Programmed ribosomal frameshifting during *PLEKHM2* mRNA decoding generates a constitutively active mediator of kinesin-1-dependent lysosome transport

Yousuf A. Khan[1,2,3,4], Raffaella De Pace[5], Irwin Jungreis[6,7], Gionmattia Carancini[8], Jonathan M. Mudge[9], Ji Wang[10,11], Manolis Kellis[6,7], John F. Atkins[8], Pavel V. Baranov[8], Andrew E. Firth[10], Juan S. Bonifacino[5], Gary Loughran[8*]*

[1]Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA.

[2]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA

[3]Department of Structural Biology, Stanford University, Stanford, CA, USA

[4]Department of Photon Science, Stanford University, Stanford, CA, USA

[5]Neurosciences and Cellular and Structural Biology Division, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA.

[6]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA.

[7]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[8]School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.

[9]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton CB10 1SD, Cambridge, UK.

[10]Department of Pathology, University of Cambridge, Cambridge, UK.

[11]Current affiliation: Institute of Precision Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, Republic of China.

*Corresponding authors: yousuf@stanford.edu, g.loughran@ucc.ie

Abstract

Programmed ribosomal frameshifting is a translational recoding phenomenon in which a proportion of ribosomes are stimulated to slip backwards or forwards on an mRNA[1], rephasing the ribosome relative to the mRNA. While frameshifting is often employed by viruses[2], very few phylogenetically conserved examples are known in vertebrate genes and the evidence for some of these is controversial[3,4]. Here we report a +1 frameshifting signal in the coding sequence of the human gene *PLEKHM2*, encoding the ARL8-dependent, lysosome–kinesin-1 adaptor protein PLEKHM2[5]. This +1 frameshifting signal, UCC_UUU_CGG, is highly conserved in vertebrates and exhibits an influenza virus-like frameshift motif with similar efficiency[6,7]. Purification and mass spectrometry of GFP-tagged trans-frame protein from cells confirms frameshifting. Structure prediction shows that the new C-terminal domain generated by this frameshift forms an alpha-helix. This additional domain relieves PLEKHM2 from autoinhibition, allowing it to move to the tips of cells via association with kinesin-1 without requiring activation by ARL8. Thus, the frameshift proteoform generates a constitutively active adaptor of kinesin-1.

Main

Ribosomes catalyze the synthesis of proteins, a tightly coordinated, regulated, and conserved process[8]. During elongation, each codon of an mRNA is matched to its amino acid-specific tRNA. The nascent peptide chain is then constructed in a sequential manner as tRNAs continue to deliver amino acids[9]. Maintenance of the reading frame is critical to ensure that proteins are correctly synthesized. A network of interactions between the ribosome, tRNAs, and other factors help to maintain translation fidelity[10]. Thus, the rate of spontaneous frameshifting, where the ribosome shifts reading frame during elongation, is exceptionally low[11].

Programmed ribosomal frameshifting (PRF) occurs when a proportion of elongating ribosomes are induced to shift their reading frame by +1, −1, or −2 nucleotides upon encountering specific signals in *cis*- or a *trans*-acting factor[1]. *Cis*-acting PRF signals may include a slippery site where tRNAs can re-pair in an alternative reading frame, a 3′ RNA-element, or a nascent peptide that can interact with the ribosome's peptide exit tunnel. *Trans*-acting stimulators of PRF include proteins[12,13] or small molecules such as polyamines[14]. Except where a stop codon is immediately encountered, PRF results in ribosomes synthesizing a distinct polypeptide from sequence 3′ of the PRF site. Several known PRF sites are 5' adjacent to an original frame stop codon resulting in synthesis of truncated and extended proteoforms with identical N-termini. In some instances, no termination codon exists adjacent to the PRF slippery site in either reading frame, leading to the synthesis of a "transframe" protein in addition to the product of standard translation. Many viruses, including SARS-CoV-2[15] and HIV-1[16], and virally-derived sequences require PRF to regulate their gene expression . In several viruses, transframe proteins are synthesized from overlapping open reading frames (ORFs)[2]. In vertebrates, there are few known examples of PRF in cellular genes. Most are retrovirally-derived −1 PRF mRNAs[17,18]. Reported cases of non-retrovirus derived −1 PRF in vertebrate genes were shown to be artefacts[3,4]. The gene for antizyme - a negative regulator of cellular

polyamine levels - requires +1 PRF to translate and synthesize a single antizyme product as part of an autoregulatory circuit where the efficiency of PRF increases in response to elevated polyamine levels[14]. While examples of translation from long overlapping reading frames in vertebrates are known[19,20], these result from leaky scanning or alternative splicing rather than PRF. An influenza virus-like +1 PRF site has been bioinformatically predicted in the vertebrate gene *ASXL1*, though this has not been experimentally validated[21]. To date, there are no experimentally confirmed PRF signals in vertebrate cellular genes that provide ribosome access to internally overlapping ORFs.

Here, we report a highly conserved +1 PRF signal and overlapping reading frame in the mRNA encoding PLEKHM2 (also known as SKIP) that mediates ARL8-dependent coupling of lysosomes to the anterograde microtubule motor kinesin-1[22]. PRF results in the synthesis of a transframe protein, PLEKHM2-FS. PLEKHM2-FS does not contain the autoinhibitory C-terminal domain of PLEKHM2-WT and instead is predicted to contain an α-helical domain that we found to promote self-association. This results in a proteoform that does not require activation by ARL8 to promote kinesin-1-driven lysosome transport.

***PLEKHM2* contains a highly conserved +1 PRF signal and overlapping ORF**

During bioinformatic scans for alternative ORFs, we noticed a long, conserved overlapping ORF in one of the coding exons of *PLEKHM2* (**Fig. 1a**). Our extensive surveying of transcriptomics datasets did not provide evidence that this overlapping frame can be accessed by alternative splicing or transcriptional initiation. Towards, the 5′ end of this alternative ORF is a potential +1 PRF cassette, UCC_UUU_CGG, almost identical to the influenza A virus UCC_UUU_CGU +1 PRF cassette, that is conserved in vertebrate *PLEKHM2* mRNAs (**Fig. 1b**). For influenza A virus PRF, slippage is thought to occur when UUU is positioned in the ribosomal P-site and CGU is positioned in a presumably empty A-site[6]. Both UUU and UUC are decoded by the same tRNA isoacceptor whose anticodon, 3′-AAG-5′, has a higher affinity for UUC in the +1 frame than for the zero-frame UUU[23]. Subjecting *PLEKHM2* transcript sequence alignments to both Synplot2[24] and MLOGD[25] analysis, we found that, in mammals, sauropsids, amphibians and teleost fish, in the region of the *PLEKHM2* gene that is overlapped by the *FS* region, there is statistically significantly enhanced synonymous site conservation, and a positive coding signature specifically in the +1 reading frame (**Fig. 1c-f**).

PhyloCSF, an algorithm used to predict coding genomic regions, measures protein coding potential by computing the log-likelihood of a multispecies alignment under coding and noncoding models of evolution[26]. The PhyloCSF tracks in the UCSC Genome Browser indicate a conserved change of reading frame in exon 9 of *PLEKHM2* followed by coding resumption in the original reading frame downstream of the *FS* stop codon, in both mammals and birds (**Fig 2a,b**), suggesting a potential PRF[27]. A whole genome alignment of 100 vertebrate species rendered using CodAlignView (https://data.broadinstitute.org/compbio1/cav.php) shows that (i) both the CCUUUC 6-mer in the key portion of the potential PRF cassette and the *FS* stop codon are present in all aligned species, (ii) there is a preponderance of synonymous substitutions

relative to the +1 frame in the latter portion of the potential dual coding region, indicating purifying selection on the amino acid sequence translated from that frame, and (iii) there are no stop codons in the +1 frame within the potential dual coding region despite many stop codons in the flanking sequences on both sides of the aligned sequences (**S1a-d**).

Given that this overlapping ORF is highly conserved, we were curious if it encoded a protein with a high-confidence predicted structure. Using ColabFold's[28] implementation of AlphaFold2[29], and assuming PRF on UCC_UUU_CGG, we input the sequence of the predicted FS protein. While ColabFold reliably predicted the N-terminal portion of the FS protein (**Fig. 2c**), it was unable to predict a structure from the portion of the protein encoded by the alternative ORF. However, given that AlphaFold2 relies on multiple sequence alignments (MSAs) from available databases to extract critical coevolutionary information, we reasoned that it would not have *a priori* knowledge on PLEKHM2-FS since its C-terminal sequence is absent from all existing databases. We generated custom MSAs for PLEKHM2-FS across all vertebrates, injected them into the model and were able to predict a high confidence α-helical domain in the transframe portion of the protein, further suggesting that this was a bona fide +1 PRF signal that enabled expression of a functional transframe protein.

Experimental evidence of +1 PRF in *PLEKHM2*

Possible PRF into the +1 reading frame was tested experimentally by inserting the putative frameshift site and flanking sequence from *PLEKHM2* between *Renilla* and *firefly* luciferase ORFs such that a +1 PRF event would give rise to a ~100 kDa fusion product of the two reporter polypeptides[30,31]. These constructs were transfected into HEK293T cells and cell lysates were immunoblotted with anti-*Renilla* antibody. This demonstrated PRF in a WT construct, but not upon mutation of the putative PRF cassette from UCC_UUU_CGG to agC_UUc_aGa (**Fig. 3a**).

To measure the PRF efficiency of *PLEKHM2*, we used a dual luciferase reporter system that we previously described[32]. Briefly, the test sequence is flanked by tandem StopGo sequences from foot-and-mouth disease virus[33] that, while allowing continued protein synthesis, prevents peptide bond formation at a specific site in its sequence, resulting in expression of the reporter as separate proteins (**Fig. 3b**). This arrangement avoids potential artefacts that can arise when the test sequence fusion alters the individual reporter activities or stabilities[3,4,32]. Furthermore, we changed a +1 frame AUG within the *PLEKHM2* cassette to minimize the possibility of firefly translation due to initiation on potential low abundance mRNAs that may be generated by cryptic splicing or cryptic promoters. Dual luciferase assays on lysates prepared from HEK293T cells transfected with *PLEKHM2* constructs indicated a PRF efficiency of ~1.3% for WT and ~1.1% for WT minus AUG (**Fig. 3b** and **Ext 1a,b**). The slight decrease in firefly activity upon mutation of the +1 frame AUG indicates some level of internal initiation on this codon and further militates for controlling against such artefacts.

Although +1 PRF by mammalian ribosomes on influenza virus-like shift sites, such as in influenza A virus and *PLEKHM2*, do not appear to require *cis*-acting stimulators, we noticed a

potential stem-loop 3′ of the *PLEKHM2* shift site that appears to display sequence conservation (**Fig. S2a**). To test the possible stimulatory role of this putative stem-loop structure, we introduced mutations to disrupt (delta SL), restore (restore SL) or strengthen (strong SL) the stem-duplex (**Fig. S2b**). Disrupting the putative stem-loop structure decreased *PLEKHM2* PRF efficiency to ~0.75%. However, WT levels of PRF were not restored by changes that were predicted to restore stem-loop formation or strengthen the stem-loop (**Fig. 3b** and **Ext 1a,b**). Although +1 PRF on the *PLEKHM2* mRNA does not appear to require the predicted stem-loop RNA secondary structure, there may be a weak stimulatory effect. It is possible that in certain cell types or under certain conditions the RNA secondary structure may modulate +1 PRF in *PLEKHM2* expression.

Polyamines stimulate the translation of the full antizyme 1 protein (encoded by *OAZ1*) by using +1 PRF to bypass an early stop codon in the zero-frame[34], a mechanism conserved from protists to mammals. Therefore, we tested whether +1 PRF in the *PLEKHM2* gene may also be stimulated by polyamines, since polyamines have been proposed to stimulate +1 PRF in a sequence-independent manner[35,36]. HEK293T cells transfected with *PLKEHM2* or *OAZ1* dual luciferase constructs were treated with spermidine and assayed for PRF efficiency. Although baseline *OAZ1* PRF (32.5%) is much higher than *PLEKHM2* (~1 %), we observed a similar 1.5-fold increase in PRF efficiency for both *OAZ1* and *PLEKHM2* when spermidine was added (**Fig. 3c** and **Ext. 1c,d**). This indicates that *PLEKHM2* PRF efficiency can be regulated. To assess whether *PLEKHM2* PRF may be more efficient in other cells or tissues, we compared ribosome read densities extracted from publicly available ribosome profiling data, upstream and downstream of the PRF casette in the *PLEKHM2* mRNA. Although subtle differences in read densities were observed among different tissues, the baseline rate of PRF is too low to determine if these differences are significant (**Fig. S3**).

To determine the precise site and direction of PRF, we aimed to identify transframe peptides by mass spectrometry. We generated a construct in which an ORF encoding green fluorescent protein (GFP) was fused in-frame to the 3′ end of the transframe ORF (**Fig. 3d,e**). An additional construct (PLEKHM2-K-GFP) introducing a lysine codon close to the slip site was used to increase the chances of identifying the junction tryptic peptide (**Ext. 1e**). Synonymous mutations in the zero-frame were introduced to alter two +1 frame AUG codons to avoid potential initiation in the +1 frame. Frameshift expression from these constructs would result in the transframe fusion PLEKHM2-GFP or PLEKHM2-K-GFP, which could be affinity-purified on GFP-Trap beads, while non-frameshift expression would result in products that do not contain GFP. Constructs were expressed in HEK293T cells, and transframe fusions were affinity-purified from cell lysates and resolved by SDS-PAGE. Both constructs directed synthesis of a specific protein migrating at the predicted molecular weight for PLEKHM2-FS-GFP (~34 kDa). Gel slices containing these proteins were excised, digested with trypsin, and the resulting peptides were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS). Tryptic peptides covering 59% of the PLEKHM2-FS-GFP construct and 85% of the PLEKHM2-K-FS-GFP construct were identified, including peptides encoded both upstream and downstream

of the shift site (**Fig. 3f and Ext. 1e**). Importantly, three peptides spanning the shift site itself were identified. These peptides, AEPPDQSFGPALPGTPR (from PLEKHM2-FS-GFP), QSFGPALPGTPR and AEPPKQSFGPALPGTPR with a missed tryptic cleavage (from PLEKHM2-K-FS-GFP), define the shift site (UCC_UUU_CGG) and direction (+1) of PRF.

The PLEKHM2-FS proteoform is constitutively active

Next, we investigated the activity of the PLEKHM2-FS proteoform. PLEKHM2[37] comprises a RUN domain that interacts with the lysosome-associated small GTPase ARL8[22], WD-WE motifs that interact with kinesin-1[38], a largely disordered region, and three autoinhibitory PH domains[5] (**Fig. 4a**). ARL8 both recruits PLEKHM2 to lysosomes[22] and relieves PLEKHM2 autoinhibition[5]. These properties enable PLEKHM2 to function as a regulated adaptor for ARL8-dependent coupling of lysosomes to kinesin-1 and consequent transport toward the plus end of microtubules[5,38].

The PRF that generates PLEKHM2-FS preserves the RUN domain and WD-WE motifs but parts of the disordered region and the three PH domains are absent (**Fig. 4a**). As previously shown[5,38], overexpression of HA-tagged PLEKHM2 (HA-PLEKHM2) in WT HeLa cells (**Ext. 2a**) results in the redistribution of a large fraction of lysosomes immunolabeled for the lysosomal markers LAMP1 (**Fig. 4b,4c** and **Ext. 3a**) and LAMTOR4 (**Ext. 2b**) to cell tips (also referred to as vertices; indicated by yellow arrows). Overexpression of HA-PLEKHM2-FS or a truncated HA-PLEKHM2 1-437 lacking the three PH domains and the FS sequence (**Fig. 4a, Ext. 2a**) similarly redistributed lysosomes to cell tips (**Fig. 4b, 4c, Ext. 2b, 3a**). All these HA-PLEKHM2 variants themselves redistributed to cell tips along with lysosomes. PLEKHM2–lysosome clusters exhibited a tighter appearance and were occasionally observed to be more centrally located for the HA-PLEKHM2-FS and HA-PLEKHM2-1-437 constructs compared to the HA-PLEKHM2 construct (**Fig. 4b, Ext. 2b, 3a**; centrally located clusters are indicated by magenta arrows), a phenotype that likely stems from the removal of the C-terminal sequences. These observations thus demonstrated that PLEKHM2-FS is as active as PLEKHM2, and that this activity does not depend on the FS sequence, when overexpressed in WT cells.

Knock out (KO) of both the ARL8A and ARL8B paralogs in HeLa cells (referred to as ARL8 KO) prevented the association of HA-PLEKHM2 with lysosomes (due to loss of ARL8-dependent recruitment) and kinesin-1 (due to loss of ARL8-dependent relief from autoinhibition), thereby impeding the overexpression-induced redistribution of both lysosomes and PLEKHM2 to cell tips (**Fig. 4c, 4d, Ext. 2a**). Importantly, although overexpressed HA-PLEKHM2-FS failed to redistribute lysosomes, it could itself relocate to the tips of ARL8-KO cells (**Fig. 4c, 4d, Ext. 2c**, green arrows). This latter phenotype differs from that observed for HA-PLEKHM2 and is likely attributable to the absence of the autoinhibitory PH domains in PLEKHM2-FS. These observations indicated that PLEKHM2-FS is constitutively active, requiring no activation by ARL8 for association with kinesin-1 and movement toward cell tips.

In ARL8-KO cells, the truncated HA-PLEKHM2-1-437 exhibited a more cytosolic distribution (**Fig. 4d, Ext 2c, S4a, S4b**) and substantially reduced movement to cell tips

compared to PLEKHM2-FS (**Fig. 4c, 4d, Ext. 2c, 3b, S4a, S4b**). These observations suggest that the FS sequence does not just replace an autoinhibitory sequence but also enhances the interaction with kinesin-1. Co-immunoprecipitation of HA- and myc-tagged constructs showed that PLEKHM2-FS interacted with itself much more than with PLEKHM2 or PLEKHM2-1-437 (**Ext. 2d**). Thus, the FS sequence likely promotes self-association, a property that may contribute to its enhanced ability to undergo kinesin-1–driven movement toward the cell tips.

We conclude that PLEKHM2-FS is a constitutively active variant of PLEKHM2 that does not require ARL8-dependent activation for kinesin-1–dependent movement toward the cell periphery (**Fig. 4e**). How the presence of this constitutively active form of PLEKHM2 affects the dynamics of kinesin-1-driven lysosomal distribution within cells remains an open question. PLEKHM2-FS could provide a basal level of lysosome transport, while PLEKHM2 would enable another level of regulated lysosome transport in response to stimuli such as starvation[39] or growth factor receptor activation[40]. Furthermore, PLEKHM2-FS may not just be constitutively activated, but could also play a role in activating the WT proteoform. The function of PLEKHM2-FS is likely critical in vertebrates, given its extensive conservation in all the species we were able to query.

The ability of polyamines to similarly stimulate both *OAZ1* and *PLEKHM2* +1 PRF lends additional evidence to significant latitude in sequences at which polyamines stimulate +1 PRF.While polyamine stimulation of *OAZ1* PRF has evolved to maintain polyamine levels within a narrow cellular concentration, it is possible that polyamine stimulation of *PLEKHM2* PRF could be incidental - perhaps polyamines directly or indirectly affect elongation rates that is especially seen in PRF events.

The discovery of the first +1 PRF signal in vertebrate cellular genes that provides ribosome access to internally overlapping ORFs thus marks a major milestone and sets the stage for further searches in vertebrate genomes. Other conserved PRF signals may exist but may not have been as highly selected for or evolved as early as the one in *PLEKHM2*, making them less readily identifiable through traditional bioinformatic searches. However, the increasing number of viral and cellular examples of PRF signals may lend itself to complementary computational techniques, such as machine learning, and improved high-throughput assays.

Figure Legends

Figure 1 | Synonymous site conservation and coding potential within the *PLEKHM2* dual coding region. a. Schematic of the main *PLEKHM2* transcript. Pink-dashed lines indicate exon-exon junctions and black bars represent untranslated regions (UTRs). The predicted PRF cassette is indicated with an arrow. The pink rectangle represents the +1 frame, conserved ORF. **b.** Sequence logo of putative PRF cassette, generated from alignment of *PLEKHM2* mRNA sequences from mammals, sauropsids, amphibians and teleost fish. **c-f.** Synplot2 and MLOGD plots. Top schematic shows stop codons (blue) in aligned sequences in +1 frame. Next row

shows stop codons in +2 frame. O/E plot shows ratio of observed to expected substitutions under a null model of neutral evolution at synonymous sites and p-val plot shows corresponding p-value. Bottom row shows MLOGD coding potential plot for all three frames (purple 0, blue +1, yellow +2). **c.** Generated plots for mammals. **d.** Generated plots for sauropsids. **e.** Generated plots for amphibians. **f.** Generated plots for teleost fish.

Figure 2 | PhyloCSF and AlphaFold2 predictions on PLEKHM2. a-b. UCSC Genome Browser tracks showing evolutionary signature of protein coding potential in each of three reading frames as measured by PhyloCSF in mammals (**a**) and birds (**b**). Part way through the dual coding region (blue) between the PRF cassette (green arrow) and the +1 frame stop codon (red arrow), the PhyloCSF signal changes to the +1 frame, indicating greater amino acid conservation in the +1 frame, then resumes in the zero frame downstream of the +1 frame stop codon. **c.** AlphaFold2 predictions of PLEKHM2-FS using the standard workflow (left) and by generating MSAs with trans-frame sequence (right). pLDDT coloring scheme is shown in the bottom right.

Figure 3 | PRF on *PLEKHM2* mRNA. a. Representative immunoblot of whole cell lysates prepared from HEK293T cells transfected with *PLEKHM2* dual luciferase constructs. WT: wild-type (UCC_UUU_CGG), SS: frameshift mutant (agC_UUc_aGa), IFC: in-frame control (agC_UUc_Ga) with both reporters in the zero-reading frame, MOCK: no plasmid transfected. **b.** PRF efficiencies (%) of *PLEKHM2* determined by dual luciferase assays. WT: wild-type (UCC_UUU _CGG), SS: frameshift mutant (agC_UUc_aGa). Change of +1 frame AUG to ACG labeled as '-AUG'. Delta SL, restore SL and strong SL changes are depicted in Supplemental Data Fig. 2b. $n = 4$. **c.** PRF efficiencies for *PLEKHM2* and *OAZ1* PRF cassettes upon addition of 2 mM spermidine (SPD). $n = 4$. **d.** Schematic depicting how mRNA of *PLEKHM2* corresponds to frameshifted fusion product. **e.** Schematic of mass-spectrometry construct. **f.** Mass-spectrometry results, with identified peptides highlighted in red font and junction peptide underlines. Junction peptide coverage shown below, with line depicting the transition from 0- to +1-frame, confirming the site of PRF.

Figure 4 | The PLEKHM2-FS proteoform is constitutively active for moving lysosomes towards the cell periphery. a. Schematic representation of PLEKHM2 variants. The different

domains/motifs are indicated by colored boxes: RPIP8, UNC-14, and NESCA (RUN); Trp-Asp (WD); Trp-Glu (WE); pleckstrin homology (PH). Arrows and dashed lines indicate positive and negative interactions, respectively, with binding partners. Amino acid numbers are also indicated. **b**. WT HeLa cells were transiently transfected with the indicated HA-tagged PLEKHM2 constructs for 48 h and analyzed by confocal immunofluorescence microscopy for the HA epitope (green) and endogenous LAMP1 (lysosomes, red). Nuclei were labeled with DAPI (blue). Cell edges were outlined by staining of actin with Alexa Fluor™ 647-conjugated phalloidin (not shown) and indicated by dashed lines. Arrows indicate clusters of PLEKHM2 together with lysosomes (yellow arrows) or without (green arrows) at cell tips or more centrally (magenta arrows). Scale bars: 20 μm. **c.** Quantification of the percentage of WT (panel **b**) and ARL8-KO cells (panel **d**) displaying accumulation of PLEKHM2 variants or LAMP1 at cell tips. More than 150 cells per sample were scored. **d.** ARL8-KO HeLa cells were transfected and analyzed as described for panel **b**. Scale bars: 20 μm. **e.** Schematic representation of the BORC-ARL8-dependent activation and recruitment of PLEKHM2 to a lysosome (top), and the recruitment of the constitutive, ARL8-independent activity but ARL8-dependent recruitment of PLEKHM2 to a lysosome (bottom).

Extended Figure 1 | PLEKHM2 luciferase assays and additional mass-spectrometry data. **a.** Absolute firefly luciferase values. **b.** Absolute *Renilla* luciferase values; all values are within the same order of magnitude. **c.** Absolute firefly luciferase values in spermidine experiments. **d.** Absolute *Renilla* luciferase values in spermidine experiments. **e.** Mass-spectrometry results of PLEKHM2-K-GFP constructs; covered peptides are shown in red font with the junction peptide underlined. The transition of the reading frame is shown below.

Extended Figure 2 | Expression, activity, and self-association of PLEKHM2-FS. a. WT or ARL8-KO HeLa cells were transiently transfected with the indicated HA-tagged PLEKHM2 constructs (see scheme in Fig. 4a) for 48 h and analyzed by SDS-PAGE and immunoblot analysis with antibodies to the HA epitope. β-tubulin was used as a loading control. The positions of molecular mass markers (Mr, in kDa) are indicated on the left. **b.** WT HeLa cells like those shown in panel **a** were analyzed by confocal immunofluorescence microscopy for the HA epitope (green) and endogenous LAMTOR4 (lysosomes, red). Nuclei were labeled with DAPI (blue). Cell edges were outlined by staining of actin with fluorescent phalloidin (not

shown) and indicated by dashed lines. Arrows indicate clusters of HA-PLEKHM2 together with lysosomes (yellow arrows) or without (green arrows) at the cell tips or more centrally (magenta arrows). Scale bars: 20 μm. Notice how all HA-PLEKHM2 constructs drive distribution of both HA-PLEKHM2 and lysosomes to cell tips. **c.** ARL8-KO HeLa cells were transfected and analyzed as described for panel **b**. Scale bars: 20 μm. Notice that HA-PLEKHM2 is unable to localize to and drive lysosomes to cell tips in ARL8-KO cells. In contrast, HA-PLEKHM2-FS can localize to, but cannot drive lysosomes to cell tips. HA-PLEKHM2-1-437 is incapable of moving either protein to cell tips. **d.** HEK293T cells were co-transfected with plasmids encoding the indicated HA-tagged constructs with myc-PLEKHM2-FS and subjected to immunoprecipitation with an antibody to the myc epitope. Immunoprecipitates (IP) and cell extracts (10%, Input) were analyzed by SDS-PAGE and immunoblotting (IB) for the HA and myc epitopes, or β-tubulin (loading control). Red arrows indicate the HA constructs that are co-immunoprecipitated with the myc-PLEKHM2-FS construct. The positions of molecular mass markers (Mr, in kDa) are indicated on the left. Notice the much greater co-immunoprecipitation of myc-PLEKHM2-FS with HA-PLEKHM2-FS relative to HA-PLEKHM2 and HA-PLEKHM2-1-437.

Extended Figure 3 | Airyscan microscopy of WT and ARL8-KO HeLa cells expressing different PLEKHM2 variants. a. WT HeLa cells were transiently transfected with the indicated HA-tagged PLEKHM2 constructs for 48 h and analyzed by immunofluorescence microscopy for the HA epitope (green), endogenous LAMP1 (red) and Alexa Fluor™ 647-conjugated phalloidin (magenta). Cells were imaged using a higher-resolution Airyscan confocal microscope. Arrows indicate clusters of PLEKHM2 together with lysosomes (yellow arrows) or without (green arrows) at cell tips outlined by Alexa Fluor™ 647-conjugated phalloidin. Scale bars: 5 μm. Notice that HA-PLEKHM2-FS or HA-PLEKHM2-1-437 form tight clusters with lysosomes. **b.** ARL8-KO HeLa cells were transfected, stained, and imaged as described for panel **a**. Notice the absence of HA-PLEKHM2 and the reduction of HA-PLEKHM2-1-437 at cell tips. Scale bars: 5 μm.

Figure S1 | Genomic alignment of 100 vertebrate species in a neighborhood of the dual coding region, with features relevant to protein-coding evolution color-coded by

CodAlignView. For clarity, insertions relative to the human sequence are not shown in (a) and (b); the same alignments including these insertions are shown in (c) and (d), respectively. (a) Codonization and color coding with respect to the +1 frame. Preponderance of synonymous substitutions (light green) in the latter portion of the dual coding region indicates purifying selection on the amino acid sequence translated from this frame. There are many TGA (yellow), TAG (magenta), and TAA (cyan) stop codons in other species in this reading frame both 5' of the PRF cassette (green arrow) and 3' of the +1 frame stop codon (red arrow), but none within the dual coding region. The CCTTTC in the PRF cassette is perfectly conserved in all aligned species. (b) Same region codonized and color-coded with respect to the 0 frame.

Figure S2 | Potential RNA secondary structure downstream of slippery site. **a.** Predicted RNA secondary structure using forna[41] and CodAlignView with secondary structure regions highlighted. Uppercase letters in the diagram indicate which nucleotides were inputted for structure prediction. **b.** Diagram showing what structural mutants were made.

Figure S3 | Ribosome profiling analysis mined from across different tissues. Individual datasets are represented by dots coloured by the tissue of origin. The orange dashed line and red solid line represent three and four z-score thresholds, respectively. The y-axis shows the log2 ratio of ribosome protected fragments mapped downstream of the +1 frame stop codon over upstream (see Methods). The x-axis shows the total number of fragments aligned to the considered region of the PLEKHM2 mRNA.

Figure S4 | Enhanced ARL8-independent localization of PLEKHM2-FS to cell tips. a. ARL8-KO HeLa cells were transiently transfected with the indicated HA-tagged PLEKHM2 constructs for 48 h and analyzed for confocal immunofluorescence microscopy for the HA epitope (green), and endogenous LAMP1 (lysosomes, red). Nuclei were labeled with DAPI (blue). Cell edges were outlined by staining of actin with fluorescent phalloidin (not shown) and indicated by dashed lines. Scale bars: 20 μm. **b.** Quantification of the percentage of cells displaying localization of HA-PLEKHM2 variants to cell tips (left panel) or cytosol (right panel) from more than 150 cells per sample. Notice that HA-PLEKHM2-1-437 is less tip-localized and more cytosolic relative to HA-PLEKHM2-FS, suggesting a role for the transframe sequence in

ARL8-independent localization of HA-PLEKHM2-FS to cell tips, indicative of interaction with kinesin-1.

Methods

Bioinformatic analysis

Reference sequences representing mammals, bird, amphibians and fish were used to generate sequences for each of these clades, in a manner previously reported[19]. For each clade, the sequences were translated to amino acids, aligned using MUSCLE[42], and the amino acid alignments were used to guide codon-respecting nucleotide sequence alignments using EMBOSS tranalign[43]. For the MLOGD and synplot2 analyses, the alignments were mapped to the coordinates of a specific reference sequence by removing alignment columns that contained a gap character in the reference sequence. The reference sequences used were the coding regions of NCBI accession numbers NM_015164.4 (*Homo sapiens*), XM_417616.8 (*Gallus gallus*), XM_041569881.1 (*Xenopus laevis*) and NM_001130783.1 (*Danio rerio*) for the mammal, sauropsid, amphibian and teleost fish alignments, respectively. These alignments were analysed with synplot2[24], using a 15-codon window size, and MLOGD[25], using a 25-codon window size.

Cell Culture and Transfections

For **Fig. 3**, HEK293T cells (ATCC) were maintained in DMEM supplemented with 10% FBS, 1 mM L-glutamine and antibiotics. For western blotting, cells were transfected with Lipofectamine 2000 reagent (Invitrogen) in 6-well plates using the 1-day protocol in which suspended cells are added directly to the DNA complexes in 6-well plates, with 1 mg of each indicated plasmid. The transfecting DNA complexes in each well were incubated with 5×10^5 cells suspended in 3 ml DMEM + 10% FBS and incubated for 36 h at 37°C in 5% CO_2.

For GFP-Trap transfections, 1 x 10^7 cells were forward-transfected with Lipofectamine 2000 reagent in 15 cm Petri dishes with 10 mg of each indicated plasmid. Cells were incubated for 48 h at 37°C in 5% CO_2.

For dual luciferase assays, cells were transfected with Lipofectamine 2000 reagent, again using the 1-day protocol described above. The following were added to each well: 25 ng of each plasmid plus 0.2 µl Lipofectamine 2000 in 25 µl Opti-Mem (Gibco). The transfecting DNA

complexes in each well were incubated with 3×10^4 cells suspended in 50 µl DMEM + 10% FBS at 37°C in 5% CO_2 for 20 h.

For polyamine experiments, cells were transfected using Lipofectamine 2000 reagent as described above. Suspended cells were supplemented to a final concentration of 1 mM aminoguanidine hydrochloride (Sigma), or 1 mM aminoguanidine hydrochloride plus 2 mM spermidine (Sigma) before adding to transfecting DNA complexes. Luciferase activities were measured 20 h after spermidine treatment.

For **Fig. 4,** HeLa and HEK293T cells were cultured in DMEM (Quality Biological, #112-319-101), supplemented with 2 mM L-glutamine (GIBCO, #25030081), 10% fetal bovine serum (BSA) (Corning, # 35-011-CV), 100 U/ml penicillin-streptomycin (GIBCO, # 15140122) (complete DMEM) in a 37°C incubator (5% CO_2, 95% air). HeLa cells grown on 6-well plates were transiently transfected with 2 µg plasmid DNA using 8 µl Lipofectamine 3000 (Invitrogen, #L3000001), according to the manufacturer's instructions. Approximately 24 h after transfection, 40,000 cells were replated onto 12-mm coverslips coated with collagen (BioTechne, #3442-050-01). The remaining unseeded cells were used for SDS-PAGE and immunoblotting. Cells were then cultured for an additional 24 h before fixation and immunofluorescent labeling. For co-immunoprecipitation experiments, HEK293T cells grown on 6-well plates were transiently transfected for 48 h with 1.5 µg of each plasmid DNA and 8 µl Lipofectamine 3000 (Invitrogen, #L3000001) according to the manufacturer's instructions.

Expression constructs

PLEKHM2 fused dual luciferase constructs (for immunoblotting) were generated by either 1-step or 2-step PCR on *PLEKHM2* gBlock1 (IDT) using primer sequences which incorporated 5′ *Xho*I and 3′ *Bgl*II restriction sites (outlined in Supplementary Table 1). PCR amplicons were digested with *Xho*I / *Bgl*II and cloned into *Psp*XI / *Bgl*II digested pDlucV2.0. pDlucV2.0 is a version of pDluc[30] generated by introducing silent mutations into the *Renilla* coding sequence to disrupt two potential donor splice sites (TGGgtaagt).

PLEKHM2-X-GFP and PLEKHM2-K-X-GFP were synthesized as gBlocks with flanking 5′ *Sac*I and 3′ *Bam*HI restriction sites and cloned into pcDNA3.4.

PLEKHM2 unfused dual luciferase constructs (for dual luciferase assay) were generated by either 1-step or 2-step PCR on a *PLEKHM2* G Block (IDT) using primer sequences which

incorporated 5′ *Xho*I and 3′ *Bgl*II restriction sites (outlined in Supplementary Table 1). PCR amplicons were digested with *Xho*I / *Bgl*II and cloned into *Psp*XI / *Bgl*II digested pSGDlucV3.0 (Addgene 119760)[32]. *OAZ1* dual luciferase expression constructs were described previously[44]. All clones were verified by Sanger sequencing (Eurofins). All constructs used for cell biology assays were directly synthesized by Twist Biosciences for overexpression in mammalian cells.

Antibodies and chemicals

The following primary antibodies were used for immunoblotting (IB) and/or immunofluorescence microscopy (IF) (catalog numbers, sources and working dilutions are in parentheses): rat anti-HA (12158167001, Roche, IF 1:500, IB 1:1,000), mouse anti-LAMP1 (H4A3, DSHB, IF 1:500), rabbit anti-LAMTOR4 (13140, Cell Signaling, IF 1:500), rabbit anti-myc-tag (2272, Cell Signaling, IB 1:1,000), mouse anti-tubulin-HRP (sc-32293, Santa Cruz, IF 1:1,000). Secondary antibodies: HRP-conjugated goat anti-rat IgG (H+L) (112-035-003, Jackson Immuno Research, IB 1:5,000), HRP-conjugated goat anti-rabbit IgG (H+L) (111-035-003, Jackson Immuno Research, IB 1:5,000), Alexa Fluor 555-conjugated donkey anti-mouse IgG (A-31570, Thermo Scientific, IF 1:1,000), Alexa Fluor 555-conjugated goat anti-rabbit IgG (A-21428, Thermo Scientific, IF 1:1,000), Alexa Fluor 488-conjugated goat anti-rat IgG (A-11006, Thermo Scientific, IF 1:1,000). Cell edges were outlined by staining of actin with Alexa FluorTM 647-conjugated phalloidin (A22287, Thermo Scientific, IF 1:500).

Dual Luciferase Assay

Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems). Transfected cells were lysed in 15 µl of 1 × passive lysis buffer (PLB: Promega) and light emission was measured following injection of 50 µl of either *Renilla* or firefly luciferase substrate[45]. Recoding efficiencies were determined by calculating relative luciferase activities (firefly/*Renilla*) from test constructs and dividing by relative luciferase activities from replicate wells of in-frame-control constructs. Four replicate biological samples were assayed each with four technical repeats. Statistical significance was determined using a two-tailed, homoscedastic Student's *t*-test.

SDS-PAGE and immunoblotting

For **Fig. 3**, transfected cells were lysed in 100 ml 1 × PLB. Proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Protran), which were incubated at 4°C overnight with primary antibodies. Immunoreactive bands were detected on membranes after incubation with appropriate fluorescently labeled secondary antibodies using a LI-COR Odyssey® Infrared Imaging Scanner.

For **Fig. 4,** cells were washed twice with ice-cold phosphate-buffered saline (PBS; Corning, #21-040-CM), scraped from the plates in 1x Laemmli sample buffer (1x LSB) (Bio-Rad, #1610747) supplemented with 2.5% v/v 2-mercaptoethanol (Sigma-Aldrich, #60-24-2), heated at 95°C for 5 min, and resolved by SDS-PAGE. Gels were blotted onto nitrocellulose membrane and blocked with 5% w/v non-fat milk in Tris-buffered saline, 0.1% v/v Tween 20 (TBS-T) for 20 min. Membranes were sequentially incubated with primary antibody and secondary HRP-conjugated antibody diluted in TBS-T. SuperSignal West Dura Reagents (Thermo Fisher, #34075) were used for detection of the antibody signal with a Bio-Rad Chemidoc MP imaging system. β-tubulin was used as a control for sample loading.

Immunoprecipitation

For **Fig. 3**, Cells were lysed in 700 ml PLB and then incubated with 20 ml of protein G Agarose beads plus anti-HA (3 mg) overnight at 4°C with gentle rocking. The beads were washed with ice-cold 1 × PLB buffer and then removed from the beads by boiling for 5 min in 2 × SDS-PAGE sample buffer for SDS-PAGE and western blotting. For GFP immunoprecipitation, GFP-Trap (Chromtek) was used following the manufacturer's instructions. Briefly, cells were lysed in 100 ml NP-40 lysis buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA; 0.5% NP-40) then 95 ml lysate was diluted to 700 ml in dilution buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) before incubation with 20 ml of GFP-Trap beads for 1 h at 4°C with gentle rocking. The beads were washed with ice-cold dilution buffer and then removed from the beads by boiling for 5 min in 2 × SDS-PAGE sample buffer for SDS-PAGE and Coomassie staining.

For **Fig. 4**, transfected HEK293T cells were lifted in ice-cold PBS and centrifuged at 500 *g* for 5 min. The pellet was washed twice with ice-cold PBS and lysed in 10 mM Tris-Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% v/v Nonidet P40 supplemented with a protease inhibitor cocktail (Roche). Cell lysates were clarified by centrifugation at 17,000 *g* for 10 min and the

supernatant (10% was saved as input) was incubated with anti-myc magnetic agarose beads (Thermo Fisher, #88842) at 4°C for 2 h. After three washes with 10 mM Tris-Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, the immunoprecipitates were eluted with 1x LSB at 95°C for 5 min. The immunoprecipitated samples and inputs were analyzed by SDS-PAGE and immunoblotting.

Immunofluorescence microscopy

Cells were seeded on collagen-coated coverslips in 24-well plates at 40,000 cells per well in regular culture medium. After 24 h, cells were fixed in 4% w/v paraformaldehyde (Electron Microscopy Sciences, #15700) in PBS for 20 min, permeabilized and blocked with 0.1% w/v saponin, 1% w/v BSA (Gold Bio, #A-421-10) in PBS for 20 min, and sequentially incubated with primary and secondary antibodies diluted in 0.1% w/v saponin, 1% w/v BSA in PBS for 30 min at 37°C. Coverslips were washed three times in PBS and mounted on glass slides using Fluoromount-G (Electron Microscopy Sciences, #17984-24) with DAPI. Z-stack cell images were acquired on a Zeiss LSM 900 inverted confocal microscope (Carl Zeiss) using a Plan-Apochromat 63X objective (NA = 1.4) with or without Airyscan detection. Maximum intensity projections were generated with Zeiss ZEN Black software, and final composite images were created using ImageJ/Fiji (https://fiji.sc/).

Ribosome profiling

Gencode v.46 transcript ENST00000375799 (4134 bp) was used as a representative mRNA of PLEKHM2. The frameshift cassette (UCC_UUU_CGG) is located between the coordinates 1545 - 1553, and frameshifted ribosomes terminate at an early stop codon in the new reading frame (+1) at coordinates 1789 - 1791 (*TAG*). Intuitively, PRF is expected to reduce the number of ribosome protected fragments (RPFs) aligning downstream of the frameshift-introduced stop codon, and as such the reduction in read density reflects the efficiency of PRF.This relationship can be mathematically represented as the ratio between read density downstream over the one upstream the frameshift cassette.

To limit the effect of aberrant pauses around translation initiation and termination sites, the 100 nt downstream and upstream, respectively, have been discarded. In a similar reasoning, the region from 50 nt upstream the frameshift cassette and up to 50 nt downstream the frameshift-introduced stop codon has been excluded from the analysis. Specifically, the region

between coordinates 340 - 1494 was used to estimate the density of RFPs upstream the frameshift cassette, while coordinates 1842 - 3199 identify the region used to estimate the density downstream of the frameshift-introduced stop codon. TPM (transcripts per kilobase million) values for corresponding regions were obtained from RiboCrypt (https://ribocrypt.org/) for each individual dataset, and to calculate the ratio the mean value of TPM for each region was considered. The ratios were then log2-transformed for the further steps of the analysis.

The classification by tissue of origin was based on the curated metadata available in RiboSeq Data Portal (https://rdp.ucc.ie). To assess the statistical significance of the ratio variation, we used a sliding window Z-score approach similar to what was described earlier[46]. Briefly, the log2 ratios from each study were sorted by ascending values of "total counts" - that is to say, the sum of all the raw counts mapping to each nucleotide position in the regions before and after the frameshift cassette. Then the ratios were grouped into bins of size 300 and descriptive statistics (mean and standard deviation) were calculated and stored. The window was then shifted with a step of 10 and the statistics recomputed, until the end of the available ratios. At that point, each ratio had a collection of means and standard deviations which were averaged and used to obtain the Z-score value for that specific log2 ratio.

To represent thresholds of Z-scores of 3 and 4 over the log_2 ratio distributions, we used the binning-and-sliding window approach mentioned before to calculate the standard deviations and means for each value, but with bin size of 50 and step of 10. The intervals between 3 and 4 standard deviations from the mean were then computed for each datapoint and plotted on the graph as two threshold lines (Fig. S3).

Acknowledgements

Y.A.K. was supported by a Knight-Hennessy Scholarship, a NSF Graduate Research Fellowship Program award, and an NIH F31 grant (F31MH134477). A.E.F. and J.W. were supported by Wellcome Trust Senior Research Fellowships (106207/Z/14/Z, 220814/Z/20/Z) and a European Research Council grant (646891) to A.E.F. J.F.A. was supported by an Irish Research Council Advanced Laureate Award (IRCLA/2019/74). P.V.B. was supported by Science Foundation Ireland Frontiers for the Future Award (20/FFP-A/8929) and SFI-HRB-Wellcome Trust Biomedical Research Partnership Investigator Award in Science (210692/Z/18/Z). Work in the J.S.B. lab was supported by the Intramural Program of NICHD, NIH (project ZIA HD001607).

J.M.M. and I.J. were supported by the National Human Genome Research Institute (NHGRI) of the U.S. National Institutes of Health (NIH) under award number (2U41HG007234). I.J. was also supported by NIH R01 HG004037. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. J.M.M. was also supported by the Wellcome Trust (108749/Z/15/Z) and the European Molecular Biology Laboratory (EMBL). Ensembl is a registered trademark of EMBL. For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted article version arising from this submission. The authors would like to acknowledge Michael Z. Palo for discussions related to this manuscript.

References

1. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–7078 (2016).
2. Hill, C. H. & Brierley, I. Structural and Functional Insights into Viral Programmed Ribosomal Frameshifting. *Annu. Rev. Virol.* **10**, 217–242 (2023).
3. Khan, Y. A. *et al.* Evaluating ribosomal frameshifting in CCR5 mRNA decoding. *Nature* **604**, E16–E23 (2022).
4. Loughran, G., Fedorova, A. D., Khan, Y. A., Atkins, J. F. & Baranov, P. V. Lack of evidence for ribosomal frameshifting in ATP7B mRNA decoding. *Mol. Cell* **82**, 3745-3749.e2 (2022).
5. Keren-Kaplan, T. & Bonifacino, J. S. ARL8 Relieves SKIP Autoinhibition to Enable Coupling of Lysosomes to Kinesin-1. *Curr. Biol. CB* **31**, 540-554.e5 (2021).
6. Firth, A. E. *et al.* Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. *Open Biol.* **2**, 120109 (2012).
7. Jagger, B. W. *et al.* An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. *Science* **337**, 199–204 (2012).
8. Ramakrishnan, V. Ribosome Structure and the Mechanism of Translation. *Cell* **108**, 557–572 (2002).
9. Chen, J., Tsai, A., O'Leary, S. E., Petrov, A. & Puglisi, J. D. Unraveling the Dynamics of Ribosome Translocation. *Curr. Opin. Struct. Biol.* **22**, 804–814 (2012).
10. Milicevic, N., Jenner, L., Myasnikov, A., Yusupov, M. & Yusupova, G. mRNA reading frame maintenance during eukaryotic ribosome translocation. *Nature* **625**, 393–400 (2024).
11. Rodnina, M. V. *et al.* Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* **48**, 1056–1067 (2020).

12. Li, Y. *et al.* Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2172-2181 (2014).

13. Napthine, S. *et al.* Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat. Commun.* **8**, 15582 (2017).

14. Matsufuji, S. *et al.* Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**, 51–60 (1995).

15. Bhatt, P. R. *et al.* Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).

16. Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280–283 (1988).

17. Shigemoto, K. *et al.* Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting. *Nucleic Acids Res.* **29**, 4079–4088 (2001).

18. Wills, N. M., Moore, B., Hammer, A., Gesteland, R. F. & Atkins, J. F. A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *J. Biol. Chem.* **281**, 7082–7088 (2006).

19. Khan, Y. A. *et al.* Evidence for a novel overlapping coding sequence in POLG initiated at a CUG start codon. *BMC Genet.* **21**, 25 (2020).

20. Loughran, G. *et al.* Unusually efficient CUG initiation of an overlapping reading frame in POLG mRNA yields novel protein POLGARF. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 24936–24946 (2020).

21. Dinan, A. M., Atkins, J. F. & Firth, A. E. ASXL gain-of-function truncation mutants: defective and dysregulated forms of a natural ribosomal frameshifting product? *Biol. Direct*

12, 24 (2017).

22. Rosa-Ferreira, C. & Munro, S. Arl8 and SKIP act together to link lysosomes to kinesin-1. *Dev. Cell* **21**, 1171–1178 (2011).

23. Eisinger, J., Feuer, B. & Yamane, T. Codon-anticodon binding in tRNAphe. *Nature. New Biol.* **231**, 126–128 (1971).

24. Firth, A. E. Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res.* **42**, 12425–12439 (2014).

25. Firth, A. E. & Brown, C. M. Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* **7**, 75 (2006).

26. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011).

27. Mudge, J. M. *et al.* Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. *Genome Res.* **29**, 2073–2087 (2019).

28. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

29. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

30. Fixsen, S. M. & Howard, M. T. Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. *J. Mol. Biol.* **399**, 385–396 (2010).

31. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA N. Y. N* **4**, 479–486 (1998).

32. Loughran, G., Howard, M. T., Firth, A. E. & Atkins, J. F. Avoidance of reporter assay

distortions from fused dual reporters. *RNA* **23**, 1285–1289 (2017).

33. Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* **13**, 928–933 (1994).

34. Ivanov, I. P. & Atkins, J. F. Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res.* **35**, 1842–1858 (2007).

35. Oguro, A. *et al.* Translation efficiency affects the sequence-independent +1 ribosomal frameshifting by polyamines. *J. Biochem. (Tokyo)* **168**, 139–149 (2020).

36. Higashi, K. *et al.* Enhancement of +1 Frameshift by Polyamines during Translation of Polypeptide Release Factor 2 in *Escherichia coli**. *J. Biol. Chem.* **281**, 9527–9537 (2006).

37. Boucrot, E., Henry, T., Borg, J.-P., Gorvel, J.-P. & Méresse, S. The intracellular fate of Salmonella depends on the recruitment of kinesin. *Science* **308**, 1174–1178 (2005).

38. Sanger, A. *et al.* SKIP controls lysosome positioning using a composite kinesin-1 heavy and light chain-binding domain. *J. Cell Sci.* **130**, 1637–1651 (2017).

39. Pu, J., Keren-Kaplan, T. & Bonifacino, J. S. A Ragulator-BORC interaction controls lysosome positioning in response to amino acid availability. *J. Cell Biol.* **216**, 4183–4197 (2017).

40. Filipek, P. A. *et al.* LAMTOR/Ragulator is a negative regulator of Arl8b- and BORC-dependent late endosomal positioning. *J. Cell Biol.* **216**, 4199–4215 (2017).

41. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).

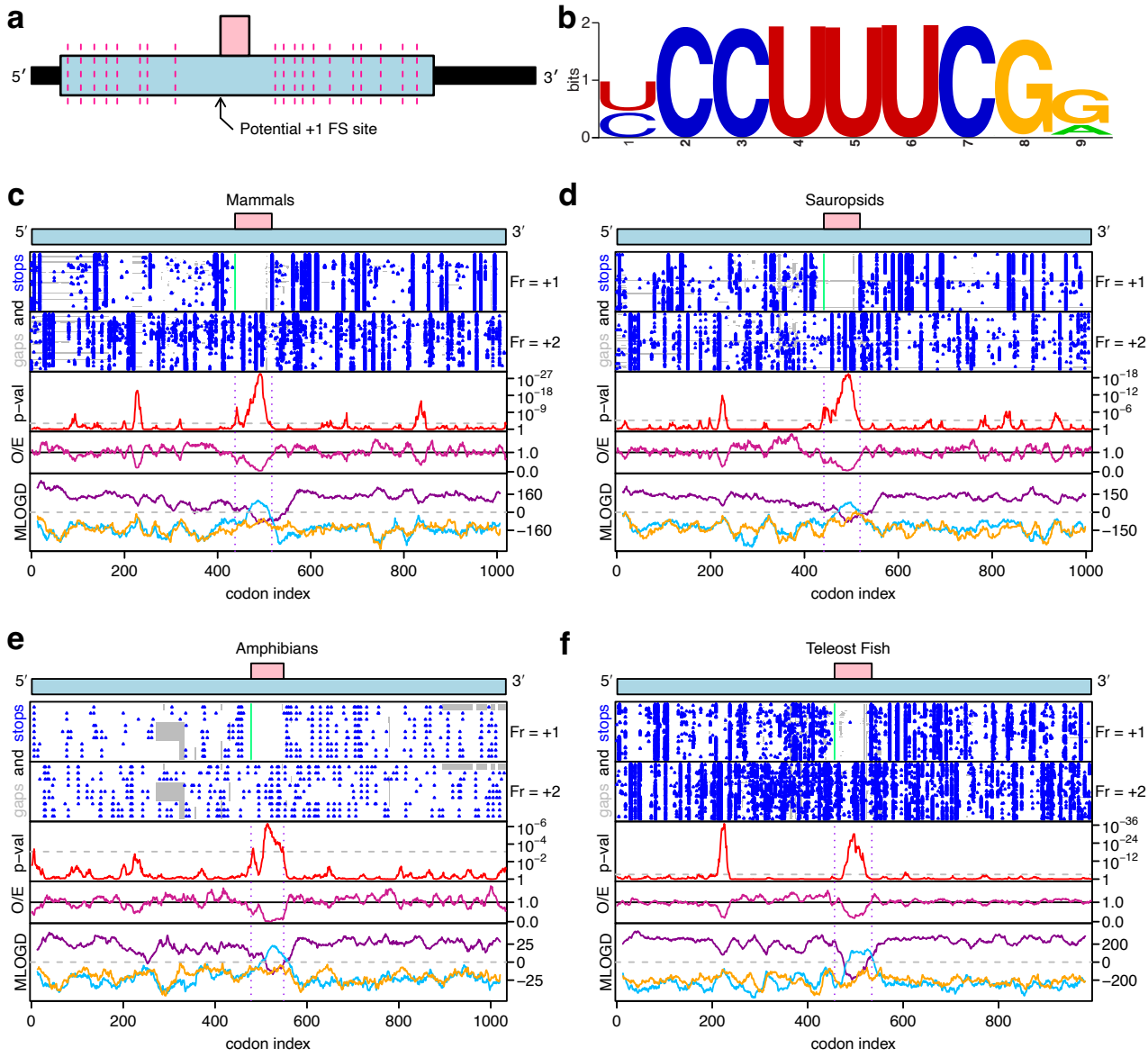
42. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

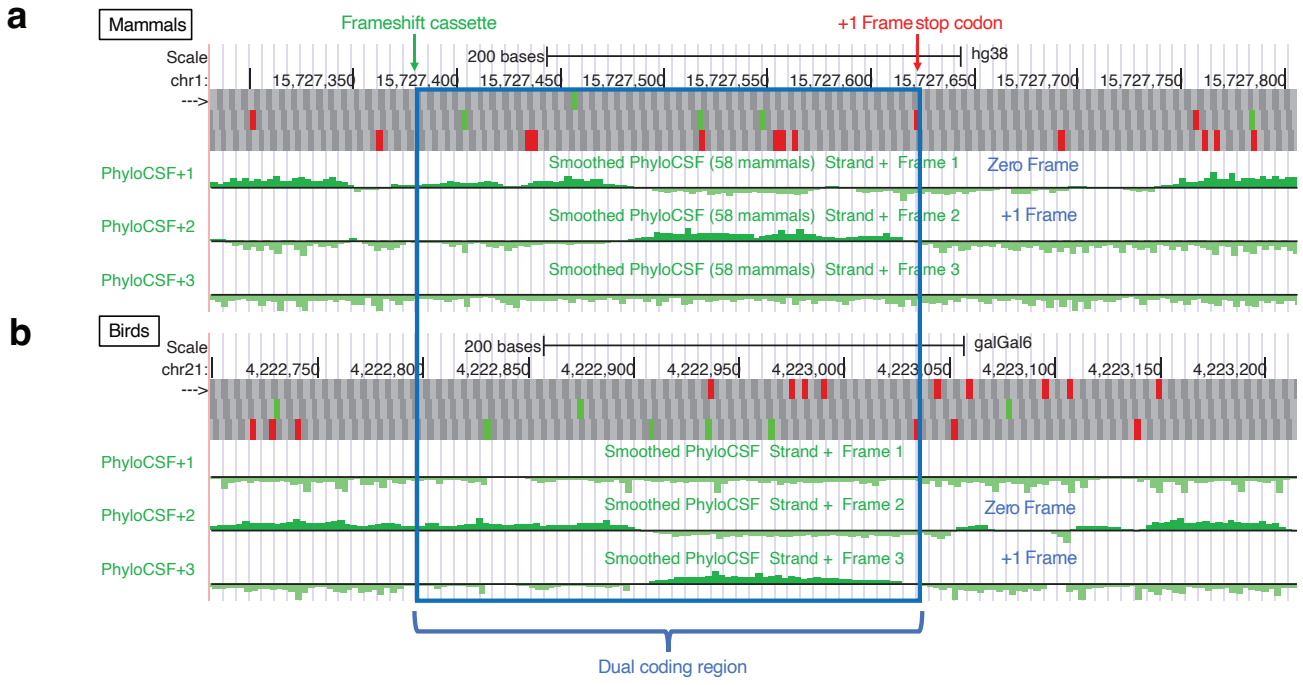
43. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG* **16**, 276–277 (2000).

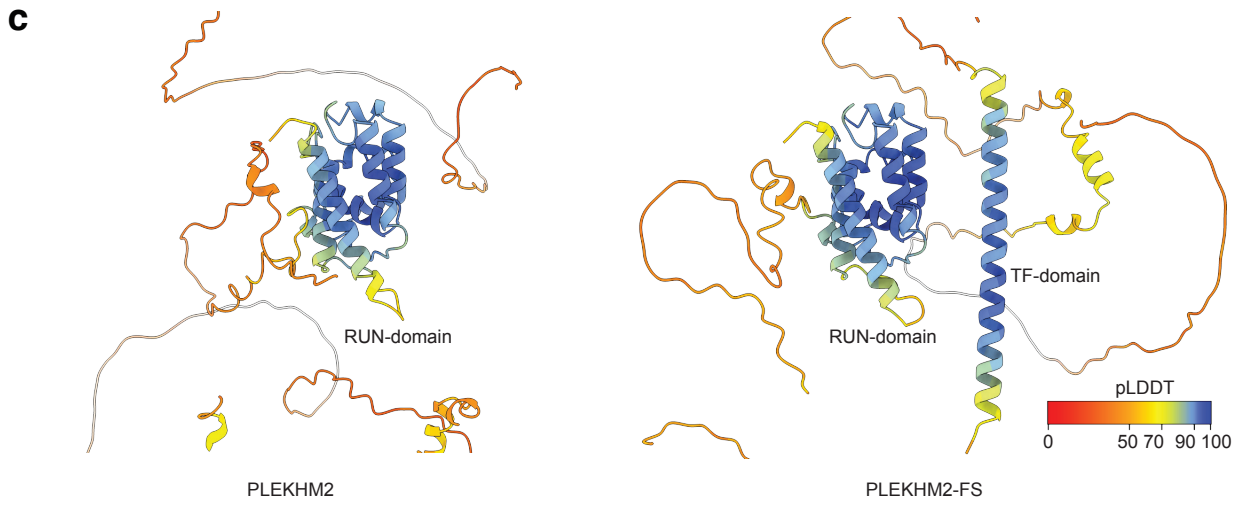
44. Loughran, G., Li, X., O'Loughlin, S., Atkins, J. F. & Baranov, P. V. Monitoring translation in all reading frames downstream of weak stop codons provides mechanistic insights into the impact of nucleotide and cellular contexts. *Nucleic Acids Res.* **51**, 304–314 (2023).

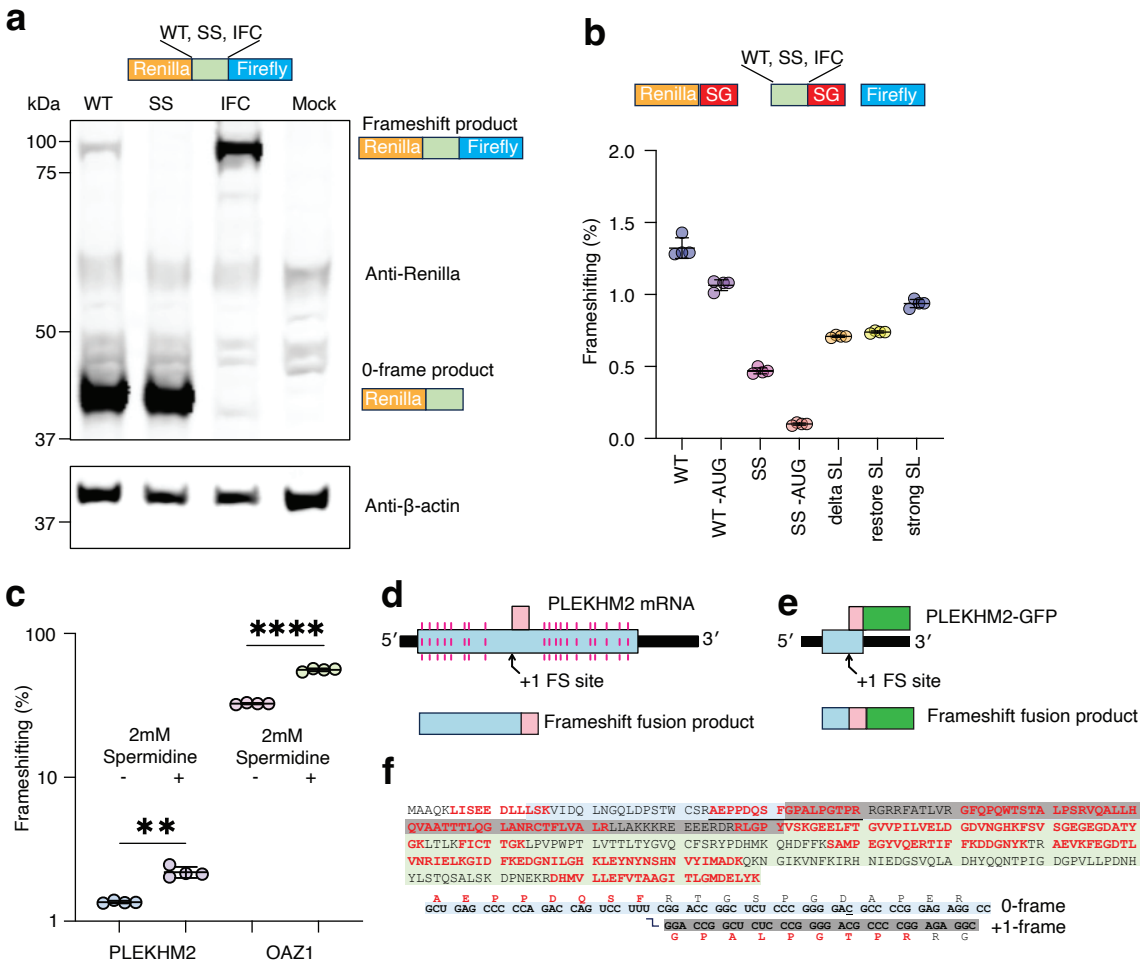
45. Dyer, B. W., Ferrer, F. A., Klinedinst, D. K. & Rodriguez, R. A noncommercial dual luciferase enzyme assay system for reporter gene analysis. *Anal. Biochem.* **282**, 158–161 (2000).

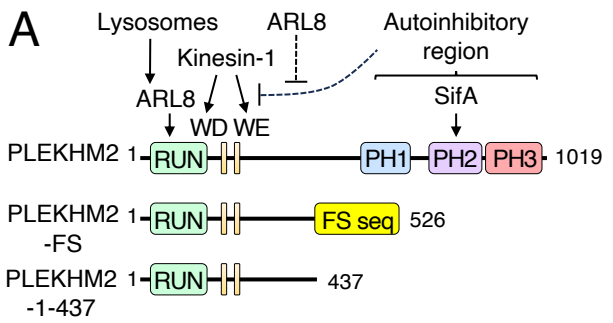
46. Andreev, D. E. *et al.* Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol.* **16**, 90 (2015).

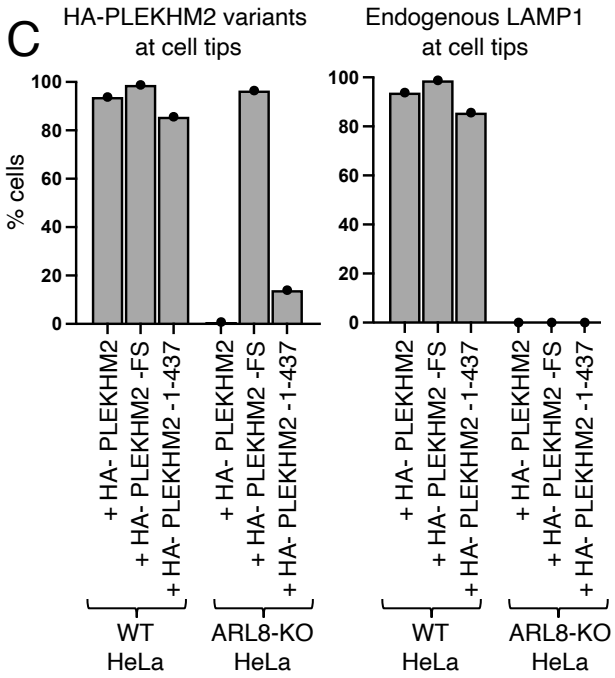

a
5′
3′
Potential +1 FS site
b
bits
c
Mammals
d
Sauropsids
e
Amphibians
f
Teleost Fish
stops and gaps
Fr = +1
Fr = +2
p–val
O/E
MLOGD
codon index

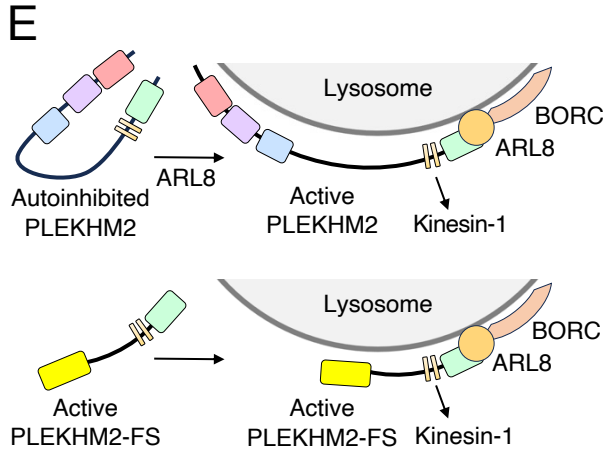

a
Mammals
Frameshift cassette
+1 Frame stop codon
Scale
chr1:
200 bases
hg38
15,727,350
15,727,400
15,727,450
15,727,500
15,727,550
15,727,600
15,727,650
15,727,700
15,727,750
15,727,800
PhyloCSF+1
Smoothed PhyloCSF (58 mammals) Strand + Frame 1
Zero Frame
PhyloCSF+2
Smoothed PhyloCSF (58 mammals) Strand + Frame 2
+1 Frame
PhyloCSF+3
Smoothed PhyloCSF (58 mammals) Strand + Frame 3
b
Birds
Scale
chr21:
200 bases
galGal6
4,222,750
4,222,800
4,222,850
4,222,900
4,222,950
4,223,000
4,223,050
4,223,100
4,223,150
4,223,200
PhyloCSF+1
Smoothed PhyloCSF Strand + Frame 1
PhyloCSF+2
Smoothed PhyloCSF Strand + Frame 2
Zero Frame
PhyloCSF+3
Smoothed PhyloCSF Strand + Frame 3
+1 Frame
Dual coding region

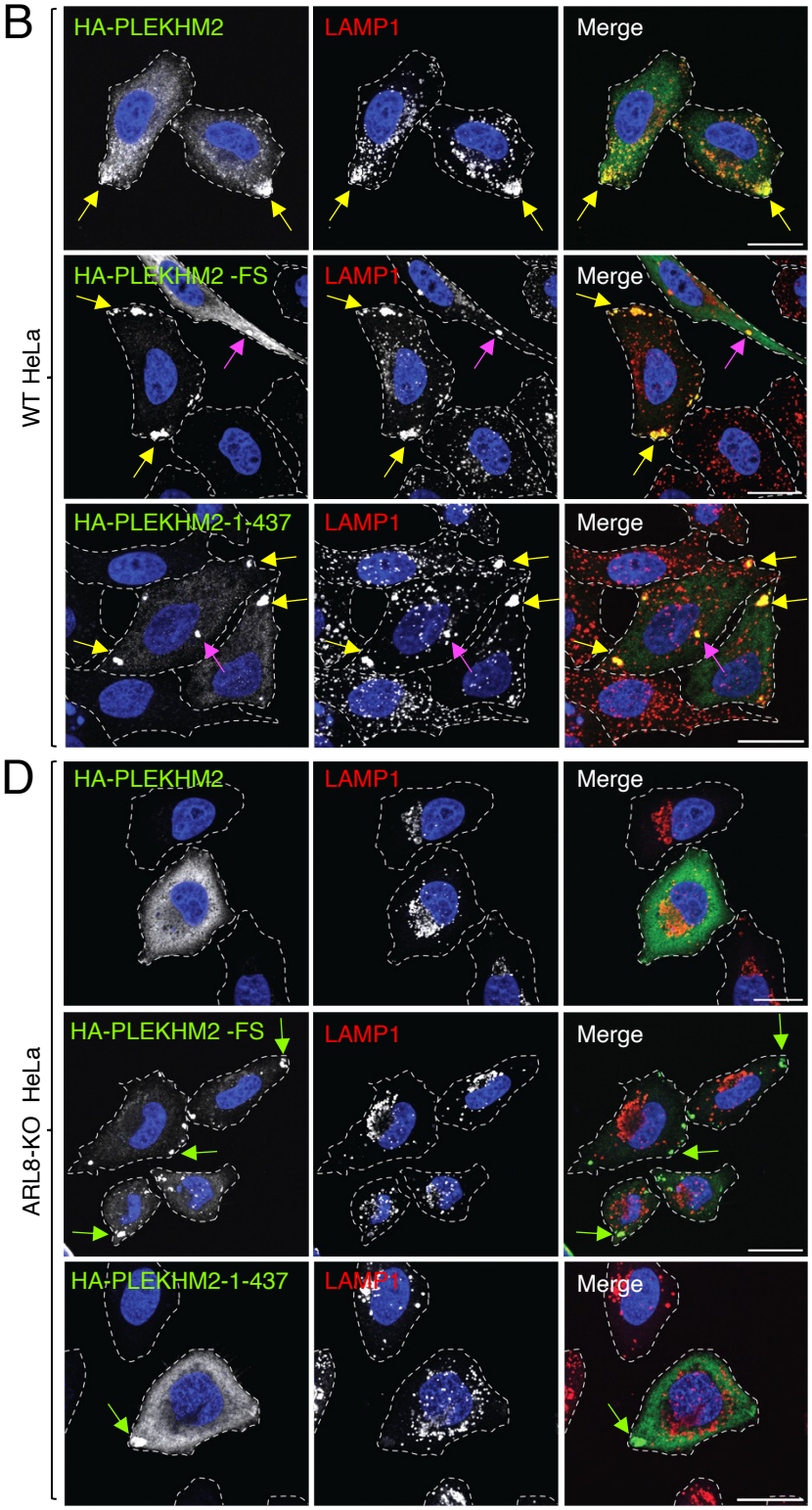

c
RUN-domain
PLEKHM2
TF-domain
RUN-domain
pLDDT
0
50 70 90 100
PLEKHM2-FS

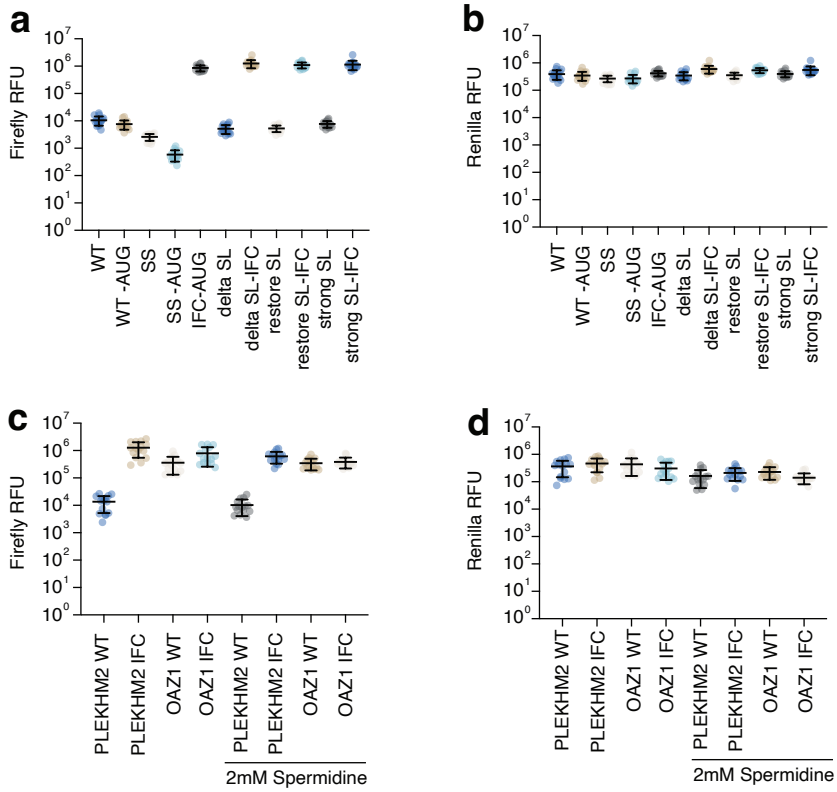

a
WT, SS, IFC
Renilla
Firefly
kDa
WT
SS
IFC
Mock
100
75
50
37
Frameshift product
Renilla
Firefly
Anti-Renilla
0-frame product
Renilla
37
Anti-β-actin
b
WT, SS, IFC
Renilla
SG
SG
Firefly
Frameshifting (%)
2.0
1.5
1.0
0.5
0.0
WT
WT -AUG
SS
SS -AUG
delta SL
restore SL
strong SL
c
Frameshifting (%)
100
10
1
****
**
2mM Spermidine
-
+
2mM Spermidine
-
+
PLEKHM2
OAZ1
d
PLEKHM2 mRNA
5′
3′
+1 FS site
Frameshift fusion product
e
PLEKHM2-GFP
5′
3′
+1 FS site
Frameshift fusion product
f
MAAQKLISEE DLLLSKVIDQ LNGQLDPSTW CSRAEPPDQS FGPALPGTPR RGRRFATLVR GFQPQWTSTA LPSRVQALLH
QVAATTTLQG LANRCTFLVA LRLLAKKKRE EEERDRRLGP YVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY
GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH
YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK
A E P P D Q S F R T G S P G D A P E R
GCU GAG CCC CCA GAC CAG UCC UUU CGG ACC GGC UCU CCC GGG GAC GCC CCG GAG AGG CC
0-frame
GGA CCG GCU CUC CCG GGG ACG CCC CGG AGA GGC
+1-frame
G P A L P G T P R R G


A
Lysosomes
ARL8
Kinesin-1
Autoinhibitory region
ARL8
SifA
WD WE
PLEKHM2 1
RUN
PH1
PH2
PH3
1019
PLEKHM2 -FS 1
RUN
FS seq
526
PLEKHM2 -1-437 1
RUN
437


C
HA-PLEKHM2 variants at cell tips
Endogenous LAMP1 at cell tips
% cells
100
80
60
40
20
0
+ HA- PLEKHM2
+ HA- PLEKHM2 -FS
+ HA- PLEKHM2 -1-437
WT HeLa
ARL8-KO HeLa

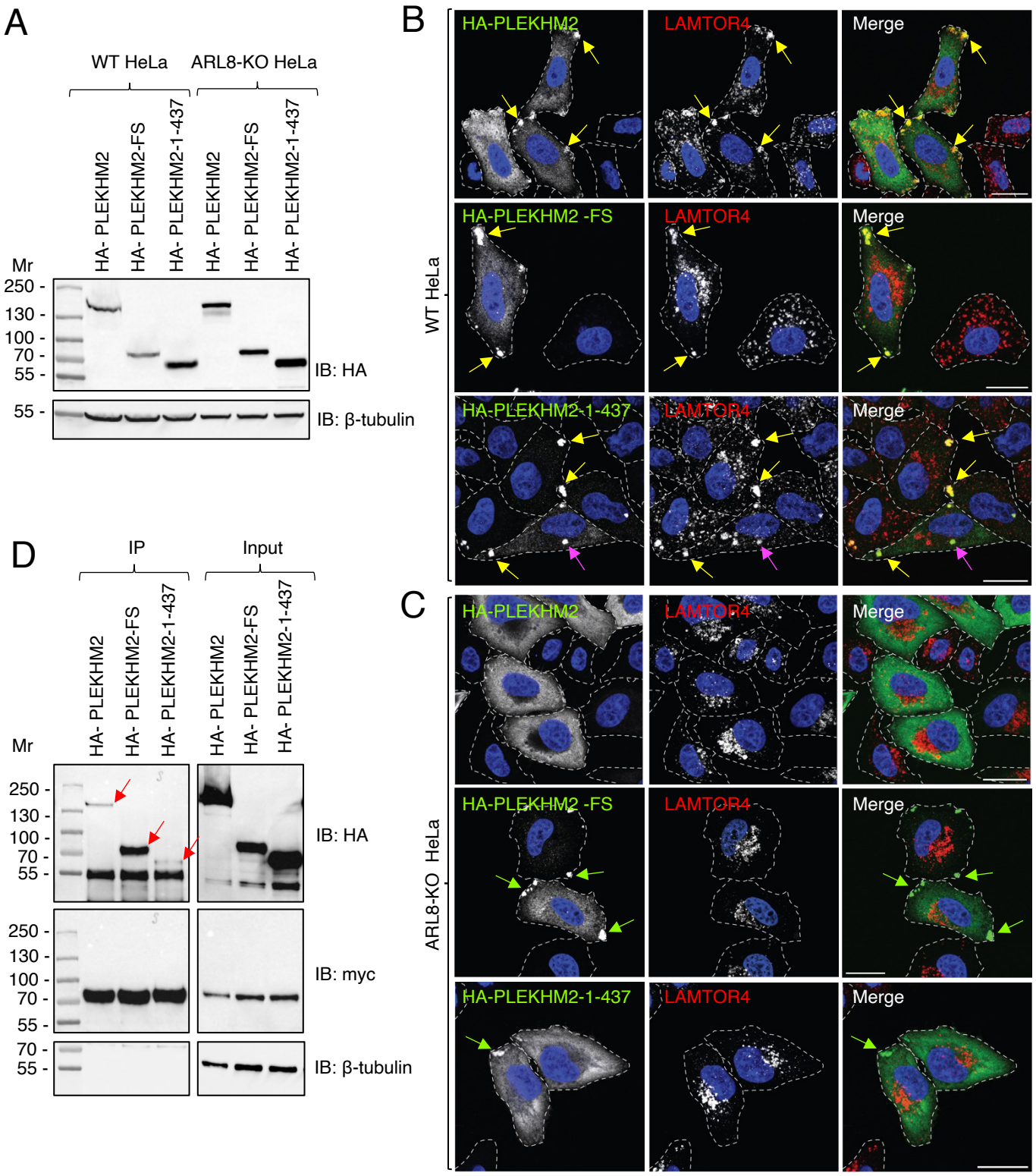

E
Autoinhibited PLEKHM2
ARL8
Lysosome
BORC
ARL8
Active PLEKHM2
Kinesin-1
Active PLEKHM2-FS
Lysosome
BORC
ARL8
Active PLEKHM2-FS
Kinesin-1

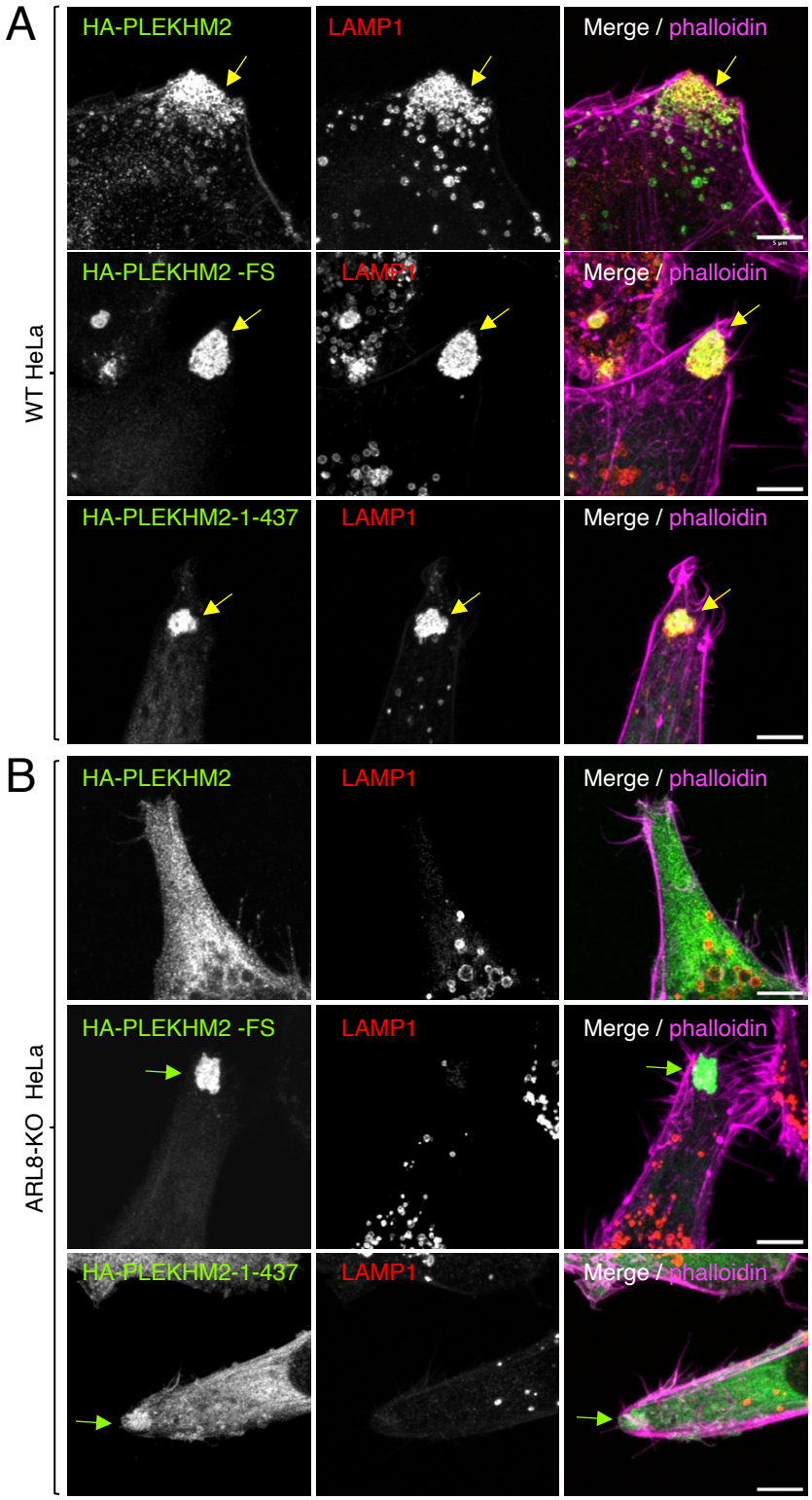

B
WT HeLa
HA-PLEKHM2
LAMP1
Merge
HA-PLEKHM2 -FS
LAMP1
Merge
HA-PLEKHM2-1-437
LAMP1
Merge
D
ARL8-KO HeLa
HA-PLEKHM2
LAMP1
Merge
HA-PLEKHM2 -FS
LAMP1
Merge
HA-PLEKHM2-1-437
LAMP1
Merge

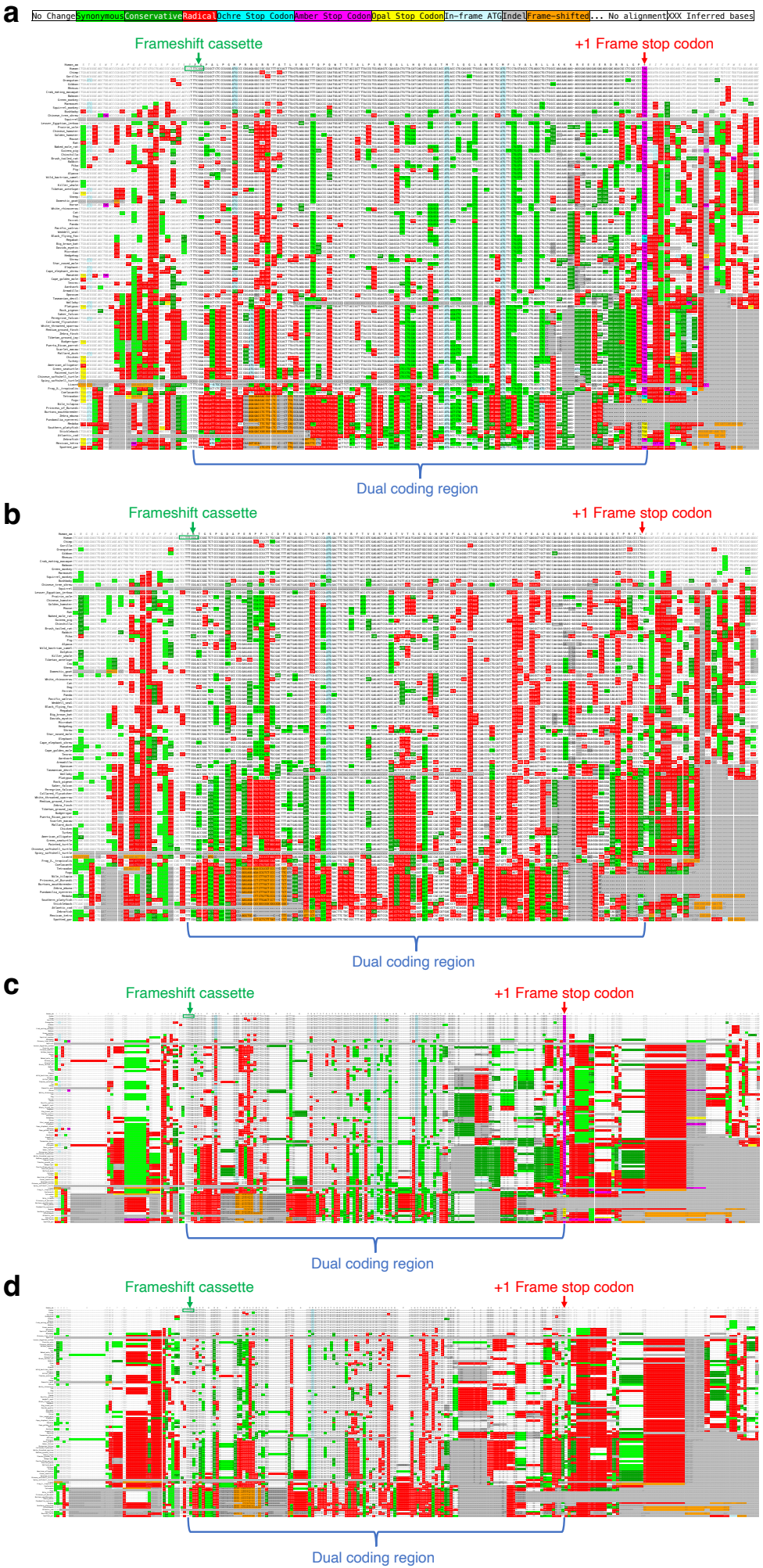


e

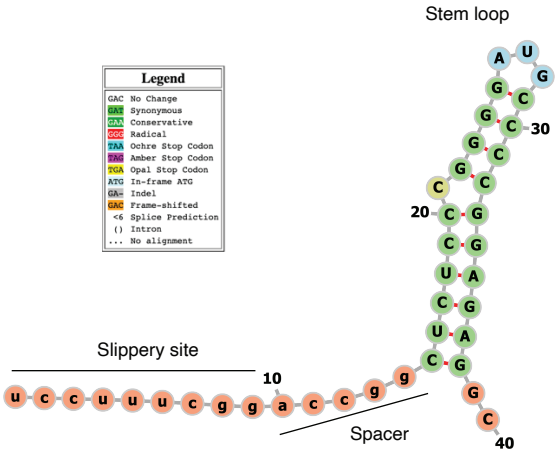
```
MAAQKLISEE DLLLSKVIDQ LNGQLDPSTW CSRAEPPKQS FGPALPGTPR RGRRFATLVR GFQPQWTSTA LPSRVQALLH
QVAATTTLQG LANRCTFLVA LRLLAKKKRE EEERDRRLGP YVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY
GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH
YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK

 A  E  P  P  K  Q  S  F  R  T  G  S  P  G  D  A  P  E  R
GCU GAG CCC CCA AAG CAG UCC UUU CGG ACC GGC UCU CCC GGG GAC GCC CCG GAG AGG CC   0-frame
                                GGA CCG GCU CUC CCG GGG ACG CCC CGG AGA GGC      +1-frame
                                 G  P  A  L  P  G  T  P  R  R  G
```

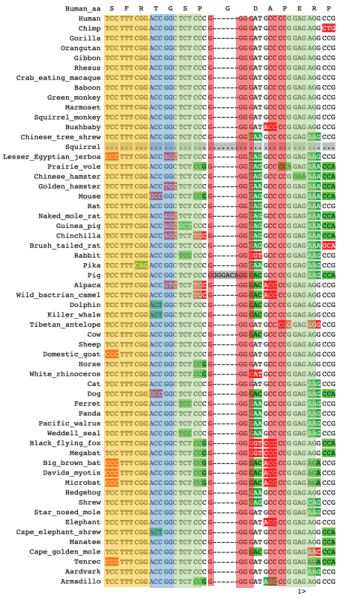

A
WT HeLa
ARL8-KO HeLa
HA- PLEKHM2
HA- PLEKHM2-FS
HA- PLEKHM2-1-437
Mr
250 -
130 -
100 -
70 -
55 -
IB: HA
55 -
IB: β-tubulin
B
WT HeLa
HA-PLEKHM2
LAMTOR4
Merge
HA-PLEKHM2 -FS
HA-PLEKHM2-1-437
C
ARL8-KO HeLa
D
IP
Input
IB: HA
IB: myc
IB: β-tubulin


A
WT HeLa
HA-PLEKHM2
LAMP1
Merge / phalloidin
HA-PLEKHM2 -FS
LAMP1
Merge / phalloidin
HA-PLEKHM2-1-437
LAMP1
Merge / phalloidin
B
ARL8-KO HeLa
HA-PLEKHM2
LAMP1
Merge / phalloidin
HA-PLEKHM2 -FS
LAMP1
Merge / phalloidin
HA-PLEKHM2-1-437
LAMP1
Merge / phalloidin

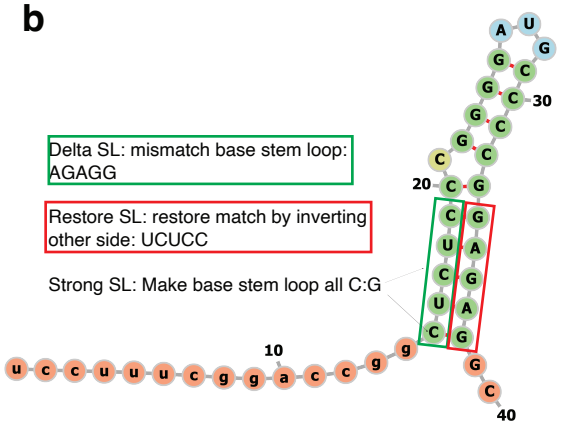

a
No Change
Synonymous
Conservative
Radical
Ochre Stop Codon
Amber Stop Codon
Opal Stop Codon
In-frame ATG
Indel
Frame-shifted
...
No alignment
XXX Inferred bases
Frameshift cassette
+1 Frame stop codon
Dual coding region
b
Frameshift cassette
+1 Frame stop codon
Dual coding region
c
Frameshift cassette
+1 Frame stop codon
Dual coding region
d
Frameshift cassette
+1 Frame stop codon
Dual coding region

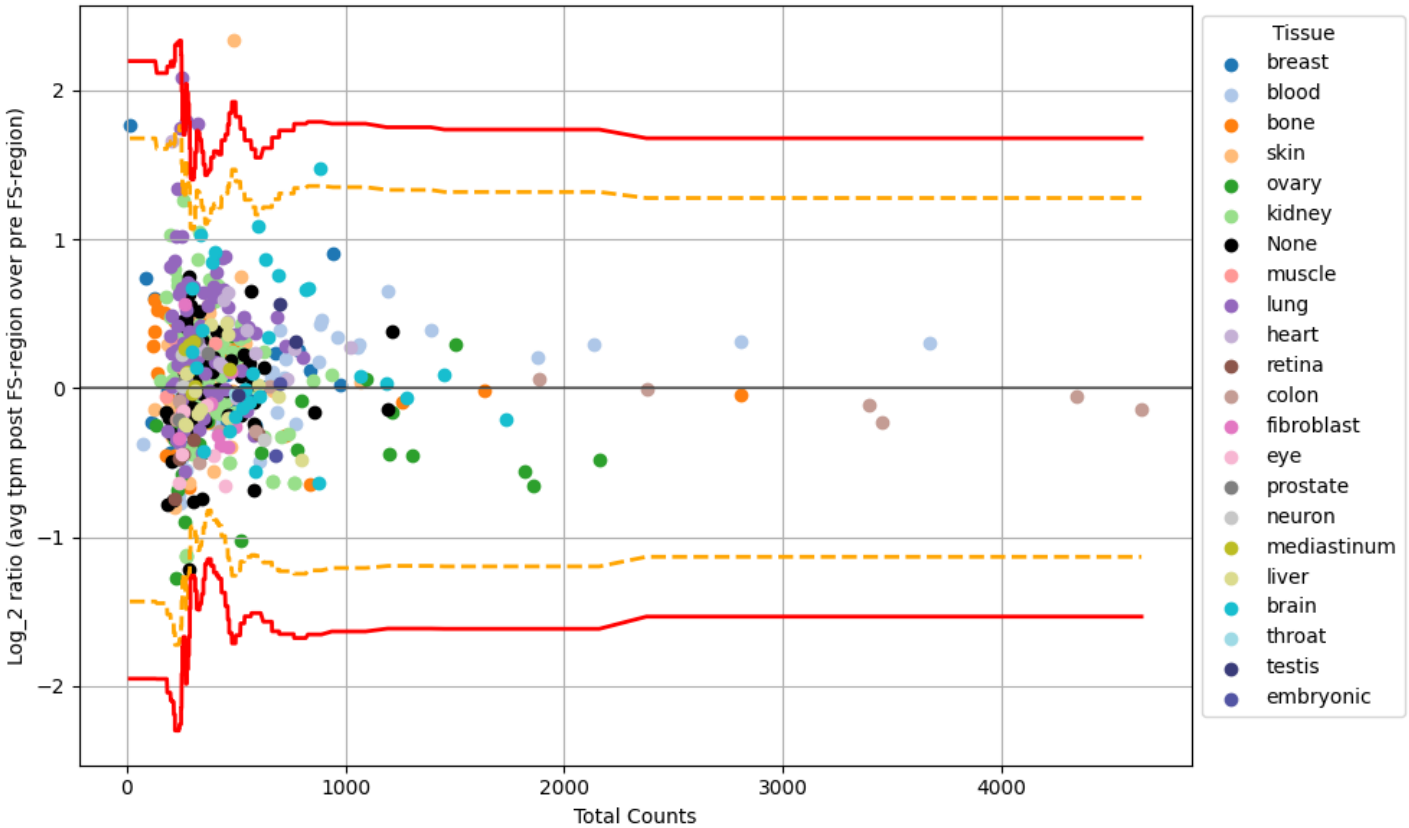
a

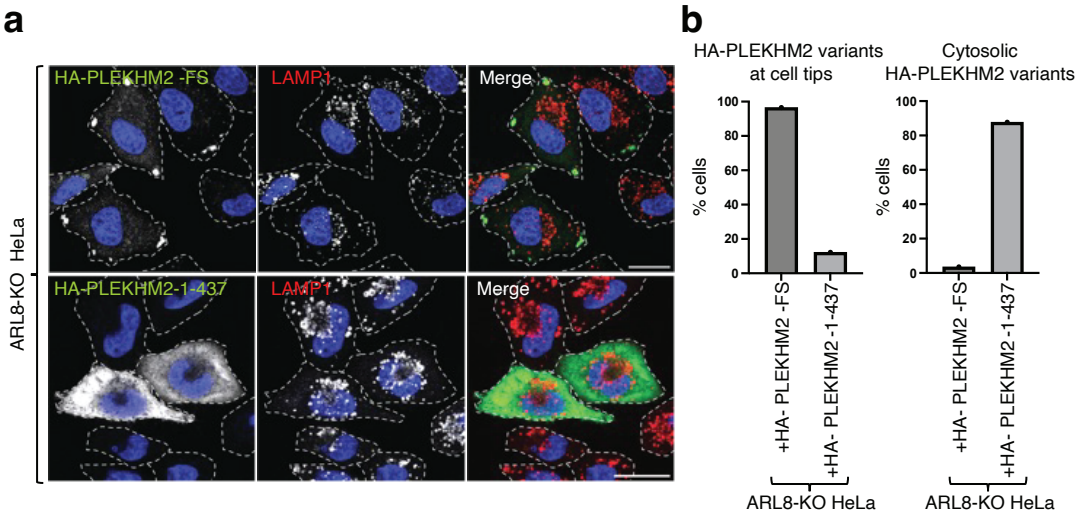

Stem loop
Legend
GAC No Change
GAT Synonymous
CAA Conservative
CGC Radical
TAA Ochre Stop Codon
TAG Amber Stop Codon
TGA Opal Stop Codon
ATG In-frame ATG
GA- Indel
GAC Frame-shifted
<6 Splice Prediction
() Intron
... No alignment
Slippery site
Spacer
10
20
30
40


Human_aa S F R T G S P G D A P E R P
Human
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Crab_eating_macaque
Baboon
Green_monkey
Marmoset
Squirrel_monkey
Bushbaby
Chinese_tree_shrew
Squirrel
Lesser_Egyptian_jerboa
Prairie_vole
Chinese_hamster
Golden_hamster
Mouse
Rat
Naked_mole_rat
Guinea_pig
Chinchilla
Brush_tailed_rat
Rabbit
Pika
Pig
Alpaca
Wild_bactrian_camel
Dolphin
Killer_whale
Tibetan_antelope
Cow
Sheep
Domestic_goat
Horse
White_rhinoceros
Cat
Dog
Ferret
Panda
Pacific_walrus
Weddell_seal
Black_flying_fox
Megabat
Big_brown_bat
Davids_myotis
Microbat
Hedgehog
Shrew
Star_nosed_mole
Elephant
Cape_elephant_shrew
Manatee
Cape_golden_mole
Tenrec
Aardvark
Armadillo

b


Delta SL: mismatch base stem loop: AGAGG
Restore SL: restore match by inverting other side: UCUCC
Strong SL: Make base stem loop all C:G
10
20
30
40


Log_2 ratio (avg tpm post FS-region over pre FS-region)
Total Counts
2
1
0
−1
−2
0
1000
2000
3000
4000
Tissue
breast
blood
bone
skin
ovary
kidney
None
muscle
lung
heart
retina
colon
fibroblast
eye
prostate
neuron
mediastinum
liver
brain
throat
testis
embryonic


a
HeLa
ARL8-KO
HA-PLEKHM2 -FS
LAMP1
Merge
HA-PLEKHM2-1-437
LAMP1
Merge
b
HA-PLEKHM2 variants at cell tips
Cytosolic HA-PLEKHM2 variants
% cells
100
80
60
40
20
0
+HA- PLEKHM2 -FS
+HA- PLEKHM2 -1-437
ARL8-KO HeLa
ARL8-KO HeLa