

RNA Localization and Translational Regulation on the Endoplasmic Reticulum

by

Chun-Chieh Hsu

Department of Biochemistry
Duke University

Date: _____

Approved:

Christopher Nicchitta, Supervisor

Michael Boyce

Richard Brennan

Michael Fitzgerald

Jack Keene

Dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy
in the Department of Biochemistry
in the Graduate School of Duke University

2016

ABSTRACT

RNA Localization and Translational Regulation on the Endoplasmic Reticulum

by

Chun-Chieh Hsu

Department of Biochemistry
Duke University

Date: _____

Approved:

Christopher Nicchitta, Supervisor

Michael Boyce

Richard Brennan

Michael Fitzgerald

Jack Keene

An abstract of a dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy
in the Department of Biochemistry
in the Graduate School of Duke University

2016

Copyright by
Chun-Chieh Hsu
2016

Abstract

mRNA localization is emerging as a critical cellular mechanism for the spatiotemporal regulation of protein expression and serves important roles in oogenesis, embryogenesis, cell fate specification, and synapse formation. Of the many cellular examples of mRNA localization, mRNA localization to the endoplasmic reticulum (ER) operates on the largest scale, and mediates that localization of the ca. 40% of the genome that encodes secretory or membrane proteins. For this cohort of mRNAs, localization requires translation and utilizes a signal in the protein, termed the signal sequence, to direct localization. Signal sequence-encoding mRNAs are localized to the endoplasmic reticulum (ER) membrane by either of two mechanisms, a canonical mechanism of translation on ER-bound ribosomes (referred to as the signal recognition particle (SRP) pathway), or a poorly understood direct ER anchoring mechanism. In this study, I identify a family of ER integral membrane proteins that function as RNA-binding proteins and play important roles in the direct anchoring of mRNAs to the ER. Specifically, I report that one of the ER integral membrane RNA-binding proteins, AEG-1 (astrocyte elevated gene-1), functions in the direct ER anchoring and translational regulation of mRNAs encoding integral membrane proteins. To identify the AEG1 mRNA interactome, I performed HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) and PAR-CLIP (photoactivatable

ribonucleoside-enhanced crosslinking and immunoprecipitation analyses of the AEG-1 mRNA interactome of human hepatocellular carcinoma cells, which revealed the high enrichment for mRNAs encoding endomembrane organelle proteins, most notably encoding transmembrane proteins. AEG-1 binding sites were highly enriched in the coding sequence and displayed a signature cluster enrichment downstream of encoded transmembrane domains. In overexpression and knockdown models, AEG-1 expression levels strongly correlated with translational efficiency and protein functions of two representative members of its bound transcripts, *MDR1* (multidrug resistance protein 1) and *NPC1* (Niemann-Pick disease, type C1). This research has revealed a molecular mechanism for the selective localization of mRNAs to the ER and identifies a novel post-transcriptional gene regulation function for AEG-1 in membrane protein expression.

Dedication

To my families in Taiwan and North Carolina - my parents, Hwakuang Shee and Si-Si Wang, my sisters, Jenny Hsu and Edwina Hsu, my wife Tiffany Tzeng and our son Samuel Hsu, without whom this never would have happened.

Contents

Abstract	iv
Contents.....	vii
List of Tables.....	x
List of Figures	xi
1. Introduction	1
1.1 Intracellular RNA localization.....	1
1.2 RNA localization to the ER	6
1.3 Non-canonical RNA-binding proteins on the ER membrane	9
2. Materials and Methods.....	13
2.1 Cell lines and antibodies.....	13
2.2 RNA cross-linking to the endoplasmic reticulum membrane proteins.....	14
2.2.1 Pre-condensation of Triton X-114.....	14
2.2.2 TX-114 phase partitioning of canine pancreas rough microsomes.....	14
2.2.3 Alkaline extraction of canine pancreas rough microsomes.....	15
2.2.4 RNA-dependent T4 RNA ligase labeling of UV-irradiated rough microsomes	16
2.3 Proteomic analysis of canine rough microsome-associated mRNPs	17
2.4 UV-cross-linking-based mRNA-binding assay of the ER membrane proteins	18
2.5 UV cross-linking and immunoprecipitation (CLIP) for AEG-1 and truncation mutants	19
2.6 Sucrose cushion centrifugation and polysome profiling	19

2.7 CLIP-Seq protocols.....	20
2.8 Data analysis for CLIP-Seq.....	21
2.8.1 Establish reference transcriptome and read mapping	21
2.8.2 Cluster identification	22
2.8.3 Cluster analysis.....	22
2.9 RNA immunoprecipitation (RIP) and quantitative PCR.....	23
3. Identification of RNA-binding ER membrane proteins	24
3.1 Introduction.....	24
3.2 Identification of RNA-binding ER integral membrane proteins	24
3.3 Proteomic analyses of candidate ER-mRNA-anchoring proteins	29
3.4 Validation of RNA-binding activity for potential mRNA-anchoring proteins	32
4. The RNA-binding activity and selectivity of AEG-1	35
4.1 Introduction.....	35
4.2 AEG-1 is a RNA-binding ER membrane protein	37
4.3 AEG-1 binds actively translated mRNA	40
4.4 Identification of non-canonical RNA-binding domains in AEG-1	44
5. Investigation of <i>in vivo</i> mRNA-binding specificity of AEG-1.....	51
5.1 Introduction.....	51
5.2 AEG-1 CLIP-Seq mapping reveals a high enrichment in coding sequence interactions	52
5.3 AEG-1 RNA interactome is enriched in endomembrane organelle protein-and transmembrane protein-encoding mRNAs	60

5.4 Validation and analysis of AEG-1 bound mRNAs	68
5.5 AEG-1 regulates the protein expression and biological function of its bound mRNAs.....	72
6. Discussion	78
6.1 Identification of the mRNA-binding function of AEG-1	78
6.2 Emerging studies of non-canonical RNA-binding proteins.....	79
6.3 Intrinsic disorder domain in RNA-binding proteins	81
6.4 mRNA-binding specificity of AEG-1	83
7. Summary and perspective	86
7.1 Summary.....	86
7.2 Future directions.....	87
7.2.1 The molecular mechanism of mRNA-AEG-1 interaction	87
7.2.2 Global post-translational regulation of AEG-1	89
7.2.3 Census of mRNA-anchoring ER membrane proteins	90
7.3 Conclusion Remarks	91
References	92
Biography	106

List of Tables

Table 1: Summary of identified mRNA-binding ER integral membrane proteins..... 32

Table 2: Summary of read mapping statistics for HITS-CLIP and PAR-CLIP 56

List of Figures

Figure 1: Mechanisms of subcellular mRNA localization.	5
Figure 2: Identification of RNA-binding proteins on the ER membrane.	26
Figure 3: UV-cross-linking reveals endoplasmic reticulum membrane proteins functioning in RNA-binding.	28
Figure 4: Proteomic Analyses of Candidate ER-mRNA-anchoring Proteins.	30
Figure 5: UV-cross-linking analysis validates the mRNA-binding activity of the candidate ER membrane proteins.	34
Figure 6: RNA interactome screens identified mRNA-binding integral membrane proteins on the endoplasmic reticulum.	36
Figure 7: Subcellular localization of endogenous and epitope-tagged AEG-1 proteins...	37
Figure 8: Co-immunoprecipitation/RNase digestion of poly(A)-binding protein (PABP) with HA-tagged AEG-1.	38
Figure 9: AEG-1 binds RNA in living cells.	40
Figure 10: AEG-1 binds translating mRNA.	42
Figure 11: AEG-1 binds translating mRNA in polysome fractions.	43
Figure 12: Mutagenesis and <i>in vivo</i> RNA-binding analysis reveals a non-canonical RNA- binding domain in the cytosolic domain of AEG-1.	45
Figure 13: Identification of RNA-binding residues in the cytosolic domain of AEG-1 via BINDN+.	46
Figure 14: Mutagenesis and ribosome ultracentrifugation analysis reveals a non- canonical RNA-binding domain in AEG-1.	48
Figure 15: Protein sequence alignment of AEG-1 homologs.	50
Figure 16: UV-cross-linking and immunoprecipitation-based AEG-1 RNA interactome analyses.	53

Figure 17: Experimental quality controls for UV-cross-linking and immunoprecipitation-based AEG-1 RNA interactome analyses.	54
Figure 18: CLIP-Seq data analysis pipeline.	55
Figure 19: Mutation plots for CLIP-Seq.	57
Figure 20: Mapping of sequence reads to RNA for HITS-CLIP replicates and PAR-CLIP.	58
Figure 21: Cluster length distributions for HITS-CLIP replicates and PAR-CLIP.	59
Figure 22: Distributions of AEG-1 binding clusters mapping to the 5' untranslated region (5' UTR), coding sequence (CDS), and 3' UTR for HITS-CLIP replicates and PAR-CLIP.	60
Figure 23: AEG-1 mRNA interactome.	61
Figure 24: Gene ontology (GO) analyses of AEG-1 RNA interactome.	63
Figure 25: Distributions of sequencing reads mapping to mRNA encoding cytosolic, secretory, and transmembrane proteins for HITS-CLIP replicates and PAR-CLIP.	65
Figure 26: Distributions of AEG-1 bound clusters on mRNAs.	67
Figure 27: Validation and analysis of AEG-1 mRNA interactome.	69
Figure 28: AEG-1 binding sites on MDR1 and NPC1 mRNAs.	71
Figure 29: Biological function of MDR1 protein in multidrug-resistant cells.	73
Figure 30: AEG-1 upregulates the protein expression of MDR1.	74
Figure 31: AEG-1 upregulates chemoresistance against anti-cancer drugs.	75
Figure 32: AEG-1 knockdown decreases the protein expression of NPC1 and causes cholesterol accumulation.	77
Figure 33: AEG-1 has an abundance of intrinsically disordered domains.	83

1. Introduction

1.1 *Intracellular RNA localization*

1.1.1 RNA localization in eukaryotic cells

RNA localization is a common cellular strategy for restricting mRNA distribution and protein synthesis to defined subcellular domains and serves established roles in biological processes as diverse as oogenesis, embryogenesis, cell fate specification, cell movements, and synapse formation in the central nervous system (Martin and Ephrussi, 2009; Palacios and Johnston, 2001). The field owes its beginnings to the early observations from the Brodeur lab, demonstrating that β -actin mRNA was localized to the vegetal pole during early Ascidian development (Jeffery et al., 1983). Later studies revealed important roles for the localization of maternal mRNA in embryonic patterning in *Xenopus* and *Drosophila* oocytes (Berleth et al., 1988; Frigerio et al., 1986; Rebagliati et al., 1985). Subsequent studies shed light on RNA localization in somatic cells, notably as a mechanism to spatially restrict the production of cytoskeleton proteins in fibroblasts (Lawrence and Singer, 1986) and neurons (Garner et al., 1988). Although early studies had only observed RNA localization for specific transcripts, recent genome-wide approaches have revealed that a substantial fraction of mRNAs is spatially localized in a variety of cell types, including the *Drosophila* embryos (Lecuyer et al., 2009), protruding pseudopodia of mouse fibroblasts (Mili et al., 2008), dendrites

(Moccia et al., 2003), axons (Andreassi et al., 2010), as well as in fungi (Heym and Niessing, 2012; Zarnack and Feldbrugge, 2010), plants (Crofts et al., 2005), and prokaryotes (Keiler, 2011). These studies suggest that mRNA localization is a prevalent biological mechanism operating on a transcriptome scale in a diverse variety of cells.

1.1.2 Molecular mechanisms for subcellular RNA localization

Several potential RNA localization mechanisms have been identified, and include active RNA transport on cytoskeleton systems, generalized RNA degradation with localized protection, as well as diffusion and entrapment (Figure 1). Active RNA transport is a prevalent RNA localization mechanism, seen in a wide variety of experimental models including *Xenopus* oocytes, *Drosophila* oocytes and embryos, fibroblasts and neurons (Gagnon and Mowry, 2011), which operates via cytoskeleton/motor protein-based active transport of translationally-repressed, *cis*-localization element-encoding mRNAs (Figure 1C) (Martin and Ephrussi, 2009; Medioni et al., 2012). In general, *cis*-localization element-encoding mRNAs are recognized by specific RNA-binding proteins (*trans*-acting factors), to subsequently form ribonucleoprotein (RNP) complexes with molecular motor proteins. These complexes then undergo active transport along the actin cytoskeleton or microtubule network, and are eventually anchored at their final subcellular destinations. To achieve a spatially

restricted protein synthesis, mRNA translation is typically inactive during mRNA transport and resumed when arriving final cellular destinations (Lecuyer et al., 2007).

Trans-acting factors play important roles in target mRNA recognition, RNP assembly, and translation repression. In the molecular mechanism of active transport, the initial step is the recognition of *cis*-encoded localization elements by *trans*-acting factors. In general, *cis*-localization elements are usually located in the 3' untranslated region (3'UTR) though sometimes found in the coding sequence (CDS) (Chartrand et al., 1999; Gonzalez et al., 1999). Multiple RNA-binding proteins have been identified as *trans*-acting RNA localization factors in a variety of cells. Zipcode binding protein (ZBP1), for example, binds to the *cis*-encoding element (zipcode) in the 3' UTR of β -*actin* mRNA and facilitates active transport via microtubule systems (Zhang et al., 2001). A series of *trans*-acting factors were identified in *Xenopus*, including Vg1RBP and VgRBPs (p33, p36, p40, p60, p69, and p78) (Mowry and Cote, 1999; Yaniv and Yisraeli, 2001). These *trans*-acting factors recognize the *cis*-localization elements in the 3' UTR of *Vg1* mRNA (VgLE) and contribute to *Vg1* mRNA localization to the vegetal pole of *Xenopus* oocyte (Mowry and Cote, 1999; Yaniv and Yisraeli, 2001).

RNA asymmetry can also be achieved via a RNA degradation/protection mechanism (Figure 1A). By this mechanism, target RNAs are initially randomly distributed and subsequently undergo generalized degradation, with the key exception

of target RNAs enriched at the appropriate subcellular destinations by a mechanism of RNA protection (Figure 1A). The RNA degradation/protection mechanism has been demonstrated in *Xenopus* and *Drosophila* oocytes, which share highly conserved RNA degradation machinery (Bashirullah et al., 1999). However, the detailed molecular mechanism of RNA protection from RNA degradation still remains to be discovered.

mRNA localization by random diffusion and entrapment has been observed in *Drosophila* and *Xenopus* oocytes, where mRNAs are diffused and entrapped in specific compartments (Figure 1B). At late oocyte stages in *Drosophila*, *nanos* mRNA is randomly distributed through cytoplasmic bulk flow mechanisms, and subsequently immobilized at the posterior end, which constrains Nanos protein synthesis to this region of the embryo (Forrest and Gavis, 2003). Interestingly, in addition to its role as a cytoskeletal component functioning in active RNA transport, actin was found to play an important role in the posterior trapping of *nanos* mRNA (Forrest and Gavis, 2003). Moreover, further studies of mRNA localization in the late stage of *Drosophila* oogenesis suggest that many *Drosophila* mRNAs utilize similar diffusion and entrapment mechanisms (Nakamura et al., 1996). Similar mechanisms were observed in early *Xenopus* oogenesis as well, where *Xcat2* and *Xdazl* mRNAs are retained at a densely packed endoplasmic reticulum (ER) concentrated in the mitochondrial cloud (Chang et al., 2004). In these

cases, local traps (e.g. actin) play important roles in spatial restriction of mRNA distribution and, eventually, in the localization of protein synthesis.

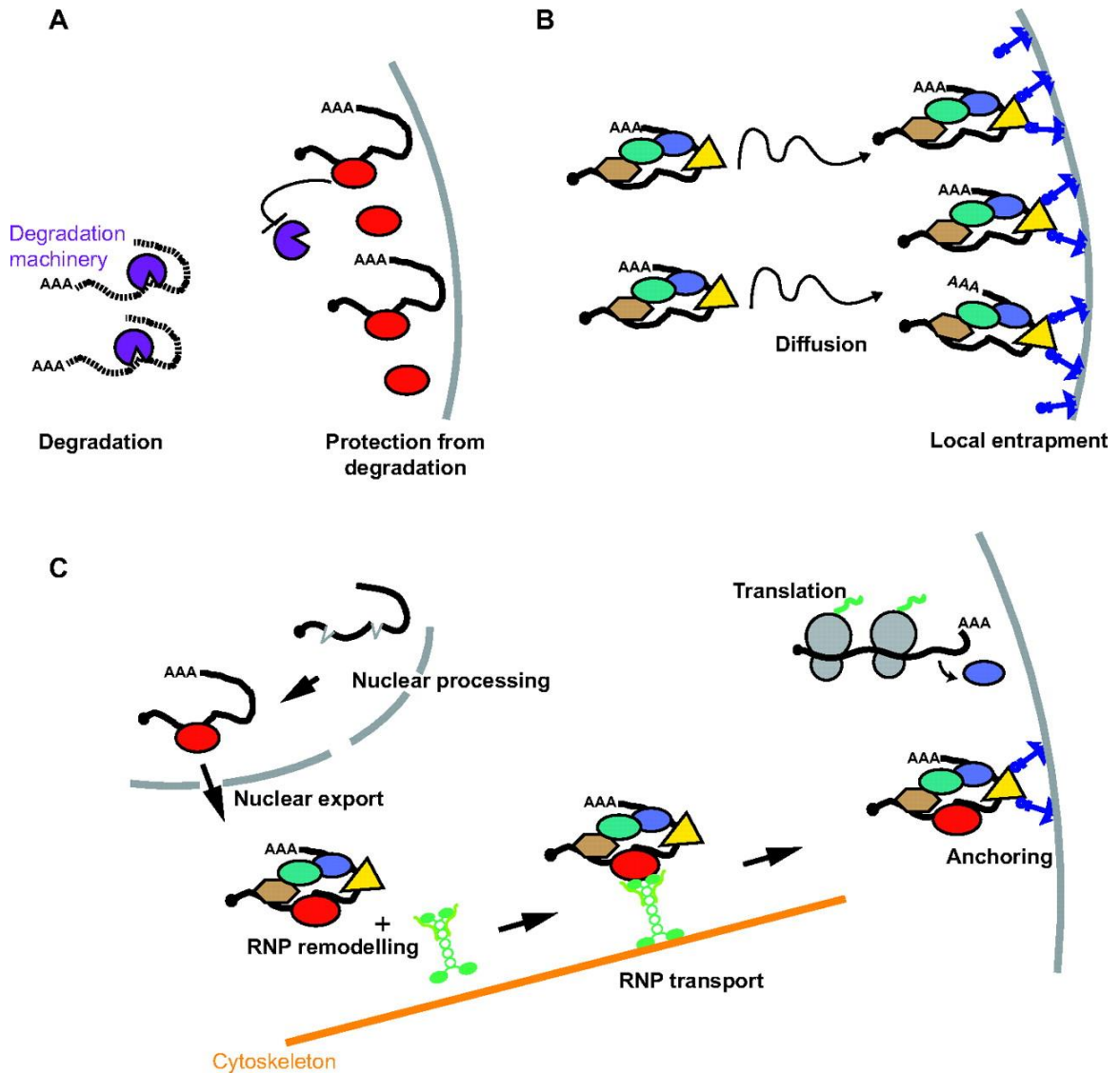


Figure 1: Mechanisms of subcellular mRNA localization.

Three mechanisms of subcellular mRNA localization have been characterized in eukaryotic cells, RNA degradation/protection mechanism (A), random diffusion and

entrapment (**B**), and active transport along cytoskeleton. Adapted from *Development* (Medioni et al., 2012).

1.2 RNA localization to the ER

1.2.1 Signal recognition particle pathway

In addition to the prominently studied RNA localization phenomena, which are restricted to specific cell types and/or developmental stages, all eukaryotic cells localize topogenic signal-encoding mRNAs to the endoplasmic reticulum (ER), the site of secretory and integral membrane protein synthesis (Blobel, 2000a; Palade, 1975; Walter and Johnson, 1994a). RNA localization to the ER operates on a remarkably large scale — ca. 30% of the human transcriptome is ER localized in HEK293 cells (Reid and Nicchitta, 2012). In current views, secretory- and integral membrane protein-encoding mRNAs are localized to the ER via a co-translational mechanism, where topogenic signals in the nascent polypeptide are bound by the signal recognition particle (SRP) and the mRNA/ribosome/nascent chain complex is subsequently localized to the ER via sequential interactions with the SRP receptor complex and the protein conducting channel/translocon complex on the ER membrane (Blobel, 2000b; Cross et al., 2009; Rapoport, 2007; Walter and Johnson, 1994b). As a translation- and nascent polypeptide-dependent RNA localization mechanism, the SRP pathway is distinct from other, established models of RNA localization as noted above, where *cis*-encoded localization

signals direct translationally-repressed mRNAs to their appropriate subcellular destinations (Blobel, 2000b; Martin and Ephrussi, 2009; Taliaferro et al., 2014; Walter and Johnson, 1994b). In contrast, RNA localization to the ER utilizes a co-translational mechanism, the SRP pathway, does not require cytoskeletal/motor protein-based active transport, and uses localization information encoded in the nascent protein rather than the mRNA (Blobel, 2000b; Rapoport et al., 1996; Walter and Johnson, 1994b; Walter and Lingappa, 1986; Zimmermann et al., 2011). Additionally, in SRP pathway, RNA localization to the ER is an indirect mRNA-ER interaction and a consequence of translation on ER-bound ribosomes (Becker et al., 2009; Pfeffer et al., 2015; Voorhees et al., 2014).

1.2.2 Ribosome-independent RNA localization

Recent investigations into the mechanisms of mRNA localization to the ER have identified a prominent exception to SRP pathway, where mRNAs encoding resident endomembrane organelle proteins (e.g., ER, Golgi, lysosomal proteins) display direct, ribosome-independent anchoring to the ER (Chen et al., 2011; Jagannathan et al., 2014; Reid and Nicchitta, 2015). In these studies, biochemical disruption of ribosome-ER membrane interactions resulted in the selective release of mRNAs encoding secretory proteins from the ER; mRNAs encoding endomembrane resident protein, in contrast,

retained their ER association in the absence of ribosomes (Chen et al., 2011; Jagannathan et al., 2014; Reid and Nicchitta, 2015). Subsequent studies into the mechanism(s) of ribosome-independent mRNA anchoring to the ER have revealed a diversity of candidate RNA interacting proteins on the ER. In one study, the ribosome receptor protein p180 (RRBP1) was identified as a general enhancer of poly(A) mRNA association with the ER, and specifically implicated in the ribosome-independent anchoring of mRNAs encoding placental alkaline phosphatase and calreticulin (Cui et al., 2012; Cui et al., 2013). An unbiased proteomic screen for the ER mRNA interactome identified numerous candidate ER resident mRNA-binding proteins, including components of the protein translocation machinery, p180, subunits of the N-linked oligosaccharyl transferase enzyme complex, reticulons, and other ER resident membrane proteins (Jagannathan et al., 2014). Concurrent proteomic screens for the mRNA interactomes of HeLa cells, HEK293 cells, and mouse embryonic stem cells (mESCs) revealed an unexpected diversity of candidate ER RNA-binding membrane proteins (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013), and also identified components of the ER protein translocation machinery, N-linked oligosaccharyl transferase complex, and other previously identified ER membrane proteins, as candidate ER RNA-binding proteins. Given the previously established diverse biochemical functions for many of these ER membrane proteins, putative function in poly(A) mRNA binding was unanticipated.

1.3 Non-canonical RNA-binding proteins on the ER membrane

RNA-binding proteins (RBPs) typically possess canonical RNA-binding domains, such as the RNA recognition motif (RRM), the K homology domain (KH), the Pumilio homology domain (PUM-HD) or the double stranded RNA-binding domain (dsRBD), and frequently such RNA-binding motifs are present as multi-domain/tandem RNA interaction sites (Gerstberger et al., 2014; Glisovic et al., 2008; Lunde et al., 2007). The candidate ER RNA-binding membrane proteins noted above lack known RNA-binding domains and thus, as has been proposed for the newly revealed RNA interactome members similarly lacking in canonical RNA-binding motifs, alternative mechanisms of protein-RNA recognition and association likely contribute to the landscape of RNA regulation in cells (Beckmann et al., 2015; Castello et al., 2012; Castello et al., 2015). Little is known, though, regarding the site or mechanism of RNA recognition by any of this newly discovered and large class of non-canonical RNA-binding proteins (Baltz et al., 2012; Castello et al., 2012; Gerstberger et al., 2014; Kwon et al., 2013; Pineiro et al., 2015). With the identification of new classes of multifunctional RNA-binding proteins, and emerging evidence that many previously identified RNA-binding proteins are multifunctional, a role for non-conventional modes of RNA localization to the ER can now be considered (Mangus et al., 2003; Markus and Morris, 2009; Sawicka et al., 2008; Turner and Hodson, 2012).

Recent study in the Nicchitta laboratory identified a ribosome- and translation-independent mRNA localization to the ER for mRNAs encoding endomembrane resident proteins (mRNA_{endo}) (Jagannathan et al., 2014). In this study, we used a sequential detergent fractionation approach to disrupt ribosome-ER membrane interactions in tissue culture cells, yielding the selective release of mRNAs encoding secretory proteins from the ER. In brief, the cells were digitonin-permeabilized to release the cytoplasmic mRNA pool and enable access to the ER compartment. The ribosome-ER membrane interactions were disrupted by Brij 35 to selectively release the ribosome-dependent mRNAs (Brij-sensitive fraction, BrS). The residual mRNAs using ribosome-independent ER-associations were released from the ER by DDM (dodecylmaltoside) (Brij-resistant fraction, BrR). Transcriptome-wide mRNA analysis of mRNA partitioning between BrS and BrR showed that the mRNA_{endo} were highly enriched in the BrR, suggesting that mRNA_{endo} were anchored to the ER in a ribosome-independent manner (Jagannathan et al., 2014).

At the time I started my research in the Nicchitta laboratory, however, little was known regarding the molecular mechanisms of mRNA localization and anchoring to the ER. Presumably, direct mRNA anchoring to the ER membrane would require RNA-binding proteins, which are known to serve diverse functions in RNA localization, translational regulation, and stability. Indeed, recent studies have identified ribosome

receptor RRBP1 (p180), a coiled-coil multidomain ER integral membrane protein, as a general ER-poly(A) mRNA-anchoring protein (Cui et al., 2012; Cui et al., 2013). p180 is reported to bind mRNAs nonspecifically, however, and so the question of how mRNAs undergo selective association with the ER membrane remains largely unanswered. Here I report the discovery of a family of ER integral membrane proteins function as RNA-binding proteins and which play important roles in the direct anchoring of mRNAs to the ER. Specifically, AEG-1 (LYRIC, metadherin), an ER integral membrane protein previously identified as an oncogene (Emdad et al., 2010; Hu et al., 2009; Lee et al., 2009; Yoo et al., 2009b) and as a candidate RNA-binding protein in a number of RNA interactome screens, functions as a selective mRNA-binding protein (Baltz et al., 2012; Castello et al., 2012; Jagannathan et al., 2014; Kwon et al., 2013). Genome-scale analyses of the AEG-1 RNA interactome by HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) and PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) revealed that AEG-1 serves a specific function in the localization of endomembrane resident protein transcripts to the ER. Notably, AEG-1 bound mRNAs are highly enriched in endomembrane protein transcripts ($p < 10^{-16}$), and intriguingly, integral membrane protein-encoding mRNAs ($p < 10^{-54}$). AEG-1 RNA interaction sites are highly enriched in coding regions and largely absent from untranslated regions (UTRs). In addition, the

translation of AEG-1 interacting transcripts is strongly and positively correlated with AEG-1 expression, independent of target transcript levels. These data reveal a novel mechanism for the direct anchoring of distinct cohorts of mRNAs to the ER and identify a critical role for AEG-1 in both RNA localization and integral membrane protein expression.

2. Materials and Methods

This section contains the materials and methods for all research, including UV-cross-link-based RNA-binding protein identification (Jagannathan et al., 2014), ER membrane-associated mRNP identification by proteomics (Jagannathan et al., 2014), the identification of non-canonical RNA-binding domain in AEG-1, and the identification and validation of AEG-1 bound RNAs by CLIP-Seq (UV-cross-linking and immunoprecipitation coupling with RNA-sequencing) (Darnell, 2010; Hafner et al., 2010; Konig et al., 2011) and RIP-qPCR (RNA immunoprecipitation and quantitative PCR) (Keene et al., 2006).

2.1 Cell lines and antibodies

AEG-1-14, PC-4 and AEG-1 KD cells were cultured at 37°C and 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM; Mediatech), supplemented with 10% fetal bovine serum (FBS; Gibco) as described (Yoo et al., 2009a). Immunoblotting and immunofluorescence staining were conducted by standard protocols using primary antibodies against HA (mouse monoclonal; 1:10,000; Thermo Fisher Scientific, #26183) ; AEG-1 (chicken polyclonal; 1:5,000) (Kang et al., 2005), PABP (rabbit polyclonal; 1:2,000; kind gift of Dr. Jack Keene, DUMC), α -tubulin (mouse monoclonal; 1:500; Iowa Hybridoma), NPC1 (rabbit polyclonal; 1:500; Novus Biologicals, #NB400-148). GAPDH (mouse monoclonal; 1:10,000; Sigma), Ribophorin I (rabbit polyclonal; 1:2,000; kind gift

of Dr. G. Kreibich, New York University). Sec61 α , TRAP α , CKAP4, and LRC59 antibodies were used as described (Jagannathan et al., 2014).

2.2 RNA cross-linking to the endoplasmic reticulum membrane proteins

2.2.1 Pre-condensation of Triton X-114

Triton X-114 (TX-114) (Sigma-Aldrich) was pre-condensed to remove hydrophilic impurities (Mathias et al., 2011). 20 g of TX-114 was dissolved in 980 mL condensation buffer containing 150 mM NaCl, 10 mM Tris HCl, pH 7.2, at 4 °C. For phase separation, the TX-114 solution was incubated overnight at 30 °C. The supernatant was discarded and replaced by same volume of condensation buffer. The phase separation was repeated twice. The pre-condensed TX-114 was stored at 4 °C. The concentration of pre-condensed TX-114 was 11.8% (w/v), as determined by UV absorbance at 274 nm with extinction coefficient of 2.51 (mg TX-114/g of solution)⁻¹cm⁻¹.

2.2.2 TX-114 phase partitioning of canine pancreas rough microsomes

The integral membrane protein fraction was prepared by either TX-114 phase partitioning or alkaline buffer extraction. The TX-114 phase partitioning protocol was modified from (Mathias et al., 2011). First, canine rough microsomes (RM) were incubated in a high salt/EDTA supplemented buffer (500mM KOAc, 15mM EDTA,

proteinase inhibitor cocktail (Sigma-Aldrich)) on ice for 30 min. Stripped RM were collected by centrifugation through a high salt sucrose cushion (0.5 M sucrose, 500mM KOAc, 15mM K-HEPES, pH 7.2, proteinase inhibitor cocktail) in the TLA-2 rotor (Beckman Instruments) at 250,000 g at 4 °C for 20 min. The RM pellet was resuspended in TX-114 lysis buffer (2% TX-114, 50mM NaCl, 10mM Tris-HCl, pH 7.2, 10mM MgCl₂, proteinase inhibitor cocktail) on ice for for 30 min with frequent vortexing. The insoluble material was discarded by centrifugation at 10,000 g at 4 °C for 10 min. The supernatant was collected and incubated at 30 °C for 10 min and centrifugation at 5,000 g for 10 min for phase separation. The upper aqueous phase (AP) and lower detergent rich phase (DP) were collected and adjusted to a final TX-114 concentration of 2% and the phase separation repeated. The proteins in AP and DP were TCA precipitated and analyzed by SDS-PAGE and immunoblot.

2.2.3 Alkaline extraction of canine pancreas rough microsomes

Alkaline extraction of canine pancreas rough microsomes was performed by a modification of the methods described in (Nicchitta and Blobel, 1993). First, RM was diluted 10-fold in alkali buffer (50 mM CAPS at pH 10.5) and incubated on ice for 30 min. The diluted RM was collected by centrifugation through a sucrose cushion (0.5M sucrose, 50mM KOAc, 15mM K-HEPES, pH 7.2, proteinase inhibitor cocktail) in the TLA-2 rotor at 250,000 g at 4 °C for 20 min. The RM pellet was resuspended in DDM lysis

buffer (1% DDM, 50mM NaCl, 10mM Tris-HCl, pH 7.2, 10mM MgCl₂, 10mg/ml proteinase inhibitor cocktail) on ice for 30 min with frequent vortexing. The insoluble material was discarded by centrifugation at 10,000g at 4°C for 10 min. The supernatant was collected, TCA precipitated and assayed by SDS-PAGE and western blot.

2.2.4 RNA-dependent T4 RNA ligase labeling of UV-irradiated rough microsomes

RMs were UV irradiated as described (Jagannathan et al., 2014). For RNA digestion, mung bean nuclease was added to a final concentration of 1U/μg RNA and incubated at 30 °C for 1 hr. RM were collected by centrifugation through a sucrose cushion in the TLA-2 rotor at 250,000 g at 4 °C for 20 min. The RM pellet was resuspended in DDM lysis buffer (1% DDM, 50mM NaCl, 10mM Tris-HCl, pH 7.2, 10mM MgCl₂, 10mg/ml proteinase inhibitor cocktail) on ice for 30 min with frequent vortexing. For T4 RNA ligase reaction, the RM lysate was adjusted to a final concentration of 50 mM Tris-HCl, pH 7.5, 10mM MgCl₂, 1mM DTT, 15% DMSO, 1mM ATP, 80μCi [5'-³²P] cytidine 3',5'-bis(phosphate), 1 U/μl T4 RNA ligase (New England Biolabs) at 4 °C. The lysate was collected, TCA precipitated and assayed by SDS-PAGE and phosphorimaging.

2.3 Proteomic analysis of canine rough microsome-associated mRNPs

Canine pancreas rough microsomes were dodecylmaltoside (DDM)-solubilized, and the polysome fraction was obtained by sucrose gradient centrifugation. Polysome fractions were pooled and concentrated by ultracentrifugation. Polysome pellets were then gently resuspended, and a binding control fraction was prepared by digestion of one-half of the sample with staphylococcal nuclease. Following the addition of 10 mM EDTA to all samples, to dissociate the ribosomal subunits and release mRNPs, poly(A) mRNPs were selected on Oligo(dT)-Cellulose Type 7 resin (GE Healthcare). Following extensive washing in EDTA-supplemented buffers, poly(A) mRNA-associated proteins were released by the addition of 10 mM ammonium bicarbonate, 15% dimethylformamide. Eluted proteins were concentrated by TCA precipitation and separated on 4–12% Bis-Tris gradient gels (Life Technologies), and in-gel digestion was performed as described (Wilm et al., 1996). Approximately one-half of each digest (5 μ l) was fractionated on a C18 column (Waters) using a gradient of 5–40% acetonitrile with 0.1% formic acid on a nanoAcquity liquid chromatograph (Waters). Electrospray ionization was used to introduce the sample in real-time to a Q-ToF Synapt G1 mass spectrometer (Waters), collecting data for each sample in data dependent acquisition (DDA) mode. Raw data were processed in Mascot Distiller (version 2.3) and searched in Mascot version 2.2 (Matrix Science) against the NCBI nr database with mammalian taxonomy. Scaffold (version 3.6.2; Proteome Software Inc.) was used to validate MS/MS-

based peptide and protein identifications. Measured peptide and protein level false discovery rate were both determined to be 0.05% using the target-decoy strategy. These identified proteins are limited to those proteins that were (1) identified in at least two of three independent replicates, (2) absent from the staphylococcal nuclease-treated controls, and (3) with high accuracy ($p < 0.05$, peptide count ≥ 2).

2.4 UV-cross-linking-based mRNA-binding assay of the ER membrane proteins

RM were UV-irradiated at 254-nm at 3.2 J/cm² (Stratalinker 24000) on ice. Irradiated RM were incubated in salt/EDTA buffer (0.5 M KOAc, 15 mM EDTA, 50 units/ml RNaseOUT, protease inhibitor mixture) for 30 min on ice, to release bound ribosomes, and subsequently collected by ultracentrifugation. The RM pellet was resuspended in equilibrium buffer (0.5% LiDS, 0.5 M LiCl, 10 mM Tris-HCl, pH 7.2, 5 mM DTT, 15 mM EDTA), heated to 65 °C for 5 min, incubated on ice, and cleared by centrifugation to remove insoluble materials. mRNA was isolated from the supernatant by Oligo(dT)-Cellulose Type 7 resin (GE Healthcare) chromatography with extensive washing in the equilibrium buffer. RNA-associated proteins were subsequently eluted by incubating the resin in elution buffer (10 mM Tris-HCl, pH 7.2, 1 mM EDTA, 0.1% Triton X-110 with 1:50 (v/v) RNase A/T1 Mix (Thermo Scientific)) at room temperature for 30 min. Eluted proteins were concentrated by TCA precipitation, resolved by SDS-PAGE, and detected by immunoblot analysis.

2.5 UV cross-linking and immunoprecipitation (CLIP) for AEG-1 and truncation mutants

Cells were UV-irradiated at 254-nm at 400 mJ/cm² (Stratalinker 24000) on ice, lysed in CLIP lysis buffer (50mM Tris-HCl, pH 7.2, 100 mM NaCl, 1 mM EDTA, pH 8, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS), DNase- and RNase-digested (Turbo DNase and RNase I; Thermo Fisher Scientific), and subsequently immunisolated by anti-HA antibody conjugated G protein Dynabeads at 4 °C overnight. The beads were washed with high-salt buffer (50mM Tris-HCl, pH 7.2, 1M NaCl, 1 mM EDTA, pH 8, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) and PNK buffer (20mM Tris-HCl, pH 7.2, 10 mM MgCl₂, 0.2% Tween-20). Protein-crosslinked RNA moiety was radioisotope-labeled by addition of 0.5U/μl of polynucleotide kinase (New England Biolabs) and 0.5 μCi/μl [³²P]- ATP in PNK buffer. Protein-RNA complexes were eluted by incubation with HA peptide and analyzed by immunoblotting and phosphorimaging.

2.6 Sucrose cushion centrifugation and polysome profiling

Sucrose cushion centrifugation and polysome profiling were performed as described (Jagannathan et al., 2014). For the sucrose cushion centrifugation assay, cells were treated with 50ug/ml cycloheximide (CHX) at 37 °C for 10 min, washed with ice-cold PBS, incubated with 50ug/ml CHX/DPBS on ice for 20 min, and lysed in sucrose

cushion lysis buffer [200mM KCl, 25 mM KHEPES, pH 7.2, 50 mM EDTA, 1mM DTT, 2% n-dodecyl- β -D-maltoside (DDM), and protease inhibitor cocktail] on ice for 15min. The lysates were cleared by centrifugation at 14,000 rpm for 15 min. Supernatants were treated with/without RNase Cocktail (Thermo Fisher Scientific) on ice for 30 min and loaded on top of sucrose cushion (500 mM sucrose, 200 mM KCl, 25 mM KHEPES, pH 7.2, 2 mM EDTA, 1mM DTT) for centrifugation at 90,000 rpm, 4 °C, 15min (TLA 100.2). Pellets were analyzed by SDS-PAGE and immunoblotting. For polysome profiling, cells were harvested as described above. The harvested cells were lysed in polysome lysis buffer (200mM KCl, 25 mM KHEPES, pH 7.2, 10 mM MgCl₂, 1mM DTT, 2% DDM, RNaseOUT, and protease inhibitor cocktail) on ice for 15min. The lysates were centrifuged at 14,000 rpm for 15 min to remove cell debris. Polyribosomes were resolved on 15–50% sucrose gradients and fractionated as previously described (Stephens and Nicchitta, 2007, 2008). The fractions were analyzed by SDS-PAGE and immunoblotting.

2.7 CLIP-Seq protocols

CLIP-Seq protocols were performed on AEG-1-14 and PC-4 cells as described above. In brief, the cells were UV cross-linked at 265nm for HITS-CLIP and 365nm for PAR-CLIP and lysed. Protein-RNA complexes were partially RNase I-digested, and immunoprecipitated with anti-HA antibody. The RNA moiety was dephosphorylated using 0.5 U/ μ l of calf intestinal alkaline phosphatase (New England Biolabs) in

dephosphorylation buffer (50 mM Tris-HCl, pH 8, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT) at 37 °C for 25 min and 5'-end radioisotope-labeled using 0.5 U/μl of polynucleotide kinase (New England Biolabs) and 0.5 μCi/μl [γ -³²P]ATP in PNK buffer supplemented with 1 mM DTT. The complexes were subsequently eluted by 2X sample buffer at 95 °C for 5 min, resolved by SDS-PAGE, transferred to nitrocellulose membranes, and analyzed by phosphorimaging. RNA fragments were extracted by proteinase K digestion from the nitrocellulose membranes and purified by phenol/chloroform extraction. cDNA libraries were constructed according to the manufacturer's instructions (New England Biolabs; E7330). For the libraries, insert lengths were checked by Bioanalyzer DNA 1000 Kit (Agilent Technologies). As a quality control step, a fraction of the library was transformed into E. coli and clones sequenced by Sanger sequencing, to confirm insert size and sequence. Libraries were then sequenced on the Illumina HiSeq with 50-nucleotide single-read runs.

2.8 Data analysis for CLIP-Seq

2.8.1 Establish reference transcriptome and read mapping

A reference transcriptome was generated based on RNA-Seq data from the ENCODE project (GEO GSM958740) using Tophat and Cufflinks (Trapnell et al., 2009; Trapnell et al., 2012). CLIP-Seq reads were adaptor removed using Cutadapt (Trapnell et al., 2010), mapped to the reference transcriptome using Bowtie with one mismatch

tolerance per read.

2.8.2 Cluster identification

The mapped reads were grouped as clusters, with cluster defined as an overlapping set of reads that contains at least six unique reads and at least one crosslinking elicited mismatch (any point mutation for HITS-CLIP, T-to-C conversion for PAR-CLIP). A cluster score was generated for each cluster according to the numbers of unique reads and mutations, which was defined as $\log_2(1000 * (1-D) * M / R)$, where M is the number of point mutations in the cluster and R is the abundance of the RNA. D represents the breadth of read distribution, defined as $(1/N)^N$, where N is the number of unique read starts or stops in the cluster. Clusters with scores of less than 0 were discarded.

2.8.3 Cluster analysis

For cluster positional analysis, total cluster score was summed relative to the start and stop codons, as well as transmembrane domains. AEG-1 bound mRNAs were defined as mRNAs that were identified in all three CLIP-Seq cluster data sets. Gene Ontology analyses were performed with the DAVID tools (Huang et al., 2009a, d). The positions of transmembrane domains were predicted using TMHMM (Krogh et al., 2001).

2.9 RNA immunoprecipitation (RIP) and quantitative PCR

RNA immunoprecipitation was performed to AEG-1-14 and PC-4 cells as described (Keene et al., 2006). In brief, the cells were lysed in RIP lysis buffer (100 mM KCl, 10 mM KHEPES, pH 7.2, 5 mM MgCl₂, 1mM DTT, 400 μM VRC, 1 mM PMSF, 10 mM NaF, 2 mM sodium orthovanadate, 2 mM β-glycerophosphate, 0.5% IGEPAL CA-630, RNaseOUT, and protease inhibitor cocktail) on ice for 10 min and centrifuged at 14,000 rpm for 15 min to remove cell debris. The supernatants were diluted with NT2 buffer (150mM NaCl, 1mM MgCl₂, 50mM Tris-HCl, pH 7.4, 0.05% IGEPAL CA-630), EDTA to 15mM, DTT to 1mM, RNaseOUT and VRC, and then immunoprecipitated at 4 °C overnight. The samples were washed with ice-cold NT2 buffer supplemented with 15mM EDTA for five times and subsequently eluted by TRIsure according to the manufacturer's instructions (Bioline). The isolated RNA was reverse-transcribed and analyzed by quantitative PCR (qPCR). The results were shown as relative fold enrichments of the qPCR signals of AEG-1-14 cells over that of PC-4 cells.

3. Identification of RNA-binding ER membrane proteins

3.1 Introduction

Recently we demonstrated that the subset of mRNAs encoding endomembrane resident proteins is anchored on the ER membrane via a ribosome-independent mechanism (Chen et al., 2011; Jagannathan et al., 2014). However, little is known regarding the molecular mechanism of this ribosome-independent mRNA-ER interaction. Given the well-established roles of RNA-binding proteins in RNA localization, the potential RNA-binding activity of the ER transmembrane proteins was evaluated. In this Chapter, I report on experiments where UV-cross-linking, a protein-nucleic acid specific cross-linking approach, was used to study protein - RNA interactomes on the ER membrane. Proteomic analysis revealed the identities of the candidate mRNA-binding ER membrane proteins. To minimize contaminations from RNA-binding proteins which were not on the ER membrane, canine pancreas rough microsomes (RM), a very highly enriched source of ER membranes, were used for the identification of mRNA-binding ER membrane proteins.

3.2 Identification of RNA-binding ER integral membrane proteins

UV-crosslinking, a relatively unbiased 'zero-length' protein-nucleic acid cross-linking method, was used to assess the diversity of RNA-binding integral ER membrane proteins (Castello et al., 2012; Pashev et al., 1991). As depicted in Figure 2, RM were UV-

irradiated to generate *in vivo* cross-link protein-RNA complexes, peripheral proteins released by salt/EDTA or alkali extraction (Nicchitta and Blobel, 1993), and the 'stripped' RM collected by ultracentrifugation. The irradiated, stripped RM was subsequently digested with mung bean nuclease to digest RNA, yielding short RNA fragments bearing free 3' OH group in the complexes. In contrast to alkaline buffer extraction, which releases both peripheral and luminal proteins, the salt/EDTA washed RM contain substantial quantities of luminal proteins. To focus the study on integral membrane proteins, multiple rounds of Triton X-114 partitioning were conducted on the salt/EDTA stripped RM (Bordier, 1981; Mathias et al., 2011), providing two independent methods for the ER integral membrane protein enrichment. Subsequently, cross-linked, digested RNA fragments in the membrane protein enriched fractions were radioisotope-labeled by addition of T4 RNA ligase, ATP and [³²P] cytidine 3'-5' bisphosphate (Kikuchi et al., 1978).

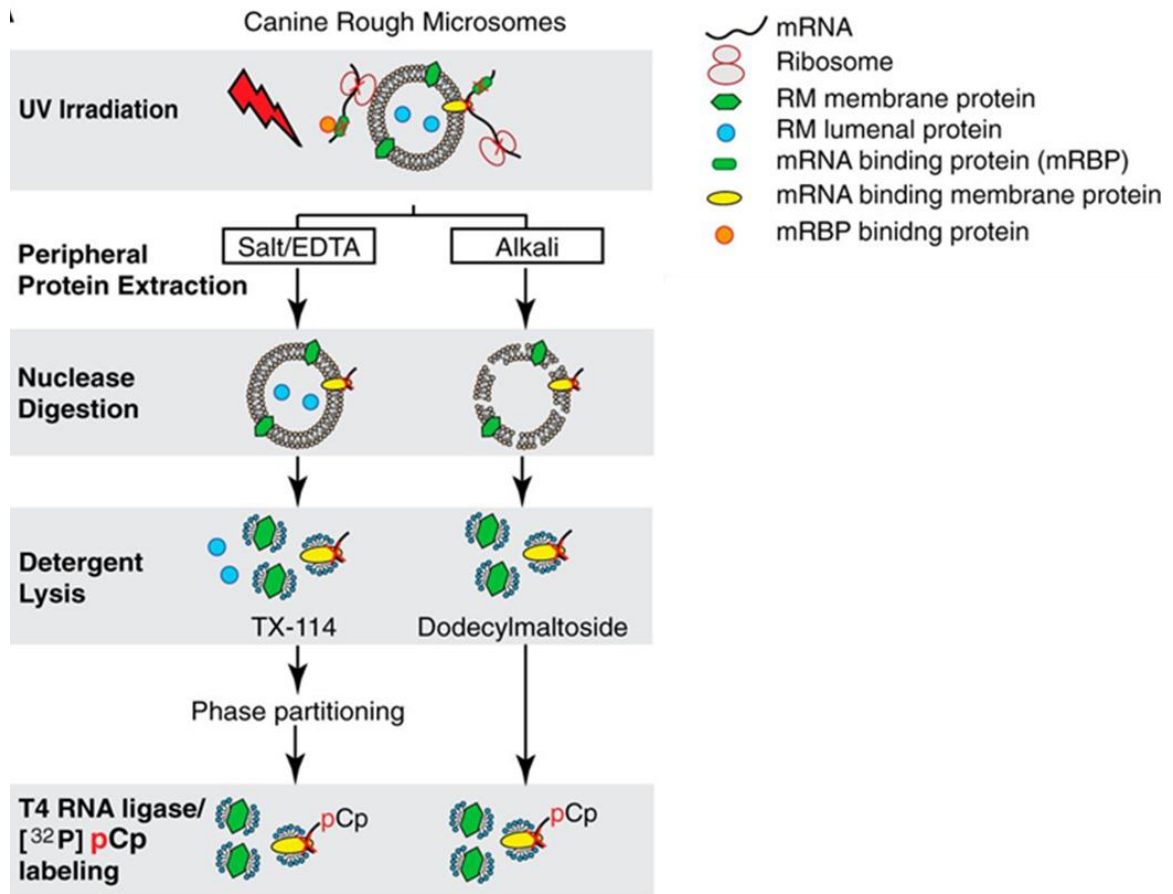


Figure 2: Identification of RNA-binding proteins on the ER membrane.

Schematic procedure for the identification of the RNA-binding activity of ER integral membrane proteins by UV-cross-linking and T4 RNA liagse/[³²P]pCp labeling assay. Rough microsomes (RM) was UV-irradiated, membrane protein isolated by either salt/EDTA wash coupled with Triton-X114 partitioning or alkali wash. Protein-RNA complexes were RNA-specific radiolabeled by T4 RNA liagse/[³²P]pCp incubation. The protein-RNA complexes were detected by phosphoimaging analysis.

The enrichments of ER integral membrane proteins using the Triton X-114 partitioning and alkali-extraction protocols were revealed by Coomassie Blue staining and immunoblot analysis. Coomassie Blue staining was conducted with either membrane equivalents, to show the abundance of ribosomal and luminal proteins and

small fraction of integral membrane proteins, or protein equivalents, to illustrate the ER membrane proteome (Figure 3A&B). The results show that Triton X-114 partitioning and alkali-extraction protocols share a similar membrane protein composition (lane DP vs. P; Figure 3B). The membrane protein enrichments of the protocols were further validated by immunoblot analysis against the ER luminal proteins (GRP94 and BiP) and the ER membrane proteins (ribophorin I, Sec61 α , and TRAP α), showing exceptional enrichments of the ER membrane proteins in the membrane fractions (Figure 3C).

The RNA-binding activity of the ER membrane proteins was evaluated by [³²P] cytidine 3'-5' bisphosphate/T4 RNA ligase method (Figure 3D). The RNA moiety in the RNA-protein complexes was radioisotope-labeled and the complexes were SDS-PAGE separated and phosphorimaged. The labeling reaction was entirely T4 RNA ligase-dependent (lane 1 vs. 2), showing the lack of non-specific radioisotope-labeled background. SDS-PAGE/phosphorimager analysis of the [³²P] pCp/T4 RNA ligase-labeled membrane fractions from the two protocols identified multiple radioisotope-labeled protein-RNA complexes ranging from 15-70 kD (Figure 4D, lanes 3 and 4). These data clearly demonstrate RNA binding activity for a diversity of ER membrane proteins. Because the [³²P] pCp/T4 RNA ligase-labeling protocol does not distinguish RNA-binding selectivity for rRNA, tRNA, mRNA, and non-translated RNAs, complementary experimental approaches were developed to identify candidate ER integral membrane proteins functioning in mRNA binding/anchoring.

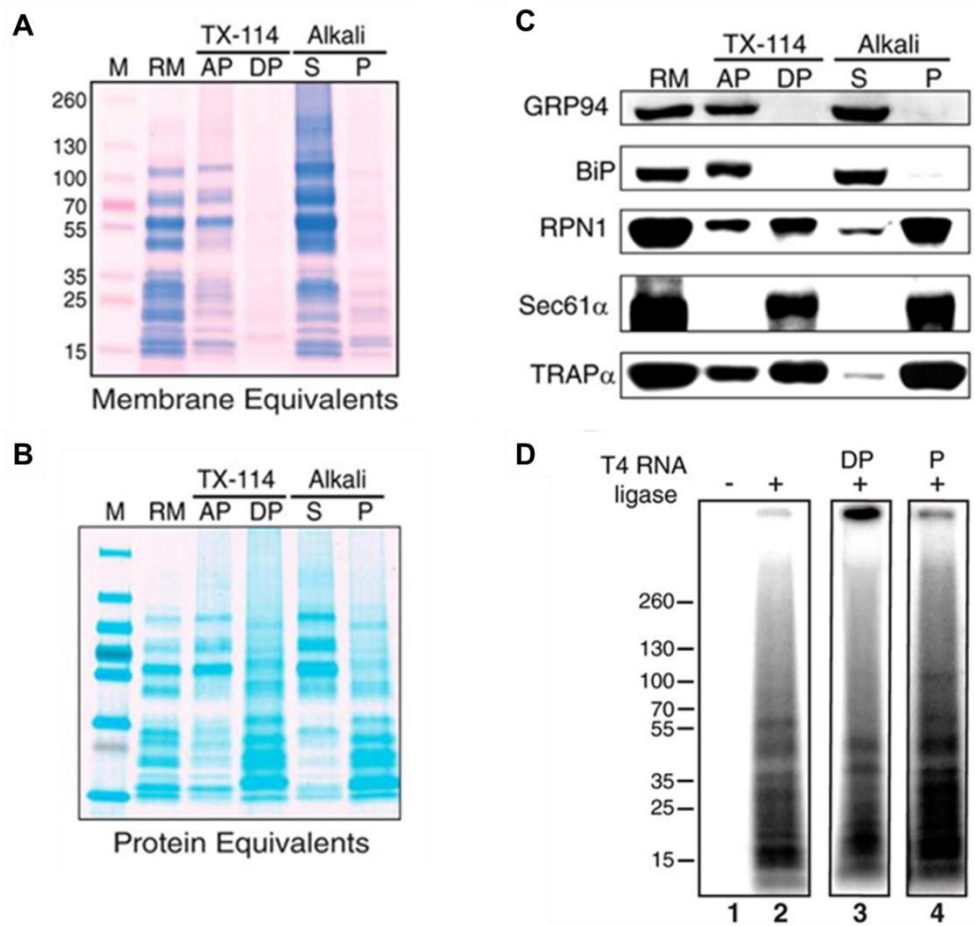


Figure 3: UV-cross-linking reveals endoplasmic reticulum membrane proteins functioning in RNA-binding.

Images of Coomassie Blue-stained SDS-PAGE show the membrane protein enrichments obtained by Triton X-114 partitioning (TX-114) or alkali extraction (Alkali), with (A) protein derived from identical quantities of rough microsome (membrane equivalents) or (B) protein loading with identical quantities of total protein (protein equivalents). RM, rough microsome; AP, aqueous phase; DP, detergent-rich phase; S, alkali-releasable proteins; P, alkali-resistant fraction. (C) Immunoblot analysis for the aqueous (AP) or detergent-rich phase (DP) of the Triton X-114-extracted RM or the soluble (S) and membrane protein (P) fractions of alkali-extracted RM against ER luminal proteins (GRP94 and BiP) and ER membrane proteins (ribophorin I (RPN1), Sec61 α , and TRAP α). (D) Phosphoimage of radioisotope-labeling via the [32 P]pCp/T4 RNA ligase protocol is ligase-dependent (lane 2) and identifies RNA-binding ER membrane proteins common to Triton X-114 (TX-114; lane 3) and alkali-extracted (lane 4) RM.

3.3 Proteomic analyses of candidate ER-mRNA-anchoring proteins

Given the strong evidence for multiple RNA-binding proteins on the ER membrane, oligo(dT) chromatography/proteomic analysis was conducted as a complementary approach to further identify mRNA-binding ER membrane proteins (Figure 4). First, rough microsome-associated polysomes were isolated by polysome profiling and subsequently incubated with 10 mM EDTA to dissociate the ribosomes. mRNA was further purified by oligo(dT) chromatography in the presence of 10 mM EDTA (Longuet et al., 1979). mRNA-associated proteins were eluted in low salt buffer and detected by protein mass spectrometry (Table 1). A nuclease-digested polysome was processed in parallel as a negative control for background binding.

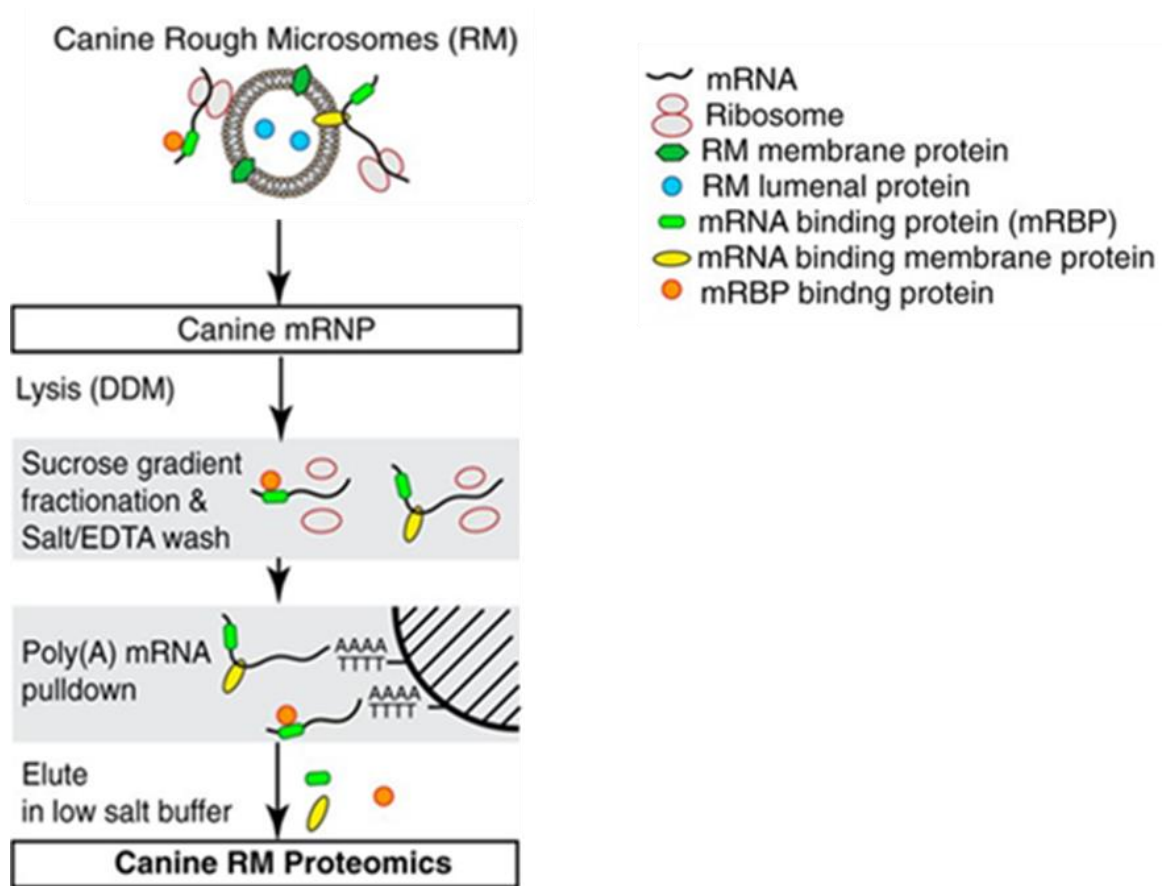


Figure 4: Proteomic Analyses of Candidate ER-mRNA-anchoring Proteins.

Schematic procedure for the identification of the mRNA-binding ER membrane proteins by native oligo(dT) chromatography and proteomic analysis. Canine rough microsome was lysed in DDM buffer. Polysomes were isolated by sucrose gradient fractionation and subsequently dissociated by 10mM EDTA incubation. Poly(A) mRNA was further isolated by oligo(dT) chromatography. mRNA-binding proteins were eluted in low salt buffer. The protein composition in the eluate was interrogated by protein mass spectrometry.

The protein composition of mRNPs was resolved by protein mass spectrometry analysis. In total, 170 proteins were identified as potential mRNA-binding proteins, including 21 ER integral membrane proteins (Table 1). Notably, even though 10 mM

EDTA was used in the purification process to dissociate the ribosomes, 26 ribosomal proteins were identified. The results are consistent with a previous study in HeLa cells where 18 of the identified ribosomal proteins were characterized as mRNA-binding protein, further suggesting direct mRNA interactions for the identified ribosomal proteins (Castello et al., 2012). Therefore, this study provides a complementary mRNA-binding protein identification focusing on the mRNA-binding proteins interacting with the ER membrane-localized mRNAs.

The candidate mRNA-binding ER membrane proteins have a variety of cellular functions (Table 1), including putative RNA/ribosome binding (p180), protein translocation (Sec61 α , Sec61 β , Sec63, TRAM1, TRAP α), mRNA targeting to the ER (SRPR, SRPRB), ER cytoskeletal binding (CKAP4, KTN1), signal peptide removal (SPCS2, SPCS3) and protein glycosylation (RPN1, RPN2, OST48, STT3A). Given the well-established functions of the ER membrane proteins, the identification as mRNA-binding proteins further suggests multifunctional roles of the ER membranes. Of the identified mRNA-binding ER integral membrane proteins, many of these proteins were independently identified as mRNA-binding proteins in HeLa cells, HEK293 cells, and mouse embryonic stem cells (mESC) (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). These mRNA-binding protein screens further support the potential mRNA-binding activity of the identified ER membrane proteins.

Table 1: Summary of identified mRNA-binding ER integral membrane proteins

mRNP	Name
AEG-1	Astrocyte elevated gene-1 (LYRIC; MTDH)
ALG5	Dolichyl-phosphate beta-glucosyltransferase
CKAP4	Cytoskeleton-associated protein 4 (Climp-63; p63)
DGAT1	Diacylglycerol O-acyltransferase 1
LRC59	Leucine-rich repeat-containing protein 59 (LRRC59)
MOGS	Mannosyl-oligosaccharide glucosidase
OST48	Oligosaccharyl transferase 48 kDa subunit
PREB	Prolactin regulatory element-binding protein
PTSS1	Phosphatidylserine synthase 1
RPN1	Ribophorin 1
RPN2	Ribophorin 2
RRBP1	Ribosome-binding protein 1 (p180)
S61A1	Sec61 subunit alpha isoform 1 (Sec61 α)
SC61B	Sec61 subunit beta (Sec61 β)
SPCS2	Signal peptidase complex subunit 2
SPCS3	Signal peptidase complex subunit3
SRPR	Signal recognition particle receptor subunit beta (SR β)
SRPRB	Signal recognition particle receptor subunit alpha (SR α)
SSRA	Translocon-associated protein subunit alpha (TRAP α)
STT3A	Oligosaccharyl transferase subunit STT3A
TRAM1	Translocating chain-associated membrane protein 1

3.4 Validation of RNA-binding activity for potential mRNA-anchoring proteins

A complementary UV-cross-linking assay was conducted to validate the candidate mRNA-binding ER membrane proteins (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). First, protein-mRNA complexes on the rough microsomes were UV-cross-linked, isolated by oligo(dT) chromatography under stringent condition (0.5 M LiCl and Li-SDS). mRNA-binding proteins were selectively eluted by RNase

digestion and detected by immunoblot analysis (Figure 5). In this study, poly(A)-binding protein (PABP), a universal mRNA-binding protein, was used as a positive control for mRNA-binding activity (Figure 5B). The UV-cross-linking assay validated the poly(A) mRNA-binding activity for three of candidate membrane proteins (Sec61 α , Sec61 β , and ribophorin I) (Figure 5B). Notably, although CKAP4 and LRC59 were identified as mRNA-binding protein in the previous mRNA interactome screens (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013), neither of these two proteins were detected in this study (Figure 5B).

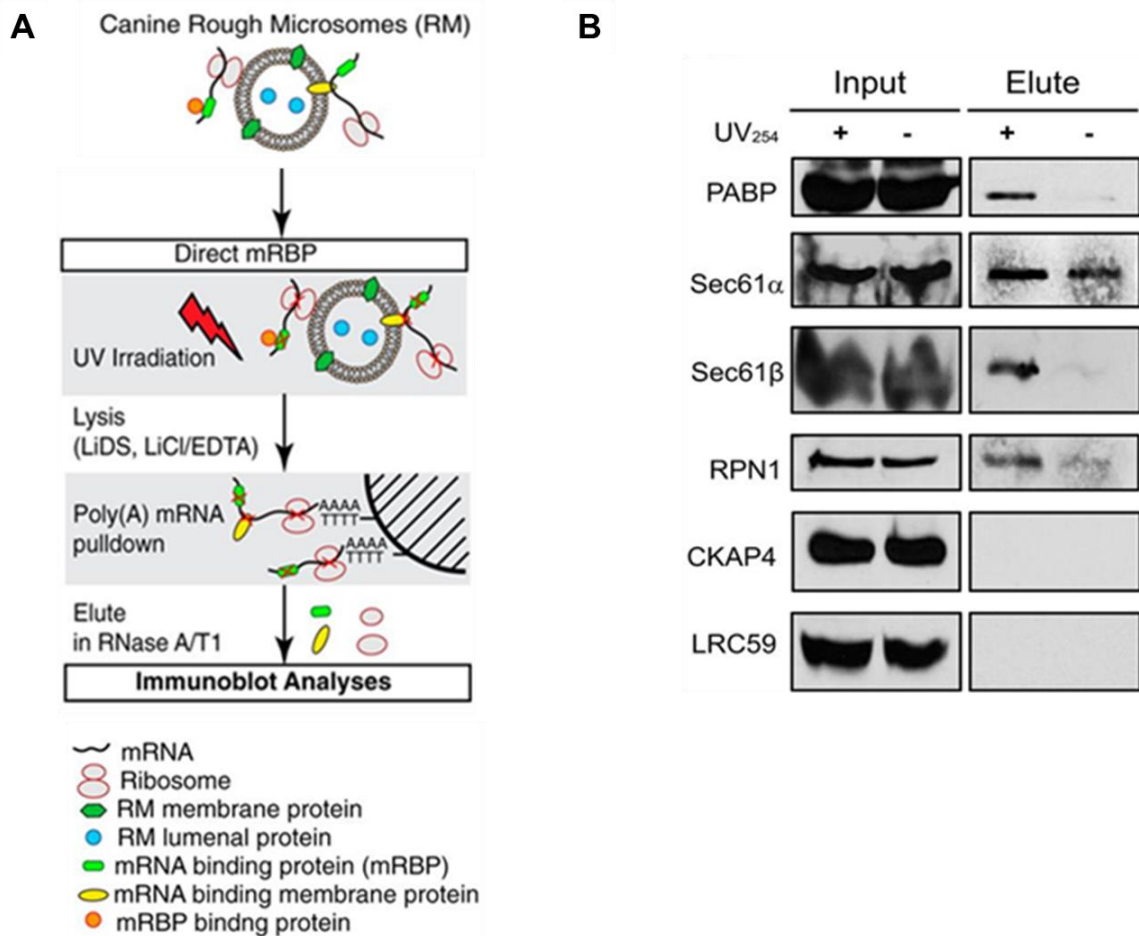


Figure 5: UV-cross-linking analysis validates the mRNA-binding activity of the candidate ER membrane proteins.

(A) Schematic procedure for the validation of the mRNA-binding activity of the candidate ER membrane proteins by UV-cross-linking and oligo(dT) chromatography. Canine rough microsome was UV-irradiated to cross-link protein-mRNA complexes and lysed in LiDS buffer. Complexes were isolated by oligo(dT) chromatography. mRNA-binding proteins were eluted by addition of RNase A/T1. (B) The protein composition in the RNase eluate was interrogated by immunoblot analysis against the ER membrane proteins (Sec61 α , Sec61 β , and ribophorin I (RPN1), CKAP4 and LRC59) and poly(A)-binding protein (PABP) as an mRNA-binding protein control.

4. The RNA-binding activity and selectivity of AEG-1

4.1 Introduction

Recent transcriptome-wide analysis revealed a subset of mRNAs directly anchored on the endoplasmic reticulum (ER) membrane (Jagannathan et al., 2014). In Chapter 3, proteomic analysis identified putative mRNA-binding proteins on the ER membrane serving as mRNA anchors (Table 1). Three independent proteomic studies in mRNA-binding protein identification have also identified putative mRNA-binding integral membrane proteins on the ER in HeLa cells, HEK293 cells, and mouse embryonic stem cells (mESC) (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). To identify conserved mRNA-binding proteins on the ER membrane, a list of putative mRNA-binding ER membrane proteins was summarized in Figure 6 (Baltz et al., 2012; Castello et al., 2012; Jagannathan et al., 2014; Kwon et al., 2013). The Venn diagram shows that AEG-1 and CKAP4 were identified in all mRNA interactome screens (Figure 6). CKAP4 is a cytoskeleton-binding protein regulating intramembrane organelle morphology (Sandoz and van der Goot, 2015). AEG-1 was identified as an overexpressed oncogene in hepatocellular carcinoma, melanoma, breast cancer and malignant glioma, implicated in HIF-1 α mediated angiogenesis, metastasis, and chemotherapy (Robertson et al., 2015a). AEG-1 has previously been identified as a potential RNA-binding protein interacting with multiple RNA-binding proteins and ribosomal proteins in a nuclease-sensitive manner (Meng et al., 2013). In this Chapter,

we focused on the RNA-binding activity and selectivity of AEG-1, and further identified a putative non-canonical RNA-binding domain in AEG-1.

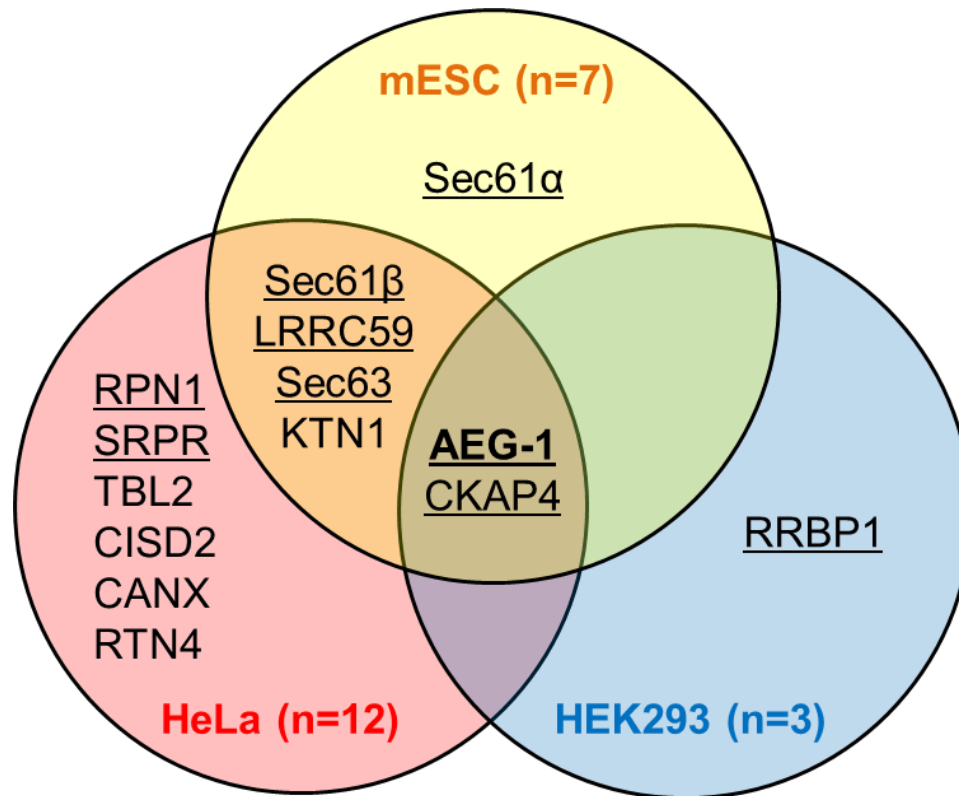


Figure 6: RNA interactome screens identified mRNA-binding integral membrane proteins on the endoplasmic reticulum.

Venn diagrams of ER integral membrane mRNA-binding proteins identified in HeLa, HEK293 and mouse embryonic stem cells (mESC) (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). The mRNA-anchoring proteins identified in canine mRNP in Table 1 are underlined. n, number of identified ER integral membrane mRNA-binding proteins in the indicated cell line.

4.2 AEG-1 is a RNA-binding ER membrane protein

AEG-1 is an ER resident membrane protein (Meng et al., 2012). The subcellular localizations of endogenous and epitope-tagged AEG-1 proteins were analyzed using immunofluorescence staining study (Figure 7). The immunofluorescence staining patterns of both endogenous and epitope-tagged AEG-1 mirrored that of the ER resident membrane protein TRAP α , indicating an ER localization for both endogenous and epitope-tagged AEG-1 (Figure 7).

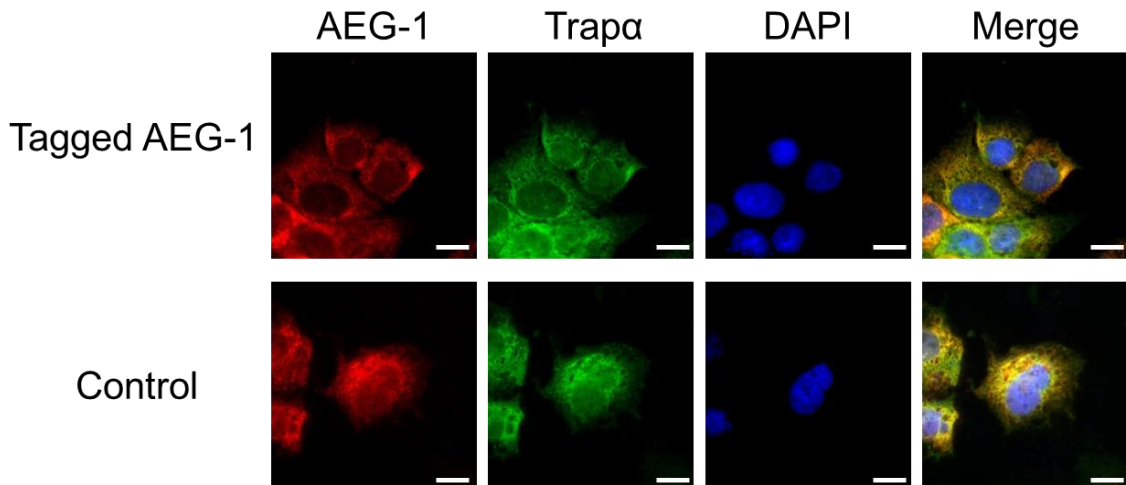


Figure 7: Subcellular localization of endogenous and epitope-tagged AEG-1 proteins. Immunofluorescence staining analysis of TRAP α (ER resident membrane protein), AEG-1, and DAPI (nucleus) distributions in AEG-1-14 (Tagged AEG-1) and PC-4 (control) cells. Scale bar = 10 μ m.

AEG-1 has previously been reported to associate with multiple RNA-binding proteins, ribosomal proteins, and translation factors (Meng et al., 2012; Yoo et al., 2011). Notably, some of these protein-protein interactions are nuclease-sensitive (Meng et al.,

2012), indicating a DNA/RNA-mediated indirect protein-protein interaction for AEG-1 and its binding partners. To further confirm these indirect interactions, the interaction of AEG-1 and poly(A)-binding protein (PABP), an RNA-binding protein common to all poly(A) mRNAs was examined by co-immunoprecipitation and RNase digestion (Figure 8). Co-immunoprecipitation against HA-epitope tag was performed to isolate epitope-tagged AEG-1. As shown in Figure 8, PABP was immunoaffinity isolated with epitope-tagged AEG-1. However, RNase incubation substantially reduced the level of co-isolated PABP, indicating a RNA-dependent, indirect interaction between AEG-1 and PABP. Since PABP binds to all poly(A) mRNA, the RNase-sensitive interaction indicates that AEG-1 may bind poly(A) mRNA to interact with RNA-binding proteins.

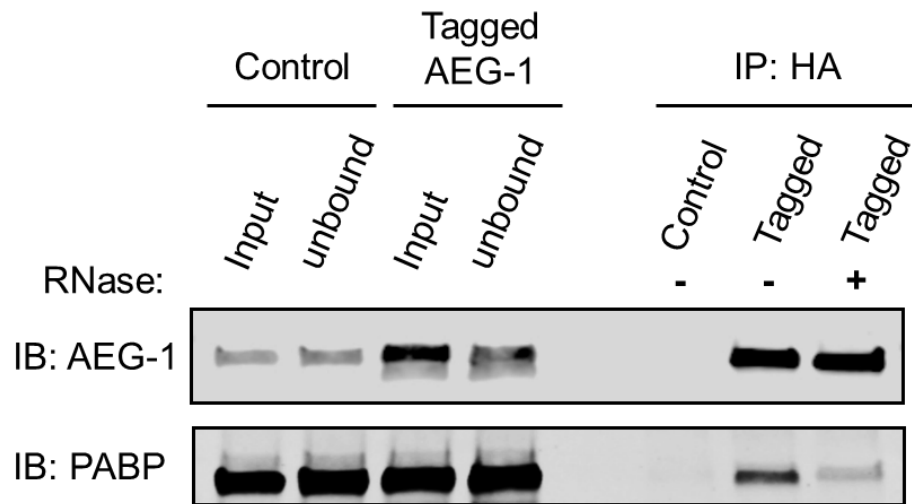


Figure 8: Co-immunoprecipitation/RNase digestion of poly(A)-binding protein (PABP) with HA-tagged AEG-1.

AEG-1-14 (Tagged AEG-1) and PC-4 (Control) cells were lysed, immunoprecipitated against HA-epitope tag, RNase digested, and subsequently analyzed by immunoblot analysis with anti-AEG-1 and anti-PABP antibodies.

To further confirm the RNA-binding activity of AEG-1 in living cells, the formation of AEG-1 and RNA complexes were examined by UV crosslinking and immunoprecipitation (CLIP) (Figure 9). As depicted in Figure 9A, Epitope-tagged AEG-1-transfected cells were UV-irradiated to covalently cross-link protein-RNA complexes. Complexes were RNase digested and immunoaffinity isolated. RNA fragments in the complexes were labeled with T4 polynucleotide kinase/[γ -³²P] ATP. A no UV irradiation sample was used as negative control.

The phosphoimage results show that the [γ -³²P] ATP-labeling reaction was strictly UV irradiation-dependent. Notably, the extensive RNase digestion reduced the diversity of RNA lengths in protein-RNA complexes, leading to a uniform protein-RNA complex in size (Figure 9B) (Konig et al., 2011). Thus, the radioisotope signal of RNA-protein complex at approximate 80 kDa is corresponding to the molecular weight of AEG-1 protein on the immunoblot (Figure 9B), indicating that AEG-1 binds RNA in living cells.

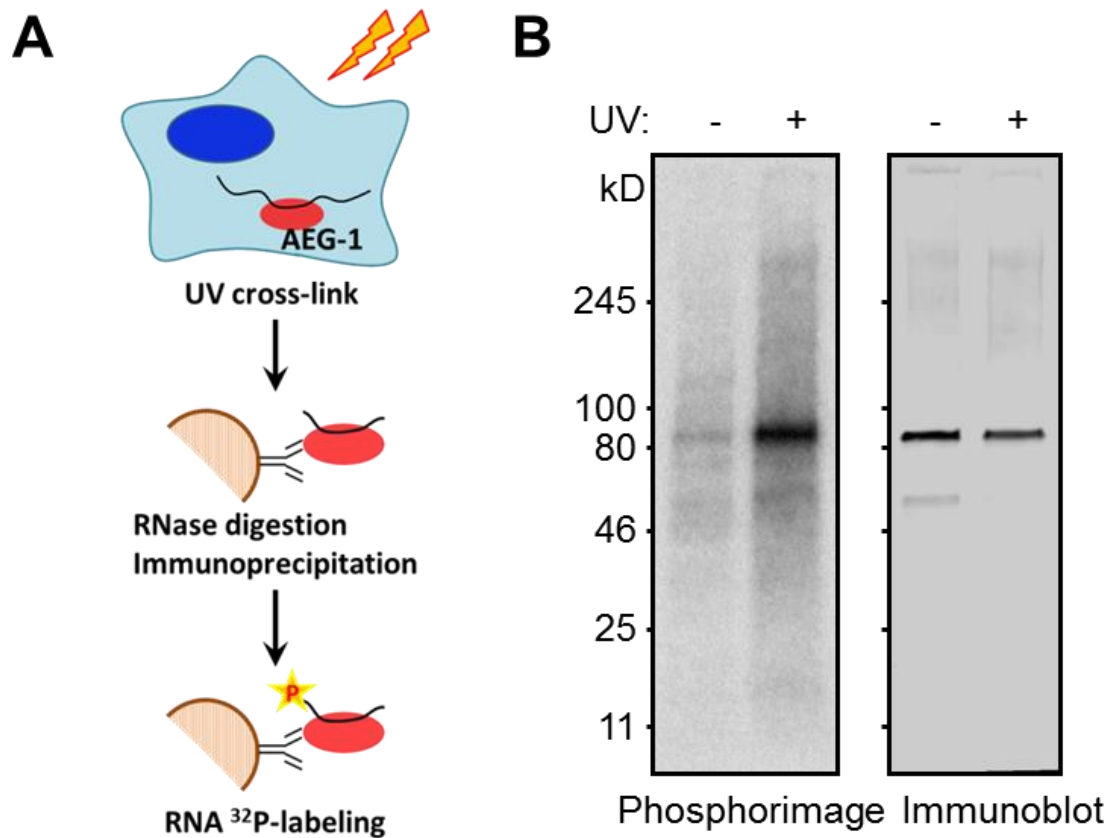


Figure 9: AEG-1 binds RNA in living cells.

(A) Schematic procedure of UV cross-linking and immunoprecipitation (CLIP) assay. Epitope-tagged AEG-1 transfected cells were UV₂₅₄ irradiated, lysed, and RNase digested. Epitope-tagged AEG-1 was immunoaffinity isolated. RNA fragments were labeled with T4 polynucleotide kinase/[γ -³²P] ATP. The AEG-1/RNA complexes were subjected to SDS-PAGE and transferred to nitrocellulose membranes. The protein and RNA in the complexes were estimated by immunoblotting and phosphorimaging. **(B)** Left, a representative phosphorimage depicting AEG-1-RNA complex formation by UV cross-linking. Right, immunoblot analysis of epitope-tagged AEG-1 for estimating sample loading.

4.3 AEG-1 binds actively translated mRNA

Having demonstrated an *in vivo* RNA-binding activity for AEG-1, further biochemical assays were performed to study its RNA selectivity. Given that recent

studies have shown that AEG-1 associates with multiple mRNA-binding proteins, including ribosomal proteins and translation factors (Meng et al., 2012; Yoo et al., 2011), AEG-1 could associate with actively translated mRNA via directly and/or indirectly interactions. To study the RNA-binding selectivity of AEG-1 for actively translated mRNAs, the formation of AEG-1 and mRNA complexes were examined by RNase digestion and ultracentrifugation (Figure 10). As depicted in Figure 10A, mRNA-binding proteins were isolated with polysomes in ribosome pellet fraction by ultracentrifugation. RNase digestion was used to degrade RNA and dissociate mRNA-binding proteins from the polysomes. In Figure 10B, immunoblotting analyses were performed to determine the mRNA-binding activity of AEG-1. Poly(A)-binding protein (PABP), a universal mRNA-binding protein, and RPL17, a ribosomal protein, were used as controls for mRNA binding and ribosome/rRNA binding, respectively. The results show that AEG-1 and PABP was recovered in the ribosome pellet fraction in the no RNase digestion condition, but the levels of proteins were substantially reduced in the RNase-digested samples, indicating that AEG-1 interacts with actively translated mRNA rather than ribosome/rRNA (Figure 10B).

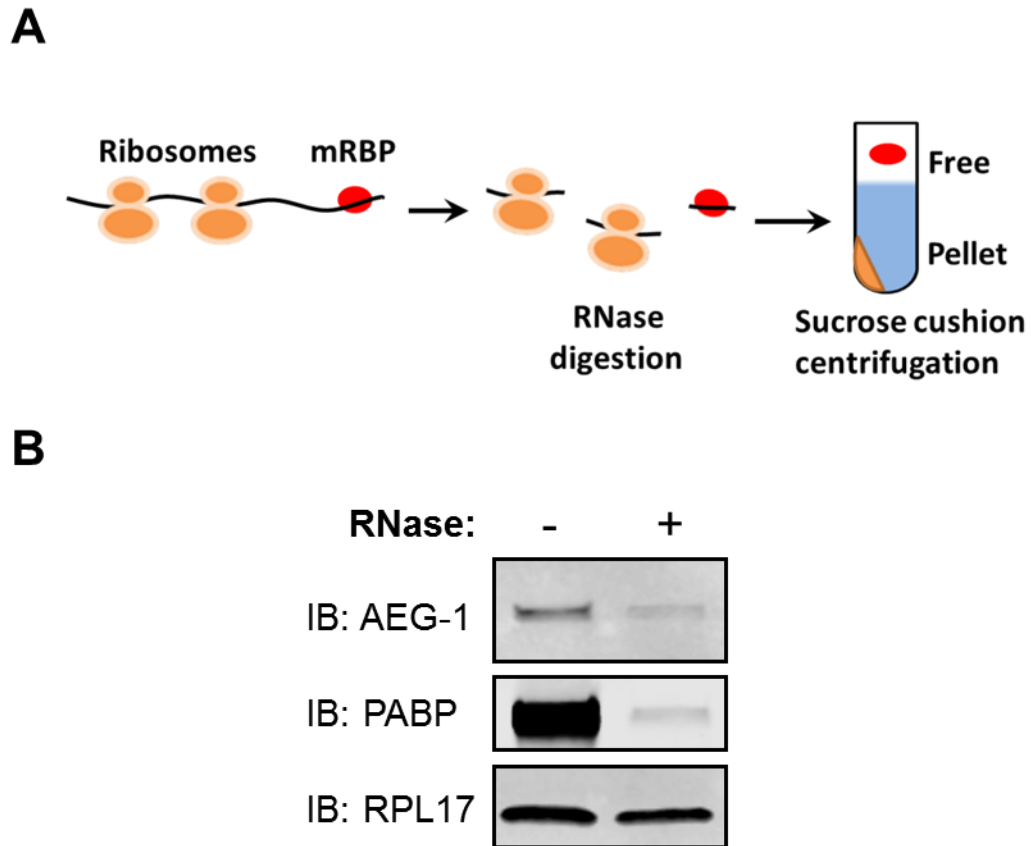


Figure 10: AEG-1 binds translating mRNA.

(A) Schematic procedure of RNase digestion/ultracentrifugation assay for studying the mRNA-binding activity of AEG-1. Cell lysate was incubated with/without RNase digestion and subjected to sucrose cushion ultracentrifugation. The ribosome pellet fraction (pellet) was analyzed by immunoblot. (B) Immunoblot analysis of AEG-1, poly(A)-binding protein (PABP; mRNA-binding protein marker), and ribosomal protein (RPL17; ribosome/ribosome-binding protein marker) in the pellet fraction of the ultracentrifugation with/without RNase digestion.

The selective mRNA-binding activity of AEG-1 was further confirmed by sucrose density gradient centrifugation/immunoblot analysis (Figure 11). As depicted in Figure 11A, polyribosomes were extracted and subjected to a sucrose density gradient. In Figure 11B, immunoblotting analyses were performed to determine the enrichment of

AEG-1 in the fractions. The results show that AEG-1 was enriched in the polyribosome fractions (Figure 11B). Ribosomal protein L17 (RPL17) was used as a control for ribosome enrichment (Figure 11B). Taken together, both RNase-sensitive polyribosome association and polysome profiling assays identify the ER integral membrane protein, AEG-1, as a poly(A) RNA-binding protein that binds actively translated mRNAs.

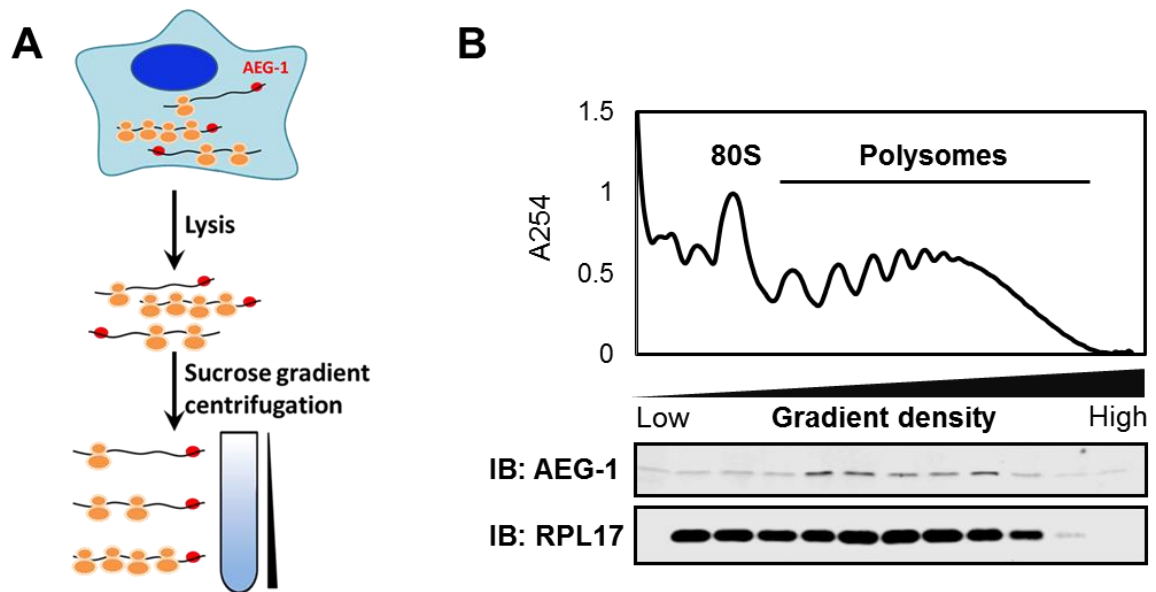


Figure 11: AEG-1 binds translating mRNA in polysome fractions.

(A) Schematic procedure of polyribosome profiling for studying the mRNA-binding activity of AEG-1. Polyribosomes in cell lysate were resolved on 15-50% sucrose gradients and fractionated. The fractions were analyzed for AEG-1 and ribosomal protein L17 (RPL17) distributions by immunoblot. **(B)** Immunoblot analysis of AEG-1 and RPL17 in sucrose gradient fractions.

4.4 Identification of non-canonical RNA-binding domains in AEG-1

Multiple RNA-binding motifs have been widely studied as functional domains with specific RNA binding activity. Having demonstrated RNA-binding activity and selectivity for AEG-1, RNA-binding domain(s) in AEG-1 was further characterized by computational, biochemical, and genetic analyses. Given the membrane protein topology of AEG-1 as a Type I transmembrane ER membrane protein (Britt et al., 2004; Kang et al., 2005; Sutherland et al., 2004), the RNA-binding domain analyses focused on the cytosolic C-terminal domain of AEG-1 (aa70-582). Sequence homology analysis shows that AEG-1 lacks any regions of significant homology to canonical RNA-binding motifs. Therefore, to identify the RNA-binding domain(s) of AEG-1, I expressed a series of AEG-1 truncation mutants (Figure 12A) in cells and examined the RNA-binding activity of the mutants by UV-cross-linking and immunoprecipitation (CLIP) and ultracentrifugation analysis as described above. The truncation mutants were designed based on the RNA-binding residue prediction algorithm, BindN+, as a reference guide (Figure 13) (Wang et al., 2010). The protein expression and subcellular localization of the mutants were further confirmed by immunofluorescence imaging and immunoblotting. As illustrated in Figure 12B, the *in vivo* RNA-binding activity of the mutants was determined by CLIP analysis. The results show that the full-length AEG-1 binds RNA *in vivo*, and the truncation mutants containing the first 350 or 462 amino acid residues of AEG-1 (1-350 and 1-462; Figure 12B) possessed the RNA-binding activity (Figure 12B).

However, a short AEG-1 truncation mutant (1-138; Figure 12B) lacked the RNA-binding activity. These results show that the protein domain between amino acid residues 139 and 350 is essential for *in vivo* AEG-1 RNA-binding activity.

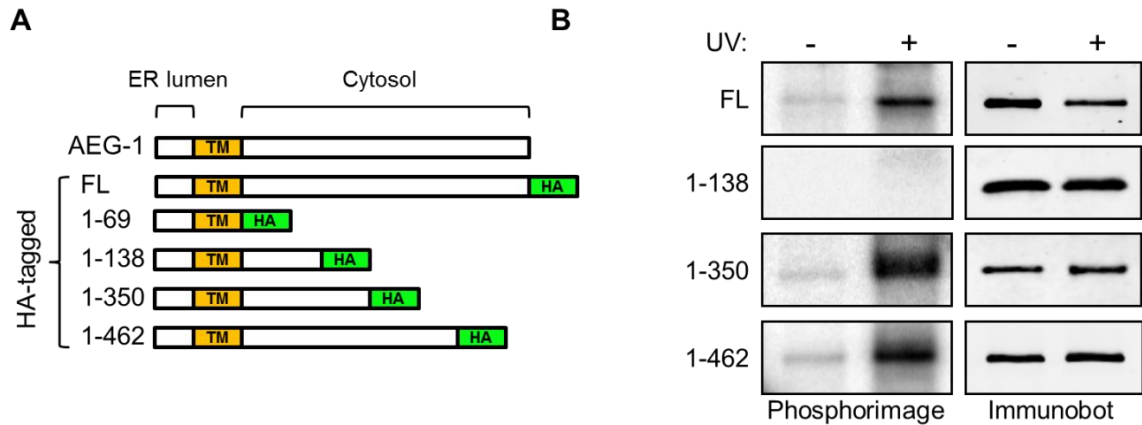


Figure 12: Mutagenesis and *in vivo* RNA-binding analysis reveals a non-canonical RNA-binding domain in the cytosolic domain of AEG-1.

(A) Schematic of AEG-1 truncation mutants. HA-tagged full length (FL) and four C-terminal truncation mutants were used for functional assays. TM, transmembrane domain. (B) The HA-tagged AEG-1 proteins were expressed in HCC cells. As describe in Figure X, UV-cross-linking and immunoprecipitation (CLIP) assay was used to estimate the RNA-binding activity of the mutants. Left, phosphorimages depicting protein-RNA complex formations for FL and truncation mutants. Right, immunoblot analysis of HA-tagged AEG-1 proteins.

```

Sequence: 70 GWAAACAGARKRRSPFRKREAAAAPDDLLKLNLRSEEQKKKNRKKLSEKPKPN
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 314442170998999889871147783433545255761376016998899800166608

Sequence: 130 GRTVEVAEGEAVRTPQSVTAKQPPEIDKKNEKSKKNKKKSKSDAKAVQNSSRHDGKEVDE
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 100636768336641003137342221678099989888886230256560802202422

Sequence: 190 GAWETKISHREKRQQRKRDVLTDSGLDSTIPGIENITVTTEQLTASFPVGSKKKNG
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 352417167969988976212633242732285275268261643752424126088870
      139 ^

Sequence: 250 DSHLNVQVSNFKSGKGDSTLQVSSGLNENLTVNGGGWNEKSVKLSSQISAGEEKWNSVSP
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 016417260026000111053712362415156001011615150438275541012421

Sequence: 310 ASAGKRKTEPSAWSQDTGDANTNGKDWGRSWSDRSIFSGIGSTAEPVSQSTTSQYQWDVS
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 181768601704245442326166707070123137611551144263232111422251

Sequence: 370 RNQFPYIDDEWSGLNGLSSADPNSDWNAPAEWGNWVDEERASLLKSQEPDPDDQKVSDDD
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 012325475213513511453122015344313325448151351246445352041233
      350 ^

Sequence: 430 KEKGEALPTGKSKKKKKKKKQGEDNSTAQDTEELEKEIREDLPVNTSKTRPKQEKAFS
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 62021034167989999999860321125364357217724752800780816116253

Sequence: 490 LKTISTSDPAEVLVKN SQPIKTLPPATSTEP SVILSKSDSKSSSQVPPILQETDKSKSN
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 611731145836481011146152244255518571113110102274466352310877

Sequence: 550 IKQNSVPPSQTKSETSWESPKQIKKKKKARRET
Prediction: -----+-----+-----+-----+-----+-----+-----+
Confidence: 110106116160630305116007999879816

*** Prediction: binding residues are labeled with '+' and in red;
                non-binding residues labeled with '-' and in green.
*** Confidence: from level 0 (lowest) to level 9 (highest).

```

Figure 13: Identification of RNA-binding residues in the cytosolic domain of AEG-1 via BINDN+.

Putative RNA-binding residues were identified in the cytosolic domain of AEG-1 by a RNA-binding residue prediction algorithm, BindN+ (Wang et al., 2010).

To further validate the mRNA-binding selectivity of the truncation mutants, the *in vivo* RNA-binding selectivity of the truncation mutants was determined by ultracentrifugation/RNase digestion analysis as illustrated in Figure 10A. The results show that the full-length AEG-1 and two truncation mutants (1-350 and 1-462) associated with translating mRNA (Figure 14). In contrast, two short truncation mutants (1-69 and 1-138) lacked the association activity (Figure 14). These results show that the protein domain between amino acid residues 139 and 350 is essential for *in vivo* AEG-1 RNA-binding activity. Taken together, both CLIP and RNase-sensitive polyribosome association assays identify an essential protein region (139-350) containing the non-canonical RNA-binding domain(s) in AEG-1.

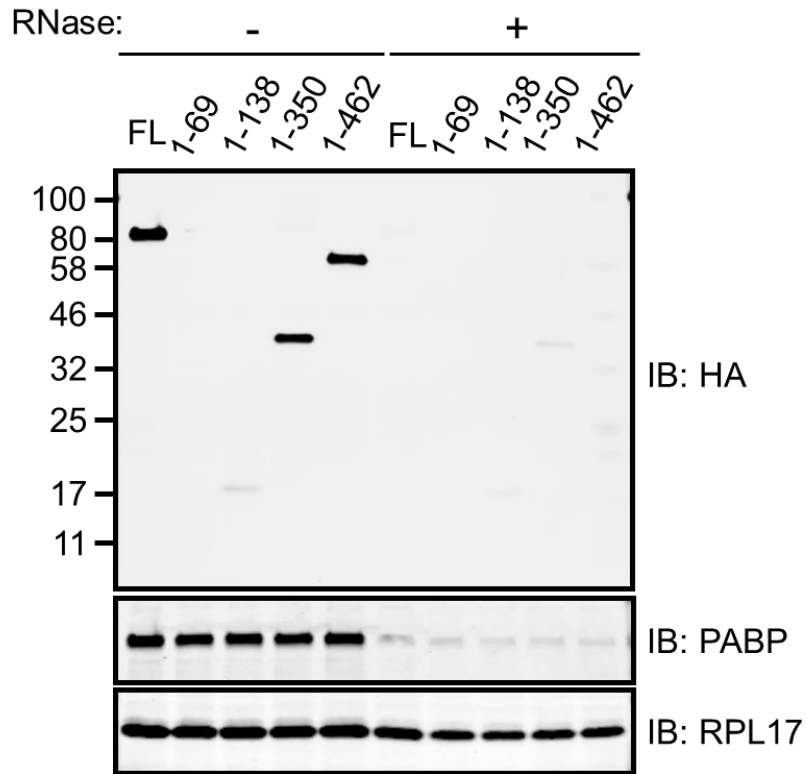


Figure 14: Mutagenesis and ribosome ultracentrifugation analysis reveals a non-canonical RNA-binding domain in AEG-1.

As depicted in Figure X, HA-tagged AEG-1 transfected cells were lysed, incubated with/without RNase digestion and subjected to sucrose cushion ultracentrifugation. The ribosome pellet fraction was analyzed for AEG-1, poly(A)-binding protein (PABP; mRNA-binding protein marker), and ribosomal protein (RPL17; ribosome/ribosome-binding protein marker) by immunoblot.

Given prior genome sequence data indicating that AEG-1 protein is highly conserved in vertebrates (Lee et al., 2013), I hypothesized that the RNA-binding domain would also display high conservation. A series of AEG-1 protein sequences, ranging from *Xenopus* to human, were aligned and analyzed by Megalign Pro (Figure 15). The primary sequences of N-terminal luminal domain (aa1-48) and transmembrane domain

(TMD; aa49-69) were highly conserved, whereas the overall sequence conservation in the cytosolic domain (aa70-582) was relatively low (46.3% identity in 587 residues between *Xenopus* and human AEG-1). Notably, multiple positively charged amino acid clusters were well-aligned in the cytosolic domain, indicating a potential RNA-binding activity for these residues. The sequence of identified RNA-binding domain (aa 139-350) is highly positively charged (27 Lys, 9 Arg residues) and the basic residues are both highly conserved and clustered between aa157 and aa209 (75.5% identity in 53 residues between *Xenopus* and human AEG-1). Combined, this high degree of sequence conservation provides additional support for the conclusion that the cytosolic domain of AEG-1 contains a non-canonical RNA-binding domain.

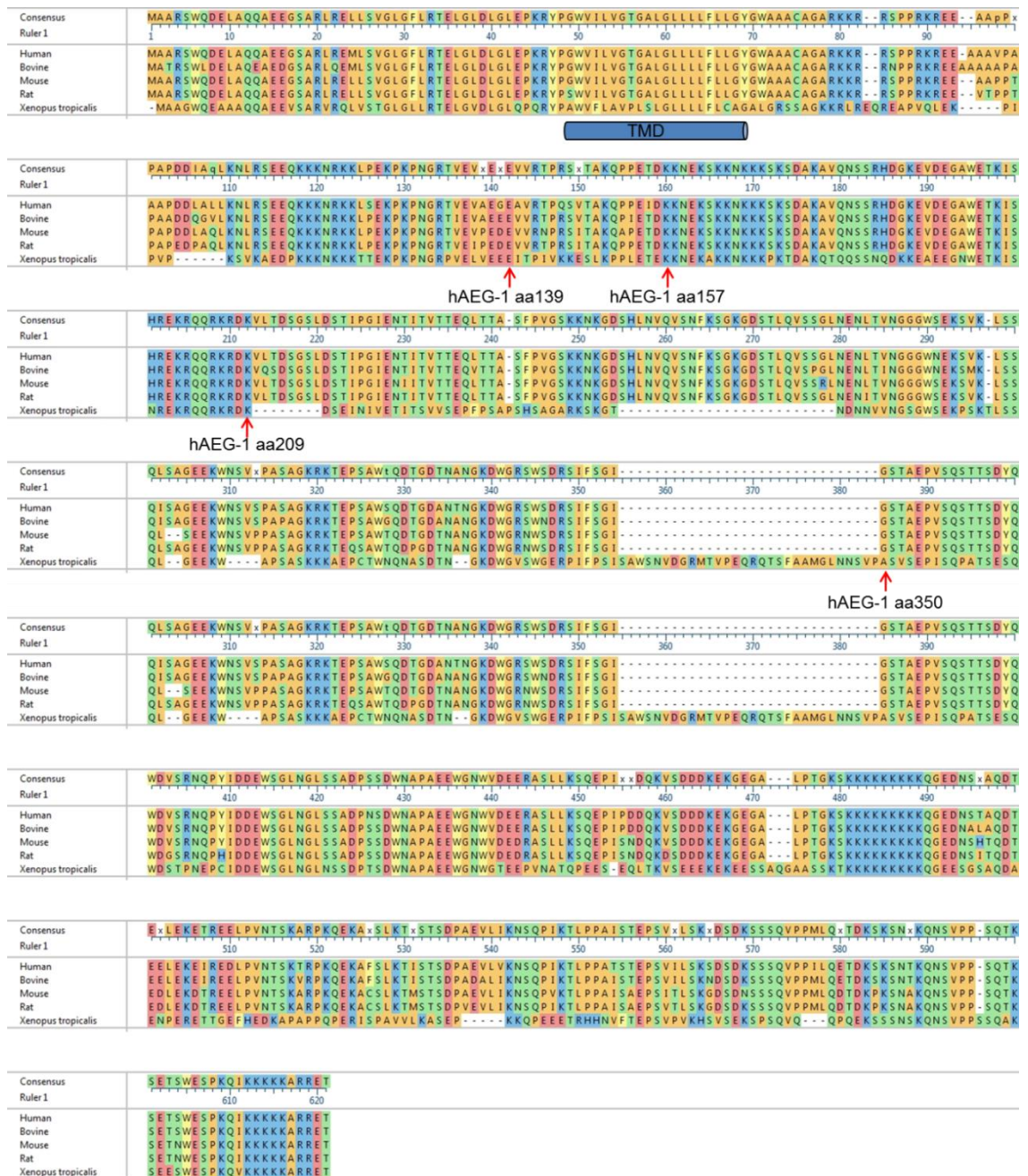


Figure 15: Protein sequence alignment of AEG-1 homologs.

Amino acid sequences of AEG-1 of human (NP_848927), bovine (NP_001039503), mouse (NP_080278), rat (NP_596889), and *Xenopus tropicalis* (NP_989164) were aligned by MegAlign Pro. A consensus sequence is shown as the uppermost sequence. The human AEG-1 truncation mutants used in Figure 12A are indicated. TMD, transmembrane domain.

5. Investigation of *in vivo* mRNA-binding specificity of AEG-1

5.1 Introduction

Crosslinking immunoprecipitation coupled to deep sequencing (CLIP-Seq) approaches have been widely applied to study binding sites of RNA-binding proteins in living cells and animal tissues (Hafner et al., 2010; Konig et al., 2010; Ule et al., 2003). Multiple CLIP-based variants have been developed to achieve reproducible and reliable identifications of RNA-binding sites on a genome-wide scale (Hafner et al., 2010; Konig et al., 2010; Ule et al., 2003). In general, CLIP-Seq approaches share a similar purification protocol, including UV-induced protein-RNA cross-linking, protein-RNA complex purification by immunoprecipitation, RNA moiety purification by proteinase K digestion, reverse transcription, library construction, and high-throughput sequencing. During reverse transcription, protein residuals at the cross-link sites cause site-specific point mutations, which can be identified by sequencing facilitating the precise identification of RNA-binding sites. In this Chapter, to identify RNA targets and binding sites for AEG-1, I applied both high-throughput sequencing-CLIP (HITS-CLIP) and photoactivatable ribonucleoside-enhanced-CLIP (PAR-CLIP) in living cells. CLIP-Seq approaches revealed a unique RNA-binding specificity for AEG-1 targeting mRNAs encoding endomembrane resident proteins and transmembrane proteins. Moreover, *in vivo* functional assays show that AEG-1 regulates the protein expression of its bound mRNAs.

5.2 AEG-1 CLIP-Seq mapping reveals a high enrichment in coding sequence interactions

To study AEG-1 bound mRNA in a genome-wide scale, I performed two CLIP variations, high-throughput sequencing-CLIP (HITS-CLIP) and photoactivatable ribonucleoside-enhanced-CLIP (PAR-CLIP) in human hepatocellular carcinoma (HCC) cells (Darnell, 2010; Hafner et al., 2010). As illustrated in Figure 16, cells expressing HA-tagged AEG-1 were UV-irradiated to *in vivo* cross-link AEG-1-RNA complexes at 254-nm and 365-nm for HITS- and PAR-CLIP, respectively. The complexes were subsequently partially RNase-digested, immunoprecipitated, RNA-specific radiolabeled, resolved by gel electrophoresis, transferred to nitrocellulose membranes, and detected by autoradiography. No UV-irradiation controls were conducted in parallel. The results show a UV-irradiation-dependent RNA-protein complex formation in both HITS- and PAR-CLIP (Figure 16). The predominant RNA-protein complex signals were at 90-100 kD, with 10–20 kD larger than AEG-1 protein (80 kDa) (Figure 17A), indicating that AEG-1 protein cross-linked to small (35- to 70-nt) RNA fragments (Huppertz et al., 2014). The small RNA fragments were released from the membranes by proteinase K digestion, reverse transcribed to cDNA, library constructed, and subjected to next-generation sequencing. The DNA concentration and library quality were determined by Bioanalyzer (Figure 17B), showing UV-irradiation-dependent cDNA libraries at 150-200 bp.

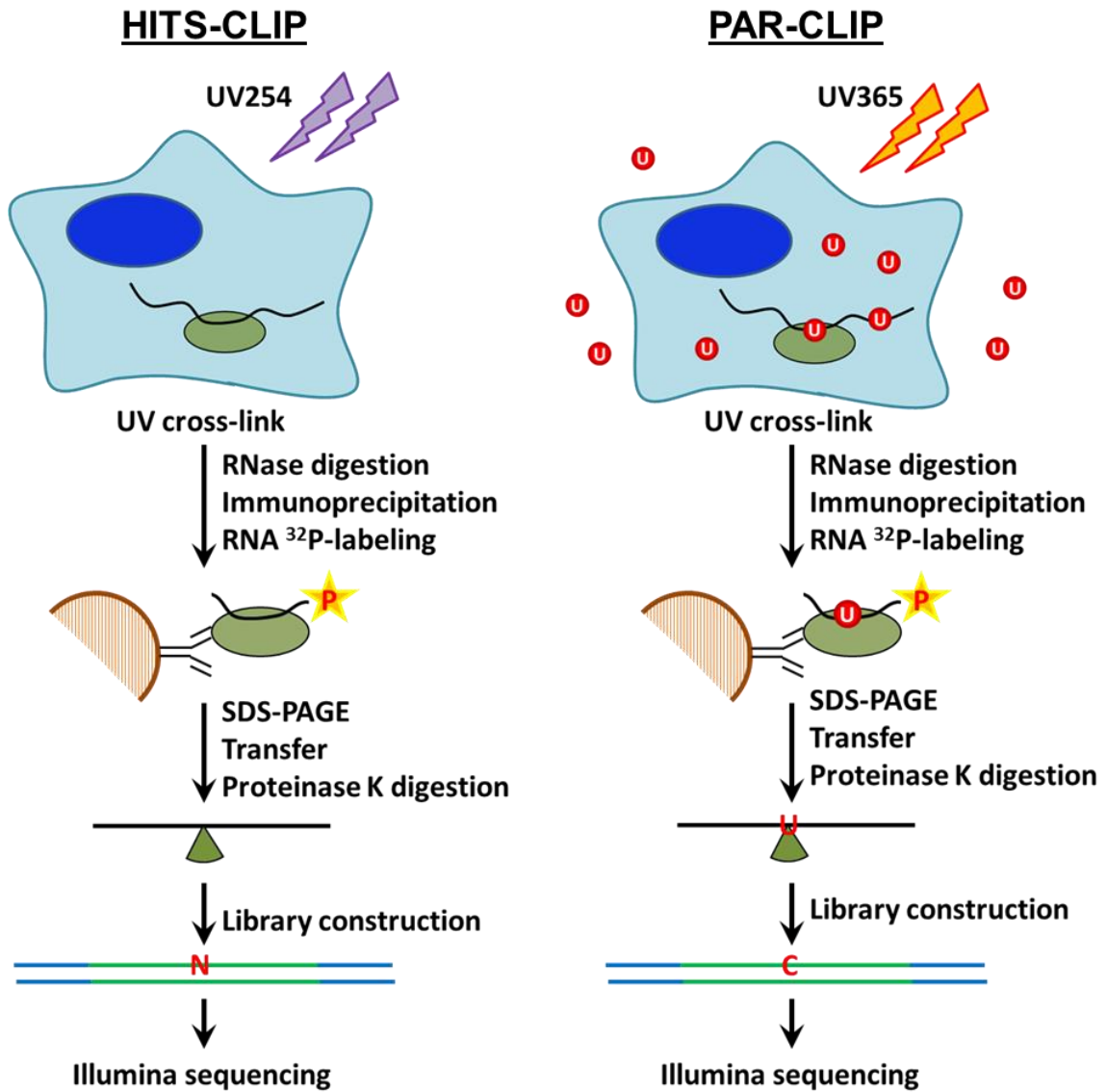


Figure 16: UV-cross-linking and immunoprecipitation-based AEG-1 RNA interactome analyses.

Schematic experimental procedures of HITS-CLIP and PAR-CLIP for the identification of AEG-1-binding sites in living cells. In both procedures, cells were UV-irradiated (at 254-nm for HITS-CLIP, at 365-nm for PAR-CLIP), lysed, and partially RNase digested. Epitope-tagged AEG-1 was immunoprecipitated, resolved by SDS-PAGE and subsequently transferred to nitrocellulose membranes. AEG-1 bound RNA fragments were released by proteinase K digestion and reverse transcribed to cDNA libraries, PCR amplified and high-throughput sequenced. The “U” in the PAR-CLIP procedure are 4-thiouridine (4SU), used as a photoactivatable ribonucleotide.

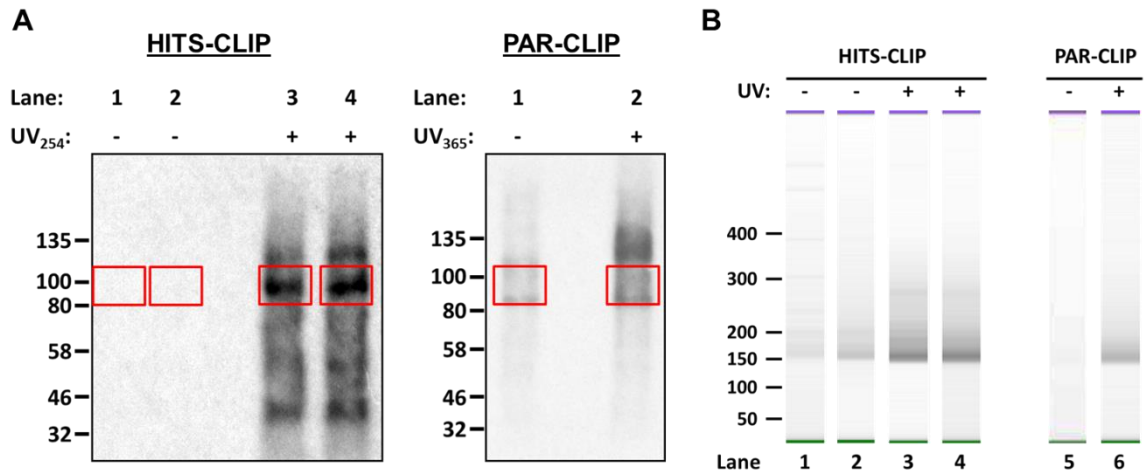


Figure 17: Experimental quality controls for UV-cross-linking and immunoprecipitation-based AEG-1 RNA interactome analyses.

(A) Autoradiographs of AEG-1-RNA complexes on nitrocellulose membranes for HITS-CLIP and PAR-CLIP. The red rectangles indicate the regions (~85-110 kD) processed for further RNA purification. **(B)** Bioanalyzer analyses of cDNA libraries for HITS-CLIP and PAR-CLIP, indicating amplified libraries at 150-200 bp.

The cDNA libraries were sequenced via Illumina HiSeq platform. cDNA sequence reads were processed through a data analysis pipeline (Figure 18). In brief, the reads were adaptor-trimmed, mapped to the human genome and HCC transcriptome, clustered, and scored according to the number of mutations (Figure 18). The analysis results were summarized in Table 2. Overall, the CLIP-Seq results generated approximately 100 to 120 million reads per sample. The reads were further mapped to human genome and transcriptome, generating approximately 4 to 16 million uniquely mapped reads in the HITS- and PAR-CLIP results, respectively (Table 2).

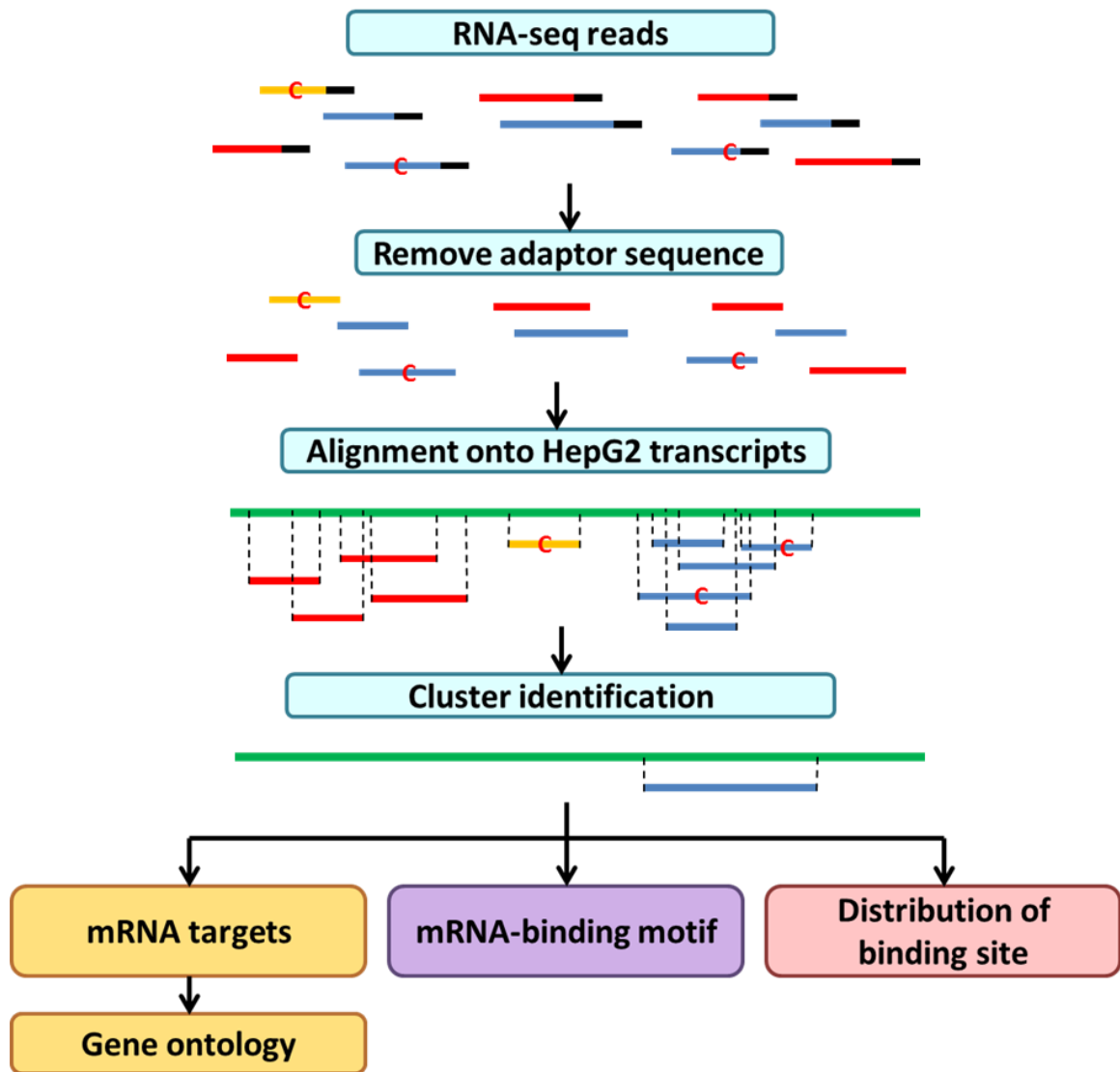


Figure 18: CLIP-Seq data analysis pipeline.

Sequence reads were acquired from next generation sequencing. The reads were adaptor removed, aligned to human genome and HCC transcriptome, and clustered for AEG-1 binding site identification. The results were further applied to the identifications of AEG-1 bound mRNA targets, mRNA-binding motif, and binding site distribution.

Table 2: Summary of read mapping statistics for HITS-CLIP and PAR-CLIP

Experiment:	HITS no UV-1	HITS-CLIP -1	HITS no UV-2	HITS-CLIP -2	PAR no UV	PAR-CLIP
Total reads	14,380,521	119,051,131	31,129,773	111,765,374	1,023,254	101,807,004
Total number of processed reads	5,905,137	64,275,654	15,028,063	65,456,270	584,903	61,099,531
Total number of uniquely mapped reads	124,791	4,112,273	257,084	3,886,390	45,878	15,466,021

Previous studies indicate that the proteinase K digested protein-RNA complexes contain small protein residues at the cross-linked sites, resulting in a diagnostic point mutation during reverse transcription (Sievers et al., 2012; Zhang and Darnell, 2011). As shown in Figure 19, multiple diagnostic point mutations were detected in the sequence reads, especially a characteristic T-to-C mutation pattern in the PAR-CLIP results (Figure 19C).

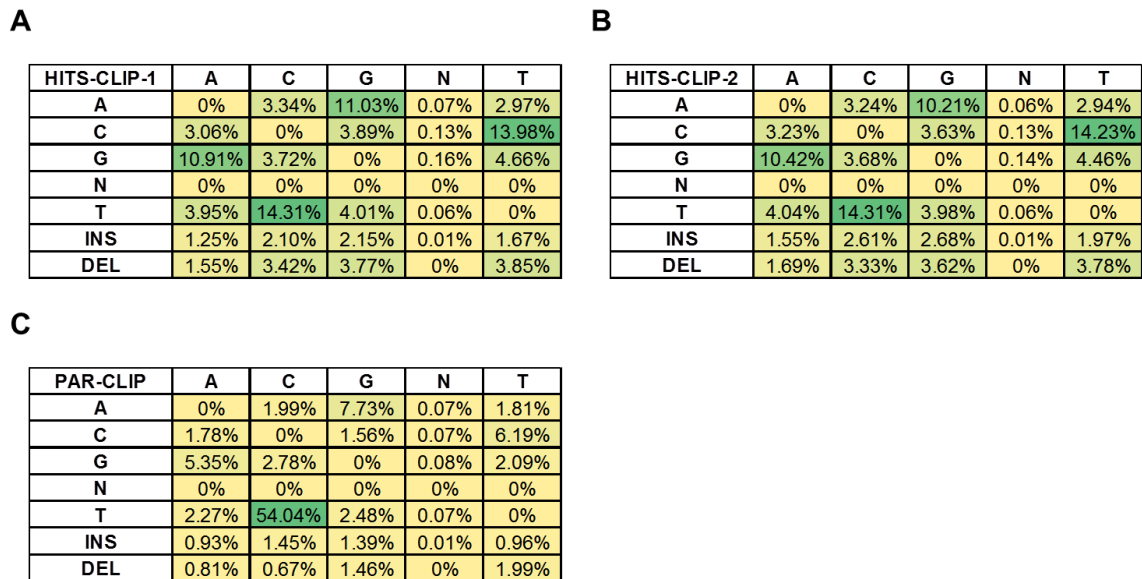


Figure 19: Mutation plots for CLIP-Seq.

(A, B) HITS-CLIP replicates and **(C)** PAR-CLIP. N, any nucleotides; INS, insertion; DEL, deletion.

To determine the RNA-binding preference of AEG-1, the sequence reads were mapped to human genome and transcriptome. Intriguingly, the sequence analysis results show high sequence read enrichments in mRNA (84-93%) in all CLIP-Seq results (Figure 20), indicating that the RNA-binding activity of AEG-1 is highly specific to mRNA. Notably, the read results from the independent HITS-CLIP duplicates demonstrated a highly similar RNA-binding preference. Moreover, the PAR-CLIP results show a higher read enrichment in mRNA, further supporting a genome-wide mRNA-binding specificity for AEG-1.

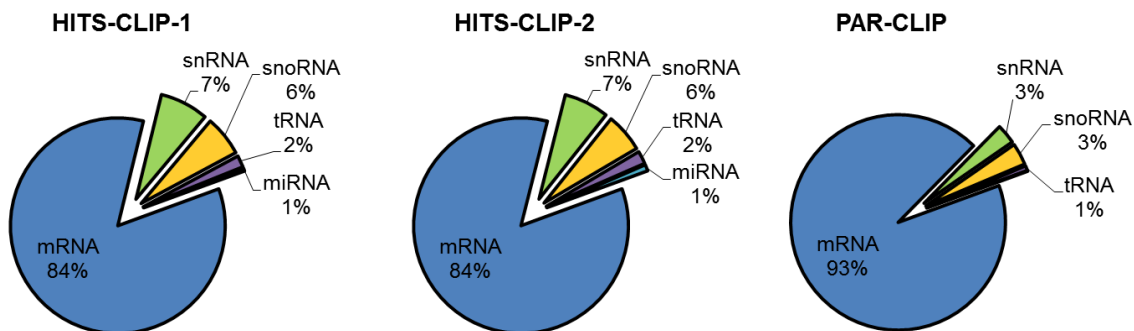


Figure 20: Mapping of sequence reads to RNA for HITS-CLIP replicates and PAR-CLIP.

To further identify AEG-1 binding sites, overlapping sequence reads were grouped as AEG-1 binding clusters if they contained at least six unique reads and at least one mutation for HITS-CLIP or at least one T-to-C conversion for PAR-CLIP (Figure 18). Following these criteria, the two HITS-CLIP experiments generated 15,998 and 17,484 clusters of 35 and 36 nt median length and the PAR-CLIP experiment possessed 99,064 clusters of 55 nt median length (Figure 21).

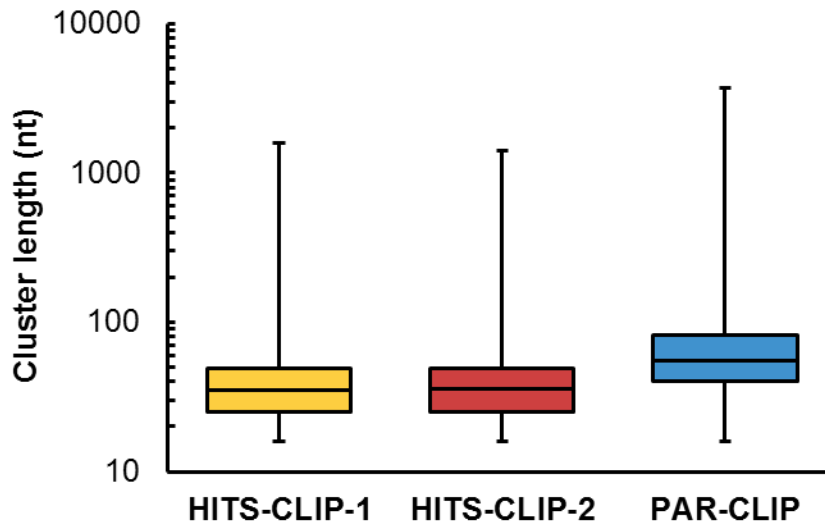


Figure 21: Cluster length distributions for HITS-CLIP replicates and PAR-CLIP.

To identify AEG-1 binding site in mRNA, we further examined the distribution of AEG-1 binding clusters. Notably, the results show that the clusters were highly enriched in the coding sequence (CDS; 68-88%). In contrast, only a small fraction of the clusters were mapped to the untranslated regions (UTR), 3-5% mapped to the 5' UTR and 6-29% to the 3' UTR (Figure 22). Intriguingly, AEG-1 binding activity is highly specific to the CDS, which demonstrates an mRNA-binding preference distinct from the majority of mRNA-binding proteins, which specifically target to the UTRs (Gebauer et al., 2012; Szostak and Gebauer, 2013). The unique CDS-binding preference may imply a distinct mechanism for AEG-1 to regulate its bound mRNA.

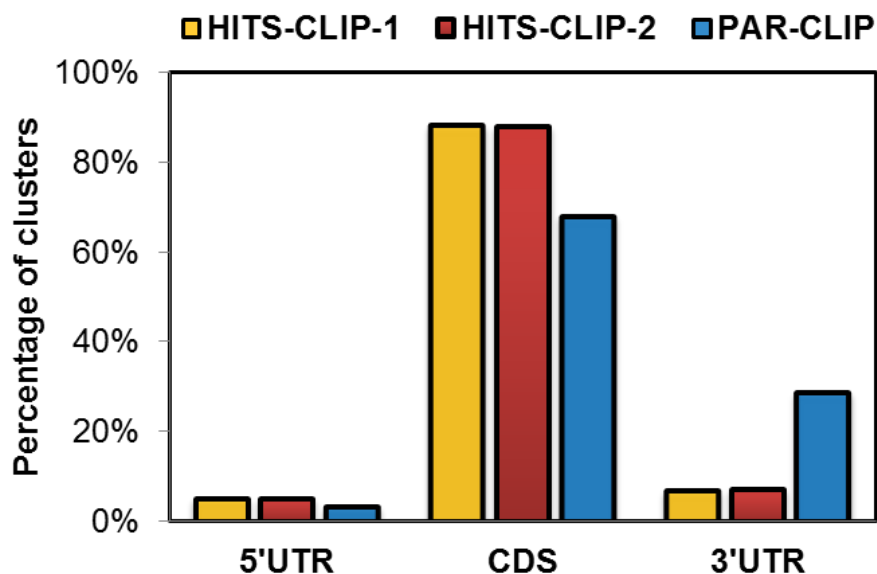


Figure 22: Distributions of AEG-1 binding clusters mapping to the 5' untranslated region (5' UTR), coding sequence (CDS), and 3' UTR for HITS-CLIP replicates and PAR-CLIP.

5.3 AEG-1 RNA interactome is enriched in endomembrane organelle protein-and transmembrane protein-encoding mRNAs

Having determined AEG-1 binding cluster, 1,234 common mRNAs were identified in the HITS- and PAR-CLIP results, which are defined as AEG-1 bound mRNAs (Figure 23). Notably, a large number of mRNAs were solely identified in the PAR-CLIP study, indicating the higher UV-cross-linking efficiency of photoactivatable ribonucleotide than that of natural nucleotides (adenosine, uridine, cytosine, guanidine) may capture more transient/weak interactions (Friedersdorf and Keene, 2014; Hafner et al., 2010). The reproducibility of the AEG-1 bound mRNA identifications from the HITS- and PAR-CLIP results was verified in Figure 23B&C. Of the 1,234 common mRNAs,

HITS-CLIP replicates shared a high reproducibility (Pearson correlation coefficient = 0.95; Figure 23B) and HITS-CLIP vs. PAR-CLIP displayed a moderate positive correlation (Pearson correlation coefficient = 0.31; Figure 23C).

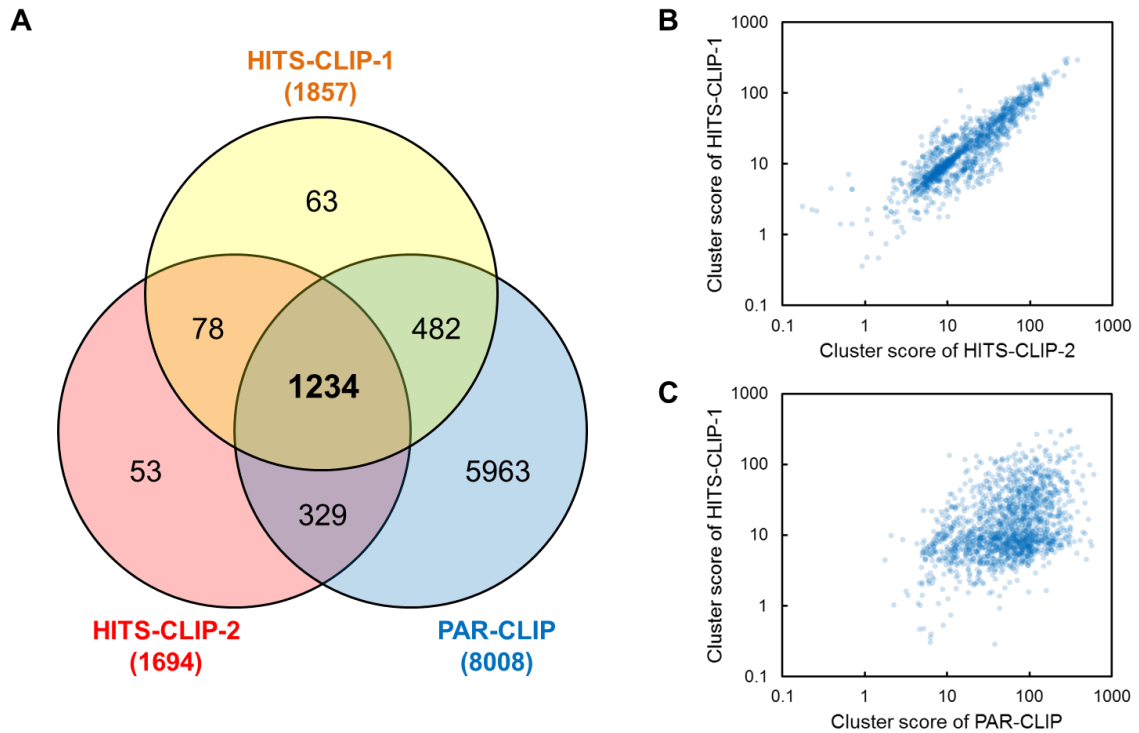


Figure 23: AEG-1 mRNA interactome.

(A) Venn diagrams of AEG-1 bound mRNAs from HITS-CLIP and PAR-CLIP studies. The numbers of identified mRNAs are indicated with brackets. Reproducibility of AEG-1 bound mRNAs, comparing HITS-CLIP duplicates (B) or HITS-CLIP vs. PAR-CLIP studies (C).

To further characterize the molecular functions and cellular distributions of the protein encoded in AEG-1 bound mRNA, the common mRNAs were analyzed by the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics tools (Huang et al., 2009a, d) (Figure 24). Of the common mRNAs, Gene

Ontology (GO) analysis show significantly enriched functional clusters in various cellular functions ranging from blood vessel development to protein maturation (Figure 24A). Remarkably, GO analysis displays an extremely significant enrichment for mRNAs encoding endomembrane organelle proteins ($p = 5.3 \times 10^{-17}$), especially for encoding the ER and lysosomal proteins ($p = 6.9 \times 10^{-66}$ and $p = 5.9 \times 10^{-16}$; Figure 24B), implying a molecular function of AEG-1 in regulating the ER and lysosome biogenesis and/or metabolism. Intriguingly, the results also show an extremely significant enrichment in mRNAs encoding transmembrane proteins ($p = 1.6 \times 10^{-55}$; Figure 24B).

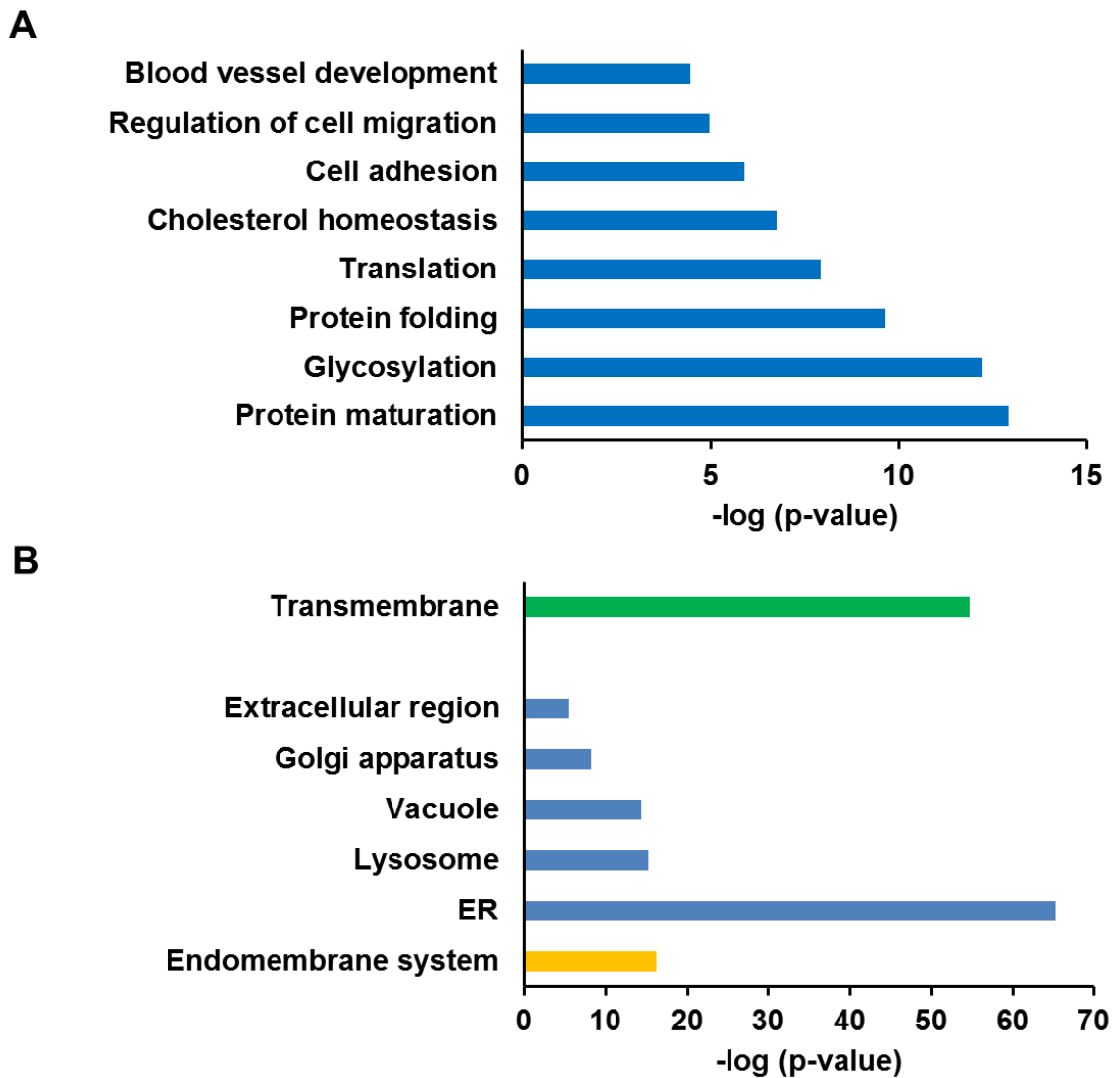


Figure 24: Gene ontology (GO) analyses of AEG-1 RNA interactome.

The AEG-1 bound mRNAs were subjected to GO analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics tools (Huang et al., 2009a, d). GO analyses of AEG-1 bound mRNAs reveal significant enrichments of GO terms in gene functions (A) and cellular component (B).

To further gain insights into the binding preference of AEG-1 for transmembrane protein-encoding mRNAs, the distribution of sequence reads for cytosolic, secretory, and transmembrane proteins were compared between RNA-Seq and CLIP-Seq results (Figure 25). Previous studies have shown that approximately 26% of human protein-coding genes encode transmembrane proteins (Almen et al., 2009; Fagerberg et al., 2010). In the HCC transcriptome, however, only a small fraction of the RNA-Seq reads was mapped to transmembrane transcripts (8.9%) (Figure 25). Perhaps most intriguingly, more than half of the sequence reads from the HITS-CLIP studies mapped to transmembrane protein genes, indicating that a strong RNA-binding preference to transmembrane transcripts (Figure 25). In contrast, the results show significant depletions of mRNA encoding secretory proteins, suggesting that the strong binding preference of AEG-1 does not simply reflect the ER-localized mRNAs (Figure 25).

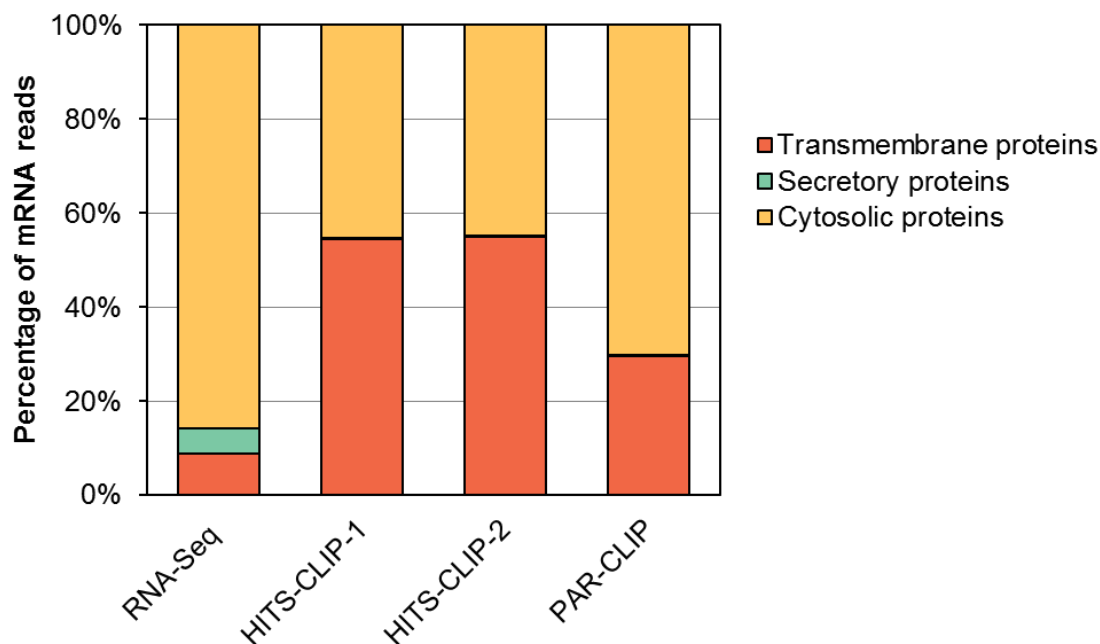


Figure 25: Distributions of sequencing reads mapping to mRNA encoding cytosolic, secretory, and transmembrane proteins for HITS-CLIP replicates and PAR-CLIP.

To further investigate the specific binding location of AEG-1 on mRNAs, the distribution of AEG-1 binding clusters on AEG-1 bound mRNAs was analyzed (Figure 26). Given the strong binding preference for the CDS, we first focused on the regions flanking the start and stop codons within 300 amino acids. However, the results show that these regions contain no over-representations of cluster density (Figure 26A&B), suggesting that AEG-1 has no binding preference for the sequences flanking start and stop codons. Since AEG-1 has a strong binding preference for mRNA encoding transmembrane proteins, we next focused on the regions flanking transmembrane domains. Remarkably, the results show significant over-representation of cluster

densities at 150-300 nucleotides downstream to transmembrane domain encoding sequences (Figure 26C). This specific binding pattern to the CDS of translating mRNAs reflects that AEG-1 not only binds to mRNAs but also regulate the translation of transmembrane proteins. To further dissect the binding preference flanking transmembrane domain encoding sequence, the enrichments of AEG-1 clusters were grouped based on targeting to the first transmembrane domains and secondary transmembrane domains (Figure 26D&E). Intriguingly, the results show similar cluster densities within first 1,200 nt downstream of first and secondary transmembrane domains, suggesting that AEG-1 has no specific binding preference to the order of transmembrane domain in mRNA.

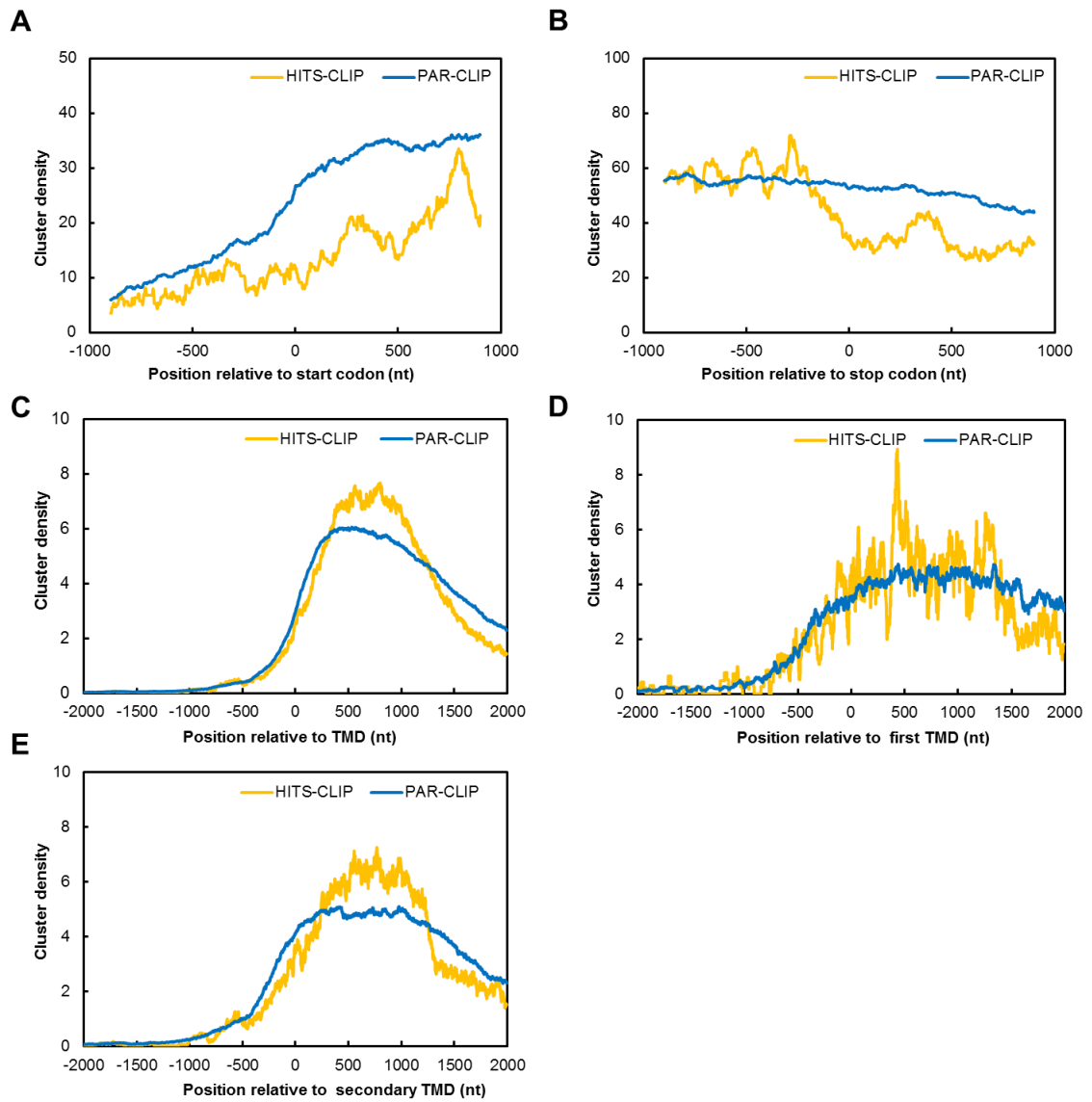


Figure 26: Distributions of AEG-1 bound clusters on mRNAs.

Cluster density plots reveal AEG-1 bound cluster distributions (A) flanking the start codon, (B) stop codon, (C) transmembrane domains, (D) first transmembrane domain, and (E) secondary transmembrane domains for HITS-CLIP (yellow) and PAR-CLIP (blue). 0, the first nucleotide of either start codons, stop codons, or transmembrane domains.

5.4 Validation and analysis of AEG-1 bound mRNAs

To validate further the interaction between AEG-1 and its bound mRNAs, representative AEG-1 bound mRNAs were examined using a complementary biochemical approach, RNA-immunoprecipitation and quantitative PCR (RIP-qPCR). The RIP-qPCR was performed as described (Keene et al., 2006). A HA-tagged AEG-1 transfected HCC cell line was used to facilitate protein-RNA complexes purification (Figure 27). In brief, the AEG-1-RNA complexes were immunoprecipitated and verified by immunoblot analysis against AEG-1 and ribophorin I (RPN1) as a loading and negative control (Figure 27A). The immunoprecipitated RNA was reverse transcribed and subsequently quantified by qPCR against AEG-1 bound mRNAs. The results show significant enrichments of the AEG-1 bound mRNAs (*MDR1*, *ATP1A1*, *NPC1*, *NPC2*, *NPC1L1*, and *HSPA5*) over a control transcript, *GAPDH* (Figure 27B). Taken together, both CLIP-Seq and RIP-qPCR approaches have independently identified that the *in vivo* mRNA-binding activity of AEG-1 is specific to mRNAs encoding endomembrane and transmembrane proteins.

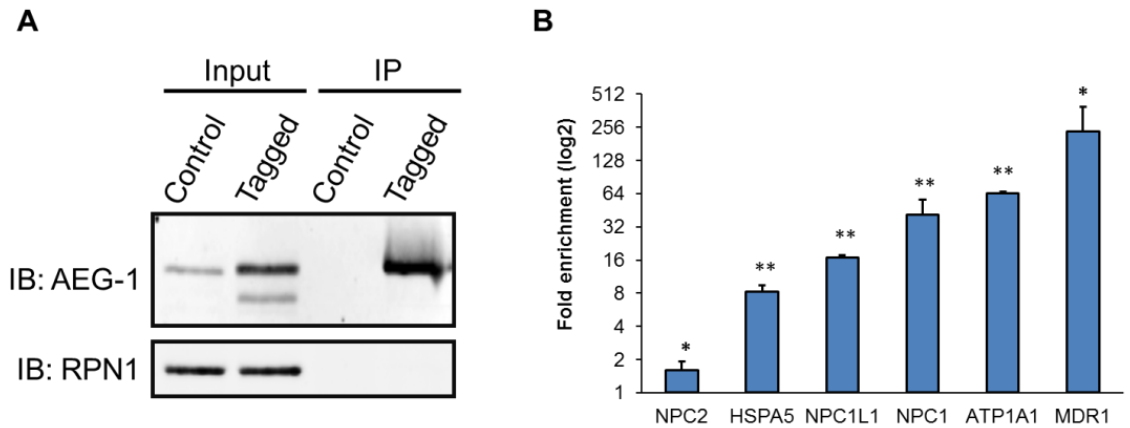


Figure 27: Validation and analysis of AEG-1 mRNA interactome.

Validation of AEG-1 bound mRNAs by RNA immunoprecipitation and RT-qPCR (RIP-qPCR) from mock-transfected cells (PC-4) and HA-tagged AEG-1 transfected cells (AEG-1-14). **(A)** Immunoblots of RIP-qPCR for AEG-1 and ribophorin I (RPN1), an ER membrane protein, as a loading control. **(B)** RIP-qPCR fold enrichments for AEG-1 bound mRNAs (*MDR1*, *ATP1A1*, *NPC1*, *NPC2*, *NPC1L1*, and *HSPA5*) from AEG-1-14 over PC-4 cells. *GAPDH* was used for normalization. * $p < 0.05$, ** $p < 0.01$.

Given the specific mRNA-binding activity of AEG-1, the AEG-1 binding sites on the AEG-1 bound mRNAs were illustrated in representative transcripts (Figure 28). As shown above, mRNAs encoding transmembrane proteins are highly enriched in the AEG-1 bound mRNAs. Here, we show the CLIP-Seq results for two representative AEG-1 bound mRNAs encoding transmembrane proteins, *MDR1* and *NPC1*. These two transcripts were independently identified to interact with AEG-1 by RIP-qPCR analysis (Figure 27B). As depicted in Figure 28A&C, in both *MDR1* and *NPC1*, multiple AEG-1 binding sites were identified, suggesting a complex binding stoichiometry for AEG-1 and mRNA interactions. Notably, as shown in Figure 28A&C, most AEG-1 binding sites were highly clustered in the CDS, whereas PAR-CLIP identified some binding sites in

the 3'UTR. As noted above, the average clusters lengths of PAR-CLIP are relatively longer than that of HITS-CLIP.

To further characterize the specific position of AEG-1 binding sites, we focused on the AEG-1 binding clusters flanking the start/stop codons and transmembrane domains. The results show, in both transcripts, AEG-1 lacked any over-representative binding site flanking the start/stop codons (Figure 28A&C). The transmembrane domains of *MDR1* and *NPC1* transcripts were depicted based on protein domain database, PROSITE-ProRule (Sigrist et al., 2005), and the transmembrane helix prediction algorithm, TMHMM (Figure 28B&D) (Krogh et al., 2001). As noted above, multiple AEG-1 binding sites resided within +300 amino acid downstream transmembrane domains (Figure 28A&C).

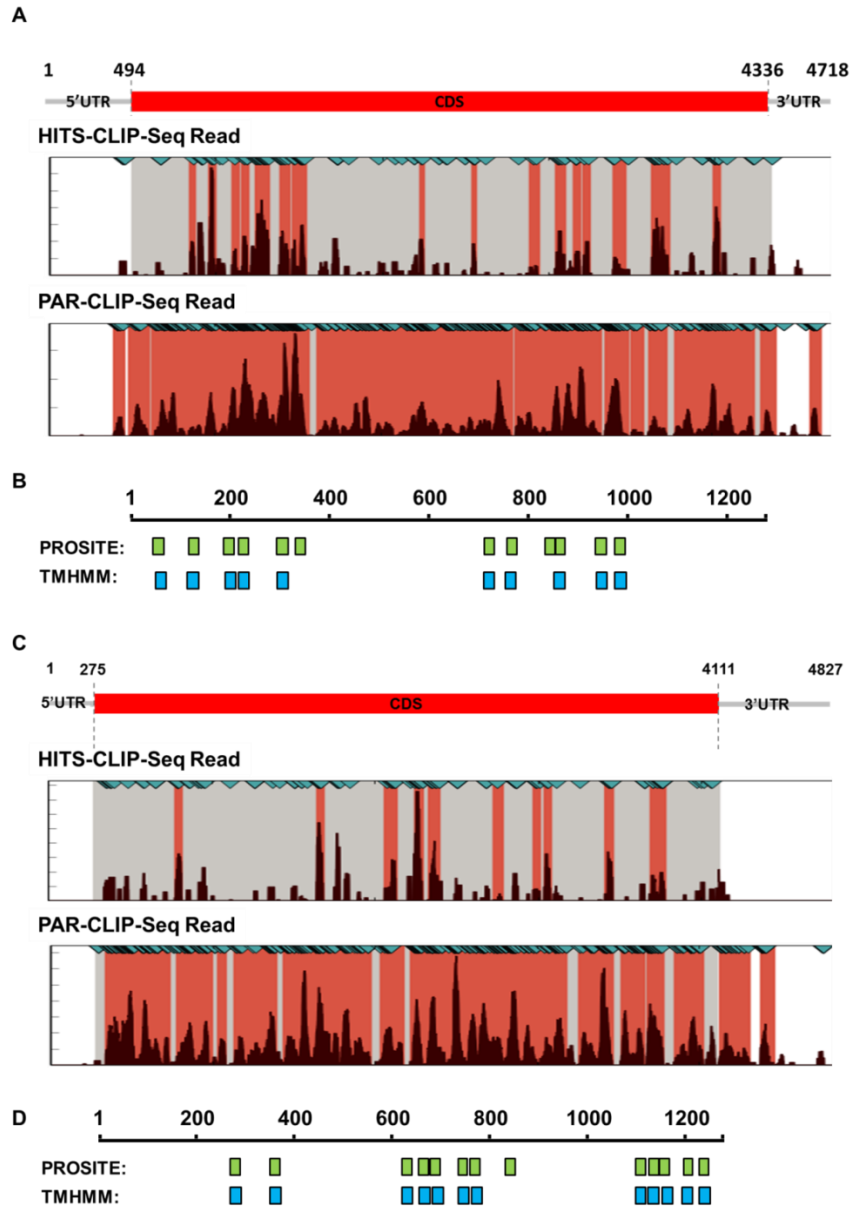


Figure 28: AEG-1 binding sites on MDR1 and NPC1 mRNAs.

Graphical representations of AEG-1 binding sites on MDR1 (A) and NPC1 (C) transcripts. Dark red peaks are sequencing read densities at indicated locations on the x-axis. Light red regions indicate AEG-1 binding sites. Grey regions indicate the coding sequences. Green triangles on the top line mark identified mismatches in the HITS-CLIP and T-to-C conversions in the PAR-CLIP sequencing results. Graphical representations of transmembrane domain distributions in MDR1 (D) and NPC1 (F) indicated by PROSITE-ProRule (green boxes) and TMHMM (blue boxes) (Krogh et al., 2001; Sigrist et al., 2005).

5.5 AEG-1 regulates the protein expression and biological function of its bound mRNAs

Given the CDS-specific binding activity of AEG-1 to actively translated mRNAs, I hypothesized that AEG-1 interacts with translating mRNAs and regulates their translation. Previous studies have shown that AEG-1 expression positively correlates with of multidrug-resistant protein 1 (MDR1) synthesis (Yoo et al., 2010). MDR1 is a plasma membrane transmembrane protein functioning in xenobiotic compound efflux with broad substrate specificity (Figure 29). The protein expression of MDR1 has been found to be significantly up-regulated in multiple cancers, resulting in the development of resistance to anticancer drugs (Yoo et al., 2010). As noted above, the CLIP-Seq and RIP-qPCR studies have demonstrated MDR1 as an AEG-1 bound mRNA, revealing a detailed map of AEG-1 binding sites on MDR1 mRNA (Figure 27&28). To gain further insights into the molecular function of AEG-1 in the regulation of MDR1 expression and function, MDR1 expression at the mRNA and protein levels were examined by RT-PCR, immunoblot, and immunofluorescence analyses (Figure 30). The RT-PCR results indicate that the expressions of MDR1 mRNA were not significantly altered by AEG-1 expression level between AEG-1 overexpression and mock-transfected control cells (Figure 30A). Intriguingly, elevated AEG-1 expression caused a highly increased MDR1 protein level in AEG-1 overexpression cells (Figure 30B&C), suggesting a translational regulation that AEG-1 regulates MDR1 expression at protein level but not at mRNA level. Moreover,

AEG-1 overexpression cells showed higher chemoresistance against anti-cancer drugs than that of control cells (Figure 31).

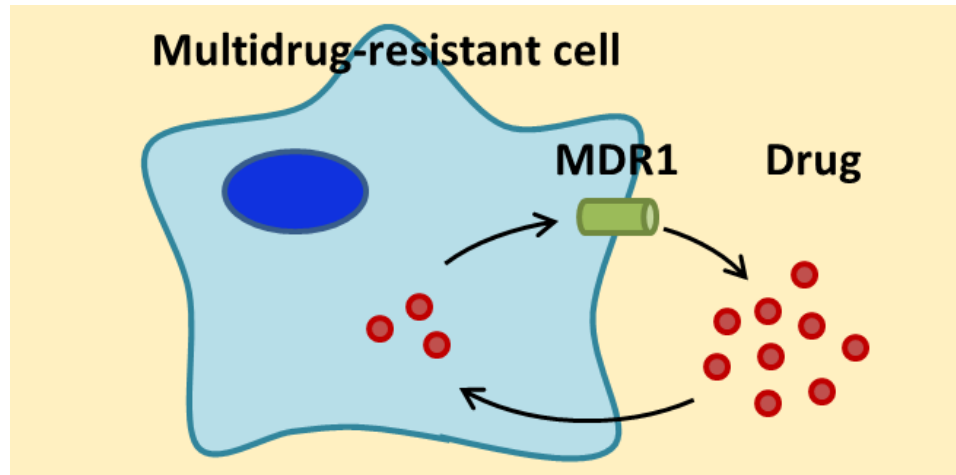


Figure 29: Biological function of MDR1 protein in multidrug-resistant cells.

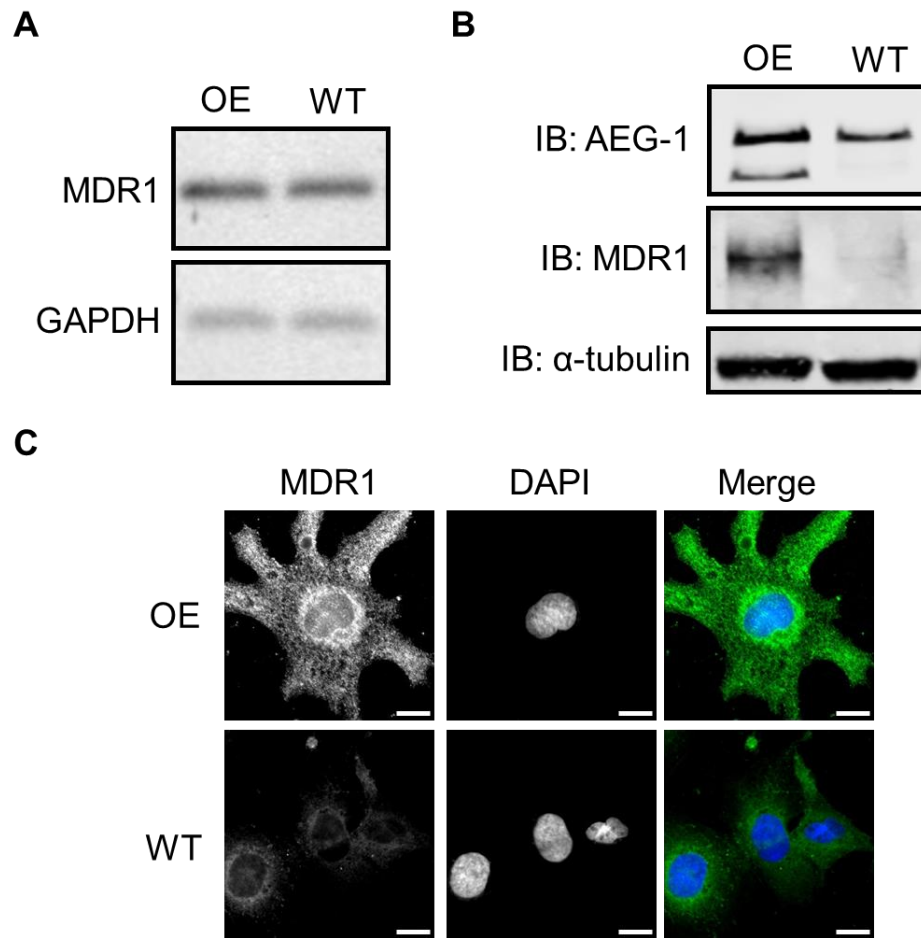


Figure 30: AEG-1 upregulates the protein expression of MDR1.

(A) RT-PCR analyses for MDR1 and GAPDH mRNAs in AEG-1 overexpression (OE) cells. WT, wild-type cells. (B) Immunoblot analyses for the protein expression levels of AEG-1, MDR1, and α -tubulin. (C) Immunofluorescence analyses for MDR1 protein and DAPI (nucleus). Scale bar = 10 μ m.

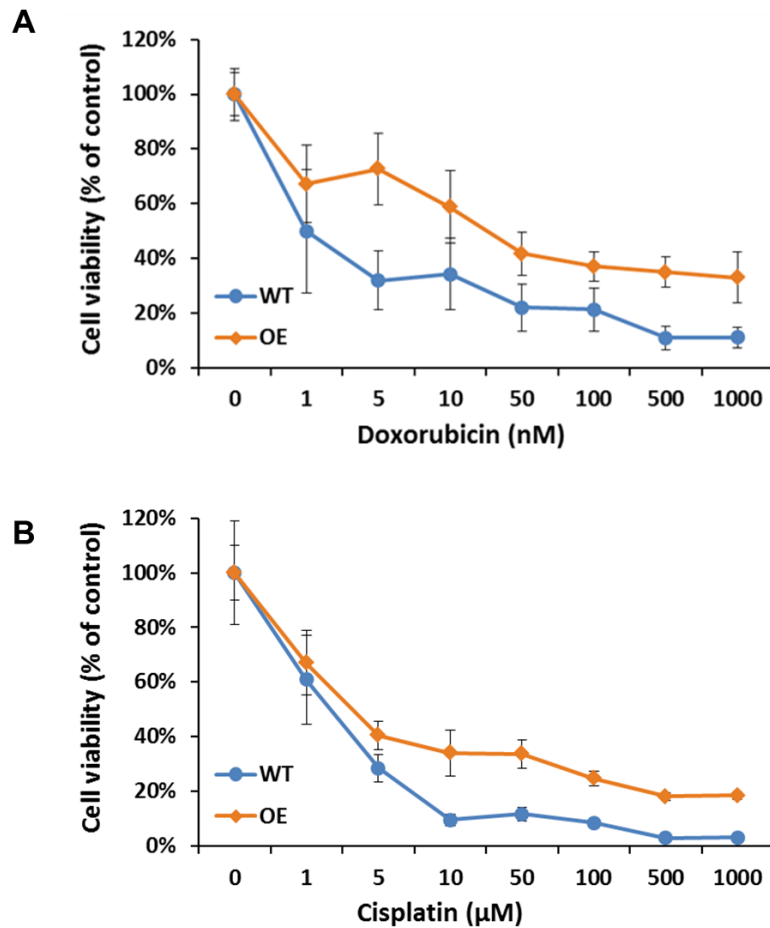


Figure 31: AEG-1 upregulates chemoresistance against anti-cancer drugs.

Cell viability assays for AEG-1 overexpression cells (OE) and wild-type cells (WT) treated with anti-cancer drugs, doxorubicin (A) and cisplatin (B), at indicated concentrations.

In recent study of an AEG-1 knockout (KO) mouse model, AEG-1 expression was linked to cholesterol homeostasis, with loss of expression resulting in decreased body weight and fat, which presumably via regulating the protein expression of intestinal cholesterol transporter NPC1L1 (Niemann-Pick C1-Like 1) and decreasing cholesterol absorption (Robertson et al., 2015b). Remarkably, the GO analysis for the CLIP-Seq

results identified a significant enrichment for genes functioning in cholesterol homeostasis ($p = 1.8 \times 10^{-7}$; Figure 24A), including the cholesterol transporters Niemann-Pick C1 (NPC1) and NPC1L1. Given the mouse phenotype and significant GO enrichment, I hypothesized that AEG-1 not only binds mRNA encoding cholesterol homeostasis proteins, but also regulates their protein expression and biological function. In this study, I focused on a lysosomal cholesterol transporter, NPC1. The cellular mRNA and protein expression levels of NPC1 were examined by RT-PCR, immunoblot, and immunofluorescence analyses. The RT-PCR results indicate that the expressions of NPC1 mRNA were not significantly altered by AEG-1 knockdown (KD) (Figure 32A). Remarkably, AEG-1 knockdown caused a reduced NPC1 protein level (Figure 32B&C), suggesting a translational regulation that AEG-1 regulates NPC1 expression at protein level but not at mRNA level. Notably, mutagenesis and knockout studies have shown that lack of functional NPC1 protein leads to cholesterol accumulation in lysosomes, resulting in a lysosomal storage disease, Niemann-Pick type C disease (Carstea et al., 1997; Neufeld et al., 1999; Reid et al., 2004; Xie et al., 1999). To gain further insights into AEG-1 function in cholesterol homeostasis, free cholesterol accumulation was demonstrated by filipin staining. In Figure 32D, filipin staining revealed a highly up-regulated cholesterol accumulation in AEG-1 KD cells, consistent with lysosomal cholesterol accumulation pattern in NPC1 deficient and knockout cells. Taken together, in both MDR1 and NPC1 studies, these results suggest a translational regulatory

function for AEG-1 to regulate the protein expression and function of its bound transcripts.

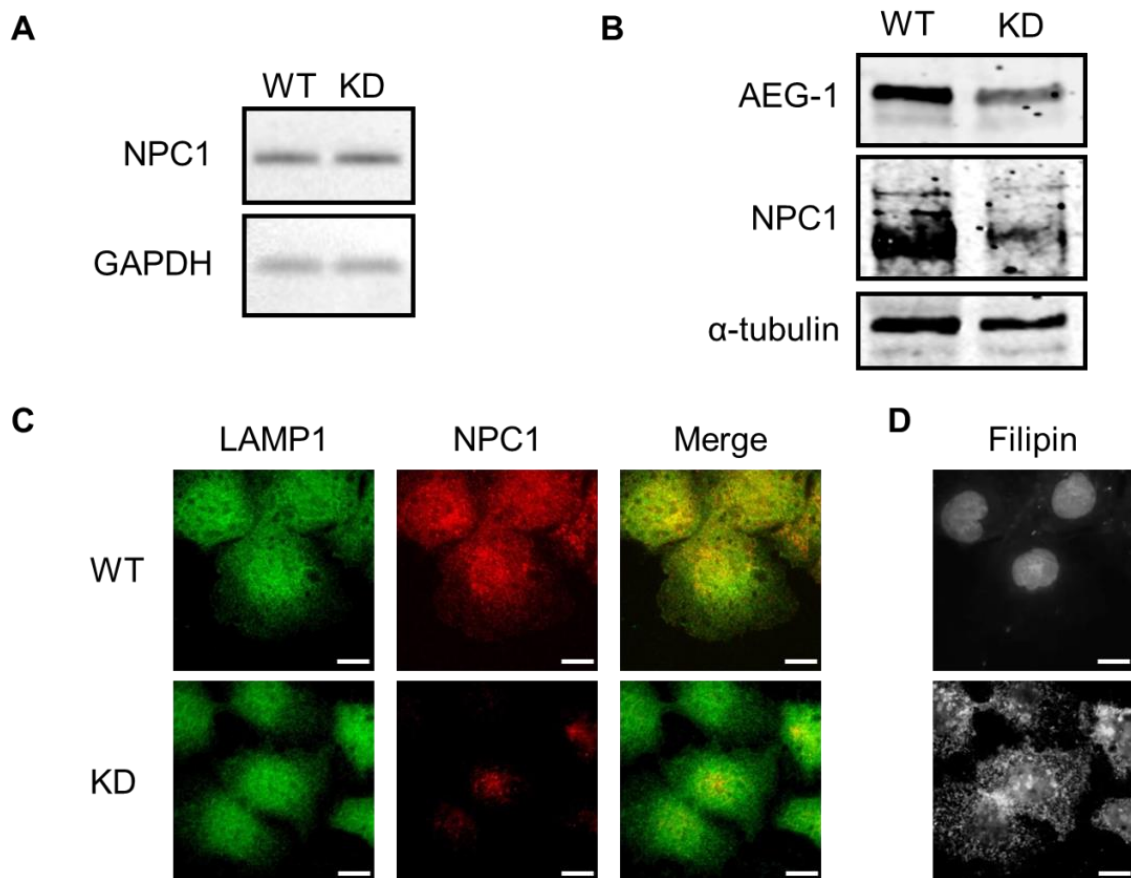


Figure 32: AEG-1 knockdown decreases the protein expression of NPC1 and causes cholesterol accumulation.

(A) RT-PCR analyses for NPC1 and GAPDH mRNAs in AEG-1 knockdown (KD) cells. (B) Immunoblot analyses for the protein expression levels of AEG-1, NPC1, and α -tubulin in AEG-1 KD cells. WT, wild-type cells. (C) Immunofluorescence analyses for the protein expression levels of LAMP1 and NPC1 expression in AEG-1 KD cells. Scale bar = 10 μ m. (D) Fluorescence images of filipin staining for cholesterol accumulation in AEG-1 KD cells. Scale bar = 10 μ m.

6. Discussion

6.1 Identification of the mRNA-binding function of AEG-1

Here I report that select endoplasmic reticulum (ER) membrane proteins play important roles as RNA-binding proteins functioning in mRNA anchoring to the ER membrane. Recent proteomic studies also identified multiple ER membrane proteins as mRNA-binding proteins (Baltz et al., 2012; Castello et al., 2012; Jagannathan et al., 2014; Kwon et al., 2013). To further characterize the mRNA-binding function of the ER membrane proteins, I focused on AEG-1, an ER integral membrane protein previously identified as an oncogene and whose expression is strongly correlated with high metastatic potential and chemoresistance, (Emdad et al., 2010; Hu et al., 2009; Lee et al., 2009; Yoo et al., 2009b). AEG-1 was recently identified as a candidate mRNA-binding protein in a number of recent RNA interactome screens (Baltz et al., 2012; Castello et al., 2012; Jagannathan et al., 2014; Kwon et al., 2013). In this study, we performed genome-scale analysis of the AEG-1 RNA interactome by HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) and PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) and discovered a novel function for AEG-1 in the membrane localization and translational regulation of resident endomembrane organelle protein-encoding mRNAs (e.g., nuclear envelope, ER, Golgi, lysosome, plasma membrane). Notably, AEG-1 bound RNAs were highly enriched in endomembrane protein transcripts, and intriguingly, integral

membrane protein-encoding mRNAs. AEG-1 RNA interaction sites were highly enriched in coding regions (CDS) and largely absent from untranslated regions (UTRs). In addition, the translation of AEG-1 bound mRNA was strongly and positively correlated with AEG-1 expression, independent of target transcript levels, suggesting that AEG-1 promotes translation, although the precise molecular mechanism for such positive regulation remains to be determined.

6.2 Emerging studies of non-canonical RNA-binding proteins

RNA-binding proteins (RBPs) play important roles in a wide variety of post-transcriptional regulations, including splicing, polyadenylation, transport, translation, RNA stabilization, and RNA degradation. Recent proteomic studies have identified a substantial fraction of the human proteome encodes RBP (approximately 7.5% of protein coding genes), including, notably, 700 RBPs bind mRNA (Baltz et al., 2012; Castello et al., 2012; Gerstberger et al., 2014; Kwon et al., 2013). In general, based on the structure of RNA-binding unit, RBPs can be classified into two groups: canonical and non-canonical RBPs. Canonical RBPs possess established RNA-binding motifs such as RNA-recognition motif (RRM), hnRNP K homology (KH) domain, double-stranded RNA-binding domain (dsRBD), and zinc finger (ZF). Canonical RBPs usually have a strong binding preference for the UTRs of mRNAs rather than CDS. In contrast, most non-canonical RBPs generally lack established RNA-binding motifs and were characterized

as multifunctional proteins (moonlighting proteins). Recent mRNA interactome studies have expanded the horizon of RBPs by identifying a large fraction of non-canonical RBP in HeLa, HEK293, and mouse embryonic stem cells (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). Notably, many “housekeeping” proteins were identified as mRNA-binding proteins, including metabolic enzymes functioning in glycolysis, the tricarboxylic acid (TCA) cycle, lipid metabolism, and deoxynucleotide biosynthesis (Castello et al., 2012). Intriguingly, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a “housekeeping” gene has been widely used as controls in many research, was identified as a RBP to regulate translation of its bound transcript in lymphocytes, which involves in T cell activation (Chang et al., 2013). Although recent mRNA interactome studies provide a broadening view of RNA-binding protein identity and functionality, the precise RNA-binding mechanism and function of the non-canonical RBP remains to be determined.

In contrast to many canonical RBPs, AEG-1 binding sites are almost wholly located in the CDS of its mRNA interactome. Intriguingly, the positive correlation observed between AEG-1 expression and target mRNA translation suggests that such interactions promote translation, either directly or indirectly. At first glance, however, it is not intuitively apparent how RBP interactions within a CDS might directly promote translation. Noting that the combined HITS-CLIP and PAR-CLIP datasets did not reveal a strong consensus binding motif; we suggest that AEG-1 affects localization and

translational regulation by a common mechanism of low affinity, high avidity interactions within the CDS. By such mechanism, a highly stable, avidity-based localization/anchoring function could be achieved and a high density of low affinity binding sites may constrain the secondary structural dynamics of the target mRNAs and thereby favor a largely unstructured CDS that would serve as a more optimal translational landscape for the elongating ribosome. Though such a detailed mechanism is plausible, and indeed there is extensive evidence for the broad utilization and mechanistic rationale for such interactions throughout biology (Leulliot and Varani, 2001; Van Roey and Davey, 2015; Varadi et al., 2015; Williamson, 2000), the selectivity and enrichment for endomembrane protein-encoding mRNAs implies the recognition of common features of this cohort of mRNAs not readily discernible at the primary sequence level.

6.3 Intrinsic disorder domain in RNA-binding proteins

Compounding this complexity, RBPs are noteworthy for their high, conserved enrichment in intrinsically disordered domains, and as a consequence a diversity of potential induced fit interactions with RNA substrates (Jarvelin et al., 2016; Leulliot and Varani, 2001; Varadi et al., 2015; Williamson, 2000). Notably, in recent identification of human mRNA-binding proteins (mRBPs), 20% of identified mRBPs are highly disordered (contain disorder amino acid residues > 80%), which enrich disorder-

favoring amino acids [glycine (G), serine (S), and proline (P)], basic amino acids [arginine (R) and lysine (K)], acidic amino acids [aspartic acid (D) and glutamic acid (E)] and tyrosine (Y) (Castello et al., 2012). RS- and RG-repeats were reported in many RBPs, which play important roles in direct RNA interaction and RNA-binding regulation functioning in splicing, nuclear RNA export and translation repression (Chen and Joseph, 2015; Golovanov et al., 2006; Xiang et al., 2013). A newly emerging type of disorder domain in RNA-binding proteins contains K and/or R enriched basic patches (Castello et al., 2012), which may form non-specific electrostatic interactions with the negatively charged sugar-phosphate backbone of RNA. For example, the arginine rich motif in HIV Rev protein is intrinsically disordered, whereas the motif folds into an α -helix upon binding to RNA (Battiste et al., 1996; Casu et al., 2013; Tan et al., 1993; Tan and Frankel, 1994; Wilkinson et al., 2004).

In this regard, the cytosolic domain of AEG-1 has an abundance of intrinsically disordered domains and positively charged patches (Figure 33). We analyzed the distribution and abundance of intrinsically disordered domains in AEG-1 using FoldIndex algorithm (Prilusky et al., 2005). The results show that the cytosolic domain (aa70-582) of AEG-1 is intrinsically disordered. Moreover, the cytosolic domain contains a signature amino acid composition of disordered mRBPs with disorder-favoring amino acids (G, S, P), basic amino acids (R, K), acidic amino acids (D,E) and Y (Figure 15), suggesting the function of AEG-1 as a disordered non-canonical mRBP. In Figure 12 and

14, deletion analyses highlight the domain between aa139 and aa350, which includes a prominent sequence of high disorder and K-enriched patches (Figure 13), displays *in vivo* RNA-binding activity. The abundance of K-enriched patches in the mRNA-binding domain suggests a potential RNA-binding mechanism for AEG-1, which may interact with negatively charged RNA backbone.

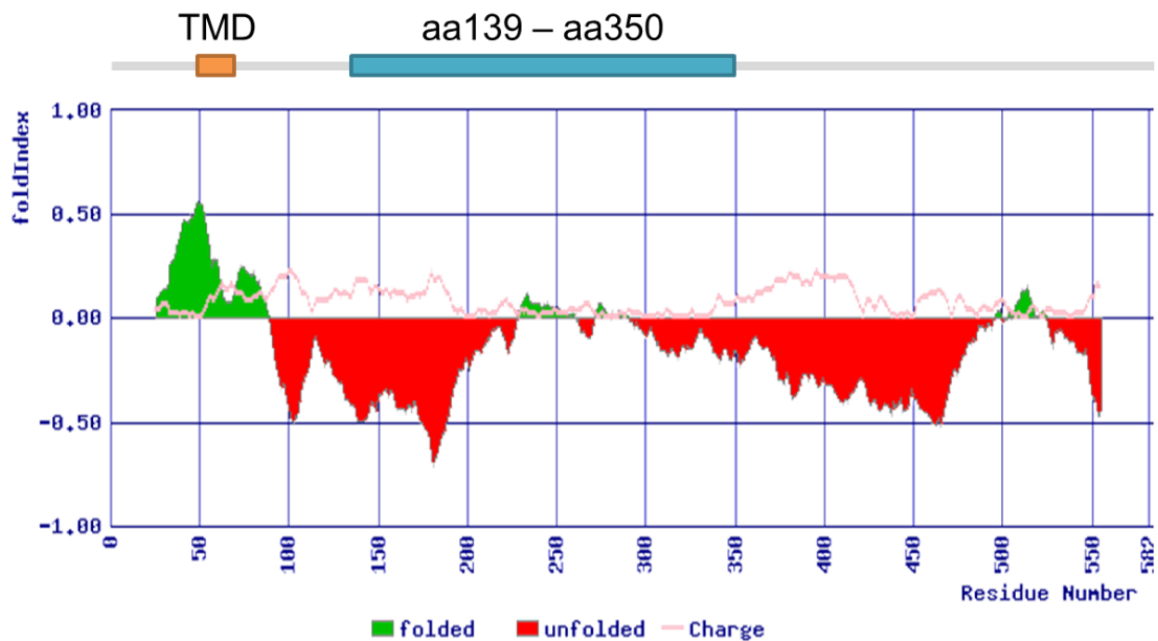


Figure 33: AEG-1 has an abundance of intrinsically disordered domains.

Primary sequences of human AEG-1 (NP_848927) was analyzed by FoldIndex (Prilusky et al., 2005). TMD, transmembrane domain.

6.4 mRNA-binding specificity of AEG-1

The location of mRBP binding sites mirrors RBP molecular function, such as mRNA processing, transport, degradation and translation. Translation initiation factors

bind to the 5' UTR, regulating translation initiation and translation efficiency (Araujo et al., 2012; Wilkie et al., 2003); splicing factors recognize and process the intron-exon junctions at single-nucleotide resolution (Gerstberger et al., 2014; Glisovic et al., 2008); and many mRBPs bind sites in the 3' UTR functioning as regulatory signals for mRNA stability, transport, and degradation (Dreyfuss et al., 2002; Gerstberger et al., 2014; Glisovic et al., 2008). Based on the HITS-CLIP and PAR-CLIP analyses, AEG-1 showed a high sequence read enrichment in the downstream region of encoded transmembrane domains in the CDS (Figure 22). It is interesting to consider that such enhanced interactions may be functionally linked to protein biogenesis, perhaps in enabling translational stalling and thereby enlarging the temporal windows for the integration of transmembrane domains.

Here we also note that studies of the mechanism of transmembrane domain integration have demonstrated that both the primary and secondary structural characteristics of transmembrane domains within the ribosome exit tunnel can serve as a signal regulating the topogenic assembly/insertion of upstream transmembrane domains in the translocon (Daniel et al., 2008; Do et al., 1996). The findings reported here may represent an additional molecular component of this ordered protein synthesis, signaling and translocation process (Cross et al., 2009; Mandon et al., 2013; Shao and Hegde, 2011; Skach, 2009). Alternatively, such cluster read enrichments may be a consequence of the binding interactions between SRP and the nascent transmembrane

domains, though how such long range interactions might influence AEG-1/mRNA interactions awaits further experimental study.

7. Summary and perspective

7.1 Summary

Prior work in the lab demonstrated that a subset of mRNAs encoding resident endomembrane organelle proteins ($\text{mRNA}_{\text{endo}}$) undergoes a SRP- and ribosome-independent localization to the ER. This work sought to further elucidate this process, including the identification of mRNA-anchoring proteins on the ER, characterizing the mRNA-binding specificity of AEG-1, and revealing a translational regulation of AEG-1 bound mRNAs.

Since we observed a SRP- and ribosome-independent mRNA localization to the ER, which is specific for $\text{mRNA}_{\text{endo}}$, we proposed that some ER membrane proteins may play important roles in mRNA anchoring to the ER. First, in canine rough microsome, we demonstrated that numerous ER membrane proteins have RNA-binding activity. The ER membrane proteins were further identified by native oligo(dT) pulldown and mass spectrometry.

In order to focus on one of the best potential candidates, I compared the putative mRNA-binding ER membrane proteins with three proteomic screens of mRNA-binding proteins in HEK293, HeLa, and mouse embryonic stem cells, thus I focused on AEG-1, an mRNA-binding ER membrane protein candidate which was identified in all the screens. First, I demonstrated that AEG-1 binds to translating mRNA in living cells. Since AEG-1 contains no known RNA-binding motifs, I further identified a novel RNA-

binding domain in the cytosolic domain of AEG-1, which is highly disordered and highly positively charged.

CLIP-Seq approaches were used to study the RNA-binding specificity of AEG-1. I found that the RNA-binding activity of AEG-1 is specific to the mRNA encoding endomembrane transmembrane proteins. Intriguingly, AEG-1 exclusively targets to the coding sequence (CDS). Further AEG-1 binding site analyses revealed that AEG-1 has specific binding sites within +300 amino acid downstream transmembrane domains. I addressed the hypothesis that AEG-1 binding may interact with translating ribosomes on its bound mRNAs and simultaneously regulate their translation. I showed that AEG-1 regulates not only the protein expressions of its bound mRNAs (*MDR1* and *NPC1*) but their cellular functions. AEG-1 is therefore an mRNA-binding protein functioning in mRNA-anchoring to the ER and translation regulation.

7.2 Future directions

7.2.1 The molecular mechanism of mRNA-AEG-1 interaction

RNA-binding proteins (RBPs) contain various canonical RNA-binding domains, such as the RNA recognition motif (RRM), the K homology domain (KH), the Pumilio homology domain (PUM-HD) and the double stranded RNA-binding domain (dsRBD) (Lunde et al., 2007), which serve as functional units to modulate RNA-protein interaction. Notably, the candidate ER membrane mRNA-binding proteins noted above

(Figure 6), however, contain no known RNA-binding domains. Moreover, based on the prediction of FoldIndex algorithm (Prilusky et al., 2005), many of these candidate RBPs are predicted to be highly disordered in their cytosolic domains where interact with RNAs. Little is known regarding the molecular mechanism of RNA recognition by any of this newly discovered non-canonical RNA-binding proteins (Baltz et al., 2012; Castello et al., 2012; Gerstberger et al., 2014; Kwon et al., 2013; Pineiro et al., 2015).

In this study, I identified important roles of AEG-1 in mRNA binding and translation regulation whereas the RNA-binding domain of AEG-1 lacks any known RNA-binding motifs. Intriguingly, the RNA-binding domain is highly positively charged and disordered, which may provide clues for further investigation on the mechanism of RNA-AEG-1 interaction. First, the basic poly-lysine patches in the RNA-binding domain may provide non-specific electrostatic interactions with the negatively charged sugar-phosphate backbone of RNA. Given the structure flexibility of disordered RNA-binding domain, the RNA-AEG-1 interaction may provide a driving force in stable protein structure formation, as known as induced fit, which is a prevalent mechanism of RNA-protein interaction (Williamson, 2000). Future work will be required to determine the protein structures of apoprotein and RNA-bound protein and to more clearly define the mechanism of RNA-AEG-1 interaction.

7.2.2 Global post-translational regulation of AEG-1

Post-translational modification is a prevalent mechanism for regulating protein functions. Protein phosphorylation is the most common post-translational modification in eukaryotic cells. Multiple phosphorylation sites have been identified in the cytosolic domain of AEG-1. Recent phosphoproteomic study revealed that AEG-1 is phosphorylated on Ser 298 by the beta subunit of I κ B kinase (IKK β), which regulates NF- κ B-dependent gene expression and cell proliferation as well as correlates with the survival of ovarian carcinoma patients (Krishnan et al., 2015).

With the identifications of phosphorylation in RNA-binding proteins and emerging evidence that phosphorylation regulates protein-RNA interactions (Thapar, 2015), a role for post-translational regulation of mRNA anchoring to the ER can be now be considered. I found that the status of global protein phosphorylation in cell lysate regulates the association of AEG-1 with ribosomes, suggesting that phosphorylation regulates the RNA-binding activity of AEG-1 (data not shown). However, further alanine mutagenesis screening of putative phosphorylation sites in AEG-1 showed no significant effect on the RNA-mediated interactions with ribosomes (data not shown) or associated proteins (Meng et al., 2012). Future work will be required to determine the phosphorylation sites functioning in RNA-binding regulatin in either AEG-1 or its binding partners and to more clearly evaluate the regulatory function of

phosphorylation in the activity and specificity of RNA binding as well as protein translation regulation.

7.2.3 Census of mRNA-anchoring ER membrane proteins

Recently, proteomic screens for the mRNA-binding proteins revealed an unexpected diversity of candidate mRNA-binding proteins (RBPs) on the ER (Figure 6). These studies suggest a complex model for mRNA anchoring to the ER, where multiple ER membrane proteins can serve as mRNA anchors to modulate ER-mRNA interaction. Systematic investigations into the candidate RBPs will gain further insights into the complex, dynamic mechanism of mRNA localization.

The candidate RBPs are non-canonical RBPs which contain no known RNA-binding motifs. Moreover, sequence alignment analysis reveals that the RBPs show no protein sequence homology in the cytosolic domains. These results suggest that the RBPs may have distinct mRNA partners. In this study, we used transcriptome-wide CLIP-Seq approaches to identify the AEG-1 bound mRNAs, enriching mRNAs encoding endomembrane organelle proteins and membrane proteins. To understand the transcriptome-wide mRNA-binding specificity, similar approaches can be applied to the candidate RBPs. Defining how the RBPs regulate mRNA anchoring and/or translation represent an intriguing new direction of research into the newly emerged non-canonical RBPs governing transcriptome-wide mRNA localization.

7.3 Conclusion Remarks

In summary, I report that multiple ER membrane proteins serve as RNA-binding proteins functioning in the selective localization and anchoring of mRNA encoding endomembrane resident proteins to the ER membrane. Many of these ER membrane proteins have been identified as mRNA-binding proteins in human and murine cell lines. Furthermore, we report that one of the identified RNA-binding ER membrane protein, AEG-1, is a non-canonical RNA-binding protein functioning in the selective localization and anchoring of mRNA encoding endomembrane transmembrane proteins to the ER. AEG-1 is a Type I transmembrane protein with a highly positively charged and largely disordered cytoplasmic domain that contains non-canonical RNA-binding domain(s).

References

Almen, M.S., Nordstrom, K.J.V., Fredriksson, R., and Schioth, H.B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *Bmc Biol* 7.

Andreassi, C., Zimmermann, C., Mitter, R., Fusco, S., De Vita, S., Saiardi, A., and Riccio, A. (2010). An NGF-responsive element targets myo-inositol monophosphatase-1 mRNA to sympathetic neuron axons. *Nature neuroscience* 13, 291-301.

Araujo, P.R., Yoon, K., Ko, D.J., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., and Penalva, L.O.F. (2012). Before It Gets Started: Regulating Translation at the 5' UTR. *Comp Funct Genom.*

Baltz, A.G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., *et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell* 46, 674-690.

Bashirullah, A., Halsell, S.R., Cooperstock, R.L., Kloc, M., Karaiskakis, A., Fisher, W.W., Fu, W., Hamilton, J.K., Etkin, L.D., and Lipshitz, H.D. (1999). Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. *The EMBO journal* 18, 2610-2620.

Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* 273, 1547-1551.

Becker, T., Bhushan, S., Jarasch, A., Armache, J.P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., *et al.* (2009). Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome. *Science* 326, 1369-1373.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., *et al.* (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* 6, 10127.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *The EMBO journal* 7, 1749-1756.

Blobel, G. (2000a). Protein targeting. *Biosci Rep* 20, 303-344.

Blobel, G. (2000b). Protein targeting (Nobel lecture). *Chembiochem* 1, 86-102.

Bordier, C. (1981). Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* 256, 1604-1607.

Britt, D.E., Yang, D.F., Yang, D.Q., Flanagan, D., Callanan, H., Lim, Y.P., Lin, S.H., and Hixson, D.C. (2004). Identification of a novel protein, LYRIC, localized to tight junctions of polarized epithelial cells. *Exp Cell Res* 300, 134-148.

Carstea, E.D., Morris, J.A., Coleman, K.G., Loftus, S.K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M.A., Pavan, W.J., Krizman, D.B., *et al.* (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* 277, 228-231.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., *et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393-1406.

Castello, A., Hentze, M.W., and Preiss, T. (2015). Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrin Met* 26, 746-757.

Casu, F., Duggan, B.M., and Hennig, M. (2013). The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophysical journal* 105, 1004-1017.

Chang, C.H., Curtis, J.D., Maggi, L.B., Jr., Faubert, B., Villarino, A.V., O'Sullivan, D., Huang, S.C., van der Windt, G.J., Blagih, J., Qiu, J., *et al.* (2013). Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* 153, 1239-1251.

Chang, P., Torres, J., Lewis, R.A., Mowry, K.L., Houliston, E., and King, M.L. (2004). Localization of RNAs to the mitochondrial cloud in *Xenopus* oocytes through entrapment and association with endoplasmic reticulum. *Mol Biol Cell* 15, 4669-4681.

Chartrand, P., Meng, X.H., Singer, R.H., and Long, R.M. (1999). Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. *Current biology* : CB 9, 333-336.

Chen, E., and Joseph, S. (2015). Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie* 114, 147-154.

Chen, Q., Jagannathan, S., Reid, D.W., Zheng, T., and Nicchitta, C.V. (2011). Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells. *Mol Biol Cell* 22, 2646-2658.

Crofts, A.J., Washida, H., Okita, T.W., Satoh, M., Ogawa, M., Kumamaru, T., and Satoh, H. (2005). The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 83, 728-737.

Cross, B.C.S., Sinning, I., Luirink, J., and High, S. (2009). Delivering proteins for export from the cytosol. *Nat Rev Mol Cell Bio* 10, 255-264.

Cui, X.A., Zhang, H., and Palazzo, A.F. (2012). p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. *PLoS Biol* 10, e1001336.

Cui, X.Y.A., Zhang, Y.J., Hong, S.J., and Palazzo, A.F. (2013). Identification of a Region within the Placental Alkaline Phosphatase mRNA That Mediates p180-dependent Targeting to the Endoplasmic Reticulum. *Journal of Biological Chemistry* 288, 29633-29641.

Daniel, C.J., Conti, B., Johnson, A.E., and Skach, W.R. (2008). Control of translocation through the Sec61 translocon by nascent polypeptide structure within the ribosome. *Journal of Biological Chemistry* 283, 20864-20873.

Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wires Rna* 1, 266-286.

Do, H., Falcone, D., Lin, J.L., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* 85, 369-378.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Bio* 3, 195-205.

Emdad, L., Sarkar, D., Lee, S.G., Su, Z.Z., Yoo, B.K., Dash, R., Yacoub, A., Fuller, C.E., Shah, K., Dent, P., *et al.* (2010). Astrocyte elevated gene-1: a novel target for human glioma therapy. *Molecular cancer therapeutics* 9, 79-88.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M., and Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics* 10, 1141-1149.

Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Current biology : CB* 13, 1159-1168.

Friedersdorf, M.B., and Keene, J.D. (2014). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol* 15, R2.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. *Cell* 47, 735-746.

Gagnon, J.A., and Mowry, K.L. (2011). Molecular motors: directing traffic during RNA localization. *Critical reviews in biochemistry and molecular biology* 46, 229-239.

Garner, C.C., Tucker, R.P., and Matus, A. (1988). Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites. *Nature* 336, 674-677.

Gebauer, F., Preiss, T., and Hentze, M.W. (2012). From Cis-Regulatory Elements to Complex RNPs and Back. *Csh Perspect Biol* 4.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature reviews Genetics* 15, 829-845.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters* 582, 1977-1986.

Golovanov, A.P., Hautbergue, G.M., Tintaru, A.M., Lian, L.Y., and Wilson, S.A. (2006). The solution structure of REF2-I reveals interdomain interactions and regions involved in binding mRNA export factors and RNA. *RNA* 12, 1933-1948.

Gonzalez, I., Buonomo, S.B., Nasmyth, K., and von Ahsen, U. (1999). ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. *Current biology* : CB 9, 337-340.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129-141.

Heym, R.G., and Niessing, D. (2012). Principles of mRNA transport in yeast. *Cellular and molecular life sciences* : CMLS 69, 1843-1853.

Hu, G., Chong, R.A., Yang, Q., Wei, Y., Blanco, M.A., Li, F., Reiss, M., Au, J.L., Haffty, B.G., and Kang, Y. (2009). MTDH activation by 8q22 genomic gain promotes chemoresistance and metastasis of poor-prognosis breast cancer. *Cancer cell* 15, 9-20.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37, 1-13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009d). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., Konig, J., and Ule, J. (2014). iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods* 65, 274-287.

Jagannathan, S., Hsu, J.C., Reid, D.W., Chen, Q., Thompson, W.J., Moseley, A.M., and Nicchitta, C.V. (2014). Multifunctional roles for the protein translocation machinery in RNA anchoring to the endoplasmic reticulum. *The Journal of biological chemistry* 289, 25907-25924.

- Jarvelin, A.I., Noerenberg, M., Davis, I., and Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Commun Signal* 14, 9.
- Jeffery, W.R., Tomlinson, C.R., and Brodeur, R.D. (1983). Localization of actin messenger RNA during early ascidian development. *Developmental biology* 99, 408-417.
- Kang, D.C., Su, Z.Z., Sarkar, D., Emdad, L., Volsky, D.J., and Fisher, P.B. (2005). Cloning and characterization of HIV-1-inducible astrocyte elevated gene-1, AEG-1. *Gene* 353, 8-15.
- Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols* 1, 302-307.
- Keiler, K.C. (2011). RNA localization in bacteria. *Curr Opin Microbiol* 14, 155-159.
- Kikuchi, Y., Hishinuma, F., and Sakaguchi, K. (1978). Addition of mononucleotides to oligoribonucleotide acceptors with T4 RNA ligase. *Proc Natl Acad Sci U S A* 75, 1270-1273.
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology* 17, 909-915.
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution. *Jove-J Vis Exp*.
- Krishnan, R.K., Nolte, H., Sun, T., Kaur, H., Sreenivasan, K., Looso, M., Offermanns, S., Kruger, M., and Swiercz, J.M. (2015). Quantitative analysis of the TNF-alpha-induced phosphoproteome reveals AEG-1/MTDH/LYRIC as an IKKbeta substrate. *Nat Commun* 6, 6658.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305, 567-580.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nature structural & molecular biology* 20, 1122-1130.

Lawrence, J.B., and Singer, R.H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* 45, 407-415.

Lecuyer, E., Yoshida, H., and Krause, H.M. (2009). Global implications of mRNA localization pathways in cellular organization. *Curr Opin Cell Biol* 21, 409-415.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131, 174-187.

Lee, S.G., Jeon, H.Y., Su, Z.Z., Richards, J.E., Vozhilla, N., Sarkar, D., Van Maerken, T., and Fisher, P.B. (2009). Astrocyte elevated gene-1 contributes to the pathogenesis of neuroblastoma. *Oncogene* 28, 2476-2484.

Lee, S.G., Kang, D.C., DeSalle, R., Sarkar, D., and Fisher, P.B. (2013). AEG-1/MTDH/LYRIC, the Beginning: Initial Cloning, Structure, Expression Profile, and Regulation of Expression. *Aeg-1/Mtdh/Lyric Implicated in Multiple Human Cancers* 120, 1-38.

Leulliot, N., and Varani, G. (2001). Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* 40, 7947-7956.

Longuet, M., Auger-Buendia, M.A., and Tavitian, A. (1979). Studies on the distribution of ribosomal proteins in mammalian ribosomal subunits. *Biochimie* 61, 1113-1123.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature reviews Molecular cell biology* 8, 479-490.

Mandon, E.C., Trueman, S.F., and Gilmore, R. (2013). Protein Translocation across the Rough Endoplasmic Reticulum. *Csh Perspect Biol* 5.

- Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4.
- Markus, M.A., and Morris, B.J. (2009). RBM4: A multifunctional RNA-binding protein. *Int J Biochem Cell B* 41, 740-743.
- Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell* 136, 719-730.
- Mathias, R.A., Chen, Y.S., Kapp, E.A., Greening, D.W., Mathivanan, S., and Simpson, R.J. (2011). Triton X-114 phase separation in the isolation and purification of mouse liver microsomal membrane proteins. *Methods* 54, 396-406.
- Medioni, C., Mowry, K., and Besse, F. (2012). Principles and roles of mRNA localization in animal development. *Development* 139, 3263-3276.
- Meng, X., Thiel, K.W., and Leslie, K.K. (2013). Drug resistance mediated by AEG-1/MTDH/LYRIC. *Adv Cancer Res* 120, 135-157.
- Meng, X., Zhu, D., Yang, S., Wang, X., Xiong, Z., Zhang, Y., Brachova, P., and Leslie, K.K. (2012). Cytoplasmic Metadherin (MTDH) provides survival advantage under conditions of stress by acting as RNA-binding protein. *The Journal of biological chemistry* 287, 4485-4491.
- Mili, S., Moissoglu, K., and Macara, I.G. (2008). Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature* 453, 115-119.
- Moccia, R., Chen, D., Lyles, V., Kapuya, E., E, Y., Kalachikov, S., Spahn, C.M., Frank, J., Kandel, E.R., Barad, M., *et al.* (2003). An unbiased cDNA library prepared from isolated Aplysia sensory neuron processes is enriched for cytoskeletal and translational mRNAs. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23, 9409-9417.
- Mowry, K.L., and Cote, C.A. (1999). RNA sorting in *Xenopus* oocytes and embryos. *FASEB J* 13, 435-445.

- Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. (1996). Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. *Science* 274, 2075-2079.
- Neufeld, E.B., Wastney, M., Patel, S., Suresh, S., Cooney, A.M., Dwyer, N.K., Roff, C.F., Ohno, K., Morris, J.A., Carstea, E.D., *et al.* (1999). The Niemann-Pick C1 protein resides in a vesicular compartment linked to retrograde transport of multiple lysosomal cargo. *The Journal of biological chemistry* 274, 9627-9635.
- Nicchitta, C.V., and Blobel, G. (1993). Luminal proteins of the endoplasmic reticulum are required to complete protein translocation. *Cell* 73, 989-998.
- Palacios, I.M., and Johnston, D.S. (2001). Getting the message across: the intracellular localization of mRNAs in higher eukaryotes. *Annu Rev Cell Dev Biol* 17, 569-614.
- Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science* 189, 347-358.
- Pashev, I.G., Dimitrov, S.I., and Angelov, D. (1991). Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci* 16, 323-326.
- Pfeffer, S., Burbaum, L., Unverdorben, P., Pech, M., Chen, Y.X., Zimmermann, R., Beckmann, R., and Forster, F. (2015). Structure of the native Sec61 protein-conducting channel. *Nature Communications* 6.
- Pineiro, D., Fernandez-Chamorro, J., Francisco-Velilla, R., and Martinez-Salas, E. (2015). Gemin5: A Multitasking RNA-Binding Protein Involved in Translation Control. *Biomolecules* 5, 528-544.
- Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., and Sussman, J.L. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21, 3435-3438.
- Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450, 663-669.

Rapoport, T.A., Jungnickel, B., and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu Rev Biochem* 65, 271-303.

Rebagliati, M.R., Weeks, D.L., Harvey, R.P., and Melton, D.A. (1985). Identification and cloning of localized maternal RNAs from *Xenopus* eggs. *Cell* 42, 769-777.

Reid, D.W., and Nicchitta, C.V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *The Journal of biological chemistry* 287, 5518-5527.

Reid, D.W., and Nicchitta, C.V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nature reviews Molecular cell biology* 16, 221-231.

Reid, P.C., Sakashita, N., Sugii, S., Ohno-Iwashita, Y., Shimada, Y., Hickey, W.F., and Chang, T.Y. (2004). A novel cholesterol stain reveals early neuronal cholesterol accumulation in the Niemann-Pick type C1 mouse brain. *J Lipid Res* 45, 582-591.

Robertson, C.L., Srivastava, J., Rajasekaran, D., Gredler, R., Akiel, M.A., Jariwala, N., Siddiq, A., Emdad, L., Fisher, P.B., and Sarkar, D. (2015a). The role of AEG-1 in the development of liver cancer. *Hepat Oncol* 2, 303-312.

Robertson, C.L., Srivastava, J., Siddiq, A., Gredler, R., Emdad, L., Rajasekaran, D., Akiel, M., Shen, X.N., Corwin, F., Sundaresan, G., *et al.* (2015b). Astrocyte Elevated Gene-1 (AEG-1) Regulates Lipid Homeostasis. *The Journal of biological chemistry* 290, 18227-18236.

Sandoz, P.A., and van der Goot, F.G. (2015). How many lives does CLIMP-63 have? *Biochemical Society transactions* 43, 222-228.

Sawicka, K., Bushell, M., Spriggs, K.A., and Willis, A.E. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochemical Society transactions* 36, 641-647.

Shao, S.C., and Hegde, R.S. (2011). Membrane Protein Insertion at the Endoplasmic Reticulum. *Annu Rev Cell Dev Bi* 27, 25-56.

Sievers, C., Schlumpf, T., Sawarkar, R., Comoglio, F., and Paro, R. (2012). Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic acids research* *40*.

Sigrist, C.J., De Castro, E., Langendijk-Genevaux, P.S., Le Saux, V., Bairoch, A., and Hulo, N. (2005). ProRule: a new database containing functional and structural information on PROSITE profiles. *Bioinformatics* *21*, 4060-4066.

Skach, W.R. (2009). Cellular mechanisms of membrane protein folding. *Nature structural & molecular biology* *16*, 606-612.

Stephens, S.B., and Nicchitta, C.V. (2007). In vitro and tissue culture methods for analysis of translation initiation on the endoplasmic reticulum. *Method Enzymol* *431*, 47-60.

Stephens, S.B., and Nicchitta, C.V. (2008). Divergent regulation of protein synthesis in the cytosol and endoplasmic reticulum compartments of mammalian cells. *Molecular Biology of the Cell* *19*, 623-632.

Sutherland, H.G., Lam, Y.W., Briers, S., Lamond, A.I., and Bickmore, W.A. (2004). 3D3/lyric: a novel transmembrane protein of the endoplasmic reticulum and nuclear envelope, which is also present in the nucleolus. *Exp Cell Res* *294*, 94-105.

Szostak, E., and Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Brief Funct Genomics* *12*, 58-65.

Taliaferro, J.M., Wang, E.T., and Burge, C.B. (2014). Genomic analysis of RNA localization. *RNA Biol* *11*, 1040-1050.

Tan, R., Chen, L., Buettner, J.A., Hudson, D., and Frankel, A.D. (1993). RNA recognition by an isolated alpha helix. *Cell* *73*, 1031-1040.

Tan, R., and Frankel, A.D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry* *33*, 14579-14585.

Thapar, R. (2015). Structural basis for regulation of RNA-binding proteins by phosphorylation. *ACS Chem Biol* *10*, 652-666.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7, 562-578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511-515.

Turner, M., and Hodson, D.J. (2012). An Emerging Role of RNA-Binding Proteins as Multifunctional Regulators of Lymphocyte Development and Function. *Adv Immunol* 115, 161-185.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212-1215.

Van Roey, K., and Davey, N.E. (2015). Motif co-regulation and co-operativity are common mechanisms in transcriptional, post-transcriptional and post-translational regulation. *Cell Commun Signal* 13, 45.

Varadi, M., Zsolyomi, F., Guharoy, M., and Tompa, P. (2015). Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. *PloS one* 10, e0139731.

Voorhees, R.M., Fernandez, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 angstrom Resolution. *Cell* 157, 1632-1643.

Walter, P., and Johnson, A.E. (1994a). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* 10, 87-119.

Walter, P., and Johnson, A.E. (1994b). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* 10, 87-119.

Walter, P., and Lingappa, V.R. (1986). Mechanism of protein translocation across the endoplasmic reticulum membrane. *Annu Rev Cell Biol* 2, 499-516.

Wang, L., Huang, C., Yang, M.Q., and Yang, J.Y. (2010). BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst Biol* 4 *Suppl 1*, S3.

Wilkie, G.S., Dickson, K.S., and Gray, N.K. (2003). Regulation of mRNA translation by 5' and 3'-UTR-binding factors. *Trends Biochem Sci* 28, 182-188.

Wilkinson, T.A., Zhu, L., Hu, W., and Chen, Y. (2004). Retention of conformational flexibility in HIV-1 Rev-RNA complexes. *Biochemistry* 43, 16153-16160.

Williamson, J.R. (2000). Induced fit in RNA-protein recognition. *Nat Struct Biol* 7, 834-837.

Wilm, M., Shevchenko, A., Houthaave, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379, 466-469.

Xiang, S., Gapsys, V., Kim, H.Y., Bessonov, S., Hsiao, H.H., Mohlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., *et al.* (2013). Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure* 21, 2162-2174.

Xie, C., Turley, S.D., Pentchev, P.G., and Dietschy, J.M. (1999). Cholesterol balance and metabolism in mice with loss of function of Niemann-Pick C protein. *Am J Physiol* 276, E336-344.

Yaniv, K., and Yisraeli, J.K. (2001). Defining cis-acting elements and trans-acting factors in RNA localization. *Int Rev Cytol* 203, 521-539.

Yoo, B.K., Chen, D., Su, Z.Z., Gredler, R., Yoo, J., Shah, K., Fisher, P.B., and Sarkar, D. (2010). Molecular mechanism of chemoresistance by astrocyte elevated gene-1. *Cancer research* 70, 3249-3258.

Yoo, B.K., Emdad, L., Su, Z., Villanueva, A., Chiang, D.Y., Mukhopadhyay, N.D., Mills, A.S., Waxman, S., Fisher, R.A., Llovet, J.M., *et al.* (2009a). Astrocyte elevated gene-1 regulates hepatocellular carcinoma development and progression. *Journal of Clinical Investigation* 119, 465-477.

Yoo, B.K., Emdad, L., Su, Z.Z., Villanueva, A., Chiang, D.Y., Mukhopadhyay, N.D., Mills, A.S., Waxman, S., Fisher, R.A., Llovet, J.M., *et al.* (2009b). Astrocyte elevated gene-1 regulates hepatocellular carcinoma development and progression. *The Journal of clinical investigation* 119, 465-477.

Yoo, B.K., Santhekadur, P.K., Gredler, R., Chen, D., Emdad, L., Bhutia, S., Pannell, L., Fisher, P.B., and Sarkar, D. (2011). Increased RNA-induced silencing complex (RISC) activity contributes to hepatocellular carcinoma. *Hepatology* 53, 1538-1548.

Zarnack, K., and Feldbrugge, M. (2010). Microtubule-dependent mRNA transport in fungi. *Eukaryot Cell* 9, 982-990.

Zhang, C.L., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29, 607-U686.

Zhang, H.L., Eom, T., Oleynikov, Y., Shenoy, S.M., Liebelt, D.A., Dichtenberg, J.B., Singer, R.H., and Bassell, G.J. (2001). Neurotrophin-induced transport of a beta-actin mRNP complex increases beta-actin levels and stimulates growth cone motility. *Neuron* 31, 261-275.

Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. *Biochim Biophys Acta* 1808, 912-924.

Biography

I was born on May 7, 1985 in Taipei City, Taiwan. I grew up in Taoyuan until I went to National Experimental High School in Hsinchu City. In 2003, I joined National Taiwan University (NTU) to pursue a Bachelor of Science in Horticulture Science. I joined the laboratory of Dr. Inn-Ho Tsai in the Institute of Biochemical Sciences (IBS) at NTU as a Research Intern to study the protein sequence of venom phospholipase A2 that led to my interest in biochemistry. I applied and was accepted into several Biochemistry MS programs in Taiwan. In 2007, I came to IBS at NTU and joined the laboratory of Dr. Rita Pei-Yen Chen in the Institute of Biological Chemistry at Academia Sinica. In her laboratory, my research focused on ultrafast protein folding kinetics using a photolabile caging strategy and time-resolved photoacoustic calorimetry¹. After my graduation in 2007, I've served a mandatory military service in Taiwanese Army as a forward observer for one year with a final rank of second lieutenant. In 2008, I returned to Dr. Chen's laboratory as a research technician. My study focused on the spectroscopic properties of a fluorescent indicator for protein aggregation, Thioflavin T². In 2011, I came to Duke University to pursue a Ph.D degree in the Department of Biochemistry. In 2012, I joined Dr. Christopher Nicchitta's research group to investigate the molecular mechanisms of mRNA anchoring to the endoplasmic reticulum^{3,4}.

Publications

1. Chen, H.L.*, **Hsu, J.C.***, Viet, M.H., Li, M.S., Hu, C.K., Liu, C.H., Luh, F.Y., Chen, S.S., Chang, E.S., Wang, A.H., Hsu, M.F., Fann, W., Chen, R.P., 2010, Studying

- submicrosecond protein folding kinetics using a photolabile caging strategy and time-resolved photoacoustic calorimetry. *Proteins*, 78:2973–2983
2. **Hsu, J.C.***, Chen, E.H.*, Snoeberger, R.C., Luh, F.Y., Lim, T.S., Hsu, C.P., Chen, R.P., 2013, Thioflavin T and its photo-irradiative derivatives: Exploring their spectroscopic properties in the absence and presence of amyloid fibrils. *Journal of Physical Chemistry B*, 117:3459–3468
 3. Jagannathan, S.*, **Hsu, J.C.***, Reid, D., Chen, Q., Thompson, W.J., Moseley, A.M. and Nicchitta, C.V., 2013, Multifunctional roles for the protein translocation machinery in RNA anchoring to the endoplasmic reticulum. *Journal of Biological Chemistry*, 289:25907-25924
 4. **Hsu, J.C.**, Reid, D.W., Sankar, D., and Nicchitta, C.V., Astrocyte Elevated Gene-1 (AEG-1) is an endoplasmic reticulum (ER) membrane protein functioning in selective mRNA anchoring to the ER. (under review).

*These authors made an equal contribution.