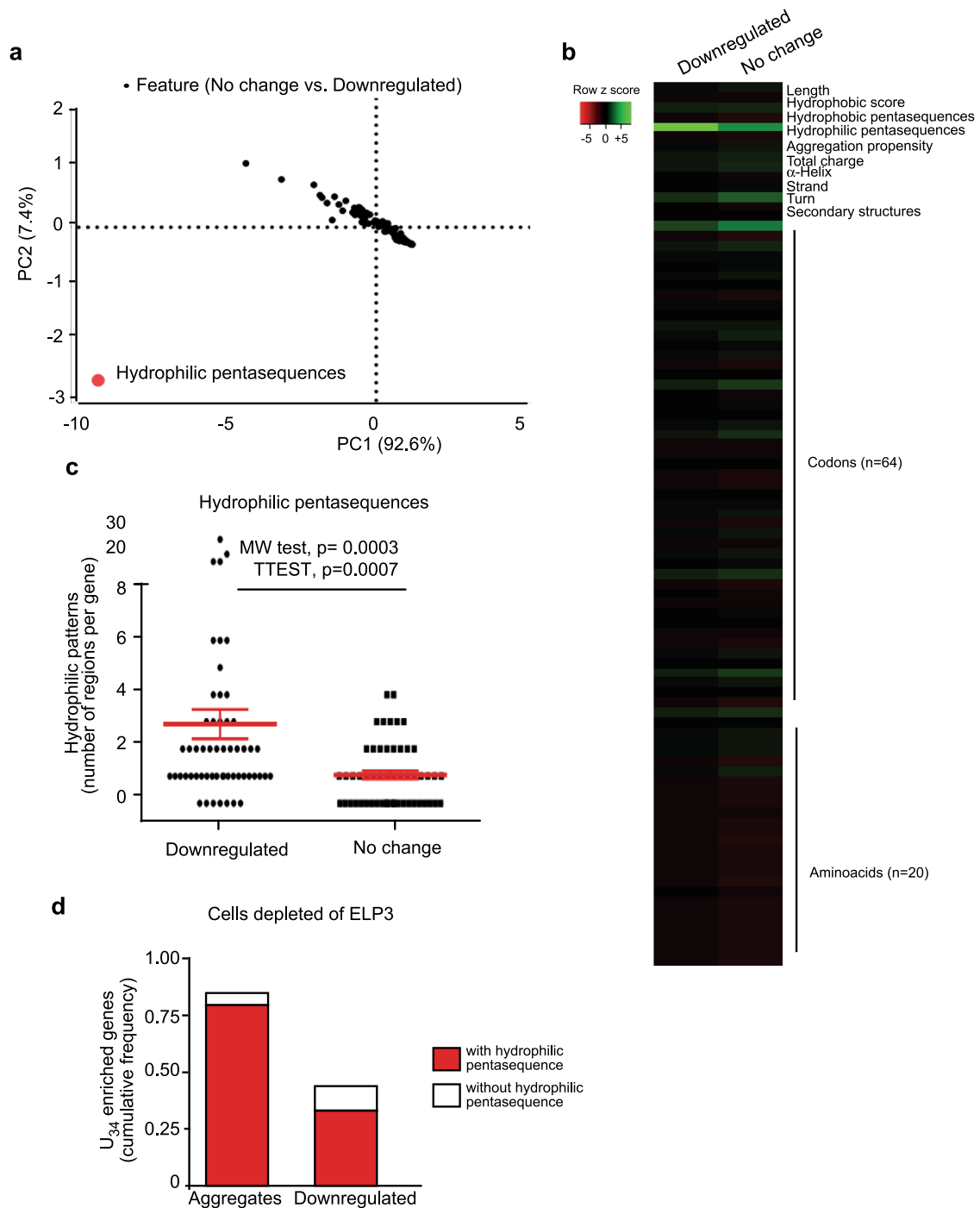
**A penta-hydrophilic amino acid pattern links translation defects to protein aggregation in targets of the U<sub>34</sub>-tRNA modification pathway.** Surprisingly, the above analyses show that defective translation (i.e., ribosome accumulation) in U<sub>34</sub>-enzymes-depleted cells does not systematically correlate with protein aggregation and with subsequent differences in protein expression levels. Indeed, a number of genes are significantly enriched in U<sub>34</sub>-codons but they remain normally expressed at the protein level in cells defective in U<sub>34</sub>-enzymes (Figs. 1e, 2a–c). We surmised that additional features may dictate protein aggregation in the context of defective codon-dependent translation. To identify these features, we performed a series of analyses to identify specific parameters that differentiate proteins that are downregulated versus unchanged upon ELP3 depletion, despite the sharing of high U<sub>34</sub>-codon content



**Fig. 3** Loss of U<sub>34</sub>-enzymes leads to protein aggregation in a codon-dependent manner. **a-c** Protein aggregation was measured by FACS in BT549 (**a**), MDA-MD231 (**b**), and MCF7 (**c**) upon loss of ELP3 or CTU2 ( $n = 3$  independent experiments, two-tail  $t$ -test, data are mean + s.d.). **d** Protein content of aggregates of MCF7 cells depleted or not of ELP3 were assessed by proteomics ( $n = 1$  independent experiment); proteins exclusively found in shELP3 or shCTR conditions are shown. Enrichment of U<sub>34</sub>-codons was assessed by  $\chi^2$  test (two-sided). **e** Levels of indicated proteins were determined by western blot in whole-cell extracts (WCE; loaded 30  $\mu$ g) and aggregates (isolated from >1 mg of protein extract) of MCF7 cells depleted of ELP3 or CTU2 ( $n = 2$  replicates). **f** Schematic representation of KIF4A mutant. The kinesin domain (green) and the globular region (blue) of the protein are indicated. **g** Protein levels of whole-cell extracts (WCE) and aggregates of MCF7 cells stably overexpressing FLAG-RPL22 and the identified KIF4A mutants in the presence or absence of ELP3 or CTU2 ( $n = 3$  replicates). **h** qRT-PCR after ribosome immunoprecipitation of indicated cells.  $n = 2$  independent experiments, data are mean + s.d.

(i.e., AAA, CAA, GAA;  $q$  value <0.05; Supplementary Fig. 4a). First, we quantified features linked to mRNAs, namely, expression levels, length, codon content, and U<sub>34</sub>-codons distribution. No significant difference is observed between the two groups (Supplementary Fig. 4b-e). Hence, for each protein of

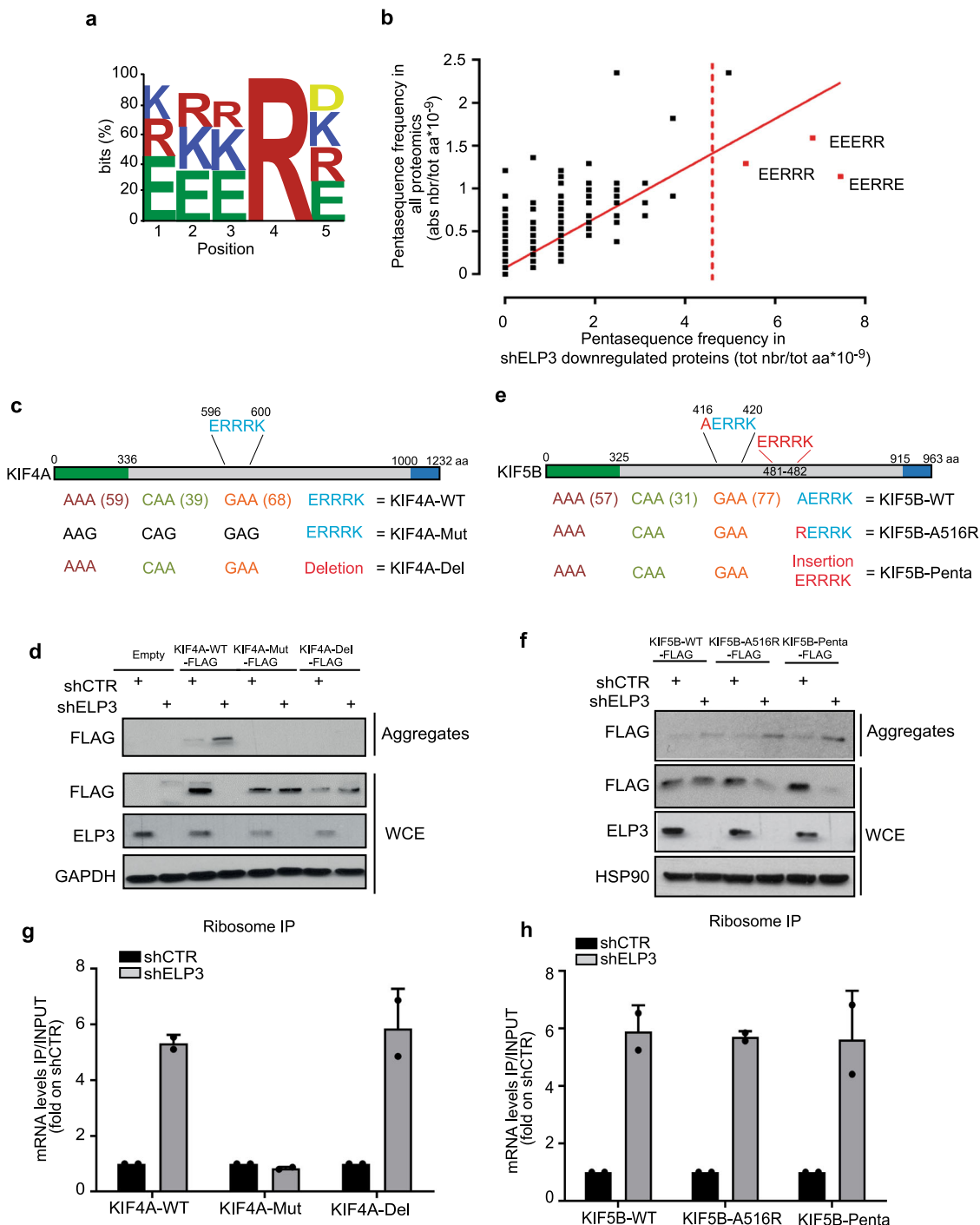
the two groups, we assessed features linked to protein structure (i.e., number of  $\alpha$ -helix and  $\beta$ -sheets), amino acid content, and protein aggregation propensity (calculated by the *Zygggregator* method<sup>34</sup>). Again here, no statistical difference between the two groups is observed (Supplementary Fig. 5a-c). Finally, we



**Fig. 4** A penta-hydrophilic amino acid pattern links translation defects to protein aggregation in  $U_{34}$ -enzymes depleted cells. **a, b** Principal component analysis (**a**) and heat map representation (**b**) of the assessed features in  $U_{34}$ -enriched ( $q$  value  $<0.05$ ) downregulated or unaffected proteins upon depletion of ELP3. **c** Number of hydrophilic pentasequences was assessed in  $U_{34}$ -enriched ( $q$  value  $<0.05$ ) downregulated ( $n = 58$ ) or unaffected proteins ( $n = 56$ ) upon depletion of ELP3. Two-tails  $t$ -test and Mann-Whitney U (MW) test were performed to assess significance. Data are mean  $\pm$  SEM. **d** Presence of hydrophilic pentasequences was assessed in  $U_{34}$ -enriched (frequency of  $U_{34}$ -codon  $>0.088$ ) downregulated proteins (BT549) or in aggregates (MCF7) cells depleted of ELP3.

further evaluated each of the parameters linked to aggregation propensity separately. Strikingly, we found that the presence of patterns of five hydrophilic amino acids significantly segregated proteins that are downregulated upon depletion of ELP3 ( $p = 0.0003$ ; Fig. 4a–c), while total charge, hydrophobic score, and patterns of five hydrophobic amino acids was similar between both groups (Supplementary Fig. 5d–f). In line with these data, proteins harboring hydrophilic pentasequences very

significantly cluster in both ELP3-downregulated proteins as well as in proteins specifically incorporated into aggregates upon ELP3 depletion, indicating a central role of the identified motif in promoting inclusion into aggregates of  $U_{34}$ -target proteins for degradation (Fig. 4d, Supplementary Data 3). Together, these analyses highlighted that the presence of hydrophilic pentasequences is a key feature to predict protein integration into aggregates upon depletion of  $U_{34}$ -enzymes.



**Fig. 5 Specific hydrophilic motifs dictate protein aggregation upon loss of  $U_{34}$ -enzymes.** **a** Identified hydrophilic motif. **b** Assessment of the specific hydrophilic pentasequences (in red) overrepresented in proteins downregulated upon ELP3 depletion. Linear regression is plotted. **c, e** Schematic representation of KIF4A (**c**) and KIF5B (**e**) proteins and mutants. Hydrophilic motifs are indicated. The kinesin domain (green) and the globular region (blue) of the proteins are indicated. **d, f** Levels of indicated proteins were determined by western blot in whole-cell extract (WCE) and aggregates of MCF7 cells overexpressing the indicated mutants of KIF4A (**d**) and KIF5B (**f**) and depleted of ELP3 ( $n = 2$  replicates). **g, h** qRT-PCR after ribosome immunoprecipitation of KIF4A (**g**) and KIF5B (**h**) mutants.  $n = 2$  independent experiments, data are mean + s.d.

**Specific hydrophilic motifs dictate protein aggregation upon loss of  $U_{34}$ -enzymes.** Having uncovered the importance of hydrophilic pentasequence patterns in the definition of protein targets of the  $U_{34}$ -tRNA modification pathway, we combined all identified pentasequences and highlighted a consensus motif [EKR]-[EKR]-[EKR]-R-[DEKR] (Fig. 5a). This motif is present in 10.16% of the protein-coding genes (Supplementary Fig. 6a) and is strongly associated with the frequency of  $U_{34}$ -codons in genes

( $\chi^2$  test:  $p = 9.84 \times 10^{-48}$ ; Supplementary Data 4). Furthermore, the motif discriminates proteins that are downregulated ( $\chi^2$  test:  $p = 3.88 \times 10^{-18}$ ; Fig. 5b) and directed into aggregates upon ELP3 depletion ( $\chi^2$  test:  $p = 8.4 \times 10^{-8}$ ; Supplementary Data 4). Finally, the combination of the two features, the  $U_{34}$ -codon content and the presence of hydrophilic motifs, better predicts protein expression levels downstream of the  $U_{34}$ -tRNA modification pathway ( $\chi^2$  test downregulated proteins vs. no change:  $p = 6.58 \times 10^{-10}$ ;

Supplementary Data 4). Of note, we did not find any correlation between the presence and the distribution of the motifs and protein length (Supplementary Fig. 6b–c). Finally, we computed the conservation scores of the identified motif (Supplementary Fig. 6d) against random sequences of five contiguous amino acids across 10 mammalian species (Supplementary Fig. 6e). Interestingly, we show that the motif is conserved ( $p = 7.05 \times 10^{-20}$ ) suggesting a selection pressure during evolution.

To further assess the biological relevance of the identified motif in determining protein fate upon  $U_{34}$ -tRNA modification-dependent translation elongation defects, we took advantage of two kinesin proteins whose mRNAs are both highly enriched in  $U_{34}$ -codons: (1) KIF4A, which harbors a consensus motif (i.e., ERRRK; Fig. 5c), and is detected in protein aggregates upon depletion of  $U_{34}$ -enzymes (Fig. 5d); and (2) KIF5B, which does not contain any consensus motif (Fig. 5e), is not downregulated and therefore not detected in protein aggregates upon depletion of  $U_{34}$ -enzymes (Fig. 5f). We deleted the hydrophilic motif in the KIF4A sequence and inserted one in the KIF5B sequence and we assessed protein expression of the constructs upon depletion of  $U_{34}$ -enzymes. Strikingly, the deletion of the ERRRK motif in KIF4A (KIF4A-Del; Fig. 5c) is sufficient to rescue its expression level and to prevent KIF4A aggregation in the absence of ELP3 (Fig. 5d). This effect does not rely on changes in mRNA translation elongation (i.e., ribosome occupancy), which remains unaffected by the absence of the motif (Fig. 5g and Supplementary Fig. 7a). Conversely, the addition of the ERRRK motif in the KIF5B cDNA sequence (either by a single amino acid permutation—KIF5B-A516R—or by insertion of the whole ERRRK sequence—KIF5B-Penta; Fig. 5e) was sufficient to promote KIF5B protein aggregation and downregulation upon ELP3 depletion (Fig. 5f). Again, the addition of the motif does not affect KIF5B mRNA translation elongation (Fig. 5h and Supplementary Fig. 7b). These data indicate that the presence of the hydrophilic motif is necessary to promote protein aggregation and downregulation of genes that undergo translation elongation defects upon depletion of  $U_{34}$ -enzymes.

We then wanted to assess if the presence of a hydrophilic motif could discriminate the proteins whose downregulation is directly dependent on the loss of the  $U_{34}$ -tRNA modification (i.e., proteins incorporated in aggregates in a codon-dependent manner) from those that are indirectly downregulated in these conditions. To this aim, we selected two proteins that share similar  $U_{34}$ -codons frequency, are downregulated upon ELP3 depletion and are part of the same kinetochore complex<sup>35</sup>. We generated  $U_{34}$ -codon mutants cDNA of NCD80 and NUF2 (as before; Fig. 6a) and we assessed their expression levels upon  $U_{34}$ -enzymes depletion. Surprisingly, even though NCD80 and NUF2 are in the same protein complex, share a similar frequency of  $U_{34}$ -codons and display comparable ribosome accumulation at their mRNA upon  $U_{34}$ -enzymes loss (Supplementary Fig. 7c), the systematic recoding of  $U_{34}$ -codons by their synonymous only rescued NCD80 expression (Fig. 6b) but did not impact NUF2 levels (Fig. 6c). Importantly, only NCD80 harbors hydrophilic motifs in its sequence (Fig. 6a) and is therefore detected in protein aggregates upon ELP3 depletion (Fig. 6d). Conversely, NUF2 sequence does not contain any hydrophilic motif and it is never detected in protein aggregates upon ELP3 depletion (Fig. 6d). Hence, NUF2 downregulation is partially reverted upon expression of the NDC80-Mut construct (Fig. 6e). This indicates that NUF2 downregulation likely results from the loss of its partner NDC80 and not from an intrinsic dependence on the  $U_{34}$ -tRNA modification pathway. These data reveal that the presence of the hydrophilic motif discriminates between direct (i.e., NDC80) and indirect (i.e., NUF2) targets of the  $U_{34}$ -tRNA modification pathway. All together this set of data demonstrates that the presence of hydrophilic motifs is necessary to mediate

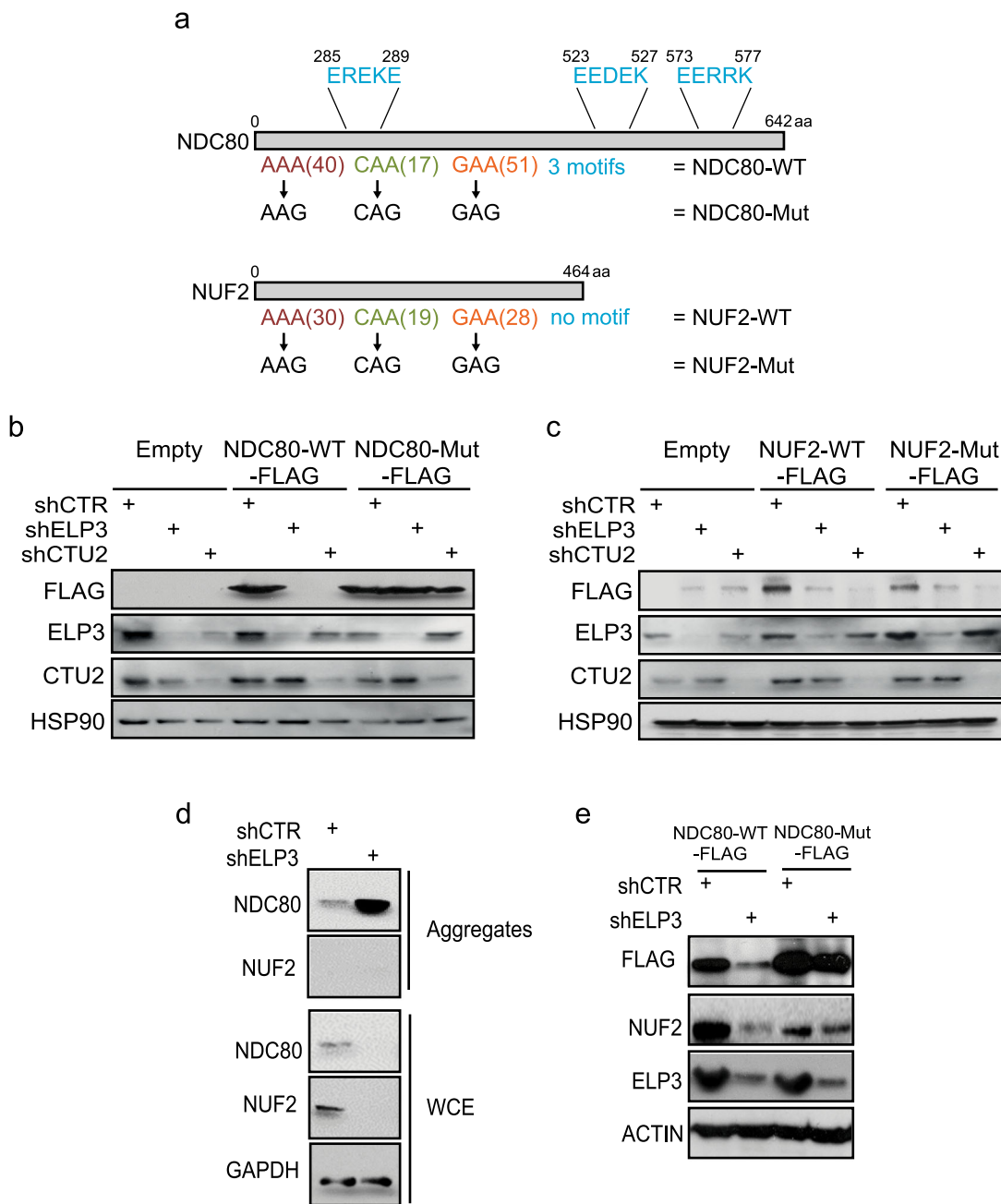
protein aggregation and clearance of  $U_{34}$ -targets upon loss of  $U_{34}$ -enzymes.

### The charged/ampholytic nature of the hydrophilic motif mediates protein aggregation upon depletion of $U_{34}$ -enzymes.

The identified hydrophilic pentamer is defined by its unique amino acid composition, exclusively enclosing charged residues, but also presenting an ampholytic nature (i.e., has the characteristics of both an acid and a base). We surmised that this unique feature might be the reason why proteins are subjected to aggregation upon depletion of  $U_{34}$ -enzymes. In line with this hypothesis, we tested two non-exclusive models that may explain the role of the hydrophilic motif in protein aggregation. First, the ampholytic nature of the motif may drive interaction with the ribosome exit tunnel during translation elongation. The ribosome exit tunnel is a negatively charged cavity from which the nascent proteins arise<sup>36</sup>. Different factors are involved in facilitating or opposing the exit of the nascent peptide from the ribosome tunnel, among these, tRNA availability, protein secondary structure formation and the presence of charged amino acids in the nascent peptide<sup>37–39</sup>. We recently computed the ribosome exit tunnel electrostatics in a model to quantify the energetic needs required to effectively exit nascent peptides from the ribosome tunnel<sup>40</sup>. Using this model, we found that the presence of the hydrophilic motif in the KIF4A protein sequence represents a region of high energetic needs in order to exit the ribosome tunnel and to progress toward translation elongation (Fig. 7a). Conversely, the deletion of the hydrophilic motif in the KIF4A protein sequence (i.e., in KIF4A-Del), which prevents KIF4A aggregation upon depletion of  $U_{34}$ -enzymes, alters the local energetic barrier, which is predicted to favor ribosome movement across this region (Fig. 7b). This observation indicates that the presence of the hydrophilic motif brings an additional energetic barrier during translation elongation, which may be the cause of protein misfolding and aggregation.

Second, recent evidences describe the propensity of low complexity regions (LCR) rich in charged amino acids to regulate protein phase behavior, to promote liquid–liquid phase separation (LLPS) and to trigger protein aggregation<sup>41–45</sup>. Therefore, we speculated that the ampholytic nature of the hydrophilic motif might trigger LLPS and auto-aggregation among the  $U_{34}$ -target proteins, upon  $U_{34}$ -enzymes depletion. We visualized KIF4A-WT or KIF4A-Del (i.e., that does not harbor the hydrophilic motif) proteins in MCF7 cells by immunofluorescence. Interestingly, upon depletion of  $U_{34}$ -enzymes, KIF4A-WT aggregates in small punctuated structures, similar to SHP2 mutants-LLPS associated structures<sup>46</sup> (Fig. 7c). Strikingly, KIF4A-Del is never found in these structures in  $U_{34}$ -enzymes depleted cells (Fig. 7c). We also assessed the amyloid aggregation propensity in KIF4A-WT and KIF4A-Del using the TANGO algorithm<sup>47</sup>. Interestingly, the hydrophilic motif we identified in KIF4A is predicted to have high propensity to undergo amyloid aggregation (Fig. 7d). Moreover, its deletion (i.e., in KIF4A-Del) causes a net reduction in amyloid aggregation propensity in KIF4A, as measured by the total area under the curve (KIF4A-WT = 163231 and KIF4A-Del = 146070), and a change in the maximal amyloid aggregation peak (KIF4A-WT max peak—intensity 34359—is found in bin 9 that harbors the motif, while KIF4A-Del max peak—intensity 24352—is found in bin 14) (Fig. 7d). Conversely, deletion of the motif did not affect KIF4A general aggregation propensity (Supplementary Fig. 8a) nor its presence was found associated with aggregation propensity calculated by the Zyggregator method (Supplementary Fig. 8b). This indicates that the presence of the charged/ampholytic hydrophilic motif may trigger





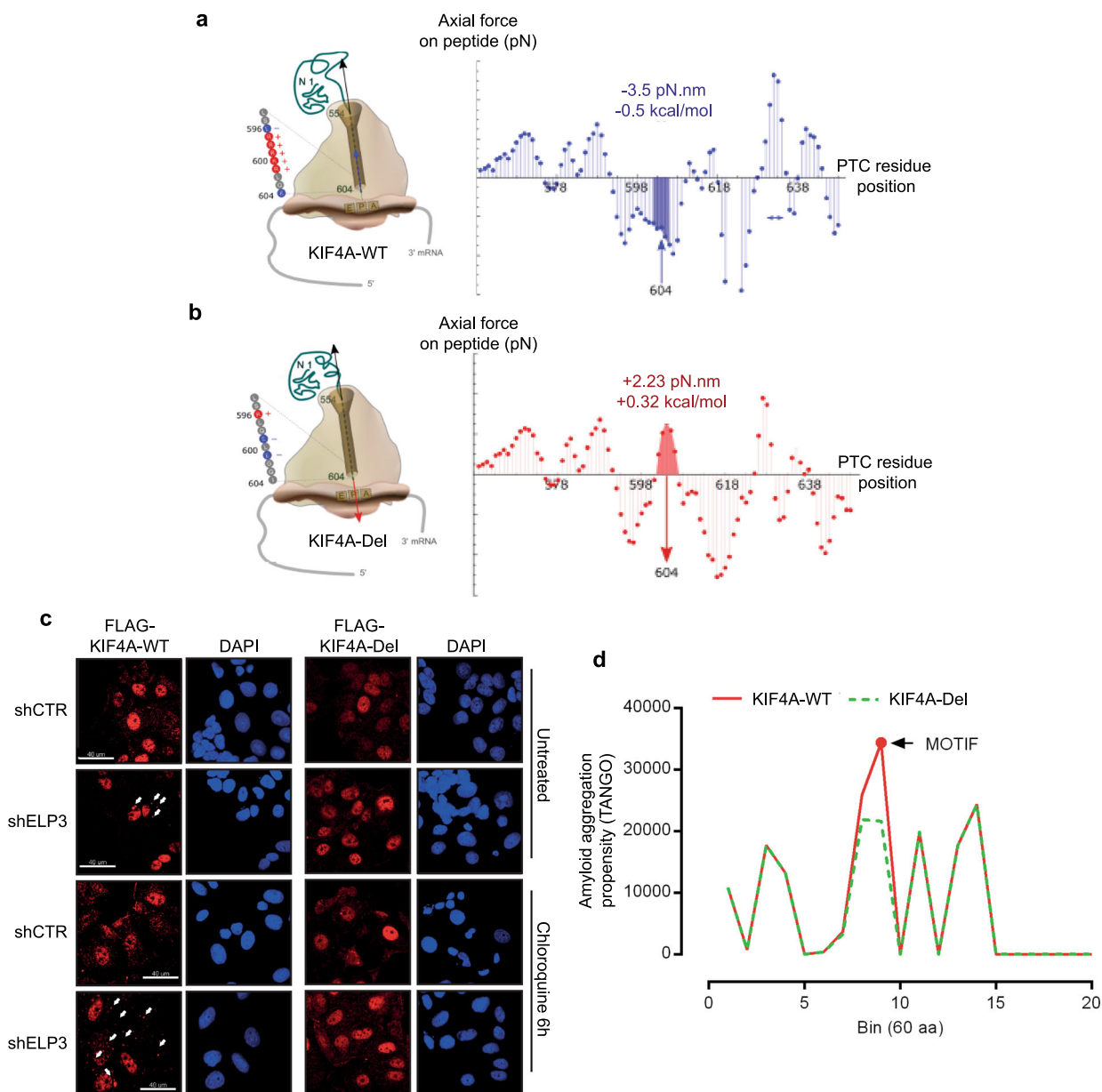
**Fig. 6 The presence of hydrophilic motifs discriminates between direct and indirect targets of the U<sub>34</sub>-tRNA modification pathway.** **a** Schematic representation of the NDC80 and NUF2 protein sequences. Hydrophilic motifs are indicated in blue. **b** Protein levels of indicated proteins in cells overexpressing control or NDC80-WT-FLAG or NDC80-Mut-FLAG constructs in the presence or absence of ELP3 or CTU2 (*n* = 2 replicates). **c** Protein levels of indicated proteins in cells overexpressing control or NUF2-WT-FLAG or NUF2-Mut-FLAG constructs in the presence or absence of ELP3 or CTU2 (*n* = 2 replicates). **d** Levels of indicated proteins were determined by western blot in whole-cell extracts (WCE) and aggregates of MCF7 cells depleted of ELP3 (*n* = 2 replicates). **e** Levels of indicated proteins were determined by western blot in cells overexpressing NDC80-WT-FLAG or NDC80-Mut-FLAG constructs in the presence or absence of ELP3 (*n* = 1 replicate).

aggregation through changes in protein phase behavior and LLPS upon depletion of U<sub>34</sub>-enzymes.

**Discussion**

Recently, the role of tRNA biology in health and disease has gained an increasing amount of attention<sup>11,48,49</sup>. The development of ribosome sequencing techniques allowed to assess the exact role of tRNA modifications, such as mcm<sup>5</sup>s<sup>2</sup> U<sub>34</sub>-tRNA modification, in translation regulation and specific codon decoding<sup>16,17,19</sup>. Currently, it is assumed that impairments in

ribosome kinetics at mRNAs are directly reflected in changes in expression of the corresponding proteins. Nevertheless, very few studies focus on the mechanisms linking codon-specific ribosome pausing and protein homeostasis. Specifically, how perturbation of codon-dependent-translation impacts subsequent protein expression output remains poorly understood. Surprisingly, here we show that codon-dependent translation elongation defects caused by the depletion of U<sub>34</sub>-enzymes are not systematically associated with changes in protein expression. While translation elongation defects upon perturbation of U<sub>34</sub>-tRNA modification

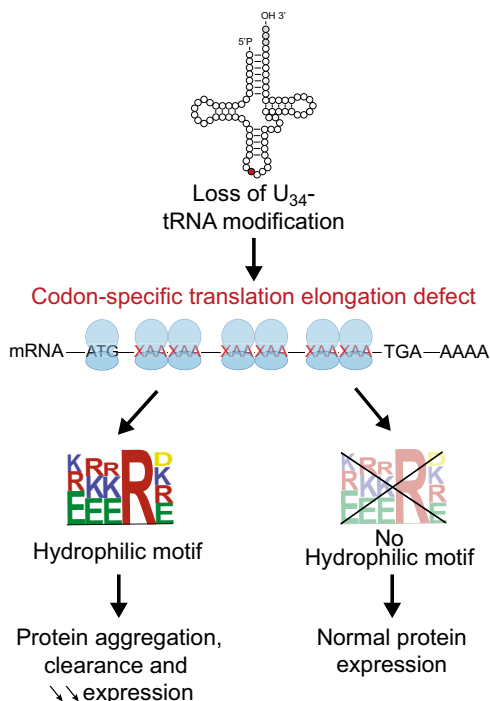


**Fig. 7** The ampholytic nature of the motif drives protein aggregation upon depletion of  $U_{34}$ -enzymes. **a, b** Axial forces profile on nascent chain caused by the electrostatic interaction in the ribosome exit tunnel (**a**) wild-type KIF4A; (**b**) ERRRK deleted KIF4A. The axial forces profiles and local mechanical work acting on the two chains are different upon incorporation of residues 596–646 at the peptidyl-transferase center (PTC). Schematic representation of ribosome–protein interaction, on the left. **c** Anti-Flag immunofluorescence of MCF7 cells depleted or not of ELP3 and overexpressing KIF4A-WT or KIF4A-Del in control or after 6 h of treatment with 100  $\mu$ M chloroquine. Arrows depict KIF4A aggregates ( $n = 2$  replicates). **d** Amyloid aggregation profiles of KIF4A-WT and KIF4A-Del calculated by using the TANGO algorithm.

are strictly dependent on codon content, the consequences of these defects on protein output are determined by other features. We identify hydrophilic amino acid motifs as essential to dictate protein aggregation and subsequent degradation in contexts of defective translation elongation at  $U_{34}$ -codons. Therefore, we propose a model in which codon-dependent regulation of translation elongation by wobble tRNA modification and maintenance of protein integrity are two independent processes (see the model in Fig. 8). Importantly, the previously identified protein targets of the  $U_{34}$ -tRNA modification pathway (i.e., SOX9<sup>25</sup>, DEK<sup>20</sup>, HIF1 $\alpha$ <sup>19</sup>, and hnRNPQ<sup>26</sup>; Fig. 1c–e) all harbor at least one hydrophilic motif in their sequences, supporting the model here proposed (Supplementary Fig. 8c). Of note, a portion of genes enriched in  $U_{34}$ -codons and downregulated

upon depletion of  $U_{34}$ -enzymes do not harbor hydrophilic motifs and are not directed toward aggregation (for instance NUF2; see Supplementary Table 3). The mechanism leading to their downregulation upon depletion of  $U_{34}$ -enzymes remains to be understood.

How the presence of hydrophilic motifs mediate protein aggregation remains to be precisely established. The motif strongly differs from known chaperone- and cochaperone-binding sites, which are generally hydrophobic<sup>50</sup>. In contrast, the here identified motif shares some similarities with the motifs responsible for chaperone-mediated autophagy (CMA) (i.e., KFERQ and derivatives)<sup>51</sup>. Nevertheless, the unbiased search for CMA motifs<sup>51</sup> showed no overlap with the hydrophilic motifs identified herein, making this mechanism of degradation unlikely



**Fig. 8 Proposed model.** Loss of U<sub>34</sub>-enzymes invariably determines pausing and accumulation of ribosomes on transcripts rich in XAA codons. The result of this pausing on protein expression is determined by the presence of hydrophilic motifs that mediate protein aggregation and subsequent clearance. See text for details.

(Supplementary Fig. 8d–e). Previous studies highlighted the potential role of hydrophilic stretches as gatekeepers for the correct folding of nascent proteins due to their proximity to aggregation-prone regions<sup>52–55</sup>. Although, we do not exclude this possibility, in the KIF4A protein sequence, the hydrophilic motif sits in the proximity of low aggregation-prone regions. Rather, our data show that the unique amino acid composition of the motif may underlie protein aggregation upon U<sub>34</sub>-enzymes depletion. Indeed, we show that the charged and ampholytic nature of the motif may affect phase behavior of the U<sub>34</sub>-protein targets, possibly triggering LLPS and consequently promoting protein aggregation upon depletion of U<sub>34</sub>-enzymes. For example, the here identified motif shares physical characteristics with fragments in the low complexity (LC) domains in the U1-70K protein, that have been shown to drive LLPS and aggregation in cells<sup>41</sup>. Also, arginine-rich domains mediate LLPS via electrostatic interactions, which therefore requires multiple proteins or motifs. Why would the perturbation of U<sub>34</sub>-tRNA modification and the subsequent codon-dependent deregulation of translation favor this mechanism is not known. It is possible that the translational defects seen upon depletion of U<sub>34</sub>-enzymes might result in protein downregulation, only if the exposure of charged/ampholytic motifs is sufficient to trigger LLPS-driven aggregation that will lead to clearance by the lysosome.

The charged/ampholytic nature of the pentamer could trigger aggregation also by its interaction with the ribosome exit channel tunnel during translation elongation. Previous studies demonstrated that factors involved in facilitating or opposing the exit of the nascent peptide from the ribosome tunnel can impact on protein folding<sup>56</sup>. Interestingly, in KIF4A, the identified hydrophilic motif represents a region of high energetic needs in order to exit the ribosome tunnel and to progress toward translation elongation (Fig. 7c). Conversely, its deletion, which prevents incorporation of KIF4A in aggregates upon

depletion of the U<sub>34</sub>-enzymes (Fig. 5d), alters the local energetic barrier, likely favoring ribosome movement across this region (Fig. 7d). It is therefore tempting to speculate that the additional energetic barrier brought by the presence of the hydrophilic motif could lead to an excessive ribosome pausing and may be the cause of protein misfolding and aggregation. Further work is required to further clarify these mechanisms.

Our data open the possibility that the uncoupling of translation elongation defects and maintenance of protein homeostasis may generally apply to other tRNA modification pathways. Understanding the mechanisms linking codon-dependent translation defects to maintenance of protein expression for all the still poorly studied tRNA modification pathways will help define the importance of this layer of regulation in proteome expression.

## Methods

**Cell lines and reagents.** Breast cancer cell lines (BT549, MDA-MB231, and MCF7) were from ATCC and were grown in Dulbecco's modified Eagle's medium (Gibco) containing 10% of fetal bovine serum, 1% glutamine, and 1% penicillin–streptomycin. Lines were tested for mycoplasma contamination on a regular basis.

The antibodies, primers, and shRNA sequences used in this study are listed in Supplementary Data 5.

**Western blot and immunoprecipitation.** Cells were lysed in lysis buffer (50 mM Tris-HCl pH 7.5, containing 2 mM EDTA, 150 mM sodium chloride, 0.5% sodium deoxycholate, 50 mM sodium fluoride, 1% NP-40, and 0.1% SDS, Complete™ proteinase inhibitor cocktail (Roche)) and protein determination was assessed using BCA Protein Assay Kit from ThermoFisher and according to manufacturer's instructions.

**qPCR.** Total RNA was extracted using the E.Z.N.A. Total RNA Kit I (OMEGA Bio-Tek) according to the manufacturer's instructions. Subsequently, 1 µg of total RNA was used to synthesize the corresponding cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher). Quantitative evaluation of selected genes was determined by SYBR-Green-based qPCR performed using the Light Cycler 480 (Roche). All determinations were performed in triplicates. The relative expression of the target gene transcript and reference gene transcript was calculated as  $\Delta\Delta Ct$ .

**Viral infections.** For stable lentiviral shRNA infections, all constructs were purchased from Sigma. The kinesins FLAG-tag overexpression plasmids (KIF4A, KIF5B, and mutants) were generated from the pLV-Puro-hPGK plasmid by Vectorbuilder. NUF2 and NDC80 FLAG-tag overexpression plasmids and codon mutants were generated from the Lego-Ig2 plasmid by GenScript. The Luciferase/Nanoluciferase constructs were derived from pLV-Puro-hPGK and generated by Vectorbuilder (maps are available upon request). Lentiviral infections were carried out as previously described<sup>57</sup>. Briefly, in all cases, HEK293-LentiX producer cells were transfected with 12 µg of lentiviral vector, 5 µg VSVG, and 12 µg R.891 and Mirus (3:1) for 48 h. Supernatant containing virus was filtrated using a 45 µm filter and polybrene was added to a final concentration of 8 µg/ml, and virus were added to target cells. Cells were selected for 3 days with 1 µg/ml puromycin. Knockdown or overexpression efficiency was evaluated by RT-qPCR and/or by western blot analysis.

**Ribosome immunoprecipitation.** Ribosome pulldown was performed as previously described<sup>19</sup>. Briefly, cells were stably transduced with RPL22-FLAG expressing lentiviral construct and treated for 15 min with 100 µg/ml cycloheximide. Then, cells were lysed in 20 mM HEPES-KOH (pH 7.3), 150 mM KCl, 10 mM MgCl<sub>2</sub>, 1% NP-40, EDTA-free protease inhibitors, 0.5 mM DTT, 100 µg/ml cycloheximide, and 10 µl/ml rRNasin and Supersin in RNase-free water, and incubated overnight with anti-FLAG beads. After incubation, beads were washed five times with a high salt buffer (350 mM KCl) and three times with a low salt buffer (150 mM KCl). Beads were then washed in RNA extracting buffer with β-mercaptoethanol as previously described and underwent RT-qPCR.

**Immunofluorescence.** 200000 MCF7 cells depleted or not of ELP3 were seeded and treated as indicated. Cells were fixed in 4% PAF and permeabilized in PBS supplemented with Triton 0.2%. Primary antibody was kept overnight at 4 °C in PBS supplemented with 0.1% TRITON and 5% BSA (dilution: 1/200). Appropriate secondary antibody (dilution: 1/1000) and DAPI were added for 1 h at room temperature in PBS supplemented with TRITON 0.1% and BSA 3%. Images were acquired by confocal microscopy (confocal Leica SP5).

**RNA sequencing.** BT549 cells were stably depleted of ELP3 by shRNA (see the “Viral infections” section); a scramble shRNA was used as control. Total RNAs from three independent infections were extracted using TRIzol (Life Technologies), according to the manufacturer’s protocol. RNA integrity was verified on the Bioanalyser 2100 with RNA 6000 Nano chips, and RIN scores were >9 for all samples. The Illumina Truseq RNA Sample Preparation kit V2 was used to prepare libraries from 500 ng total RNAs. PolyA RNAs were purified with polyT-coated magnetic beads, chemically fragmented, and used as template for cDNA synthesis using random hexamers. cDNA ends were subsequently end-blunted, adenylated at 3’OH extremities, and ligated to indexed adaptors. Finally, the adaptors ligated library fragments were enriched by PCR following Illumina’s protocol and purified. Libraries were validated on the Bioanalyser DNA 1000 chip and quantified by qPCR using the KAPA library quantification kit. Sequencing was performed on HiSeq2000 in paired-end 2 × 100 base protocol. For data analysis, the nf-core-rnaseq pipeline was used.

**Proteomic study.** BT549 cells depleted or not of ELP3 by shRNA were lysed in SDS 2% buffer containing 50 mM Tris-HCl pH 8, 150 mM sodium chloride, 10 mM sodium fluoride, and 1 mM trisodium phosphate, Complete™ proteinase inhibitor cocktail (Roche). At least 1 mg of each sample was sent for mass spectrometry analysis using LC/MS/MS (MsBioworks, LLC). Samples were TCA precipitated overnight at –20 °C in 90/10 acetonitrile/water incorporating 20% TCA. Precipitates were washed, pelleted, and resuspended in 5 M urea, 50 mM Tris-HCl, pH 8.0, Complete™ proteinase inhibitor cocktail (Roche) and quantitated by Qubit fluorometry (Life Technologies). A 50 µg aliquot per sample was reduced using dithiothreitol, alkylated with iodoacetamide, and then subjected to digestion with trypsin (Promega) for 18 h at 37 °C at an enzyme-to-substrate ratio of 1:50. The digestion was quenched with formic acid and peptides cleaned using solid-phase extraction (SPE) using the Empore C18 plate (3 M). A 2 µg aliquot of each sample was analyzed by nano LC/MS/MS with a Waters M-Class HPLC system interfaced to a Fusion Lumos tandem mass spectrometer (ThermoFisher). Peptides were loaded on a trapping column and eluted over a 75 µm analytical column at 350 nL/min; both columns were packed with Luna C18 resin (Phenomenex). A 4 h gradient was employed. The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 60,000 FWHM resolution and 15,000 FWHM resolution, respectively. APD was turned on. The instrument was run with a 3-s cycle for MS and MS/MS. The experiment was done in triplicate.

Thermo RAW files were processed using MaxQuant v1.6.2.3 (Max Planck Institute). Production data were searched against the combined forward and reverse Swissprot H. sapiens protein database using Andromeda. Data were filtered 1% protein and peptide-spectrum match (PSM) level false discovery rate (FDR) and requiring at least one unique peptide per protein. Label-free protein quantitation was based on the extracted ion chromatograms of detected peptides; data were normalized using the embedded LFQ algorithm.

**Protein aggregates determination.** Protein aggregates formation was monitored by FACS using PROTEOSTAT® Aggresome Detection Kit (ENZO) according to manufacturer instructions and as previously described<sup>19</sup>. Fluorescence was measured by FACS by FL3 channel. Aggresome propensity factor (APF) was calculated using the formula:  $100 \times (\text{mean fluorescence intensity depleted} - \text{mean fluorescence intensity control}) / \text{mean fluorescence intensity depleted}$ .

Isolation of protein aggregates was performed as previously described<sup>19</sup>. Briefly, cells were pelleted, snap-frozen in liquid nitrogen, and resuspended in lysis buffer (20 mM Na-phosphate, pH 6.8, 10 mM DTT, 1 mM EDTA, 0.1% Tween, 1 mM PMSF, protease inhibitor cocktail [Roche] and DNase I). Lysates were kept 20 min at room temperature and sonicated. After centrifugation at  $200 \times g$  for 20 min at 4 °C, supernatants were collected and protein dosage performed. Protein extracts were then centrifuged for 20 min  $16,000 \times g$  at 4 °C. Pellets were washed (20 mM Na-phosphate, pH 6.8, 1 mM PMSF, protease inhibitor cocktail, 2% NP-40) and sonicated. Final pellets were washed with buffer with no NP-40 and sonicated again. For protein detection by western blot, pellets were resuspended in 1% SDS.

For proteomics analysis, aggregates were purified from the cytoplasmic fraction after cyto-nuclear extraction. Briefly, cells were lysate with cytoplasmic lysis buffer (1 M Tris-HCl pH 7.9, 0.2 CaCl<sub>2</sub>, 0.25 M EDTA, 2 M MgCl<sub>2</sub>, protease inhibitors) and supernatant was collected after centrifugation (4 °C for 10 min at  $17,000 \times g$ ). After aggregates isolation, pellets were washed twice in 1 × PBS, and analyzed by LC/MS/MS.

**Luminescence.** Luminescence of indicated stably transduced MDA-MB231 cell lines was measured using Promega Dual-Luciferase® Reporter Assay according to manufacturer instructions. Data were expressed as fold of nanoluc/firefly ratio.

**Computational analysis of gene-specific codon usage (GSCU) patterns.** We downloaded 23,099 homo sapiens gene sequences from the human genome database (HGD), and then analyzed each gene using our GSCU algorithm. The GSCU algorithm interrogated a gene from the start to the stop codon. The frequency of a specific codon (i.e., AAA, GAA, and CAA) used in a gene was counted and further compared with the codon frequency in another gene, which was randomly picked up. 10,000 times comparison was done. By such comparison we got the *p*-value,

which means the possibility that the randomly picked up gene has higher frequency of that specific codon than the interest one. We then used *q* package with the R program, and we generated the *q* value for each gene. Using  $q = 0.05$  as cut off, we picked up the genes enriched for the U<sub>34</sub>-codons AAA, GAA, and CAA.

**Codon enrichment analysis.** Genes clustering was performed by filtering for statistical increase in specific codons content ( $q < 0.05$ ). Genes with AAA, CAA, and GAA *q* value <0.05 subsequently underwent Toppgene analysis in order to uncover enriched biological process and gene families.

**Statistical analysis.** Gene set enrichment analysis was performed using Toppfun software and GO\_biological process database. A FDR *q* value lower than 0.05 was taken into account. Bubble graphs were created in excel: distance between the clusters was calculated as the frequency of overlapping genes between clusters. The first 30 significant pathways were considered.

Distribution of U<sub>34</sub>-codons across mRNAs was assessed by computing the U<sub>34</sub>-codons present in genes. To this end, we divided into 100 bins the 116 genes of interest and calculated the frequency of the U<sub>34</sub> codons for each bin. Kolmogorov–Smirnov test was performed to assess significance.

Principal component analyses were performed using the FactoMineR and Factoextra R packages. The script is available upon request. Chi-square test was performed to assess significance.

Aggregation propensity was evaluated by two methods: (1) the Zyggregator method<sup>34</sup>, (2) the TANGO algorithm<sup>47</sup>. Aggregation by the zyggregator method was calculated by using the  $\log(k) = \alpha_0 + \alpha^{\text{hydr}} + \alpha^{\text{ss}} + \alpha^{\text{ch}} + \alpha^{\text{pat}}$  formula; where  $\alpha_0$  indicates pH and temperature (here set to 0),  $\alpha^{\text{hydr}}$  indicates the Hydrophobic score calculated as the sum of the hydrophobic value from ref.<sup>58</sup>,  $\alpha^{\text{ss}}$  indicates the absolute number of secondary structures such as  $\alpha$ -helix and  $\beta$ -sheets according to Uniprot database,  $\alpha^{\text{ch}}$  indicates the total charge of each protein calculated by adding the net charge of +1 or –1 for positively or negatively charged amino acids, respectively,  $\alpha^{\text{pat}}$  indicates a pattern of five consecutive hydrophobic or hydrophilic residues in the sequence; a value of +1 was given to each pentasequence.

Aggregation calculated by the TANGO algorithm was performed assuming the following parameters: pH = 7; temperature = 298.15 K; ionic strength is 0.02 M; concentration = 1 M; no protection of N and C-term. Both beta-sheet aggregation and amyloid aggregation were assessed.

Analysis of the chaperone-mediated autophagy motifs was done by the KFERQ-finder software<sup>31</sup>.

The mapping of the electrostatic potential within the ribosomal exit tunnel and the energy needed by the nascent peptide to exit the tunnel was performed as in ref.<sup>40</sup>.

**Motif search.** Hydrophilic pentasequences ( $n = 161$ ) were aligned by ClustalW in UNIGE to find a consensus motif. Motif search ([EKR]-[EKR]-[EKR]-R-[DEKR]) in different datasets was performed by Motif search tool by Genome.net.

For the conservation analysis, 1466 multiple-alignments were conducted for each subset of 10 orthologs. The human protein was used as source for all the 9 target mammalian orthologs in order to be aligned using the CLUSTAL-W algorithm in Bio-Python. Conservation scores for all the hydrophilic pentasequences and for random sequence of 5 contiguous residues were calculated (Shannon entropy conservation score).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Proteomics data are available at ProteomeXchange Consortium (PXD019590; DOI: 10.6019/PXD019590 and PXD019620; DOI: 10.6019/PXD019620) and RNA-seq data are available at ArrayExpress under accession code E-MTAB-9206. Open access databases used in this work include Swissprot (H. sapiens protein database, [https://www.uniprot.org/uniprot/?query=\\*&fil=organism%3A%22Homo+sapiens+%28Human%29+%5B9606%5D%22+AND+reviewed%3Ayes](https://www.uniprot.org/uniprot/?query=*&fil=organism%3A%22Homo+sapiens+%28Human%29+%5B9606%5D%22+AND+reviewed%3Ayes)); Human genome database (<https://www.ncbi.nlm.nih.gov/grc/human>), GO\_biological process database (<http://geneontology.org/>). The data supporting the findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Received: 18 May 2020; Accepted: 5 March 2021;

Published online: 15 April 2021

## References

- Kim, S. J. et al. Translational tuning optimizes nascent protein folding in cells. *Science* <https://doi.org/10.1126/science.aaa3974> (2015).
- Yu, C. H. et al. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2015.07.018> (2015).

3. Collart, M. A. & Weiss, B. Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz763> (2019).
4. Thommen, M., Holtkamp, W. & Rodnina, M. V. Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* <https://doi.org/10.1016/j.sbi.2016.11.020> (2017).
5. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/nsmb.2466> (2013).
6. Zhang, G. & Ignatova, Z. Folding at the birth of the nascent chain: Coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* <https://doi.org/10.1016/j.sbi.2010.10.008> (2011).
7. Patil, A. et al. Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. *RNA Biol.* **9**, 37–41 (2012).
8. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* <https://doi.org/10.1016/j.cell.2008.05.042> (2008).
9. Stein, K. C. & Frydman, J. The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.REV118.002814> (2019).
10. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* <https://doi.org/10.1126/science.aac4354> (2016).
11. Pereira, M. et al. Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19123738> (2018).
12. Huang, B., Johansson, M. J. O. & Byström, A. S. An early step in wobble uridine tRNA modification requires the Elongator complex. *RNA* **11**, 424–436 (2005).
13. Leidel, S. et al. Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature* **458**, 228–232 (2009).
14. Songe-Møller, L. et al. Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. *Mol. Cell. Biol.* **30**, 1814–1827 (2010).
15. Dewez, M. et al. The conserved Wobble uridine tRNA thiolase Ctu1-Ctu2 is required to maintain genome integrity. *Proc. Natl Acad. Sci. USA* **105**, 5459–5464 (2008).
16. Zinshteyn, B. & Gilbert, W. V. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* **9**, e1003675 (2013).
17. Nedialkova, D. D. & Leidel, S. A. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* **161**, 1–13 (2015).
18. Lagusse, S. et al. A dynamic unfolded protein response contributes to the control of cortical neurogenesis. *Dev. Cell* **35**, 553–567 (2015).
19. Rapino, F. et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* **558**, 605–609 (2018).
20. Delaunay, S. et al. Elp3 links tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis in breast cancer. *J. Exp. Med.* **213**, 2503–2523 (2016).
21. Leidel, S. et al. Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature* **458**, 228–232 (2009).
22. Rapino, F., Delaunay, S., Zhou, Z., Chariot, A. & Close, P. tRNA modification: is cancer having a wobble? *Trends Cancer* **3**, 249–252 (2017).
23. Freeman, S. et al. Proteostasis is essential during cochlear development for neuron survival and hair cell polarity. *EMBO Rep.* <https://doi.org/10.15252/embr.201847097> (2019).
24. Hawer, H. et al. Roles of elongator dependent tRNA modification pathways in neurodegeneration and cancer. *Genes* <https://doi.org/10.3390/genes10010019> (2019).
25. Ladang, A. et al. Elp3 drives Wnt-dependent tumor initiation and regeneration in the intestine. *J. Exp. Med.* **212**, 2057–2075 (2015).
26. Xu, S. et al. Genome-wide CRISPR screen identifies ELP5 as a determinant of gemcitabine sensitivity in gallbladder cancer. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-13420-x> (2019).
27. Chan, C. T. Y. et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* **3**, 937 (2012).
28. Endres, L., Dedon, P. C. & Begley, T. J. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. *RNA Biol.* **12**, 603–614 (2015).
29. Goffena, J. et al. Elongator and codon bias regulate protein levels in mammalian peripheral neurons. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-03221-z> (2018).
30. Patil, A. et al. Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. *Cell Cycle* **11**, 3656–3665 (2012).
31. Frenkel-Morgenstern, M. et al. Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.* <https://doi.org/10.1038/msb.2012.3> (2012).
32. Jacobson, G. N. & Clark, P. L. Quality over quantity: optimizing co-translational protein folding with non-‘optimal’ synonymous codons. *Curr. Opin. Struct. Biol.* <https://doi.org/10.1016/j.sbi.2016.06.002> (2016).
33. Nedialkova, D. D. & Leidel, S. A. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* <https://doi.org/10.1016/j.cell.2015.05.022> (2015).
34. Tartaglia, G. G. & Vendruscolo, M. The Zygggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* <https://doi.org/10.1039/b706784b> (2008).
35. DeLuca, J. G. et al. Hec1 and Nuf2 are core components of the kinetochore outer plate essential for organizing microtubule attachment sites. *Mol. Biol. Cell* <https://doi.org/10.1091/mbc.E04-09-0852> (2005).
36. Lu, J., Kobertz, W. R. & Deutsch, C. Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2007.06.038> (2007).
37. Sharma, A. K., Ahmed, N. & O’Brien, E. P. Determinants of translation speed are randomly distributed across transcripts resulting in a universal scaling of protein synthesis times. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.97.022409> (2018).
38. Tuller, T. et al. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.* <https://doi.org/10.1186/gb-2011-12-11-r110> (2011).
39. Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L. & Roubos, J. A. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkv199> (2015).
40. Joiret, M., Rapino, F., Close, P. & Geris, L. Ribosome exit tunnel electrostatics. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.20.346684> (2020).
41. Xue, S. et al. Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aax5349> (2019).
42. Lin, Y. H., Forman-Kay, J. D. & Chan, H. S. Theories for sequence-dependent phase behaviors of biomolecular condensates. *Biochemistry* <https://doi.org/10.1021/acs.biochem.8b00058> (2018).
43. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.TM118.001190> (2019).
44. Babinchak, W. M. & Surewicz, W. K. Liquid-liquid phase separation and its mechanistic role in pathological protein aggregation. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2020.03.004> (2020).
45. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* <https://doi.org/10.1016/j.cell.2015.09.015> (2015).
46. Zhu, G. et al. Phase separation of disease-associated SHP2 mutants underlies MAPK hyperactivation. *Cell* <https://doi.org/10.1016/j.cell.2020.09.002> (2020).
47. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt1012> (2004).
48. Santos, M., Fidalgo, A., Varanda, A. S., Oliveira, C. & Santos, M. A. S. tRNA deregulation and its consequences in cancer. *Trends Mol. Med.* <https://doi.org/10.1016/j.molmed.2019.05.011> (2019).
49. Torres, A. G., Battle, E. & Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol. Med.* **20**, 306–314 (2014).
50. Koldewey, P., Stull, F., Horowitz, S., Martin, R. & Bardwell, J. C. A. Forces driving chaperone action. *Cell* <https://doi.org/10.1016/j.cell.2016.05.054> (2016).
51. Kirchner, P. et al. Proteome-wide analysis of chaperone-mediated autophagy targeting motifs. *PLoS Biol.* <https://doi.org/10.1371/journal.pbio.3000301> (2019).
52. Pawar, A. P. et al. Prediction of ‘aggregation-prone’ and ‘aggregation-susceptible’ regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2005.04.016> (2005).
53. Rousseau, F., Serrano, L. & Schymkowitz, J. W. H. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2005.11.035> (2006).
54. Varadi, M., Zsoltyomi, F., Guharoy, M., Tompa, P. & Levy, Y. K. Functional advantages of conserved intrinsic disorder in RNA-binding proteins. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0139731> (2015).
55. Boeynaems, S. et al. Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2017.02.013> (2017).
56. Lu, J. & Deutsch, C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2008.08.089> (2008).
57. Close, P. et al. Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol. Cell* **22**, 521–531 (2006).

58. Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. & Hodges, R. S. Relationship of sidechain hydrophobicity and  $\alpha$ -helical propensity on the stability of the single-stranded amphipathic  $\alpha$ -helix. *J. Pept. Sci.* <https://doi.org/10.1002/psc.310010507> (1995).

### Acknowledgements

We are grateful to the GIGA-proteomics, imaging, genomics, and viral vector facilities for their assistance. This study was supported by an Incentive Grant for Scientific Research (MIS F:4532.13) from the FNRS, the Concerted Research Action Program (tRAME), and Special Research Funds (FSR: C-15/44) at the University of Liege, the Belgian Foundation against Cancer (F/2016/840 and F/2019/097), as well as by the Walloon Excellence in Life Sciences and Biotechnology (WELBIO). We also thank the National Natural Science Foundation of China (NSFC, No. 81803581) and Shanghai Institutions of Higher Learning-2018 (TP2018080, JZ2018005). We are also grateful to the "Foundation Leon Fredericq" for its financial support. N.E.H. is Research Fellow, F.R. and C.D. are Research Associate, A.C. is Research Director, and P.C. is Senior Research Associate at the FNRS, respectively.

### Author contributions

F.R. and P.C. designed the study and analyzed experimental data; F.R., Z.Z., G.V., P.L., M.J., M.G., and P.C. performed software and bioinformatic analyses; F.R., Z.Z., A.M.R.S., N.E.H., S.L., and K.S. performed experiments; G.M. and D.B. performed proteomics, P.C. provided funding. A.C., G.H., C.D., L.G., and P.C. provided resources; F.R. and P.C. wrote the paper. All authors discussed the results and commented on the manuscript.

### Competing interests

A.C., P.C., and F.R. are named on a Belgian patent application (BE2021/5237) related to the cell-based method described in this manuscript. The other authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22254-5>.

**Correspondence** and requests for materials should be addressed to F.R. or P.C.

**Peer review information** *Nature Communications* thanks Juana Diez Anton and other, anonymous, reviewer(s) for their contributions to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021