

System-wide analysis of RNA and protein subcellular localisation dynamics

Author list

Eneko Villanueva^{1,†}, Tom Smith^{1,2,†}, Mariavittoria Pizzinga^{2,3†}, Mohamed Elzek², Rayner M. L. Queiroz¹, Robert F. Harvey², Lisa M Breckels¹, Oliver M. Crook⁴, Mie Monti², Veronica Dezi², Anne E. Willis^{*2} & Kathryn S. Lilley^{*1}

Affiliations

1. Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, CB2 1QR, Cambridge, UK
2. MRC Toxicology Unit, University of Cambridge, University of Cambridge, CB2 1QR, Cambridge, UK
3. Structural Biology Research Centre, Human Technopole, 20157 Milan, Italy
4. Department of Statistics, University of Oxford, OX1 3LB, Oxford, UK

Author List footnotes

* Lead contacts

† These authors contributed equally and are in descending alphabetical order

Correspondence

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts, Kathryn Lilley k.s.lilley@bioc.cam.ac.uk or Anne Willis aew80@cam.ac.uk

Abstract

Although the subcellular dynamics of RNA and proteins are key determinants of cell homeostasis, their characterisation is still challenging. Here we present an integrative framework to simultaneously interrogate the dynamics of the transcriptome and proteome at subcellular resolution by combining two methods: Localisation of RNA (LoRNA) and a streamlined density-based Localisation of Proteins by Isotope Tagging (dLOPIT) to map RNA and protein to organelles (nucleus, ER and mitochondria) and membraneless compartments (cytosol, nucleolus and cytosolic granules). Interrogating all RNA subcellular locations at once enables system-wide quantification of the proportional distribution of RNA. We obtain a cell-wide overview of localisation dynamics for 31839 transcripts and 5314 proteins during the unfolded protein response (UPR), revealing that ER-localised transcripts are more efficiently recruited to cytosolic granules than cytosolic RNAs, and that the translation initiation factor eIF3d is key to sustaining cytoskeletal function. Overall, we provide the most comprehensive overview to date of RNA and protein subcellular localisation dynamics.

Main Text

Introduction

Compartmentalisation of eukaryotic cells and the dynamic distribution of macromolecules such as RNA and proteins across these compartments are vital for cell function. The regulation of protein production involves spatially restricted RNA-protein interactions to regulate post-transcriptional processes¹, including the creation of translation “hotspots” so that newly synthesised proteins can act at precise locations without disturbing cellular protein homeostasis². The ability to determine RNA and proteins subcellular localisation, and how they relocalise upon perturbation, is thus key to understanding cellular homeostasis^{3,4}. While cell-wide methods to study protein localisation are well established^{5–7}, equivalent methods to study RNA subcellular localisation are limited.

The localisation of a single RNA transcript can be determined using smFISH⁸. Alternatively, the RNA content of specific niches can be explored using proximity-dependent biotinylation techniques, such as APEX-RIP and APEX-seq^{9–11}, or by multiplexing smFISH in combination with immunofluorescence for a localisation marker¹². However, all these approaches are restricted to the interrogation of a single subcellular compartment per experiment and thus cannot provide a cell-wide view of RNA localisation. Combining multiple APEX experiments to generate a more complete understanding of RNA localisation is time-consuming and does not quantify the proportion of RNAs at each localisation. Previous attempts to determine the complete localisation of RNA in a cell-wide manner have been restricted by technical limitations, including localisation-independent RNA clustering and length-dependent RNA localisation biases^{13–15}. Therefore, there is a need to develop robust methods to obtain comprehensive RNA subcellular localisation information, ideally in a framework which also enables determination of protein subcellular localisation, so that the two biomolecules can be studied simultaneously.

Here, we develop such a framework that returns simultaneous cell wide spatial overviews of the transcriptome and proteome. To enable the creation of this framework, we have firstly developed a transformative method to study cell-wide subcellular localisation of RNA (termed LoRNA). LoRNA builds on the principle of reduction and reconstruction of the cellular content¹⁶ to study RNA localisation. It involves reordering the subcellular constituents by their density, quantification of RNA abundance profiles across the density fractions, and using localisation-specific RNA correlation profiles to determine the subcellular localisation of the complete transcriptome. Importantly, unlike previous methods which estimate localisation-specific enrichments, LoRNA allows the estimation of RNA proportions in each localisation, providing more biologically meaningful information. Secondly, we have engineered LoRNA to allow the simultaneous interrogation of the subcellular proteome, using a novel streamlined Localisation Of Proteins by Isotope Tagging (LOPIT^{5,6}) approach, which we refer to as density-based LOPIT (dLOPIT).

This integrative approach for the cell-wide quantification of RNA and protein distribution is especially valuable to study molecular relocalisation processes. Therefore, we chose to examine the redistribution of RNA and protein during the activation of the unfolded protein

response (UPR). The UPR is triggered by the accumulation of unfolded proteins in the ER lumen¹⁷. Its activation reduces global protein synthesis rates, resulting in the loss of RNA targeting to the ER, the formation of stress granules (SGs), and the upregulation of stress-response genes¹⁸. Importantly, UPR deregulation is associated with disease states including neurodegeneration¹⁹, cancer progression²⁰ and diabetes²¹. Combining LoRNA and dLOPIT, we simultaneously quantified the extensive reorganisation of the transcriptome and proteome upon the activation of the UPR and resultant inhibition of translation, including the loss of RNAs from the ER and the relocalisations of proteins from the secretory pathway. Our integrative approach enabled the discovery of cytosolic granules in our multiomic framework and reveals that ER-localised RNAs are recruited to granules to a greater extent than previously thought^{22–24}. Our data also show that RNAs encoding cytoskeletal proteins are retained and targeted to the periphery of organelles during UPR and that this process involves eIF3d. Importantly, while messenger RNAs undergo a profound subcellular reorganisation upon UPR activation, non-coding RNAs sharing the same properties (eg. size and nucleotide composition) do not, indicating the importance of *trans* factors in determining RNA localisation. Altogether, this uniquely integrative approach has allowed us to generate the most comprehensive analysis of RNA and protein cell-wide localisation dynamics to date, which can be readily explored using dedicated open-source resources: https://proteome.shinyapps.io/density_lorna_rnaloc_gene/ and https://proteome.shinyapps.io/density_dlopit_protloc/.

Results

Cell-wide RNA and protein subcellular localisation

Organelles can be separated based on their physicochemical properties to determine the localisation of macromolecules^{5,25,26}. Unfortunately, the application of this principle to interrogate RNA has so far failed to capture the complete subcellular distribution of RNA^{13,14}. To resolve this issue, we developed a new approach to sort the entire cellular content based on the density of its constituents (Fig. 1a). We applied this framework to U-2 OS cells. Cells were first mechanically lysed with a ball bearing homogeniser, using a 12 µm pore to lyse the cells while preserving the integrity of the intracellular organelles. The complete cell lysate was then loaded into a precast equilibrium gradient (1 to 1.6 g/ml, see experimental details). Crucially, unlike previous methods that use density gradients to resolve subcellular compartments¹³, the cell lysate was loaded at the density of free RNA (1.17 g/ml). This ensured that organelle-associated RNAs migrated upwards in the density gradient, while cytosolic RNA-protein complexes migrated downwards, allowing free RNA and organelles to reach equilibrium faster and avoiding cross-contamination between cellular compartments (Supplementary Fig. 1a). After 16h of centrifugation at 100,000 g, three distinct bands were observed in the gradient, which were enriched in i) Endoplasmic Reticulum (ER) and mitochondria, ii) nucleus, iii) cytosolic proteins, respectively (Supplementary Fig. 1b,c).

To determine whether this subcellular fractionation approach allowed analysis of the subcellular localisation of RNA, we performed RNA-Seq along the gradient and evaluated the sedimentation profile of gene products known to localise to specific subcellular compartments based on previous targeted RNA localisation experiments^{10,27,28} and known nuclear and cytoplasmic lncRNAs (see experimental details). Importantly, when analysing correlation profiles, the fractions obtained are not purified organelle/compartments; distinct sedimentation profiles for each subcellular niche are required to determine localisation. We observed distinct profiles resulting in separate clustering within the data for RNAs known to localise to the mitochondria, ER, nucleolus, nucleus and cytosol (Fig. 1b-d Supplementary Fig. 1D, Supplementary table 1, 2). Further to these expected profiles, a novel sub-cytosolic RNA sedimentation profile was discovered by semi-supervised clustering. The RNAs of this profile are enriched at a density higher than the organelles, but lower than the cytosolic ribosomes (Fig. 1b, Supplementary Fig. 1d-f). Thus, we refer to it as 'cytosol light'. To our knowledge, this is the first time the complete subcellular localisation of RNA has been successfully resolved at a cell-wide level. We refer to this method to interrogate the localisation of RNA as LoRNA.

Next, we evaluated if this fractionation approach was compatible with the simultaneous analysis of the proteome localisation, using LOPIT analysis methods²⁹. We extracted protein from the same fractions used for LoRNA, and quantified the protein abundance by mass spectrometry (Fig 1e). To assess the accuracy of our protein localisation, we used a support vector machine to classify marker proteins for all major compartments based on their abundance profiles. F1 scores (the harmonic mean of precision and recall) of 0.71 - 1 were observed, demonstrating high resolution of protein subcellular localisations, based purely on the density of the different compartments (dLOPIT, Fig. 1f, g, Supplementary Fig. 1g).

Remarkably, dLOPIT performs comparably to our previous high-resolution LOPITs^{5,6} that were engineered to exclusively interrogate the protein subcellular localisation (Supplementary Fig. 1h). These results demonstrate that this integrative framework allows, for the first time, the simultaneous analysis of the subcellular localisation of the transcriptome and proteome.

Proportional quantification of RNA localisation

Relative RNA abundances were adjusted with respect to spike-in RNAs to account for the total RNA content per fraction and obtain profiles of absolute RNA abundance. This allowed the deconvolution of the RNA profiles into the constituent contributions from each localisation, as determined from the profiles of RNAs with known localisations (Fig. 2a). In this way, we estimated the proportional localisation of 31839 transcript isoforms (13142 genes) in every localisation simultaneously. Additionally, transcript-level RNA localisation analysis retains specific information on the differential localisation of splicing variants (Sup Fig. 2a). In order to account for splice isoform localisation variability, we use transcript-level proportions, unless specified otherwise. Importantly, we took advantage of the extreme localisation of the 13 mitochondrially-encoded mRNAs to accurately quantify transcript membrane association, using these RNAs as a reference for 100% membrane localisation. As anticipated, RNA is distributed along two major axes: Nucleus:Cytosol and Cytosol:Membrane, with expected proportions obtained for RNAs known to be extremely enriched in specific locations (Fig. 2b).

To comprehensively validate our system-wide RNA localisation results, we developed an orthogonal method to sort the cellular content by a different physicochemical property, namely the sedimentation coefficient of the different organelles instead of density. Previous attempts to study RNA localisation using this concept applied high *g*-forces^{14,30}, and analysis of the published data highlights that this creates an RNA length-dependent localisation bias; longer RNAs precipitate with organelles and are depleted from the cytoplasm (Supplementary Fig. 2b,c). We developed a fractionation approach that separates the major subcellular localisations while avoiding the RNA length-dependent sedimentation bias. (Fig. 2c & Supplementary Fig. 2b,c,d, Supplementary table 1, 2). While the low *g*-force required to avoid this bias precludes sub-cytosolic resolution, differential centrifugation provides better separation between mitochondria and ER than density centrifugation (Fig. 2d). To assess the agreement between our two orthogonal approaches, we projected the density-based RNA proportions onto the sedimentation-based data. Notably, the two fractionation approaches showed a remarkable agreement for membranes, nucleus, and cytosol RNAs (Fig. 2e). In addition to the excellent separation of the main RNA localisation niches, reliable estimates of cytosol and membrane proportions were also achieved (Fig. 2f, Supplementary Fig. 2e). These proportions show Pearson's correlations of 0.77 and 0.83, respectively, confirming the reliability of our estimates. Importantly, LoRNA provides accurate estimates, even when compared to targeted single-localisation enrichment techniques, including APEX-seq and immunofluorescence-aided MERFISH (Supplementary Fig. 2f-i). Thus, LoRNA provides an accurate quantitative system-wide overview of RNA localisation.

Key RNA features drive subcellular localisation

We next explored the RNA features associated with proportional localisation. While mRNAs are observed throughout the cell, and relatively depleted from the nucleus, lncRNAs were

observed to be strictly within the nucleus:cytosol axis (Fig. 3a). Many lncRNAs have well described nuclear functions e.g. as scaffolds for histone modification complexes²⁷, and nuclear/cytoplasm fractionation RNA-Seq experiments indicate lncRNAs are relatively enriched in the nucleus²⁷. As expected, in our data set lncRNAs were more nuclear localised than mRNAs³¹. However, in line with a recent re-analysis of fractionation RNA-Seq which accounted for the total RNA content per fraction³², our quantification of localisation proportions indicates that lncRNAs are still predominantly cytosolic (Fig. 3b). A comparison of the proportions between our two cell fractionation approaches demonstrated that 59.7% of lncRNAs were consistently cytosolic. It has been shown previously that some lncRNAs have coding potential³³. Consistently, we identified that ribosome association correlates with greater lncRNA cytosol localisation³⁴ (Supplementary Fig. 3a). Importantly, we found that cytosolic lncRNAs are generally shorter, have a lower AU content, and are more likely to be polyadenylated, compared to nuclear lncRNAs (Fig. 3c). A simple model using these features is as accurate as a more complex model built from a penalised regression over a wide range of potential features (ROC AUC = 0.86 for both models; Supplementary Fig. 3b). We therefore suggest that these are the key features driving cytosolic localisation for lncRNAs (Fig. 3c).

While generally we found mRNAs predominantly in the cytosol, RNAs encoding proteins with signal peptides and/or transmembrane domains (TMs) were much more prominently membrane localised, consistently with their co-translational targeting to the ER (Fig. 3d, e). Furthermore, our data confirmed the relationship between the distance from the first signal peptide/TM domain to the stop codon and the membrane proportion (Supplementary Fig. 3c)^{28,35}. Surprisingly, we also identified 87 membrane-localised mRNAs that did not encode a signal peptide or TM (Supplementary table 2). These included the mRNA encoding the ER-localised signal recognition receptor subunit beta, which co-translationally binds to the alpha subunit³⁶, and mRNAs encoding for mitochondrial proteins (Supplementary Fig. 3d), but most of the RNAs were not previously described as membrane localised. GO analysis revealed a significant enrichment for terms associated with membrane or cytoskeleton localisation (Fig. 3f), suggesting potential localised translation at the surface of the membranes. To confirm ER-association of these RNAs, we combined smFISH/IF with a digitonin-based cytosol extraction approach, focusing on two mRNAs which don't encode signal peptides or TM domains, Microtubule-actin cross-linking factor 1 (MACF1) and Dystonin (DST) RNAs. Both mRNAs are only marginally affected by digitonin treatment, similarly to known ER-associated transcript PIGT (Fig. 3g, h). By contrast, cytosolic transcript Calpain2 is strongly affected by the treatment. These results are in agreement with LoRNA data and indicate that LoRNA can accurately uncover novel, non-canonical ER-associated transcripts.

Assessing global transcriptome and proteome relocation

System-wide RNA localisation is especially suited for studying RNA redistribution upon stimulation, where transcripts can migrate to unexpected localisations. As a showcase, we used LoRNA to interrogate RNA relocation upon activation of the UPR, which was induced via inhibition of the SERCA Ca^{2+} pump by treating U-2 OS cells with 250nM of Thapsigargin for 1 h. This reduces calcium in the ER lumen, impairing protein folding and activating the UPR, which, in turn, rapidly induces a global translational shutdown through phosphorylation of eIF2 α (Supplementary Fig. 4a,b). This results in the formation of stress granules (SGs; Supplementary Fig. 4c) and the increased expression of UPR genes including XBP1 and

CHOP (Supplementary Fig. 4d). LoRNA allows quantification of the cell-wide redistribution of RNA upon inhibition of translation, clearly showing a profound RNA migration from the membranes towards the cytosol (Fig. 4a, Supplementary Fig. 4e). Surprisingly, we found a pronounced relocalisation of RNAs to the 'cytosol light', both from the cytosol and from membranes (Fig. 4b). We therefore hypothesised that the cytosol light profile may represent RNP granules.

To test this hypothesis and gain a deeper understanding of the cellular reorganisation upon UPR activation, we exploited our fractionation approach to yield simultaneous data on proteome relocalisation by applying dLOPIT to the same gradient fractions that had been used for LoRNA. A Bayesian analysis framework³⁷ was applied to identify protein relocalisation 1hr post UPR activation, at three levels of confidence depending on the differential localisation probabilities, 'highly confident' (>99%), 'confident' (>95%) and 'candidate' (>85%). This identified 73 proteins that were differentially localised with confidence (Fig. 4c,d; Supplementary table 3). The Golgi apparatus was the most affected organelle, as expected given its role in the trafficking of newly synthesised proteins. Interestingly, four stress granule proteins (STAU2, UPF1, PABPC1 and PABPC) relocalised away from the ribosomes towards the fraction that differentiates the cytosol light (Fig. 4e, Supplementary Fig. 4f). We used these 4 proteins to identify a further 22 proteins with correlated UPR profiles; including 5 members of the P-body associated CCR4-NOT complex³⁸, P-body proteins DCP1A/B³⁹ and stress granule components YBX3⁴⁰, SECISBP2⁴⁰ and CASC3⁴¹, further suggesting that the cytosol-light profile represents cytosolic granules. To confirm that cytosolic granules are enriched in the fractions discriminating the cytosol-light profile, we engineered U-2 OS cells to express GFP-tagged DCP2 and G3BP1 proteins. DCP2 is a canonical P-body protein⁴², while G3BP1 is a well-known constituent of SGs⁴³. By performing density-based cell fractionation experiments on cells expressing these two engineered proteins and imaging them along the gradient we confirmed that cytosolic granules sediment at the density ranges of 'cytosol light' (Supplementary Fig. 4g, h).

RNA features driving granule localisation

The proteins relocalising to the cytosol light upon UPR are associated with RNA condensation into granules, therefore we further characterised the RNA composition of this fraction and interrogated the features of the enriched RNAs. RNAs with higher cytosol light abundance in unstressed cells were correlated with lower ribosome association and longer transcript length (Supplementary Fig. 5a, b), both features associated with RNAs partitioning to granules^{44,45}. Furthermore, RNAs enriched in P-bodies⁴⁶ and TIS granules⁴⁷ using fluorescence-activated particle sorting have a similar distribution to the cytosol light profile (Supplementary Fig. 5c, d). Finally, we observed a positive correlation between the cytosol light proportions upon UPR activation and SG enrichment upon arsenite treatment using targeted SG purification²³ (Supplementary Fig. 5e). Altogether, these data provide strong evidence that our cytosol light profiles represent cytosolic ribonucleoprotein-containing granules, and we henceforth refer to this localisation as 'granule'.

As expected, mRNAs are recruited to granules upon UPR activation in a length and AU content-dependent manner. Surprisingly, lncRNAs do not relocalise to granules, irrespective of these features (Fig. 5a-c). Importantly, LoRNA provides cell-wide quantification of RNA

localisation prior and post stimulation. This allowed us to uncover that upon activation of the UPR, although many cytosolic mRNAs migrate to granules, membrane mRNAs do so at a higher proportion (Fig. 5d). However, not all membrane mRNAs relocate to granules (Fig. 4b, Supplementary Fig. 5f). The recruitment of specific mRNAs to granules was confirmed with smFISH (Fig. 5e) and agrees with a recent targeted study of the relocalisation of six RNAs upon UPR induction (Supplementary Fig. 5g)⁴⁸. To fully characterise the features driving mRNAs towards granules upon UPR activation, we modelled the contribution of a broad range of RNA features, including transcript length, localisation prior to stress, RBP binding (from eCLIP data⁴⁹), k-mer content, codon usage, and presence of IRES or upstream ORFs. Overall transcript length, especially longer 3'UTRs in membrane mRNAs, was observed to be the greatest positive predictor of relocalisation to granules (Fig. 5f). RBPs whose binding according to eCLIP was positively predictive of RNA relocalisation to granules included the granule proteins FAM120A, IGF2BP1, and G3BP1 (Fig. 5g). Moreover, the binding of canonical stress granule proteins TAIL1, TIA1 and IGFBP3 to membrane-localised mRNAs also showed a significant predictive value. Intriguingly, unexpected associations between RBP binding and relocalisation to granules were observed, including ZNF622, which is involved in ribosome subunit joining⁵⁰ and ribosome stalling⁵¹, and upregulated upon viral infection and UPR stress⁵². Importantly, ZNF622 was a 'candidate' relocalising protein, with movement from the cytosol to the ER upon UPR (Supplementary Fig. 5h, i), a step that may be required for the membrane-RNA recruitment to granules.

A higher AU content in the coding and 3'UTR regions was predictive of greater relocalisation of mRNAs to granules (Fig. 5h). The latter is consistent with the known role of 3'UTR AU-rich elements (AREs) in regulating RNA stability. Interestingly, higher G content, but not C content, in the coding region was a predictor of lower granule relocalisation (Fig. 5h). In addition, a higher frequency of AUG k-mers in the 5'UTR was predictive of lower relocalisation, although presence of annotated uORFs was not. This may indicate that the presence of an AUG at the 5'UTR inhibits granule association, regardless of the ORF capacity of the sequence. Finally, while overall codon optimality was not a selected predictive feature, the frequency of UUA, GAU, AAU, and CAC codons were predictive of granule relocalisation (Fig. 5i). Three of these are NAC/U codons, decoded by queuosine-modified tRNAs. Notably, the two NAU codons are predictive of greater relocalisation to granules and the NAC codon predictive of lower relocalisation, pointing to a potential role of this modification in the stress response and RNA recruitment to cytosolic granules.

RNAs retaining organelle association upon UPR

Despite the global loss of RNA from the membranes, some RNAs remain membrane-associated upon UPR activation (Fig. 6a). We modelled this non-linear relationship using a generalised additive model (GAM) with a cubic regression spline and used the residuals from the GAM to identify RNAs with unexpectedly high membrane localisation after the induction of the UPR. Notably, RNAs with longer sequences between the signal sequence and stop codon are retained more efficiently in membranes, suggesting the involvement of the signal recognition particle (SRP) for these RNAs³⁵ and therefore active reassociation with the membranes (Fig. 6b). However, many of the membrane-localised RNAs that do not encode a transmembrane domain or signal peptide (Fig. 3) were retained at the membrane upon UPR activation (Supplementary table 2), indicating an SRP-independent mechanism for their retention. To gain an understanding of the mechanisms of RNA retention in the periphery of organelles upon UPR activation, we investigated protein-RNA interactions using publicly available eCLIP data⁴⁹. Applying a Lasso penalisation model to select the proteins whose binding predicts RNA localisation, we observed that the binding of 11 out of 177 RBPs was predictive of greater membrane localisation, including eIF3h and eIF3d (Fig. 6c, Supplementary Fig. 6a). Surprisingly, genes whose translation is affected by the knockdown of the core eIF3 component, eIF3e⁵³ do not show a differential membrane retention upon UPR activation (Supplementary Fig. 6b, c), suggesting that membrane retention may be specific to specialised eIF3 components. Interestingly, eIF3d has been found to maintain translation upon ER stress⁵⁴ and can directly bind at the 5' cap⁵⁵. We confirmed that RNAs known to be bound by eIF3d at their 5' cap⁵⁶ (Fig. 6d) are more retained in the membranes upon UPR activation, indicating that 5' cap binding of eIF3d may contribute to the maintenance of this localisation. Membrane-localised RNAs bound by eIF3d show an over-representation for GO terms relating to the actin cytoskeleton, cell morphogenesis, and focal adhesions (Fig. 6e). These membrane-localised eIF3d-bound RNAs include two cytoskeletal protein coding transcripts, MACF1 and DST transcripts, which we previously showed are amongst the cytoskeletal-protein encoding transcripts unexpectedly associated with the ER (Fig. 3g). Using detergent-based subcellular fractionation, we further validated that the membrane-localisation of MACF1 and DST is eIF3d-dependent (Fig. 6f, Supplementary Fig. 6e,f).

Finally, we hypothesised that the retention of eIF3d bound transcripts in the ER upon UPR induction may be required to remodel the cytoskeleton in processes that depend on localised translation e.g. cell migration, particularly during activation of the UPR. In agreement with this, specifically decreasing eIF3d expression reduces cell migration overall, with a significantly stronger effect in cells undergoing UPR (Fig. 6g,h, Supplementary Fig. 6 g-j). While eIF3d has been shown to bind cytoskeletal RNAs⁵⁶ and cytoskeleton remodelling is required for calcium homeostasis upon UPR induction⁵⁷, to our knowledge, this is the first time these two processes have been connected as potentially being part of the same integrated response. Altogether, our results suggest that eIF3d is required for continued translation of actin cytoskeleton components localised in the periphery of membranes during the UPR.

Discussion

In this work, we present an integrative multiomics subcellular localisation framework composed of two different methods (LoRNA and dLOPIT) to generate the first cell-wide analysis of RNA and protein subcellular localisation in membranous (nucleolus, ER and mitochondria) and membraneless (cytosol, nucleolus and cytosolic granules) compartments. By precisely reconstructing RNA localisation, our transformative approach allows the quantitative determination of the complete subcellular distribution of each RNA. We combined LoRNA and dLOPIT to characterise the dynamic transcriptome and proteome subcellular redistribution upon UPR. This established the RNA features driving relocalisation to granules, identify those transcripts that are targeted to the periphery of the organelles during the integrated stress response, and reveal the role of eIF3d in maintaining cytoskeleton function upon UPR.

When interpreting LoRNA proportions, it is important to consider that the RNA marker profiles that are used to estimate localisation proportions do not represent spatial coordinates in the cell, but rather the typical profile for RNAs which predominantly reside in that localisation. For example, the distribution of the cytosolic RNA markers includes nascent RNA copies that are also nuclear-localised. As such, proportions are with respect to RNAs that are paradigmatic representatives of localisation, and do not represent absolute localisations. Another important consideration is that, while RNA nuclear export is a tightly regulated process, small proteins diffuse freely. Therefore, when interpreting cell fractionation experiments it is important to consider that small nuclear proteins with weak interactions can diffuse out from the nucleus⁵⁸. Fixing protein-protein interactions could overcome this limitation, however, cross-linking agents like formaldehyde limit organelle separation. Development of reagents to efficiently stabilise protein-protein interactions without affecting organelle integrity and lipid content should overcome this limitation. Here, we identified nucleoplasm protein markers reflecting this phenomenon to ensure nucleoplasm proteins are correctly classified to the nucleus. Furthermore, in cases where gradients are of interest, such as across the cytosol of a polarised cell, targeted methods such as FISH or IF are required as any high-throughput method involving cell lysis will necessarily result in a loss of subcellular coordinates. When RNA compartments not amenable to biochemical fractionation are of interest (e.g. the nucleopore), alternative methods like APEX-Seq may be applicable. However, it is worth noticing that multiple APEX-based experiments need to be combined to determine RNA localisation at a cell-wide level, making this approach incompatible with the proportional estimation of RNA localisation and especially challenging to apply in dynamic systems or to study multiple biological systems. Nonetheless, APEX and LoRNA can function as complementary approaches to interrogate molecular subcellular distribution at different resolutions.

One of our most striking findings regarding RNA relocalisation after UPR activation is that membrane-localised mRNAs are recruited more efficiently to granules than cytosolic mRNAs, while lncRNAs do not relocalise to granules. We show that the two features most associated with mRNA recruitment to granules, length and AU content do not affect lncRNA relocalisation to granules. This suggests that RNA recruitment to granules may depend on more than RNA length and ribosome occupancy as previously stated²³ or that there is a mechanism to exclude lncRNAs from granules. Notably, membrane-localised mRNA recruitment to granules conflicts with transcriptomic studies that have claimed ER-targeted RNA are depleted from SGs²³.

However, targeted methods to purify SGs typically focus on specific engineered bait proteins and require multiple purification steps, while LoRNA recovers granules regardless of their specific composition, based on their distinct density sedimentation profile, which may explain the apparent discrepancy between SG transcriptomics and LoRNA. Since our results suggest ER-RNA relocalisation to granules during stress is widespread, we speculate that novel subtypes of stress-induced granules may be the ultimate destination for membrane-associated RNAs upon UPR activation, opening new avenues for future research. Furthermore, we observed previously uncharacterised predictors of RNA relocalisation to granules. Notably, the relocalisation of RNA to granules is correlated with frequencies for 3 NAU/C codons that are decoded by queuosine-modification of the G in the wobble position of the cognate tRNA, with opposite effects for NAU and NAC codons. Considering the emerging role that tRNA modifications play in translation^{59,60}, it would be interesting to explore if a tRNA-specific co-translational mechanism could regulate the recruitment of RNAs to SGs.

Using dLOPIT under UPR induction, we revealed the relocalisation of specific proteins, many of which having been previously implicated in the formation of P-bodies and stress granules. We also observed relocalisation of ZNF622 to the ER upon UPR and that RNAs containing ZNF622 binding sites relocalise more readily to granules. Given that ZNF622 has been recently shown to inhibit ribosome subunit joining⁵⁰, and may be involved in the ER associated degradation (ERAD) machinery or ER quality control following viral infection⁵², ZNF622 relocalisation to the ER may be required to inhibit ER-localised translation and drive ER-targeted RNAs to granules upon UPR activation.

The precise measurement of RNA proportionality afforded by LoRNA allowed characterisation of the relocalisation of RNA away from membranes in unprecedented detail. Intriguingly, we found that despite the general loss of mRNA from the ER upon induction of the UPR⁶¹, many RNAs are partially retained in association with organelles, or even migrate to the ER. Notably, we found that many of these mRNAs encode for proteins involved in cytoskeletal remodelling. Furthermore, eIF3d binding correlates with membrane localisation and decreasing its expression impairs cell migration, a process highly dependent on cytoskeletal remodelling. Actin cytoskeleton remodelling plays a key role to restore Ca^{2+} homeostasis in the ER upon UPR induction through promotion of ER-PM contact⁵⁷. Separately, eIF3d has been found to maintain RNA translation upon stress⁵⁴. Altogether, we speculate that the cytoskeleton remodelling required to overcome UPR is eIF3d-dependent and that it involves the *in-situ* translation of cytoskeletal mRNAs.

In summary, our framework, combining LoRNA and dLOPIT, provides a quantitative system-wide determination of RNA and protein relocalisation. This has allowed us to evaluate the transcriptome and proteome dynamics during UPR induction with unprecedented resolution. The resulting data has invaluable potential for future studies characterising the functions of specific RNAs and proteins. To facilitate this, we have generated a user-friendly graphical interface to explore our data available at https://proteome.shinyapps.io/density_lorna_rnaloc_gene/ and https://proteome.shinyapps.io/density_dlopit_protloc/. Notably, the LoRNA method allows for the unbiased cell-wide determination of RNA compartmentalisation. This will support a paradigm shift from the study of RNA localisation through relative enrichments between two localisations, towards a cell-wide analysis of RNA proportional distribution. We anticipate others will build upon LoRNA. For example, coupling it with metabolic labelling such as 4-

Thiouridine, and direct RNA sequencing methods, would enable the precise temporal interrogation of the role of RNA modifications in the regulation of RNA localisation. The simultaneous characterisation of proteome and transcriptome relocalisation represents a transformative approach to study how the cell coordinately responds to physiological signals, stress conditions, exogenous cues, or infectious pathogens. Importantly, this approach can contribute to the translation of molecular biology observations to medical benefits by fostering new studies into the role of RNA and protein localisation dynamics in cell homeostasis and disease.

Acknowledgements

We thank Catarina De Matos Ferraz Franco of the MRC-Toxicology Proteomics Facility for collection of all mass spectrometry data and Dee Scadden for kindly providing the G3BP1-GFP expression plasmid, B. Fisher for kindly sharing equipment and Maria Marti-Solano for assistance with the manuscript writing. We also thank the Advanced Light Imaging facility at the MRC Toxicology Unit and the Light Imaging Facility at the Human Technopole for their support in the acquisition of light microscopy data. E.V., T.S., R.M.L.Q., M.P. and M.E. are supported by Wellcome Trust, grant numbers 110170/Z/15/Z and 110071/Z/15/Z awarded to A.E.W. and K.S.L.; R.F.H., V.D., and M.M. are supported by the Medical Research Council, grant number 5TR00. L.M.B. is supported by EU Horizon 2020 programme INFRAIA project EPIC-XS (Project 823839). O.M.C. was supported by a Wellcome Trust Mathematical Genomics and Medicine studentship funded by the Cambridge School of Clinical Medicine and by the Todd-Bird Junior Research Fellowship from New College, Oxford.

Author Contributions Statement

Conceptualization, E.V., T.S., A.E.W. and K.S.L.; Methodology, E.V., T.S., M.P. and M.E.; Investigation, E.V., T.S., M.P., M.E., R.M.L.Q. and R.F.H.; Writing - Original Draft, E.V., T.S. and M.P.; Resources, L.M.B.; Data Curation, T.S.; Writing - Review & Editing, E.V., T.S., M.P., M.E., R.M.L.Q., R.F.H., O.M.C., M.M., V.D., A.E.W. and K.S.L.; Visualization, E.V., T.S. and M.P.; Supervision, A.E.W. and K.S.L.; Project administration, E.V., T.S., A.E.W. and K.S.L.; Funding acquisition, A.E.W. and K.S.L.

Competing Interests

All authors declare no competing interests.

Copyright

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Figure titles and legends

Figure 1. Simultaneous analysis of RNA and protein subcellular localisation

a, Schematic representation of the subcellular fractionation framework. Cells are lysed and fractionated by density equilibrium centrifugation. RNA and protein are extracted from each fraction to perform LoRNA and dLOPIT. Cell lysate and the gradient banding pattern are represented in yellow. Aqueous and organic phases are blue and red respectively. **b, c**, Application of LoRNA to U-2 OS cells. Mean profiles for RNA markers along pooled gradient fractions in a single experiment. Shaded regions denote $\pm$ one standard error. **d**, PCA projection of RNA profiles across three replicate experiments, with marker RNAs highlighted. **e**, Application of dLOPIT to U-2 OS cells. **f**, Distributions of F1 scores for protein markers for each localisation using support vector machine classification. $n = 50$ iterations. **g**, t-SNE projections for protein localisation profiles across three replicate experiments, with marker RNAs highlighted.

Figure 2. System-wide quantification of RNA localisation

a, Schematic representation of how localisation proportions are estimated. New profile (left) is decomposed into different proportions of each marker profile (middle) to approximate the original profile (right). **b**, Cytosol, membrane and nuclear (including nucleolus) proportions for all transcripts (left) and well-documented localisation markers (right). **c**, Schematic representation of cell fractionation by sedimentation coefficient and linear profiles of the localisation markers. **d**, PCA projections of RNA profiles in (D), with markers highlighted. **e**, Membrane, nucleus and cytosol proportions obtained by equilibrium density centrifugation-based LoRNA projected on the sedimentation-based-RNA localisation results. **f**, Correlations between proportions obtained using different fractionation approaches. Cytosol proportion by density is the sum of cytosol and cytosol-light proportions. Membrane proportion by sedimentation is the sum of mitochondrial and ER proportions.

Figure 3. Features driving RNA localisation

a, Cytosol, nucleus and membrane proportions for mRNAs and lncRNAs. **b**, Empirical cumulative frequency distributions for nucleus proportions for mRNAs and lncRNAs. **c**, Coefficient estimates for logistic regression model of lncRNA cytosol proportions. P-values derived from two-tailed z-test. **d**, Proportions for mRNAs encoding signal peptides and/or transmembrane (TM) domains. **e**, Membrane proportions for mRNAs shown in (D). **f**, GO-terms significantly enriched in mRNAs not encoding signal peptides and/or TM domains but over 35% membrane-localised. P-values derived from a two-tailed hypergeometric test and adjusted for multiple testing using the Benjamini–Hochberg FDR correction. **g** smFISH/IF images of digitonin-extracted (Digitonin +) and control (Digitonin -) U-2 OS cells co-stained for the ER (calnexin, magenta), and MACF1 and DST RNAs (green). Calpain2 and PIGT RNAs are shown as examples of cytosolic transcripts and known ER transcripts, respectively. Scale bar = 10 μ m. **h**, Quantification of RNA transcripts in digitonin-treated cells, using control (untreated) cells for normalisation. At least 30 cells were quantified per condition. **** $P < 0.0001$ (two-tailed Mann-Whitney U test).

Figure 4 Transcriptome and proteome subcellular redistribution upon UPR

a, PCA projection of RNA profiles in control and UPR. Colour and intensity indicate RNA proportions for the primary localisation. PC2 in UPR is inverted. **b**, Cytosol, membrane and cytosol light proportions in control and UPR. Point colour indicates RNA major localisation (excluding nucleus and nucleolus) in control, with colour intensity denoting proportion. **c**, **d** Confident differential protein localisation between control and UPR. **e**, Abundance profiles for stress granule proteins relocalising from ribosomes to the fraction discriminating the cytosol-light-RNA profile under UPR, grey arrows mark cytosol light discriminating fraction.

Figure 5. Analysis of the characteristics driving RNAs to granules

a, Distributions for changes in granule proportions upon UPR activation for lncRNA and mRNAs. Results from a two-sample Kolmogorov–Smirnov test that the distributions are different are shown. **b**, Relocalisation to granules upon UPR activation. RNAs are binned by transcript length and split by gene biotype. lncRNAs and mRNAs are represented as grey and pink dots respectively. Black bar=median. * = $p < 0.001$ two-tailed Wilcoxon rank sum test for lncRNAs vs mRNAs. For exact P-values and n numbers, see supplementary table 4. **c**, As per (b), with bins by AU content. **d**, Relocalisation to granules upon UPR activation. RNAs are binned by transcript length and split by localisation in control condition. Box extends to the 25th and 75th percentiles. Whiskers extend to range, excluding outliers. Outliers are defined as greater than 1.5 x the interquartile range from the box. * = $p < 0.001$ two-tailed Wilcoxon rank sum test for Membrane vs Cytosol. For exact P-values and n numbers, see supplementary table 4. **e**, Representative smFISH/IF images showing PIGT and CD9 RNA localisation relative to the ER (calnexin) in control conditions and relative to SGs (G3BP1) upon UPR activation, n=3 independent experiments. Scale bar = 10 μ m. **f**, Coefficients for features selected by lasso regression model of mRNA relocalisation to granules. Coefficients for RNA length features. **g**, Coefficients for features describing RBP binding or the interaction between RBP binding and localisation in control. RBP eCLIP cell line indicated in the feature name. Positive coefficients mean increased relocalisation to granules. **h**, Coefficients and AU content for k-mer features. **i**, Coefficients for codon features.

Figure 6. Analysis of the RNAs retaining membrane association under UPR

a, Membrane proportions in control and UPR. Yellow line indicates fit from generalised additive model (GAM) with cubic regression spline. **b**, Relationship between residual from GAM and the distance between the first signal or TM domain and stop codon. Blue line indicates smoothed fit by LOESS local regression. **c**, Coefficients for RBPs which are predictors of GAM residuals for RNAs which do not encode a signal peptide/TM domain. RBP eCLIP cell line indicated in the feature name. P-value from ANOVA comparing GAM +/- stratification by eIF3d-binding. **d**, Membrane proportions as per a, for RNAs bound by eIF3d according to eCLIP or Subunit-Seq. **e**, GO terms enriched in membrane RNAs bound by eIF3d, relative to all membrane RNAs. **f**, Cytosolic and membrane associated RNA extraction using digitonin under UPR induction. MTX1, SSR2 and PSK1N are non-eIF3d ER-associated controls. All RNAs are normalised to the MT-ND6 housekeeping transcript. Changes in RNA localisation upon eIF3d siRNA are relative to the RNA localisation in the scrambled siRNA condition. Error bars represent SD. (n=5 independent cytosol and membrane RNA extractions per condition) ** $P=0.0079$ (two-tailed Mann-Whitney U test). **g**, Representative U-2 OS cell migration time series for eIF3d knock-down and siRNA Control (siCt) cells in control and UPR conditions. Cell index denotes impedance measured from migrating cells. **h**, Quantification of cell migration from (g) at 24 h post UPR induction in eIF3d knock-down and siRNA Control cells (n=5 independent experiments). P-value derived from a two-tailed paired t-test.

References

1. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* **8**, 533–543 (2007).
2. Chouaib, R. *et al.* A Dual Protein-mRNA Localization Screen Reveals Compartmentalized Translation and Widespread Co-translational RNA Targeting. *Dev. Cell* **54**, 773-791.e5 (2020).
3. Christopher, J. A., Geladaki, A., Dawson, C. S., Vennard, O. L. & Lilley, K. S. SUBCELLULAR TRANSCRIPTOMICS & PROTEOMICS: A COMPARATIVE METHODS REVIEW. *Mol. Cell. Proteomics MCP* 100186 (2021) doi:10.1016/j.mcpro.2021.100186.
4. Christopher, J. A. *et al.* Subcellular proteomics. *Nat. Rev. Methods Primer* **1**, 32 (2021).
5. Christoforou, A. *et al.* A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 9992 (2016).
6. Geladaki, A. *et al.* Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* **10**, 331 (2019).
7. Mulvey, C. M. *et al.* Spatiotemporal proteomic profiling of the pro-inflammatory response to lipopolysaccharide in the THP-1 human leukaemia cell line. *Nat. Commun.* **12**, 5773 (2021).
8. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
9. Engel, K. L. *et al.* Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *Nucleic Acids Res.* gkab1185 (2021) doi:10.1093/nar/gkab1185.
10. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26 (2019).
11. Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L. & Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-

- RNA crosslinking. *eLife* **6**, e29224 (2017).
12. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci.* **116**, 19490–19499 (2019).
 13. Adekunle, D. A. & Wang, E. T. Transcriptome-wide organization of subcellular microenvironments revealed by ATLAS-Seq. *Nucleic Acids Res.* **48**, 5859–5872 (2020).
 14. Benoit Bouvrette, L. P. *et al.* CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in *Drosophila* and human cells. *RNA* **24**, 98–113 (2018).
 15. Wang, E. T. *et al.* Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* **150**, 710–724 (2012).
 16. de Duve, C. Tissue fraction-past and present. *J. Cell Biol.* **50**, 20 (1971).
 17. Li, H. & Sun, S. Protein Aggregation in the ER: Calm behind the Storm. *Cells* **10**, 3337 (2021).
 18. Pizzinga, M. *et al.* The cell stress response: extreme times call for post-transcriptional measures. *WIREs RNA* **11**, e1578 (2020).
 19. Hughes, D. & Mallucci, G. R. The unfolded protein response in neurodegenerative disorders - therapeutic modulation of the PERK pathway. *FEBS J.* **286**, 342–355 (2019).
 20. Madden, E., Logue, S. E., Healy, S. J., Manie, S. & Samali, A. The role of the unfolded protein response in cancer progression: From oncogenesis to chemoresistance. *Biol. Cell* **111**, 1–17 (2019).
 21. Doultzinos, D. *et al.* Control of the Unfolded Protein Response in Health and Disease. *SLAS Discov. Adv. Life Sci. R D* **22**, 787–800 (2017).
 22. Unsworth, H., Raguz, S., Edwards, H. J., Higgins, C. F. & Yagüe, E. mRNA escape from stress granule sequestration is dictated by localization to the endoplasmic reticulum. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **24**, 3370–3380 (2010).
 23. Khong, A. *et al.* The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol. Cell* **68**, 808-820.e5 (2017).
 24. Matheny, T., Rao, B. S. & Parker, R. Transcriptome-Wide Comparison of Stress

- Granules and P-Bodies Reveals that Translation Plays a Major Role in RNA Partitioning. *Mol. Cell. Biol.* **39**, e00313-19 (2019).
25. Barylyuk, K. *et al.* A Comprehensive Subcellular Atlas of the Toxoplasma Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell Host Microbe* **28**, 752-766.e9 (2020).
26. De Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R. & Appelmans, F. Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochem. J.* **60**, 604–617 (1955).
27. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
28. Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014).
29. Breckels, L. M., Mulvey, C. M., Lilley, K. S. & Gatto, L. A Bioconductor workflow for processing and analysing spatial proteomics data. *F1000Research* **5**, 2926 (2016).
30. Markmiller, S. *et al.* Persistent mRNA localization defects and cell death in ALS neurons caused by transient cellular stress. *Cell Rep.* **36**, 109685 (2021).
31. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
32. Carlevaro-Fita, J. & Johnson, R. Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. *Mol. Cell* **73**, 869–883 (2019).
33. Chen, J. *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).
34. Zeng, C., Fukunaga, T. & Hamada, M. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics* **19**, 414 (2018).
35. Chartron, J. W., Hunt, K. C. L. & Frydman, J. Cotranslational signal independent SRP preloading during membrane targeting. *Nature* **536**, 224–228 (2016).
36. Young, J. C. & Andrews, D. W. The signal recognition particle receptor alpha subunit assembles co-translationally on the endoplasmic reticulum membrane during an mRNA-

- encoded translation pause in vitro. *EMBO J.* **15**, 172–181 (1996).
37. Crook, O. M. *et al.* Inferring differential subcellular localisation in comparative spatial proteomics using BUNDLE. *Nat. Commun.* **13**, 5948 (2022).
38. Cougot, N., Babajko, S. & Séraphin, B. Cytoplasmic foci are sites of mRNA decay in human cells. *J. Cell Biol.* **165**, 31–40 (2004).
39. Ingelfinger, D., Arndt-Jovin, D. J., Lührmann, R. & Achsel, T. The human LSM1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrn1 in distinct cytoplasmic foci. *RNA* **8**, 1489–1501 (2002).
40. Youn, J.-Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69**, 517-532.e11 (2018).
41. Baguet, A. *et al.* The exon-junction-complex-component metastatic lymph node 51 functions in stress-granule assembly. *J. Cell Sci.* **120**, 2774–2784 (2007).
42. van Dijk, E. *et al.* Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* **21**, 6915–6924 (2002).
43. Tourrière, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell Biol.* **160**, 823–831 (2003).
44. Campos-Melo, D., Hawley, Z. C. E., Droppelmann, C. A. & Strong, M. J. The Integral Role of RNA in Stress Granule Formation and Function. *Front. Cell Dev. Biol.* **9**, 621779 (2021).
45. Mazroui, R. *et al.* Inhibition of ribosome recruitment induces stress granule formation independently of eukaryotic initiation factor 2 α phosphorylation. *Mol. Biol. Cell* **17**, 4212–4219 (2006).
46. Hubstenberger, A. *et al.* P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol. Cell* **68**, 144-157.e5 (2017).
47. Horste, E. L. *et al.* Subcytoplasmic location of translation controls protein output. *bioRxiv* (2022) doi:10.1101/2022.11.04.515216.
48. Child, J. R., Chen, Q., Reid, D. W., Jagannathan, S. & Nicchitta, C. V. Recruitment of endoplasmic reticulum-targeted and cytosolic mRNAs into membrane-associated stress

- granules. *RNA N. Y. N* **27**, 1241–1256 (2021).
49. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
50. Lv, K. *et al.* HectD1 controls hematopoietic stem cell regeneration by coordinating ribosome assembly and protein synthesis. *Cell Stem Cell* **28**, 1275-1290.e9 (2021).
51. Liaud, N. *et al.* Cellular response to small molecules that selectively stall protein synthesis by the ribosome. *PLOS Genet.* **15**, e1008057 (2019).
52. Shaban, M. S. *et al.* Multi-level inhibition of coronavirus replication by chemical ER stress. *Nat. Commun.* **12**, 5536 (2021).
53. Lin, Y. *et al.* eIF3 Associates with 80S Ribosomes to Promote Translation Elongation, Mitochondrial Homeostasis, and Muscle Health. *Mol. Cell* **79**, 575-587.e7 (2020).
54. Guan, B.-J. *et al.* A Unique ISR Program Determines Cellular Responses to Chronic Stress. *Mol. Cell* **68**, 885-900.e6 (2017).
55. Lee, A. S., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536**, 96–99 (2016).
56. Lamper, A. M., Fleming, R. H., Ladd, K. M. & Lee, A. S. Y. A phosphorylation-regulated eIF3d translation switch mediates cellular adaptation to metabolic stress. *Science* **370**, 853–856 (2020).
57. van Vliet, A. R. *et al.* The ER Stress Sensor PERK Coordinates ER-Plasma Membrane Contact Site Formation through Interaction with Filamin-A and F-Actin Remodeling. *Mol. Cell* **65**, 885-899.e6 (2017).
58. Wühr, M. *et al.* The Nuclear Proteome of a Vertebrate. *Curr. Biol.* **25**, 2663–2671 (2015).
59. Chan, C., Pham, P., Dedon, P. C. & Begley, T. J. Lifestyle modifications: coordinating the tRNA epitranscriptome with codon bias to adapt translation during stress responses. *Genome Biol.* **19**, 228 (2018).
60. Tuorto, F. *et al.* Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.* **37**, e99777 (2018).

61. Reid, D. W., Chen, Q., Tay, A. S.-L., Shenolikar, S. & Nicchitta, C. V. The Unfolded Protein Response Triggers Selective mRNA Release from the Endoplasmic Reticulum. *Cell* **158**, 1362–1374 (2014).

Online Methods

Cell culture and UPR induction

U-2 OS cells were obtained from the American Type Culture Collection (ATCC), maintained in McCoy's A5 media (Gibco-BRL) supplemented with 10% of FBS (Gibco-BRL), at 37 °C and 5% CO₂, and regularly tested for mycoplasma contamination with negative results. UPR was induced by directly adding 250 nM of Thapsigargin (TG, UPR) or equivalent volume of DMSO (control) to cells at 90% confluency. Cells were incubated with TG or DMSO for 1 h at 37 °C unless specified otherwise.

Density based cell fractionation

7ml discontinuous density gradients of 15, 25 and 25% iodixanol (Optiprep, Stemcell Technologies), 0.25 M sucrose, 75 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA free protease inhibitor were prepared in polyallomer optiSeal ultracentrifuge tubes (11.2 ml capacity; Beckman Coulter) gradients were allowed to diffuse 1h at 20 °C. Partially diffused gradients were stored at 4°C for 1 h while cells were prepared for fractionation. Cells were cultured in 500 mm² plates until 90% confluence, using a single plate per replica per condition. Cells were treated with TG or DMSO for 1 h. After treatment, cells were washed twice with PBS and detached using EDTA-free trypsin (Thermo Scientific) for 10 min. Trypsin was quenched by adding equal volumes of media (supplemented with DMSO or TG). Detached cells were transferred to a 50 ml tube and spun down 10 min at 250 g. Cell pellets were washed twice with ice cold PBS and resuspended in 900 µl of lysis buffer (0.25 M sucrose, 75 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA-free protease inhibitor) and lysed with a ball bearing homogenizer (Isobiotec) on ice. 50 µl of lysate was stored at -80°C as total cell lysate. Cell lysate iodixanol and ion concentration was adjusted by adding 1.5 ml of 50% iodixanol solution (in 75 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA free protease inhibitor) to a 1 ml cell lysate, and underlayered in the previously prepared density gradient with a 2.5 ml syringe and a wide-bore blunt-end needle (Sigma-Aldrich). Finally, a 40% iodixanol (in 75 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA-free protease inhibitor) cushion was underlayered until the tube was filled. Density gradients were centrifuged in a NVT65 fixed-angle near-vertical ultracentrifuge rotor (Beckman Coulter) in a Optima L-80 XP ultracentrifuge (Beckman Coulter) for 16 h at 100,000 g and collected using an auto Densi-Flow peristaltic pump fraction collector with a meniscus-tracking probe (Labconco) to obtain 20 fractions of 500 µl each. The refractive index (RI) of each fraction is measured with a Hand-held refractometer (Reichert) and the iodixanol concentration was calculated as $\text{iodixanol}\% = (\text{RI} / 0.83) - 10.111$, and fraction density calculated by $d = m / V$. The iodixanol concentration per fraction was adjusted to 30% in a volume of 600 µl. All fractions were frozen and dried by sublimation using Vacuum centrifuge with cold trap (Labconco, Refrigerated CentriVac concentrator). Dried pellets were solubilised in 1 ml of trizol (Thermo Scientific) and stored at -80 °C.

Differential sedimentation speed based cell fractionation

Cells were cultured in 500 mm² plates until 90% confluence, using a single plate per replicate. Five replicates were performed for the differential sedimentation speed based cell fractionation experiment. Cells were washed twice with PBS and detached using trypsin/EDTA (0.05%) (Thermo Scientific) for 5 min. Trypsin was quenched with an equal volume of media. Detached cells were transferred to a 50 ml falcon tube and spun down for 5 min at 200 g at 4 °C. The

pellets were washed twice with ice cold PBS and resuspended in 1 ml of lysis buffer (0.25 M sucrose, 75 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA-free protease inhibitor) and homogenised on ice using ball bearing homogenizer (Isobiotec). A total lysate sample of 75 µl was obtained and stored at -80 °C. The remaining sample was fractionated into five consecutive fractions at centrifugation speeds (100, 500, 2000, and 5000 g) using the supernatant of every centrifugation as starting material for the next, with an Eppendorf Centrifuge 5424R. The supernatant of the last centrifugation was retained as the final fraction.

RNA and protein sample precipitation

RNA and protein were obtained from trizol solubilised fractions by adding 200 µl of chloroform and phase partitioning the sample for 15 min at 12000 g at 4 °C. RNA was purified by collecting and transferring the trizol/chloroform upper aqueous phase to a new tube and the RNA precipitated with 750 µl of isopropanol (Sigma Aldrich) for 10 min at 16000 g. RNA pellets were washed twice with 70% ethanol and solubilised in 200 µl of RNase-free water (Thermo Scientific). RNA samples were treated with DNase in RNeasy columns (Qiagen) according to manufacturer instructions using the RNase-free DNase Set kit (Qiagen). RNA concentration was measured with a DS-11 UV Spectrophotometer (Denovix). RNA samples were pooled as indicated in the supplementary table 4. Protein was purified by precipitating the trizol/chloroform lower organic phase (and interface) using 9:1 v:v of methanol:sample. Samples were solubilised in 1% SDS (Thermo Fisher Scientific) 100 mM TEAB (Sigma-Aldrich) using a Bioruptor sonicating bath (Diagenode). Protein concentration was measured with a Pierce BCA Protein concentration assay kit (Thermo Fisher Scientific) on a spectrophotometer plate reader (Molecular Devices, SpectroMax M2).

RNA sequencing

RNA sequencing libraries for the density fractionation were generated using 1 µg of RNA as starting material and depleting ribosomal RNA using Ribocop V3 (Lexogen). Post ribosomal RNA depletion, RNA content was measured with a Bioanalyzer pico kit (Agilent). 1% RNA SIRVs were spiked-in (Lexogen), and RNA seq libraries were generated and amplified using the CORALL RNA sequencing kit (Lexogen) according to the manufacturer instructions. All CORALL generated libraries were sequenced in parallel on 4 Novaseq S4 lanes (Illumina). RNA sequencing libraries for the differential sedimentation speed based cell fractionation experiment were performed using QuantSeq 3' mRNA-Seq kit (Lexogen) according to manufacturer instructions. A total of 400 ng of total RNA and 0.1% RNA SIRVs (Lexogen) were used for library preparation of each fraction. All libraries were balanced, multiplexed and pooled, and run on two single-end Novaseq SP lanes (Illumina).

Proteomic sample preparation

Samples were resuspended in 100 µL of 100 mM TEAB (Sigma-Aldrich), reduced with 10 mM DTT (Sigma-Aldrich) at room temperature for 60 min and alkylated with 40 mM iodoacetamide (Sigma-Aldrich) at room temperature in the dark for 60 min. Samples were digested overnight at 37 °C with 1 µg of trypsin (Promega). Subsequently, 1 µg of modified trypsin (Promega) was added, and the samples were incubated for 3-4 h at 37 °C. Samples were then acidified with TFA (0.5% (v/v) final concentration (Sigma-Aldrich) and centrifuged at 21,000 g for 10 min. The supernatant was immediately desalted.

For peptide clean-up and quantification, 200 μ L of Poros Oligo R3 (Thermo Fisher Scientific) resin slurry (approximately 150–200 μ L resin) was packed into Pierce centrifuge columns (Thermo Fisher Scientific) and equilibrated with 0.1% TFA. Samples were loaded, washed twice with 200 μ L 0.1% TFA and eluted with 300 μ L 70% acetonitrile (ACN) (adapted from ⁶³). From each elution, 10 μ L was taken for Qubit protein assay (Thermo Fisher Scientific) quantification and the remaining sample was retained for MS.

TMT-10plex or TMTpro-16plex (Thermo Fisher Scientific) labelling from desalted peptides was performed according to the manufacturer's protocol. Equal amounts of desalted peptides were labelled immediately after being quantified with Qubit protein assay (Thermo Fisher Scientific). Multiplexed TMT samples were then fractionated using high-pH reverse phase chromatography. In detail, the TMT-labelled peptide samples were resuspended in 100 μ L of 20 mM ammonium formate pH 10 (Buffer A). The total volume of each sample was injected onto an Acquity UPLC BEH C18 column (2.1-mm i.d. $\times$ 150-mm; 1.7- μ m particle size) on an Acquity UPLC System with a diode array detector (Waters) and the peptides were eluted from the column using a linear gradient of 4–60% (v/v) acetonitrile in 20 mM ammonium formate pH 10 over 50 min and at a 0.244 mL/min flow rate (with a total run time of 75 min). The gradient was set up as follows: 0 min–95% Buffer A–5% Buffer B (20 mM ammonium formate pH 10 + 80% (v/v) acetonitrile), 10 min–95% Buffer A–5% Buffer B, 60 min–25% Buffer A–75% Buffer B, 62 min–0% Buffer A–100% Buffer B, 67.5 min–0% Buffer A–100% Buffer B, 67.6 min–95% Buffer A–5% Buffer B. Approximately 40–50 1-min fractions, representing peak peptide elution, were collected starting from initial peptides elution, and were reduced to dryness by vacuum centrifugation shortly thereafter. For downstream MS analysis, the fractions were concatenated into 20 samples by combining pairs of fractions which eluted at different time points during the gradient.

Each sample was analysed in an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific). Mass spectra were acquired in positive ion mode applying data acquisition using synchronous precursor selection MS3 (SPS-MS3) acquisition mode⁶⁴ triggered using Real-time Search against human protein sequences from UniProt/Swiss-Prot. Carbamidomethylation of cysteine and TMT-6plex (total proteome samples) or TMTpro-16plex tagging (Subcellular fractionated samples) of lysine and peptide N terminus were set as static modifications, with oxidation of methionine as a variable modification. Scoring thresholds were set as follows: Xcorr=1.4, dCn=0.1 and Precursor ppm=10.

MS spectra processing and peptide and protein identification

Raw data were viewed in Xcalibur v.2.1 (Thermo Fisher Scientific), and data processing was performed using Proteome Discoverer v2.3 (Thermo Fisher Scientific). The raw files were submitted to a database search using Proteome Discoverer with SequestHF and MS Amanda⁶⁵ algorithms against the *Homo sapiens* database downloaded in June 2020 from UniProt/Swiss-Prot. Common contaminant proteins (several types of human keratins, BSA and porcine trypsin) from the common Repository of Adventitious Proteins (cRAP) v1.0 (48 sequences, adapted from the Global Proteome Machine repository, <https://www.thegpm.org/crap/>) were added to the database. The spectra identification was performed with the following parameters: MS accuracy, 10 ppm.; MS/MS accuracy of 0.5 Da;

up to two missed cleavage sites allowed; carbamidomethylation of cysteine and TMT-6plex (total proteome samples) or TMTpro-16plex tagging (Subcellular fractionated samples) of lysine and peptide N terminus as a fixed modification; and oxidation of methionine and deamidated asparagine and glutamine as variable modifications. Percolator was used for false discovery rate estimation and only rank 1 peptide identifications of high confidence (FDR < 1%) were accepted. TMT reporter values were assessed through Proteome Discoverer v2.3 using the Most Confident Centroid method for peak integration and integration tolerance of 20 ppm. Reporter ion intensities were adjusted to correct for the isotopic impurities of the different TMT reagents (following manufacturer specifications). Sample labels for each TMT tag are presented in supplementary table 4.

Spatial proteomics

Previous marker proteins defined for U-2 OS were annotated in⁶. 12 / 733 markers were deemed outliers based on manual curation of their profile and consideration of localisation assigned in HPA⁶⁶ and GO. To ensure accurate assignment of nuclear proteins, nuclear markers were annotated *de novo*, utilising the COMPARTMENTS database⁶⁷. COMPARTMENTS localisations were mapped to the localisations defined by the standard LOPIT marker sets. Proteins with a score of 5 for the 'Nucleus' and no score over 2 for any other localisation were denoted as exclusively nuclear. The profiles for these nuclear exclusive proteins were split into three groups, which were separated based on the following thresholds on the row-sum normalised abundances: Group 1 = over 0.3 in pooled fraction 4. Group 2 = over 0.2 in pooled fraction 5. Group 3 = over 0.4 in pooled fraction 8. A GO enrichment analysis, using a Hypergeometric test, was then used to determine the enriched functionalities, relative to the background of all quantified proteins. This indicated that the first group of nuclear proteins were highly enriched in nucleolus, chromatin and ribosome biogenesis proteins, whereas the other two were enriched in nucleoplasm proteins but no more specific GO terms. Furthermore, proteins annotated as nuclear membrane or nuclear lamina in GO were more closely associated with group 1. Thus, group 1 was denoted as 'Nucleus' and the other groups were referred to as 'Nucleoplasm-1' and 'Nucleoplasm-2'. Finally, the 'Proteosome' marker set was expanded to define a set of markers for 'Protein complexes' by including proteins annotated in GO as part of the MARS complex ('GO:0017101'), COP9 signalosome ('GO:0008180') or eIF3 complex ('GO:0005852'), with 5/42 of the additional protein complex markers excluded as outliers. SVM classification in basal conditions was performed as described in²⁹, with hyperparameters selected by grid search and 50 iterations.

BUNDLE³⁷ v1.0 was used to identify differentially localised proteins between the control condition and UPR. Differential localisation analysis was performed on each replicate separately, with 10,000 MCMC iterations, 5,000 burn-in iterations, 4 chains and 1/20 thinning. Off diagonal values for the matrix of Dirichlet priors were set at 0.01, with default values used for the penalised complexity priors. MCMC chains were inspected for convergence based on the reported number of localisation outliers, as suggested in³⁷, and all chains were found to converge. Localisations in each condition were determined by setting the following threshold: bundle localisation probability * (1 - bundle outlier probability) > 0.95. Where this threshold was not met, the localisation was deemed 'Undefined'. Proteins were deemed differentially localised if they were never assigned to the same localisation across the two conditions and at least 2/3 replicates had BUNDLE differential localisation probabilities over the following

thresholds to define three levels of confidence were for relocalisation: ≥ 0.99 = 'Highly confident', ≥ 0.95 = 'Confident' and ≥ 0.85 = 'Candidate'.

Comparison of dLOPIT with other methods

Subcellular location data (The Cell Atlas) was downloaded from Human Protein Atlas (HPA; version 22.0) and compared to dLOPIT localisation classifications obtained with BUNDLE and hyperLOPIT and LOPIT-DC classifications⁶. LOPIT and HPA classifications were standardised to 8 localisations that were interrogated with both approaches; cytosol, endoplasmic reticulum, Golgi apparatus, lysosome, mitochondria, nucleus, peroxisome and plasma membrane. This involved the following conversion for the HPA classifications: cytoplasmic bodies = cytosol, nuclear bodies, nuclear membrane, nuclear speckles, nucleoli fibrillar center, nucleoli and nucleoplasm = nucleus, and the following conversion for the LOPIT classifications: nucleoplasm-1, nucleoplasm-2 and chromatin = nucleus, protein complex, proteasome, ribosome 40S, ribosome 60S and ribosome = cytosol. All other classifications not part of the standard localisations were discarded. Proteins not classified by both LOPIT and HPA were excluded.

HPA may classify a single protein with multiple localisations. For each protein, agreement between HPA and LOPIT was deemed to have occurred when the LOPIT classification was within the set of HPA classifications. HPA classifies at increasing levels of confidence, from Approved, to Supported, Enhanced and Uncertain. We separately considered all HPA classifications meeting a minimal level of confidence from Approved to Enhanced.

RNA-seq data processing for CORALL RNA-Seq samples

RNA-Seq fastq processing and quantification was performed using bespoke pipelines built with CGAT-core⁶⁸. Fastq files were demultiplexed using idemux (<https://github.com/Lexogen-Tools/idemux>) and concatenated into one fastq per sample. To assess PCR duplication rate, UMIs were extracted from the read sequences using UMI-tools v1.1.0⁶⁹. Reads were aligned to a concatenation of the hg38 reference genome and the artificial SIRV genome using hisat v2.2.1⁷⁰ using default settings. Secondary reads and reads with MAPQ<10 were discarded. Duplicate reads were identified with UMI-tools dedup using default settings. Transcript isoform quantification was performed from the fastqs without deduplication, using Salmon v1.4.0⁷¹ against a concatenation of the ensembl v102 human transcriptome and the artificial SIRV annotations, using default settings.

Data analysis Data analyses post RNA and peptide quantification steps above were performed using R v4.0.3⁷² and R markdown notebooks⁷³, making extensive use of the tidyverse v1.3.2 R packages⁷⁴, ggplot2 v3.4.2, tidyr v1.2.0, dplyr v1.0.8 and MSnbase v2.20.4⁷⁵, pRoloc v1.34.0⁷⁶ and camprotR v0.0.0.900 (<https://github.com/CambridgeCentreForProteomics/camprotR>).

Differential abundance analysis

Gene-level quantifications (transcripts per million; TPM) were parsed using tximport⁷⁷. Differential gene abundance was tested using DESeq2 v1.34.0⁷⁸ with default settings, with a False Discovery Rate threshold of 5% used to identify significant changes in abundance.

Spatial transcriptomics

Transcript isoform and gene-level quantifications (transcripts per million; TPM) were parsed using tximport v1.22.0⁷⁷ to generate objects to hold the quantification estimates for a single experiment. To take advantage of the spatial proteomics functionalities available through pRoloc⁷⁶, we stored the RNA quantification data in MSnSets. Separate objects were created to hold transcript isoform and gene-level abundance estimates. Transcripts and genes with average TPM < 0.5 or TPM==0 in $\geq 2/3$ samples in a given condition were discarded. Post TPM-filtering, 41547 transcript isoforms and 14203 genes were quantified in at least 2 replicates in both conditions.

Localisation markers were identified using a combination of *a priori* markers and semi-supervised clustering by non-negative matrix factorisation (NNMF). *A priori* markers were defined as follows: Nuclear markers were > 16-fold nuclear enriched in nuclear/cytoplasm fraction RNA-Seq²⁷ or manually defined known nuclear lncRNAs (XIST, MALAT1, MEG3, DLX6-AS1, PINCR, UCHL1-AS1, NEAT1). Cytosol markers were significantly enriched in NES APEX-Seq¹⁰, or manually defined from known cytosolic lncRNAs (LINCMD1, NORAD, H19, NKILA, SNHG5, DANCER, OIP5-AS1, SNHG1). ER markers were defined as having enrichment in ER-Ribo-Seq > $2^{0.5}$ ²⁸ and significant enrichment > 8-fold in KDEL APEX-Seq¹⁰ and a predicted signal peptide or transmembrane domain according to ensembl. Mitochondrial markers were mitochondrially-encoded mRNAs. To determine the optimal number of clusters (k) for NNMF, the imputation-based approach was used⁷⁹, with the selected k value minimising the mean squared error for NNMF-imputation of randomly added missing values. This gave k=5 for the basal condition experiment and k=4 for the UPR experiment. NNMF cluster assignments in basal conditions were then compared to the *a priori* markers to define a set of NNMF-guided gene-level markers, in which each marker set was associated with the NNMF cluster with greatest overlap and all markers in the NNMF cluster were retained. In addition, the NNMF cluster containing the nuclear markers was further used to define a nucleolus marker set by intersecting it with the top 30 most abundant snoRNAs according to the total RNA-Seq. Finally, a novel NNMF cluster was observed in both basal condition and UPR, which contained few *a priori* markers. The genes assigned to the novel NNMF cluster in both basal and UPR conditions were used to define a novel profile, which was observed to have the greatest relative abundance in between the nucleus and cytosol profile peaks, and was hence denoted as cytosol-light. Gene-level markers were used to generate transcript-level markers by taking all the transcript isoforms for each gene-level marker where the ensembl transcript biotype and gene biotype matched. The final marker sets were manually curated to remove 14 / 135 gene-level markers and 37 / 202 transcript-level markers that were deemed to be outliers. The higher proportion of transcript-level markers removed reflects the markers having been built at the gene level and extended naively to transcript-level markers by assuming all transcripts with the same biotype as the gene biotype will have the same localisation.

Estimation of localisation proportions

RNA content per fraction was estimated using the relative abundance of all SIRV features and human RNA and the proportion of the fraction used for RNA-Seq library preparation. Relative RNA-Seq quantifications in each fraction were adjusted with respect to the RNA content per fraction and abundance estimates row-sum normalised across the 8 fractions per replicate. Non-negative least squares regression was used to estimate localisation proportions by separately modelling the profile of each transcript/gene as a non-negative linear combination of the average profile for the markers of each localisation. Since most of the ER markers

relocalised upon UPR and the mitochondrial markers represent 100% membrane localisation in both conditions (no RNA copies in the cytosol), mitochondrial markers were used to estimate the membrane proportion. The proportion estimates were bootstrapped 100 times by re-sampling the markers, with resampling. Proportion estimates for a given transcript/gene were discarded where the model did not account for at least 90% of the variance or the absolute value or the intercept was greater than 0.05. The mean proportions across replicates were estimated for transcripts/genes with estimates from at least 2/3 replicates. In total, proportions for 27368 transcript isoforms and 13274 genes were retained.

Reads were frequently observed to bridge annotations between lncRNAs and neighbouring protein-coding, which resulted in mis-estimation of lncRNA proportions. To avoid this, we excluded all lncRNAs with a protein-coding gene within 15Kb downstream or 30Kb upstream.

Comparison of LoRNA with other methods

The comparison of LoRNA proportions with other methods was performed using gold-standards for nuclear and ER localisation, namely nuclear/cytoplasm fractionation RNA-Seq and ER Ribo-Seq. Comparisons were performed with gene-level data, since transcript-level data was not available for all experiments. Gene-level quantification (TPM) from fractionation polyA + RNA-Seq²⁷ was obtained from www.encodeproject.org. Mean nuclear/cytoplasm ratios were calculated across the cell lines and genes with ratios quantified in at least 2/9 cell lines were retained. ER Ribo-Seq data²⁸ was converted from UNIPROTKB identifiers to ensembl gene IDs using the R biomaRt package. APEX-Seq¹⁰ data was obtained from GEO GSE116008. APEX-RIP¹¹ data was obtained from GEO GSE106493. CeFra-Seq¹⁴ data was taken and nuclear/cytosol and membrane/cytosol ratios calculated. Relative enrichments from LoRNA were calculated from proportions as e.g $\text{LocalisationX}/(1-\text{LocalisationX})$, with nuclear proportions being the sum of nucleus and nucleolus proportions. All enrichments were log2-transformed and Pearson correlation coefficients against gold-standard ratios calculated.

The classification of ribosome-associated lncRNAs was obtained from³⁴. lncRNAs with multiple transcript IDs were excluded. Ribosome profiling data was obtained from⁸⁰ and CDS length obtained from ensembl. Ribosome density was then calculated as $\text{RNaseI_CDS} / \text{CDS length}$. P-body RNAs were obtained from⁴⁶. RNAs with over 8-fold higher abundance in the sorted fraction and Benjamini-Hochberg⁸¹ adjusted p-value less than 0.01 were deemed P-body RNAs.

MERFISH data for ER and Nuclear enrichment was obtained from¹². This was presented at the level of RNA transcripts and was therefore compared to transcript-level LoRNA proportions.

Defining signal peptide and transmembrane domain features

Signal peptide and transmembrane (TM) domain annotations were obtained from ensembl v102 and the distance between the start of the first signal peptide or TM domain and the stop codon was determined. For gene-level analyses, the minimum distance across all transcript isoforms was used. A conservative annotation of presence/absence of either signal peptide and/or TM domain was obtained by taking the union of the ensembl annotations with UNIPROT annotations.

eCLIP binding data

eCLIP data in narrowPeak (BED6+4) format was obtained from ENCODE using ENCODEExplorer v2.16.0. Genomic coordinates were converted to transcriptomic coordinates using the mapToTranscripts function in GenomicFeatures v1.46.5 R package⁸².

Modelling lncRNA cytosol localisation

To model the cytosolic localisation of lncRNAs, transcript length, spliced status, eCLIP binding data, kmers frequencies, AU content, RNA modifications, PolyA status, presence in FANTOM5 robust catalogue, and abundance were considered. Transcript length and splicing status was obtained from ensembl v102. eCLIP data was obtained as described above. 64 RBPs with > 10 lncRNA targets were retained and binding data converted to binary 0=unbound, 1=bound. Kmers (k=1-7) were counted and expressed as frequencies. RNA modifications were obtained from m6A-atlas⁸³, with modifications with > 10 lncRNA targets retained, namely, m6A, m5c, m1A and Psi. The FANTOM5 robust catalogue and PolyA status were obtained from⁸⁴. PolyA status was converted to a binary 0=non-polyadenylated, 1=polyadenylated, with 'undetermined' encoded as 0 and 'bimorphic' encoded as 1. Abundance was calculated as the mean TPM from the total RNA-Seq of basal condition samples. Cytosol proportions were converted to a binary feature, where 1 = cytosolic ($\geq \frac{2}{3}$ cytosol) and 0 = not-cytosolic ($< \frac{1}{3}$ cytosol). The data was then split 80:20 into training and test data. Elastic net logistic regression was used to model the cytosol variable, with 10-fold cross-validation, using the glmnet v4.1-3 R package⁸⁵. By default glmnet scales the dependent variables but returns the coefficients on the original scale. Alpha values were varied between 0-1, in steps of 0.1. Cross-validation folds were pre-computed to ensure they were the same for each alpha value. The alpha value with the minimum mean cross validation error was identified as 1, e.g lasso regression. Lambda was selected to give the most parsimonious model, using the 'one-standard-error' rule⁸⁶. The final model contained 45 / 21918 non-zero coefficients. The predictive accuracy of this model was assessed using the hold-out test data and compared to the accuracy of a simpler logistic regression model of just transcript length, polyA status and AU content.

Modelling UPR-resistant membrane localisation

Membrane proportion in UPR was modelled as being dependent upon membrane proportion in control conditions using a Generalised Additive Model (GAM) with a cubic regression spline with shrinkage, using the mgcv R package. The residual from the GAM was taken to represent the degree of UPR resistant localisation relative to the overall reduction in membrane proportions. Lasso regression (glmnet R package⁸⁵) was used to model UPR resistance, with RBP binding from eCLIP data (see above) being the dependent variables. 10-fold cross-validation was used to select the lambda value that minimised the mean cross validation error, with the 'one-standard-error' rule used to select the most parsimonious model⁸⁶. Transcripts with signal peptides/TM domains were separately modelled transcripts from those without.

Modelling relocalisation to granules

Granule relocalisation (UPR granule proportion - Control granule proportion) was modelled according to the following features which were annotated to each mRNA: localisation in basal conditions, RNA length, RBP binding, sequence kmers, 5' AUGs, codon features, upstream open reading frames (uORFs) and internal ribosome entry site (IRES). One-hot encoding was used for the features describing the localisation in basal conditions, where the highest proportion localisation was taken to be the single localisation for the RNA. Transcript, 5' UTR, 3'UTR and coding sequence lengths were obtained from ensembl v102. eCLIP data was

obtained as described above. 177 RBPs with > 100 mRNA transcript targets were retained and binding data converted to binary 0=unbound, 1=bound. K-mers (k=1-6) were counted and expressed as frequencies. 5' AUGs were identified with separate features to encode in-frame and out-of-frame AUGs. Codon, dinucleotide and wobble base frequencies were computed from the coding sequence. Codon optimality (MILC⁸⁷) was computed, with separate features for the comparison against all transcripts and just the top 1% most abundant. uORFs were identified from⁸⁸, retaining only ORFs with a score > 5 and an AUG start codon. Annotated IRES were obtained from IRESbase⁸⁹. uORF and IRES were converted to binary presence/absence features for each transcript. In addition, pairwise interactions between the localisation in basal conditions and the transcript, 3'UTR, 5'UTR and coding length features and RBP binding features were added as separate features.

Granule relocalisation data was split 80:20 into training and test data. Elastic net regression was used to model the granule relocalisation, with 10-fold cross-validation, using the glmnet R package⁸⁵. By default glmnet scales the dependent variables but returns the coefficients on the original scale. Alpha values were varied between 0-1, in steps of 0.1 Cross-validation folds were pre-computed to ensure they were the same for each alpha value. The alpha value with the minimum mean cross validation error was identified as 1, e.g lasso regression. Lambda was selected to give the most parsimonious model, using the 'one-standard-error' rule⁸⁶. The final model contained 74 / 16895 non-zero coefficients.

Data processing and analysis for differential centrifugation-based LoRNA

Data processing and analysis for samples quantified using 3' Quant Seq was identical to CORALL RNA-Seq samples, except for the following: Gene read counts were normalised to counts per million (CPM) using the total number of assigned reads per sample. Genes with average CPM < 1 or CPM==0 in > 20% of the samples were discarded. Markers identified from equilibrium density centrifugation-based LoRNA were annotated and proportions calculated as indicated above, except that proportions were estimated separately for mitochondria and ER and summed to give membrane proportions, and 'cytosol-light' proportions were not estimated. The mean proportions across replicates were estimated for transcripts/genes with estimates from at least 3/5 replicates.

Assessing the technical bias for RNA length and sedimentation

Gene-level CeFra-Seq quantification data (TPM) was downloaded using the ENCODExplorer v2.16.0 R package. We computed the gene length as the mean length of all transcript isoforms included in the GENCODE basic gene set and with Transcript Support Level 1 in ensembl. Mean abundances per fraction were computed for the cytosol and membrane fractions, from which ratios were then computed compared to the total RNA-Seq samples. High-confidence cytosolic and membrane localised RNAs were obtained from LoRNA, setting a threshold of proportion > 80%. Pearson product-moment correlation coefficients were computed for gene length vs cytosol/total for cytosolic RNAs and membrane/total for membrane RNAs. The same procedure was used for the differential sedimentation-based RNA localisation approach estimates, with supernatant/total correlated with gene length for cytosolic RNAs.

Gene-Ontology enrichment analyses

All Gene-Ontology (GO) enrichment analyses were performed using the goseq v1.46.0 R package⁹⁰. For the analysis of GO terms enriched in eIF3d-bound membrane RNAs, the membrane proportion in basal conditions was included as the bias factor and enrichment effect

sizes accounting for biasing factors were calculated using the probability weight functions obtained with goseq using the estimate_go_overrep function in camprotR. For the analysis of GO terms enriched in the membrane RNAs not encoding a signal peptide or TM domain, no bias factor was included and a Hypergeometric test was instead used. For both analyses, p-values were adjusted for multiple testing using the Benjamini-Hochberg procedure⁸¹ and GO terms with adjusted p-value > 0.05 (5% FDR) or accounting for fewer than 10% of the foreground genes were excluded. Redundant GO terms were removed using the remove_redundant_go function in camprotR R package.

Single molecule FISH probe design and synthesis

Subcellular RNA localisation was assessed by using an adaptation of the single-molecule inexpensive FISH protocol⁹¹. Z (CTTATAGGGCATGGATGCTAGAAGCTGG) and Y (AATGCATGTCGACGAGGTCCGAGTGTA) FLAP DNA handles, labelled at the 5' and 3' ends with Atto488, were purchased from Sigma (HPLC-purified) and resuspended to a concentration of 100 μ M in nuclease-free water. For each target RNA, 30-48 DNA probes of 20 nucleotides (nt) were designed with a minimum spacing length of 2 nt and a guanine-cytosine content of 40-65%. Each gene-specific sequence was flanked by a 28 nt sequence complementary to either a Z- or Y-FLAP sequence. The resulting 48 nt probes were purchased from Sigma (standard desalt purification, 100 μ M in nuclease-free water). Fluorescently labelled gene-specific probes were then generated as follows: 200 pmol of an equimolar mixture of all gene-specific oligos for each gene were mixed with 250 pmol of the appropriate FLAP oligo in 1x NEBuffer 3 (New England Biolabs, B7003), then incubated in a Thermocycler (BioRad) for 3 min at 85 °C, 3 min at 65 °C, and 5 min at 25 °C (lid 99°C).

smiFISH/Immunofluorescence and Immunofluorescence

For smiFISH/IF experiments, 8×10^4 U 2-OS cells were seeded in each well of a 12-well plate, on top of no 1.5 glass coverslips previously washed in 1 M HCl. The following day, cells were treated with either 250 nM Tg or the corresponding volume of DMSO, rinsed three times with PBS (with $MgCl_2$ and $CaCl_2$, Sigma D8662), then fixed for 10 min in 3% Methanol-free PFA (Alfa Aesar, 43368) in PBS at RT. The fixative was then quenched in 100 mM Glycine (Sigma) in PBS for 10 min at RT, samples were then washed twice in PBS for 10 min and permeabilised in 70% EtOH at 4 °C for at least 1 h. Samples were then prepared for hybridisation by washing in 10% Formamide (Sigma), 1 U/ μ l RNasin Plus (Promega) in 2x Saline-sodium citrate (SSC) buffer for 10 min. From this step onwards, samples were protected from direct light. Hybridisation was performed by incubating coverslips with 100 – 250 nM probes diluted in hybridisation buffer (2X SSC buffer, 10% Dextran sulfate (Sigma), 10% Formamide, 2 mM ribonucleoside vanadyl complexes (Sigma), 200 μ g/ml bovine serum albumin (Roche), 1 mg/ml *E. coli* tRNA (Roche), 1 U/ μ l RNasin Plus (Promega)), for 3 h at 37 °C in a humid chamber. Coverslips were then transferred to a clean 12-well plate and washed twice in 10% Formamide, 1 U/ μ l RNasin Plus, 2x SSC buffer, for 10 min at 37 °C. Further washes were then performed (three times in 2X SSC with no incubation, twice in PBS for 10 min), before incubation with blocking buffer (3% nuclease-free bovine serum albumin (Sigma) in PBS) for 30 min at RT. Samples were then incubated with primary antibody diluted 1:500 in blocking buffer, for 2 h at RT. They were then washed three times in PBS for ten min, incubated with a secondary antibody diluted 1:2000 in blocking buffer, for 1 h at RT. After three additional washes in PBS, the nucleus was stained by incubation with 4',6-diamidino-2-phenylindole (DAPI, Sigma, 200 ng/ml in PBS) for 1 min at RT. Coverslips were then washed twice in PBS

for 5 min before being mounted onto glass microscope slides with a drop of ProLong™ Glass Antifade Mountant (Invitrogen).

For consistency, IF experiments were performed following the smiFISH/IF procedure with the omission of the FISH hybridisation step and the washes in 10% formamide, 2X SSC.

Digitonin cytosol extraction for smFISH/IF experiment was performed following the protocol described in⁹². Briefly, cells grown on glass coverslips were washed twice in warm CHO buffer (115 mM KAc, 25 mM HEPES pH 7.4, 2.5 mM MgCl₂, 2 mM EGTA, 150 mM sucrose). The coverslips were then placed (cell-side down) on a droplet of 0.025% Digitonin in CHO buffer, on top of a metal plate warmed to 40 °C, for 20 s. Cells were then immediately blocked in 4% PFA/PBS for 15 minutes at room temperature, then extensively washed with PBS and immersed in ice-cold MeOH for 30 min. After additional washes in PBS and two washes in 10% Formamide/SSC buffer, hybridisation with FISH probes and IF were performed as described above.

Confocal microscopy

Images were acquired on a Zeiss Axio Observer.Z1 LSM 980 microscope equipped with Airyscan 2, using ZEN Blue software (version 3.3), in Airyscan super-resolution imaging mode. A C-Plan-Apochromat 63X /1.4 NA oil objective was used with Zeiss Immersol 518F (23 °C) immersion oil. 20 Z-slices per image were acquired at an interval of 0.13 µm.

Image processing and quantification

Airyscan processing with standard parameters was applied to each image in Zen Blue software. Further processing and analysis was performed using Fiji software⁹³ (version 2.3.051).

siRNA transfection

Transfections were performed in detached cells post trypsinization to ensure maximum exposure. 10nM of each siRNA were transfected per condition using Lipofectamine RNAiMAX (Thermo Fisher) according to the manufacturer's instructions. Knockdown experiments were performed using ON-TARGETplus SMARTpool siRNAs (Horizon): control siRNA #1 #D-001810-01-05; eIF3d siRNA #L-017556-00-0005; eIF3j siRNA #L-019532-00-0005. All subsequent experiments were performed 48h post transfection.

Cell migration assay

Cell migration assays were performed using the xCELLigence RTCA DP instrument (ACEA Biosciences) according to the manufacturer's instructions. Cells were treated with thapsigargin for 1 hour, collected by trypsinisation and washed in PBS to remove all traces of serum. 30,000 cells were seeded in the upper chamber of a 16 well migration plate (CIM-16 plate) in 100 µl of media containing 0.1% serum. Cells migrate to a lower chamber containing 160 µl of media supplemented with 10% serum. As cells migrate across the microelectrodes into the lower chamber they generate impedance measurements (cell index), which enables label free quantification of cell migration for 24 hours. Cell indexes for Tg-treated samples were expressed as fold changes relative to the DMSO control sample. Significance testing between eIF3d and control siRNA knockdowns was performed with a paired Student's t-test.

Digitonin cytosolic RNA extraction

Digitonin extraction protocol was adapted from⁹⁴. Cytosolic and membrane-associated RNA fractionation experiments were performed in 80% confluent 12 multiwell plates. Cells were washed twice in ice cold PBS and incubated for 5 min in 200 μ l of digitonin extraction buffer (0.03% of Digitonin (Thermo Fisher) 5 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA free protease inhibitor). The supernatant was transferred to a new tube and the cells were incubated for 5 min with 200 μ l of digitonin wash buffer (0.004% of Digitonin (BRAND) 5 mM KCl, 5 mM MgCl₂, 50 nM CaCl₂, 10 mM HEPES pH 7.4 and EDTA free protease inhibitor). The second supernatant was transferred to the same tube as the first one. This combined fraction is enriched in cytosol transcripts. The plates were then scrapped in 400 μ l of digitonin extraction buffer. This second fraction is enriched in membrane-associated RNAs. Cytosolic and membrane associated RNA fractions were frozen and dried by sublimation using a vacuum centrifuge with cold trap. RNA was extracted and DNase treated using the RNeasy kit according to the manufacturer's instructions. A total of 200 μ g of RNA were reverse transcribed using Moloney murine leukemia virus reverse transcriptase and random decamers (Ambion). qPCR amplification reaction was performed using the LightCycler 480 SYBR Green I Master Mix (Roche Diagnostics) and the following set of primers: MACF1 Fw 5'-TAGAGATGACTGCTGTGGC-3' and MACF1 Rv 5'-TGTCTTGTAACCTCATCTTCGA-3', DST Fw 5'-ATTGGTACAGAGGGTTGCA-3' and DST Rv 5'-CGTCCTTTGCTGTACACAG-3', TMX1 Fw 5'-ACGGACGAGAACTGGAGAGA-3' and TMX1 Rv 5'-ATTTTGACAAGCAGGGCACC-3', SSR2 Fw 5'-GTTTGGGATGCCAACGATGAG-3' and SSR2 Rv 5'-CTCCACGGCGTATCTGTTCA-3', PSK1N Fw 5'- -3' and PSK1N Rv 5'- -3'. Quantitative expression data were normalised to the mitochondrial transcript MT-ND6 using the MT-ND6 Fw 5'-GGGTTGAGGTCTTGGTGAGT-3' and MT-ND6 Rv 5'-ACCAATCCTACCTCCATCGC-3' primers. TMX1, SSR2, PSK1N and MT-ND6 primers were obtained from¹¹.

Cell death assay

All flow cytometry data was acquired using a BD LSRFortessa (BD bioscience) and analysed with BD FACSDiva software (version 9.0.1). 10,000 counts were acquired for each experimental condition. Cells were collected in Annexin binding buffer (BD biosciences #556454) and cell death was determined after incubation with Annexin-V-FITC (Thermo Fisher, BMS500FI-100) and Draq7 (Abcam, ab 109202).

LC-MS/MS acquisition

TMT-labelled samples were analysed in an Orbitrap Eclipse coupled to a nanoLC Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific). Peptides were trapped on a 100 μ m $\times$ 2 cm, C18, 5 μ m, 100 trapping column (Acclaim PepMap 100) in μ l-pickup injection mode at 15 μ l/min flow rate for 3 min. Samples were then loaded on a Rapid Separation Liquid Chromatography, 75 μ m $\times$ 50 cm nanoViper C18 3 μ m 100 column (Acclaim, PepMap) at 50 °C retrofitted to an EASY-Spray source with a flow rate of 300 nl/min. Analytical chromatography was performed over 120 min (buffer A, HPLC H₂O, 0.1% formic acid; buffer B, 100% ACN, 0.1% formic acid; 0–3 min: at 2% buffer B, 3–105 min: linear gradient 2% to 40% buffer B, 105–105.3 min: 40% to 90% buffer B, 105.3–110 min: at 90% buffer B, 110–110.3 min: 90% to 3% buffer B, 100.3–120 min: at 3% buffer B). Each MS1 scan was performed in the Orbitrap analyser (mass range = m/z 400–1500, resolution = 120,000). Precursors with

charge between 2 and 6 and intensity above 5000 were selected for collision induced dissociation (CID)-MS2 fragmentation, with normalised AGC target of 200% and maximum accumulation time of 50 ms. Mass filtering was performed by the quadrupole with 0.7 m/z transmission window, followed by CID fragmentation in the linear ion trap with 30% normalised collision energy. Selected fragmented ions were dynamically excluded for 60 s. Real-time Search (RTS) was used to trigger SPS-MS3 acquisition. RTS used Human UniProt/Swiss-Prot database, carbamidomethylation of cysteine and TMT-6plex (total proteome samples) or TMTpro-16plex tagging (Subcellular fractionated samples) of lysine and peptide N terminus as static modification, oxidation of methionine as variable modification, as scoring thresholds Xcorr=1.4, dCn=0.1, Precursor PPM=10, one missed cleavage and maximum search time of 35 ms. SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–1,500 m/z range and were set to preclude selection of the precursor ion and TMT or TMTpro ion series. Normalised AGC targets and maximum accumulation times were set to 200% and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalised collision energy and were analysed in the Orbitrap with a nominal resolution of 50,000. The number of SPS-MS3 spectra acquired between full scans was restricted to a duty cycle of 3 s.

Proteomics data processing

PSM-level quantifications were filtered to conservatively remove peptides from common contaminants. Alongside the cRAP proteins, further potential contaminants were identified by considering all proteins that shared an observed peptide with a cRAP protein as further contaminants. PSMs without a unique master protein assigned or more than 20% missing values were excluded. Remaining missing values were imputed using knn imputed (k=10), with sum normalisation prior to imputation and de-normalisation post imputation, to ensure nearest neighbours shared similar abundance profiles over the fractions, rather than similar average abundance. We then identified and removed outlier PSMs which had a median euclidean distance over 0.2 from all other PSMs for the same master protein. In doing so, PSMs with higher co-isolation and lower average signal/noise were selectively removed where these low quality PSMs disagreed with other PSMs. Remaining PSMs were then median-centre-normalised. Protein-level abundances were estimated by summing PSM-level abundances, for proteins with at least two PSMs. Protein abundances were row-sum normalised such that the total abundance across all fractions from a given replicate equaled one.

Polysome profiling

10%-50% (w/v) sucrose gradients were prepared in gradient buffer (100 mM NaCl, 5 mM MgCl₂, 15 mM tris-HCL pH 7.5, 1 mM DTT, 0.1 mg/ml cycloheximide). Cells were washed in PBS-cycloheximide (100 g/ml) and scraped into lysis buffer (100 mM NaCl, 5 mM MgCl₂, 15 mM tris-HCL pH 7.5, 1 mM DTT, 0.2 M sucrose, 0.1 mg/ml cycloheximide, 0.5% IGEPAL, 5 µl RNasin per 1 ml). Lysates were incubated on ice for three min and cells pelleted by centrifugation at 1300 x g for 5 min. The supernatant was layered on top of a gradient and centrifuged at 247,767 g (acceleration 9, deceleration 6) for 2 h at 4 °C using a Beckman Coulter ultracentrifuge. Polysome profiles were obtained by measuring absorbance at 254 nm using a UA-6 UV-VIS detector (Presearch Ltd).

Reporter gene expression analysis

DCP2-GFP tagged mammalian cell expression plasmid was directly acquired from Adgene (pT7-EGFP-C1-HsDCP2, ref 25031). G3BP1-GFP plasmid was kindly provided by Dr. Dee Scadden (University of Cambridge). Plasmids were purified using EasyPep™ Mini MS Sample Prep Kit (Quiagen). One plate of 500 mm² of 90% confluent U-2 OS cells were transfected per plasmid, per condition using 100 µg of plasmid with Lipofectamine 3000 Transfection Reagent (Thermo Scientific) according to manufacturer instructions. Two days post-transfection, cells were treated with DMSO or TG as indicated above and lysed and fractionated as specified in the “density based cell fractionation” methods section. Gradient fractions were imaged on a Zeiss Axio Observer.Z1 LSM 980 microscope and fluorescence intensity quantified with ImageJ 1.8.0_172 for Mac OS X.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁶² partner repository with the dataset identifier PXD030456. The RNA-Seq data have been deposited in ENA with study accession PRJEB49479. The processed data including the data underlying the Shiny app are available as serialised R objects in RDS format from https://github.com/CambridgeCentreForProteomics/LoRNA_UPR (v1.0 and archived with zenodo, DOI: 10.5281/zenodo.8375646). RNA annotations were obtained from ensembl v102 using biomaRt R package. FANTOM5 data was downloaded from <https://www.nature.com/articles/s41587-021-00936-1>. eCLIP and CeFra-Seq data were downloaded from <https://www.encodeproject.org>. The hyperLOPIT and LOPIT-DC data were obtained using R package pRolocdata (object names: hyperLOPITU2OS2018, lopitdcU2OS2018). The following additional publicly available datasets were used. In all cases the data used is available from https://github.com/CambridgeCentreForProteomics/LoRNA_UPR v1.0 in the directory 1_external. Human Protein Atlas data was downloaded from <https://www.proteinatlas.org/about/download> on 03 May 2023. MERFISH data was downloaded from <https://www.pnas.org/doi/10.1073/pnas.1912459116#data-availability>. Ribosome-associated lncRNAs classification was obtained from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5975437>. ER Riboseq data was obtained from <https://www.science.org/doi/10.1126/science.1257521>. P-body enriched RNA data was obtained from DOI: <https://doi.org/10.1016/j.molcel.2017.09.003>. APEX-Seq data was obtained from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116008>. APEX-RIP data was obtained from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106493>. IRES data was downloaded from IRESbase (<http://reprod.njmu.edu.cn/cgi-bin/iressbase/download.php#human>). uORF data was downloaded from <https://doi.org/10.1093/nar/gky188>. GO term annotations were downloaded from ensembl v102 using biomaRt R package. Data pertaining to the effect of eIF3D Knockdown on translation was downloaded from <https://doi.org/10.1016/j.molcel.2020.06.003> (Supplemental Information; Data S1). 5' cap eIF3d binding data was obtained from DOI: 10.1126/science.abb0993. Signal peptide and transmembrane domain annotations were downloaded from Uniprot (<https://www.uniprot.org/>) using the search string '(annotation:(type:signal) OR annotation:(type:transmem)) AND organism:"Homo sapiens (Human) [9606]"' on 06 September 2021. Data pertaining to stress granule enrichment upon arsenite treatment were downloaded from <https://doi.org/10.1016/j.molcel.2017.10.015>. Data pertaining to TIS-granules were downloaded from DOI: 10.1101/2022.11.04.515216. RNA modification data was downloaded from m6A Atlas (<http://180.208.58.19/m6A-Atlas/download.html>) on 02 November 2022.

Code availability

Data analysis code is available from:

https://github.com/CambridgeCentreForProteomics/LoRNA_UPR (v1.0) was archived with zenodo, DOI: 10.5281/zenodo.8375646.

Methods-only references

62. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
63. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom. JMS* **34**, 105–116 (1999).
64. McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
65. Dorfer, V. *et al.* MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **13**, 3679–3684 (2014).
66. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
67. Binder, J. X. *et al.* COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database J. Biol. Databases Curation* **2014**, bau012 (2014).
68. Cribbs, A. P. *et al.* CGAT-core: a python framework for building scalable, reproducible computational biology workflows. Preprint at <https://doi.org/10.12688/f1000research.18674.2> (2019).
69. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
70. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
71. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods*

- 14**, 417–419 (2017).
72. R Core team. *A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2020).
73. Allaire, J. J. *et al.* *rmarkdown: Dynamic Documents for R*. (2021).
74. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
75. Gatto, L., Gibb, S. & Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J. Proteome Res.* **20**, 1063–1069 (2021).
76. Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T. & Lilley, K. S. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinforma. Oxf. Engl.* **30**, 1322–1324 (2014).
77. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
79. Lin, X. & Boutros, P. C. Optimization and expansion of non-negative matrix factorization. *BMC Bioinformatics* **21**, 7 (2020).
80. Miettinen, T. P. & Björklund, M. Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res.* **43**, 1019–1034 (2015).
81. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
82. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
83. Tang, Y. *et al.* m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.* **49**, D134–D143 (2021).
84. Lorenzi, L. *et al.* The RNA Atlas expands the catalog of human non-coding RNAs. *Nat. Biotechnol.* **39**, 1453–1465 (2021).

85. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
86. Hastie, T., Tibshirani, R. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*. (Springer, 2009).
87. Supek, F. & Vlahovicek, K. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC Bioinformatics* **6**, 182 (2005).
88. McGillivray, P. *et al.* A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Res.* **46**, 3326–3338 (2018).
89. Zhao, J. *et al.* IRESbase: A Comprehensive Database of Experimentally Validated Internal Ribosome Entry Sites. *Genomics Proteomics Bioinformatics* **18**, 129–139 (2020).
90. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
91. Tsanov, N. *et al.* smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).
92. Wu, J. J. & Palazzo, A. F. Visualization of Endoplasmic Reticulum-Associated mRNA in Mammalian Cells. *Methods Mol. Biol. Clifton NJ* **2166**, 35–49 (2020).
93. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
94. Jagannathan, S., Nwosu, C. & Nicchitta, C. V. Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation. *Methods Mol. Biol. Clifton NJ* **714**, 301–321 (2011).

figure 1

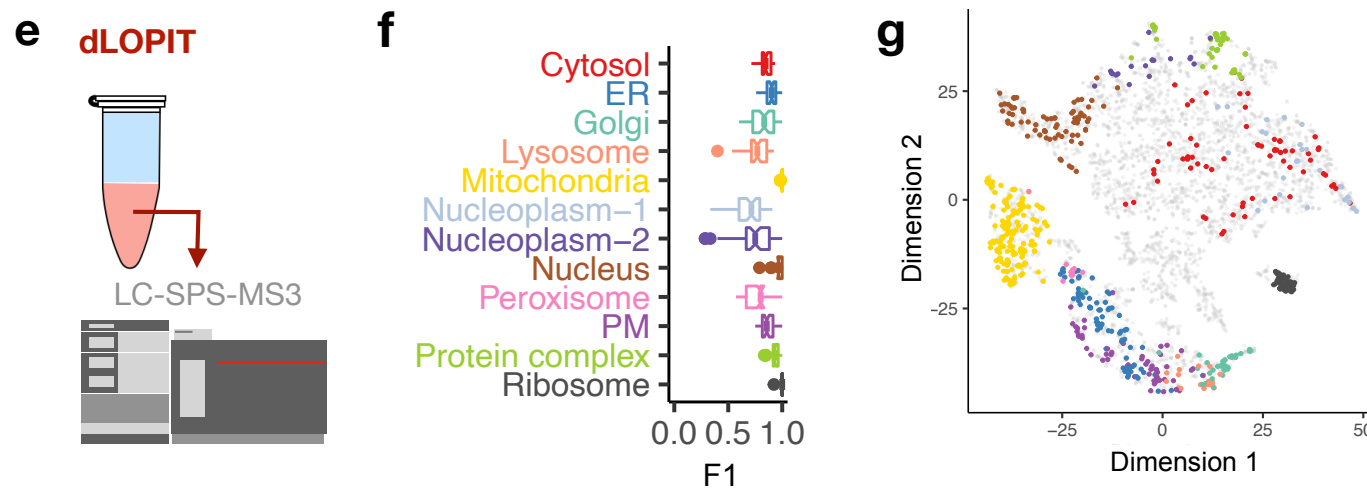
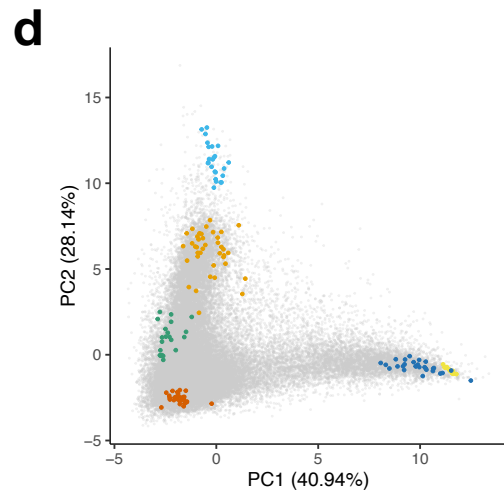
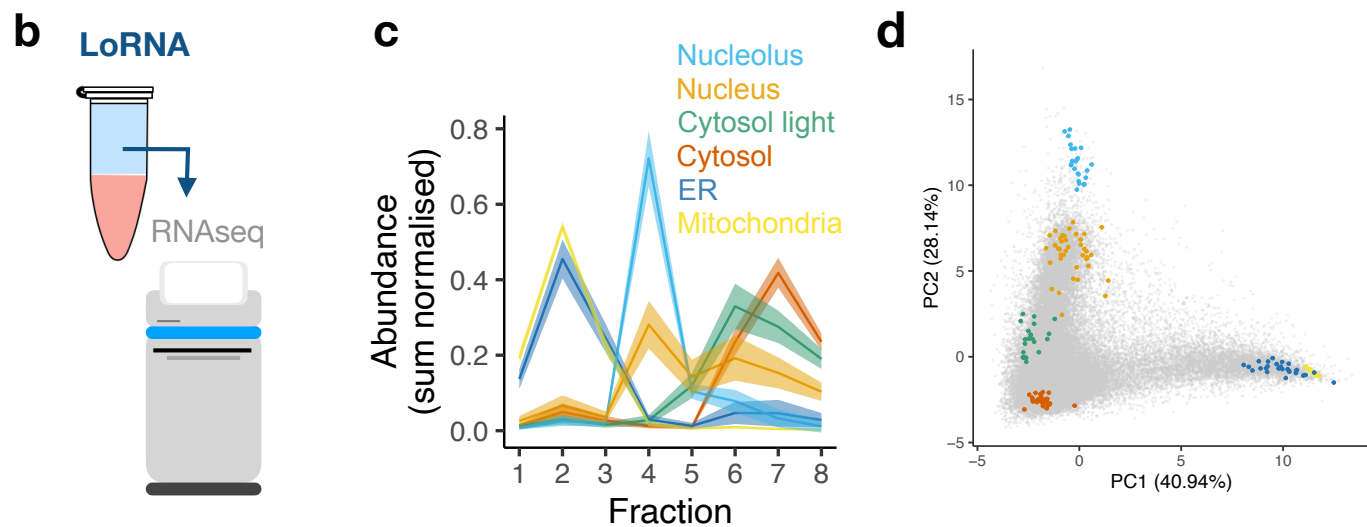
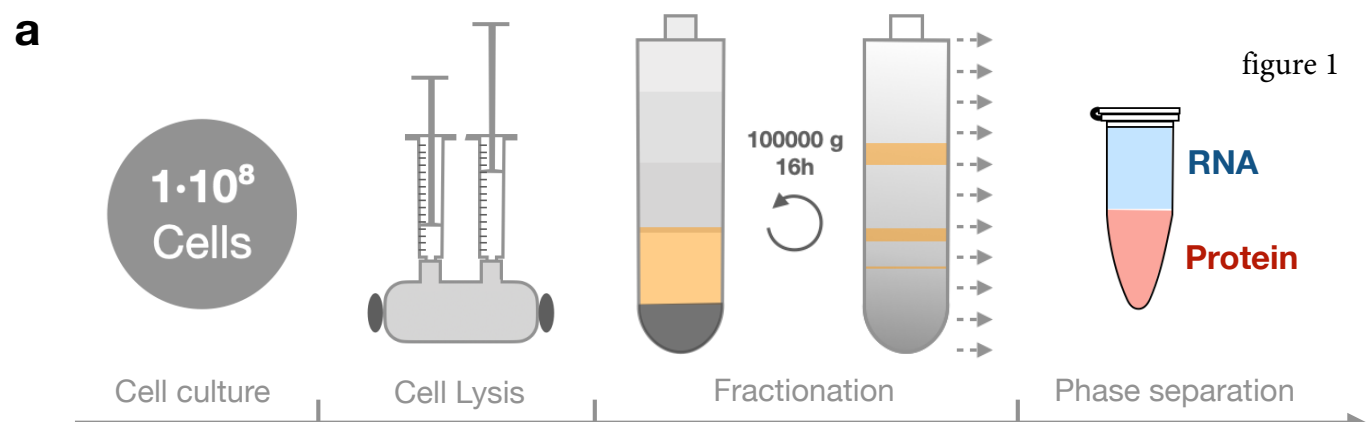


figure 2

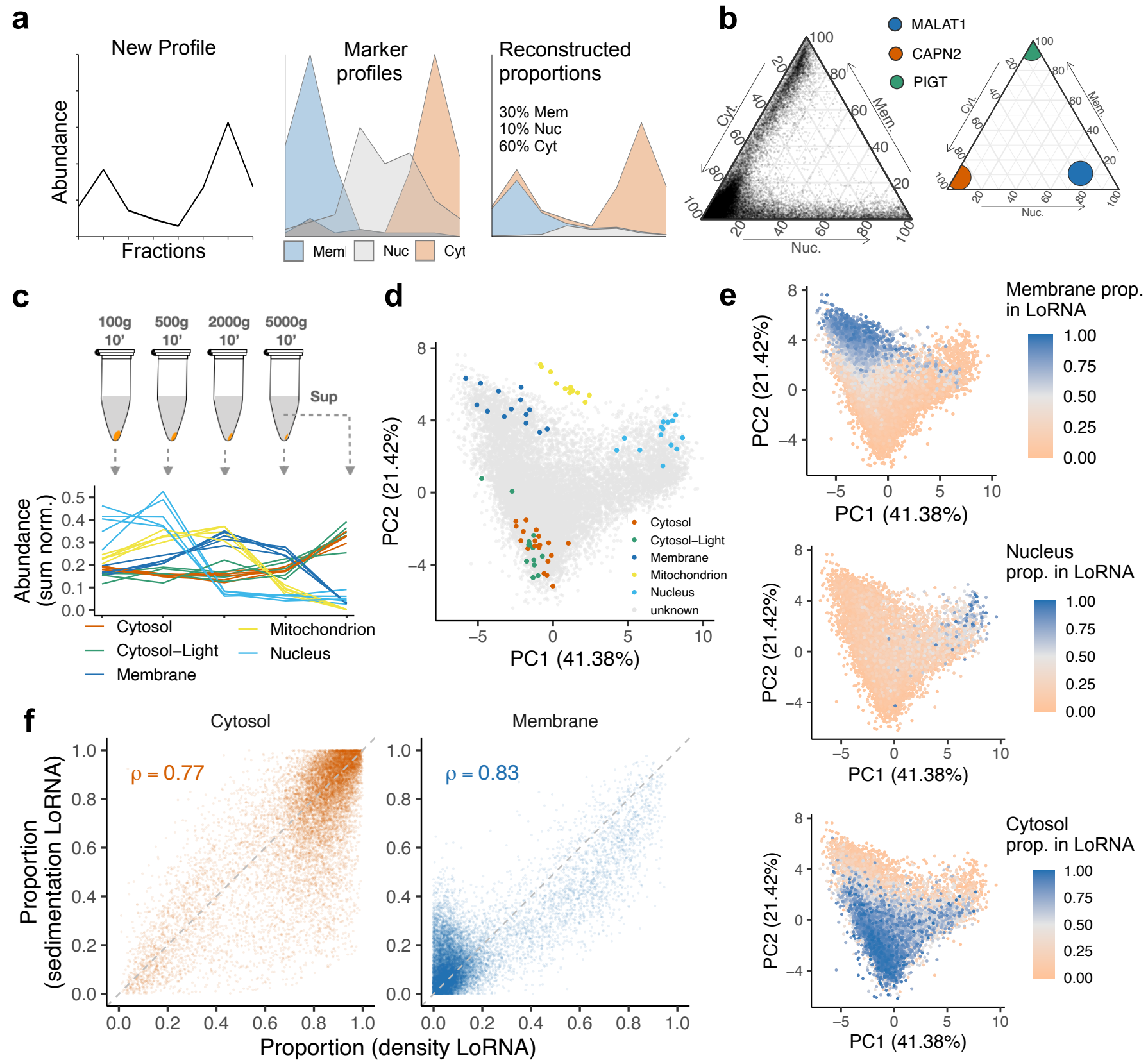


figure 3

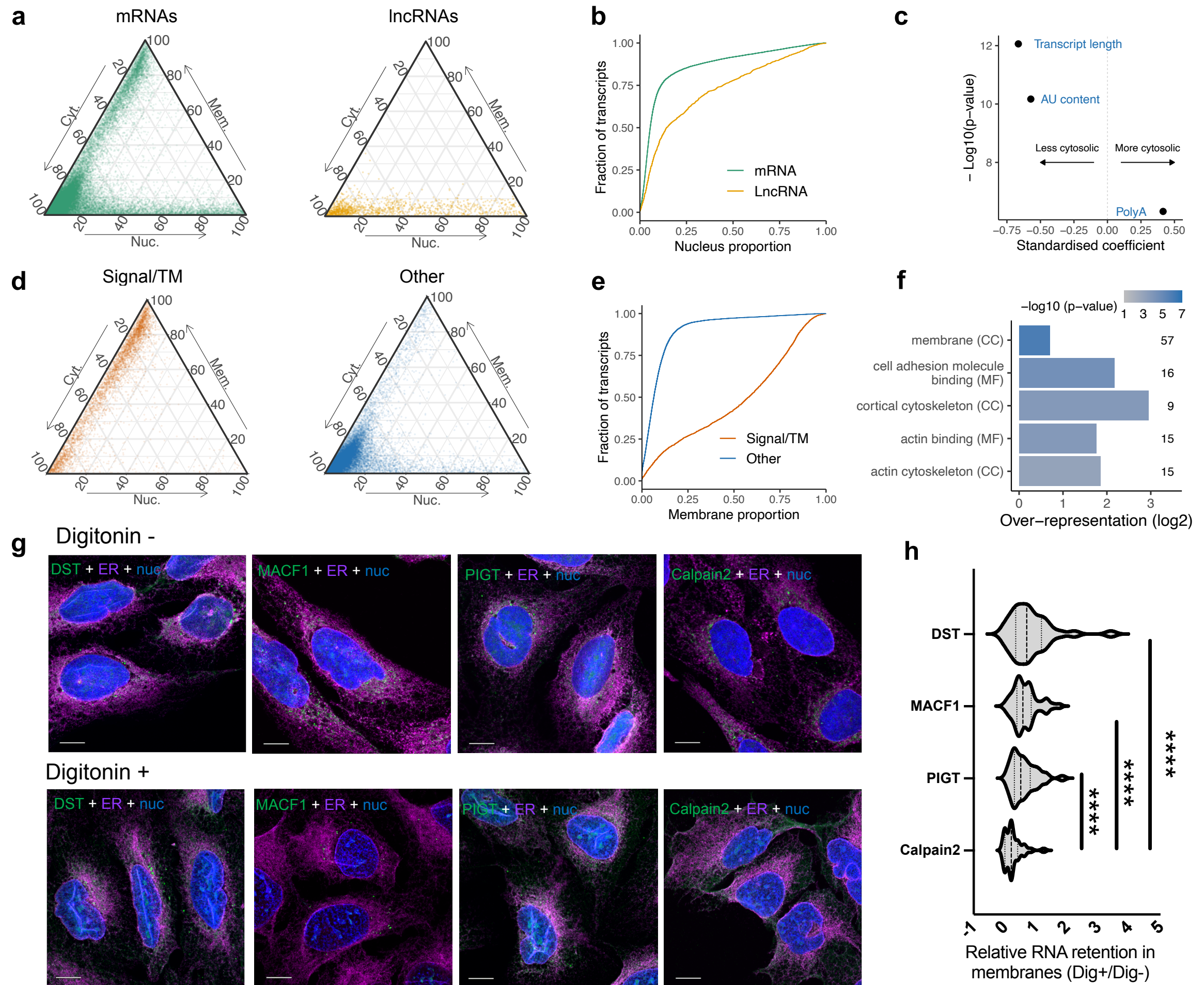


figure 4

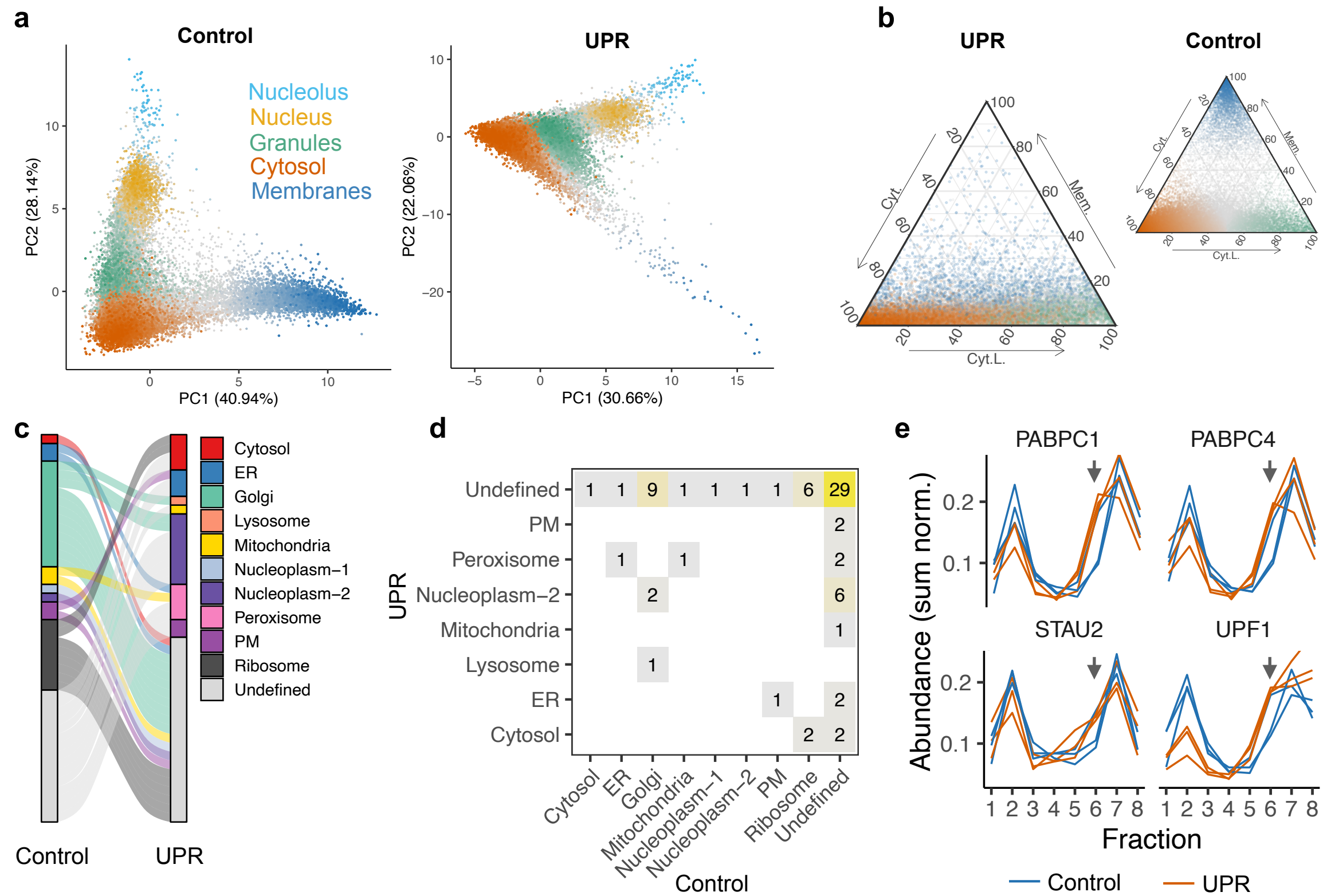


figure 5

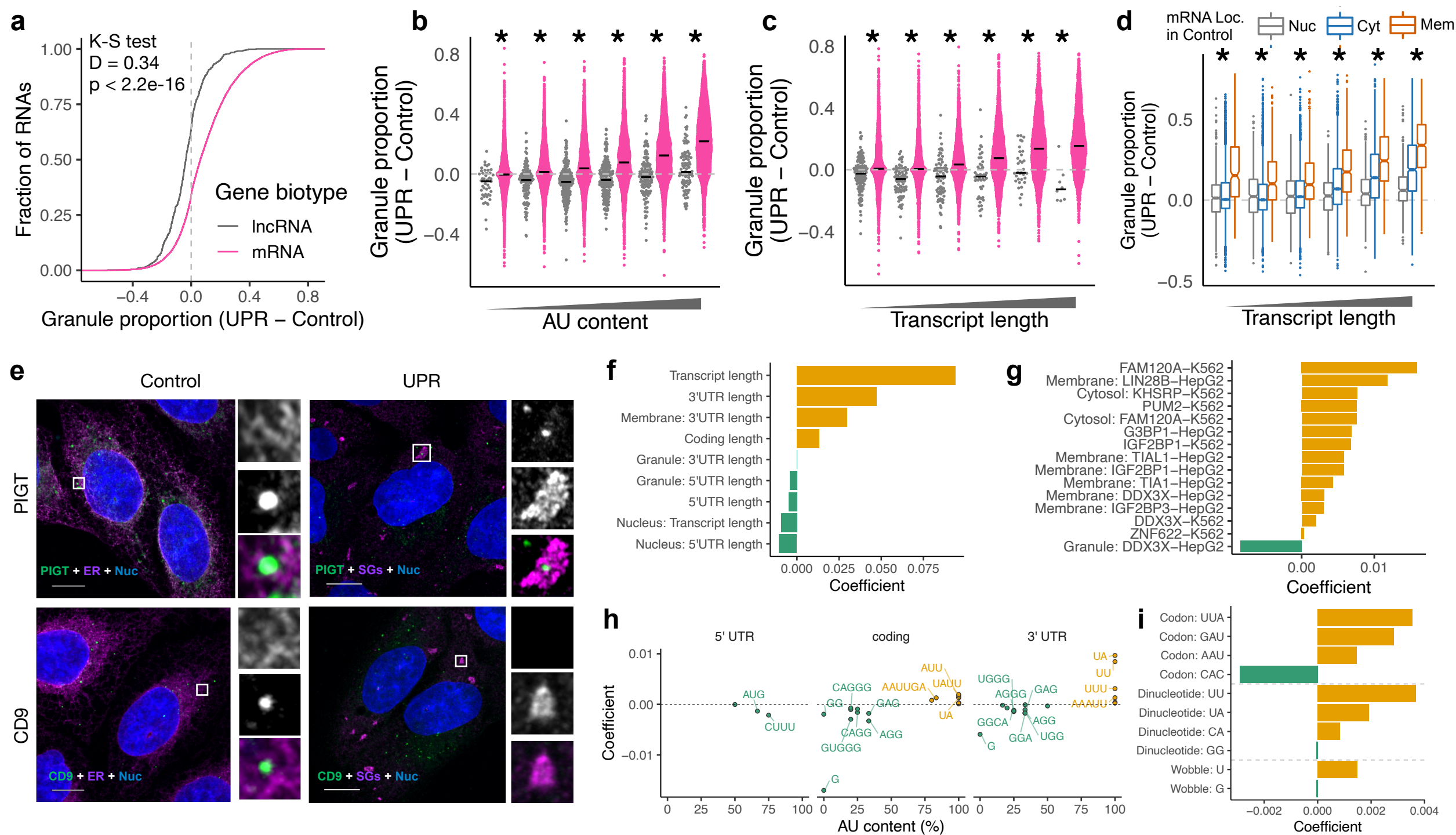
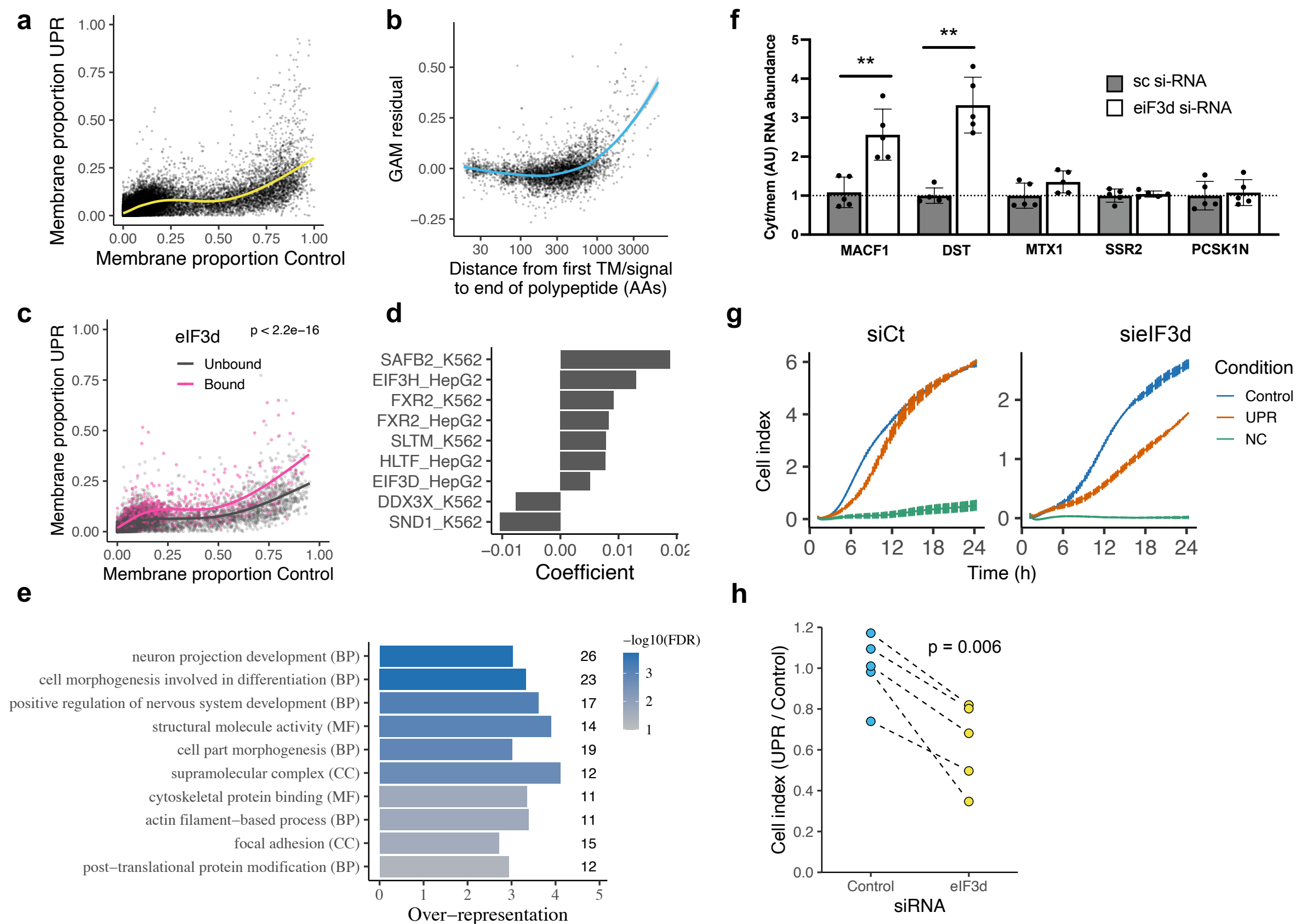
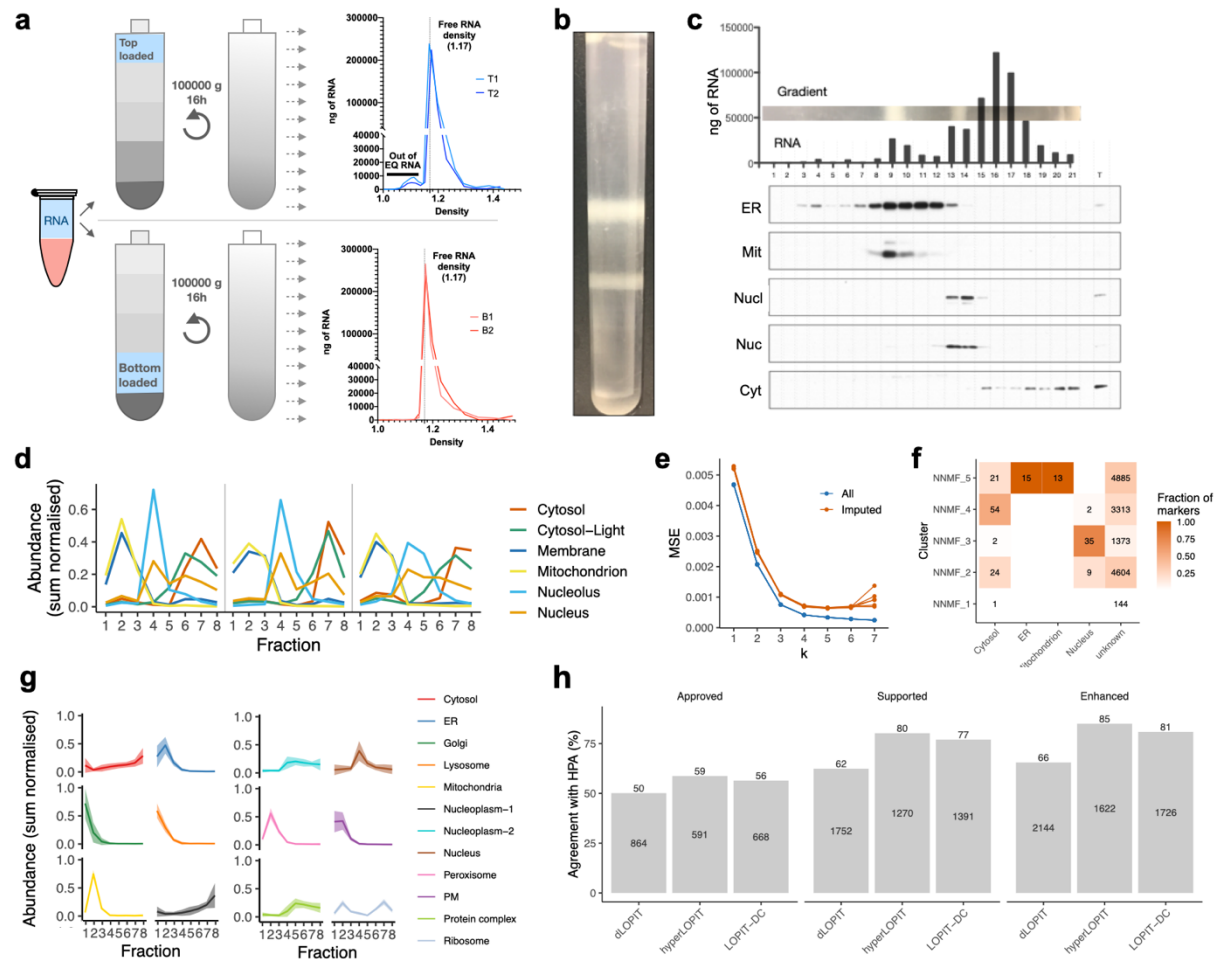


figure 6

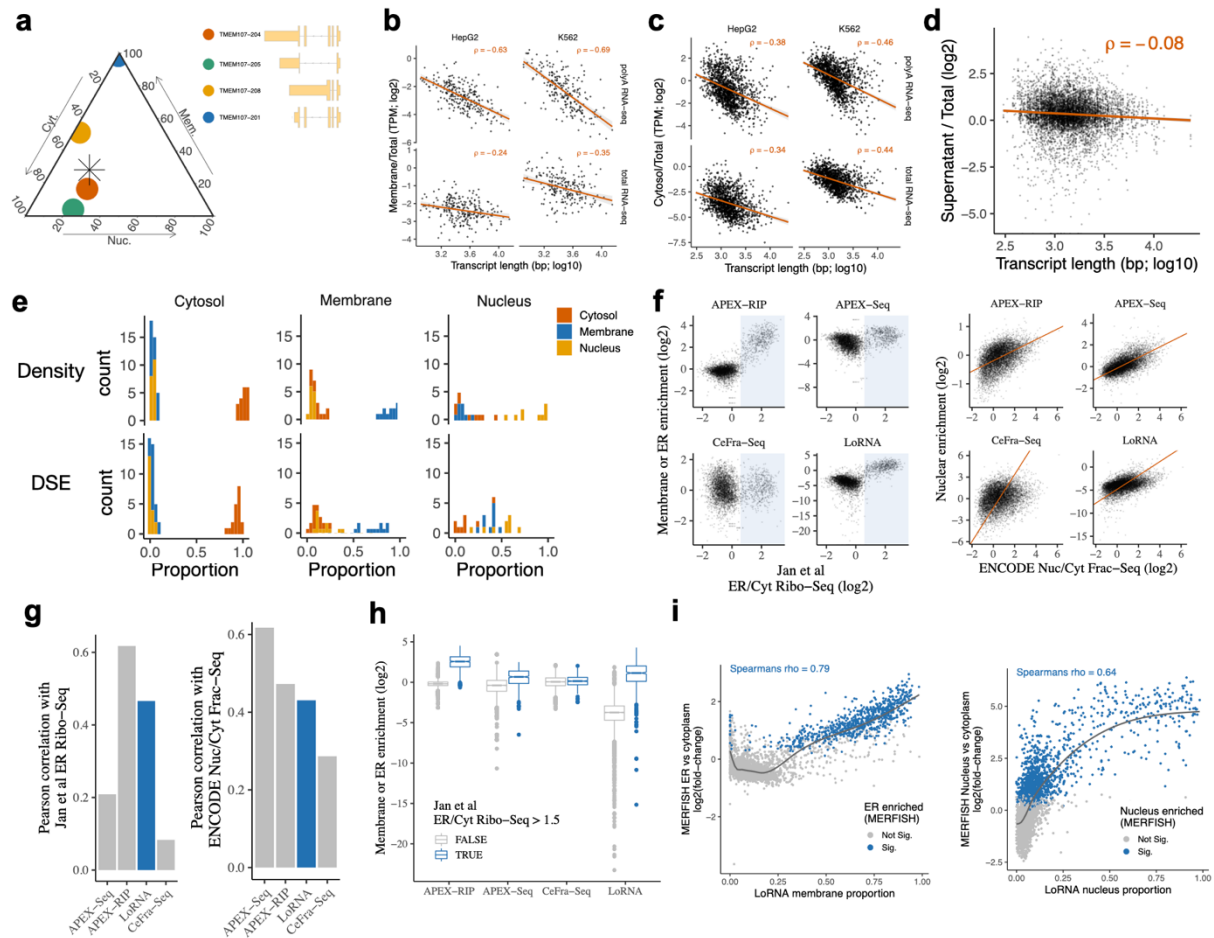


Supplementary figures and information



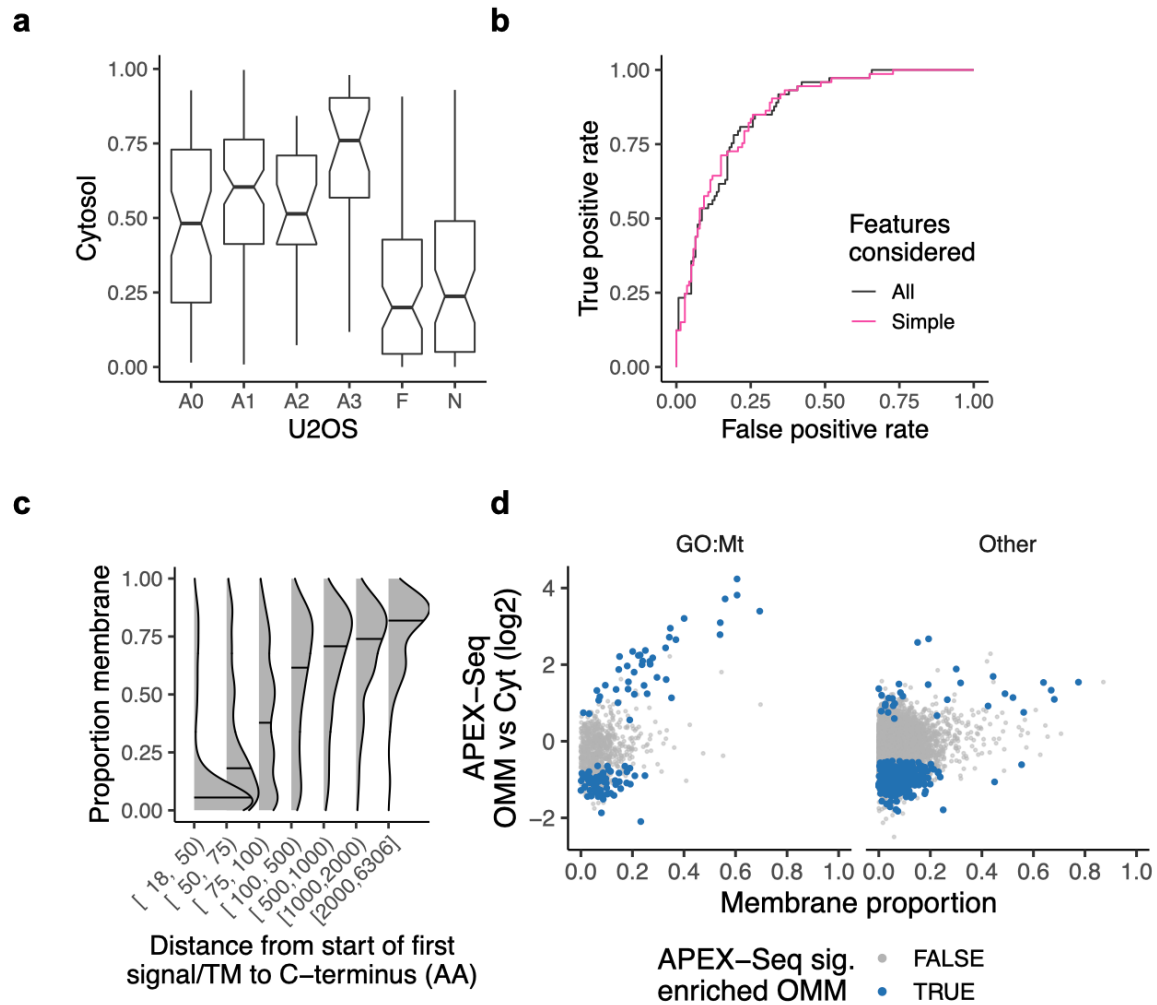
Supplementary figure 1. Simultaneous maps of RNA and protein subcellular localisation

a, Cell-free RNA sedimenting profile along the density gradient. T = top loading (blue), B = bottom loading (red, at 1.17 g/ml). Fractions with density < 1.17 are enriched in organelles. **b**, Representative image of a density-based cell fractionation gradient post centrifugation. **c**, Quantification of RNA (in ng) at the different gradient fractions, and a representative western blot of ER (Calreticulin), mitochondria (Mit, Cytochrome c oxidase subunit 4), nucleolus (Nucl, Fibrilarin), nucleus (Nuc, Histone H3) and cytosolic (b-Actin) compartment markers. **d**, Linear profiles of the indicated RNA markers per replica. **e**, Selection of optimal K for NNMF clustering of gene-level profiles in control condition. MSE = Mean Standard Error. Imputed = data points that were removed and imputed by NNMF. MSE lowest for imputed data points at k=5. **f**, Intersections between highest scoring NNMF cluster and a priori markers, where 'unknown' denotes non-markers. **g**, Protein quantification by LC-MS3 and representative protein marker profiles for one replicate. Line and shaded region represents mean +/- one standard error. **h**, Agreement between LOPIT classifications and Cell Atlas data from Human Protein Atlas, where agreement means the LOPIT classification is within the Cell Atlas annotations. Minimum Cell Atlas annotation confidence denoted above bar plot.



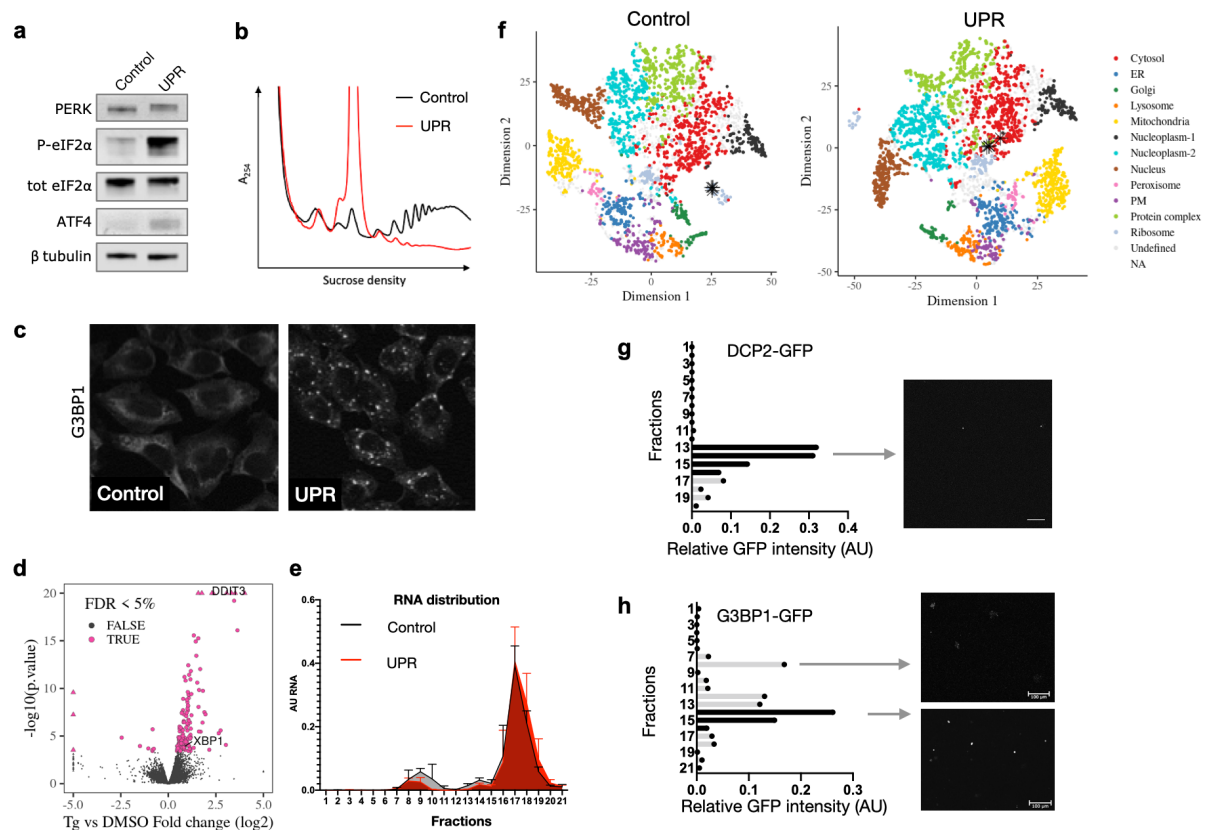
Supplementary figure 2. System-wide quantification of RNA localisation

a, Cytosol, membrane and nucleus proportions for TMEM107 transcript isoforms. Gene-level proportion indicated with asterisk (left). Transcript models (right). **b**, Transcript length vs. CeFra-Seq membrane / total for RNAs with > 80% membrane proportion according to equilibrium density centrifugation-based LoRNA. **c**, As per **b**, except cytosol / total for RNAs with > 80% cytosol proportion. **d**, As per **b**, except differential centrifugation speed-based LoRNA final fraction / total for RNAs with > 80% cytosol proportion. **e**, Proportions for cytosol, membrane and nucleus markers using equilibrium density centrifugation ('Density') and differential sedimentation ('DSE') based cell fractionation. **f**, Comparison of methods for detecting membrane (CeFra-Seq/LoRNA) or ER (APEX-Seq/APEX-RIP) enriched RNAs, against ER Ribo-Seq, and nuclear enrichment (all methods) against nucleus/cytoplasm fractionation RNA-Seq. Shaded area denotes RNAs enriched > 1.5 fold, line represents the total least squares regression. **g**, Pearson correlation vs. ER Ribo-Seq and Nuc/Cyt RNA-Seq. Correlation for ER Ribo-Seq was calculated for RNAs with ER/Cyt Ribo-Seq > 1.5 (blue box in **f**). **h**, Membrane/ER enrichment for RNAs with ER/Cyt Ribo-Seq > 1.5. **i**, Comparison between LoRNA proportions and MERFISH ER and Nucleus enrichments. Spearman rank correlation coefficient calculated for RNAs which were significantly enriched in MERFISH.



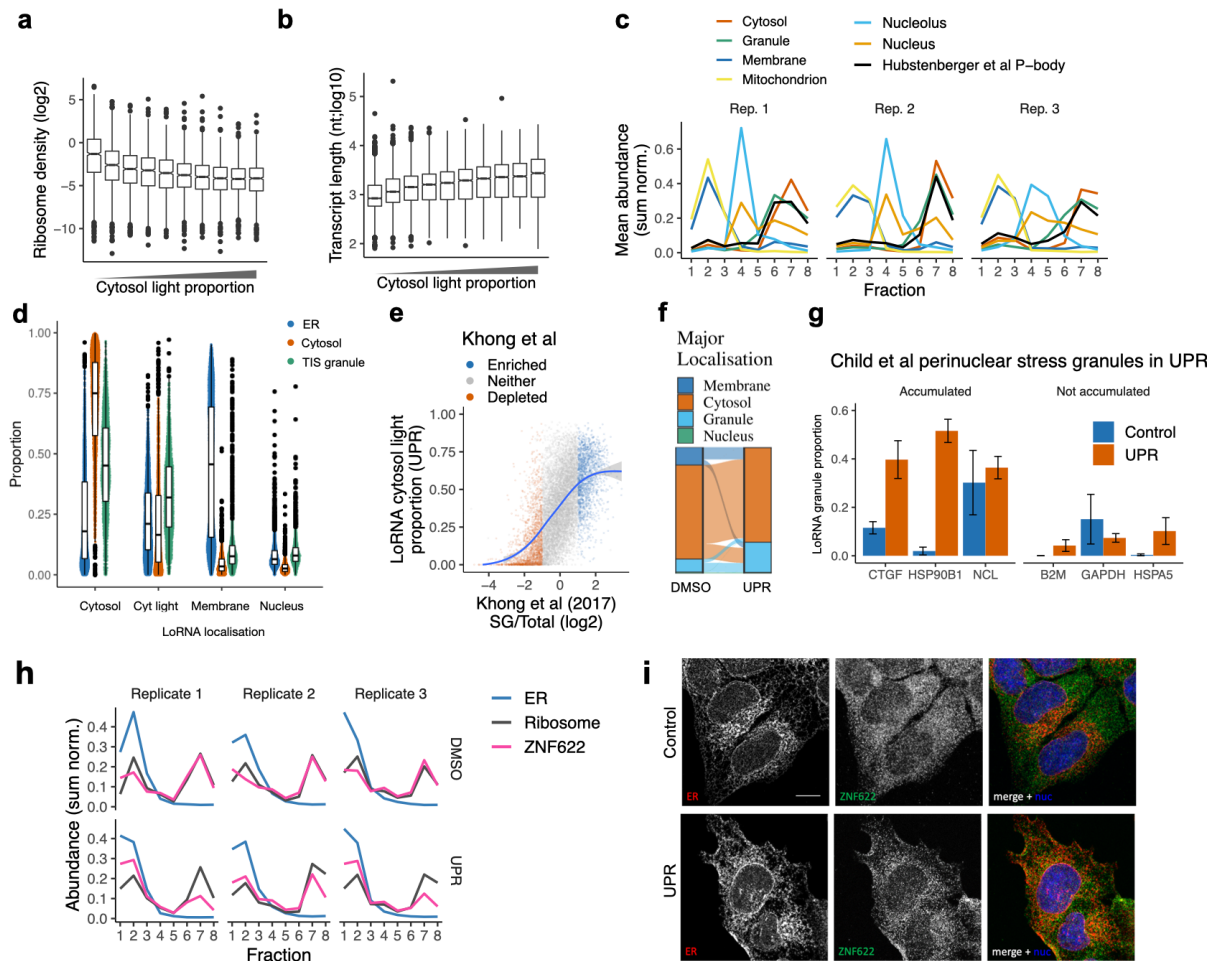
Supplementary figure 3. Features driving RNA localisation

a, Cytosol proportions for lncRNAs annotated by their ribosome association³². A0-A3 represent increasingly confident ribosome association. F=Not associated. N=Undetermined. **b**, Receiver operating characteristic curves for logistic regression models of lncRNA cytosol localisation. All=Lasso regression using all features. Simple=Just using AU content, polyA status and transcript length. **c**, Relationship between the membrane proportion of an mRNA and the distance from the first signal peptide or transmembrane domain to the stop codon of the polypeptide product. **d**, RNAs annotated as encoding mitochondrial proteins in GO and also enriched in APEX-Seq OMM¹⁰ (vs cytosol) have high membrane proportions in LoRNA.



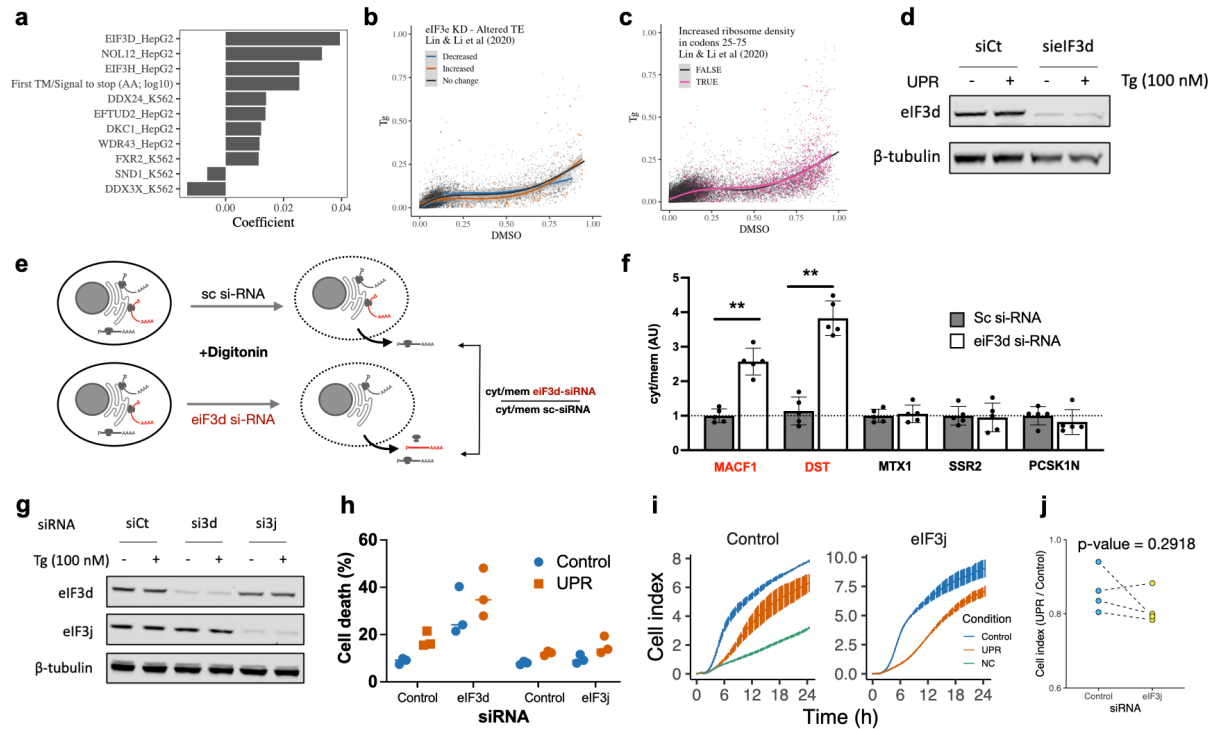
Supplementary figure 4. Transcriptome and proteome subcellular redistribution upon UPR

a, Representative western blot for samples treated with DMSO or TG showing canonical UPR markers. TG treatment induces: a shift in PERK size, consistent with phosphorylation; phosphorylation of eIF2α; increased expression of stress-specific transcription factor ATF4. **b**, Representative polysome profiles in control (DMSO) and UPR (1 h at 250 mM thapsigargin). **c**, Representative Z-projection images of control (DMSO) and UPR-induced (TG 1h) U-2 OS cells immuno-labelled with an antibody for stress granule marker G3BP1. **d**, RNA abundance changes between UPR and control. Significant changes (5% FDR) highlighted. DDIT3/CHOP and XBP1 labelled. **e**, RNA profile along unpooled density gradient fractions in control and UPR conditions. Fractions 1:11 represent light organelles (including mitochondria and ER), fractions 13 and 14 nucleus, and 15 to 21 cytosol (n = 3 per condition). **f**, t-SNE projections for protein localisation profiles in Control and UPR. The four proteins that relocalise away from the ribosome and towards the cytosol light profile are highlighted with asterisks. Proteins are coloured by their localisation allocation from BUNDLE. Proteins with inconsistent localisation allocation are denoted as 'Undefined'. **g**, DCP2-GFP sedimentation profile along unpooled density gradient fractions measured as GFP fluorescence intensity per fractions. Fractions corresponding to the density range of the cytosol light (pooled fraction 6 in Fig 1) in black (left), and representative images of the peaking fraction containing DCP2 granules respectively (right). Scale bar = 100 μm. **h**, G3BP1-GFP sedimentation profile along unpooled density gradient fractions under UPR stimulation measured as GFP fluorescence intensity per fractions. Fractions corresponding to the density range of the cytosol light in black (left), and representative images of fractions 7 and 14 representing membrane enclosed protein aggregates and G3BP1 containing granules respectively (right). Scale bar = 100 μm.



Supplementary figure 5. Analysis of the characteristics driving RNAs to granules

a, b Ribosome density and transcript length for RNAs split into deciles of cytosol light proportion, shown as Tukey boxplots. **c**, Mean profiles for RNA markers (including P-body proteins) across the 3 replicates in control condition. **d**, LoRNA proportions for RNAs classified as cytosolic (CY), canonical rough ER (ER) and TIS granules (TIS). **e**, Agreement between arsenite-induced SG enrichment and LoRNA cytosol light proportion upon UPR activation. **f**, Primary localisation of transcripts in control and upon UPR activation. **g**, Granule proportions for RNAs found to accumulate/not accumulate in stress granules upon UPR activation⁴⁶. **h**, Abundance profile of ZNF622 in Control and UPR, with average profiles for ER and ribosome markers. **i**, Representative IF images showing localisation of ZNF622 relative to the ER (calnexin) in control conditions (left panel) and upon UPR activation (right panel). Scale bar = 10 μ m.



Supplementary figure 6. Analysis of the RNAs retaining membrane association under UPR

a, Coefficients for RBPs which are predictors of GAM residuals for RNAs which encode a signal peptide/TM domain. **b**, Membrane proportions in control and UPR. RNAs with altered transcriptional efficiency (TE) upon eIF3e knockdown⁵¹ are highlighted and appear indistinct from other RNAs. **c**, Membrane proportions in control and UPR. RNAs with increased ribosome binding in codons 25-75 upon eIF3e knockdown⁵¹ are highlighted and appear indistinct from other RNAs. **d**, Western blot validating eIF3d knockdown. **e**, Schematic representation of the cytosolic and membrane associated RNA extraction using digitonin. **f**, Cytosolic and membrane associated RNA extraction using digitonin in control inductions. MTX1, SSR2 and PSK1N are non-eIF3d ER-associated controls. All RNAs are normalised to the MT-ND6 housekeeping transcript. Changes in RNA localisation upon eIF3d siRNA are relative to the RNA localisation in the scrambled si-RNA condition. (n=5). **g**, Western blot validating eIF3d and eIF3j knockdown. **h**, Quantification of cell death using Annexin V-FITC and Draq7 staining following eIF3d and eIF3j knockdown. Cells were treated with a 1 hour pulse of thapsigargin and allowed to recover for a further 24 hours (n=3). **i**, Representative U-2 OS cell migration time series for eIF3j knock-down and siRNA Control (siCt) cells in control and UPR conditions. **j**, Quantification of cell migration at 24 h post UPR induction in eIF3j knock-down and siRNA Control cells (n=4).

Table S1. Normalised RNA abundances, related to figures 1 and 2

Tables comprise gene and transcript-level abundances for control and upr conditions for density-based LoRNA and gene-level abundances for sedimentation-based LoRNA. Column name suffixes denote replicate and fraction number. Abundances are row-sum normalised within each replicate.

Table S2. Localisation proportions, related to figures 2, 3 and 4.

Tables are included for gene and transcript-level proportions for control and upr conditions for density-based LoRNA and gene-level proportions for sedimentation-based LoRNA. Further tables list the subsets of genes whose protein product is not predicted to encode a signal peptides or transmembrane domain but have > 35% membrane localisation in control or upr conditions. Column markers denotes whether a gene/transcript was a marker for a localisation. Columns prefixed with sd_ describe the standard deviation for the localisation proportion across the 3 replicates. Column signal_or_TM denotes whether the gene/transcript is predicted to encode a signal peptides or transmembrane domain.

Table S3. Protein subcellular localisation and relocalisation upon UPR activation, related to figure 4

The table includes the predicted subcellular localisation of 5315 protein in control and UPR conditions. Column markers denote whether a protein was a marker for a localisation. Columns tsne_dim1 and tsne_dim2 denote the tSNE coordinates, as shown in supplementary figure 4J. Columns primary_localisation_per_rep and primary_localisation denote the most probable localisation per replicate and, where replicates agree, the most probable localisation for the protein. Column diff_loc_level indicates the confidence level of relocalisation events, Columns n_diff_loc_reps and diff_loc_r_reps indicate how many replicates relocalisation was observed in and which replicates, respectively.

Table S4. Experimental conditions, related to all figures.

Tables comprise TMT labelling schemas, gradient fraction merging schemas for RNA and protein and antibodies used.