bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24

Xiongwen Cao[1,2,5], Alexandra Khitun[1,2,5], Zhenkun Na[1,2], Thitima Phoodokmai[3], Khomkrit Sappakhaw[3], Elizabeth Olatunji[2], Chayasith Uttamapinant[3], Sarah A. Slavoff[1,2,4*]

[1]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[2]Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States

[3]School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand

[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06529, United States

[5]These authors contributed equally

*Correspondence: sarah.slavoff@yale.edu

Abstract

While thousands of previously unannotated small and alternative open reading frames (alt-ORFs) have recently been revealed in the human genome, the functions of only a handful are currently known, and no post-translational modifications of their polypeptide products have yet been reported, leaving open the question of their biological significance as a class. Using a proteomic strategy for discovery of unannotated short open reading frames in human cells, we report the detection of alt-RPL36, a 148-amino acid protein co-encoded with and overlapping human RPL36. Alt-RPL36 interacts with TMEM24, which transports the phosphatidylinositol 4,5-bisphosphate [PI(4,5)P_2] precursor phosphatidylinositol from the endoplasmic reticulum to the plasma membrane. Knock-out of alt-RPL36 in HEK 293T cells increased PI(4,5)P_2 levels in the plasma membrane and upregulated the PI3K-AKT-mTOR signaling pathway. Remarkably, we find that four serine residues of alt-RPL36 are phosphorylated, and mutation of these four serines to alanine abolished the interaction with TMEM24 and regulation of PI3K signaling. These results implicate alt-RPL36 as a novel regulator of PI(4,5)P_2 synthesis upstream of the PI3K-AKT-mTOR signaling pathway, and the first example of a phosphorylated alt-ORF product. More broadly, both alt-RPL36 and RPL36 regulate protein synthesis and cell growth via different molecular mechanisms – PI3K signaling and ribosome composition, respectively. One human transcript can therefore express two sequence-independent polypeptides from overlapping ORFs that regulate the same processes via distinct mechanisms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Recent advances in genomic and proteomic technologies have revealed that mammalian genomes harbor thousands of previously unannotated small (smORFs, <100 amino acids) and alternative open reading frames (alt-ORFs, >100 amino acids)[1-3]. These genes previously escaped annotation not just because of their short length, but because, as a class, they exhibit low homology to proteins of known function, and are enriched for initiation at near-cognate non-ATG start codons (approximately 50%)[4,5]. Interestingly, while many smORFs and alt-ORFs lie in regions of RNA previously annotated as non-coding, up to 30% of the alt-ORFs identified by LC-MS/MS overlap an annotated protein-coding sequence in a different reading frame[3]. A rapidly increasing number of smORFs have been shown to play important roles in mammalian biology[6,7]. For example, NBDY regulates the mRNA decapping complex in human cells[8], PIGBOS regulates the endoplasmic reticulum (ER) stress response[9], and the MIEF1 microprotein regulates mitochondrial protein translation[10]. In mouse, AW112010 is a required component of the innate immune response[11], and myoregulin interacts with SERCA calcium channels in muscle[12]. These findings demonstrate that assignment of functions to small proteins represents a major opportunity to gain new insights into biology. However, all previously characterized smORF and alt-ORF-encoded proteins are expressed from a dedicated mRNA, except MIEF1 microprotein, which is encoded in a 5'UTR, and only one alt-ORF, Aw112010, initiates at a near cognate start codon. Of the dozens of non-ATG-initiated, overlapping alt-ORFs with proteomic evidence in human cells, none have yet been characterized, and the question of whether this class of alt-ORFs is functional remains open.

Multiple human ribosomal proteins are encoded in complex genes that co-express two independent functional proteins. For example, ribosomal proteins L40 and S27A are synthesized as preproteins fused to ubiquitin, and S30 as a fusion to ubiquitin-like protein[13]. A prior proteogenomics study revealed translation of a sequence-independent alt-ORF co-encoded with human 60S ribosomal protein L36 (RPL36)[14], which has nonetheless remained unannotated, likely due to a lack of information about its start codon. In this study, we provide new molecular and proteomic evidence for its existence and define the complete coding sequence of this novel protein, which we term alt-RPL36. We also demonstrate that the human RPL36 gene is dual coding via initiation of the alt-ORF at a near cognate start codon that overlaps the RPL36 nucleotide sequence in an alternative reading frame.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We further demonstrate that alt-RPL36 interacts with and regulates TMEM24. TMEM24 is anchored to the ER membrane and mediates ER-plasma membrane (PM) contacts, transporting the PI(4,5)P_2 precursor phosphatidylinositol from the ER to the PM[15]. This activity is required to replenish PI(4,5)P_2 after its phosphorylation by phosphoinositide 3-kinase (PI3K) enzymes to generate phosphatidylinositol-3,4,5-trisphosphate (PI(3,4,5)P_3) and downstream signaling lipids[16], which activate the AKT-mTOR pathway to control cell growth and protein synthesis[17,18]. The location of TMEM24 to ER-PM contacts is regulated by dynamic phosphorylation of the C-terminal region of TMEM24. Ca^{2+}-stimulated phosphorylation causes TMEM24 dissociation from the PM, and dephosphorylation allows TMEM24 to re-associate with the PM[15]. TMEM24 is mainly expressed in the brain and pancreatic islets[19], and loss of TMEM24 in insulin-secreting cells leads to a defect in insulin release[15,20].

In this study, we provide evidence for expression of alt-RPL36 from an upstream non-ATG start codon in human *RPL36* transcript variant 2, and show that it partially localizes to the cytosolic face of the ER membrane in human cells using chemical protein tagging strategies. We identify and map four phosphorylation sites that are present at high stoichiometry in alt-RPL36, and show that phosphorylation is required for its interaction with TMEM24. Finally, we engineered specific alt-RPL36 knockout and rescue cell lines to demonstrate that loss of alt-RPL36 promotes increased plasma membrane PI(4,5)P_2 levels, as well as increased activation of the PI3K-AKT-mTOR pathway and upregulation of its transcriptional outputs. These results implicate alt-RPL36 as a novel upstream regulator of PI3K-AKT-mTOR signaling and demonstrate that overlapping alt-ORFs may play biological roles in *trans*.

Results

A GTG-initiated alternative protein is translated from *RPL36* variant 2

Using a previously reported proteogenomic strategy for unannotated small protein discovery[3], we identified two tryptic peptides that mapped to an alternative reading frame of human *RPL36* transcript variant 2 in HEK 293T cells, which we name alt-RPL36 (Figure 1a,b, and Supporting Table 1). These tryptic fragments were also previously identified in the supporting information of a proteogenomic study of A431 cells but were not characterized[14]. Compared to variant 1 (NCBI RefSeq NM_033643), *RPL36* variant 2 (NCBI RefSeq NM_015414 and Figure 1a) contains a longer 5'UTR, which we hypothesized to contain a start codon initiating alt-RPL36 translation. Interestingly, the first stop codon in frame with the observed tryptic peptides is downstream of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the *RPL36* stop codon, meaning that alt-RPL36 is longer than RPL36 and completely encompasses its coding sequence. However, since alt-RPL36 is translated in the -1 reading frame relative to RPL36, the amino acid sequences of these two proteins are completely different (Figure 1a).

To confirm expression and identify the start codon of alt-RPL36, the cDNA sequence comprising the 5'UTR of *RPL36* transcript variant 2 through the stop codon of the putative alt-ORF was cloned into a mammalian expression vector with a myc tag appended to the 3' end of the putative alt-ORF. This construct produces two anti-myc immunoreactive bands (~22 kDa and ~20 kDa apparent molecular weight, due to phosphorylation, *vide infra*) from the alt-RPL36 reading frame when transiently transfected into HEK 293T cells (Figure 1d, lanes 1 and 4). Because there is no upstream ATG start codon in frame with the observed alt-RPL36 tryptic peptides, we hypothesized that alt-RPL36 initiates at a near-cognate start codon, which can initiate protein translation with methionine at a fractional efficiency relative to ATG[21,22]. We searched the upstream, in-frame DNA sequence of alt-RPL36, and found two possible near-cognate start codons in a strong Kozak sequence context: G_{114}TG and C_{165}TG (numbered relative to the first nucleotide of the cDNA). Deletion of G_{114}TG, but not C_{165}TG, abolished the expression of alt-RPL36 (Figure 1d, lanes 2, 3 and 5), and mutation of G_{114}TG to the more efficient start codon A_{114}TG increased the expression of alt-RPL36 (Figure 1d, lane 6), indicating that G_{114}TG is the start codon of alt-RPL36.

Epitope-tagging of the RPL36 coding sequence revealed that the annotated ribosomal protein is also translated from *RPL36* transcript variant 2, and that translation of alt-RPL36 has small (~40%) inhibitory effect on RPL36 synthesis, consistent with co-regulated expression of both ORFs from this transcript (Figure 1e, f). Taken together, these results indicate that human *RPL36* transcript variant 2 generates both RPL36 and an alternative protein, alt-RPL36, that initiates from G_{114}TG, in overlapping reading frames.

To determine whether alt-RPL36 is conserved among species, RPL36 mRNAs from different species were obtained from NCBI nucleotide database, then translated in the +1, +2 and +3 frames using the ExPaSy translate tool. Cognate or near-cognate start codons within Kozak consensus motifs in frame with sequences homologous to human alt-RPL36 were identified in

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the 5'UTR of each transcript in order to predict the full-length sequence of hypothetical alt-RPL36 homologs (Supporting Figure 1a-c). ClustalW alignment of these hypothetical homologs against human alt-RPL36 revealed significant sequence similarity (Supporting Figure 1d), though protein-level existence for these hypothetical homologs does not currently exist.

A number of human disease-associated mutations in the *RPL36* gene have been previously reported, and we queried whether any of these mutations could be predicted to alter the sequence of alt-RPL36 (Supporting Table 2). Of 10 cancer-specific coding sequence (CDS) mutations in *RPL36* in the COSMIC database (cancer.sanger.ac.uk)[23], 2 alter the amino acid sequence of RPL36 but are synonymous in the alt-RPL36 reading frame; 4 affect both the RPL36 and alt-RPL36 amino acid sequences; and, remarkably, 4 are synonymous with respect to RPL36 but create nonsynonymous mutations in the alt-RPL36 reading frame. While it is possible that these alt-RPL36 specific missense mutations could cause disease through alternative mechanisms, such as altering the mRNA or protein levels of RPL36, these observations could be consistent with the existence of alt-RPL36-specific cancer-associated mutations.

Alt-RPL36 partially localizes to the endoplasmic reticulum membrane and the plasma membrane

We next investigated the subcellular localization of alt-RPL36. Microproteins can mis-localize upon overexpression[24,25], so we opted to visualize alt-RPL36 in live cells at low overexpression levels with genetic code expansion. A single amino acid residue of alt-RPL36 was replaced with a bicyclononyne-lysine (BCNK) unnatural amino acid via an engineered amber-suppressor pyrrolysyl tRNA (Pyl $tRNA_{CUA}$)/Pyl tRNA synthetase (PylRS) pair[26]. In this system, expression of full-length recombinant alt-RPL36 is dependent on BCNK concentration present in the cell media as well as BCNK incubation time, allowing precise control of alt-RPL36 expression levels (Sappakhaw et al., manuscript submitted). Alt-RPL36 bearing BCNK is subsequently derivatized with a fluorophore via inverse-electron-demand Diels-Alder reaction[27] for visualization.

To label and image alt-RPL36, we selected several amino acid positions within the protein to be replaced with BCNK. Among the positions tested, only one amber variant—alt-$RPL36^{L18TAG}$—showed strong BCNK-dependent expression to produce full-length alt-RPL36 (Figure 2a). We found that incubating cells with low concentrations of BCNK (60 μM) for ~40 hours was

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sufficient to produce alt-RPL36^{L18BCNK} at levels detectable by the Diels-Alder reaction with tetrazine-cy5. Cy5-labeled alt-RPL36^{L18BCNK} was primarily cytosolic, with partial staining present at the nuclear envelope (Field of view 1, Figure 2b) as well as the extended ER network in the cytosol (Field of view 2). Anti-myc staining of alt-RPL36^{L18BCNK} also highlighted a subpopulation of alt-RPL36 present at the plasma membrane (white arrows in Fields of view 1 and 3), which was not stained with tetrazine-cy5. It is possible that the L18BCNK label site—which is adjacent to one of alt-RPL36 phosphorylation sites (*vide infra*)—is buried in a protein-protein interaction interfaces and therefore not sterically accessible to the tetrazine-cy5 reagent, or that labeling affects phosphorylation of alt-RPL36.

We provided further evidence for the localization of alt-RPL36 to the cytosol and, likely, the cytosolic face of the ER membrane by performing APEX fingerprinting[28]. In this assay, proximity-dependent biotin-phenol labeling patterns generated from APEX-fusion proteins will appear distinct, depending on the subcellular location of the fusion protein. We generated alt-RPL36 fusions with APEX at either N- or C-terminus, and performed biotin-phenol labeling in cells alongside APEX constructs targeted to different compartments of the cell. Both alt-RPL36-APEX and APEX-alt-RPL36 produced labeling patterns similar to those of APEX targeted to the cytosolic face of ER (ERM, lane 1, top) as well as cytosolic APEX (NES, lane 4, top), and clearly different from APEX targeted to the ER lumen (ER lumen, lane 2, top), ruling out alt-RPL36 localization to the latter (Figure 2c).

Four serine residues of alt-RPL36 are phosphorylated

To test whether alt-RPL36 exhibits two bands by Western blot analysis (Figure 1d, lane 1) due to protein phosphorylation[29], alt-RPL36 was immunopurified from human cells and treated with a nonspecific phosphatase. Phosphatase treatment eliminated the upper band, suggesting that it represents a phosphorylated form of alt-RPL36 (Figure 3a). To identify the phosphorylated residues, we performed LC-MS/MS with immunopurified, digested alt-RPL36, and identified four candidate phosphoserine residues (S19, S22, S140 and S142) (Figure 3b, Supporting Table 3).

Because no phosphoantibodies specific to alt-RPL36 exist, we employed Phos-tag SDS-PAGE and Western blotting, combined with mutational analysis, to confirm the phosphorylation sites identified by LC-MS/MS. As shown in Figure 3c (lane 2), wild-type, constitutively phosphorylated

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alt-RPL36 exhibited four bands. Only the bottom band remained after nonspecific phosphatase treatment, indicating that it represents unphosphorylated alt-RPL36 (Figure 3c, lane 1). Mutation of S19 to alanine abolished the second band, and mutation of S22 to alanine eliminated both the first and the second bands (Figure 3c, lanes 3 and 4), suggesting that the second band represents pS19, the first band is pS22, and pS22 is required for phosphorylation of S19. Single mutation of S140 or S142 to alanine attenuated the signal of the third band (Figure 3c, lanes 5 and 6), and double mutation of S140 and S142 to alanine abolished the third band entirely (Figure 3c, lane 7), suggesting the third band represents both S140 and S142-phosphorylated alt-RPL36. The quadruple mutant S19A S20A S140A S142A exhibited a single band comigrating with phosphatase-treated wild-type alt-RPL36, further confirming that these four serine residues are the phosphorylation sites. Quantitation of replicate Western blots revealed that approximately 86% of alt-RPL36 is constitutively phosphorylated (Figure 3d). High-occupancy phosphorylation of four specific sites in alt-RPL36 is consistent with a functional role for the protein.

Phosphorylated alt-RPL36 interacts with TMEM24 via the SMP and C2 domains

Because many small proteins characterized to date bind to and regulate other proteins[30], we performed a two-step co-immunoprecipitation (co-IP) of dually FLAG- and HA-tagged alt-RPL36 from HEK 293T cells and quantified proteins specifically enriched over untransfected controls via quantitative proteomics. After excluding common contaminants and proteins nonspecifically enriched by multiple previously reported, intrinsically disordered microproteins[24,25,31,32], we found that phospholipid transfer protein TMEM24/C2CD2L is specifically enriched by alt-RPL36 (Figure 4a, Supporting Table 4). To confirm the LC-MS/MS results, we performed reciprocal co-IP and observed enrichment of alt-RPL36-myc by TMEM24-FLAG over controls (Figure 4b, c).

TMEM 24 consists of an N-terminal transmembrane domain that resides in the ER membrane, an SMP domain, a C2 domain, and an unstructured C-terminal region that interacts with the plasma membrane (Figure 4d)[15]. To determine the domain(s) of TMEM24 with which alt-RPL36 interacts, we generated a series of truncation mutants of TMEM24 (Figure 4d), and examined their ability to interact with alt-RPL36 by co-IP and Western blotting. Deletion of the N-terminal or the C-terminal regions of TMEM24 maintained the interaction (Figure 4e, f), while further

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deletion of the C2 domain or the SMP domain abolished the interaction (Supporting Figure 2a, b), indicating that both SMP and C2 domains are required for the interaction with alt-RPL36.

Since alt-RPL36 is constitutively phosphorylated at high occupancy, we asked whether these modifications regulate its interaction with TMEM24. As shown in Figure 4g, while wild-type, phosphorylated alt-RPL36 interacts with TMEM24, the quadruple alanine mutant, which abolishes all phosphorylation of alt-RPL36, eliminated the interaction. These results indicate that the interaction between alt-RPL36 and the SMP and C2 domains of TMEM24 is specific and requires phosphorylation of alt-RPL36.

We hypothesized that, since alt-RPL36 exists in complex with TMEM24 in cells, the same kinase may recognize and phosphorylate both proteins. The C-terminal region of TMEM24 is phosphorylated by protein kinase C (PKC)[15]. Treatment of purified recombinant alt-RPL36 with PKC *in vitro* led to generation of lower-migration phosphorylated species in a Phos-tag Western blot (Supporting Figure 4). While *in vitro* kinase assays cannot establish the cellular relevance of a kinase-substrate interaction, these results suggest that PKC may be competent to phosphorylate purified alt-RPL36, consistent with the finding that alt-RPL36 interacts with TMEM24.

Phosphorylated alt-RPL36 regulates PI(4,5)P_2 transport and the PI3K-AKT-mTOR signaling pathway

TMEM24 is an ER-anchored membrane protein that transports the PI(4,5)P_2 precursor phosphatidylinositol to the plasma membrane (PM) at ER-PM contact sites via its SMP domain[15]. Phosphatidyinositol is the precursor to PI(4,5)P_2, which is converted by PI3K to PI(3,4,5)P3 to regulate downstream AKT and mTOR signaling pathways[16-18]. To determine whether alt-RPL36 regulates TMEM24-dependent phenotypes, we generated an alt-RPL36-specific knock-out (KO) HEK 293T cell line using CRISPR-Cas9. The CRISPR/Cas9 strategy targeted 210 nucleotides surrounding the alt-RPL36 start codon without perturbing the RPL36 coding sequence (Supporting Figure 3a). Expression of the transcript variant 2-specific exon containing the alt-RPL36 start codon was undetectable via mRNA-seq and RT-qPCR in KO cells, but detectable in wild-type HEK 293T cells (Supporting Figure 3a, b). In the KO cells, expression of RPL36 was unchanged at the protein level indicated by Western blotting

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Supporting Figure 3c). To determine that any observed phenotypic effects are specific to (phospho)alt-RPL36 ablation in the KO, and not off-target effects, we also generated a "rescue" cell line in which the alt-RPL36 coding sequence was stably reintroduced on the KO background, as well as "4S-A rescue" cells, in which the non-phosphorylatable mutant of alt-RPL36 was reintroduced on the KO background.

We hypothesized that the effect of alt-RPL36 on TMEM24-dependent phosphatidylinositol transport could be detectable at three levels: PI(4,5)P_2 production in the plasma membrane immediately downstream of phosphatidylinositol transport, AKT-mTOR activity, and transcriptional outputs of the AKT-mTOR pathway. First, we assayed PI(4,5)P_2 levels at the plasma membrane in wild-type, alt-RPL36 knock-out, rescue and 4S-A rescue HEK 293T cell lines stably expressing PH-PLC-GFP, which is a PI(4,5)P_2 sensor[33]. Knock-out of alt-RPL36 increased the PH-PLC-GFP signal intensity compared with wild-type HEK 293T (Figure 5a-c). The increase can be rescued by reintroduction of the coding sequence of wild-type, but not phosphorylation-incompetent, alt-RPL36. These results suggest that phosphorylated alt-RPL36 inhibits TMEM24-dependent phosphatidylinositol transport, therefore inhibiting downstream production of PI(4,5)P_2. We note that the KO and 4S-A rescue HEK 293T cells show a round morphology (Figure 5a, rows 2 and 4), which may be caused by PI3K-dependent activation of Rho/Rac-Rock signaling pathway[34-36], which induces cytoskeletal reorganization, and that rescue with alt-RPL36 ameliorates this phenotype (Figure 5a, row 3).

To determine whether alt-RPL36 regulates the AKT-mTOR signaling pathway, we measured phosphorylation of these kinases and downstream protein synthesis regulators S6K1 and S6, which correlates with their activation[17,18]. As demonstrated in Figure 5d, knock-out of alt-RPL36 increased the phosphorylation of AKT, mTOR, S6K1 and S6. Furthermore, this effect can be rescued by complementary expression of wild-type, but not phosphorylation-incompetent, alt-RPL36, consistent with the PI(4,5)P_2 assay results.

Finally, we applied mRNA-seq to quantify the transcriptional state of wild-type vs. alt-RPL36 KO HEK 293T cells. We identified 329 genes significantly upregulated and 70 genes downregulated in the KO relative to wild-type ($p < 0.05$, $n = 2$, Figure 5e). KEGG analysis of genes upregulated in the KO revealed that GO term PI3K-AKT signaling pathway is enriched (Figure 5f). To

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

confirm the mRNA-seq data, we performed conventional RT-qPCR targeting three upregulated genes in the PI3K-AKT signaling pathway. As shown in Figure 5g, all three genes are upregulated upon alt-RPL36 knockout, compared with wild-type HEK 293T cells, consistent with the mRNA-seq results. The upregulation can be partially rescued by complementary expression of wild-type, but not phosphorylation-incompetent, alt-RPL36. Taken together, these results suggest that phosphorylated alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway, most likely through its inhibitory interaction with TMEM24.

Discussion

In this work, we have identified a previously unannotated protein, alt-RPL36, that is co-encoded with human ribosomal protein L36 in the -1 reading frame of *RPL36* transcript variant 2. Alt-RP36 initiates at a non-ATG start codon upstream of the RPL36 initiation site and terminates downstream of the RPL36 stop codon, and both of these sequence-independent proteins are expressed from the same mRNA. Previous reports have demonstrated that upstream ORFs with non-ATG start codons tend to exhibit positively correlated expression regulation with downstream coding sequences[37], while upstream ATG start codons inhibit expression of downstream cistrons[38]. We therefore propose that the upstream non-ATG start codon enables co-expression of alt-RPL36 and RPL36 from the same mRNA with minimal perturbation of RPL36 translation. Despite their co-regulation, alt-RPL36 is likely present in cells at a significantly lower concentration than RPL36. The transcript variant 2-specific exon is ~10-fold less abundant than an exon specific to transcript variant 1, which only encodes RPL36, by quantitative RNA-seq (Supporting Figure 3a). Further taking into account ~10-fold less efficient translation initiation at GTG start codons relative to ATG[39], there may be at least 100-fold less alt-RPL36 in the cell than RPL36. Notwithstanding their differing cellular concentrations, it is remarkable that RPL36 – a ribosomal protein – and alt-RPL36 – a negative regulator of the AKT-mTOR pathway – regulate the same process (translation) via different mechanisms, and in different directions. To date, only one example of a functional smORF encoded in the same mRNA as a known protein – MIEF1 microprotein and MIEF1, which both play roles in mitochondria – has been reported, and these ORFs are co-encoded bicistronically rather than overlapping[10,40]. More broadly, the examples of the *MIEF1* and *RPL36* genes demonstrate that dense co-encoding of functional proteins within the same transcript - and even within overlapping nucleotide sequences - can occur in the human genome – a phenomenon previously thought to occur primarily in viral genomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Post-translational modifications (PTMs) of smORF or alt-ORF-encoded proteins have not been reported to date. The discovery that alt-RPL36 is highly phosphorylated, and that its association with its target protein requires these phosphoserine residues, may suggest that additional sm/alt-ORF-encoded proteins could bear functional modifications. However, important questions remain about alt-RPL36 phosphorylation. Most importantly, dephosphorylation of alt-RPL36 would be required to alleviate its apparent inhibition of TMEM24 activity in cells, but constitutive alt-RPL36 phosphorylation was observed under the conditions of this study. it is possible that the phosphorylation state of alt-RPL36 is dynamic, as has been previously reported for TMEM24[15,19]. Considering that TMEM24 is primarily expressed in electrically excitable cells with secretory functions, future functional studies of alt-RPL36 phosphorylation dynamics should be conducted in these contexts.

The subcellular localization of alt-RPL36 may be relevant to its function, but several questions remain to be answered. TMEM24 has been reported to localize to ER-PM contact sites, and to relocalize throughout the ER when its C-terminal domain is phosphorylated. The observation that alt-RPL36 localizes to the ER surface and cytosol in the fluorescent labeling channel, but is enriched at PM contact sites in the immunostaining channel, could be consistent with TMEM24 binding, but the discordance in these channels could suggest that bio-orthogonal Cy5 labeling visualizes a minor pool of the alt-RPL36 protein, while anti-myc staining visualizes the entire population, some of which is not Cy5-labeled. This could be a result of (1) over-expression artifacts, (2) inaccessibility of the labeling site to bio-orthogonal derivatization when in complex with TMEM24, (3) perturbation of alt-RPL36 phosphorylation upon bio-orthogonal labeling, or (4) perturbation of TMEM24 localization when in complex with phosphorylated alt-RPL36, among other possibilities. Further work is needed to elucidate the subcellular localization of alt-RPL36.

Finally, our functional data suggest that phosphorylated alt-RPL36 inhibits TMEM24. However, the functional assays employed in this study are indirect readouts of TMEM24 activity. Therefore, the mechanism by which phosphorylated alt-RPL36 inhibits TMEM24 remains to be established, and it further remains possible that some observed phenotypes could be due to non-TMEM24-dependent actions of alt-RPL36. For example, we cannot rule out interaction of alt-RPL36 with C2CD2, a close homolog of TMEM24, or other cellular proteins, at this time. Regardless, given the direct evidence for alt-RPL36-TMEM24 interaction, and the necessity of phosphorylated alt-RPL36 to rescue wild-type inositide synthesis and function, it is reasonable

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to hypothesize that alt-RPL36 inhibits TMEM24. It is possible that the alt-RPL36 interaction with the SMP and C2 domains of TMEM24 affects its lipid transport activity, either by competitive binding or conformational change. It is also possible that phosphorylated alt-RPL36 contributes to electrostatic inhibition of the TMEM24-PM interaction, which has been shown to occur upon phosphorylation of the TMEM24 C-terminus[15,19], and could interfere with phospholipid transport between these membranes. Simpler mechanisms, such as alteration of the stability, folding or localization of TMEM24 by alt-RPL36, are also possible. Cell-based, biochemical and structural studies will be required to differentiate these possible mechanisms of action.

Online methods

Data analysis. Statistics, two-tailed *t*-test was performed using Excel or Prism, and equal variance between samples being compared was established using an *F*-test.

Antibodies. Primary antibodies for Western blotting include the following anti-FLAG (Sigma, F3165); anti-myc (Rockland, 600-401-381, Cell Signaling, 2276 (mouse), or Cell Signaling, 2278 (rabbit)); anti-HA (Invitrogen, 71-5500); anti-His (Thermo Fisher, MA1-21315); anti-V5 (Cell Signaling, 13202); anti-β-actin (Invitrogen, BA3R); anti-p-mTOR(S2448) (Cell Signaling Technology, 2971); anti-mTOR (Cell Signaling Technology, 2972); anti-p-AKT(S473) (Cell Signaling Technology, 9271); anti-AKT (Cell Signaling Technology, 9272); anti-p-S6K1(T389) (Cell Signaling Technology, 9234); anti-S6K1 (Cell Signaling Technology, 9202); anti-p-S6(S235/236) (Cell Signaling Technology, 2211); anti-S6 (Cell Signaling Technology, 2217); anti-RPL36 (Bethyl Laboratories, A305065A-M). Immunoprecipitation was performed with the following antibody beads: anti-FLAG M2 affinity gel (Sigma, A2220); anti-myc tag agarose beads (Sigma, A7470); anti-HA tag magnetic beads (Thermo Fisher, 88836).

Cloning and genetic constructs. A construct comprising the full 5'UTR of human *RPL36* transcript variant 2 through the stop codon of alt-RPL36 was synthesized by Genscript with a myc epitope tag appended to the 3' end of the alt-RPL36 coding sequence, then subcloned into pcDNA3. For generation HEK 293T cells stably expression alt-RPL36, a dual FLAG and HA tag were appended to the 3' end of alt-RPL36 by PCR, and the GTG start codon was mutated to ATG. The dually tagged coding sequence was amplified by PCR, then cloned into pLJM1. The serine point mutations with a dual FLAG and HA tag were generated by ligating the PCR products into AgeI and EcoRI cloning sites in the pLJM1 vector. The full-length TMEM24 clone

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with a C-terminal FLAG epitope tag in pcDNA3 was purchased from Genscript. The truncations of TMEM24 were generated by PCR with a C-terminal myc epitope tag, then subcloned into pcDNA3. The cDNA clone expressing PH-PLC-GFP was purchased from Addgene (a gift from Tobias Meyer, Stanford), then cloned into pLJM1 by PCR. pPB-MmPylRS-AF-4xPylT$_{CUA}$, which expresses *M. mazei* pyrrolysyl tRNA synthetase$^{Y306A/Y384F}$ (PylRS-AF) and 4 copies of pyl tRNA$_{CUA}$, was a gift from Jason Chin (MRC Laboratory of Molecular Biology, UK). pPB-alt-RPL36-4xPylT$_{CUA}$ was cloned via subcloning alt-RPL36 gene to replace MmPylRS-AF via restriction cloning. Site-directed mutagenesis to introduce L18TAG, A30TAG, and V31TAG mutations to alt-RPL36 were performed on pPB-alt-RPL36-4xPylT with standard protocols. Constructs expressing subcellular APEX2--pcDNA5-V5-APEX2-NES (cytosolic targeting), pcDNA5-V5-APEX2-NLS (nuclear targeting), pcDNA5-V5-APEX2-Sec61β (targeting to cytosolic face of ER membrane), and pcDNA3-Sec61β-V5-APEX2 (ER lumen targeting) were gifts from Hyun-Woo Rhee (Seoul National University, South Korea). pcDNA5-alt-RPL36-APEX2 and pcDNA3-APEX2-alt-RPL36 were cloned via subcloning alt-RPL36 to replace Sec61β portion of pcDNA5-V5-APEX2-Sec61β and pcDNA3-Sec61β-V5-APEX2 respectively, via restriction cloning.

Cell culture and transfection. HEK 293T and HeLa cells were purchased from ATCC and early-passage stocks were established in order to ensure cell line identity; cells were maintained up to only 10 passages. HEK 293T and HeLa cells were cultured in DMEM (Corning, 10-013-CV) with 10% FBS (Sigma, F0392) and 1% penicillin-streptomycin (VWR, 97063-708) in a 5% CO_2 atmosphere at 37℃. Plasmid transfection was performed with Lipofectamine 2000 or Lipofectamine 3000 and Opti-MEM (GIBCO, 31985-070) according to the manufacturer's instructions, or polyethyleneimine (PEI, Polysciences, 23966-1) according to established protocol[41].

Lentivirus production and stable cell line generation. Lentivirus was produced as previously described[42]. Briefly, HEK 293T cells were co-transfected using Lipofectamine 2000 or polyethyleneimine with expression construct in pLJM1, along with pMD2.G and psPAX2, and growth media were replaced after 3-5 h. 48 h post-transfection, media containing viruses was harvested, filtered through a 0.45-µm filter, and infection was performed by mixing with two volume fresh media containing suspended HEK 293T cells. 24 h post-infection, the growth media was replaced. 48 h post-infection, stably expressing cells were selected with 6 µg/mL

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

puromycin for 2 days. Early stocks of stable cell lines were established after selection. Stable cell lines were released from puromycin for 2 days prior to use in experiments.

Alt-RPL36 labeling and imaging. HEK 293 cells were seeded onto 6-well plates for Western blot analysis or on glass coverslips placed in 24-well plates for imaging studies. The coverslips were incubated for 1 hour with 0.1 mg/mL poly-D-lysine. Upon reaching ~ 80% confluency, the media was exchanged for Opti-MEM containing transfection mixture of DNA (plasmid ratio of pPB-alt-RPL36-4x$PylT_{CUA}$: pPB-mMPylRS-AF-4x$PylT_{CUA}$ = 9 : 1), lipofectamine-3000 and its accessory reagent. After incubation for 6 hours at 37 °C, the media was exchanged for DMEM supplemented with 10% FBS, 1% Pen-strep and 60 uM BCNK (Sichem). Subsequently, cells were grown in presence of BCNK for 45 hours. To analyze BCNK-bearing alt-RPL36 expression, cells were lysed with RIPA buffer, and lysates resolved on polyacrylamide gels. Proteins were transferred onto nitrocellulose membranes, and Western blotting with anti-myc antibody was performed using standard protocols and imaged with ImageQuant LAS500 (GE Healthcare).

To fluorescently label alt-RPL36 via inverse-electron-demand Diels-Alder cycloaddition, cells were fixed with 4% formaldehyde then methanol, blocked with 3% BSA in PBS, then labeled for 10 minutes with 500 nM tetrazine-Cy5 (Jena Bioscience) in 5% skimmed milk in PBS at room temperature. Cells were further immunostained with anti-myc antibody and appropriate fluorescently labeled secondary antibodies. Imaging were performed with Zeiss LSM800 confocal microscope with appropriate laser and filter settings, and analyzed with ZEN Software.

APEX fingerprinting. HEK 293 cells were seeded onto 6-well plates and upon reaching ~ 80% confluency, transfected with the plasmid encoding appropriate APEX2-fusion proteins via lipofectamine 3000. 24 hours post-transfection, cells were incubated with 500 uM biotin-phenol for 30min, then treated with 1mM H_2O_2 for 1min. Cells were immediately washed with PBS thereafter, and lysed in RIPA buffer. Lysates were resolved on polyacrylamide gels, and protein content transferred onto nitrocellulose membranes. Biotinylated proteins on the membranes were detected with streptavidin-HRP (Abcam), while APEX2 expression was detected with anti-V5 antibody and appropriate secondary antibodies conjugated to HRP. Chemiluminescent blot imaging was performed with ImageQuant LAS500.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PH-PLC-GFP imaging. HEK 293T cells stably expressing PH-PLC-GFP were grown to 80% confluency on coverslips in 12-well plates. Coverslips were inverted and imaged in pre-warmed DMEM with 10% FBS, 1% penicillin-streptomycin in MatTek imaging dishes. Confocal imaging was performed on a Leica SP8 LS confocal microscope with 63× oil immersion objective under atmosphere-controlled stage at 37℃. Quantification was performed with Image J using standard parameters.

Co-immunoprecipitation. HEK 293T cells were grown to 50% confluency in 15 cm dishes, then transiently transfected with the indicated plasmids. 48 h after transfection, cells were harvested and suspended in 1 mL lysis buffer (Tris-buffered saline (TBS) with 1% Triton X-100 and Roche Complete protease inhibitor cocktail tablets). Then cells were sonicated (50% intensity, 5 s pulse with 25 s rest, 5×, MICROSON XL 2000) on ice followed by centrifugation at 15,000 r.p.m., 4℃, 10 min. 1% of lysate samples were saved for analysis of loading. A 25 µL aliquot of anti-Flag agarose beads (clone M2, Sigma) was washed with 1 mL lysis buffer, collected by centrifugation for 2 min at 1,000 g, then suspended in the cell lysate supernatant. Beads suspensions were rotated at 4℃ for 1 h, then washed 3 times with wash buffer (Tris-buffered saline (TBS) with 1% Triton X-100 and 350 mM NaCl), then 1 time with lysis buffer. Proteins were eluted by adding 40 µL SDS-PAGE loading buffer and boiling.

Immunoprecipitation and proteomics. Control HEK 293T cells or HEK 293T cells stably expressing alt-RPL36-FLAG-HA were grown to 80-90% confluency in 15 cm dishes. Cells were harvested and immunoprecipitation were performed as described above. After the final wash, elution was in 40 µL of 3× FLAG peptide (Sigma), at a final concentration of 100 µg/mL in lysis buffer at 4℃ for 1 h. Beads were removed by centrifugation and the entire supernatant were collected. A 50 µL aliquot of anti-HA magnetic beads (Cat.88836, Thermo Fisher) was washed with 1 mL lysis buffer, then suspended in the elution. Beads suspensions were rotated at 4℃ for 4 h, then washed 3 times with lysis buffer. Elution was in 40 µL of HA peptide (Covance), at a final concentration of 400 µg/mL in lysis buffer at 4℃ for 2 h. The eluted proteins were subjected to MS-MS analysis.

Proteomics and database searches. Protein-containing gel slices were digested with trypsin at 37℃ for 14-16 h. The resulting peptide mixtures were extracted from the gel, dried, followed with ethyl acetate extraction to remove residual detergent, then re-suspended in 15 µl of 3:8 70% formic acid:0.1% TFA. A 5 µL aliquot of each sample was injected onto a pre-packed

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

column attached to a nanoAcquity UPLC (Waters) in-line with an LTQ Orbitrap Velos (Thermo Scientific) and a 90-min gradient was used to further separate the peptide mixtures as follows (solvent A: 0.1% formic acid; solvent B: acetonitrile with 0.1% formic acid): Single pump trapping was turned on for 6 min at a flow rate of 2.5 µL/min at 98% A. Isocratic flow was maintained at 0.3 µL/min at 2% B for 10 min, followed by linear gradients from 2% B to 10% B over 2 min, 10% B to 25% B over 58 min, 25% B to 40% B over 10 min, 40% B to 95% B over 2 min. Isocratic flow at 95% B was maintained for 5 min, followed by a gradient from 95% B to 2% B over 10 min. The column flow rate was 0.3 µL/min. The full MS was collected over the mass range of 298-1,750 m/z with a resolution of 30,000. MS/MS data was collected using a top 10 high-collisional energy dissociation method in data-dependent mode with a normalized collision energy of 33.0 eV and a 2.0 m/z isolation window. The first mass was 100 m/z in fixed mode. MS/MS resolution was 7,500 and dynamic exclusion was 60 seconds.

For co-IP proteomics searches and quantitative analysis, oxidation of methionine and N-terminal acetylation were set as variable modifications. And files were analyzed using MaxQuant. For phosphoproteomics searches, phosphorylation of Ser, Thr and Tyr, oxidation of methionine and N-terminal acetylation were set as variable modifications. ProteoWizard MS Convert was used for peak picking and files were analyzed using Mascot. For all analysis, a mass deviation of 20 p.p.m. was set for MS1 peaks, and 0.6 Da was set as maximum allowed MS/MS peaks with a maximum of two missed cleavages. Maximum false discovery rates (FDR) were set to 1% both on peptide and protein levels. Minimum required peptide length was five amino acids. Protein quantitation was accomplished via spectral counting, where the number of total peptides observed for each identified protein was taken as the total spectral counts and compared for the IP vs. negative control samples.

Phos-tag SDS PAGE. Phos-tag SDS-PAGE was performed with 10% polyacrylamide gels containing 50 µM Phos-tag acrylamide and 100 µM $MnCl_2$. After electrophoresis, Phos-tag acrylamide gels were washed with transfer buffer containing 10 mM EDTA for 10 min with gentle shaking and then with transfer buffer without EDTA for 10 min according to the manufacturer's protocol. Proteins were transferred to nitrocellulse membranes followed by a standard Western bloting protocol[8].

Phosphatase treatment. Alt-RPL36-FLAG-HA expressed and immunopurified from HEK 293T cells was incubated with 0.4 U/ µL alkaline phosphatase, calf intestinal (CIP) (NEB, cat. M0290)

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

at 37℃ for 1 h, or 8 U/ µL lambda protein phosphatase (NEB, cat. P0753) at 30℃ for 0.5 h according to the manufacturer's instructions in a total reaction volume of 20 µL before Western blotting.

Recombinant expression and purification of human alt-RPL36. His-MBP-tagged human alt-RPL36 in pET21a was transformed into an *E. coli.* BL21 (DE3) strain. Overnight cultures were diluted 1:500 in Luria-Bertani (LB) media supplemented with 100 µg/mL of ampicillin and grown at 37°C. Expression was induced by adding 0.5 mM of isopropyl β-d-1-thiogalactopyranoside (IPTG) when OD 600 reached 0.6, followed by shaking at 18°C overnight. After cell harvest and resuspension in lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 2 mM imidazole) with 1 mM β-mercaptoethanol and Roche Complete protease inhibitor cocktail tablets, cells were lysed by sonication (30% intensity, 10 s pulse with 50 s rest on ice, 20×). The solution was clarified by centrifugation at 14,000 rpm for 20 min at 4°C and loaded onto a column containing 1 mL of Co^{2+}-TALON resin (Takara, 635606) pre-equilibrated with lysis buffer. Following incubation for 1 h at 4°C, the resin was washed with 10 mL of lysis buffer and 30 mL wash buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 10 mM imidazole), and proteins were eluted with elution buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 250 mM imidazole). Eluted proteins mixed with 15% glycerol were snap-frozen and stored at -80°C.

In vitro phosphorylation. In vitro alt-RPL36 phosphorylation by PKCα (Invitrogen, P2232) was performed by incubating various concentrations (0-0.3 µM) of PKCα with 2 µM purified His-MBP-alt-RPL36 in reaction buffer (20 mM HEPES-KOH pH 7.4, 16.7 mM CaCl2, 10 mM DTT, 100 mM MaCl2, 100 µM ATP and 0.6 mg/mL phosphatidylserine) at 30℃ for 0.5 h in a total reaction volume of 10 µL.

Generation of alt-RPL36 knock out cells. Alt-RPL36 KO HEK 293T cells were generated using CRISPR-Cas9, the guide RNAs (gRNAs) were designed with the guide design tool from the Zhang lab (crispr.mit.edu) to target the RPL36 genomic region (gRNA1: 5'-CCGGGATATCTACTCGGCTC-3'; gRNA2: 5'-GAGTACCGGCTCAGTTCCCG-3'). Double-stranded DNA oligonucleotides corresponding to the gRNAs were inserted into pSpCas9(BB)-2A-GFP vector (Addgene, as a gift from F. Zhang, MIT, Cambridge, MA). An equal mixture of the two gRNA plasmids were transfected into HEK 293T cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions, and GFP-positive cells were sorted with flow cytometry. Loss of alt-RPL36 expression was confirmed by genomic DNA PCR and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequencing. In the alt-RPL36 KO cell line used in this study, the two alleles were disrupted by a 210-nt homozygous deletion, deleting the alt-RPL36 start codon GTG, but not the RPL36 start codon ATG.

Alt-RPL36 conservation analysis. For alt-RPL36 conservation analysis, RPL36 mRNAs from different mammalian species were obtained from the NCBI nucleotide database, then translated in the +1, +2 and +3 frames using ExPaSy translate tool. The Cognate or near-cognate start codons within Kozak consensus motifs in frame were identified in the 5'UTR of each transcript, and were considered as the first amino acid (Methionine) of alt-RPL36. The hypothetical proteins thus derived were aligned with Clustal Omega using standard parameters.

mRNA-seq and data analysis. Whole RNA was isolated from 2.5 x 10^6 control or alt-RPL36 knockout (n = 2) HEK 293T cells using Qiagen RNeasy Mini Kit spin columns, then treated in solution with DNase I prior to Qiagen column clean-up according to the manufacturer's protocol. Whole RNA was submitted to the Yale Center for Genomic Analysis for preparation according to the standard Illumina protocol for paired-end sequencing with enrichment of poly-A RNA. Samples were multiplexed and 75 bp fragments were sequenced on the Hiseq2500 sequencer. The reads were mapped to the human genome (hg19) using TopHat (v2.0.11). To identify the genes regulated by alt-RPL36, we counted the RNA reads in exons and calculated the reads per kilobase per million reads (RPKM) for each gene as a measure of expression using Cufflinks (v.2.2.1) with the Cuffdiff tool using default parameters. A $\log_2$ (fold change) cutoff of 0.2 with p value ≤ 0.05 was used to identify the genes with downregulated or upregulated expression in the alt-RPL36 knockout cells, compared with that in the control cells. Gene ontology analysis was performed using g:Profiler with standard parameters[43]. For qRT-PCR, total RNA was isolated using Trizol (Invitrogen) according to the manufacturer's protocol. Reverse transcription was performed with iScript (Bio-Rad) and qPCR was performed with iTaq Universal SYBR Green Supermix (Bio-Rad), with quantitation by a relative Ct method. qPCR primer sequences are provided in Supplementary Table 5.

Data availability. The mRNA-seq data have been deposited in the NCBI Gene Expression Omnibus under accession GSE144979.

Author Contributions

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

X.C., A.K. and E.O. designed and performed experiments and analyzed data. Z.N. designed and constructed CRISPR/Cas9 knockout cells. T.P. and K.S. performed imaging and APEX fingerprinting. C.U. designed experiments and analyzed data. S.A.S. conceived the project, designed experiments, and analyzed data. X.C., C.U., and S.A.S. wrote the manuscript, and all authors edited and approved the final version of the manuscript.

Acknowledgment

We thank Benjamin Turk, Pietro de Camilli, and the members of the Slavoff lab for helpful discussions. No competing financial interests have been declared. This work was supported by an American Cancer Society Institutional Research Grant Individual Award for New Investigators (IRG-58-012-57), the Searle Scholars Program, an Odyssey Award from the Richard and Susan Smith Family Foundation, and Yale University West Campus start-up funds (to S.A.S.). X.C. was supported in part by a Rudolph J. Anderson postdoctoral fellowship from Yale University. A.K. was in part supported by an NIH Predoctoral Training Grant (5T32GM06754 3-12). T.P., K.S., and C.U. were supported by student and research assistant funds from VISTEC (to T.P. and K.S. respectively), a startup grant from VISTEC (to C.U.), Thailand Research Fund (MRG6280177, to C.U.), and Seed Awards in Science from Wellcome Trust (to C.U.).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

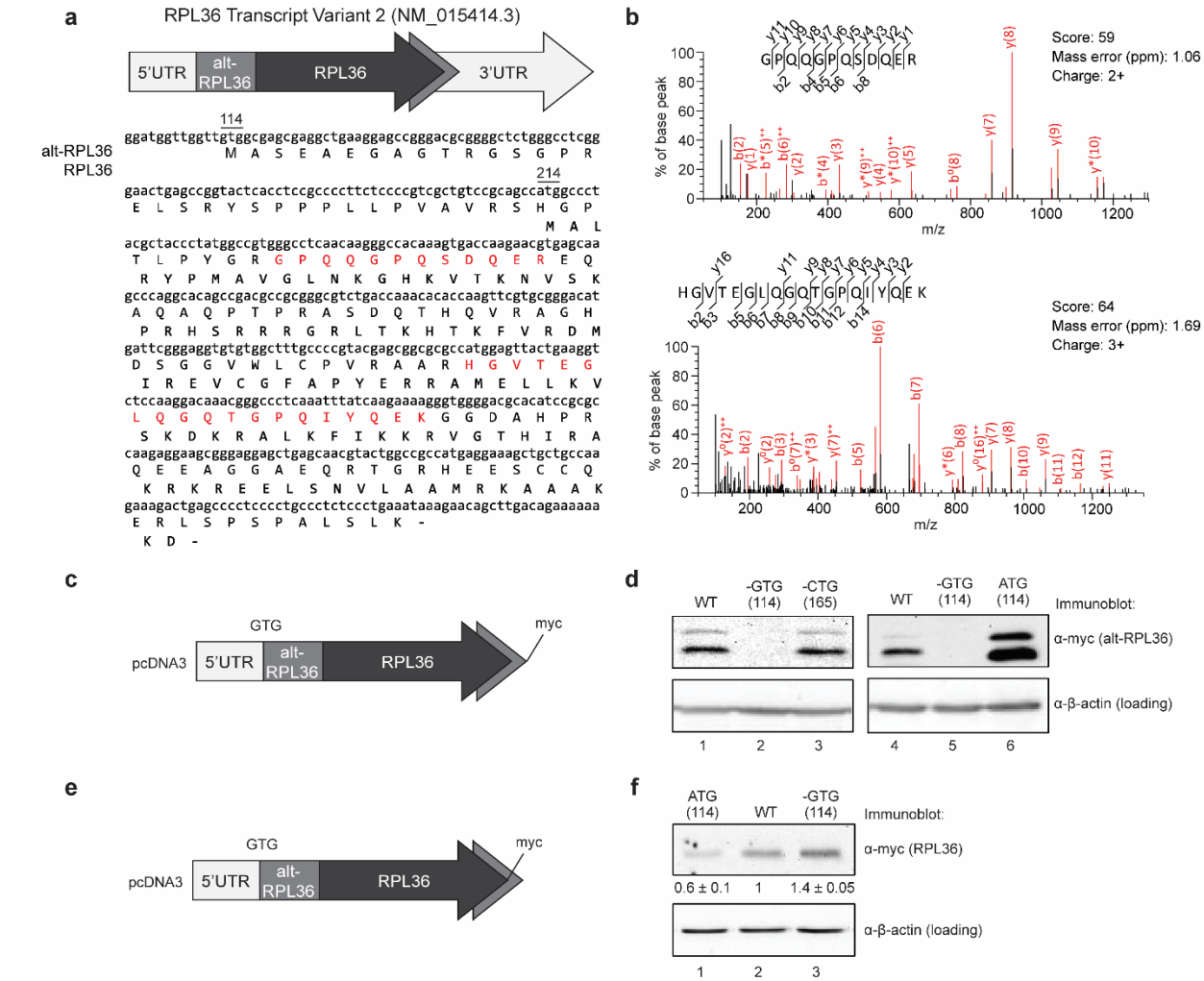


Figure 1 | A GTG-initiated alternative protein is translated from *RPL36* variant 2. (**a**) Top, a schematic representation of human RPL36 transcript variant 2 (tv2); light gray arrow, 5' and 3' untranslated regions (UTR); mid gray, alternative open reading frame (alt-ORF) encoding alt-RPL36; dark gray, annotated *RPL36* coding sequence. Bottom, the cDNA sequence of human *RPL36* transcript variant 2 is shown with the protein sequences of alt-RPL36 and RPL36 (bold) indicated below. The GTG start codon of alt-RPL36 and ATG start codon of RPL36 are numbered above the cDNA sequence. Highlighted in red are two tryptic peptides of alt-RPL36 detected by MS/MS. (**b**) MS/MS spectra of two alt-RPL36 tryptic peptides detected via peptidomics in HEK 293T cells. (**c,d**) Expression of a construct containing the full 5'UTR and alt-RPL36 coding sequence derived from *RPL36* tv 2, with a myc tag appended to the C-terminus of alt-RPL36 (**c**), in HEK 293T cells, was followed by lysis and Western blotting with the antibodies indicated to the right (**d**). (**e,f**) Expression of a construct containing the full 5'UTR and alt-RPL36 coding sequence derived from *RPL36* tv 2, with a myc tag appended to the C-terminus of RPL36 (**e**), in HEK 293T cells, was followed by lysis and Western blotting with the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

antibodies indicated to the right (**f**). Quantitative analysis of the Western blot signal of RPL36-myc are indicated at the bottom. Data represent mean values ± standard error of the mean (s.e.m.) of three biological replicates. All Western blots are representative of results obtained in three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

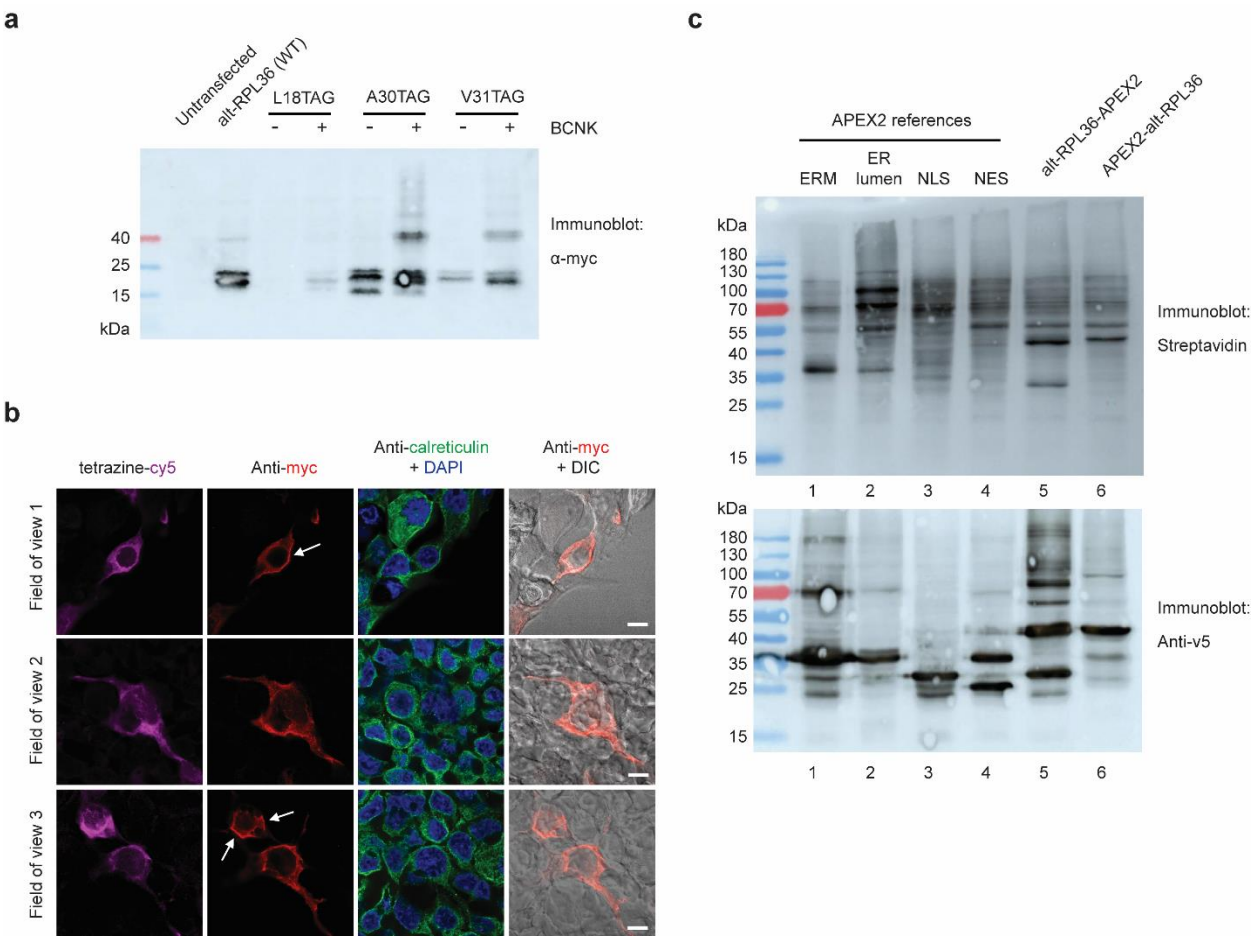
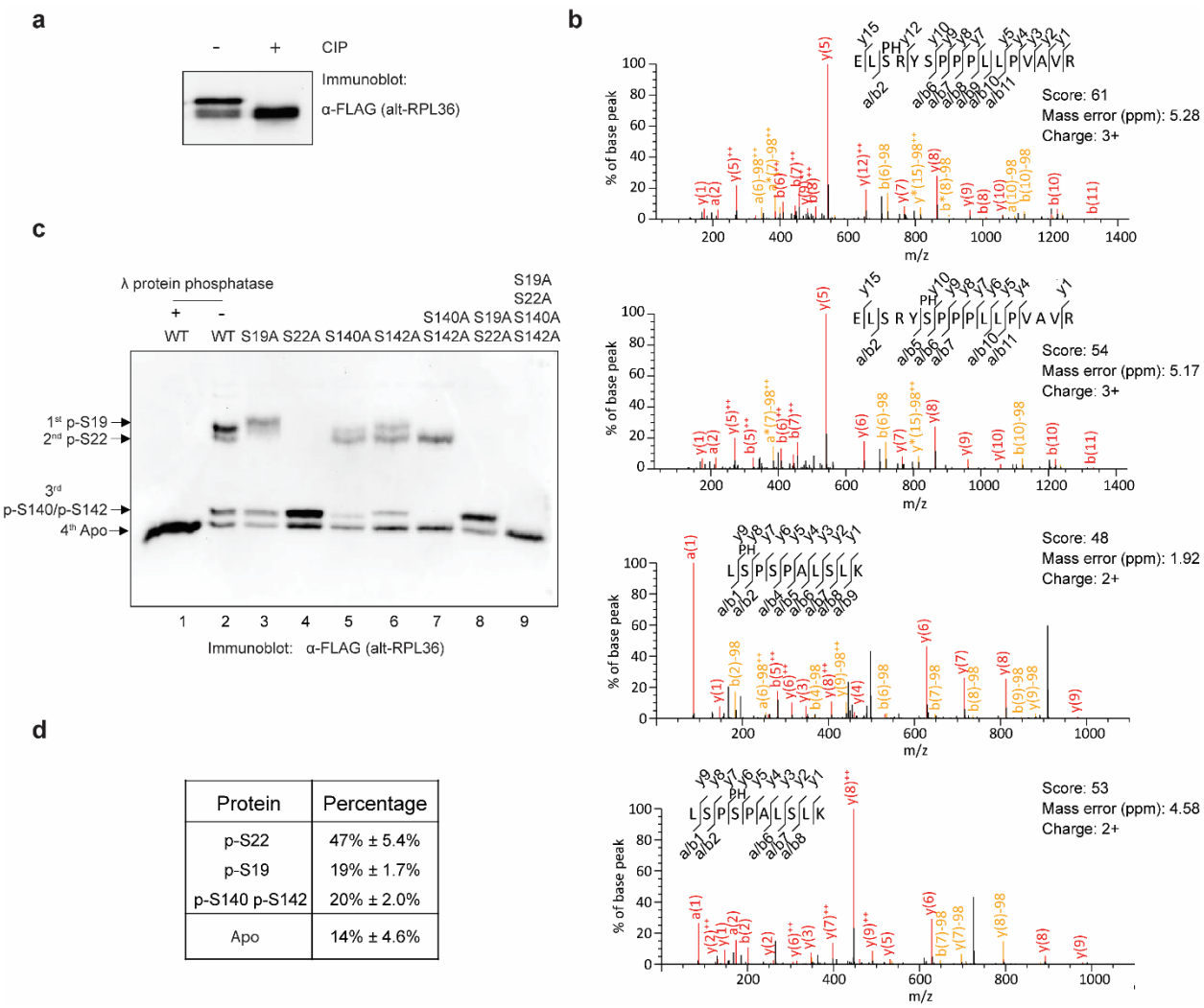


Figure 2 | Alt-RPL36 partially localizes to the endoplasmic reticulum membrane and the plasma membrane. (**a**) Expression of alt-RPL36 variants via genetic code expansion. HEK 293 cells were transfected with transgenes encoding *M. mazei* pyrrolysyl-tRNA synthetase bearing Y306A/Y384F mutations (PylRS-AF), amber-suppressing Pyl tRNA ($PylT_{CUA}$), and variants of amber codon (TAG)-bearing, c-myc-tagged alt-RPL36. Cells were incubated in the presence or absence of 60 μM bicyclononyne-lysine (BCNK) for 45 hours at 37 ºC, lysed, and analyzed by Western blotting with anti-c-myc antibody. Controls with untransfected cells and cells expressing wild-type alt-RPL36 are shown. Data are representative of two independent biological replicates. (**b**) Visualizing alt-$RPL36^{L18TAG}$ expression via genetic code expansion. HEK 293 cells were transfected with transgenes encoding PylRS-AF, $PylT_{CUA}$, and c-myc-tagged alt-$RPL36^{L18TAG}$ expressing plasmids. Cells were incubated with 60 μM BCNK for 45 hours, fixed, labeled with 500 nM tetrazine-cy5 (purple) for 30 min, and immunostained with anti-myc (red), anti-calreticulin (green, an ER marker) and DAPI (blue). Scale bar, 5 μm. Data are representative of three independent biological replicates; three fields of view from the same experiment are shown. White arrows indicate plasma membrane localization of alt-RPL36(BCNK). (**c**) APEX fingerprinting to

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capture proximal proteome environments of alt-RPL36. HEK cells expressing APEX2 fused to alt-RPL36 or subcellular APEX2 controls were labeled with biotin-phenol for 1min, then lysed. Biotinylated proteins were detected via streptavidin blot (top), while anti-V5 blot detected APEX2 fusion protein expression (bottom). APEX2 controls were targeted to the following subcellular locations: cytosolic face of the endoplasmic reticulum (ERM); ER lumen; nucleus (NLS); and cytosol (NES). Data are representative of two independent biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Protein	Percentage
p-S22	47% ± 5.4%
p-S19	19% ± 1.7%
p-S140 p-S142	20% ± 2.0%
Apo	14% ± 4.6%

Figure 3 | Four serine residues of alt-RPL36 are phosphorylated. (**a**) Western blot analysis of immunoprecipitated alt-RPL36-Flag protein treated with or without calf intestinal alkaline phosphatase (CIP) and run on an SDS-PAGE gel. Data are representative of three biological replicates. (**b**) MS/MS spectra of alt-RPL36 phosphopeptides. S(PH) denotes phosphoserine. See Supporting Table 3 for complete phosphoproteomics data. (**c**) To confirm phosphoserine assignments from MS/MS, wild-type alt-RPL36-FLAG or alanine point mutants were immunopurified from HEK 293T cells, resolved with Phos-tag SDS-PAGE, and detected with anti-FLAG Western blotting. A sample of wild-type alt-RPL36-FLAG was treated with λ protein phosphatase before Phos-tag SDS-PAGE (lane 1). (**d**) Quantitative analysis of the Western blot signal of wild-type alt-RPL36-Flag. Data represent mean values ± standard error of the mean (s.e.m.) of three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

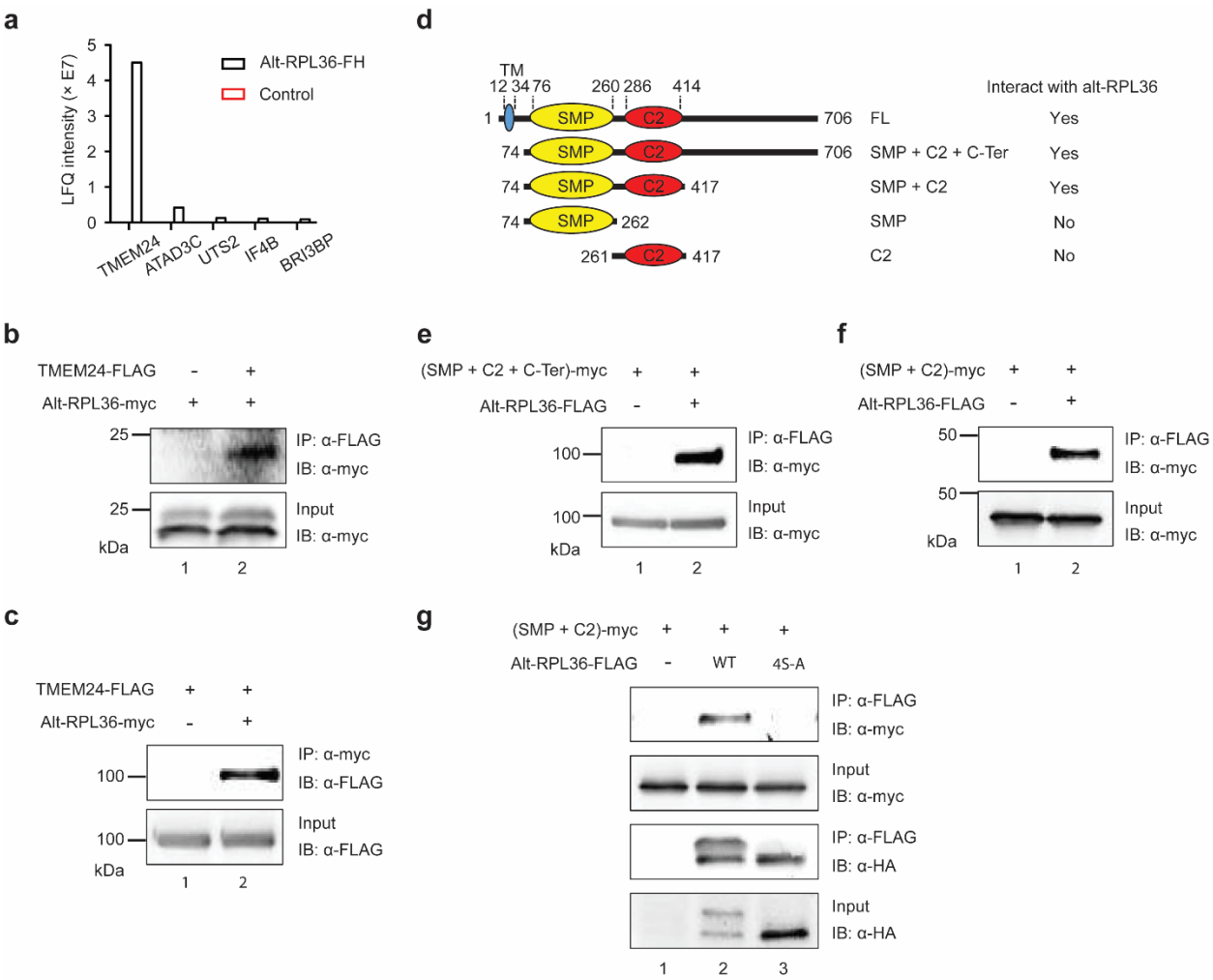


Figure 4 | Phosphorylated alt-RPL36 interacts with TMEM24 via the SMP and C2 domains. (**a**) Quantitative proteomics (n = 1) of alt-RPL36-Flag-HA (Alt-RPL36-FH) immunoprecipitates from HEK 293T lysates identified putative alt-RPL36 interaction partners enriched over untransfected HEK 293T cells (control). For complete quantitative proteomics results, see Supporting Table 4. (**b** and **c**) HEK 293T cells were transfected with alt-RPL36-myc only (**b**, lane 1), TMEM24-Flag only (**c**, lane 1) or both plasmids (**b** and **c**, lane 2), and immunoprecipitation (IP) was performed with anti-Flag (**b**) or anti-myc (**c**) antibody, followed by immunoblotting (IB) with antibodies indicated to the right. Cell lysates (1%) before IP (input) were used as the loading controls. (**d**) Schematic representation of the domain structures of the wild-type and truncation mutants of TMEM24, with amino acid residue numbers above. Alt-RPL36 interaction status of each construct is listed to the right. (**e** and **f**) Control HEK 293T cells (lane 1) or HEK 293T cells stably expressing alt-RPL36-Flag-HA (lane 2) were transfected with TMEM24 truncation mutants, and IPs were performed with anti-Flag antibody followed by IB with anti-myc. (**g**) Control HEK 293T cells (lane 1), HEK 293T cells stably expressing wild-type

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alt-RPL36-Flag-HA (lane 2), and HEK 293T cells stably expressing non-phosphorylatable 4S-A mutant alt-RPL36-Flag-HA (lane 3) were transfected with the TMEM24 SMP + C2 domain construct, followed by IP with anti-Flag antibody and immunoblotting with antibodies indicated on the right. Cell lysates (1%) before IP (input) were used as the loading controls. All Western blots are representative of three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

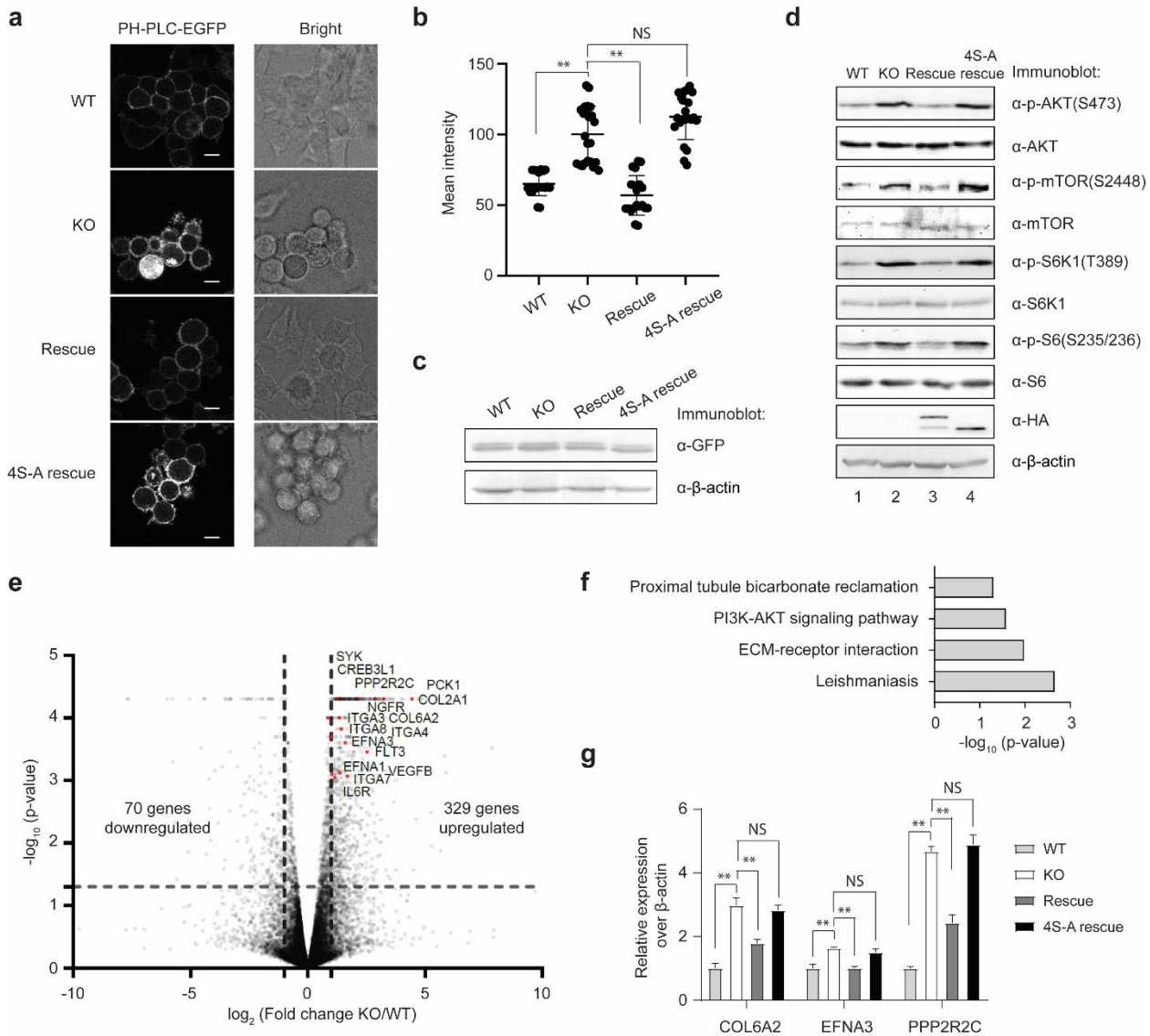


Figure 5 | Phosphorylated alt-RPL36 regulates the PI3K-AKT-mTOR signaling pathway. (**a**) Confocal imaging of wild-type (WT), alt-RPL36 knock-out (KO), rescue with wild-type alt-RPL36 (Rescue) or rescue with nonphosphorylatable 4S-A mutant alt-RPL36 (4S-A rescue) HEK 293T cells stably expressing PH-PLC-GFP. Scale bar, 10 µm. (**b**) Quantitation of the PH-PLC-GFP signals in the four cell lines described above. At least 20 fields of view were analyzed, totaling > 400 cells for each measurement. Significance was evaluated with two-tailed *t*-test. ** $p < 0.01$. NS, not significant. (**c**) Western blot of the four cell lines with antibodies indicated on the right for comparison of PH-PLC-EGFP expression. Data are representative of three biological replicates. (**d**) Western blot analysis of the wild-type (WT), alt-RPL36 knock-out (KO), rescue with wild-type alt-RPL36-Flag-HA (Rescue) or rescue with 4S-A mutated alt-RPL36-Flag-HA

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(4S-A rescue) HEK 293T cells with antibodies indicated on the right. Data are representative of three biological replicates. (**e**) Volcano plot of mRNA-seq from wild-type (WT) and alt-RPL36 knock-out (KO) HEK 293T cells. Upregulated transcriptional targets of the PI3K-AKT-mTOR signaling pathway identified via KEGG analysis are indicated in red and gene names are labeled. (**f**) KEGG analysis of genes upregulated upon alt-RPL36 knock-out; enriched biological pathways are plotted by significance (Fisher's exact test). (**g**) Quantitative RT-PCR results of the mRNA transcripts of the wild-type (WT), alt-RPL36 knock-out (KO), rescue with wild-type alt-RPL36-Flag-HA (Rescue) or rescue with 4S-A mutated alt-RPL36-Flag-HA (4S-A rescue) HEK 293T cells. Data represent mean values ± standard error of the mean (s.e.m.) of four biological replicates. Significance was evaluated with two-tailed *t*-test. ** $p < 0.01$. NS, not significant.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

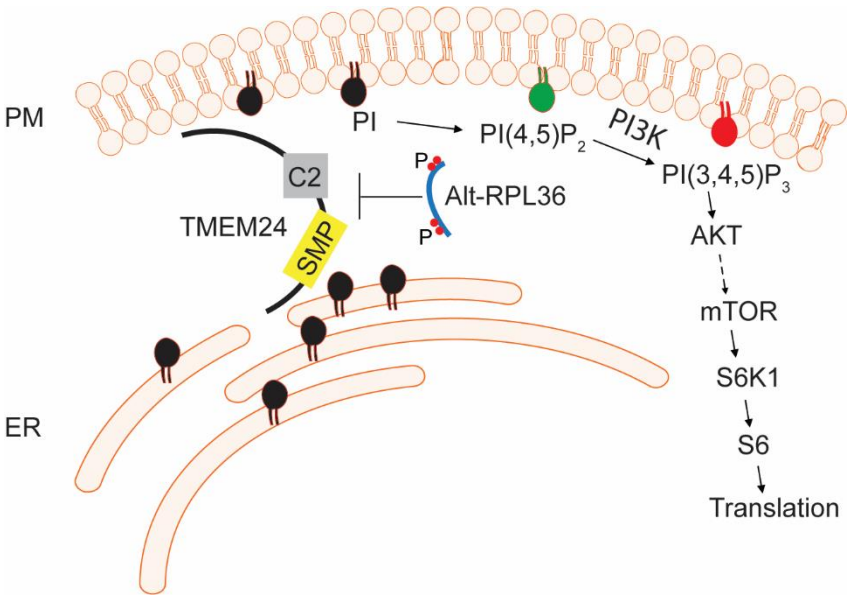


Figure 6 | Model of alt-RPL36 regulatory pathway. Phosphorylated (P) alt-RPL36 interacts with TMEM24 via the SMP and C2 domains to negatively regulates TMEM24-dependent phosphatidylinositol 4,5-bisphosphate [PI(4,5)P_2] precursor phosphatidylinositol (PI) transport. In cells depleted of alt-RPL36, more PI is transported to the plasma membrane (PM) from the endoplasmic reticulum (ER), where it is synthesized. At the PM, PI is converted to PI(4,5)P_2 then phosphorylated by PI3K to produce PI(3,4,5)P_3. PI(3,4,5)P_3 then activates AKT which, indirectly via inhibition of TSC1/2, activates downstream mTOR signaling. mTOR phosphorylates S6K1, and activated S6K1 phosphorylates S6, finally regulating protein translation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-23 (2009).
2. Martinez, T.F. et al. Accurate annotation of human protein-coding small open reading frames. *Nat Chem Biol* (2019).
3. Slavoff, S.A. et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* **9**, 59-64 (2013).
4. Lee, S. et al. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A* **109**, E2424-32 (2012).
5. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).
6. Saghatelian, A. & Couso, J.P. Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat Chem Biol* **11**, 909-16 (2015).
7. Couso, J.P. & Patraquim, P. Classification and function of small open reading frames. *Nat Rev Mol Cell Biol* **18**, 575-589 (2017).
8. D'Lima, N.G. et al. A human microprotein that interacts with the mRNA decapping complex. *Nat Chem Biol* **13**, 174-180 (2017).
9. Chu, Q. et al. Regulation of the ER stress response by a mitochondrial microprotein. *Nat Commun* **10**, 4883 (2019).
10. Rathore, A. et al. MIEF1 Microprotein Regulates Mitochondrial Translation. *Biochemistry* **57**, 5564-5575 (2018).
11. Jackson, R. et al. The translation of non-canonical open reading frames controls mucosal immunity. *Nature* **564**, 434-438 (2018).
12. Anderson, D.M. et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595-606 (2015).
13. Wool, I.G., Chan, Y.L. & Gluck, A. Structure and evolution of mammalian ribosomal proteins. *Biochem Cell Biol* **73**, 933-47 (1995).
14. Zhu, Y. et al. Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow. *Nat Commun* **9**, 903 (2018).
15. Lees, J.A. et al. Lipid transport by TMEM24 at ER-plasma membrane contacts regulates pulsatile insulin secretion. *Science* **355**(2017).
16. Fruman, D.A. et al. The PI3K Pathway in Human Disease. *Cell* **170**, 605-635 (2017).
17. Manning, B.D. & Toker, A. AKT/PKB Signaling: Navigating the Network. *Cell* **169**, 381-405 (2017).
18. Saxton, R.A. & Sabatini, D.M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960-976 (2017).
19. Sun, E.W. et al. Lipid transporter TMEM24/C2CD2L is a Ca(2+)-regulated component of ER-plasma membrane contacts in mammalian neurons. *Proc Natl Acad Sci U S A* **116**, 5775-5784 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Pottekat, A. et al. Insulin biosynthetic interaction network component, TMEM24, facilitates insulin reserve pool release. *Cell Rep* **4**, 921-30 (2013).
21. Kearse, M.G. & Wilusz, J.E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev* **31**, 1717-1731 (2017).
22. Peabody, D.S. Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031-5 (1989).
23. Tate, J.G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947 (2019).
24. D'Lima, N.G. et al. A human microprotein that interacts with the mRNA decapping complex. *Nature Chemical Biology* **13**, 174-180 (2017).
25. Slavoff, S.A., Heo, J., Budnik, B.A., Hanakahi, L.A. & Saghatelian, A. A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *J Biol Chem* **289**, 10950-7 (2014).
26. Uttamapinant, C. et al. Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. *J Am Chem Soc* **137**, 4602-5 (2015).
27. Blackman, M.L., Royzen, M. & Fox, J.M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J Am Chem Soc* **130**, 13518-9 (2008).
28. Lee, S.Y. et al. APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell Rep* **15**, 1837-47 (2016).
29. Wegener, A.D. & Jones, L.R. Phosphorylation-induced mobility shift in phospholamban in sodium dodecyl sulfate-polyacrylamide gels. Evidence for a protein structure consisting of multiple identical phosphorylatable subunits. *J Biol Chem* **259**, 1834-41 (1984).
30. Orr, M.W., Mao, Y., Storz, G. & Qian, S.B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* (2019).
31. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* **10**, 730-6 (2013).
32. Chu, Q. et al. Identification of Microprotein-Protein Interactions via APEX Tagging. *Biochemistry* **56**, 3299-3306 (2017).
33. Varnai, P. & Balla, T. Visualization of phosphoinositides that bind pleckstrin homology domains: calcium- and agonist-induced dynamic changes and relationship to myo-[3H]inositol-labeled phosphoinositide pools. *J Cell Biol* **143**, 501-10 (1998).
34. Lien, E.C., Dibble, C.C. & Toker, A. PI3K signaling in cancer: beyond AKT. *Curr Opin Cell Biol* **45**, 62-71 (2017).
35. Wang, W., Eddy, R. & Condeelis, J. The cofilin pathway in breast cancer invasion and metastasis. *Nat Rev Cancer* **7**, 429-40 (2007).
36. Sanz-Moreno, V. et al. Rac activation and inactivation control plasticity of tumor cell movement. *Cell* **135**, 510-23 (2008).
37. Brar, G.A. et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-7 (2012).
38. Calvo, S.E., Pagliarini, D.J. & Mootha, V.K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* **106**, 7507-12 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.977314; this version posted March 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Wei, J., Zhang, Y., Ivanov, I.P. & Sachs, M.S. The stringency of start codon selection in the filamentous fungus Neurospora crassa. *J Biol Chem* **288**, 9549-62 (2013).
40. Delcourt, V. et al. The Protein Coded by a Short Open Reading Frame, Not by the Annotated Coding Sequence, Is the Main Gene Product of the Dual-Coding Gene MIEF1. *Mol Cell Proteomics* **17**, 2402-2411 (2018).
41. Longo, P.A., Kavran, J.M., Kim, M.S. & Leahy, D.J. Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol* **529**, 227-40 (2013).
42. Tiscornia, G., Singer, O. & Verma, I.M. Production and purification of lentiviral vectors. *Nat Protoc* **1**, 241-5 (2006).
43. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019).