이학박사학위논문

다세포생물체 내에서의 RNA-단백질 간의 상호 작용에 대한 단백체학적 연구

Proteome-wide study on RNA-protein interactions in multicellular organisms

2021년 2월

서울대학교 대학원
생명과학부
나용우

Proteome-wide study on RNA-protein interactions in multicellular organisms

Advisor: Professor V. Narry Kim

Submitting a doctoral thesis of philosophy
Feburary, 2021

Graduate School of Seoul National University
School of Biological Sciences
Yongwoo Na
Confirming the doctoral thesis written by Yongwoo Na
December, 2020

Chair Young-Tae Kang (seal)

Vice Chair NARRY KIM (seal)

Examiner Ji-Young Lee (seal)

Examiner Jong-Seo Kim (seal)

Examiner SANGWON CHA (seal)

Abstract

Proteome-wide study on RNA-protein interactions in multicellular organisms

Yongwoo Na
School of Biological Sciences
The Graduate School
Seoul National University

From synthesis to decay, mRNA is bound with tens of RNA binding proteins (RBPs) and exists as a messenger ribonucleoprotein (mRNP) complex. Proper expression and function of mRNA in the biological systems are dependent on highly coordinated and dynamic change in the profile of RBPs over the course of mRNA transcription, translation, and decay. Distinct target and context specific RNA-protein interactions are thus central to many of the post-transcriptional gene regulatory mechanisms. Recently developed RBP profiling techniques rely on the use of various methods that crosslink RNA-protein interactions *in vivo*. Currently, UV light induced crosslinking (UVX) is the most widely utilized *in vivo* crosslinking method in RNA biology.

Nonetheless UVX has notable limitations, including the limited applicability to the tissues of multicellular organisms, due to its limited depth of penetration in the biological systems.

Here I introduce formaldehyde crosslinking (FAX) as an alternative chemical crosslinking method for RNA interactome capture (RIC). FAX-RIC captured the RNA-protein interactions with high specificity and efficiency in cell culture. Further analysis of the UVX or FAX preferentially enriched RBPs revealed the distinct crosslinking specificity of the two methods.

FAX can be readily reversed with the high temperature. Utilizing this unique property, I developed the peptide-level FAX-RIC method to directly identify the RNA crosslinking site within the RBPs with tryptic peptide resolution.

FAX-RIC method was then applied to the *Xenopus laevis* oocytes and embryos and compared with the respective result obtained by the UVX-RIC. The results demonstrate that FAX-RIC can enable highly comprehensive and relatively unbiased RNA interactome capture in multicellular organisms *in vivo*. Furthermore, quantitative comparison of the oocyte and embryo FAX RNA interactome revealed the dynamic remodeling of RNA-protein complex during oocyte to embryo transition (OET). FAX-RIC result was also compared to the total protein expression level change in OET. From this analysis I

defined the OET specific dynamic RBPs whose enrichment rate change in FAX-RIC cannot be explained by the change in their protein expression level. Notably, I observed the significant change in the critical translation initiation factors during the OET, for instance, from canonical eIF4E to non-canonical eIF4E3.

I utilized the peptide-level FAX-RIC method and developed a strategy for both reliable and versatile RNA interactome capture experiment in mammalian tissue samples. The newly developed protocol was applied to mouse liver for the determination of both poly A and total RNA interactome profile.

Taken together, I developed and thoroughly investigated the utility of FAX based RIC method in wide range of biological samples from cultured human cell lines to *X. laevis* embryo and *M. musculus* liver. The result in HeLa cells demonstrate how use of either UV light and formaldehyde based *in vivo* crosslinking methods can significantly change the outcome of high throughput system wide RBP profiling methods. I then provide highly comprehensive RNA interactome profile of vertebrate embryo and mammalian tissue for the first time *in vivo*. Dynamic RNP complex remodeling in the animal oocyte to embryo transition revealed in this study can be the basis for further study into the individual RBP's regulatory mechanism and exact importance in the early

animal development. Applicability of the FAX-RIC approach to the mammalian tissue warrants its broad future use in the studies of animal tissue homeostasis and human diseases.

Key words

messenger RNA (mRNA); RNA binding protein (RBP); UV crosslinking (UVX); formaldehyde crosslinking (FAX); liquid chromatography (LC); tandem mass spectrometry (MS/MS); *X. laevis*; oocyte to embryo transition (OET); *M. musculus*; liver

CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

DNA, deoxyribonucleic acid

dsRNA, double stranded RNA

FAX, formaldehyde crosslinking

FDR, false discovery rate

GO, gene ontology

GITC, guanidinium isothiocyanate

LC-MS/MS, liquid chromatography coupled tandem mass spectrometry

LFQ, label free quantification

mRNA, messenger RNA

OET, oocyte to embryo transition

qRT-PCR, quantitative real-time polymerase chain reaction

RBD, RNA-binding domain

RBP, RNA –binding protein

RIC, RNA interactome capture

RNA, ribonucleic acid

RNP, ribonucleoprotein complex

rRNA, ribosomal RNA

tRNA, transfer RNA

UVX, ultraviolet crosslinking

Chapter I.

Introduction

I-1. RBPs and the post transcriptional regulation of mRNA

Upon transcription, a messenger RNA is bound with the tens of RBPs and thereby constitutes a messenger ribonucleoprotein (mRNP) complex (Dreyfuss et al., 2002). Dynamic transformation of the RBP profile then drive subsequent processing steps of mRNA, namely, capping of the 5' end, splicing, and 3' end processing and polyadenylation by the cleavage and polyadenylation specificity factors (CPSFs) and poly(A) polymerases (PAPs) (Dreyfuss et al., 2002) (**Figure I-1**). Successfully processed, mature mRNA is exported to the cytoplasm with the help of mRNA transport factors such as the transcription-export (TREX) complex proteins (Carmody and Wente, 2009). In the cytoplasm, proper exchange of the nuclear and cytoplasmic RBPs on mRNA is required for translation and protein expression to occur (Hentze and Kulozik, 1999). The nonsense mediated decay (NMD) pathway, is a notable example for the mRNA quality control mechanisms that are based on the proper change in the profile of RBPs. In NMD, incomplete removal of exon junction complex (EJC) proteins by the translating ribosomes, due to the presence of the premature termination codons (PTCs), act as the signal for the removal of such aberrant mRNA (Dreyfuss et al., 2002). Fully functional mature mRNA is also subject to various mRNA decay pathways, which can determine the stability, and thus half-life, of mRNA (Houseley and Tollervey,

2009).

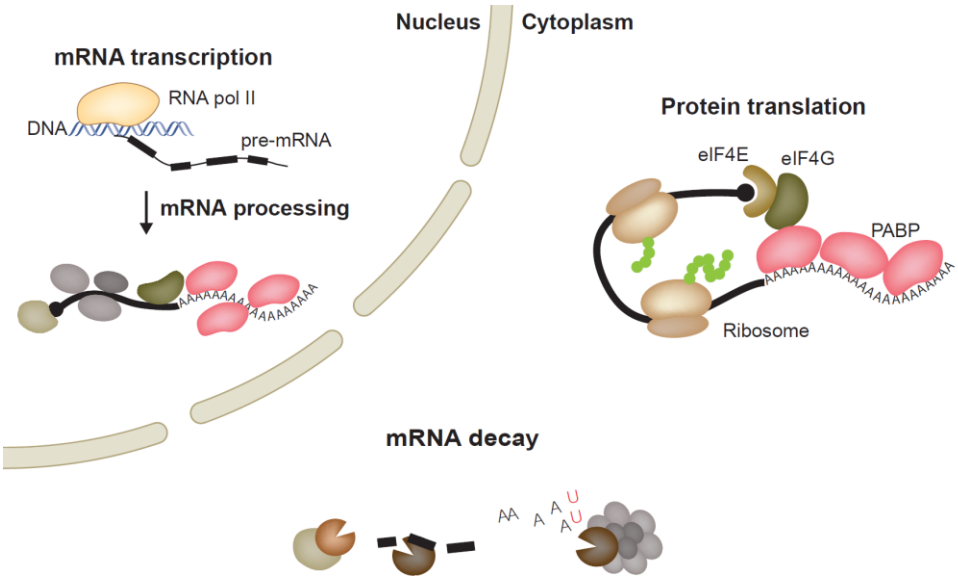


Figure I-1. The life of mRNA

Name of the distinct RBPs, involved in mRNA transcription and translation, are indicated.

Primary function of mRNA is to code for the protein. Highly intricate translational machinery, such as the ribosomes and the transfer RNAs (tRNAs), mediate highly efficient and accurate protein synthesis, and their association with mRNA is dependent on the presence of distinct translation initiation factors (Chu et al., 2016; Dreyfuss et al., 2002). Many of the post transcriptional regulatory mechanisms thus target the interactions between mRNA and the translation initiation factor proteins. One of the well-recognized targets for regulation is the phosphorylation state of the 4EBP1 protein, which affects the interaction between the 4EBP1 and the EIF4E proteins (Richter and Sonenberg, 2005). Phosphorylated 4EBP1 protein dissociates from the EIF4E protein, which can then bind mRNA cap to promote the translation of mRNA in canonical translation pathway (Richter and Sonenberg, 2005). In recent years, phase separation of RNP complexes has also been widely studied as the mRNA translation regulatory mechanisms (Langdon and Gladfelter, 2018). Most notably, in the stress conditions such as oxidative stress, the mRNA molecules are sequestered in the stress granule, whose formation is mediated by the RBPs such as the G3BP1 protein (Somasekharan et al., 2020). Taken together, change in the repertoire of RBPs can influence both the expression level and translation activity of mRNA, and thus central to posttranscriptional-regulation of gene expression level.

I-2. RNA interactome capture and the repertoire of RBPs in cultured cells and organisms

To understand the network of RNA-protein interactions, several methods have been developed to profile RBPs at the proteomic scale (Baltz et al., 2012; Castello et al., 2012; Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019). Among them, RNA interactome capture (RIC) method is based on the oligo-dT bead pulldown and mass spectrometry of the RNA-protein conjugates (Baltz et al., 2012; Castello et al., 2012; Hentze et al., 2018). Development and application of the RIC technique has significantly expanded the RBP repertoire by discovering hundreds of the unorthodox RBPs without any known RNA-binding domains (RBDs) or RNA-related functions (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012; Hentze et al., 2018; Perez-Perri et al., 2018). In the previous RIC experiments, *in vivo* crosslinking was induced by irradiation of ultraviolet light (254 or 365 nm) on cells (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012; Hentze et al., 2018; Perez-Perri et al., 2018). However, UV light has an innate weakness owing to its highly limited penetration depth (Elinson and Pasceri, 1989; Sysoev et al., 2016). With UVX-RIC, it is hard to profile the RNA interactome of large or opaque samples due to its low efficiency and the inevitable bias toward the molecules on the surface, as pointed out previously in a UVX-based study on

Drosophila embryo (Sysoev et al., 2016). This obvious limitation of current RIC method suggest that development of a new method is required for the future study on the RNA interactome profile of multicellular organisms.

I-3. Formaldehyde crosslinking and RNA-protein interactions

Formaldehyde crosslinking (FAX) can be a promising alternative because of the high membrane permeability of formaldehyde (Thavarajah et al., 2012). Although FAX has been used to characterize RNA-protein (RNP) complexes in a number of studies (Chu et al., 2015; Kim et al., 2017; Knoener et al., 2017; Panhale et al., 2019; Yong et al., 2010), the primary concern has been its selectivity (Panhale et al., 2019; Wheeler et al., 2018) because it is generally thought that formaldehyde crosslinks promiscuously many biomolecules with various functional groups. However, the reported mechanisms of FAX indicate that formaldehyde is highly selective to crosslinking between nucleophiles such as amines via Schiff base formation and nucleophilic addition (Gavrilov et al., 2015; Hoffman et al., 2015). The majority of RNA nucleobases retain exo-amino groups, and the RNA-binding motifs often contain amino groups such as lysine (Beckmann et al., 2015). Formaldehyde is a small chemical crosslinker with short molecular span which is thought to be ~2 Å apart (Hoffman et al., 2015). Such knowledge suggests that amino acids closely

located to the nucleobases of RNA can be preferentially crosslinked by formaldehyde treatment, compared to random and transient interactions with other macromolecules. Nevertheless, FAX has only been used sparsely in RNA biology and the use has also been largely restricted to the probing of structured or duplex RNA-protein interactions (Kim et al., 2017; Kim et al., 2014; Singh et al., 2012). Therefore, both the specificity and potential benefits of FAX for probing *in vivo* RNA-protein interactions remained to be investigated through comprehensive system-wide analysis.

Here, I report that FAX can enable comprehensive and reliable RIC studies within the diverse biological systems, from cultured cells to *X. laevis* oocyte and embryo, and *M. musculus* liver. Systematic and quantitative comparison between RNA interactome profiles from the FAX-RIC and UVX-RIC in HeLa cells disclosed the distinct characteristics of two crosslinking methods and suggested the relatively high specificity of FAX-RIC. I further demonstrate the specificity of FAX to RNA binding domains, by developing a peptide-level FAX-RIC protocol. Using FAX-RIC, I profiled the changes in the RNA interactome landscape during oocyte-to-embryo transition (OET) for the first time *in vivo*. Furthermore, I tested the applicability of the FAX-RIC approach to mammalian tissues by utilizing mouse liver as a model. The findings of this

study will significantly broaden our understanding of mRNP complex remodeling in multicellular organisms *in vivo*.

Chapter II.

Materials and methods

HeLa cell culture

HeLa cells were cultured in DMEM (Welgene) supplemented with 9% fetal bovine serum (Welgene) and maintained in a humid incubator at 37 °C in a 5% CO2 environment to reach a density of 1 × 106 cells/mL. The HeLa cell line is a modified HeLa with TUT4 gene deletion.

Formaldehyde crosslinking (FAX) and RNA interactome capture (RIC) in HeLa Cell

Formaldehyde treatment condition was first optimized in HeLa cell. Formaldehyde in phosphate buffered saline (PBS), prepared with 16% formaldehyde (w/v), methanol-free (Thermo Scientific), with increasing concentration of 0.1, 0.3, 0.5, 0.7 and 1.0% were applied to HeLa cell for 5 min and oligo-dT beads (NE Biolabs) enriched protein amount from each condition was checked by SDS page gel running and silver staining. For respective replicate of RIC experiment, we plated the HeLa cells on 5 x 145 cm^2 dish to reach confluence overnight, resulting in the cell number of ~12.5 × 10^7 before crosslinking. FAX was done by directly applying formaldehyde in PBS to HeLa cells on plate. Briefly, cells were washed twice with PBS at room temperature (RT) and incubated with 0.5% formaldehyde in PBS for 5

min at RT. Formaldehyde solution was removed at 5 min and the cells were washed twice with ice-cold 200 mM Tris in PBS, with 30 seconds incubation time each, to quench the residual formaldehyde reaction. Cells were collected from the plate by scraping and washed twice with ice cold PBS by centrifuge. Cells were lysed in the lysis buffer (0.5% (w/v) lithium dodecyl sulfate (LDS), 500 mM lithium chloride (LiCl), 10 mM Tris (pH 7.5), 2 mM EDTA, and 5 mM dithiothreitol (DTT) (all Sigma)) and sheared by passing through 21-gauge needle syringe for ten times. Oligo-dT bead was added to the lysate and incubation were done with over and over rotation for 1 hour at 4 ˚C. Beads were separated from the cell lysate using DynaMag (Thermo scientific) and washed twice in each buffer, the lysis buffer, low LDS lysis buffer (lysis buffer made with 0.1% (w/v) LDS), high salt buffer (lysis buffer made without LDS) and low salt buffer (lysis buffer made without LDS and 200 mM LiCl). Oligo-dT beads were then incubated with Turbo DNase (Thermo Scientific) in Turbo DNase buffer supplemented with 200 mM of LiCl, for 30 min with over and over rotation at RT. Beads were then washed twice with each wash buffer. Elution of the poly(A) RNA by heat was done twice in TE buffer, by incubation on Thermomixer C (Eppendorf) at 65 ˚C for 3minutes with mixing at 800 rpm. Peptide sample preparation for LC-MS/MS analysis were done with Microcon-30kDa Centrifugal Filter Unit (Millipore), following the previously described protocol (Castello et al.). Final peptide sample was

desalted using Discovery DSC-18 SPE Tube (Supelco) and prepared for LC-MS/MS analysis.

Peptide-level FAX-RIC

Formaldehyde treatment and cell lysate preparation was done as described previously for protein-level FAX-RIC. HeLa cell lysate was then diluted with Tris EDTA (TE) buffer to make LDS 0.07% (w/v) and LiCl 70 mM. MS grade trypsin (Thermo scientific) was added to the cell lysate at protein weight ratio of 1:100 followed by incubation with over and over rotation for 8 hours at RT. The cell lysate was then made to contain 0.5% LDS and 500 mM LiCl for the oligo-dT bead pulldown. Oligo-dT bead pulldown and washing steps were done as described previously for protein-level FAX-RIC. RNAs were eluted from the bead by adding 8 M urea in TE buffer, twice. Samples were then filtered down 4 times with 8 M urea TE buffer and 3 times with TE buffer in 100 kDa Amicon filter (Millipore). The sample in TE buffer was incubated at 65 ˚C on Thermomixer C overnight. The samples were subjected to the conventional trypsin digestion protocol for LC-MS/MS analysis and desalted using Millipore ZipTip with C18 resin.

RIC via UV crosslinking (UVX) in HeLa Cell

Cell culture condition and scale was identical to that described previously for the FAX-RIC. For UVX, cells were washed twice with PBS at RT and after removing the PBS, irradiated with 450 mJ (~60 seconds) of UV light (254 nm), using the Spectrolinker XL-1500 UV crosslinker (Spectronics) on ice, as it was previously described (3). Following procedures for oligo-dT capture were identical to the above described procedure for the FAX-RIC.

***Xenopus laevis* oocyte and embryo preparation**

X. laevis oocytes were obtained from excised ovary of female *X. laevis* as previously described (Sive et al., 2010). *X. laevis* embryos collection were also done as described previously (Sive et al., 2007). Briefly, human chorionic gonadotropin was injected into a female frog 12 hours before collecting eggs. The eggs were obtained in 1X Marc's Modified Ringer's (MMR) solution and in vitro fertilized using excised testes from a male frog.

RNA interactome profiling in *X. laevis* oocyte and embryo

X. laevis embryo collected as described above were washed three times with PBS and then treated with 2% formaldehyde in PBS for 10 min with gentle

rotation. Formaldehyde reaction was quenched by treating the embryo with 200 mM tris in PBS for 5 min and then washed with ice cold PBS three times. Crosslinked *X. laevis* embryo was first lysed with high salt lysis buffer (0.5% lithium dodecyl sulfate, 1 M LiCl, 20 mM Tris (pH 7.5), 5 mM DTT, 1 mM EDTA), and then 8 M urea lysis buffer (8 M urea, 0.5% lithium dodecyl sulfate, 20 mM Tris (pH 7.5), 5 mM DTT, 1 mM EDTA) was added to adjust the concentration of urea and lithium chloride in sample lysate to become 4 M and 500 mM, respectively. I found that use of high salt lysis buffer and the urea in sample lysis buffer is crucial to the integrity of RNA for the preparation of *X. laevis* oocyte and embryo lysate (data not shown). RIC experiment in *X. laevis* oocyte and embryo were all performed in triplicate. Following procedures for RIC was identical to the above described procedure for the HeLa cell. FAX-RIC for *X. laevis* oocyte was identical to that of the protocol for the embryo. For UVX, prepared *X. laevis* oocytes and embryos were placed in minimal amount of PBS to cover the whole embryo on plate and irradiated with 500 mJ of UV light (254 nm) on ice for four times, with agitation of plate after each irradiation to turn around the embryo, resulting in total 2 J of UV light irradiation. Embryos were washed with PBS twice after UV light irradiation. Embryo lysis and oligo-dT capture procedure was identical to the above described procedure for *X. laevis* FAX-RIC.

TMT based quantitative profiling of global protein expression level *in X. laevis* oocyte

Protein samples in 0 h and 8 h after the progesterone induced oocyte maturation were prepared as previously reported (Peuchen et al., 2017). Instrumental settings for LC-MS3 analysis and data analysis parameters were largely the same as the case of our previous work (Jung et al., 2019). The final protein quantification results were taken from the biological duplicate experiments.

RNA interactome profiling in mouse liver using peptide-level FAX-RIC

Mice liver samples were a kind gift from Hyun-Woo Rhee Lab (SNU). Animal experiments with C57BL/6 mice were performed in accordance with the governmental and institutional laws and recommendations (Approval no. SNU-180521-2-3). Mouse liver tissue was cut by a lobe and submerged in the formaldehyde solution of 4% in PBS for 10 minutes with occasional shaking. Crosslinking reaction was quenched by submerging the tissue in 200 mM Tris (pH 7.0) in PBS buffer for 5 minutes. The tissue was weighed on scale and ~20 mg of liver samples was used for each of three replicate dT pull down experiments and total RNA interactome extraction experiments via RNeasy

column (Qiagen). The tissue was first lysed in 4M GuSCN, 800 mM LiCl, 10 mM DTT, 5 mM EDTA and 20 mM HEPES (pH 7.5), and 3 volumes of ethanol was added to the sample followed by 1hr incubation at -20 °C. Precipitated samples were centrifuged down at x16000 g for 15 minutes and supernatant was removed followed by 70% ethanol wash twice at x8000 g. The sample was reconstituted with the 0.05% LDS, 100 mM LiCl, 10 mM DTT, 5 mM EDTA, and 20 mM Tris (pH 7.5) and trypsin was added 1:50 (w/w) and incubated at RT for 4 hours. Prepared lysate was subjected to peptide-level FAX-RIC protocol as described above for HeLa cells. For total RNA interactome experiment, after the trypsin treatment twice the volume of RLT buffer was added to the sample and RNA sample preparation by RNeasy column was done as described in the product manual followed by heat incubation 65 °C overnight and LC-MS/MS analysis.

LC-MS/MS analysis

Analytical capillary columns (100 cm x 75 µm i.d.) and trap columns (3 cm x 150 µm i.d) were packed in-house with 3 µm Jupiter C18 particles (Phenomenex, Torrance, CA). The long analytical column was placed in a column heater (Analytical Sales and Services, Pompton Plains, NJ) regulated to a temperature of 45 °C. Ultimate 3000 nanoRSLC system (Thermo

Scientific, Sunnyvale, CA) was operated at a flow rate of 350 nL/min over 2 hours with linear gradient ranging from 95% solvent A (water with 0.1% formic acid) to 40% of solvent B (acetonitrile with 0.1% formic acid). The enriched samples were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) equipped with an in-house customized nanoelectrospray ion source. Precursor ions were acquired (m/z 300 – 1500 at 120k resolving power and the isolation of precursor for MS/MS analysis was performed with a 1.4 Th. Higher-energy collisional dissociation (HCD) with 30% collision energy was used for sequencing with a target value of 5E4 ions determined by automatic gain control. Resolving power for acquired MS2 spectra was set to 30k at m/z 200 with 150 ms maximum injection time.

Protein identification

MS raw data files were processed with MaxQuant (version 1.5.3.30) (Tyanova et al., 2016). Enzyme specificity was set to trypsin/P and a maximum of two missed cleavages were allowed. Cysteine carbamidomethylation and methionine oxidation were selected as fixed and variable modifications, respectively. The derived peak list was searched using the built-in Andromeda search engine in MaxQuant against the human UniProt database (version 2/2018) or ‘X. laevis protein (Xenbase)’ fasta file (2/14/2018 version)

downloaded from Xenbase website. Initial maximal allowed mass tolerance was set to 20 ppm for peptide masses, followed by 6 ppm in the main search, and 0.5 Da for fragment ion masses. The minimum peptide length was set to six amino acid residues, and three labelled amino-acid residues were allowed. A false discovery rate (FDR) < 0.01 was required at both the protein-level and the peptide-level. Label free quantification was turned on and applied to each experimental group separately, except for the comparison between *X. laevis* oocyte and embryo FAX.

Statistical analysis for label-free quantification (LFQ)

Statistical analysis for defining RNA interactome and quantitative comparison between each experiments were done with PERSEUS software (Tyanova et al.). For defining the RNA interactome, protein groups with LFQ value in two or more replicates in each experiment were used for statistical analysis. Missing LFQ values were imputed with a normal distribution shifted by -4 and sharpened with a standard deviation factor of 0.3. Student's t-test was performed to test if any log2 fold-change ratio is different from 0 and protein groups with Benjamini Hochberg FDR < 0.01 were considered as RNA interactome. Same statistical analysis procedure was used to test for differentially captured protein groups between individual RNA interactome.

Protein groups with Benjamini Hochberg FDR < 0.05 were considered to be significantly enriched in certain RNA interactome.

Gene ontology (GO) and domain analysis

GO and PFAM domain annotations of the identified proteins were retrieved using ENSEMBL Biomart and the Ensembl Human release 81 (GRCh38.p3). GO term enrichment analysis was done using the DAVID tool. Human orthologues of X. laevis proteins were retrieved from UniProt database by matching the gene name of X. laevis proteins with that of human proteins.

SDS–PAGE and western blotting

Input cell lysate and oligo-dT enriched samples were all treated with benzonase and RNase A/T1 and sonicated for 15 min by Bioruptor (COSMO BIO). Protein samples were separated on Bolt 4-12% Bis-Tris Plus Gel (Thermo Fisher Scientific) and transferred to polyvinylidene fluoride membrane (GE Healthcare). Western blotting was performed as previously described (32). The following antibodies were used for western blotting at 1:1000 dilution in PBS containing 1% skim milk and 0.1% Tween 20 (USB): anti-AGO1 (Cell signaling), anti-EIF4E (Cell signaling), anti-PABPC1 (Gift

from Dr. Dreyfuss' lab), anti-Tubulin (Abcam), DDX19B (Novus Bio), GAPDH (Santa Cruz), and ENO1 (Proteintech Group).

qPCR analysis

Oligo-dT pulled down of RNAs were all treated with proteinase K in proteinase K buffer (0.5% SDS, 100 mM NaCl, 1 mM EDTA and 50 mM Tris (pH 7.5) at 65 °C for 4 hours. For total RNAs, HeLa cell lysate in lysis buffer for oligo-dT pull down was treated with the proteinase K at 65 °C for 4 hours. The RNAs were then purified by the Trizol reagent (Invitrogen) and reverse-transcribed with 5x RT Master Mix (Takara) and the RNA levels were measured with SYBR Green assays (Applied biosystems) with primers against 18S rRNA (forward: GAAACTGCGAATGGCTCATTAAA, reverse: CACAGTTATCCAAGTGGGAGAGG), EEF2 (f: AACTTCACGGTAGACCAGATCC, r: TCGTCCTTCCGGGTATCAGTG), and TS (f: GGCAGAATACAGAGATATGGAATCAGA, r: TCGTCAGGGTTGGTTTTGATG).

Chapter III.

Results

III-1. FAX-RIC profiled the known RBPs with high specificity in HeLa cell

In order to find the optimal FAX condition for RNA-protein interactions, I screened a series of mild FAX conditions in HeLa cells and the RNP was pulled down with oligo-dT (**Figure III-1**). The total amount of the precipitated proteins was significantly greater than that of UVX-RIC even at concentrations as low as 0.1-0.5% formaldehyde (w/v) for 5 min (**Figure III-2**). The protein profiles were relatively constant across the tested FAX conditions in comparison with UVX-captured protein profiles (**Figure III-2**). Through the comparative qPCR analysis of the input and oligo-dT enriched RNA samples, I found that FAX did not impair the RNA pull down efficiency nor the specificity (**Figure III-3**).

I performed western blot analysis for the initial assessment of the relative specificity and the enhanced efficiency of FAX-RIC to the representative RBPs along with a negative control protein, Tubulin A (**Figure III-4**). Moreover, when the RBPs were eluted using RNase A/T1 treatment, instead of heat treatment, to prevent potential reversal of formaldehyde crosslinking, the western blot bands did not shift upward (**Figure III-5**). This suggested that their enrichment via FAX-RIC is dependent on RNA-protein crosslinking rather than protein-protein crosslinking.

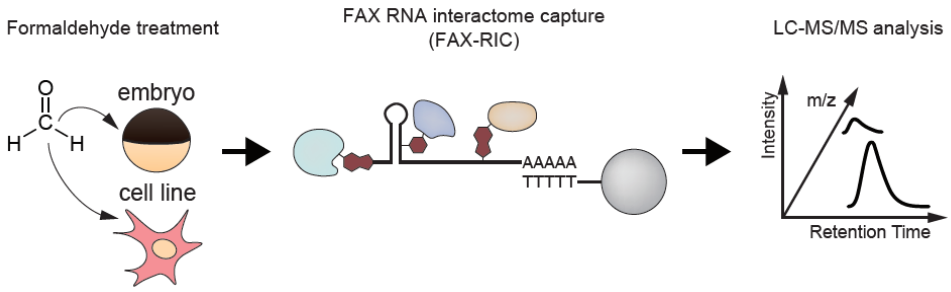


Figure III-1. Schematic outline of the FAX based RNA interactome capture (FAX-RIC)

Schematic outline of the FAX based RNA interactome capture (FAX-RIC) method. Biological samples are treated with formaldehyde solution in PBS at conditions that are separately optimized to form covalent bond between proximal RNA-protein interactions. RNA crosslinked proteins are enriched through oligo-dT pulldown of poly A tailed RNA. Profile of enriched protein samples are obtained via LC-MS/MS analysis.

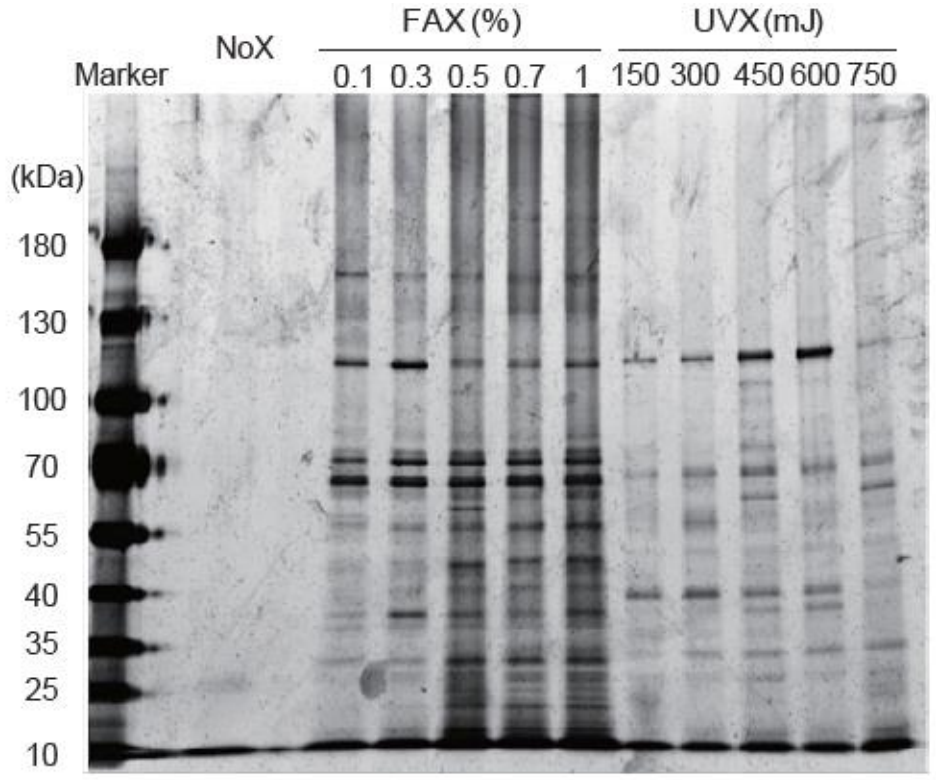


Figure III-2. Optimization of crosslinking conditions in HeLa cell

SDS-PAGE and silver staining of the oligo-dT pulldown samples from the lysate of HeLa cells that are treated with indicated formaldehyde and UV light crosslinking conditions.

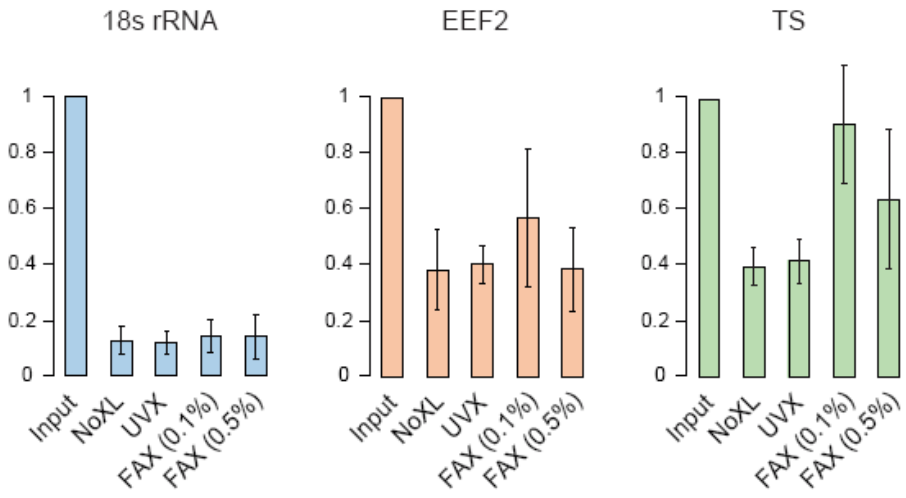


Figure III-3. mRNA specificity of the FAX-RIC

Relative enrichment of EEF2 and TS mRNA transcripts in both UVX- and FAX-RIC enriched RNA over the input cell lysate compared to 18s rRNAs are confirmed by RT-qPCR.

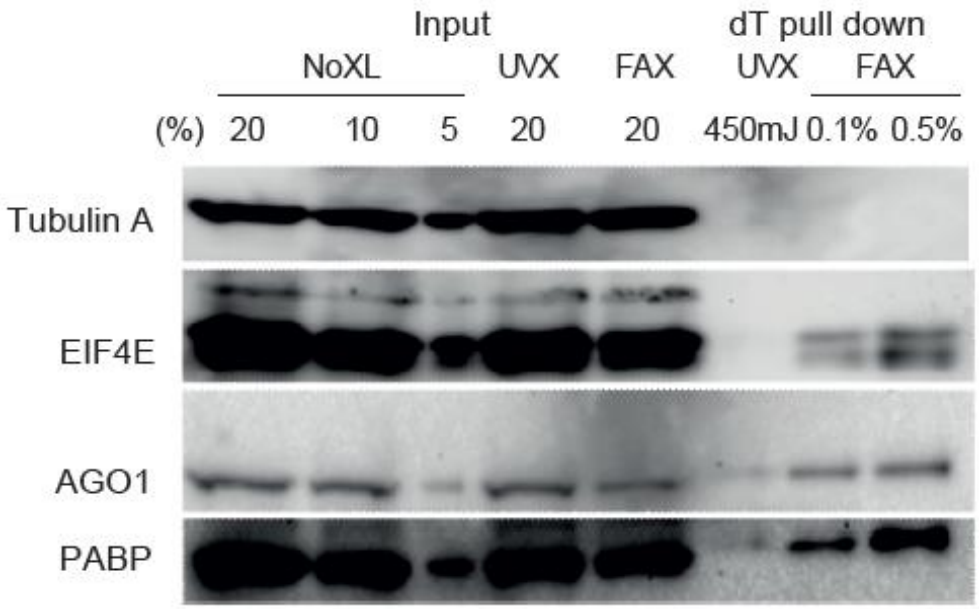
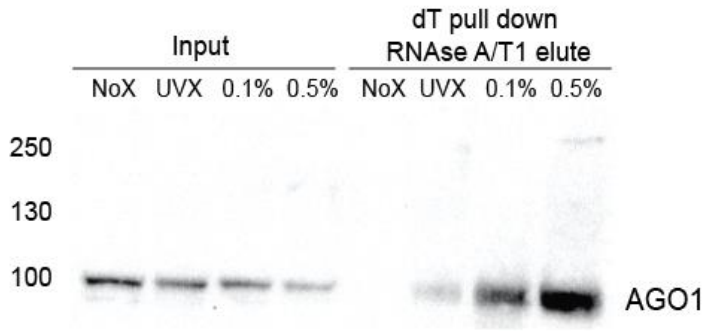
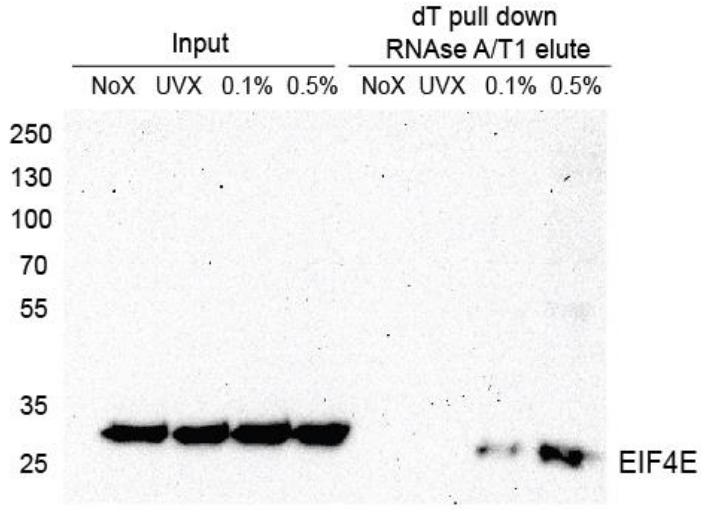


Figure III-4. Western blot analysis for representative RBPs

Western blot analysis for representative RBPs (EIF4E, AGO1 and PABP) and a negative control protein (Tubulin A).


Input
dT pull down
RNAse A/T1 elute
NoX UVX 0.1% 0.5% NoX UVX 0.1% 0.5%
250
130
100
AGO1


Input
dT pull down
RNAse A/T1 elute
NoX UVX 0.1% 0.5% NoX UVX 0.1% 0.5%
250
130
100
70
55
35
25
EIF4E

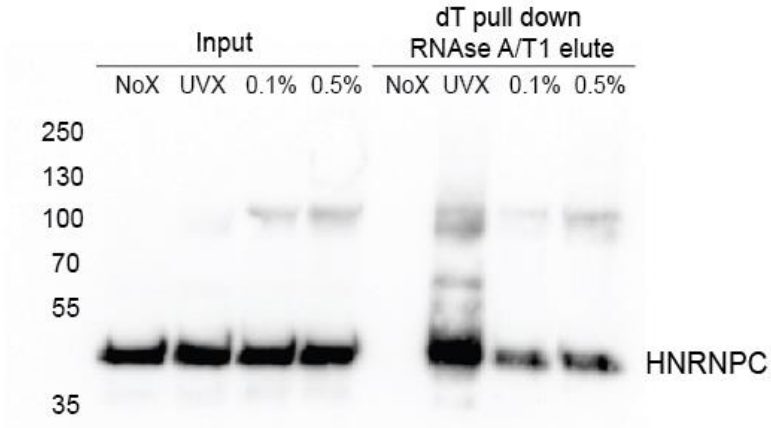

Input
dT pull down
RNAse A/T1 elute
NoX UVX 0.1% 0.5% NoX UVX 0.1% 0.5%
250
130
100
70
55
35
HNRNPC

Figure III-5. Specificity of FAX-RIC to the direct RNA-protein interactions

Western blot analysis of AGO1, EIF4E and HNRNPC for the input and oligo-dT enriched and RNAse A/T1 eluted protein samples from HeLa cells treated with indicated crosslinking conditions. Both input and enriched proteins were subjected to SDS-PAGE without the heat treatment to prevent the reversal of formaldehyde crosslinking.

I carried out quantitative proteomic profiling of FAX-captured proteins (**Figure III-6**). A total of 912 proteins passed the LFQ intensity-based quantitative filtering criteria against the no crosslinking (NoX) control at a < 0.01 false discovery rate (FDR). These proteins are regarded as the "FAX RNA interactome" in HeLa cells (**Figure III-7**). I found that > 94% (861 proteins) of the FAX RNA interactome have been reported as RBPs in previous RBP profiling studies (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012; Queiroz et al., 2019; Trendel et al., 2019) (**Figure III-8**). The proportions of RBPs with known RBDs (Castello et al., 2012) (**Figure III-9A**) or those detected by RBDmap (Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019) (**Figure III-9B**) were comparable between the FAX RNA interactome and the combined list of the UVX RNA interactomes (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012), which I referred to as "REF-UVX", demonstrating that FAX-RIC can capture both conventional RBPs and unorthodox RBPs.

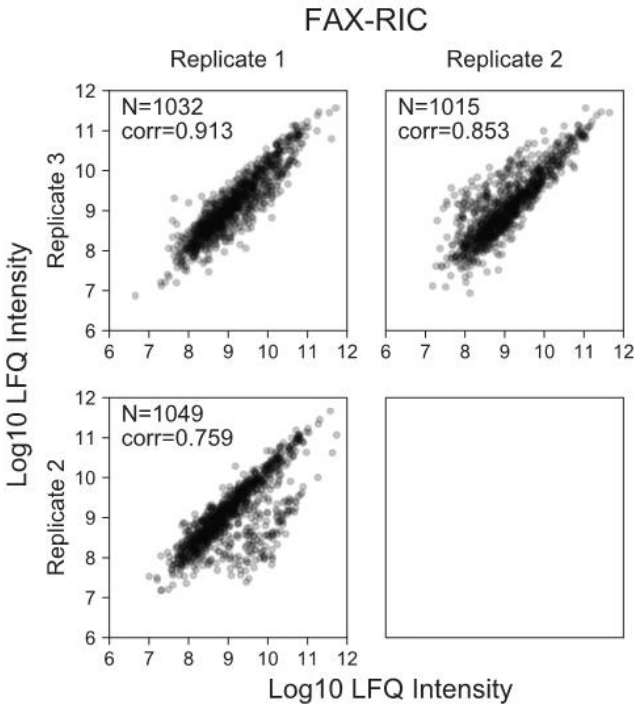


Figure III-6. Reproducibility of FAX-RIC in HeLa cell

Scatter plots of protein LFQ intensities between replicate FAX-RIC experiments in HeLa cell. Protein number (N) and correlation coefficient (corr) for the plot are inserted.

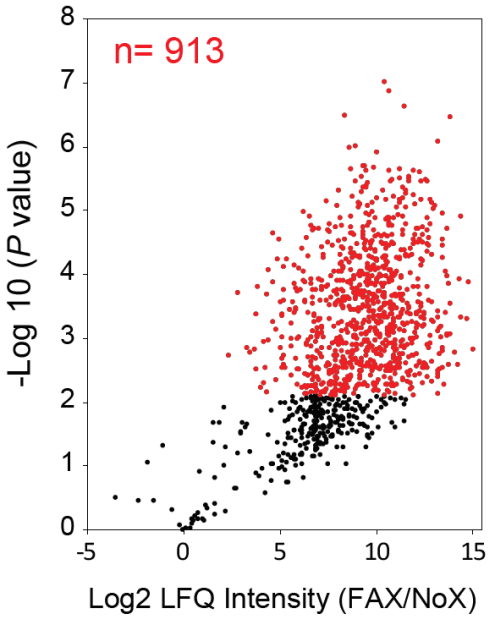


Figure III-7. Defining the high confidence FAX RNA interactome in HeLa cell

Defining the high confidence FAX RNA interactome in HeLa cell. Volcano plot displaying the fold-change of average LFQ intensity (FAX-RIC/NoX-RIC) (x-axis) and the –log10 student's t-test P value (y-axis) for all the proteins quantified in at least two out of three replicate FAX-RIC experiment. Proteins with log2 fold-change >1 and statistically significant enrichment over the NoX-RIC (P value < 0.01, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red.

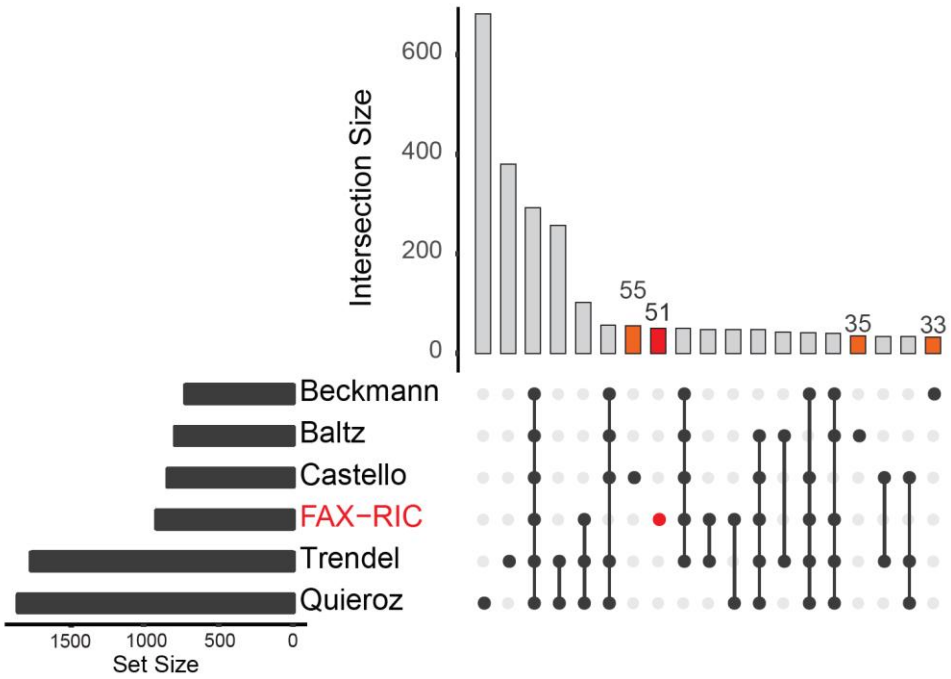

Intersection Size
600
400
200
0
55
51
35
33
Beckmann
Baltz
Castello
FAX−RIC
Trendel
Quieroz
1500
1000
500
0
Set Size

Figure III-8. FAX-RIC and the reported RNA interactome

UpSet plot for the number of proteins that are identified in indicated group of RBP profiles obtained from the FAX-RIC in HeLa cell and previous UVX based RBP profiling studies (Baltz et al., 2012; Beckmann et al., 2016; Castello et al., 2012; Queiroz et al., 2019; Trendel et al., 2019). Each bar on the plot represent the number of proteins that were identified in single or multiple RNA interactome profiles denoted by the black dots below for respective studies whose name and the number of identified proteins are indicated on the left column. For example, the first and second bar on the plot represent the number of RBPs that were exclusive to Trendel et al. (Trendel et al., 2019) and Queiroz et al. (Queiroz et al., 2019) study, respectively. The third bar represent the number of RBPs that were common to all 6 RNA interactome profile, indicated by the 6 black dots joined by a solid line. Number of the proteins that are exclusive to FAX-RIC and three representative RIC experiments (Baltz et al., 2012; Beckmann et al., 2016; Castello et al., 2012) are highlighted in red or orange, respectively.

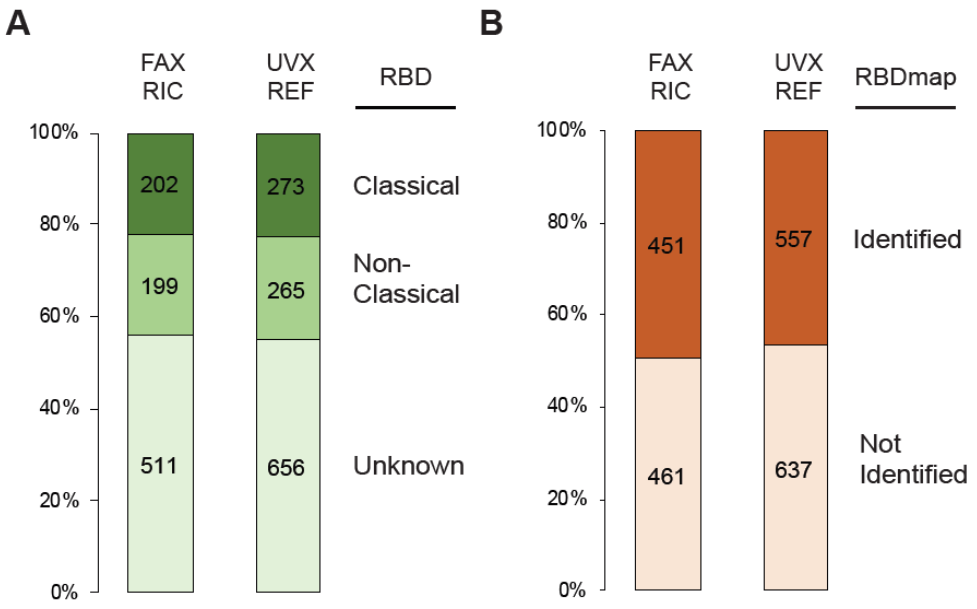


Figure III-9. Proportion of the RBPs with known RBDs in FAX-RIC and previous RIC

(A) Number and proportion of proteins annotated with the known RBDs, either 'classical' or 'non-classical' as defined previously (Castello et al.). 'UVX-REF' include all the proteins identified in three representative RIC experiment (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012) (B) Composition of the proteins with or without RNA interacting region defined in previous RBDmap based studies (Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019).

III-2. Peptide-level FAX-RIC profiled the sites of RNA-protein interaction

Unlike UV-induced crosslink, formaldehyde-induced crosslink can be reversed at high temperature (Panhale et al., 2019). Taking advantage of this property, I modified FAX-RIC by treating cell lysates with trypsin before oligo-dT pulldown (**Figure III-10**). Subsequently, the peptide fragments directly linked to RNA were purified via oligo-dT pulldown, size selective filtration, and decrosslinking by heat. The non-crosslinked peptides are depleted via stringent washing and filtration. This method allows the direct identification RNA-binding regions of RBPs.

The results from peptide-level FAX-RIC were highly reproducible between replicates (**Figure III-11**), resulting in 382 significantly enriched proteins (compared to the NoX control; < 0.05 FDR) (**Figure III-12**). Over 80% of RNA interactome obtained by peptide-level FAX-RIC overlapped with that of the protein-level FAX-RIC. I compared the protein profile from both FAX-RICs to that from RBDmap (Castello et al., 2016) (**Figure III-12**). RBDmap determines the potential RBDs through the identification of the peptides which are adjacent to the RNA-crosslinked sites by UVX. The peptide-level FAX-RIC and the UVX-based RBDmap identified ~200 RBPs exclusively. Incomplete overlap of the protein profile from the two methods

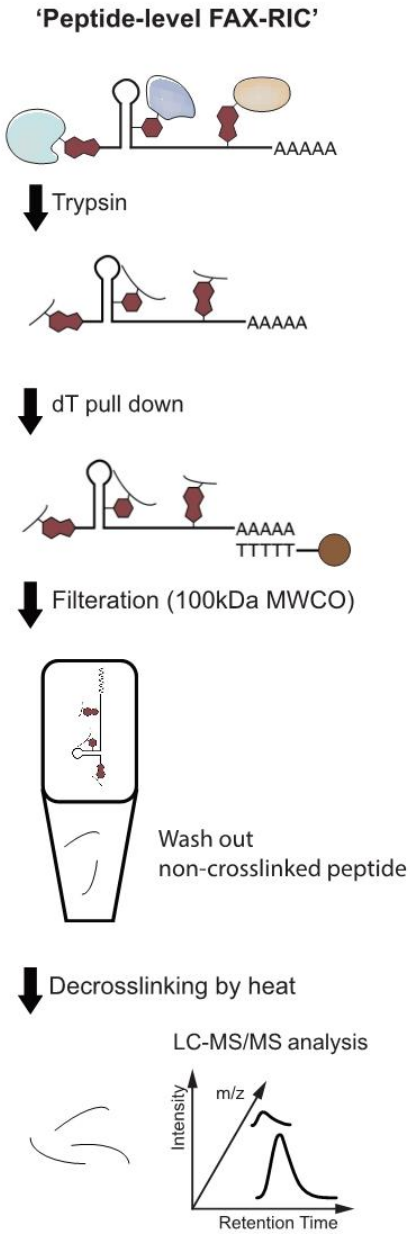


Figure III-10. Experimental scheme for peptide-level FAX-RIC

Abbreviations: kDa, kilo Dalton; MWCO, molecular weight cut off.

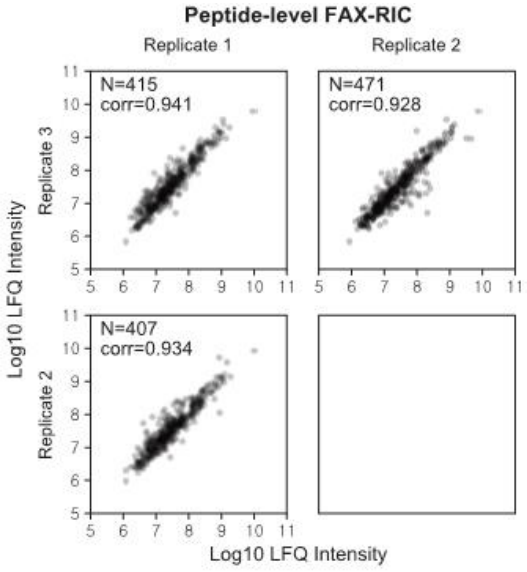


Figure III-11. Reproducibility of peptide-level FAX-RIC

Scatter plot displaying LFQ intensity for peptide identified in two replicates of peptide-level FAX-RIC experiments. Peptide number (N) and correlation coefficient (corr) for the plot are inserted.

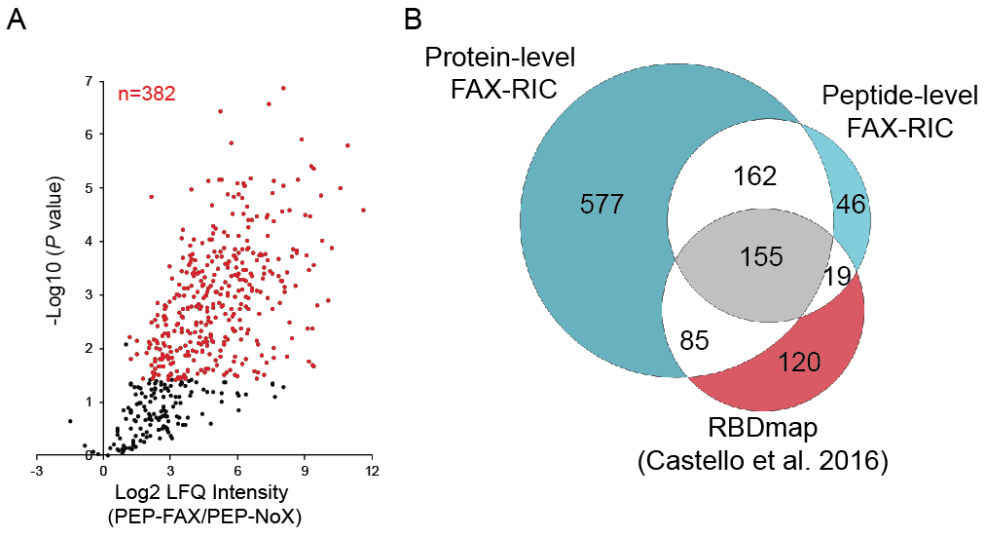


Figure III-12. Peptide-level FAX RNA interactome

(A) Volcano plot displaying the fold-change of average LFQ intensity (peptide-level FAX-RIC/NoX-RIC) (x-axis) and the –log 10 student's t-test P value (y-axis) for all the proteins quantified in at least two out of three replicate peptide-level FAX-RIC experiments. Proteins with log2 fold-change >1 and statistically significant enrichment over peptide-level NoX-RIC (P value < 0.05, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red. (B) Overlap between the protein-level FAX-RIC, peptide-level FAX-RIC and the RBDmap by Castello et al. 2016 (Castello et al., 2016).

suggest that the use of two RBD profiling methods using different crosslinking method may complement each other for the deeper coverage of RNA binding region in RBPs. Compared with protein-level FAX-RIC, peptide-level FAX-RIC showed a higher proportion of peptides mapped within known RBDs or RNA-interacting regions experimentally defined by the RBDmap approach (Castello et al., 2016) (**Figure III-13**). Notably, the peptide-level FAX-RIC approach offers higher resolution than RBDmap (**Figure III-14**). The identified peptides are often found within classical RBDs such as CSD, dsrm, KH, and RRM domains, but they also reveal novel RNA-protein interaction sites within other structural features such as intrinsically disordered regions (**Figure III-14**). These results provide a further basis for the FAX specificity to proximal and stable RNA-protein interactions within the known RBDs, and demonstrate a potential of peptide-level FAX-RIC for the discovery and comparative analysis of the RNA-binding motifs within the RBPs.

III-3. Quantitative comparison of UVX- and FAX-RIC

To further characterize the merits and efficacy of FAX as an alternative *in vivo* crosslinking method in terms of capture specificity and efficiency, I performed a quantitative comparison of the UVX- and FAX-captured proteins based on the LFQ intensity. For fair comparison, UVX RNA interactome (657 proteins)

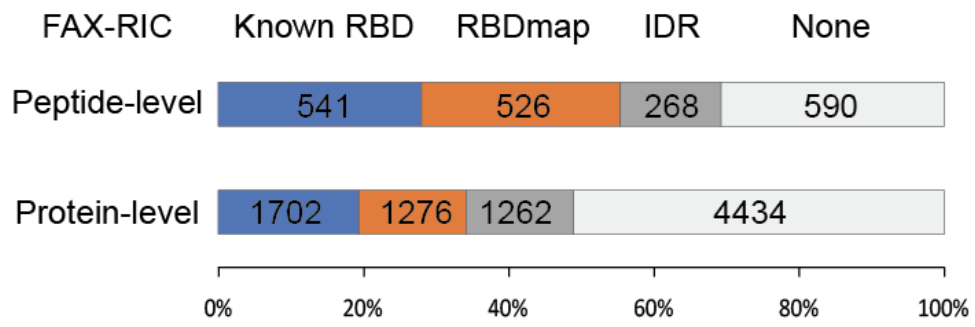
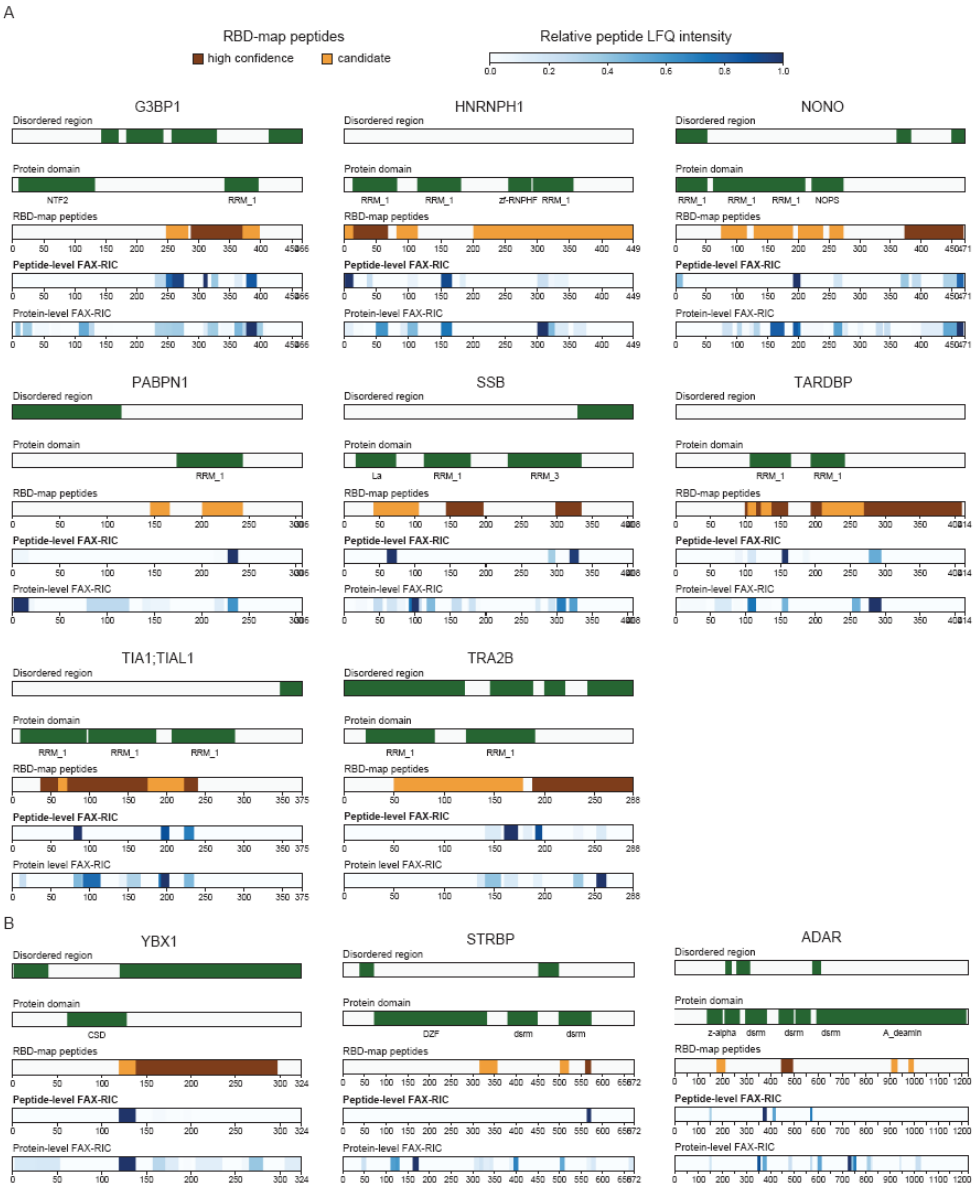


Figure III-13. Proportion of peptides mapped to the known RBDs

Proportion and number of identified peptides in each experiment that are mapped to the known RBD (Gerstberger et al., 2014), RBDmap identified peptides (Castello et al., 2016), and intrinsically disordered regions (IDRs), by MobiDB-lite (Necci et al., 2017).


A
RBD-map peptides
high confidence
candidate
Relative peptide LFQ intensity
0.0 0.2 0.4 0.6 0.8 1.0
G3BP1
HNRNPH1
NONO
PABPN1
SSB
TARDBP
TIA1;TIAL1
TRA2B
Disordered region
Protein domain
RBD-map peptides
Peptide-level FAX-RIC
Protein-level FAX-RIC
NTF2
RRM_1
zf-RNPHF
NOPS
La
RRM_3
B
YBX1
STRBP
ADAR
CSD
DZF
dsrm
z-alpha
A_deamin

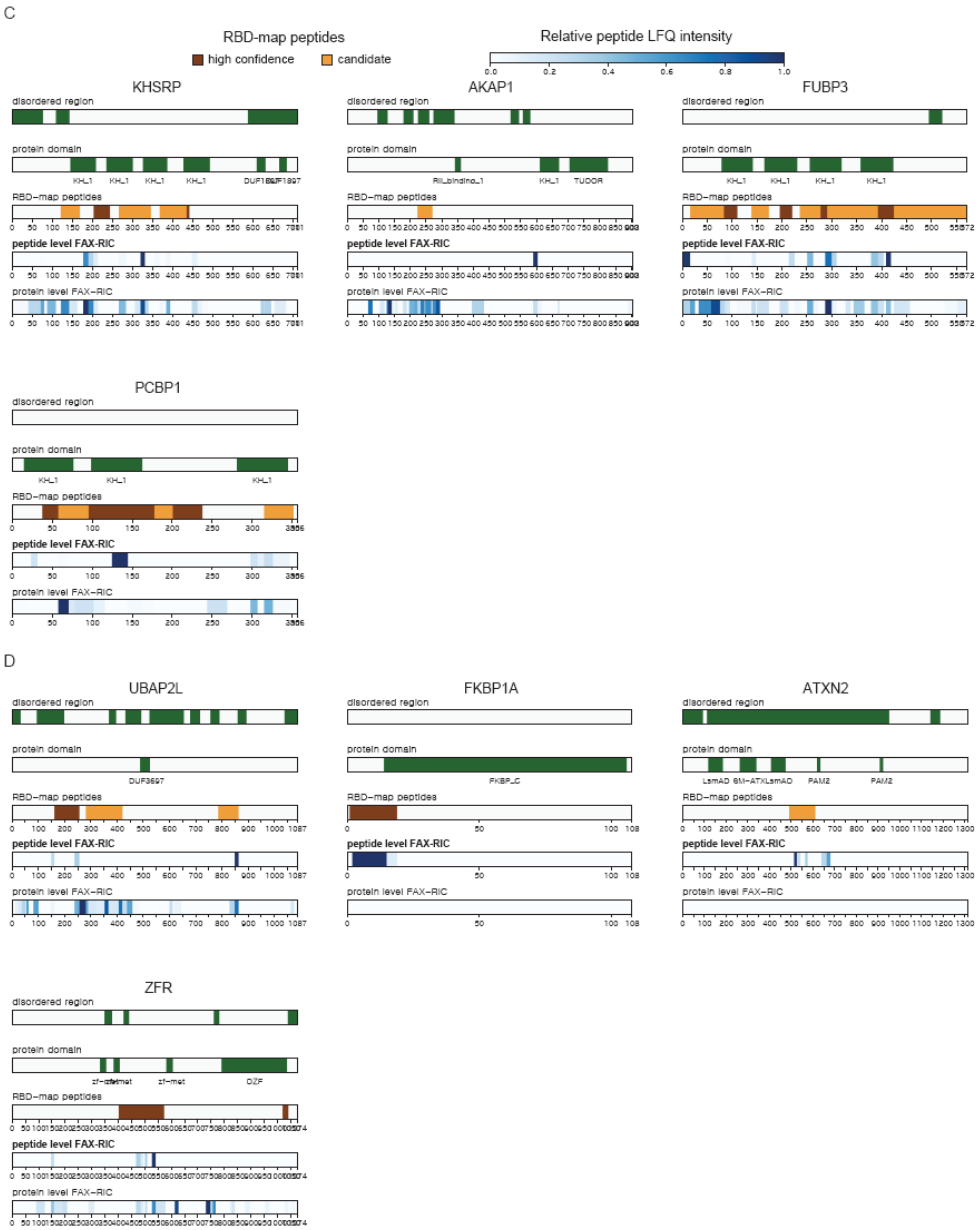

C
RBD-map peptides
high confidence
candidate
Relative peptide LFQ intensity
KHSRP
AKAP1
FUBP3
PCBP1
D
UBAP2L
FKBP1A
ATXN2
ZFR
disordered region
protein domain
RBD−map peptides
peptide level FAX−RIC
protein level FAX−RIC

Figure III-14. Peptide-level FAX-RIC results for the exemplary RBPs

(A-D) Peptide- and protein-level FAX-RIC results for the exemplary RBPs annotated with RRM (A), CSD and dsrm (B), KH (C) and no classical RBDs (D). PFAM annotated protein domain and disordered region (green), and RBDmap identified peptides21 (brown), including the 'candidate' (light brown). Heat map (blue) for the relative peptide intensity normalized by the sum of all identified peptide intensity of that protein in each experiment.

was generated according to the conventional UVX-RIC protocol at an identical sampling scale to that of the FAX-RIC protocol (**Figure III-15**). FAX-RIC enabled more comprehensive profiling, generating ~40% more proteins than UVX-RIC did (**Figure III-16**). Consistent with the previous analysis on the FAX-RIC's specificity to the known RBPs, I found that both UVX- and FAX-RIC had similar quantitative enrichment rate (LFQ intensity from the RIC divided by the relative protein intensity in the total proteome) for the RBPs with known RBDs or those defined via 'RBDmap' (Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019) (**Figure III-17**). No significant change in the relative amount of ribosomal proteins obtained by UVX- and FAX-RIC, suggest that there was no significant increase in the rRNA binding proteome contamination in the FAX-RIC and thus they had comparable specificity to the poly(A)+ RNA interactome (**Figure III-17B**). This quantitative comparison showed that FAX-RIC profiled a larger number of RBPs with specificity comparable to that of UVX-RIC.

Next, I expanded the quantitative comparison to the whole identified RNA interactome, i.e., including the unorthodox RBPs. In the 555 common RNA interactome proteins, 57% of the proteins were more efficiently captured (> 2-fold) by FAX-RIC, whereas only 9.3% of proteins showed better efficiency in UVX-RIC (**Figure III-18A**). Interestingly, I found that a number

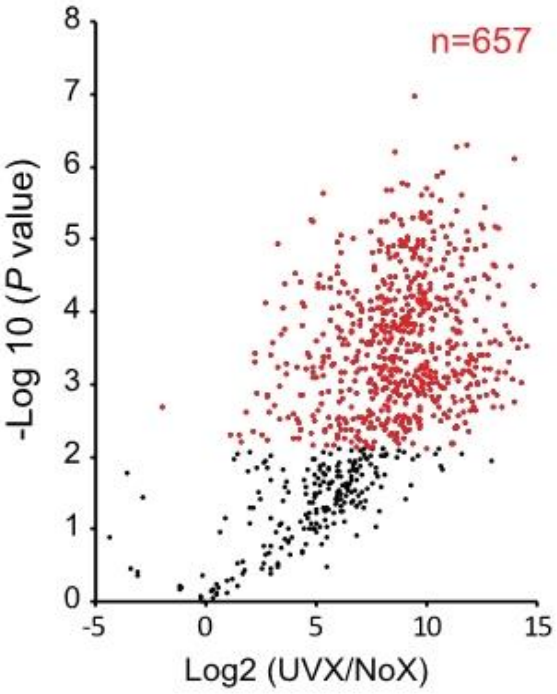


Figure III-15. Defining the high confidence UVX RNA interactome in HeLa cell

Volcano plot displaying the fold-change of average LFQ intensity (UVX-RIC/NoX-RIC) (x-axis) and the –log 10 student's t-test P value (y-axis) for all the proteins quantified in at least two out of three replicate UVX-RIC experiments. Proteins with log2 fold-change >1 and statistically significant enrichment over the NoX-RIC (P value < 0.01, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red.

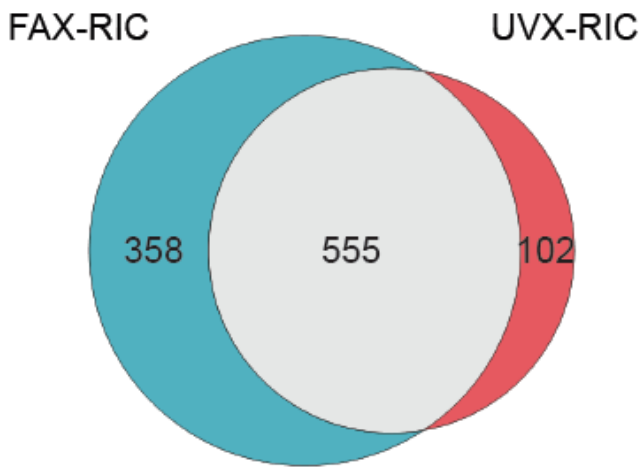


Figure III-16. Overlap between the UVX and FAX RNA interactome

Overlap between the UVX- and FAX-RIC based high confidence RNA interactome profiles.

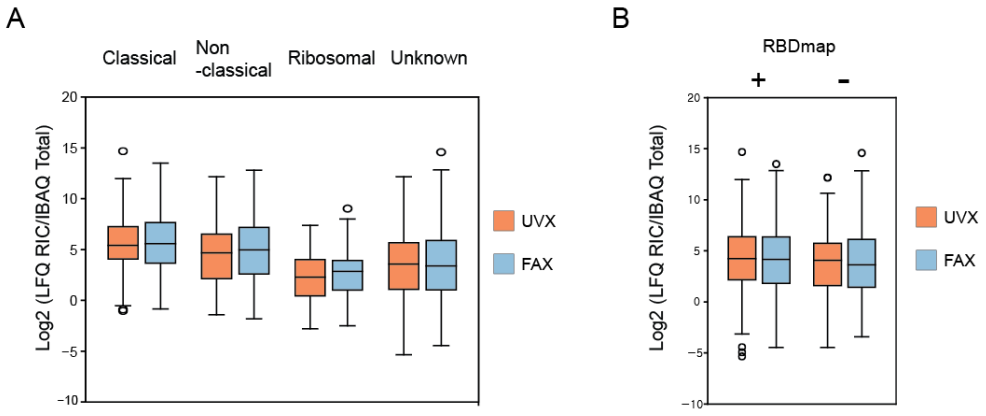


Figure III-17. Relative enrichment rate of the RBPs with different RBDs in UVX and FAX RNA interactome

(A) Boxplots for the relative LFQ intensity level obtained from UVX- and FAX-RIC experiments, normalized by the IBAQ intensity from LC-MS/MS analysis of the total HeLa proteome. RBPs are grouped by the annotation with the indicated class of RBDs3. The median (center line), first and third quartiles (lower and upper box limits, respectively), and 1.5 times the interquartile range (whiskers) are shown in boxplots. (B) Same as (A), but RBPs are grouped by with (+) or without (-) RNA interacting regions identified in previous RBDmap experiments.

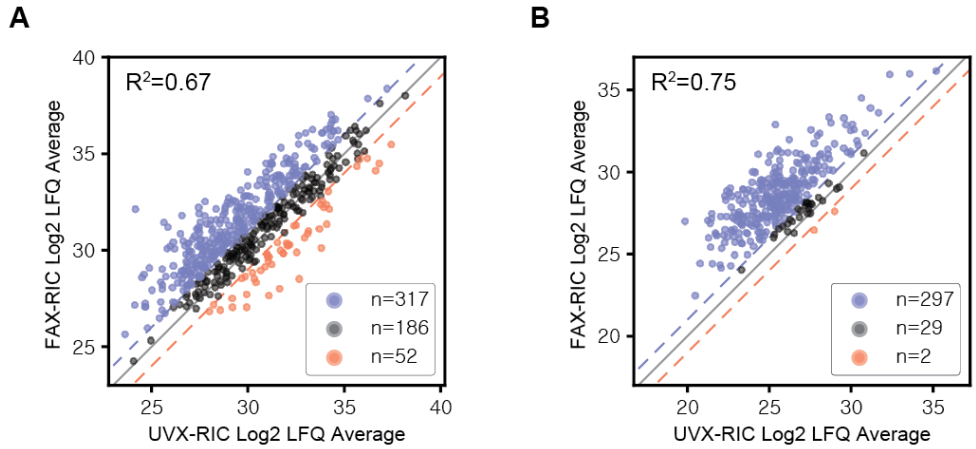


Figure III-18. Quantitative comparison of FAX and UVX RNA interactome

(A-B) Scatter plot of average LFQ intensity between UVX- and FAX-RIC experiments, drawn for RBPs, common to UVX and FAX RNA interactome (A) and exclusive to FAX RNA interactome (B).

of FAX-exclusive RNA interactome were actually identified and quantified in the other UVX-RIC experiment (328 out of 358), even though they failed to pass the quantitative criteria to be defined as UVX RNA interactome. I included these proteins to the quantitative comparison. As expected, over 90% of the FAX-exclusively determined RNA interactome showed a higher intensity value over that obtained by UVX-RIC (**Figure III-18B**), Thus, these RBPs may not have been detected by the UVX-RIC due to the low efficiency of UVX. Of note, I found a strong correlation between protein intensity signals obtained via UVX- and FAX-RIC.

The system-wide comparison between UVX- and FAX-RIC generated a quantitatively UVX- or FAX-preferred RBP list (> 2-fold relative ratio; 82 or 613 proteins for UVX- or FAX-preferred RBPs, respectively). Using this protein list, we further tried to dissect the molecular characteristics of both methods based on the known RBD information of those RBPs, since the different crosslinking mechanism of each method should be preferential to distinct modes of RNA-protein interactions. Notably, I found a strong overrepresentation of the RRM domain in UVX-preferred RBPs (**Figure III-19**) that was consistent with the known preference of UVX on uracil-aromatic amino acid residues (Hockensmith et al., 1986), a major contributor to the affinity

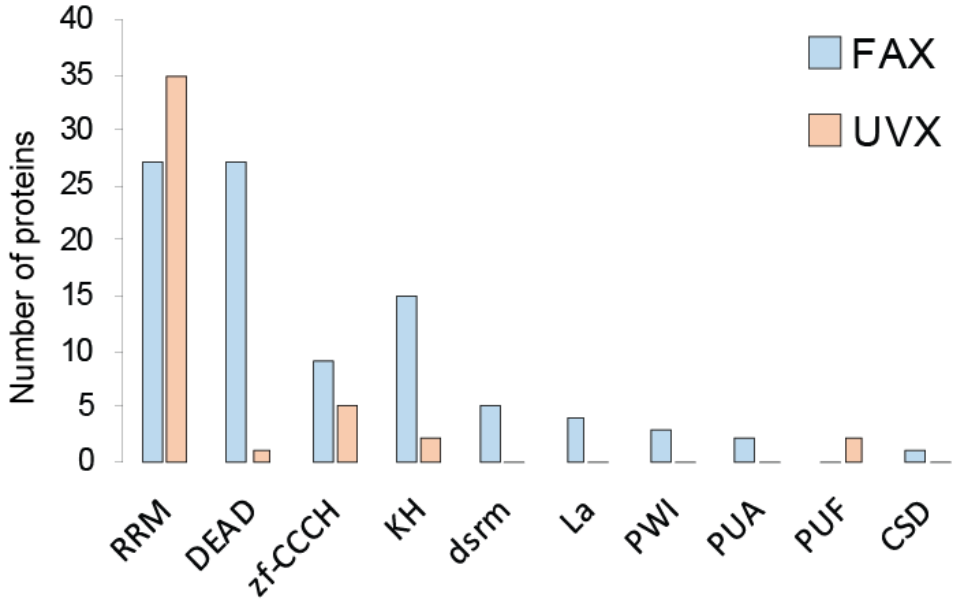


Figure III-19. Number of preferentially enriched RBPs annotated with the known RBDs

Number of proteins with significantly greater LFQ intensity (Log2 >1) in UVX- or FAX-RIC, annotated with indicated classical and non-classical RBDs.

between RRM domains and RNAs (Lunde et al., 2007; Maris et al., 2005). On the contrary, FAX-preferred RBPs RBPs included a broad range of canonical RBPs annotated with the classical RBDs such as RRM, KH, DEAD, La, and PWI domains (**Figure III-19**). Furthermore, I found that the representative RBPs with preferences for distinct RNA sequences or structures were profiled with significantly greater protein intensity by FAX-RIC. These included RBPs associated with RNA duplexes (e.g., AGO1 and STAU1), helicases (e.g., EIF4A3 and UPF1), and uracil-poor RNA sequences such as the mRNA cap (e.g., NCBP3 and EIF4E), N6-methyladenosine (e.g., YTHDF1/2/3), and poly-adenosine (e.g., PABP1/2 and ZC3H14) (**Figure III-20**). It is important to note that the RBPs such as AGO1 and EIF4E were conspicuously missing from many of the previous interactome lists, as well as the list of proteins from UVX-RIC experiment in this study.

I classified the RNA interactome into two groups; conventional RBPs annotated with the known RBDs and unorthodox RBPs, not annotated with the known RBDs. The relative capture efficiencies of FAX over UVX in terms of LFQ intensity were represented via one-dimensional scatter plotting (i.e., beeswarm plot) versus identification frequency among the nine UVX-based RBP profiling studies (Baltz et al., 2012; Castello et al., 2012; Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019) in human cancer cell lines for

conventional (**Figure III-21A**)

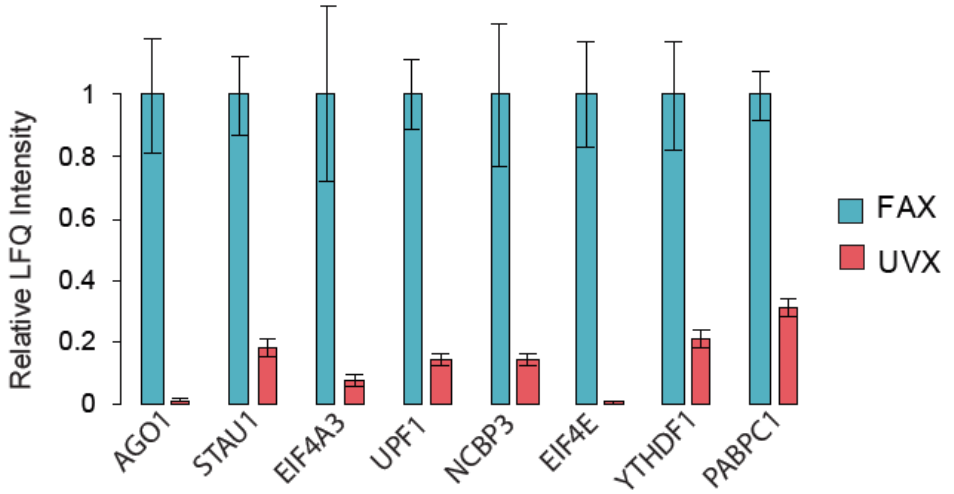
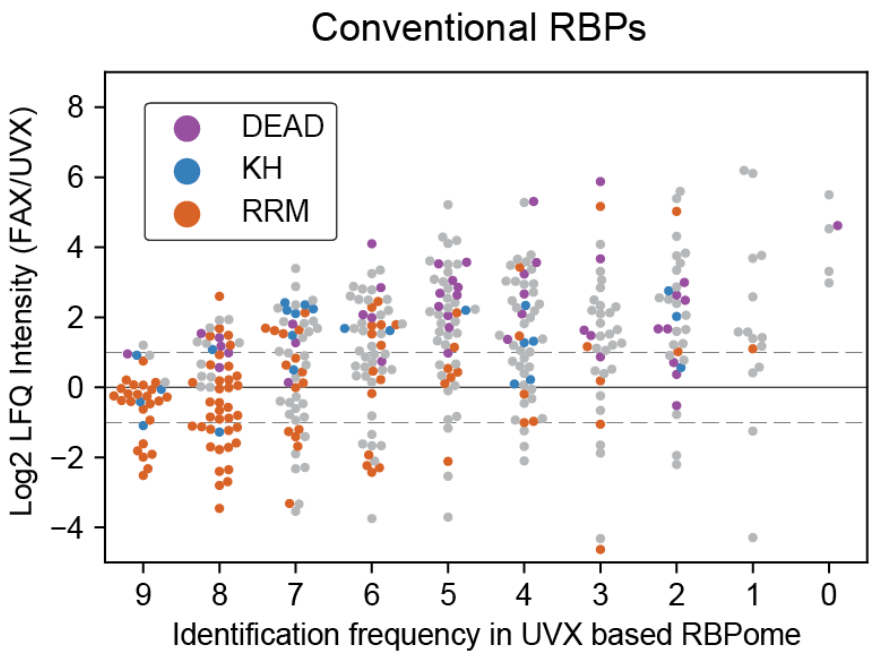


Figure III-20. Protein intensities of the representative RBPs in UVX and FAX-RIC

Relative LFQ intensities from the UVX- and FAX-RIC for the representative RBPs. Error bars represent mean standard error from three independent experiments.

A


Conventional RBPs
DEAD
KH
RRM
Log2 LFQ Intensity (FAX/UVX)
8
6
4
2
0
−2
−4
9
8
7
6
5
4
3
2
1
0
Identification frequency in UVX based RBPome

B

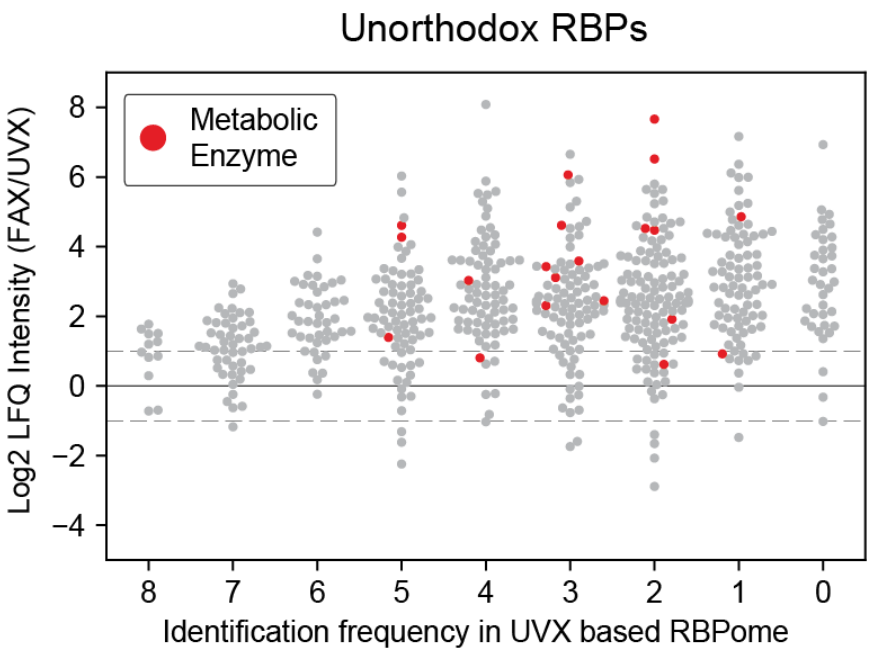

Unorthodox RBPs
Metabolic Enzyme
Log2 LFQ Intensity (FAX/UVX)
8
6
4
2
0
−2
−4
8
7
6
5
4
3
2
1
0
Identification frequency in UVX based RBPome

Figure III-21. Preferential FAX-RIC enrichment and identification frequency in the previous RNA interactome

(A-B) Average LFQ intensity fold-change (FAX/UVX) for the RBPs, annotated with either 'classical' or 'non-classical RBDs' (A) and not annotated with such well characterized RBDs (B). RBPs were grouped by the identification frequency in total of nine UVX based RBP profiling studies (Hentze et al., 2018; Queiroz et al., 2019; Trendel et al., 2019).

and unorthodox RBPs (**Figure III-21B**). The plots showed that less frequently identified RBPs in UVX-RIC studies were more favorably captured by FAX-RIC in both conventional and unorthodox RBPs. Most notably, RBPs with an RRM domain were identified with the most predominant frequency, consistent with the previously determined quantitatively UVX-preferred character of those proteins (**Figure III-19**). In contrast, RBPs with other classical RBDs such as the DEAD and KH domains were identified with relatively low frequency (**Figure III-21A).** Strong overrepresentation of RRM domain in the UVX-based studies highlight the strong bias of UVX method toward a particular RBD (**Figure III-21A**). Presumably owing to the bias of UVX, the significant underrepresentation of the FAX-preferred RBPs with the well-defined RBD annotation such as KH and DEAD in the previous UVX-based RBP profiling studies (**Figure III-21A**) suggests that enhancing crosslinking efficiency using FAX can be a key solution for the robust mapping of such RBPs.

Among the well-known unorthodox RBPs are the metabolic enzymes (Castello et al., 2015; Hentze et al., 2018). Interestingly, I found that nearly all metabolic enzymes identified in UVX-RIC (Hentze et al., 2018; Liao et al., 2016), including SHMT2, GAPDH, and ENO1, showed strongly enhanced capture efficiency in FAX-RIC (**Figure III-21B**). A number of these

unorthodox RBPs were found to bind RNA through their nicotinamide adenine dinucleotide binding domains (Castello et al., 2015; Liao et al., 2016), indicating that those metabolic enzymes could retain affinity to RNA adenine nucleotides. These results thus suggested that these RBPs may have been significantly underrepresented in previous UVX-RIC studies owing to the strong nucleobase bias of UVX. I further validated the RNA-crosslinking dependent enrichment of the representative FAX-RIC exclusive RBPs, DDX19B, ENO1, and, GAPDH, through the WB analysis (**Figure III-22**). Near exclusive detection of the protein signal for all three proteins in the FAX-RIC is consistent with the quantitative proteomics analysis in this study and their absence in all of the 7 representative UVX based RNA interactome profiles of the human cell lines (Hentze et al., 2018). In summary, system-wide quantitative analysis between FAX- and UVX-RIC revealed that FAX-RIC can enable robust RNA interactome mapping with broader and more balanced coverage of RBPs that are both conventional and unorthodox.

I extended the comparative analysis to a more comprehensive reference of UVX RNA interactome in HeLa cells (Castello et al., 2012) (referred to as REF-RIC), which was determined with greater input cell amount (~10 times) and more in-depth profiling manner for LC-MS/MS analysis. Despite the significant differences in the experimental settings, the

correlation between

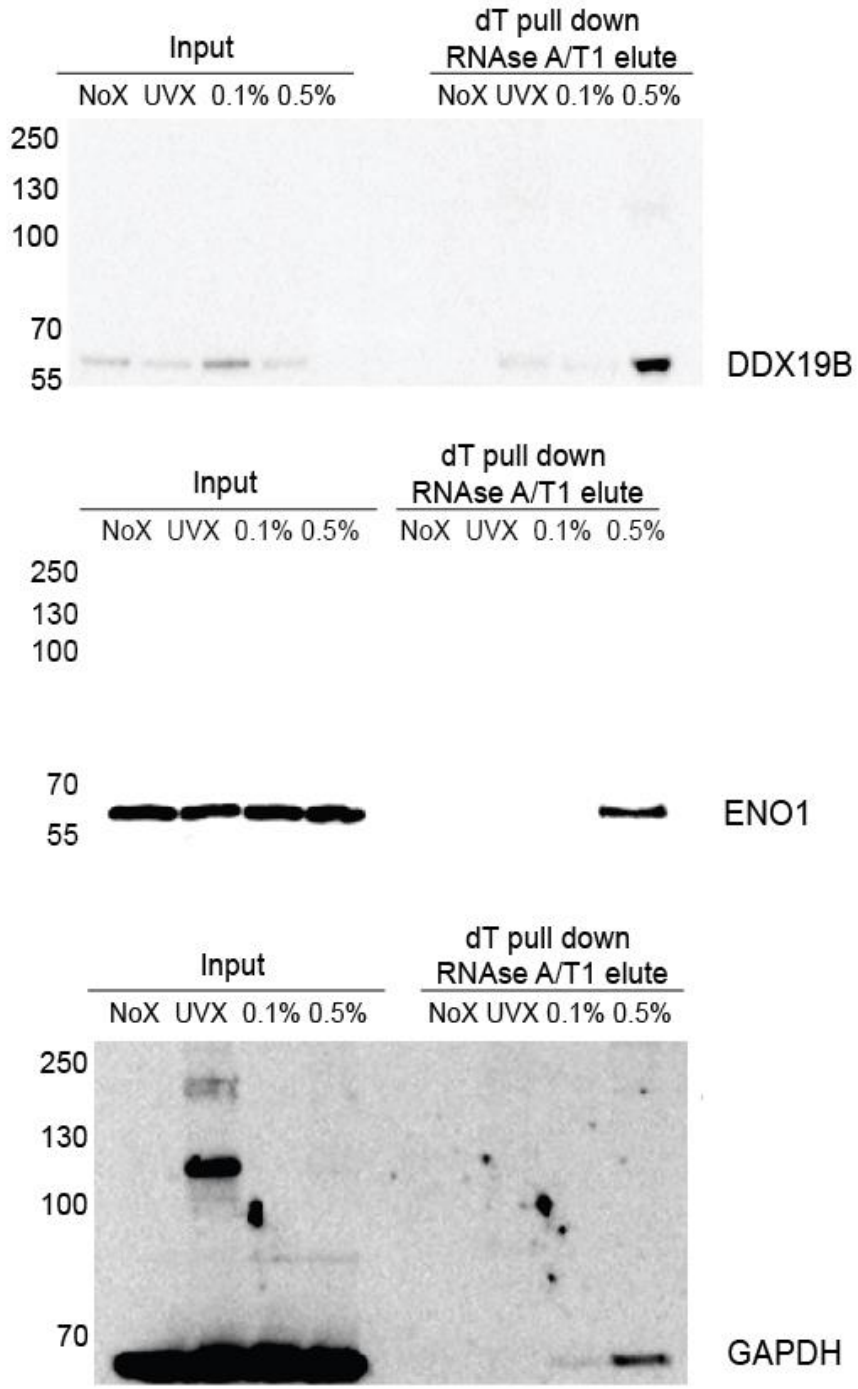
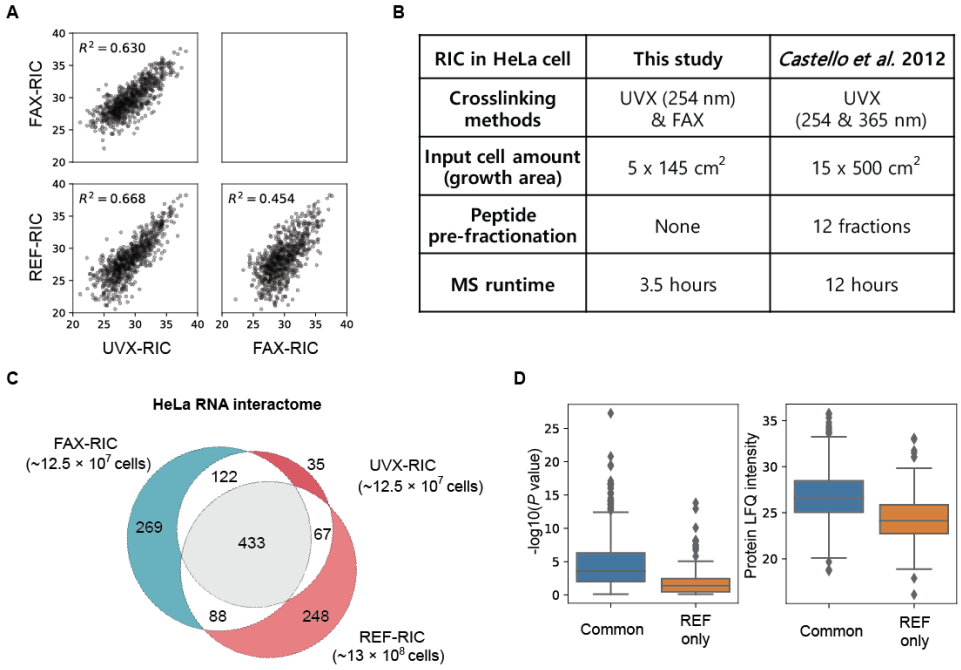


Figure III-22. RNA-protein interaction specificity to the unorthodox RBPs

Western blot analysis of DDX19B, ENO1, and GAPDH for the input and oligo-dT enriched and RNAse A/T1 eluted protein samples from HeLa cells treated with indicated crosslinking conditions. Both input and enriched proteins were subjected to SDS-PAGE without the heat treatment to prevent the reversal of formaldehyde crosslinking.

the protein LFQ intensities from the REF- and UVX-RIC experiments were high, suggesting that UVX-RIC largely reproduced the relative protein intensity value obtained by the REF-RIC (**Figure III-23A-B**). I also found that majority of the REF-RIC exclusive RBPs had higher adjusted P-value from the DEseq analysis, and the relatively low average LFQ values, compared to those of the commonly identified RBPs, suggesting that UVX-RIC reliably identified most of the high-confidence RBPs reported by the REF-RIC study (**Figure III-23C**). The size of REF and FAX RNA interactome were comparable, and yet there was a significant difference in the relative protein intensity of the individual RBPs (**Figure III-24A**). Enhanced profiling of the metabolic enzymes by FAX-RIC is also consistent with previous analysis and result (**Figure III-24A**). Furthermore, overrepresentation of the distinct classical RBDs in the set of REF- (RRM) or FAX-RIC preferred RBPs (DEAD, KH and dsrm) also closely recapitulate previous result on the same analysis (**Figure III-24B**). Taken together, the comparative analyses using previously reported UVX-RIC data further validated the similarities and differences between the UVX- and FAX-RIC approaches.



RIC in HeLa cell	This study	*Castello et al.* 2012
Crosslinking methods	UVX (254 nm) & FAX	UVX (254 & 365 nm)
Input cell amount (growth area)	5 x 145 cm^2	15 x 500 cm^2
Peptide pre-fractionation	None	12 fractions
MS runtime	3.5 hours	12 hours

Figure III-23. Comparison with the reference UVX RNA intearctome capture experiment

(A) Table outlining the significant differences between this study and the REF-RIC, Castello et al. 2012 (Castello et al., 2012) (B) Correlation of the average LFQ intensity obtained from each experiment. (C) Protein profile comparison of the three HeLa RNA interactome. (D) Confidence and average LFQ intensity of the RBPs identified in REF-RIC, which were either common to this study or exclusive to the REF-RIC.

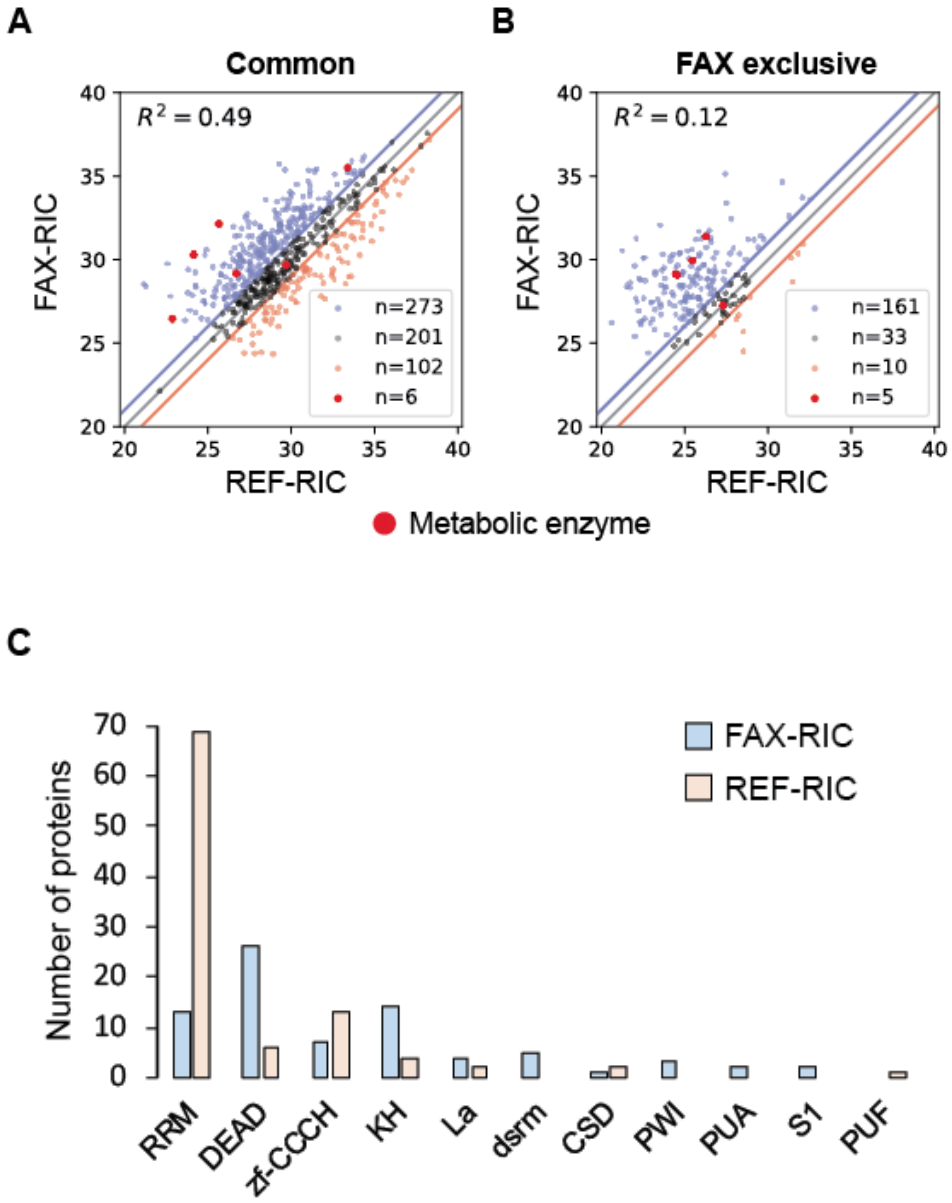

A
Common
$R^2 = 0.49$
FAX-RIC
REF-RIC
n=273
n=201
n=102
n=6
B
FAX exclusive
$R^2 = 0.12$
FAX-RIC
REF-RIC
n=161
n=33
n=10
n=5
Metabolic enzyme
C
Number of proteins
FAX-RIC
REF-RIC
RRM
DEAD
zf-CCCH
KH
La
dsrm
CSD
PWI
PUA
S1
PUF

Figure III-24. FAX-RIC and UVX-RIC (*Castello et al.* 2012)

(A) Average LFQ intensity for common RBPs in FAX-RIC and REF-RIC and (B) the same for FAX-RIC exclusive RBPs. R squared value for correlation are inserted. (C) Number of RBPs annotated with indicated canonical RBDs in FAX or *Castello et al.* preferred RBPs.

III-4. FAX-RIC enables comprehensive and unbiased RNA interactome profiling in multicellular organisms *in vivo* (*X. laevis* oocytes and embryos)

The most evident limitation of UVX is its highly limited penetration depth into opaque biological systems (Elinson and Pasceri, 1989; Sysoev et al., 2016). Due to this innate limitation of UVX, to my knowledge, no previous study has shown whether one can achieve sensitive and comprehensive profiling of the RNA interactome in multicellular organisms. Owing to the good membrane permeability of formaldehyde (Thavarajah et al., 2012), FAX-RIC readily overcomes this limitation of conventional UVX-RIC, enabling robust RNA interactome profiling in multicellular tissues and organisms. Thus, I performed a comparative analysis of UVX- and FAX-RIC in *X. laevis* oocytes (stage VI) and embryos (stage 8-9), which are large (1-2 mm in diameter), opaque, and partially pigmented. I found that at a newly optimized FAX condition for *X. laevis* samples (2% formaldehyde (w/v) for 10 min), FAX-RIC captured a significant amount of proteins whose profile was distinct from that of both the input and NoX control (**Figure III-25**). On the contrary, the protein staining pattern of UVX-RIC experiments done with relatively high UV-irradiation energy (2 J) was indistinguishable from that of the NoX control samples (**Figure III-25**).

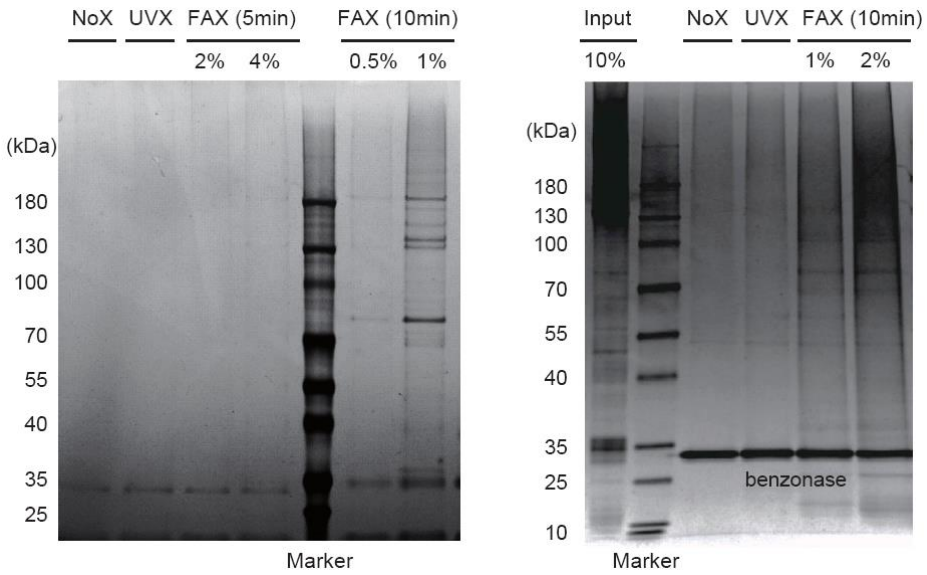


Figure III-25. Optimization of FAX-RIC in *X. laevis* embryo

Silver staining of proteins enriched via oligo-dT pulldown from the lysate of *X. laevis* embryos treated with the indicated crosslinking conditions.

FAX-RIC and UVX-RIC experiments were carried out in triplicate using 50 oocytes or embryos per replicate along with the NoX control. The LC-MS/MS results of FAX-RIC experiments were highly reproducible between the replicate experiments in both oocytes and embryos (**Figure III-26**). FAX-RIC then identified 693 and 541 RNA interactome of *X. laevis* oocytes and embryos, respectively, at < 0.01 FDR (**Figure III-27**). I found that collectively ~80% of these FAX RNA interactome with obvious human orthologues in oocytes or embryos were reported as RBPs with relevant records such as 'RNA-binding' Gene Ontology (GO) terms or known RBDs (Gerstberger et al., 2014) (**Figure III-28**). These results clearly demonstrated that the aforementioned advantage and specificity of FAX-RIC in HeLa cells could be readily reproduced even in large and opaque samples, enabling the comprehensive profiling of *in vivo* RNA interactome landscapes.

In contrast, UVX-RIC identified no more than 94 and 85 proteins as RNA interactome at the same oocyte and embryo scales, respectively, most of which were also covered by the FAX-RIC results (**Figure III-29A and B**). I found that the RBPs localized to nucleus in *X. laevis* oocytes, which were defined by Wuhr *et al.* (Wuhr et al., 2015), were significantly underrepresented compared to those of embryos in UVX-RIC (**Figure III-29C**, left). This observation illustrates the inability of UVX-RIC to form detectable amounts

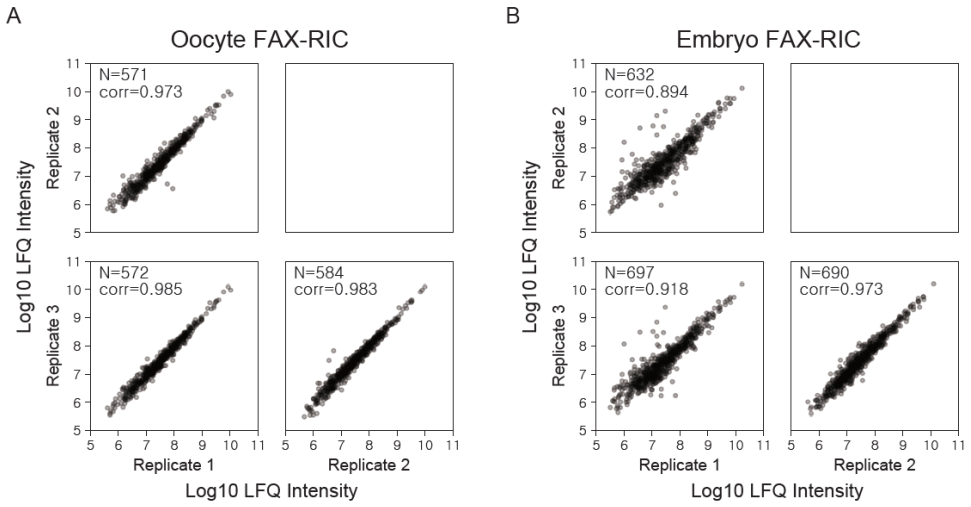
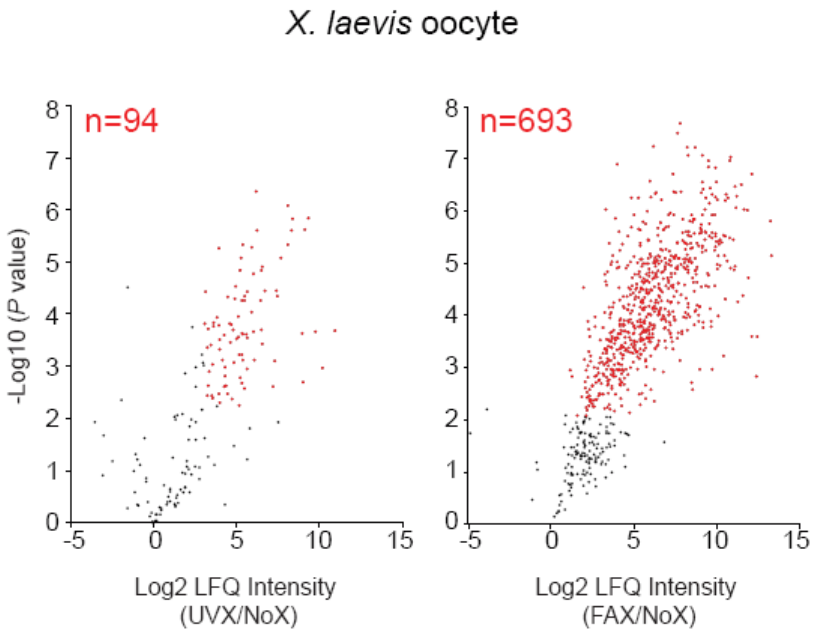


Figure III-26. Reproducibility of FAX-RIC in *X. laevis* oocyte and embryo

Scatter plots of protein LFQ intensities between replicates for oocyte and embryo FAX-RIC, respectively. Protein number (N) and correlation coefficient (corr) for the plot are inserted.

A


X. laevis oocyte
n=94
n=693
-Log10 (P value)
Log2 LFQ Intensity
(UVX/NoX)
Log2 LFQ Intensity
(FAX/NoX)

B

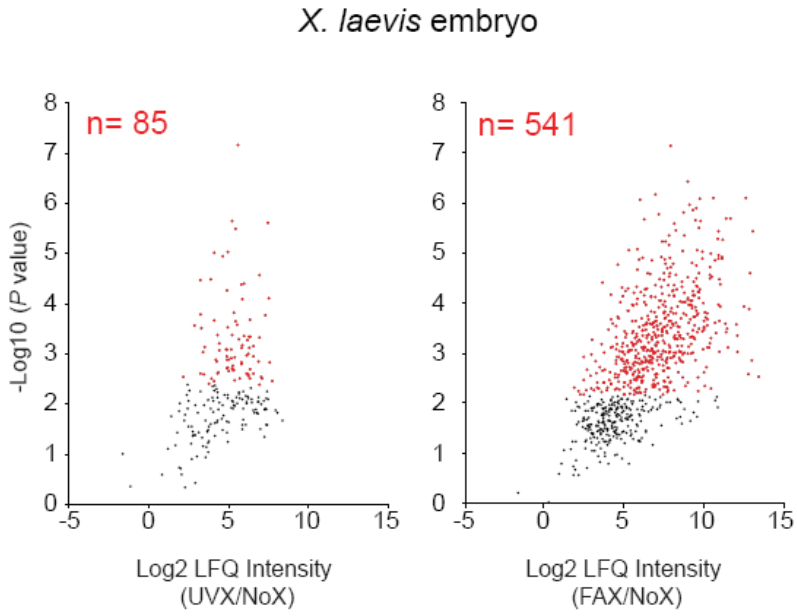

X. laevis embryo
n= 85
n= 541
-Log10 (P value)
Log2 LFQ Intensity
(UVX/NoX)
Log2 LFQ Intensity
(FAX/NoX)

Figure III-27. Defining the high confidence RNA interactome in *X. laevis* oocyte and embryo.

(A-B) Defining the high confidence UVX and FAX RNA interactome in *X. laevis* oocyte (stage VI) (A) and embryo (stage 8-9) (B). Volcano plots displaying the log2 fold-change of average LFQ intensity (x-axis) and the –log10 student's t-test P value (y-axis) for all the proteins quantified in at least two out of three replicate UVX- or FAX-RIC experiments. Proteins with log2 fold-change >1 and statistically significant enrichment over the NoX-RIC experiments (P value < 0.01, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red.

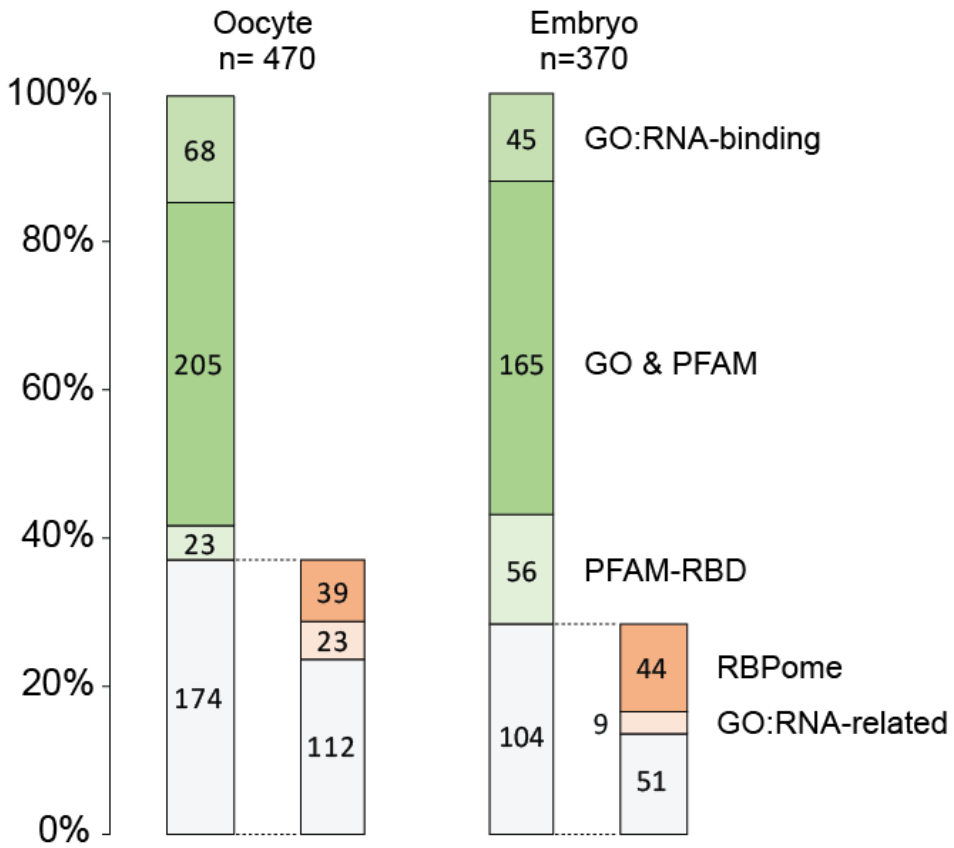


Figure III-28. Known RBPs in FAX *X. laevis* RNA interactome and comparison to UVX RNA interactome

Proportion and number of human orthologous proteins in *X. laveis* RNA interactome that are annotated with GO:'RNA-binding', previously defined (Gerstberger et al., 2014) 'RNA-related GO' and the 'RBD', or identified as RBPs in previous RBP profiling studies (Hentze et al., 2018) RBPome.

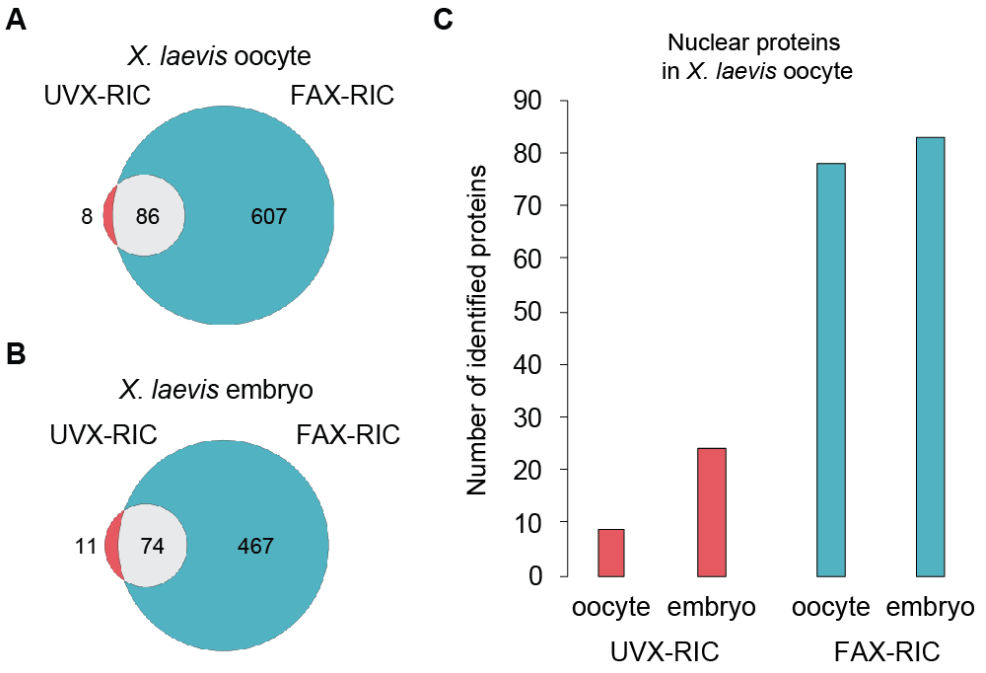


Figure III-29. Comparison of FAX and UVX RNA interactome in *X. laevis*

(A-B) Overlap between UVX-RIC and FAX-RIC RNA interactome in oocyte (A) and embryo (B). (C) Number of the *X. laevis* oocyte nucleus enriched proteins, defined by having > 0.5 protein amount ratio in *X. laevis* oocyte nucleus compared to the cytoplasm (29), identified in two or more replicate UVX- and FAX-RIC experiments in oocyte or embryo.

of crosslinks for the RNP complexes localized to the oocyte nucleus. During the OET, these complexes become more evenly distributed throughout the egg (Radford et al., 2008) and thus become more available to be captured and identified via UVX-RIC in embryo. As a result, comparative analysis of UVX-RIC experiments can result in inaccurate conclusions for some of such stage-specifically localized RBPs, e.g., overestimation of their embryo stage-specific RNA-binding activities. In contrast, FAX-RIC unbiasedly identified a larger number of stage-specifically localized RBPs from both oocytes and embryos (**Figure III-29C**, right). These results illustrated how the use of FAX-RIC was critical to the comprehensive and unbiased profiling of the RNA interactome landscape in physiologically distinct multicellular organisms and tissues *in vivo*.

III-5. FAX-RIC reveals the landscape of mRNP remodeling in *X. laevis* oocyte-to-embryo transition

Quantitative comparison of the RNA interactome profiles from FAX-RIC can characterize RNA interactome landscape transformation in the *X. laevis* OET. There were 295 FAX-captured proteins with significant changes in their LFQ intensity level at < 0.05 FDR and > 1.5 log2 fold-change between oocytes and

embryos (**Figure III-30, Table III-1**). The differences in the two RNA

Table III-1. Proteins with significant change in *X. laevis* oocyte and embryo FAX-RIC

Statistical analysis result for the proteins with significant change in FAX-RIC in *X. laevis* oocyte and embryo

GI number	Uniprot ID (*X. laevis*)	Uniprot ID (*H. sapiens*)	Log2 FAX-RIC (embryo/oocyte)	*P*-value (-log10 scale)
147906019	Q7ZXH6_XENLA	1433T_HUMAN	-1.740	2.428
147899826	Q8AVW5_XENLA	1433Z_HUMAN	3.225	2.628
148225470	Q6DE09_XENLA	4ET_HUMAN	-3.581	3.717
147901618	Q7ZWW5_XENLA	ABCF2_HUMAN	1.670	1.972
147904130	Q6DKB9_XENLA	ACON_HUMAN	-3.184	5.635
148231177	A0A1L8EXC8_XENLA	ACTB_HUMAN	-1.527	3.226
148234583	ADAD1_XENLA	ADAD1_HUMAN	-1.846	3.848
147899117	AGO4_XENLA	AGO4_HUMAN	6.475	2.854
148223025	Q6GQ58_XENLA	ANXA3_HUMAN	-3.352	2.377
148231277	Q90X16_XENLA	ANXA4_HUMAN	-4.352	3.478
147899113	Q5BJ37_XENLA	AP2A2_HUMAN	-2.569	1.829
148231388	Q7ZXU8_XENLA	ASCC2_HUMAN	1.513	1.808
148223147	Q68EY5_XENLA	ATPA_HUMAN	-2.482	4.525
148235415	BIC1B_XENLA	BICC1_HUMAN	6.087	2.220
147900243	Q75T15_XENLA	CAPR1_HUMAN	1.875	2.269
189217806	B1WBD5_XENLA	CAPR2_HUMAN	-5.799	4.243
212286118	A1A5J6_XENLA	CDC5L_HUMAN	3.800	3.684
148235959	A0A1L8FE96_XENLA	CDK1_HUMAN	-2.732	3.066
148224750	CEL1B_XENLA	CELF1_HUMAN	-2.970	3.372
148228597	CEL1A_XENLA	CELF1_HUMAN	-2.846	2.457
27735378	Q7ZTR6_XENLA	CH60_HUMAN	-2.814	2.293
147905967	CKAP5_XENLA	CKAP5_HUMAN	2.050	1.896
148231021	Q6GNR4_XENLA	CLH1_HUMAN	-1.774	2.474
147898757	A0A1L8EVQ1_XENLA	CNOT9_HUMAN	-4.554	3.159
148226821	A0A1L8GCW3_XENLA	COF1_HUMAN	-2.333	2.281
148230798	Q6DJD1_XENLA	COPA_HUMAN	-3.577	2.714
148223736	Q498K6_XENLA	COPB_HUMAN	-1.875	3.892
148230452	Q6DJG5_XENLA	COX42_HUMAN	-4.191	3.886
213623874	CPE1A_XENLA	CPEB1_HUMAN	-5.269	3.392
213625306	CPE1A_XENLA	CPEB1_HUMAN	-8.423	2.726
148233068	A0A1L8GT79_XENLA	CPEB1_HUMAN	-6.897	5.727
147901518	CPSF2_XENLA	CPSF2_HUMAN	-5.789	3.149
147905468	Q5XGZ1_XENLA	CPSF3_HUMAN	-4.571	4.135
51258747	Q68F10_XENLA	CSTF2_HUMAN	-5.891	3.989
350538533	G1FF47_XENLA	DBX2_HUMAN	-4.098	2.056
148236875	Q6IRC3_XENLA	DCTN6_HUMAN	1.796	1.703
148228442	Q66JB0_XENLA	DDX17_HUMAN	5.652	2.217
148236143	B1WBD2_XENLA	DDX25_HUMAN	-3.585	3.606
147902002	B7ZQ46_XENLA	DDX4_HUMAN	-5.304	2.209
148230348	Q801P9_XENLA	DDX5_HUMAN	1.950	3.449
148222264	DENR_XENLA	DENR_HUMAN	1.781	2.931
60389570	DND1_XENLA	DND1_HUMAN	1.652	2.725
148231835	Q7ZX48_XENLA	DX39A_HUMAN	4.297	4.061
148228466	Q641G9_XENLA	DX39B_HUMAN	3.500	2.779
148230603	Q6GPR8_XENLA	DYL1_HUMAN	-3.356	2.222
148238090	Q6GPS9_XENLA	ECHA_HUMAN	-3.865	2.533
148226545	Q7ZTL0_XENLA	ECHB_HUMAN	-2.973	4.125
148228671	A0A1L8F5S5_XENLA	EDF1_HUMAN	2.832	3.256
148223782	A0A1L8FTL8_XENLA	EIF3E_HUMAN	3.448	2.279
148226152	A0A1L8H2S1_XENLA	EIF3G_HUMAN	2.281	3.811
148236035	A0A1L8GNT3_XENLA	EIF3L_HUMAN	2.107	3.528
148234853	Q5U577_XENLA	EMAL1_HUMAN	-2.443	3.027
148223399	ESRP1_XENLA	ESRP1_HUMAN	-5.085	3.004
148223293	Q66KU0_XENLA	ESTD_HUMAN	-2.889	2.231
351542215	Q6PB18_XENLA	EXD2_HUMAN	-3.462	4.697
189217744	Q08AX0_XENLA	EXD2_HUMAN	-2.354	3.121
148237404	Q6AXA1_XENLA	FA98A_HUMAN	1.828	4.646
49118982	Q6GNA5_XENLA	FBRL_HUMAN	3.648	3.330
148230426	A0A1L8GJ67_XENLA	FEN1_HUMAN	3.957	2.591
205277362	Q3B8M1_XENLA	FKBP3_HUMAN	1.533	2.297

GI number	Uniprot ID (*X. laevis*)	Uniprot ID (*H. sapiens*)	Log2 FAX-RIC (embryo/oocyte)	*P*-value (-log10 scale)
148228350	Q6NRX8_XENLA	FOXE1_HUMAN	-1.973	2.061
147906092	Q7ZXS1_XENLA	FUBP1_HUMAN	2.364	2.833
147898765	Q6IRB9_XENLA	FUBP1_HUMAN	1.929	3.114
171847306	B1WBB7_XENLA	FUBP2_HUMAN	5.439	3.202
147902659	Q8AX85_XENLA	FUBP2_HUMAN	4.380	1.929
147899778	Q6PAA0_XENLA	FUBP3_HUMAN	2.115	2.179
147902226	Q7ZXQ2_XENLA	FUS_HUMAN	1.625	3.667
147899332	Q6DFJ1_XENLA	GLGB_HUMAN	-4.222	2.883
147900646	Q7ZX34_XENLA	GRP75_HUMAN	-2.325	4.065
147902718	GTPB6_XENLA	GTPB6_HUMAN	2.005	2.394
147901873	Q6NRY1_XENLA	HABP4_HUMAN	1.905	2.809
147903585	Q6PB22_XENLA	HABP4_HUMAN	1.603	2.545
148235933	HIBCH_XENLA	HIBCH_HUMAN	-2.938	2.060
148223673	Q91764_XENLA	HMGB2_HUMAN	3.633	3.899
148229870	Q7ZY24_XENLA	HMGB3_HUMAN	4.510	2.391
148235301	HNDLB_XENLA	HNRDL_HUMAN	-3.747	1.899
148237217	Q6IRC9_XENLA	HNRH1_HUMAN	1.652	1.954
147906713	Q6DD57_XENLA	HNRH1_HUMAN	2.239	4.419
148233462	Q640J5_XENLA	HNRH3_HUMAN	2.633	2.547
147900289	Q6NS22_XENLA	HNRPQ_HUMAN	1.933	4.060
54311369	Q5U5E2_XENLA	HNRPU_HUMAN	2.720	2.054
28302346	Q7ZTJ7_XENLA	HNRPU_HUMAN	4.193	3.490
1150850	Q91662_XENLA	HS90A_HUMAN	-3.842	2.187
147902812	IBTK_XENLA	IBTK_HUMAN	2.580	1.852
148230953	Q6NU16_XENLA	IF1AX_HUMAN	2.561	2.334
148539602	Q6Q4H9_XENLA	IF2B_HUMAN	2.179	2.553
148229563	IF23B_XENLA	IF2B3_HUMAN	2.231	3.551
148226749	A0A1L8FW10_XENLA	IF2B3_HUMAN	1.948	3.686
54311504	I4A3A_XENLA	IF4A3_HUMAN	2.281	3.418
148234108	Q5HZ86_XENLA	IF4E_HUMAN	-2.315	3.470
148222442	I4E3A_XENLA	IF4E3_HUMAN	1.866	2.251
161169050	Q5KTT8_XENLA	IF4G2_HUMAN	1.977	3.091
148238002	Q6AZF3_XENLA	IF4H_HUMAN	-2.943	2.278
148227544	Q7ZXG3_XENLA	IF5A1_HUMAN	-2.707	2.194
147906156	Q7ZY44_XENLA	ILF2_HUMAN	2.850	3.294
290463410	ILF3B_XENLA	ILF3_HUMAN	2.607	2.969
10834850	ILF3A_XENLA	ILF3_HUMAN	7.061	3.386
147907116	Q4QR45_XENLA	IMA8_HUMAN	2.180	2.890
155369253	Q2NL56_XENLA	IPO5_HUMAN	-1.570	1.750
148234658	K118B_XENLA	K1C18_HUMAN	-3.226	2.022
148238295	Q05AX6_XENLA	K1C19_HUMAN	-3.017	1.667
147901918	Q8AVH2_XENLA	KCRB_HUMAN	-1.625	2.071
147904322	KIF2C_XENLA	KIF2C_HUMAN	2.699	2.269
148228012	Q32N81_XENLA	LA_HUMAN	5.374	2.255
147900941	Q566K4_XENLA	LARP1_HUMAN	3.609	3.282
148224363	A1L3M5_XENLA	LARP4_HUMAN	-1.638	2.036
168693499	A9ULY1_XENLA	LARP6_HUMAN	-4.395	4.413
148225879	Q7ZSY1_XENLA	LEG3_HUMAN	-2.497	3.136
148237916	Q52L12_XENLA	LPP_HUMAN	-2.634	2.827
148223501	L14AB_XENLA	LS14A_HUMAN	-4.641	3.637
148226583	L14AA_XENLA	LS14A_HUMAN	-2.204	2.427
147900474	L14BB_XENLA	LS14B_HUMAN	-3.892	2.590
148222186	A0A1L8ESZ1_XENLA	LS14B_HUMAN	-5.411	2.252
148886747	A5D8N1_XENLA	LSM1_HUMAN	-4.957	2.273
147903851	A3KMU6_XENLA	LSM3_HUMAN	-3.000	3.033
148228551	Q6GN59_XENLA	MAMD1_HUMAN	-2.984	2.530
148230813	Q9YI90_XENLA	MAP4_HUMAN	1.674	3.123
148223495	Q8UW78_XENLA	MARE1_HUMAN	-3.202	1.762
327412335	Q6GP10_XENLA	MATR3_HUMAN	6.210	4.486
147899408	MIF_XENLA	MIS_HUMAN	-3.177	3.221

GI number	Uniprot ID (*X. laevis*)	Uniprot ID (*H. sapiens*)	Log2 FAX-RIC (embryo/oocyte)	*P*-value (-log10 scale)
148224889	Q2VPL8_XENLA	MMSA_HUMAN	-3.243	2.268
148228873	MTHSD_XENLA	MTHSD_HUMAN	-3.645	2.781
602760	Q91793_XENLA	MYL1_HUMAN	-1.794	2.902
148229387	A0A1L8GJR5_XENLA	NAA40_HUMAN	3.689	3.436
147906322	Q7ZWY8_XENLA	NH2L1_HUMAN	3.391	2.524
147898427	NSUN2_XENLA	NSUN2_HUMAN	6.119	3.458
64937	NUCL_XENLA	NUCL_HUMAN	3.070	2.459
148232824	PABPA_XENLA	PABP1_HUMAN	2.694	4.222
147899475	A0A1L8FZR3_XENLA	PABP1_HUMAN	2.891	2.390
148232768	PATL2_XENLA	PATL2_HUMAN	-8.751	4.476
148237123	Q6GPT4_XENLA	PDIP3_HUMAN	3.937	3.897
147902093	Q566H6_XENLA	PDIP3_HUMAN	2.454	3.439
148227308	Q52KT0_XENLA	PEPL1_HUMAN	-2.276	3.177
148229663	Q63ZK6_XENLA	PFD5_HUMAN	-1.834	2.653
189217597	Q08B76_XENLA	PLAP_HUMAN	-3.340	1.671
148237548	Q800A4_XENLA	PRI1_HUMAN	-3.030	5.202
147900604	Q7ZXW4_XENLA	PRP19_HUMAN	2.161	1.810
147904940	A2VD02_XENLA	PRP4B_HUMAN	1.542	2.131
62739333	PTCD3_XENLA	PTCD3_HUMAN	1.892	2.437
350538385	PUM2_XENLA	PUM2_HUMAN	1.925	2.225
148227760	Q6AZK2_XENLA	PURA_HUMAN	1.816	2.879
147899952	A0A1L8GRZ5_XENLA	PURB_HUMAN	2.705	3.006
148233332	A0A1L8HI53_XENLA	PYM1_HUMAN	3.760	3.937
147903173	Q6PAA8_XENLA	RBPMS_HUMAN	-3.699	3.066
148223487	A0A1L8GT66_XENLA	RBPS2_HUMAN	-5.249	2.302
148223321	Q6GNR2_XENLA	RENT1_HUMAN	2.748	4.019
147904963	A0A1L8FGH8_XENLA	RL22_HUMAN	2.293	1.788
49670406	Q6DJQ0_XENLA	RL26_HUMAN	-1.627	3.303
213627669	Q7ZYT3_XENLA	RL3_HUMAN	2.378	2.792
147903020	A0A1L8H8M5_XENLA	RL31_HUMAN	3.095	2.800
148230945	Q66KW5_XENLA	RL32_HUMAN	1.687	2.828
148224674	Q4FZQ7_XENLA	RL35_HUMAN	2.416	1.663
148228014	RL37A_XENLA	RL37A_HUMAN	3.586	1.986
147905658	Q3B8I3_XENLA	RL7_HUMAN	1.607	2.137
147903617	Q5U556_XENLA	RM12_HUMAN	2.121	1.896
214241	ROA1_XENLA	ROA1_HUMAN	2.833	4.351
191256845	B3DLK9_XENLA	RRP5_HUMAN	3.237	4.121
148232006	Q6IP37_XENLA	RRS1_HUMAN	2.320	2.156
148227702	Q7SZU3_XENLA	RS10_HUMAN	-1.861	3.172
148228908	Q7SZ77_XENLA	RS11_HUMAN	1.949	2.609
148237954	Q7SZT6_XENLA	RS13_HUMAN	3.914	1.804
871774	RS13_XENLA	RS13_HUMAN	1.721	2.964
123959702	Q3KQC9_XENLA	RS16_HUMAN	1.704	2.258
147905776	Q6NTT2_XENLA	RS18_HUMAN	2.259	1.827
148223451	Q6PHL6_XENLA	RS19_HUMAN	1.965	3.907
148232341	Q7T0R9_XENLA	RS2_HUMAN	1.738	2.589
148226202	A0A1L8ELM9_XENLA	RS21_HUMAN	-3.150	2.354
147901123	Q6DJJ1_XENLA	RS25_HUMAN	2.090	3.249
148230110	Q6AZL7_XENLA	RS26_HUMAN	2.922	2.239
155369251	A0A1L8HJY0_XENLA	RS3_HUMAN	2.420	4.247
147901291	A0A1L8G4X1_XENLA	RS7_HUMAN	1.540	2.374
148232048	Q641E8_XENLA	RU2A_HUMAN	4.892	2.384
148228438	SAS6_XENLA	SAS6_HUMAN	-3.052	2.454
147905097	Q641I5_XENLA	SC24C_HUMAN	-2.583	2.043
147899706	SDHB_XENLA	SDHB_HUMAN	-1.686	3.180
147905376	A0A1L8EPY7_XENLA	SF3B1_HUMAN	4.694	3.171
148226626	Q7ZX30_XENLA	SF3B4_HUMAN	3.385	1.957
148235953	SLBP2_XENLA	SLBP_HUMAN	-2.586	2.022
63101213	SMC3_XENLA	SMC3_HUMAN	2.496	2.430
147902353	Q6NU78_XENLA	SMN_HUMAN	-2.987	1.827

GI number	Uniprot ID (*X. laevis*)	Uniprot ID (*H. sapiens*)	Log2 FAX-RIC (embryo/oocyte)	*P*-value (-log10 scale)
148222234	Q7ZXI5_XENLA	SNRPA_HUMAN	4.372	2.111
148231125	A2VDA2_XENLA	SORCN_HUMAN	-4.312	1.985
148229557	Q6GNA6_XENLA	SRP72_HUMAN	3.170	2.851
148227107	Q6INS4_XENLA	SRPK1_HUMAN	3.182	1.872
148226156	SSRP1_XENLA	SSRP1_HUMAN	3.667	1.903
147900893	STRBP_XENLA	STRBP_HUMAN	2.072	2.597
147906027	Q63ZJ1_XENLA	SUCA_HUMAN	-2.376	1.844
148233113	SUMO3_XENLA	SUMO3_HUMAN	4.905	1.835
147900889	Q6AX83_XENLA	SYLC_HUMAN	2.548	2.094
147902998	A0A1L8F3F9_XENLA	SYMPK_HUMAN	-2.922	3.863
148231213	Q6DD18_XENLA	SYNC_HUMAN	1.654	2.158
6634718	TACC3_XENLA	TACC3_HUMAN	-2.188	1.825
147905746	TBB4_XENLA	TBB4A_HUMAN	-2.404	3.158
148232786	Q6NTQ5_XENLA	TCPA_HUMAN	1.893	2.002
147901319	TDRD3_XENLA	TDRD3_HUMAN	1.563	2.848
148230292	Q5EAU7_XENLA	TDRKH_HUMAN	-4.259	3.149
148234217	A0A1L8GYW8_XENLA	TES_HUMAN	-2.611	2.953
125858782	A2VD71_XENLA	TF3A_HUMAN	-6.367	3.808
148234082	THILB_XENLA	THIL_HUMAN	-1.790	2.832
147898399	A0A1L8ETL9_XENLA	THOC4_HUMAN	4.928	4.089
148229158	Q9I8J7_XENLA	TKTL2_HUMAN	-1.806	2.140
148235969	Q6PAZ0_XENLA	TOM70_HUMAN	-3.200	3.076
148236601	Q9IAM5_XENLA	TSN_HUMAN	3.532	3.485
148234072	Q7SZ15_XENLA	TSNAX_HUMAN	2.759	4.328
148223762	Q32NM8_XENLA	U2AF1_HUMAN	2.106	2.773
147902896	Q7ZY06_XENLA	U2AF2_HUMAN	1.859	2.125
148235050	Q4KLF0_XENLA	UBIM_HUMAN	1.760	2.714
148229156	Q3B8B3_XENLA	UCHL1_HUMAN	-1.804	2.443
147900383	Q6P704_XENLA	UFD1_HUMAN	-1.870	3.288
147903082	Q5U511_XENLA	VAPB_HUMAN	-2.329	2.967
147900356	Q6NRE3_XENLA	WASL_HUMAN	-2.294	1.761
147903761	Q8AVY9_XENLA	YBOX1_HUMAN	3.278	3.888
27503841	Q8AVK9_XENLA	YBOX1_HUMAN	5.279	3.619
148226902	Q66J89_XENLA	YTHD3_HUMAN	6.175	4.184
147906220	Q6DDZ7_XENLA	Z3H7A_HUMAN	1.526	2.268
148231819	B7ZPG0_XENLA	ZAR1_HUMAN	-7.516	3.771
148232651	A4QNS6_XENLA	ZC11A_HUMAN	4.736	4.370
27503878	ZFR_XENLA	ZFR_HUMAN	-2.592	2.235
148233294	A0A1L8GGA3_XENLA	ZN326_HUMAN	2.529	3.256
227908840	C0SPG1_XENLA		-9.998	4.728
66911762	Q4V7X3_XENLA		4.461	3.626
147906624	Q9I910_XENLA		4.089	1.742
148232254	Q7ZYU4_XENLA		1.745	3.091
350538247	B7ZQZ7_XENLA		3.344	2.517
240849376	Q6GM69_XENLA		1.934	2.884
155369239	Q7SZF6_XENLA		-1.823	3.212
226529318	A0A1L8GLL5_XENLA		-1.631	2.399
148230857	Q7T226_XENLA		-5.289	4.707
976219	Q91855_XENLA		-2.061	3.868
147901035	Q66KP7_XENLA		2.402	3.037
50417514	Q6DDS8_XENLA		4.390	2.619
147903185	Q6AX08_XENLA		5.843	5.082
148231444	Q6IRP2_XENLA		-3.489	2.671
147900426	Q6NRQ9_XENLA		-2.893	2.144
46249488	Q6NUC0_XENLA		-3.464	2.895
148232485	Q6GP41_XENLA		-2.492	2.204
148232692	Q6GQ64_XENLA		1.914	1.924
147905250	Q6PA57_XENLA		-4.114	1.678
76780030	Q3KPP1_XENLA		4.258	2.302
148234565	Q3KQE0_XENLA		2.193	2.980

GI number	Uniprot ID (*X. laevis*)	Uniprot ID (*H. sapiens*)	Log2 FAX-RIC (embryo/oocyte)	*P*-value (-log10 scale)
171847160	B1WBA2_XENLA		-1.883	4.171
122936449	Q505M7_XENLA		-2.168	2.002
83318359	Q2VPE9_XENLA		2.915	1.853
83318221	Q2VPL7_XENLA		-1.506	3.591
147904914	A1L2G5_XENLA		-3.916	4.354
125858778	A2VD56_XENLA		4.524	3.440
67677974	Q4QR27_XENLA		2.048	4.756
147901143	Q0IH65_XENLA		-2.152	2.267
76780301	Q3KQG3_XENLA		1.961	2.566
147904018	Q640C8_XENLA		-3.167	3.077
51258671	Q68F01_XENLA		2.144	4.131
62740166	Q52KU9_XENLA		-2.196	4.043
67678468	Q6GNF9_XENLA		2.204	2.567
49115752	Q6GPN8_XENLA		-2.929	2.737
48734658	Q6ING0_XENLA		-2.000	2.508
47940239	Q6IP25_XENLA		-5.863	3.963
169641978	B1H1P6_XENLA		-3.806	4.531
111185526	Q6PAB6_XENLA		1.589	2.903
38014465	Q6PA30_XENLA		-1.907	4.205
29124415	Q801N0_XENLA		3.666	2.566
169642415	A1L3F1_XENLA		3.726	2.231
49522060	Q6DKN1_XENLA		2.313	4.786
28277337	Q7ZX45_XENLA		2.316	2.064
47122884	Q6NRZ5_XENLA		2.053	2.658
213623691	B7ZR96_XENLA		2.815	3.350
213623661	B7ZR52_XENLA		4.513	3.188
169642030	B1H1X2_XENLA		1.791	2.495
169642427	A9ULX3_XENLA		3.738	2.210
152012986	A7E222_XENLA		1.728	4.316
169641962	B1H1N0_XENLA		-3.761	1.731
147903151	A2RV71_XENLA		-3.115	3.563
110468094	A5XAW2_XENLA		3.141	2.597
132424622	B6RBR2_XENLA		-6.806	2.160
64737	B4_XENLA		6.180	1.858
16974926	NUPL_XENLA		6.723	2.020
4929470	Q9W6R9_XENLA		-1.606	2.086
147905268	Q7ZX82_XENLA		4.723	3.840
134104559	H32_XENLA		5.432	4.185
148233322	Q7ZXF2_XENLA		5.001	3.314
148222886	H2A1_XENLA		4.598	2.028
66911543	Q4V7V5_XENLA		3.449	3.556
3293344	O93335_XENLA		2.949	1.694
32396222	Q7T225_XENLA		-1.761	1.683
64561	VITB2_XENLA		-2.132	2.641
64501	VITA1_XENLA		-2.170	1.672
28175406	Q7ZX71_XENLA		-2.318	3.821
214556	K2C8_XENLA		-4.250	2.142

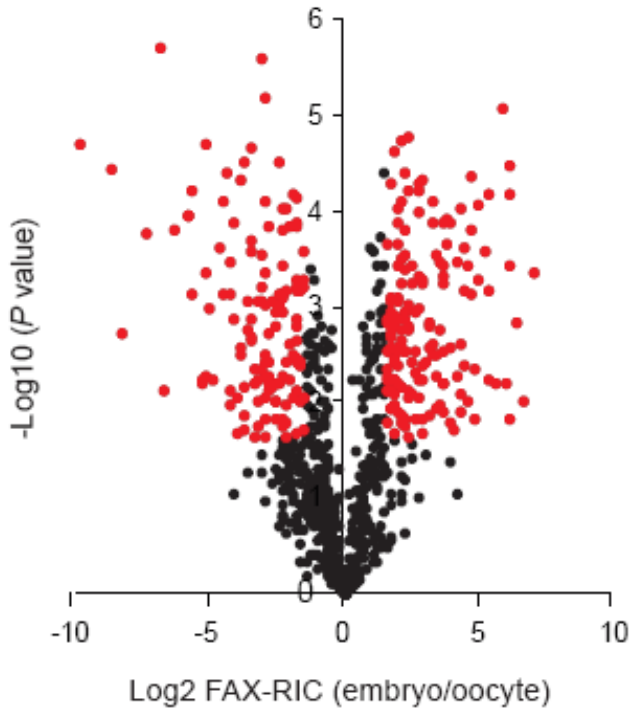


Figure III-30. Transformation of mRNP complex landscape in *X. laevis* oocyte-to-embryo transition (OET).

Volcano plot displaying the log2 fold-change of average LFQ intensity and the –log10 student's t-test P value (y-axis) for all the proteins identified as FAX RNA interactome in *X. laevis*. Proteins with log2 fold-change >1.5 and had statistically significant enrichment in oocyte or embryo FAX-RIC experiments (P value < 0.05, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red.

interactomes may reflect the changes in both the protein expression level and/or those in RNA-binding activity of these RBPs. Thus, I investigated how influential the protein expression changes were to the captured protein changes during the transition (**Figure III-31**). I found that a relatively small number of differentially captured RBPs represented stage-specific protein expression patterns (≥ 1.5 in log2 protein fold-change), while the majority of differentially captured RBPs were largely stable in their expression level, strongly indicating the alterations in their RNA-binding activity (**Figure III-31**).

Among those RBPs with an oocyte-specific expression pattern, Cpeb1, Caprin2, Eif4enif1, Zar1, and Patl1 were highly enriched via FAX-RIC experiments and all of them, with the notable exception of Caprin2, are known to have essential regulatory functions in *X. laevis* OET (Nakamura et al., 2010; Radford et al., 2008; Standart and Minshall, 2008; Wu et al., 2003) (**Figure III-32**). Interestingly, CAPRIN2 and TDRKH were reported as the only two proteins with significant downregulation during human oocyte maturation *in vitro* (Virant-Klun et al., 2016), and both proteins were consistently observed in this study (**Figure III-32**), suggesting that the regulatory mechanism and/or importance of these RBPs in OET may be conserved from *X. laevis* to humans. In contrast, I found relatively few RBPs with an embryo-specific expression pattern (**Figure III-32**), consistent with the previous report that the early

embryonic proteome of

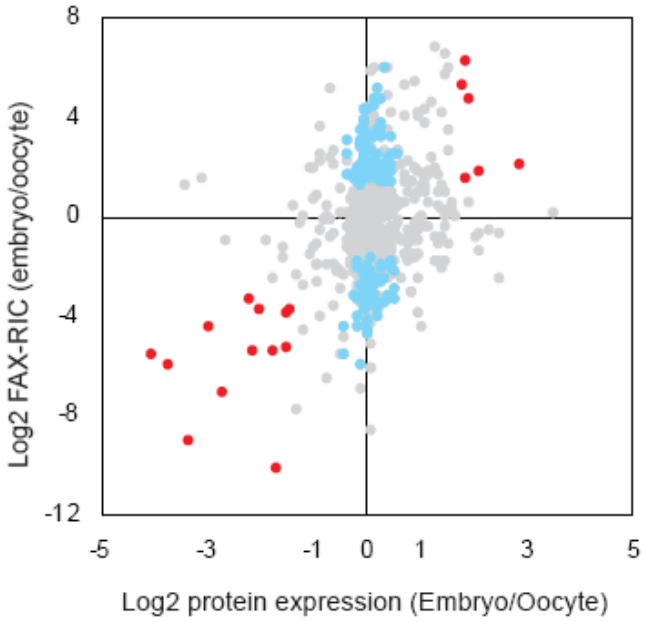


Figure III-31. FAX-RIC enrichment level change in *X. laevis* OET and the respective change in total protein expression level.

Scatter plot displaying the sum of the log2 protein expression level changes in oocyte maturation and early embryo development, previously reported (Jung et al., 2019) (x-axis) and the respective change in average FAX-RIC LFQ intensity level (y-axis). Red dots represent the proteins whose FAX-RIC captured protein amount change can be explained by their respective change in total protein abundance during OET. Light blue dots represent the 'dynamic RBPs', whose FAX-RIC enriched protein amounts are significantly changed

while their respective total protein abundance were changed with log2 fold-change < 0.5.

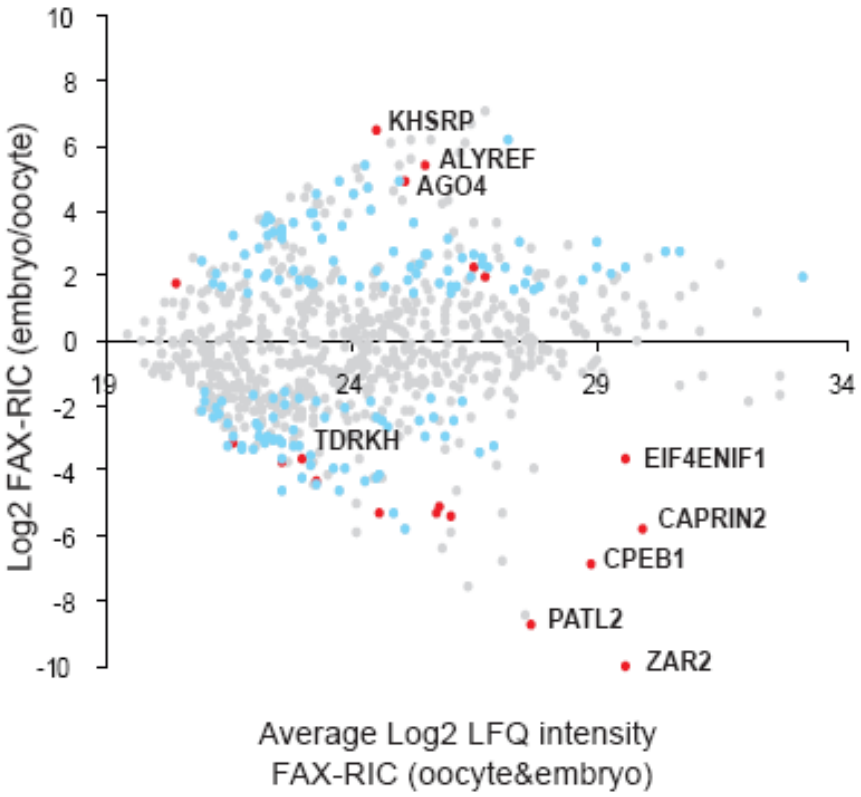


Figure III-32. Relative protein intensity and the fold change in OET

Scatter plot displaying average LFQ intensity from the FAX-RIC experiments in oocyte and embryo FAX-RIC (x-axis) and the change in FAX-RIC enrichment level between oocyte and embryo FAX-RIC (y axis). RBPs are marked with light blue or red, as described in Figure II-28. The protein names are inserted for the targets with most significant change and/or enrichment

level in FAX-RIC, along with the Tdrkh whose human homologue had similar change during the maturation of human oocyte.

X. laevis largely consists of maternally deposited proteins (Peshkin et al., 2015). Nevertheless, among those RBPs, I found AGO4 and KHSRP proteins whose embryo-specific expressions and functions were reported previously (Kroll et al., 2002; Lund et al., 2011) (**Figure III-32**).

As aforementioned, over 141 differentially captured RBPs showed no significant change in their protein expression level during OET (light blue dots in **Figure III-31**) and could be considered dynamic RBPs whose RNA-binding activities were significantly changed during *X. laevis* OET. Translational repression and subsequent activation of the specific target mRNAs (Radford et al., 2008; Richter, 2007) are one of the most distinctive events in mRNA biology of oocyte and embryo. Accordingly, GO term analysis of the dynamic RBPs identified the 'translational initiation' as the most significantly enriched GO term (**Figure III-33A**). It has been known that translational activation of mRNAs during OET occurs through the first dissolution of translation-repressive oocyte-specific mRNP complexes containing Eif4e and the subsequent formation of the canonical Eif4e complex for translation initiation (Radford et al., 2008; Richter, 2007). Most intriguingly, however, I found that the FAX-captured levels of Eif4e were

significantly downregulated in OET while those of other noncanonical translational pathway related RBPs such as Eif4g2, Denr, and Eif4e3 were upregulated (**Figure III-33B**). It is noteworthy that in

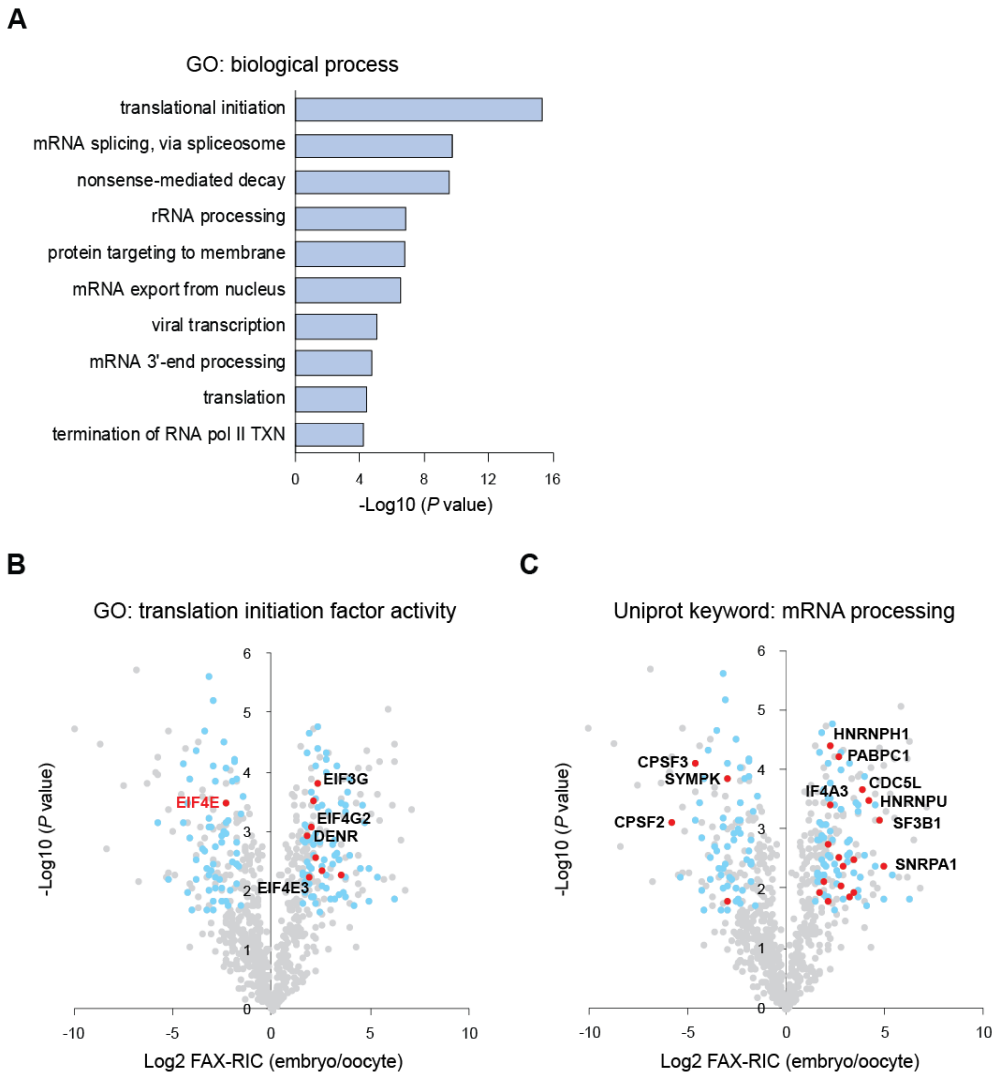


Figure III-33. Representative GO terms and RBPs of the dynamic RBPs

(A) Most significantly enriched biological process GO terms in the ‘dynamic RBPs’. (B) Volcano plot same as Figure II-27, but the ‘dynamic RBPs’ are highlighted with light blue and the dynamic RBPs annotated with ‘GO: translation initiation’ are highlighted with red. The protein Eif4e is highlighted for its unexpected change. (C) Same as (B) but RBPs annotated with the ‘UniProt keyword: mRNA processing’ are highlighted with red.

mice, Eif4g2 proteins are essential for the development of early mouse embryos (Yamanaka et al., 2000) and the differentiation of mouse embryonic stem cells (Sugiyama et al., 2017). These results indicate that upon the dissociation of oocyte-specific Eif4e-containing mRNP complexes, noncanonical translation via Eif4e-independent mechanisms, either cap-independent (Eif4g2 (Shatsky et al., 2018; Takahashi et al., 2005) and Denr (Skabkin et al., 2010)) or alternatively cap-dependent (Eif4e3 (Volpon et al., 2013)), are activated in part along with the canonical translation pathway during the OET.

Furthermore, I found that most of the RBPs related to 'RNA processing', most of which were splicing factors, had a significantly increased FAX-RIC levels in embryos (**Figure III-33C**). This result is consistent with the fact that zygotic genome transcription activation in early embryos (Tadros and Lipshitz, 2009) likely requires the involvement of RNA processing factors to process the newly produced zygotic pre-mRNAs. This result is also in line with the conclusions drawn from a similar study profiling 'dynamic RBPs' in the drosophila maternal to zygotic transition (Sysoev et al., 2016). Intriguingly, notable exceptions occurred for Cpsf2, Cpsf3, and Sympk, all of which are cleavage and polyadenylation related factors (Charlesworth et al., 2013) (**Figure III-33C**). In oocytes, these RBPs are known to form oocyte-specific

mRNP complexes with polyadenylated mRNAs in the cytoplasm (Charlesworth et al., 2013). Unlike oocytes, these RBPs are known to interact with newly transcribed mRNAs prior to polyadenylation in the embryos (Charlesworth et al., 2013) and most other somatic cells (Gruber et al., 2014), and therefore should not be identified as part of the RNA interactome via oligo-dT pulldown. Accordingly, these RBPs were notably absent from the UVX and FAX RNA interactome in HeLa cells and all of the reviewed human RNA interactome studies (Hentze et al., 2018), despite the ubiquitous expression of those RBPs in human cancer cell lines (Thul and Lindskog, 2018). Results from study thus suggest that a significant majority of these RBPs become part of the embryo stage-specific RNP complexes at or before the embryo stage. Of note, such dynamic RBPs could not be revealed by gene expression analysis alone but by the integrated analysis of the RNA interactome and global proteomics. Collectively, all of these results demonstrate that the FAX-RIC method can enable robust mapping of physiologically distinct changes in RNP complex formations in multicellular organisms *in vivo*.

III-6. FAX-RIC based RNA interactome profiling in mouse liver

I further demonstrate that FAX can enable RNA interactome profiling in the mammalian tissue samples. For this, I utilized *M. musculus* liver samples. Analyzing the profile of the nucleic acids obtained from the liver samples, treated with 2% or 4% formaldehyde (10 minutes), I found distinct and near complete change in migration pattern of genomic DNAs and RNAs at 4% FAX condition (**Figure III-34A**). Furthermore, consistent with the known characteristics of protein crosslinked RNA (Trendel et al., 2019), most of the FAX liver RNAs were found at the interphase of acid guanidinium thiocyanate-phenol chloroform extraction, suggesting that extensive RNA-protein crosslinking had occurred. On the contrary, the profiles of the extracted liver RNAs in both negative control and UVX (2 J) were largely indistinguishable from each other (**Figure III-34B**). Taken together, I found that the FAX (at 4% 10 minutes) can both penetrate and crosslink mouse liver RNA interactome with high efficiency.

To reduce the potential interference of protein-to-protein crosslinking products at such strong FAX condition, I applied the peptide-level FAX-RIC strategy, in which the peptides that are not directly crosslinked to RNAs are removed via trypsin digestion and washing. This strategy ensures the specificity of the RIC to direct RNA-protein interactions. Of note, I adapted the GITC based lysis buffer to prevent RNA degradation (Chomczynski et al.) ,

instead of the conventional LDS lysis buffer, which can be critical to animal

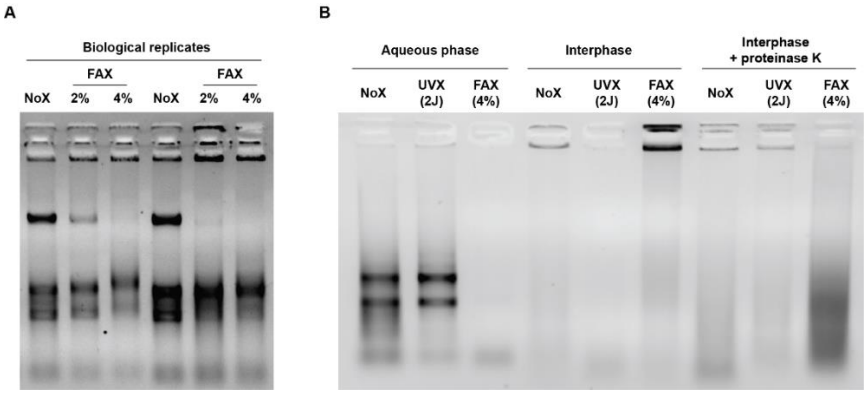


Figure III-34. Crosslinking condition test for RIC in mouse liver *in vivo*

(A) Mouse liver samples were treated with the indicated condition and the total liver lysate was run on the agarose gel, EtBr staining, two biological replicates. FAX was done for 10 minutes at the indicated FA concentration in PBS. At 4%, we observed relatively slight and yet systematic upward shift of the RNAs along with the complete disappearance or upward shift of the band that likely corresponds to the genomic DNAs, indicating that FA penetration was complete and efficient protein-nucleic acids crosslinking was achieved. Notably, there were residual amount of unfixed DNAs with FAX at 2%. (B) RNAs obtained from the indicated phase of acid guanidinium thiocyanate-phenol-chloroform extraction of the mouse liver samples treated with the indicated crosslinking conditions. All samples were treated with DNAse before gel running.

tissue samples, especially the nuclease-rich mammalian organs such as the liver (Chomczynski et al.). After tissue lysis, GITC removal, and trypsin digestion, I could effectively retrieve the RNA-peptide conjugates using oilgo-dT beads or silica columns (Asencio et al., 2018), for the profiling of poly A RNA interactome or total RNA interactome, respectively (**Figure III-35**).

Poly A RNA interactome from mouse liver was relatively small but largely consisted of the RBPs annotated with known RBDs (**Figure III-36A and C**), suggesting that the peptide-level FAX-RIC protocol can profile the tissue RNA interactome with high specificity (**Table III-2**). On the other hand, the total RNA interactome capture identified 761 significantly enriched proteins (**Figure III-36B**), the majority of which are the known RBPs (**Figure III-36C**). Significant increase in the relative protein intensity of the representative mRNA and rRNA binding proteins from the poly A and total RIC experiments, respectively, further suggested the methods' specificity to the target RNA interactome (**Figure III-37**). In sum, I introduced a new versatile strategy to enable RNA interactome profiling of the mammalian tissue samples *in vivo*.

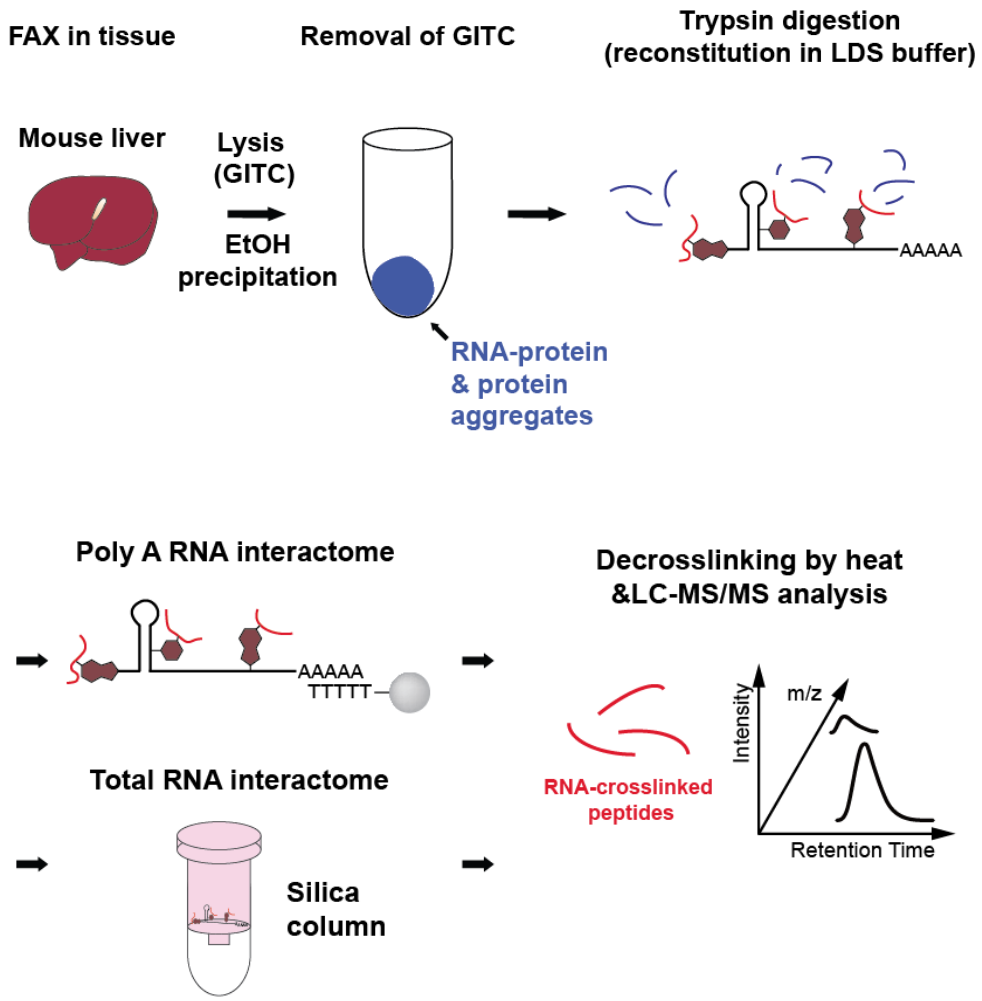


Figure III-35. Experimental scheme based on peptide-level FAX-RIC in mouse liver tissue

Oligo-dT bead and silica column (RNeasy) were used for the enrichment of poly A and total RNA interactome, respectively.

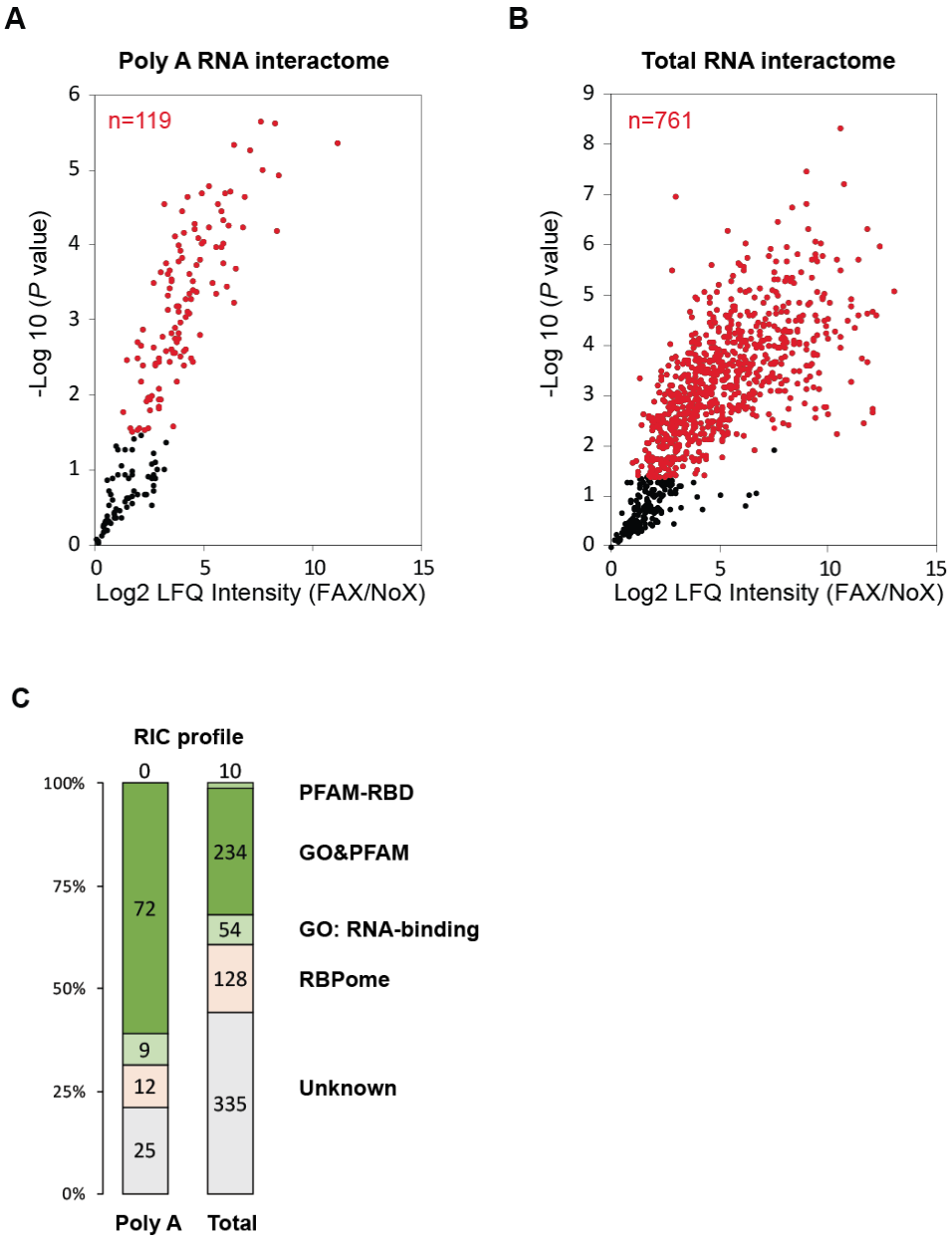

A
Poly A RNA interactome
n=119
-Log 10 (P value)
Log2 LFQ Intensity (FAX/NoX)
B
Total RNA interactome
n=761
-Log 10 (P value)
Log2 LFQ Intensity (FAX/NoX)
C
RIC profile
100%
75%
50%
25%
0%
0
10
72
9
12
25
234
54
128
335
Poly A
Total
PFAM-RBD
GO&PFAM
GO: RNA-binding
RBPome
Unknown

Figure III-36. Poly A and Total RNA interactome and their specificity to the RBPs in mouse liver

(A-B) Defining the high confidence RNA interactome in mouse liver for poly A (A) and total RNA interactome (B). Volcano plots displaying the log2 fold-change of average LFQ intensity (x-axis) and the –log10 student's t-test P value (y-axis) for all the proteins quantified in all three replicates. Proteins with log2 fold-change >1 and statistically significant enrichment over the NoX-RIC experiments (P value < 0.05, Student's t–test, adjusted by Benjamini-Hochberg method) are highlighted in red. (C) Proportion of known RBPs within the mouse liver poly A and total RNA interactome profiles.

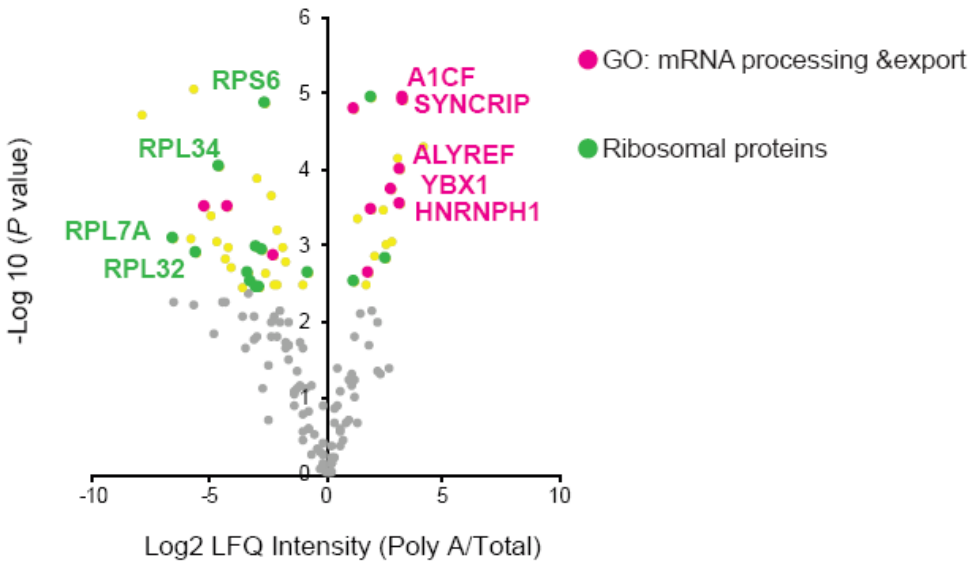


Figure III-37. RBPs with significantly increased protein intensity in poly(A)+ or Total RNA interactome

Volcano plots displaying the log2 fold-change of average LFQ intensity (x-axis) and the –log10 student's t-test P value (y-axis) for all the proteins identified in both poly(A)+ and total RNA interactome. RBPs with increased relative LFQ intensity in poly A or total RNA interactome profiles (FDR <0.01) are highlighted in yellow, among them representative mRNA binding proteins, all GO: mRNA processing and transport, and ribosomal proteins are marked with pink and green, respectively, name of the some of the proteins with most significant changes are inserted.

Table III-2. Mouse liver poly A RNA interactome

Proteins with statistically significant enrichment in the mouse liver poly A RNA interactome capture experiment are marked with +.

P19253	Rpl13a	-11.1	5.4
Q9CPR4	Rpl17	-8.4	4.9
P62960	Ybx1	-8.3	4.2
Q9Z2X1	Hnrnpf	-8.2	5.6
Q78PY7	Snd1	-7.6	5.0
Q6A0A9	FAM120A	-7.6	5.7
P62245	Rps15a	-7.1	5.3
P16460	Ass1	-6.8	4.6
P17742	Ppia	-6.8	4.2
P42669	Pura	-6.4	3.7
P62908	Rps3	-6.4	3.2
Q8VDJ3	Hdlbp	-6.3	5.3
Q9D8E6	Rpl4	-6.1	4.7
P97351	Rps3a	-6.1	4.3
E9Q3S4	Map3k19	-6.0	3.4
Q99PL5	Rrbp1	-5.9	4.7
Q5YD48	A1cf	-5.9	4.0
P62849	Rps24	-5.8	3.8
O88569	Hnrnpa2b1	-5.8	4.3
Q8BG05	Hnrnpa3	-5.8	4.5
Q9Z204	Hnrnpc	-5.7	4.0
P25444	Rps2	-5.6	4.6
Q6ZWY9	Hist1h2bc	-5.5	4.0
P09405	Ncl	-5.5	3.3
P62983	Rps27a	-5.3	3.5
Q8C3F2	Fam120c	-5.2	4.2
P12710	Fabp1	-5.1	4.8
Q91VS7	Mgst1	-4.9	4.1
P26883	Fkbp1a	-4.9	4.0
P61979	Hnrnpk	-4.8	4.7
P62751	Rpl23a	-4.8	3.8
P07901	Hsp90aa1	-4.8	2.8
P10126	Eef1a1	-4.7	4.1
Q6ZWV3	Rpl10	-4.6	3.4
O35490	Bhmt	-4.6	3.7
P62082	Rps7	-4.5	4.3

P56959	Fus	-4.5	4.2
P47915	Rpl29	-4.5	3.4
P84099	Rpl19	-4.4	2.4
Q9EPU0	Upf1	-4.4	3.5
Q9CR57	Rpl14	-4.4	2.5
P70333	Hnrnph2	-4.3	3.3
Q6PDM2	Srsf1	-4.3	3.1
P43274	Hist1h1e	-4.2	3.3
Q7TMK9	Syncrip	-4.2	3.6
Q8C196	Cps1	-4.2	3.1
O35737	Hnrnph1	-4.2	4.6
Q61176	Arg1	-4.1	3.0
P0DN91	0	-4.1	2.6
P50580	Pa2g4	-4.1	3.3
P62855	Rps26	-4.0	2.4
Q9CY58	Serbp1	-4.0	4.2
P00186	Cyp1a2	-4.0	2.6
O08583	Alyref	-4.0	3.8
Q60668	Hnrnpd	-3.9	4.4
Q921F2	Tardbp	-3.9	3.9
P84104	Srsf3	-3.9	2.5
P62301	Rps13	-3.9	3.0
Q6PB66	Lrpprc	-3.8	3.2
Q8BL66	Eea1	-3.8	4.0
P61358	Rpl27	-3.8	3.8
P60867	Rps20	-3.8	2.8
Q62093	Srsf2	-3.8	3.1
P16858	Gapdh	-3.7	2.8
P62911	Rpl32	-3.7	2.2
Q64458	Cyp2c29	-3.7	2.7
P12970	Rpl7a	-3.7	3.2
P29341	Pabpc1	-3.6	2.6
P63325	Rps10	-3.6	2.8
P17225	Ptbp1	-3.6	2.9
Q8R081	Hnrnpl	-3.6	4.1
P11499	Hsp90ab1	-3.6	1.6

Q61656	Ddx5	-3.5	2.6
P97494	Gclc	-3.5	3.5
P70694	Akr1c6	-3.5	2.6
Q9DBR1	Xrn2	-3.4	2.8
P31786	Dbi	-3.4	3.5
Q9DCF9	Ssr3	-3.4	3.4
P63242	Eif5a	-3.4	3.7
P60335	Pcbp1	-3.4	2.5
P27661	H2afx	-3.3	3.3
Q8BGD9	Eif4b	-3.3	3.1
Q8VEK3	Hnrnpu	-3.3	3.6
Q91VM5	Rbmxl1	-3.2	3.8
Q6ZWQ0	Syne2	-3.2	2.6
Q9CXW4	Rpl11	-3.1	4.5
Q8BJW6	Eif2a	-3.1	2.8
Q80WJ7	Mtdh	-3.0	1.9
O35295	Purb	-2.9	3.6
P19783	Cox4i1	-2.9	2.1
Q61133	Gstt2	-2.9	1.8
Q91Y97	Aldob	-2.9	2.5
Q99020	Hnrnpab	-2.9	1.9
P63038	Hspd1	-2.9	2.4
Q64310	Surf4	-2.8	1.8
Q8BL97	Srsf7	-2.7	2.4
Q920L1	Fads1	-2.7	2.4
Q924T2	Mrps2	-2.7	3.5
P62918	Rpl8	-2.6	2.6
Q6ZWY3	Rps27l	-2.5	2.0
P97461	Rps5	-2.5	1.9
Q8VC52	Rbpms2	-2.4	1.8
Q8K1H1	Tdrd7	-2.4	1.6
Q05920	Pc	-2.3	2.0
Q8BMJ3	Eif1ax	-2.3	1.9
Q9D1R9	Rpl34	-2.2	1.5
Q64374	Rgn	-2.1	2.4
P58252	Eef2	-2.1	2.9

P62754	Rps6	-2.1	2.2
Q64442	Sord	-2.1	2.7
P70670	Naca	-2.0	1.6
Q9D0E1	Hnrnpm	-1.9	2.7
P35979	Rpl12	-1.9	2.5
Q922Q8	Lrrc59	-1.9	1.5
Q9JII6	Akr1a1	-1.7	1.5
P62717	Rpl18a	-1.5	1.6
P97379	G3bp2	-1.4	2.5
Q9DB15	Mrpl12	-1.2	1.8

Chapter IV.

Conclusion

UVX has long been regarded as the gold standard of *in vivo* crosslinking method and highly specific for RNA-protein interaction, in particular due to its 'zero-length' characteristic to the interaction. However, it has been reported that UVX can irreversibly crosslink protein-protein interactions as well (Itri et al.; Leo et al.), indicating that a certain degree of false positive RBPs will be profiled via protein-protein crosslinking even from UVX-RIC experiment as a similar concern exists in FAX. Moreover, UVX can induce RNA damaging or fragmentation (Beckmann et al.; Kladwang et al.), which can reduce the RNA recovery of oligo-dT pulldown and thereby decrease the overall quantity of captured proteins. These drawbacks of UVX have been largely neglected in the field of RNA biology and it would be appropriate to re-evaluate the value of UVX and FAX for the researches of RNA-protein interactions.

In this study, I report that FAX can capture RNA-protein interaction with high specificity and efficiency not only in cultured cells but also in multicellular organisms. Through the first system-wide and quantitative comparison of two *in vivo* crosslinking methods (FAX vs. UVX), I solidly demonstrated that FAX-RIC was more efficient and as highly specific as UVX-RIC for the mapping of the *in vivo* RNA-protein interactions, while possessing advantages over UVX, particularly in opaque samples such as *X.*

laevis embryos and mammalian tissue samples. Furthermore, for the first time I also performed the systematic de novo analyses of RNA interactome transformations during OET using the FAX-RIC in addition to global proteome profiling data. The majority of the differentially enriched RBPs had no significant change in protein expression level, underscoring the importance of the RIC method in discovering functionally regulated RNA-protein interactions *in vivo*. The significant changes in the FAX-captured RBP profiles were clearly reflective of the known changes in RNP complex functions during early animal development (Richter and Lasko; Sysoev et al.) and disclosed some of the under-evaluated components of RNP complexes, such as those associated with noncanonical translational pathway.

Despite significant expansion of our knowledge in the RBP repertoire, many aspects of context-dependent RNA-protein interactions, e.g., such 'dynamic RBPs' in OET, demand to be further elucidated especially in human physiology and diseases. FAX-RIC can allow for the profiling of the context-dependent 'dynamic RBPs' in various human tissues in combination with global proteome profiling or/and post-translational modification (PTM) proteomics data. Such integrated approach would serve as a powerful platform for discovering novel key regulatory RBPs and PTMs of such RBPs (i.e., RBP-code). I thus expect that FAX-RIC would significantly broaden our

understanding of the dynamic mRNP formation in multicellular organisms and human tissues *in vivo* as I demonstrate with mouse liver tissues.

국문초록

다세포생물체 내에서의 RNA-단백질 간의 상호작용에 대한 단백체학적 연구

서울대학교 대학원 생명과학부

나용우

생성부터 분해에 이르기 까지 mRNA는 수십개 이상의 RNA 결합 단백질과 상호작용하며 mRNP 복합체 형태로 존재한다. 그와 같은 복합체 안에서 역동적으로 변화하고 긴밀하게 조절되는 RNA 결합 단백질의 구성은 mRNA 의 전사와 번역 그리고 분해에 이르는 모든 단계를 조절 한다. 그러므로 RNA와 단백질간의 상호작용의 조절은 전사 후 유전자 발현 조절에 필수 요소로 작용한다. 최근에 개발되고 소개된 RNA 결합 단백질 프로파일링 방법들은 생물체내에서의 교차 결합의 생성을 통해 가능하게 되었다. 자외선 빛 조사에 기반한 생물체 내에서의 RNA와 단백질간의 교차결합은 현재 RNA 생물학 전반에서 가장 널리 쓰이는 방법이지만 조직이나 다세포생물체 안으로 침투하는 깊이가 얕아 그와 같은 실험 환경에 대한 연구에서는 효과적이지 않다는 뚜렷한 한계점이 있다.

이 학위논문에서 나는 포름알데히드 기반의 생물체내 교차결합 방법을 이용하여 RNA 결합 단백질 구성요소를 프로파일링 하는 방법론을 개발하고 그 방법의 효용성을 확인하였다. 먼저 새롭게 개발된 방법론은 배양된 세포내에서 RNA와 직접결합하는 것으로 알려져있는 인간 단백질에 높은 특이성을 가짐을 확인할수 있었다. 나아가서 자외선 조사 기법 또는 포름알데히드에 기반한 방법에 의해 비교적 높은 효율로 분석된 RNA 결합 단백질들의 확인을 통해서 두 방법론의 특이점을 확인할수 있었다.

포름알데히드에 의해서 생성된 교차결합은 높은 온도에서는 빠르게 역반응을 일으키고 없어진다는 특징을 가지고 있다. 이와같은 성질을 이용하여 나는 펩타이드 수준에서의 RNA 결합 단백질 동정 방법론을 개발하여 포름알데히드에 의해서 RNA와 교차결합되는 부근의 단백질 서열을 확인하는 것 또한 가능하게 하였다.

또한 포름알데히드에 기반한 RNA 결합 단백질 동정 방법은 *Xenopus laevis* 의 난자와 배아에 적용 되었고 그 결과는 자외선 조사 기반의 방법을 통하여 얻은 결과와 비교되었다. 새로운 방법론이 훨씬 더 포괄적이고 정확한 RNA 결합 단백질 동정을 가능하게 함이 그 비교 실험결과를 통하여 확인되었다. 나아가서 두개의 다른 발생단계에서 동정된 단백질들의 양적 비교를 통하여 난소에서 초기 배아 발생과정에서 동적으로 변화된 RNA와 프로틴간의 상호작용을 확인 할수 있었다. 그 중 특히 주목할만한 결과는 중요한 단백질 번역 개시 인자인 eIF4E 와 eIF4E3 의 변화 등이었다.

펩타이드 수준에서의 RNA 결합 단백질 동정 방법론은 또한 포유류의 조직내에서의 RNA 결합 단백질 동정 방법을 개발하는데 사용이 되었다. 새롭게 개발된 방법은 쥐의 간에 적용되어 아데닌 꼬리를 가진 RNA와 전체 RNA에 특이적인 RNA 결합 단백질을 확인하는것을 가능케 하였다.

종합적으로 포름알데히드에 기반한 RNA 결합 단백질 분석 방법론이 개발되었고, 이 방법론의 효용성이 배양된 인간 세포, *X. laevis* 의 배아, *M. musculus* 의 간 조직 등에서 심도 있게 조사되고 확인되었다. HeLa 세포주에서 확인한 결과는 자외선 조사와 포름알데히드 라는 두개의 교차결합 방법의 차이가 RNA 결합 단백질들을 동정하는데 있어 큰 차이를 줄 수 있음을 시사한다. 이 연구에서 확인된 동물의 난자에서 배아로의 전환 과정에서의 역동적인 RNA 단백질간의 상호작용 변화는 앞으로 생명체 발생과정에서의 RNA 결합 단백질의 중요성을 연구하는데 있어서 단초 역할을 할것으로 기대된다. 마지막으로 이 방법론이 포유류의 조직에도 잘 적용될수 있다는 것을 확인한 것은 앞으로 이 학위논문에서 소개된 방법론이 다세포 생물 조직의 향상성 유지와 인간의 질병 발생 등을 연구하는데에 널리 사용될 수 있음을 시사한다.

REFERENCES

Asencio, C., Chatterjee, A., and Hentze, M.W. (2018). Silica-based solid-phase extraction of cross-linked nucleic acid–bound proteins. Life Science Alliance 1, e201800088.

Baltz, A.G., Munschauer, M., Schwanhausser, 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. Mol Cell 46, 674-690.

Beckmann, B.M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. Pflugers Arch 468, 1029-1040.

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.

Carmody, S.R., and Wente, S.R. (2009). mRNA nuclear export at a glance. J Cell Sci 122, 1933-1937.

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., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell 63, 696-710.

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

Charlesworth, A., Meijer, H.A., and de Moor, C.H. (2013). Specificity factors in cytoplasmic polyadenylation. Wiley Interdiscip Rev RNA 4, 437-461.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell 161, 404-416.

Chu, J., Cargnello, M., Topisirovic, I., and Pelletier, J. (2016). Translation Initiation Factors: Reprogramming Protein Synthesis in Cancer. Trends in Cell Biology 26, 918-933.

Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., Straube,

K., Gerstein, M.B., Butter, F., and Neugebauer, K.M. (2017). Dynamic RNA–protein interactions underlie the zebrafish maternal-to-zygotic transition. Genome Research.

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

Elinson, R.P., and Pasceri, P. (1989). Two UV-sensitive targets in dorsoanterior specification of frog embryos. Development 106, 511-518.

Gavrilov, A., Razin, S.V., and Cavalli, G. (2015). In vivo formaldehyde cross-linking: it is time for black box analysis. Brief Funct Genomics 14, 163-165.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat Rev Genet 15, 829-845.

Gruber, A.R., Martin, G., Keller, W., and Zavolan, M. (2014). Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. Wiley interdisciplinary reviews RNA 5, 183-196.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nature Reviews Molecular Cell Biology 19, 327.

Hentze, M.W., and Kulozik, A.E. (1999). A Perfect Message: RNA Surveillance and Nonsense-Mediated Decay. Cell 96, 307-310.

Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., and von Hippel, P.H. (1986). Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. Journal of Biological Chemistry 261, 3512-3518.

Hoffman, E.A., Frey, B.L., Smith, L.M., and Auble, D.T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem 290, 26404-26411.

Houseley, J., and Tollervey, D. (2009). The Many Pathways of RNA Degradation. Cell 136, 763-776.

Jung, J., Jeong, K., Choi, Y., Kim, S.A., Kim, H., Lee, J.W., Kim, V.N., Kim, K.P., and Kim, J.S. (2019). Deuterium-Free, Three-Plexed Peptide Diethylation for Highly Accurate Quantitative Proteomics. Journal of proteome research 18, 1078-1087.

Kim, B., Jeong, K., and Kim, V.N. (2017). Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates. Mol Cell 66, 258-269 e255.

Kim, Y., Lee, J.H., Park, J.-E., Cho, J., Yi, H., and Kim, V.N. (2014). PKR is activated by cellular dsRNAs during mitosis and acts as a mitotic regulator. Genes & development 28, 1310-1322.

Knoener, R.A., Becker, J.T., Scalf, M., Sherer, N.M., and Smith, L.M. (2017). Elucidating the in vivo interactome of HIV-1 RNA by hybridization capture and mass spectrometry. Scientific Reports 7, 16965.

Kroll, T.T., Zhao, W.M., Jiang, C., and Huber, P.W. (2002). A homolog of FBP2/KSRP binds to localized mRNAs in Xenopus oocytes. Development 129, 5609-5619.

Langdon, E.M., and Gladfelter, A.S. (2018). A New Lens for RNA Localization: Liquid-Liquid Phase Separation. Annual review of microbiology 72, 255-271.

Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C.K., Ragan, C., Kurscheid, S., Pagler, E., Yang, H., et al. (2016). The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. Cell Rep 16, 1456-1469.

Lund, E., Sheets, M.D., Imboden, S.B., and Dahlberg, J.E. (2011). Limiting Ago protein restricts RNAi and microRNA biogenesis during early development in Xenopus laevis. Genes Dev 25, 1121-1131.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8, 479-490.

Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif,

a plastic RNA-binding platform to regulate post-transcriptional gene expression. Febs j 272, 2118-2131.

Nakamura, Y., Tanaka, K.J., Miyauchi, M., Huang, L., Tsujimoto, M., and Matsumoto, K. (2010). Translational repression by the oocyte-specific protein P100 in Xenopus. Dev Biol 344, 272-283.

Necci, M., Piovesan, D., Dosztányi, Z., and Tosatto, S.C.E. (2017). MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. Bioinformatics 33, 1402-1404.

Panhale, A., Richter, F.M., Ramírez, F., Shvedunova, M., Manke, T., Mittler, G., and Akhtar, A. (2019). CAPRI enables comparison of evolutionarily conserved RNA interacting regions. Nature Communications 10, 2682.

Perez-Perri, J.I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., and Hentze, M.W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nature Communications 9, 4408.

Peshkin, L., Wuhr, M., Pearl, E., Haas, W., Freeman, R.M., Jr., Gerhart, J.C., Klein, A.M., Horb, M., Gygi, S.P., and Kirschner, M.W. (2015). On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Developmental cell 35, 383-394.

Peuchen, E.H., Cox, O.F., Sun, L., Hebert, A.S., Coon, J.J., Champion, M.M., Dovichi, N.J., and Huber, P.W. (2017). Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. Scientific Reports 7, 15647.

Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). Nature Biotechnology 37, 169-178.

Radford, H.E., Meijer, H.A., and de Moor, C.H. (2008). Translational control by cytoplasmic polyadenylation in Xenopus oocytes. Biochim Biophys Acta 1779, 217-229.

Richter, J.D. (2007). CPEB: a life in translation. Trends in Biochemical Sciences 32, 279-285.

Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 433, 477-480.

Shatsky, I.N., Terenin, I.M., Smirnova, V.V., and Andreev, D.E. (2018). Cap-Independent Translation: What's in a Name? Trends in Biochemical Sciences 43, 882-895.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell 151, 750-764.

Sive, H.L., Grainger, R.M., and Harland, R.M. (2007). Xenopus laevis In Vitro Fertilization and Natural Mating Methods. Cold Spring Harbor Protocols 2007, pdb.prot4737.

Sive, H.L., Grainger, R.M., and Harland, R.M. (2010). Isolation of Xenopus Oocytes. Cold Spring Harbor Protocols 2010, pdb.prot5534.

Skabkin, M.A., Skabkina, O.V., Dhote, V., Komar, A.A., Hellen, C.U.T., and Pestova, T.V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. Genes & development 24, 1787-1801.

Somasekharan, Syam P., Zhang, F., Saxena, N., Huang, Jia N., Kuo, I.C., Low, C., Bell, R., Adomat, H., Stoynov, N., Foster, L., et al. (2020). G3BP1-linked mRNA partitioning supports selective protein synthesis in response to oxidative stress. Nucleic Acids Research 48, 6855-6873.

Standart, N., and Minshall, N. (2008). Translational control in early development: CPEB, P-bodies and germinal granules. Biochem Soc Trans 36, 671-676.

Sugiyama, H., Takahashi, K., Yamamoto, T., Iwasaki, M., Narita, M., Nakamura, M., Rand, T.A., Nakagawa, M., Watanabe, A., and Yamanaka, S. (2017). Nat1 promotes translation of specific proteins that induce differentiation of mouse embryonic stem cells. Proceedings of the National Academy of Sciences 114, 340-345.

Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A., and Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat Commun 7, 12128.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development 136, 3033-3042.

Takahashi, K., Maruyama, M., Tokuzawa, Y., Murakami, M., Oda, Y., Yoshikane, N., Makabe, K.W., Ichisaka, T., and Yamanaka, S. (2005). Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). Genomics 85, 360-371.

Thavarajah, R., Mudimbaimannar, V.K., Elizabeth, J., Rao, U.K., and Ranganathan, K. (2012). Chemical and physical basics of routine formaldehyde fixation. J Oral Maxillofac Pathol 16, 400-405.

Thul, P.J., and Lindskog, C. (2018). The human protein atlas: A spatial map of

the human proteome. Protein Sci 27, 233-244.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell 176, 391-403.e319.

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols 11, 2301.

Virant-Klun, I., Leicht, S., Hughes, C., and Krijgsveld, J. (2016). Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. Molecular & cellular proteomics : MCP 15, 2616-2627.

Volpon, L., Osborne, M.J., Culjkovic-Kraljacic, B., and Borden, K.L.B. (2013). eIF4E3, a new actor in mRNA metabolism and tumor suppression. Cell Cycle 12, 1159-1160.

Wheeler, E.C., Van Nostrand, E.L., and Yeo, G.W. (2018). Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley Interdiscip Rev RNA 9.

Wu, X., Wang, P., Brown, C.A., Zilinski, C.A., and Matzuk, M.M. (2003). Zygote arrest 1 (Zar1) is an evolutionarily conserved gene expressed in vertebrate ovaries. Biol Reprod 69, 861-867.

Wuhr, M., Guttler, T., Peshkin, L., McAlister, G.C., Sonnett, M., Ishihara, K., Groen, A.C., Presler, M., Erickson, B.K., Mitchison, T.J., et al. (2015). The Nuclear Proteome of a Vertebrate. Curr Biol 25, 2663-2671.

Yamanaka, S., Zhang, X.Y., Maeda, M., Miura, K., Wang, S., Farese, R.V., Jr., Iwao, H., and Innerarity, T.L. (2000). Essential role of NAT1/p97/DAP5 in embryonic differentiation and the retinoic acid pathway. The EMBO journal 19, 5533-5541.

Yong, J., Kasim, M., Bachorik, J.L., Wan, L., and Dreyfuss, G. (2010). Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. Mol Cell 38, 551-562.