Identification and comparative analysis of the β-actin mRNA interactome by RNA-proximity labeling in mouse embryonic fibroblast

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Joyita Mukherjee
aus Kolkata, India

Tübingen
2015

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	16.09.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Ralf-Peter Jansen
2. Berichterstatter:	Prof. Dr. Gabriele Dodt

Declaration

This thesis describes my work conducted in the laboratory of Prof. Ralf-Peter Jansen at Interfakultäres Institut für Biochemie (IFIB), University of Tübingen, Tübingen, Germany from April 2015 to August 2019. The work was co-supervised by Prof. Dr. Gabriele Dodt also at IFIB, University of Tübingen, Tübingen, Germany. I declare that this thesis is the product of my own work. The parts that have been published or where other sources have been used were cited accordingly.

Contents

1. Abbreviations:

43S	**43S** pre-initiation complex
aa	**A**mino **a**cid
ADAR1	**A**denosine **d**eaminase **1** **a**cting on **R**NA
ADAR2	**A**denosine **d**eaminase **2** **a**cting on **R**NA
ALKBH5	α-ketoglutarate-dependent dioxygenase **alkB** **h**omolog **5**
APEX	**A**scorbate **pe**ro**x**idase
APEX2	**A**scorbate **pe**ro**x**idase **2**
APA	**A**lternative **p**oly**a**denylation
APOBEC	**Apo**lipoprotein **B** mRNA **e**diting **e**nzyme, catalytic polypeptide-like
Arp2/3	**A**ctin-**r**elated **p**rotein complex **⅔**
BDNF	**B**rain-**d**erived **n**eurotrophic **f**actor
BirA	**Bi**otin-[acetyl-CoA-carboxylase] ligase and a biotin-operon **r**epressor (BirA-biotinoyl-5'-**A**MP)
BioID	Proximity-dependent **bio**tin **id**entification
CBP	**C**ap **b**inding **p**roteins
CDAR	**C**ytosine **d**eaminase **a**cting on **R**NA
CDS	**C**o**d**ing **s**equence
CLIP	**C**ross**l**inking and **i**mmuno**p**recipitation
CPSF	**Cl**eavage and **p**olyadenylation **s**pecific **f**actor
CstF	**C**leavage **st**imulation **f**actor
Dcp	**D**e**c**apping **p**rotein
DNA	**D**eoxyribo**n**ucleic **a**cid
DSIF	**D**RB **s**ensitivity **i**nducing **f**actor
GMP	**G**uanosine **m**ono**p**hosphate
GTFs	**G**eneral **t**ranscription **f**actors
GTP	**G**uanosine **t**riphos**p**hate
eIF	**e**ukaryotic translation **i**nitiation **f**actor
EJC	**E**xon **j**unction **c**omplex
FBPs	**F**ar upstream element-**b**inding **p**rotein

FG	Phenylalanine-glycine
FISH	**F**luorescent **i**n **s**itu **h**ybridization
FMR1	**F**ragile X **m**ental **r**etardation **1**
FTO	Fat mass and obesity-associated
FUBP1	**F**ar **u**pstream element-**b**inding **p**rotein **1**
FUBP2	**F**ar **u**pstream element-**b**inding **p**rotein **2**
FUBP3	**F**ar **u**pstream element-**b**inding **p**rotein **3**
hnRNP	**H**eterogeneous **n**uclear **r**ibo**n**ucleo**p**rotein
HSP	**H**eat-**s**hock **p**rotein
HuR	**Hu**man antigen **R**
IGF2BP1	**I**nsulin-like **g**rowth **f**actor **2** mRNA **b**inding **p**rotein **1**
IGF2BP2	**I**nsulin-like **g**rowth **f**actor **2** mRNA **b**inding **p**rotein **2**
IGF2BP3	**I**nsulin-like **g**rowth **f**actor **2** mRNA **b**inding **p**rotein **3**
IMP1	**Imp**rintor **1**
KHDRBS1	**KH d**omain-containing, **R**NA-**b**inding, **s**ignal transduction-associated protein **1**
KHSRP	**KH-**Type **S**plicing **R**egulatory **P**rotein
MAP2	**M**icrotubule-**a**ssociated **p**rotein **2**
MARTA1	**MA**P2-**R**NA **t**rans-**a**cting protein **1**
MARTA2	**MA**P2-**R**NA **t**rans-**a**cting protein **2**
MBS	**M**S2 **b**inding **s**ites
MBP	**M**yelin **b**asic **p**rotein
MCP	**M**S2 **c**oat **p**rotein
MEFS	**M**ouse **e**mbryonic **f**ibroblast**s**
METTL	**Met**hyl **t**ransferase **l**ike protein
miRNA	**mi**cro **RNA**
MRE	**m**iRNA **r**egulatory **e**lement
mRNA	**M**essenger **r**ibo**n**ucleic **a**cid
ncRNA	**N**on**c**oding **RNA**
NELF	**N**egative **e**longation **f**actor

NGD	**N**o-**g**o **d**ecay
NMD	**N**on-sense **m**ediated **d**ecay
NPC	**N**uclear **p**ore **c**omplex
NSD	**N**on-**s**top **d**ecay
NXF	**N**uclear e**x**port **f**actor
nt	**n**ucleo**t**ide
ODN	**O**ligo**d**eoxy**n**ucleotides
OligodT	**D**eoxy**t**hymidine **oligo**nucleotides
ORF	**O**pen **r**eading **f**rame
PABP	**p**oly(**A**) **b**inding **p**rotein
PAP	**p**oly(**A**) **p**olymerase
PAS	**p**oly**a**denylation **s**ite
PCR	**P**olymerase **c**hain **r**eaction
PDGF	**P**latelet-**d**erived **g**rowth **f**actor
PIP3	**P**hosphat**i**dylinositol (3,4,5)-tris**p**hosphate
pre-mRNAs	**Pre**cursor **mRNAs**
P-TEFb	**P**ositive **t**ranscription **e**longation **f**actor **b**
PTMs	**P**ost-**t**ranscriptional **m**odification**s**
qRRM	**q**uasi **R**NA **r**ecognition **m**otif
RACK1	**R**eceptor of **a**ctivated protein **C** **k**inase **1**
Ran	**RA**s-related **N**uclear protein
RBPs	**R**NA **b**inding **p**rotein**s**
RCA	**R**olling circle **a**mplification
REF	**R**NA and **e**xport **f**actor binding protein
RGC	**R**etinal **g**anglion **c**ell
RLR	**R**NA **l**ocalization **r**egion
RNA	**R**ibo**n**ucleic **a**cid
RNP	**R**ibo**n**ucleo**p**rotein
RRM	**R**NA **r**ecognition **m**otif
rRNA	**R**ibosomal **RNA**
RTS	**R**NA **t**ransport **s**equence

SAGE	**S**erial **a**nalysis of **g**ene **e**xpression
Sam68	**S**rc-**a**ssociated in **m**itosis **68** KDa
smFISH	**S**ingle-**m**olecule **f**luorescence **i**n **s**itu **h**ybridization
snRNAs	**s**mall **n**uclear **RNAs**
snoRNAs	**s**mall **n**ucle**o**lar **RNAs**
SRP	**S**ignal **r**ecognition **p**article
SS	**S**plice **s**ites
SSH	**S**uppression of **s**ubtractive **h**ybridization
STAU1	**S**taufen **1**
STAU2	**S**taufen **2**
RDA	**R**epresentational **d**ifference **a**nalysis
TFIIH	**T**ranscription **f**actor **II H**
TREX	**Tr**anscription/**ex**port complex
tRNA	**T**ransfer **RNA**
UHM	**U**2AF **h**omology **m**otif
UTR	**U**n**t**ranslated **r**egion
VICKZ	**V**g1 RBP/Vera, **I**MP-1,2,3, **C**RD-BP, **K**OC, **Z**BP-1
ZBP1	**Z**ipcode **b**inding **p**rotein **1**
ZBP2	**Z**ipcode **b**inding **p**rotein **2**

2. Summary:

The temporal and spatial expression of genes is required for maintaining cellular asymmetry, proper embryonic development, neuronal function, and cell fate. In mouse embryonic fibroblasts (MEFs) this cellular asymmetry is generated by localizing various cellular mRNAs to the protrusions (lamellipodia/filopodia). Among those mRNAs, β-actin mRNA plays a major role in defining cellular asymmetry by its localization to the cell periphery. Upon mRNA localization and translation, β-actin protein helps the cells to respond to extracellular cues and to move during extracellular matrix remodeling to maintain tissue homeostasis and tissue repair, traversing changes in local tissue environments as needed in tissue degradation, repair or regeneration.

Under normal trophic conditions, the localization of β-actin mRNA to the cellular protrusions of fibroblasts or growth cones in neurons is regulated by a cis-acting localization element or localization signal known as zipcode (in case of β-actin, this is a 54 nt long sequence in the 3'UTR of the β-actin mRNA adjacent to its stop codon) together with trans-acting factors, mainly RNA-binding proteins (RBPs) that either bind directly to the zipcode or regulate the binding of other RBPs to it. In the case of the motor-driven movement of these localized mRNAs, such RNA-protein complexes are then tethered to molecular motors such as kinesin, dynein, or myosin, to form transport or locasome complexes. Thus, messenger ribonucleoprotein particles (mRNPs) that act as functional units not only contain the information for an encoded polypeptide but also determine the precise spatio-temporal regulation of its translation, thereby facilitating the correct subcellular localization of the translation product.

It has been shown that the localization of β-actin mRNA is dependent on the binding of the zipcode-binding protein ZBP1 (an RBP of the conserved VICKZ RNA-binding protein family) to its cognate site present in the 3'UTR of the mRNA. ZBP1 (also called IGF2BP1 or IMP1) interacts with the zipcode via two K-homology (KH) RNA-binding domains by RNA looping mechanism and is required for β-actin mRNA localization in migrating cells including fibroblasts and neurons. In addition, in fibroblasts, it is also known that it controls the translation of β-actin by blocking the assembly of ribosomes at the start codon. Apart from ZBP1, the RBPs IGF2BP2, RACK, KHDRBS1/Sam68, and FMR1 play important roles during the localization of the mRNA. To obtain a complete picture of the associated proteome of any mRNA has been challenging. The high throughput methods available so far (like CLIP, MS2 pull down) mainly fail in the identification of indirect or transient interactors of specific RNAs. To solve these issues, I applied a BioID method where a protein of interest is fused to a mutant version of the E.coli biotin ligase BirA (BirA*), which biotinylates accessible lysine residues of proteins present in its vicinity. After cell lysis, biotinylated proteins can be isolated by streptavidin affinity purification and identified using standard mass spectrometry techniques.

In this thesis, I report that tethering of BirA* to a specific localized, MS2- tagged mRNA does not only allow the identification of its associated proteins but can also be used to probe the environment of this mRNA. This approach allows, with high confidence, to identify novel functional β-actin interactors like FUBP3/MARTA2, STAU1, and STAU2. FUBP3 is an RBP from the conserved FUBP family of proteins. FUBP3 shown to mediate the dendritic targeting of MAP2 mRNA in neurons. In this thesis, I report FUBP3 to bind to and facilitate localization of β-actin mRNA to fibroblast protrusions. By immunoprecipitation and in vitro binding assays, I could demonstrate that it binds 460 nt downstream of the stop codon in the β-actin 3' UTR and participates in the localization of the mRNA to the cellular protrusions. Apart from BirA* I also applied direct MS2-MCP pull-down APEX2-mediated biotin labeling of beta-actin associated proteins and compared the obtained datasets of the proteins that bind to the β-actin mRNA directly or via transient interactions.

The established method convincingly shows

1. Additional proteins which could be a part of the β-actin localization complex. Amongst all these proteins, FUBP3 has shown to be a part of the β-actin localization complex for the first time.

2. FUBP3 to bind to downstream of the localization element at the 3'UTR of β-actin mRNA and is essential for the localization of β-actin mRNAs at the protrusions of fibroblasts.

3. Comparison of the β-actin proteome under serum-starved and unstarved conditions and the difference between the associated RNA interacting proteome under these two conditions.

2.1 Zusammenfassung

Die zeitliche und räumliche Expression von Genen dient dazu, zelluläre Asymmetrie zu etablieren, der Determination des Zellschicksals, aber auch zur Embryonalentwicklung oder der neuronalen Funktion. In embryonalen Fibroblasten der Maus (MEFs) wird die zelluläre Asymmetrie durch Lokalisierung verschiedener zellulärer mRNAs an die Lamellipodien oder Filopodien etabliert. Unter diesen mRNAs spielt β-Actin-mRNA eine wichtige Rolle. Durch die lokale Translation seiner mRNA unterstützt das β-Actin-Protein die Zellen, bei ihrer Reaktion auf extrazelluläre Signale und sich während des Umbaus der extrazellulären Matrix zu bewegen, um die Gewebehomöostase und die Gewebereparatur aufrechtzuerhalten.

Unter normalen trophischen Bedingungen wird die Lokalisierung von β-Actin-mRNA zu den zellulären Vorsprüngen von Fibroblasten oder Wachstumskegeln in Neuronen durch ein cis-wirkendes Lokalisierungselement (LE) reguliert, das auch als „Zipcode“bekannt ist (im Fall von β-Actin ist dies eine 54 Nukleotide lange Sequenz hinter dem Stop-Codon in seiner 3'UTR). Dieses Element arbeitet zusammen mit trans-wirkenden Faktoren, hauptsächlich RNA-bindenden Proteinen (RBPs), die entweder direkt an den „Zipcode“ binden oder die Bindung anderer RBPs regulieren. Im Falle des aktiven Transports dieser lokalisierten mRNAs werden solche RNA-Protein-Komplexe dann an molekulare Motoren wie Kinesin, Dynein oder Myosin gebunden, um Transport- oder Locasom-Komplexe zu bilden. So enthalten Boten-Ribonukleoprotein-Partikel (mRNPs), die als funktionelle Einheiten fungieren, nicht nur die Information für ein kodiertes Polypeptid, sondern bestimmen auch die genaue räumlich-zeitliche Regulation seiner Translation, wodurch die korrekte subzelluläre Lokalisierung des Translationsprodukts erleichtert wird.

Es wurde gezeigt, dass die Lokalisierung von β-Actin-mRNA von der Bindung des „Zipcode“-bindenden Proteins ZBP1 (ein RBP der konservierten VICKZ-RNA-bindenden Proteinfamilie) abhängt. ZBP1 (auch IGF2BP1 oder IMP1 genannt) interagiert mit dem LE über zwei K-Homologie (KH) -RNA-Bindungsdomänen und ist für die Lokalisierung von β-Actin-mRNA in migrierenden Zellen, einschließlich Fibroblasten und Neuronen, erforderlich. Darüber hinaus ist bei Fibroblasten auch bekannt, dass es die Translation von β-Actin kontrolliert, indem es die Assemblierung von Ribosomen am Startcodon blockiert. Neben ZBP1 spielen die RBPs IGF2BP2, RACK, KHDRBS1 / Sam68 und FMR1 weitere wichtige Rollen bei der Lokalisierung der mRNA. Bisher war es allerdings eine Herausforderung, ein vollständiges Bild des assoziierten Proteoms einer spezifischen mRNA zu erhalten. Die bisher verfügbaren Hochdurchsatzmethoden (wie CLIP oder MS2 Affinitätsreinigung) versagen hauptsächlich bei der Identifizierung indirekter oder transienter Interaktoren spezifischer RNAs. Um diese Probleme zu lösen, habe ich eine BioID-Methode

angewendet, bei der ein Protein mit einer mutierten Version der E. coli-Biotinligase BirA (BirA *) fusioniert wird, die zugängliche Lysinreste von Proteinen in der Nähe biotinyliert. Nach der Zelllyse können biotinylierte Proteine mittels Streptavidin-Affinitätsreinigung isoliert und unter Verwendung von Standard-Massenspektrometrietechniken identifiziert werden.

In dieser Arbeit zeige ich, dass die Anheftung von BirA * an eine bestimmte lokalisierte, MS2-markierte mRNA nicht nur die Identifizierung der assoziierten Proteine ermöglicht, sondern auch zur Untersuchung der Umgebung dieser mRNA verwendet werden kann. Mit diesem Ansatz können mit hoher Sicherheit neuartige funktionelle β-Actin-Interaktoren wie FUBP3 / MARTA2, STAU1, STAU2 identifiziert werden. FUBP3 ist ein RBP aus der konservierten FUBP-Proteinfamilie. Bekannt war, dass FUBP3 die dendritische Lokalisierung von MAP2-mRNA in Neuronen vermittelt. Wie in dieser Arbeit gezeigt, bindet FUBP3 an β-Actin-mRNA hilft bei der Lokalisierung der mRNA in Fibroblasten. Durch Immunpräzipitation und in-vitro-Bindungsassays konnte ich zeigen, dass es 460 Nukleotide stromabwärts des Stop-Codons in der β-Actin-3-UTR bindet. Des weiteren wurde die Biotinylierungsmehtode mittels BirA* mit einer (direkten) MS2-MCP-Affiitätsreinigung, sowie einer APEX2-vermittelten Biotin-Markierung verglichen, um seine Effektivität abschätzen zu können. Die erhaltenen Datensätze der Proteine ermöglichen Rückschlüsse auf direkte Bindungspartner oder solche, die über transiente Wechselwirkungen an die β-Actin-mRNA binden.

Die etablierten Methoden zeigen letztendlich:

1. Es existieren zusätzliche Proteine, die Teil des β-Actin-Lokalisierungskomplexes sein könnten. Unter all diesen Proteinen ist FUBP3 zum ersten Mal als Teil des β-Actin-Lokalisierungskomplexes nachgewiesen.

2. FUBP3 bindet stromabwärts des bekannten Lokalisierungselements an der 3'UTR von β-Actin-mRNA und ist für die Lokalisierung von β-Actin-mRNAs an den Vorsprüngen von Fibroblasten essentiell.

3. Der Vergleich des β-Actin-Interaktoms unter Bedingungen mit und ohne Serummangel im Medium zeigt Differenzen zwischen dem assoziierten RNA-interagierenden Proteom unter diesen beiden Bedingungen.

Diese Studie zeigt die Vorteile der Proximity-Biotinylierungsmethode für das Verständnis des Verhaltens jeder mRNA (in diesem Fall β-Actin), indem unser Wissen über ihre Proteininteraktoren erweitert wird.

3. Introduction:

3.1 The regulation of gene expression in eukaryotes: an overview

For most living organisms, deoxyribonucleic acid (DNA) is the key genetic determinant of life. Being the most important entity of eukaryotic life, DNA is protected within the nucleus, in a supercoiled form, within a double membrane barrier. This architecture results in selective accessibility by other cellular components, mainly by proteins. From a fertilized egg to all the differentiated cell types in the adult organism, the DNA remains unchanged, but the organism's complexity comes from selective and specialized use of functional units called genes. The gene, which can produce a transacting diffusible chemical product and has a cis-acting regulatory element, regulates the whole cellular function by a method called gene expression. The great study by Jacob and Monod in the 1960s (Jacob and Monod, 1961) regarding how single-cell organisms adapt themselves to nutrient supply led to the discovery of *lac* operon which led the groundwork for the modern understanding of gene expression and gene regulation.

The different steps of gene expression (DNA to RNA to protein) are collectively called the central dogma of molecular biology. The first step of gene expression is transcription. Being stationary within the nucleus, the DNA makes mobile copies of itself termed as mRNA that later on serves as a template for protein synthesis, a process called mRNA translation, which is the last big step in gene expression. Since gene expression has such a tremendous impact on cell fate and physiology, a huge number of control mechanisms have been installed on the epigenetic, transcriptional, and translational levels to control regulation of gene expression, summarized in Fig. 1.

The first level of regulation in gene expression is found in the DNA itself, by modifying its building blocks via adding or altering different chemical groups known as **epigenetic modifications**. Without altering the DNA sequence, by, e.g., methylation of cytosine in the DNA and modification in the major chromatin proteins called histones, the cell alters the function of the gene. Structurally the cell controls the unwinding of supercoiled DNA by controlling octameric protein complexes known as nucleosomes (Bell et al., 2011) that act as a barrier for nuclear events like meiotic recombination, DNA repair, transcription, and replication. The histone proteins in nucleosomes can undergo editable changes like phosphorylation or permanent modifications like methylation results in selective unwinding, or more packed conformations of the proteins result in complete inaccessibility or partial accessibility of the DNA by the transcription factors (Rossetto et al., 2012). The DNA itself can undergo chemical modifications like methylation of cytosines at CpG islands (Moore et al., 2013).

The next level in gene regulation is **transcriptional regulation**. Transcription is the next critical step in gene regulation where the genetic information in the DNA gets

transcribed into RNA that carries the genetic message from the nucleus to the cytoplasm. It is essential to control where and how much RNA has to be made upon cellular needs. Transcription by RNA polymerases can be regulated by a set of cis and trans-acting factors. Although there is variability in nomenclature and their presence and absence between prokaryotes and eukaryotes, generally cis-acting factors like promoters, operators, and enhancers are DNA sequences that control transcription by proteins that bind to them. For example, prokaryotic repressors bind to the operator region and control the rate of RNA polymerase binding. Eukaryotic general transcription factors (GTFs) bind to the promoter region of the gene, recruit and help in releasing the polymerase for transcription. Activators can enhance this interaction between promoter and transcription factors. Enhancers are generally recognized by another class of activators. Bringing the activators close to the promoters by looping facilitates their interaction with GTFs. Finally, silencers are another important cis-acting region that controls transcription and when bound by the corresponding transcription factors, changes conformations and silences expression of the genes (Dimas et al., 2009).

During **post-transcriptional gene regulation**, the stability and distribution of the transcripts can be controlled on a variety of levels. Via 5’ end capping the cells protect the mRNA from 5’ exonuclease mediated degradation, and the cap also helps in ribosomal assembly during translation initiation. The 5’-m7G-cap also acts as a signal (Ramanathan et al., 2016) for mRNA export from the nucleus to the cytoplasm through the nuclear pore complex (Ramanathan et al., 2016). During splicing the cell removes noncoding sequences (introns) from the precursor mRNAs (pre-mRNAs) and converts the pre-mRNAs to mRNAs by ligating coding sequences (exons) via the spliceosome machinery, a multi-megadalton ribonucleoprotein (RNP) complex (Lee and Rio, 2015). For some pre-mRNAs, the different exons of a single gene can be spliced out in various combinations and for most of the pre-mRNAs this alternatively spliced out phenomenon gives rise to different isoforms or diversity of the same gene (Nilsen and Graveley, 2010).

After splicing, the 3’ tail of the mRNA undergoes **polyadenylation**, which is the addition of multiple adenosine moieties at a newly formed mRNA end, to protect the mRNA from degradation mediated by 3’exonucleases. Proteins that bind to the polyadenylation (poly(A)) tail like poly(A)-binding protein (PABP), also encourage translation initiation by mediating the interaction between the eukaryotic initiation factors eIF4E and eIF4G. **RNA editing** is another mode of gene regulation where enzymes such as adenosine deaminase acting on RNA (ADAR) or cytosine deaminase acting on RNA (CDAR) (collectively known as APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) protein family) act on adenine or cytosine residues and convert them by hydrolytic deamination in the case of adenosine to

inosine (A>I) (read as guanosine by the translation machinery), or in the case of cytosine to uracil (C>U). There are more than 50,000 editing sites have been identified in the human transcriptome (Wang et al., 2013). This mechanism gives rise to multiple transcript variants of the same mRNA.

microRNAs (miRNAs) serve as additional regulators of posttranscriptional gene regulation. They are a class of small (~22-nt long), non-protein coding, regulatory RNAs that regulate gene expression of more than ~60% of all human genes (Friedman et al., 2009). miRNAs mostly target 3'UTRs of mRNAs through imperfect base pairing with miRNA regulatory elements (MRE). 2-8 nucleotides at the 5'end of the miRNAs are usually the perfect complement to the target mRNA referred to as seed sequences. Via recruiting argonaut proteins, these interactions can lead to cleavage of the mRNAs, destabilization of mRNAs through shortening of the poly(A) tail, or influence ribosomal binding on the mRNA resulting in less efficient translation (Wilson and Doudna, 2013). Though it is known that the target site for miRNAs is in the 3'UTR of the gene, they can also target the 5'UTR or the coding sequence (CDS) region (Bartel, 2009).

The next step of gene regulation can be found on the level of mRNA **translation**, which is often combined with **mRNA transport** and **mRNA localization**. Translation controls the protein levels in the cell. Most of the mRNAs in eukaryotic cells have a long half-life (>2hr) (Raghavan et al., 2002). The abundance of proteins is therefore controlled by translational efficiency and degradation of the protein. The stability of the mRNA and, thus, its translation is, among others, controlled by the length of the poly(A) tail (Park et al., 2016). Usually, highly expressed mRNAs have a shorter poly(A) tail and decay rapidly, whereas low expressed mRNAs have a longer poly(A) tail which gives stability to the mRNAs (Lima et al., 2017). By mRNA localization, the cell can generate high levels of protein at the site of the transcript's localization or can restrict protein production to a particular region of the cell where it may not have deleterious effects. At the site of localization, activation of translation of the corresponding mRNA occurs mostly by releasing proteins that have blocked ribosomes during translation initiation stage while the mRNA has undergone transport (Lipshitz and Smibert, 2000).

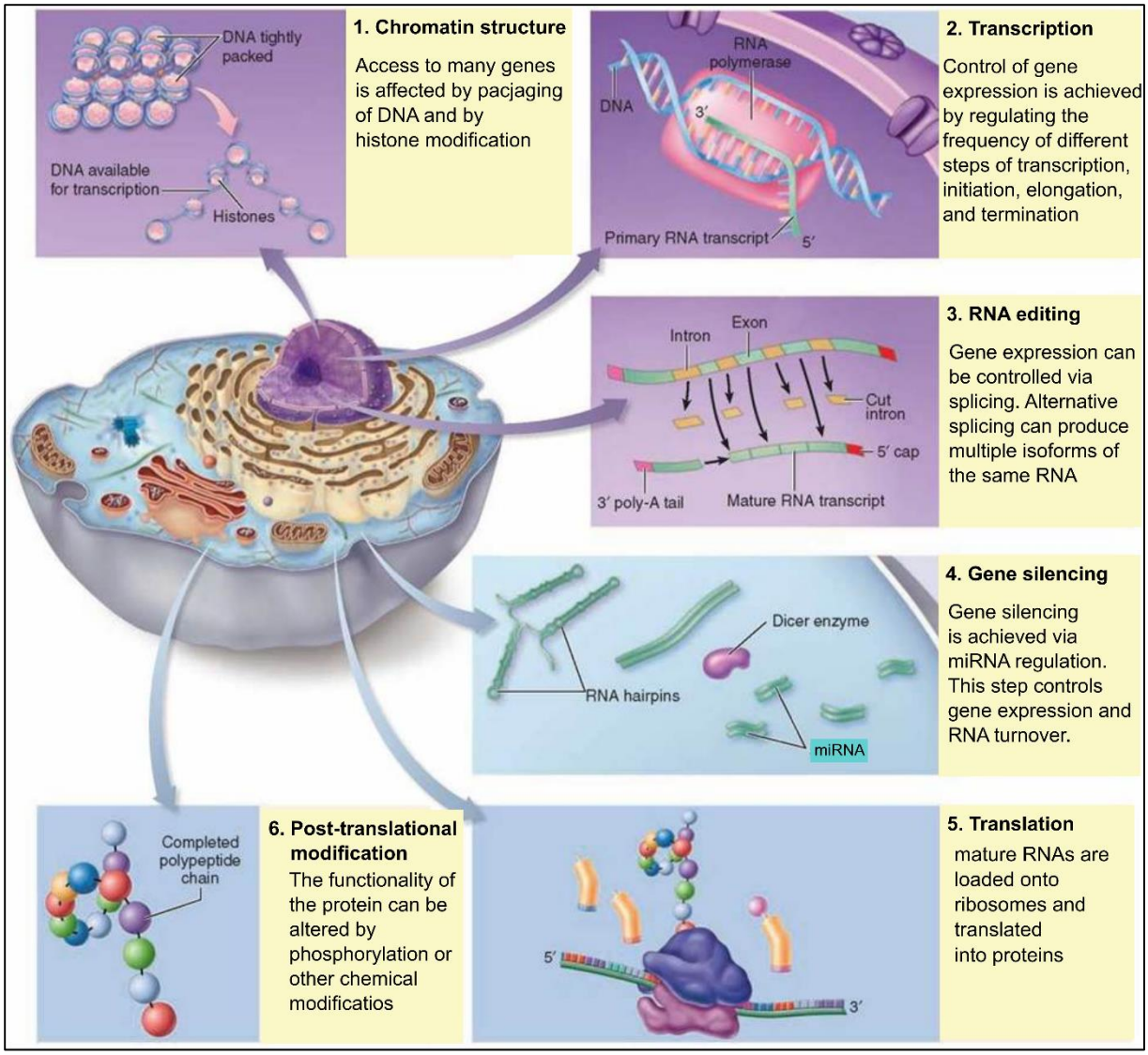


Fig 1. The six major regulators of gene regulation.

The above is the schematic representation of gene regulation describing the six major basic steps of gene regulation. The image is self-explanatory; however, in brief, **1.** after DNA replication, genes are packaged into chromatin. **2.** Then the genes in the chromatin can be transcribed into RNA, and **3.** pre-mRNA gets spliced into functional protein-coding mRNA. **4.** and **5.** RNA translation into proteins can be controlled by miRNAs, which helps in expression and turnover of RNAs. Spliced and mature RNAs are loaded onto ribosomes and become translated into proteins. **6.** The translated proteins then undergo post-translational chemical modifications like phosphorylation, SUMOylating, for their proper function. The picture was modified from https://schoolbag.info/biology.

3.1.1 Chromatin modification

The DNA (more than 2 meters long consists of approximately 3.2×10^9 nucleotides) in the human genome is distributed over 24 different chromosomes. Each chromosome is made of DNA molecules associated in a thread like structures with proteins and packed into a compact structure known as chromatin. In 1974, Kornberg and group proposed that chromatin, a DNA scaffold, is made up of repeating units of nucleosomes that contain two of each copies of four different core histones (histones H2A, H2B, H3, and H4) and, wrapped around the histone core approximately 1.7 turns of DNA or 146 bp of DNA (Kornberg, 1974). An additional histone H1 is recruited to form the 30 nm fiber by acting as a stabilizer via bridging between nucleosomes and biding to the spacer DNA not wrapped around the histone core. The nucleosome assembles in a stepwise manner. First, dimers of H3 and H4 are loaded together as a tetramer on a ~80 bp long DNA. In the next step individual dimers of H2A and H2B incorporate and form the complete nucleosome particle with DNA coiled around this octameric core, forming the so-called "beads on a string" structure. In this assembly, the multiple contacts between the histones and the DNA around the entire length of the DNA bring stability to the structure. Only the N-terminal tails of the histones protrude out of the core particle and are exposed to numerous modifications that can regulate the structure and function of the entire chromatin during DNA replication, DNA damage repair, and gene expression (Shilatifard, 2006).

The N-terminal tails of the histones can undergo different post-transcriptional modifications (PTMs), like methylation, phosphorylation, ADPribosylation, acetylation, succinylation, butyrylation, ubiquitylation and sumoylation (Huang et al., 2014). These modifications can act in the case of acetylation by neutralizing the positive charge of lysine side chains, which weakens the binding between the negatively charged DNA and histones, allowing the access of the regulatory proteins on the DNA. Other sites of modifications can be recognized by various sets of proteins (more than 100 so far) known as ‘readers,’ that influence access of transcription factors or compacting of chromatin.

3.1.2 Gene regulation during transcription

Transcription is an essential mechanism in all living organisms where the DNA molecule makes a sister molecule called RNA by copying the same information of the parental DNA molecule but with altered single-stranded nucleotide bases.

During transcription, the gene regulation occurs either via the ubiquitous DNA sequences at the promoters, close to the transcription start sites (TSS) encodes

information to recruit RNA polymerases and control the transcription initiation or by specific cis-acting features on the DNA which could be distal than the genes (operators, enhancers, silencers) but boost or downregulate the expression under right conditions (Rach et al., 2011).

Transcription requires the help of multimeric RNA polymerase enzymes. In eukaryotes, there are three major RNA polymerases. RNA polymerase I (RNA pol I) is a 590 kDa enzyme consisting of 14 different protein subunits. It is responsible for transcribing ribosomal RNA (rRNA) precursors. RNA pol II is a 550 kDa enzyme of 10-12 different subunits and responsible for transcription of messenger RNAs (mRNAs), small nuclear and nucleolar RNAs (snRNAs and snoRNAs), miRNA precursors, large non-coding RNAs (ncRNAs), and cryptic unstable transcripts (CUTs). RNA pol III is the largest polymerase complex with a molecular weight of ~700 kDa and composed of 17 different protein subunits. It helps in transcribing 5S rRNA, tRNA (transfer RNA) and U6 spliceosomal snRNA. In addition to these three enzymes, there are two plant-specific pol II versions known as RNA pol IV, and RNA pol V. The RNA pol IV enzyme synthesizes small interfering RNA (siRNA) in plants, and pol V is responsible for a siRNA-directed DNA methylation pathway (Haag and Pikaard, 2011). A complete transcription cycle for each of these polymerases involves three steps, initiation, elongation, and termination (Engel et al., 2018).

Eukaryotic **transcription initiation** of RNA pol II transcribed genes starts by assembling a pre-initiation complex (a complex of around 100 proteins) at the transcription start sites (TSS) at the 5’ site of the gene, embedded within the promoter region. The promoter region usually extends from 50 bp upstream to 50 bp downstream of the TSS. The preinitiation complex consists of the six general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH as well as additional regulatory complexes (like mediators, coactivators, chromatin remodeling complexes) (Luse, 2013). Recruitment of RNA pol II during transcription initiation is the most rate-limiting step. The activity of the core promoters can be regulated by enhancers and additionally controlled by chromatin modifications (Shlyueva et al., 2014; Zabidi and Stark, 2016). Recent genome-wide studies concluded that many regions in the genome outside the annotated gene could influence the start of transcription.

In the next step of transcription, **elongation**, promoter clearance of RNA pol II results in the removal of some components of the PIC, while others remain bound to the transcription start sites to be used as a scaffold for the next round of initiation (Yudkovsky et al., 2000). Promoter proximal pausing of the elongating RNA Pol II is another important step during transcription elongation. Pol II pauses after transcribing 25 - 50 nucleotides and this process are established by the DSIF (DRB sensitivity

inducing factor), and NELF (negative elongation factor) protein and the pausing is released by P-TEFb (positive transcription elongation factor b) (Wenzel et al., 2008). For example, for the heat-shock protein (HSPs) coding genes and 'immediate early genes' like c-myc, c-fos, or c-jun, pol II pausing helps to poise for the rapid induction of transcription as the expression of these genes needs to respond quickly to environmental stimuli like stress (Galbraith et al., 2013; Plet et al., 1995). The release of paused pol II may also be controlled via histone modifications. For example, serum induction promotes phosphorylation of serine 10 of histone H3 tail in the fos1I gene, which is then recognized by an adaptor protein that finally recruits P-TEFb to release Pol II from the paused transcripts (Zippo et al., 2009).

The **termination** step of transcription is a critical regulatory stage for the RNA, not only because the termination is coupled with polymerase recycling and retention, but also with RNA 3'end processing followed by polyadenylation. The three different polymerases use three different termination signals. For example, for mRNAs, an intact poly(A) signal is enough to terminate the pol II-mediated transcription (Park et al., 2004). Failure of transcription may lead to a read-through transcript formation with the downstream gene which cannot be capped as already being part of a polycistronic transcript, this error also known as transcriptional interference (Proudfoot, 2016; Shearwin et al., 2005).

3.1.3 Post-transcriptional gene regulation

The pre-mRNA undergoes co-transcriptional processing at different levels. In the nucleus, the co-transcriptional processing includes 5' capping (Galloway and Cowling, 2019), removal of intronic regions by splicing (Lee and Rio, 2015),
followed by 3'UTR polyadenylation or alternative polyadenylation (Tian and Manley, 2013), and nucleotide modifications (Fig. 2). After nuclear export, the RNA can be localized and/or translated. To control the spatiotemporal regulation in this multistep process, cells use specialized RBPs that can shuttle between the nucleus and the cytoplasm also participate in almost every step of RNA regulation like splicing, transcription elongation also during translation (example SR proteins). Cells also regulate their mRNAs post-transcriptionally via chemical modifications within specific nucleobases on the mRNAs, which leads to secondary and tertiary structural changes of the mRNA and can result in binding a new set of RBPS (Lewis et al., 2017).

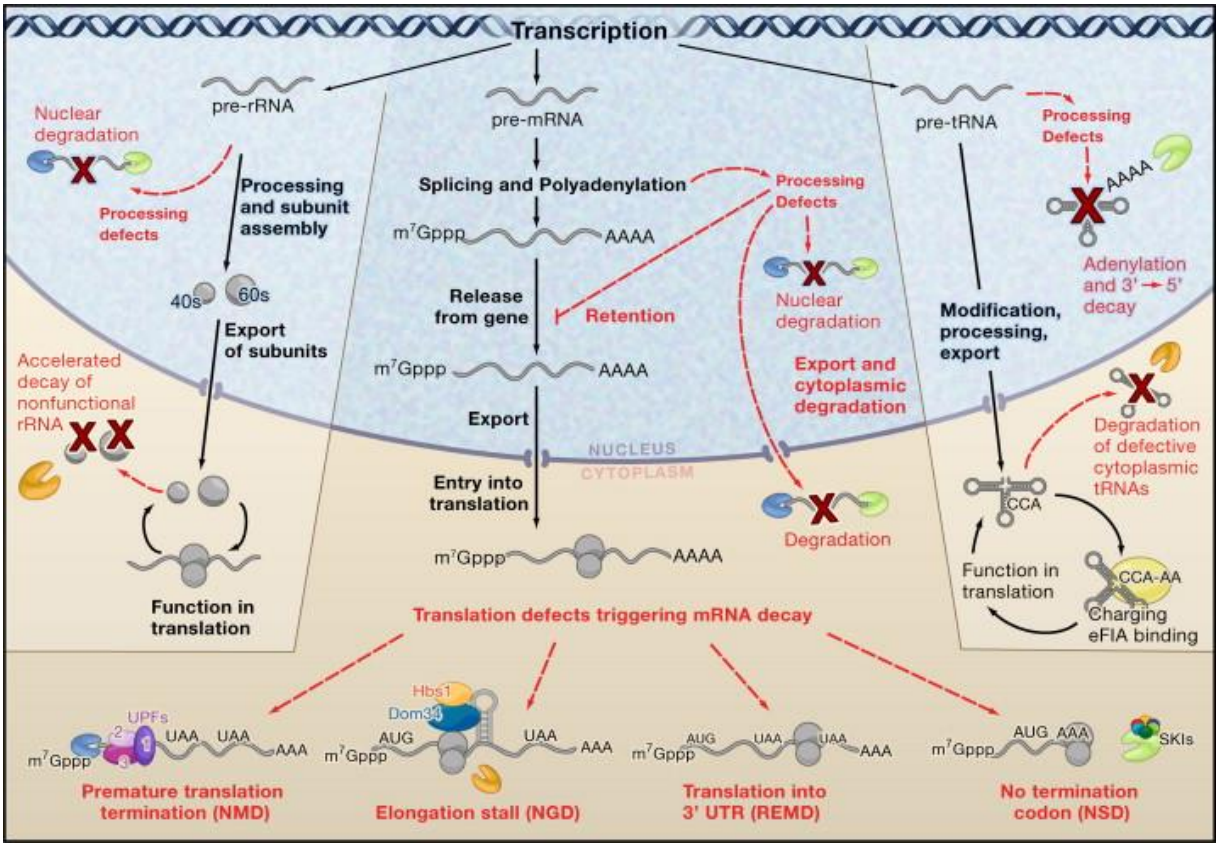


Fig 2: Post-transcriptional mechanisms of RNA function and turnover:

The picture summarizes the different steps of the fate and functions of three different major RNAs, rRNA, tRNA, and mRNA. The rRNA and tRNA were synthesized in the nucleus, and defects during their processing lead to their nuclear degradation. The tRNAs gets modified in the nucleus and the rRNAs assembled into ribosomal subunits, and both are exported to the cytosol and participate in mRNA translation. Pre-mRNAs become spliced and modified by forming an m7G-cap at the 5' end and polyadenylated at the 3' end. They are also exported to the cytosol and loaded on the ribosomes for translation. Translational defects and turnover of the mRNAs subjected to degradation are indicated in the picture and described in the text. Nonfunctional or defective rRNAs and tRNAs can also be degraded in the cytosol. Taken from (Doma and Parker, 2007), with permission from the authors.

3.1.3.1 Capping

Capping is a co-transcriptional event and the first post-transcriptional modification. After synthesis of the first 20 nucleotides by RNA pol II, hydrolysis occurs at the 5' terminal by removal of the γ- phosphate by an RNA triphosphatase to generate 5′ diphosphate RNA. Then RNA guanyltransferase transfers guanosine monophosphate (GMP) from guanosine triphosphate (GTP) to the modified diphosphate terminus, and finally, the guanosine residue is methylated via an RNA methyltransferase (Inose et al., 2015). These three enzymes are collectively known as capping enzyme. Further

methylation at the ribose can convert the so-called cap 0 structure to cap 1 and cap 2 forms. Two major protein complexes have been identified to interact with the cap structure: (i) eukaryotic translation initiation factor 4E (eIF4E), which interacts with the cap in the cytoplasm and ii. Nuclear cap-binding protein complex (nCBP), which consists of CBP20 and CBP80 (Topisirovic et al., 2011). CBP20 binds to the cap via its RNP domain and stabilizes the interaction between CBP20's N-terminal tail with CBP80 (Mazza et al., 2001). The CBP80 subunit interacts with RNA pol II's c-terminal domain and a component of the transcription-export complex (TREX) known as RNA and export factor binding protein/Aly protein (Cheng et al., 2006; Lejeune et al., 2002; Zhou et al., 2000). CBP80 assembles on the transcript simultaneously with capping and also orchestrates processes such as binding of the spliceosome for pre-mRNA splicing, 3'end processing, nuclear export, miRNA biogenesis, and nonsense-mediated decay (Hocine et al., 2010)(Fig. 3).

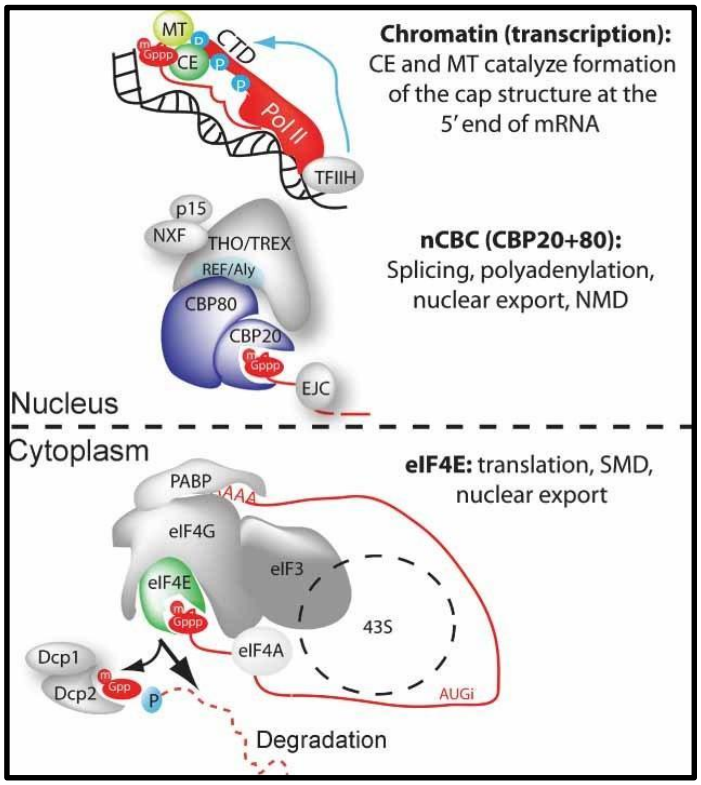


Fig. 3 Schematic representation of cap-dependent processes governed by cap-binding proteins.

Capping occurs co-transcriptionally where capping enzyme (CE) consisting of the RNA triphosphatase, guanylyltransferase, and RNA (guanine-N7-) methyltransferase (MT) are recruited to nascent transcripts through interaction with the TFIIH-phosphorylated C-terminal domain (CTD) of polymerase II (Pol II). After the cap formation, nuclear cap-binding complex (nCBC) consisting of cap-binding proteins CBP 20 and 80 binds the mRNA and with other protein complexes mediates its effects on the subsequent steps of mRNA metabolism. After the mRNA is exported from the nucleus, eIF4E binds the cap and recruits it to the small ribosomal subunit. In addition, eIF4E was suggested to export a subset of mRNAs from the

nucleus and to play a role in Staufen-mediated decay (SMD). Finally, the cap is removed by decapping enzyme (Dcp1 and 2), after which mRNA is rapidly degraded. Taken from (Topisirovic et al., 2011) with permission by the authors.

3.1.3.2 Splicing

In 1978, Walter Gilbert first time discovered exon shuffling by proposing the existence of introns as 'junk DNA' (Gilbert, 1978). After the human genome project in 2000, the hypothesis about the splicing mechanism was proposed by looking at the difference between the total number of protein-coding genes in human (~25,000) compared to the different proteins that are found in the cell (>90,000) (Lander et al., 2001).
Later on, it was identified that more than 95% of human genes undergo splicing events, and it is necessary for tissue development. The splicing events are broadly classified into two different major classes.

(i) Constitutive Splicing is a highly dynamic, mainly co-transcriptional nuclear event where via a multistep catalytic process the spliceosome machinery removes introns (non-protein coding regions) from the pre-mRNA and joins exons (protein-coding regions) in the same order that they appear in the gene, to form mature mRNA.
(ii) Alternative or differential splicing is a mechanism where certain exons can be skipped either via exon shuffling or exonization of transposable elements or constitutively spliced exons, resulting in various forms of mRNAs from the same pre-mRNA (Lee and Rio, 2015; Zheng et al., 2005). Alternative splicing increases the coding potential for a cell and greatly increases proteome complexity.

Spliceosomal splicing involves the formation of macromolecular ribozyme machinery, the spliceosome, which the phosphodiester transfer reactions. The spliceosome is composed of over 100 proteins and five small snRNAs (U1, U2, U4, U5, and U6). About 50 spliceosomal proteins are stably bound to the snRNAs, forming snRNPs (small nucleolar ribonucleoproteins). The sequential assembly of these snRNPs on the consensus splicing signals results in the specific spliceosomal complexes (E, A, B and others) together with eight conserved DExD/H-type RNA dependent ATPase/helicases (Fig. 4) (Kastner et al., 2019). For 99% of the pre-mRNAs, the beginning of the intron is determined by the consensus dinucleotide GU sequence (splice donors or 5' splice site (5'SS)), the end of the intron is determined by a polypyrimidine tract followed by an AG at the end (splice acceptor or 3' splice site (3'SS)) 5' of the string of pyrimidines lies an adenosine that determines the branch point (Kastner et al., 2019).

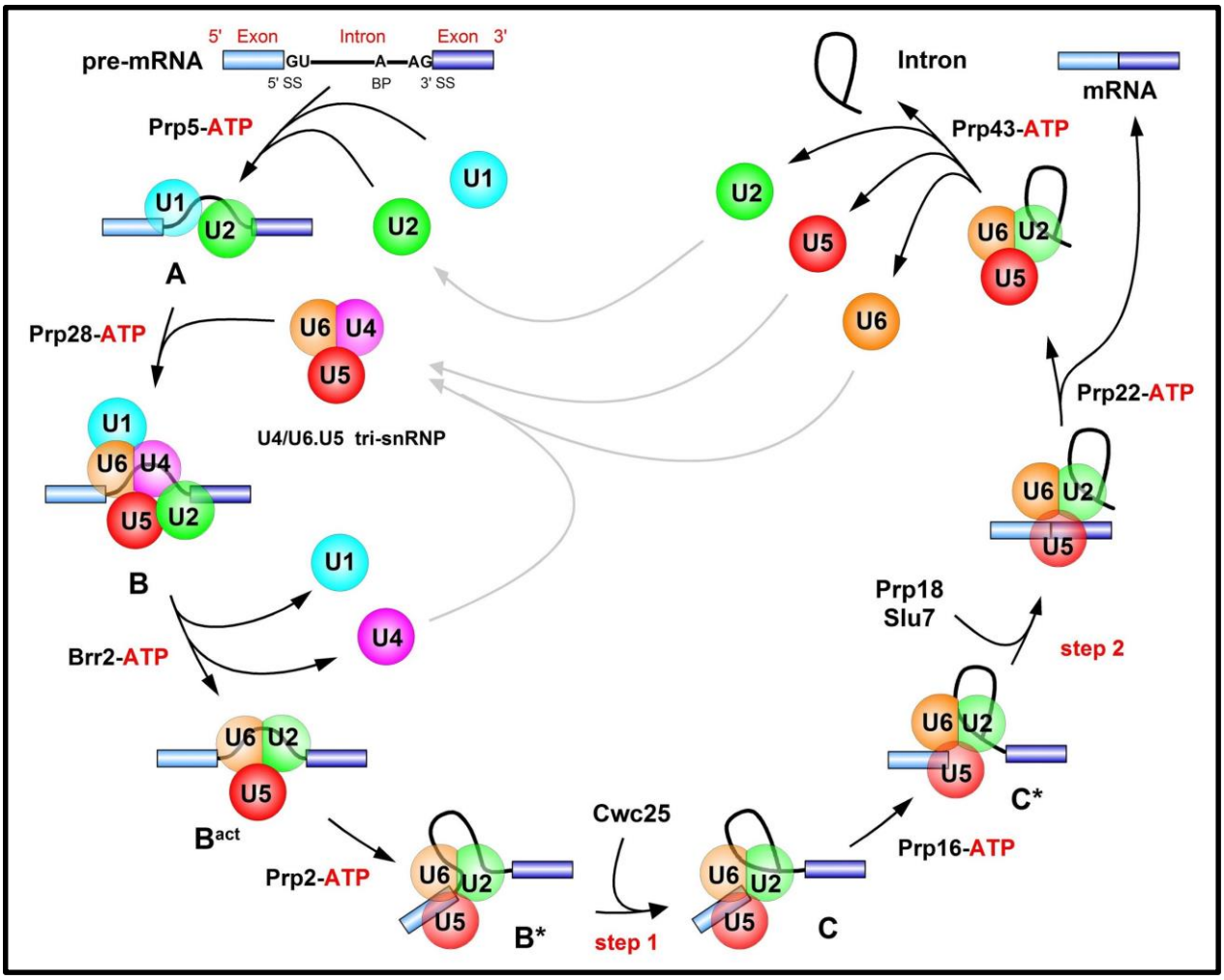


Fig. 4. Assembly of the spliceosome by the stepwise binding of the snRNPs to the pre-mRNA.

In the first stage of spliceosome assembly, the U1 snRNP binds to the 5' splice site (5' SS), and the U2 snRNP binds to the branch point (BP: close to the 3' end of the intron). This spliceosome assembly intermediate is known as A complex. Next, the binding of the U4/U6, U5 tri-snRNP complex gives rise to the pre-catalytic B complex. The catalytic activation of the spliceosome is a two-step process. At first, the RNA helicase Brr2 acts to produce the Bact complex, and in the second, the RNA helicase Prp2 facilitates the formation of the B* complex. This follows the recruitment of the protein Cwc25; which is also the first step of splicing. In this step, the phosphodiester bond at the 5' splice site is cleaved and, at the same time, the 5' end of the intron becomes linked to the 2' hydroxyl group of adenosines at the branch point. In the next step, the RNA helicase Prp16 converts the spliceosome to the C* complex, which – with the help of the proteins Prp18 and Slu7 – carries out the second catalytic step of the splicing reaction. In this step, the phosphodiester bond at the 3' splice site (3' SS: where the intron ends, and exon 2 begins) is cleaved and at the same time, the two exons are joined to one another. The intron is released from the spliceosomal complex in the form of a lasso (lariat), and the snRNPs are recycled for subsequent rounds of splicing. The dissociation phase of the spliceosome requires catalysis by the RNA helicases Prp22 and 43. "ATP" indicates the steps that require ATP molecules as a source of chemical energy. Taken from (Kastner et al., 2019) with permission from the authors.

3.1.3.3 Alternative polyadenylation

At the end of transcription, the pol II bound nascent mRNAs are cleaved and polyadenylated at the 3'end. Up to ~250 adenosines in mammalian cells and ~50 A's in yeast can be added (Zheng and Tian, 2014) by the enzyme poly(A) polymerase (PAP). The whole adenylation machinery consists of five major several protein complexes, for example, cleavage and polyadenylation specific factor (CPSF) and cleavage stimulation factor (CstF) complexes. Several other proteins are also involved in this mechanism like nuclear poly(A) binding protein 1 (PABP1) (Braunschweig et al., 2013). Many genes have more than one site leading to cleavage and polyadenylation. These sites are known as alternative polyadenylation sites (APA). Apart from bringing maturation to the nascent transcript, APA can regulate gene expression on multiple levels, including nuclear export, mRNA stability, localization of the mRNA, and translation efficiency. Via dynamic regulation, the length of the poly(A) tail plays an important role in early development and neuronal function. Most importantly APA generates alternative isoforms of the same transcript. If the PAS (aka polyadenylation site; for example, AAUAAA) is located in an internal intron or exon region, then APA can generate functional proteins with distinct C-termini from the same gene. In contrast, if the alternative pA sites are all in the 3'UTR, then APA generates multiple isoforms of the same gene with the same protein-coding features but different UTRs (UTR-APA aka distal APA) (Zhang, 2004). UTR-APA changes the UTR lengths of the same gene resulting in different translation efficiency and even localization of the translated protein. The longer UTR can have more miRNA binding sites which could influence the half-life of that transcript. As the longer isoform can have a more complex secondary structure could act as a scaffold of the interaction of other RBPs, influencing the localization of the transcript and the translation site (Berkovits and Mayr, 2015; Elkon et al., 2013).

3.1.3.4 Nuclear export

Nuclear export is essential to translocate mRNP complexes through the nuclear pore complex (NPC) before releasing the mRNP in the cytoplasm for translation (Carmody and Wente, 2009). The NPC (~4,000 NPCs/nucleus in a mammalian cell) is the biggest (60- to 100-MDa) protein complex in the nuclear envelope and consists of 30 distinct proteins, which are altogether known as nucleoporins. With these 30 proteins, the NPC forms a cylindrical ring-like structure (~10 nm in diameter at rest and can be expanded up to ~25 nm) (Fig. 5A) with the central channel containing proteins with phenylalanine-glycine (FG) repeat sequences. These repeats form a dense hydrophobic meshwork that acts as a barrier limiting the exchange of soluble macronucleus between nucleus

and cytoplasm, and also acts as a docking site for the mRNP transport complex (Burns and Wente, 2014). From the side view, the NPC looks like a triple ring structure with a 120nm horizontal and 75nm vertical axes (Fig. 5B).To be able to pass through this hydrophobic environment, the mRNP particles interact with transport receptors like export factors (NXF1 and NXF 2), karyopherins, transportin, exportins and the small GTPase Ran (RAs-related Nuclear protein) (dictates the directionality of this cargo) to channel the mRNP particle as well as small RNAs like tRNAs, miRNAs, small uridine rich nuclear RNAs and small non coding RNAs to the cytoplasm. Exportins are used for some mRNAs but mainly for ribosomes, tRNAs, miRNAs, snRNAs (Katahira, 2015). One of the best model examples of the functionality of NPC shown in Fig. 5C to describe how viral RNAs use nuclear pore complex proteins as receptors to export their RNAs from nucleus to cytosol (Fig. 5C).

Nuclear export is a surveillance checkpoint for immature mRNPs. For mRNPs containing unspliced transcripts retains in the nucleus due to the absence of spliceosomal proteins like SR proteins, NXF1 proteins, which helps the cargo to tether on the FG repeat sequences. For other transcripts, (without splicing events) hnRNPs (heterogeneous nuclear ribonucleoproteins) regulate the export via nuclear pores (Reed and Hurt, 2002). In both cases, immature or defective mRNAs are retained and undergo decay mostly via the 3'-5' degradation machinery in the nucleus (Cullen, 2000).

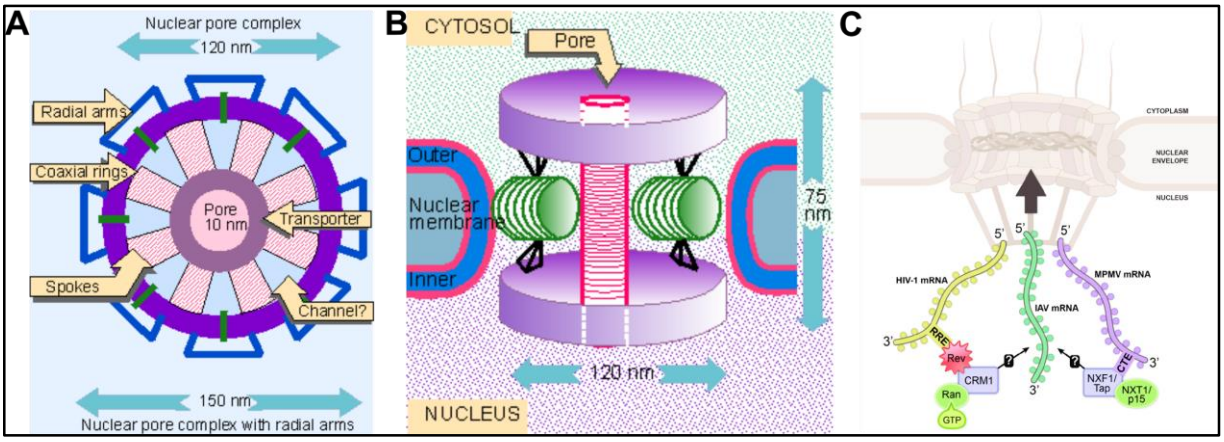


Fig 5. The basic structure of nuclear pore complex.

(A) The basic structure of the nuclear pore complex with about 10 nm of the inner ring and 120 nm of an outer ring connected by spokes. It has outer coaxial rings which contain radial arms and increase the diameter up to 150 nm. **(B)** The side view shows a triple ring architecture of the nuclear pore complex at the nuclear envelope. It shows both 120 nm horizontal and 75 nm vertical axes. **(C)** Example of the functional role of the nuclear pore complex is embedded in the nuclear envelope. In the case of the HIV RNA transport, the HIV1 Rev Protein binds to the CRM1 export receptor. The RNA-Rev-CRM1 complex binds to RanGTP and translocates through the NPC. In the case of MPMV (Mason-Pfizer monkey virus),

the mRNA export NXF1/Tap proteins heterodimerize, bind the constitutive transport element (CTE) the NXT1/p15 protein, which acts as a nuclear export receptor. In the case of Influenza A virus, it was reported to use both CRM1 and NXF1/Tap pathways. Modified from www.genes.atsoace.org and (Kuss et al., 2013).

3.1.3.5 RNA modifications

RNA modification is an evolutionarily conserved mechanism found in all types of RNAs in the cell (like, tRNAs, ncRNAs, mRNAs, rRNAs) and essential for ploidy maintenance (Geula et al., 2015), mRNA localization and translation (Meyer and Jaffrey, 2014), mRNA stability (Wang et al., 2014) and alternative polyadenylation (Ke et al., 2015). 163 different types of chemical RNA modifications have been identified so (Boccaletto et al., 2018). Amongst all the modifications (excluding the 5′ 7-methylguanosine cap and 3′ poly(A) tail), m6-methyladenosine (m^6A) is the most abundant one (80% of all RNA bases modified in eukaryotic cells (Desrosiers et al., 1974; Dominissini et al., 2012).

Modifications on the RNA require a set of enzymes. For m6A modifications, the multiprotein methyltransferase complex containing catalytic subunits (like, methyltransferase like protein 13 (METTL3), methyltransferase like protein 14 (METTL14)) and regulatory subunits (like WTAP) helps in recruiting the complex on the RNA, whereas m6A demethylase (erasers) like FTO (fat mass and obesity-associated) and ALKBH5 (α-ketoglutarate-dependent dioxygenase alkB homolog 5) can reverse RNA modifications and RNA m6A modifications control cellular homeostasis. RNA binding proteins (for example human antigen R (HuR), hnRNPs, YTH domain-containing proteins) known as ‘reader’ proteins selectively bind to the RNA with m6A and participate in functions like promoting translation RNA degradation (Yue et al., 2015; Zhang et al., 2019). Interestingly the site for modifications are mostly at the 3’UTR, close to the stop codon and target scan predicted miRNA-binding sites (Meyer et al., 2012).

One of the first post-transcriptional modifications discovered was Pseudouridine (Ψ) formation in RNA. A base specific isomerization which is catalyzed by cellular Ψ synthetase to modify uridine to Ψ. So far, Ψ has been depicted as the most abundant and highly conserved RNA modification present in a wide range of RNAs in the cell. Ψ formation is mostly found in tRNAs and rRNAs; however, initial studies showed its important function on mRNAs as well (Summarized in Fig. 6).

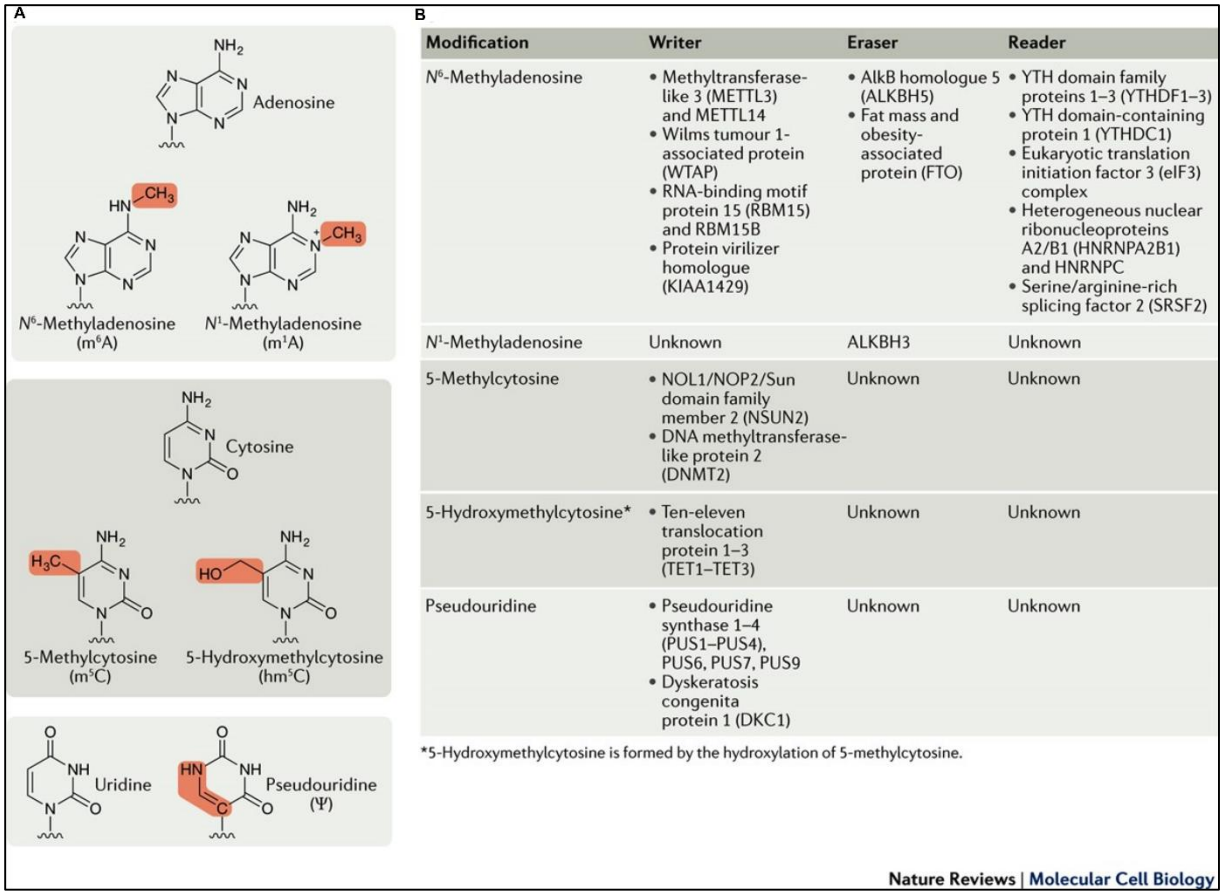


Fig 6: RNA modifications and their modifier proteins:

(A) Chemical structures of unmodified RNA bases (top) and the modified structures (below). **(B)** Table of writers, readers, and erasers of RNA modifications. Taken from (Lewis et al., 2017) with permission from the authors.

3.1.3.6 RNA Editing

RNA modification is a process by which the cell can change the chemical composition of the RNA molecules (mRNAs, tRNAs, ncRNAs, rRNAs) via RNA editing. Editing occurs by creating precise mutations on the RNA level and thus alter their function or stability. It can also lead to an alteration of the encoded protein. Editing occurs via two classes of double-stranded RNA specific editing enzymes: (i) ADAR (adenosine deaminase acting on RNA) and (ii) APOBEC1 (C-to-U editing mediated by apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1).

The ADAR family of proteins consists of the three deaminases ADAR1, ADAR2, and ADAR3. Though both ADAR1 and ADAR2 expressed ubiquitously in mammalian cells,

the majority of the editing activity is performed via ADAR1 by converting A to I (adenosine to inosine), which is then recognized as guanosine by the translation machinery (Nishikura, 2010). Editing is, for the diversification of the protein function by changing the sequences in the mRNA level. Editing regulates the stability of the RNA structure as well. A:U to I:U creates a wobble base pair which destabilizes the RNA structure, whereas editing of mismatch A:C to I:C results in Watson-Crick base pair and provide stability to the RNA secondary structure (Kung et al., 2018; Wang et al., 2013).

3.1.3.6 mRNA degradation and quality control

mRNA decay or degradation is key to regulate mRNA turnover and thus, protein production. It controls different aspects of life, like cell growth and differentiation or responding to environmental stimuli. In eukaryotes, the half-life of the mRNAs ranges from several minutes to days (Yu and Russell, 2001). Short-lived mRNAs with rapid decay like those encoding cytokines, cell cycle regulators, proto-oncogenes are present in low abundance but can increase in their levels by making a rapid adjustment to the transcriptional rate. General mRNA degradation is a stochastic event and depends on several cellular machineries and surveillance pathways (Houseley and Tollervey, 2009). However, several factors, like specific RBPs, can influence the likelihood of degradation of specific transcripts. How different degradation enzymes regulate at different stages of gene expression and maintain RNA quality control has been mentioned below.

During the co-transcriptional capping mechanism, failure to add proper 5’ cap or removal of the 5’cap by decapping enzyme Dcp2 and Nudix Hydrolase 16 (Nudt16) leads to degradation via 5’exonuclease Rat1/(5'-3' exoribonuclease 2) Xrn2 in the nucleus (Song et al., 2010; West et al., 2004). During transcriptional elongation step, modification of the C-terminal domain of RNA pol II from serine 5 phosphorylation to serine 2 phosphorylation promotes clustering if nardilysin1 and 3 (Nrd1-Nab3) complexes which promote transcription termination which is then removed by the 3’exonucleolytic degradation (Gudipati et al., 2012). During splicing, before removal, the intron forms a 5’-3’ lariat structure (also known as circular RNA) which is first debranched by Drb1 followed by degradation by exonucleases from both ends (Chapman and Boeke, 1991). The transcription termination is a final post-transcriptional mRNA quality control step. While the polymerase still on the elongating RNA, the inability to make 5’cap allows the 5’exonulease Rat1 to chase and ‘torpedo’ the polymerase (West et al., 2008). Failure of proper packaging before nuclear export, the mRNA undergoes degradation via different 5’ and 3’ exonucleases (Rougemaille

et al., 2008). After nuclear export, each mRNA undergoes multiple rounds of translation. During each round of translation, the poly(A) tail gets shorter by three major deadenylases known as C-C motif chemokine receptor 4 (Ccr4), chromatin assembly factor-1 (Caf1), and poly(A)-specific ribonuclease, (PARN) (Kim and Richter, 2006). During translational elongation, stalled ribosome leads to non-stop decay (NSD) or no-go decay (NGD), which acts as a major mRNA surveillance pathway (Summarized in Fig. 7).

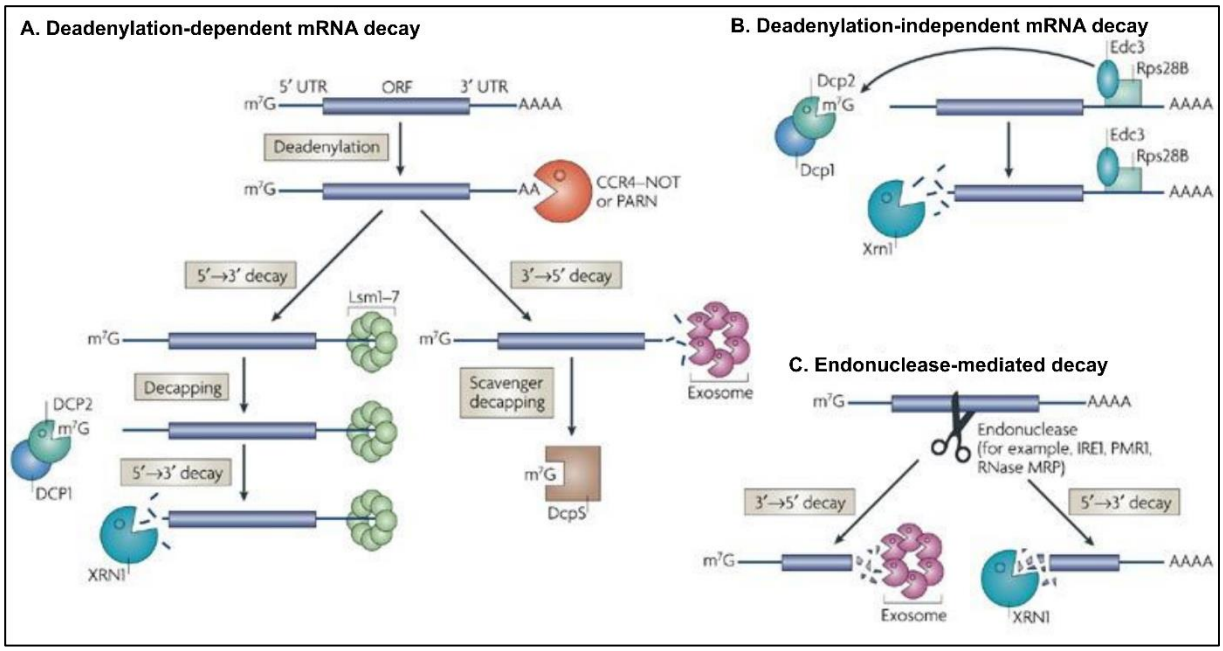


Fig 7. Summary of the major mRNA decay pathways.

(A) Description of deadenylation dependent mRNA decay. The PARN and CCR4/NOT complexes degrade the poly A tail from 3' end of the mRNA, followed by decapping using the DCP enzymes and degradation by XRN1 with the help of binding of LSM protein complex at the 3'end. In a minor pathway, deadenylation is followed by 3'-5' degradation by the exosome, leaving the cap structure only. (**B)** In the deadenylation independent mRNA decay, EDC3:RPS28B or similar proteins bind to the 3' UTR of the mRNA and induce DCP1:XRN1 proteins to remove the 5'cap and degrade the mRNA from the 5'end of the mRNA. (**C)** The third major mode of mRNA decay is mediated by endonucleases. RNA endonucleases like IRE1 or PMR1 cleave the mRNA generating two fragments with unprotected 3' or 5' end. The open 3' end fragment is degraded by the exosome and the unprotected 5' end fragment degraded by XRN1. Modified from (Garneau et al., 2007).

3.1.4 Gene regulation by mRNA localization

The first evidence of RNA localization by visualization was made in ascidian embryos. Here, in situ hybridization with DNA probes against actin (Jeffery et al., 1983)

demonstrated an uneven distribution of the mRNA 45% of the mRNA was localized in the myoplasm, 40% in the ectoplasm and 15% in the endoplasm. Starting with this simple result, extensive studies in the field of RNA localization for the last 35 years showed that the mRNA localization is present from bacteria to mammals (Bashirullah et al., 2002; Medioni et al., 2012).

mRNA localization can be advantageous for a cell or an organism.

1. It is an energy-saving process for the cells to make use of a local source or template for multiple rounds of translation from the same mRNA.

2. The local production of proteins minimizes inappropriate interaction with unwanted proteins.

3. It facilitates local protein complex assembly.

For proper targeting of mRNAs and non-coding RNAs to a cellular location, cells use different pathways. Some of the used ways are

a. Co-translational (signal recognition particle (SRP)) mediated localization at the ER: mRNAs encoding membrane and secreted proteins are targeted to the ER for local translation and post-translational modifications. This localization is mediated by recognition of signal peptide in the encoded protein during translation by SRP, an RNP complex consists of six distinct polypeptides and one small cytoplasmic 7sL-RNA. It is binding to the nascent peptide chain in the context of the translating ribosome results in an 'elongation arrest' state for the mRNA. The whole complex (SRP, ribosome, and nascent chain complexes) are transported to the protein-conducting channel of endoplasmic reticulum (ER) (ER lumen) where SRP is recognized by its cognate SRP receptor then released and the protein during re-initiated translation inserted into the ER or its membrane (Fig. 8A , see also Cui and Palazzo, 2014; Walter and Blobel, 1983).

b. SRP independent localization at ER: SRP independent mRNA localization to the surface of the ER is poorly understood. From studies done mostly in *yeast* (Mutka and Walter, 2001), it has been identified that the mRNA can associate with the ER even in the absence of ribosomes (non-translating mRNAs), and deletion of some components of the SRP pathways has surfaced a compensatory pathway to localize and translate secretory proteins at the surface of ER (Fig. 8B, See also Lerner and Nicchitta, 2006).

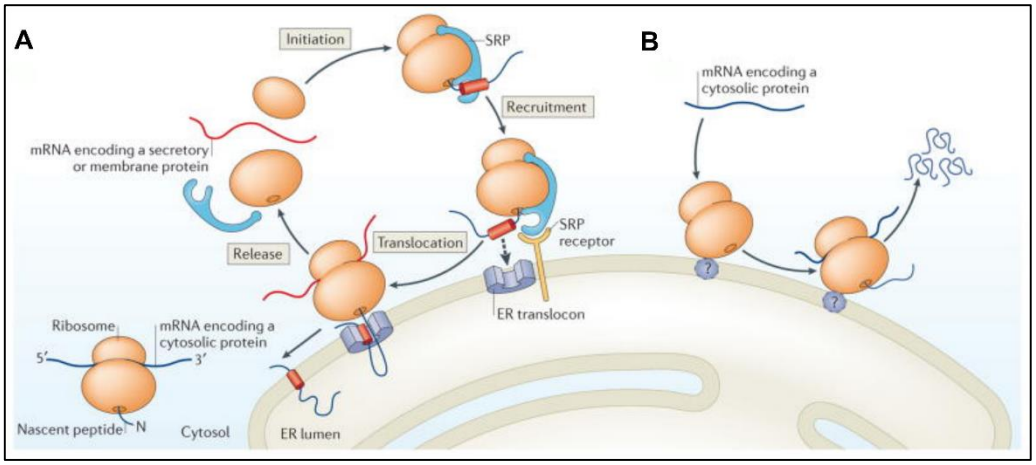


Fig 8. Pathways of RNA localization.

(A) SRP dependent localization of at the ER. Ribosomes translating mRNAs that encode a protein containing a signal peptide or transmembrane domain (such as secretory and integral membrane proteins) are targeted to the ER co-translationally by the SRP. Briefly, following translation initiation in the cytosol, the emerging topogenic signal serves as a targeting signal to the ER. Following docking on the ER translocon, secretory and membrane proteins are translocated across or into the ER membrane. Upon completion of protein synthesis, ribosomal subunits are recycled into the cytosol. **(B) SRP independent localization at the ER.** mRNAs that encode cytosolic proteins can also be translated by ER-bound ribosomes. Thus, a large fraction of the proteome can be translated by ER-associated ribosomes. Such a diverse and selective translation of mRNAs redefines this ubiquitous organelle as a primary site of proteome synthesis in the cell. Ribosomes translate mRNAs encoding cytosolic proteins associate with the ER via an unknown ribosome receptor (indicated by the question mark). Taken from (Reid and Nicchitta, 2015) with permission from the authors.

c. RNA localization via zipcodes: Most of the localized mRNAs encoding cytosolic proteins rely on the presence of cis-acting zipcodes in their sequence and transacting factors that recognize these. The trans-acting proteins perform multiple functions throughout localization, starting from translational repression of the mRNA, directed transport, anchoring, and translation induction (Bullock, 2012; Eliscovich et al., 2013). To shuttle to specific areas of the cell, localizing mRNAs move along cytoskeletal elements like microtubules or actin filaments. The active translocation of the localizing mRNAs depends on motor proteins like kinesin, myosin, and dynein families (Shav-Tal and Singer, 2005).

d. Diffusion-coupled local entrapment: In oocytes from *Drosophila* (*nanos and osk* mRNAs) or *Xenopus* (*Xcat2* mRNA) motor independent localization has been identified (Palacios, 2007). It has been postulated that the localized RNAs are organized in particles and move via diffusion throughout the cell. At the target site, they become trapped by anchoring proteins (King et al., 2005).

e. Localized protection from degradation is another less studied method to localize RNA in *Drosophila* embryos. For example, Hsp83 mRNA is uniformly distributed in early fertilized eggs, but at a later stage of development, localization is restricted to the posterior pole of germplasm. In the absence of a functional RNA degradation machinery, the mRNA is stabilized and loses its selective accumulation at the posterior pole (Ding et al., 1993).

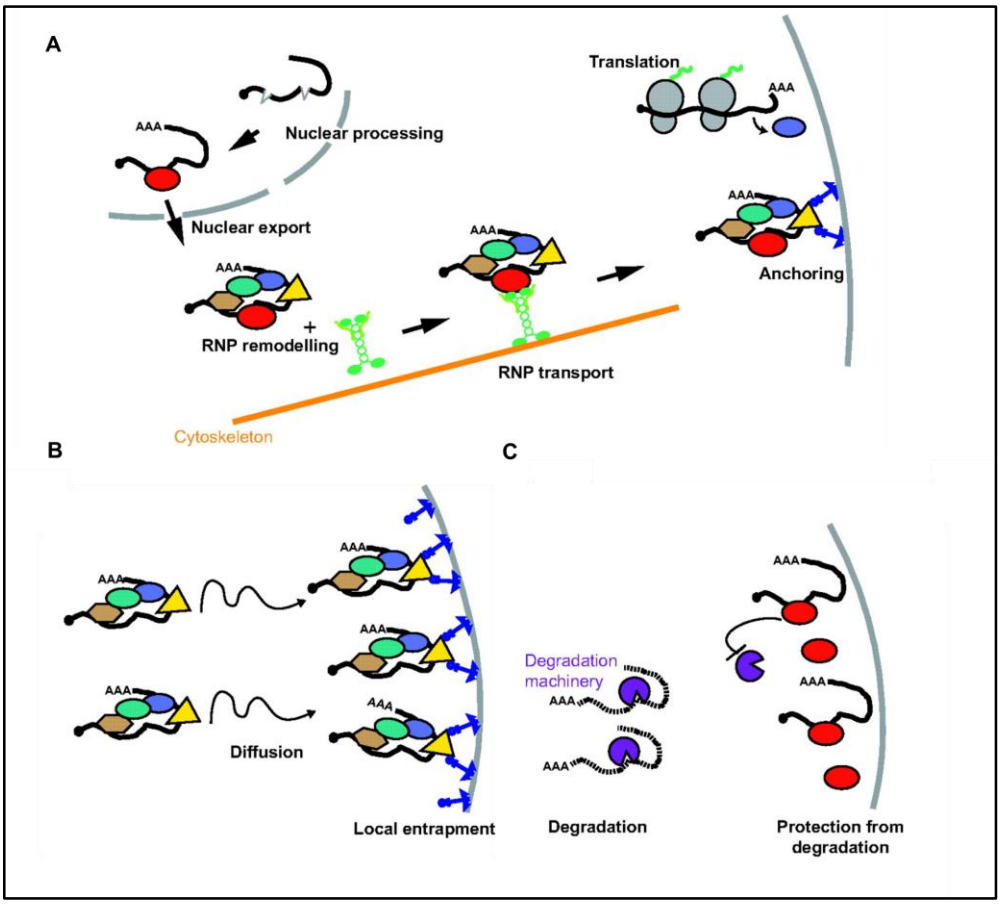


Fig 9. Pathways of RNA localization.

(A) mRNA localization via zipcodes. Zipcodes of the mRNAs destined for directional transport is recognized by specific trans-acting factors in the nucleus, where RNPs undergo different maturation steps. Upon export to the cytoplasm, RNP complexes are remodeled, and cytoplasmic factors ensuring coupling with molecular motors and transport along a polarized cytoskeleton are recruited. Once at the final destination, mRNAs are anchored, and their translation is activated. **(B) Diffusion coupled local entrapment.** mRNAs freely diffuse in the cytoplasm and are locally entrapped, at the cell cortex**. (C) Localized protection from degradation.** Non-localized mRNAs are targeted by the degradation machinery, whereas localized mRNAs are protected by yet unknown mechanisms. Modified from (Medioni et al., 2012).

3.1.4.1 Factors influencing mRNA localization

Targeting of mRNAs to a specific subcellular location involves multiple regulatory steps (Martin and Ephrussi, 2009). Interaction between the cis-acting zipcodes on the mRNAs with RBPs or accessory proteins (like motor proteins, anchoring proteins at the target site of the localized mRNAs) is important for the proper localization of the mRNA in the cell. Small regulatory RNAs play an important role in the localization as well. Binding of the regulatory RNAs influences the folding of the RNA which facilitates the association of a series of auxiliary proteins which eventually produces a ribonucleoprotein (RNP) transport particle (Kloc et al., 2002). These mRNP particles are cell type-specific for example, chromobodies are present only in the male germ cells, neuronal granules can be found in neurons and oocytes (Kloc et al., 2002). Apart from the zipcodes and the trans-acting factors, the localization can also be influenced by external cues. For example, the localization of β-actin mRNA to the neuronal growth cone is stimulated via glutamate-induced calcium signals (Yoon et al., 2016). Localization of β-actin mRNA in the protrusions of chicken and mouse primary fibroblasts can be stimulated by serum induction (Latham et al., 1994) or localization of the brain-derived neurotrophic factor (BDNF) transcript at the distinct dendritic compartment of a neuron and its local translation is important for survival and plasticity of the motor neurons (Eom et al., 2003).

3.1.4.1.1 Role of RNA binding proteins (RBPs) in mRNA localization

RBPs control different stages of post-transcriptional gene regulation like transcription, splicing, nuclear export, viral infection, silencing through RNA processing, RNA editing, and localization. The RBPs control mRNA localization by interacting with cis-acting regions at the 3’UTR of the target mRNA mostly at the 3’UTR but can also be present at the 5’UTR or CDS forming as a whole an mRNP particle (Di Liegro et al., 2014). Depending on the consequences on the mRNA fate, RBPs can associate with the RNA in the nucleus or assembled with it later at the stage of a cytoplasmic mRNP particle forming a complex for example with motor proteins like kinesin or myosin that help in localizing the mRNP cargo to its destination (Dreyfuss et al., 2002). For the specific interaction between the RBP and the mRNA, RBPs contain one or more RNA-binding domains that recognize specific sequences on the mRNA or a secondary structure, mostly a hairpin or stem-loop like structure, that are located in the cis-element in the mRNA (Jambhekar and DeRisi, 2007). The best-characterized canonical RNA binding domains of such RBPs are KH-domains, RRM (RNA recognition motif) domains, or double-stranded RNA binding motif (DSRM) domains (Clery and Allain, 2015). To date, compared to around 54,000 different proteins annotated in the smart db database, a

total of 8767 proteins contain RRM domains, 115166 contain KH domains, and 36103 DSRM motif-containing protein have been identified in eukaryotes (according to SMART db database (http://smart.embl-heidelberg.de/))

RRM (RNA recognition motif) is the largest group of single-stranded RNA binding domain-containing protein with an eight amino acid ribonucleoprotein-1 (RNP-1) consensus sequence (Bandziulis et al., 1989). RRMs are ~ 90 amino acids long domains with a four-stranded β-sheet packed against two α-helices ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta$ topology) and contains eight and six conserved amino acids sequences known as RNP1 (located in β_3 sheet) and RNP2 (located in β_1 sheet) (Valverde et al., 2008).

RRM-ligand interaction is highly dynamic. In general, the RNP1 and RNP2 sequences are the important ones as they expose three conserved aromatic residues that serve as an RNA binding surface. The RNP1 and RNP2 motifs can interact with different numbers of nucleotides like two in the case of RRM2 (Allain et al., 2000) in nucleolin to up to eight nucleotides for spliceosomal protein U2B. Based on the availability of the number of aromatic residues during interaction with the RNA, the RRMS can be subcategorized into quasi RRM (qRRM), pseudo-RRMA (ΨRRM) or U2AF homology motif (UHM). Structurally, the RRM domains, being so small and flexible, can interact with RNA, other proteins and also with its own domains, which explains the diversity of the interaction possibilities by RRM domains with other biomolecules and why this is also such a highly conserved motif during evolution (Cléry et al., 2008).

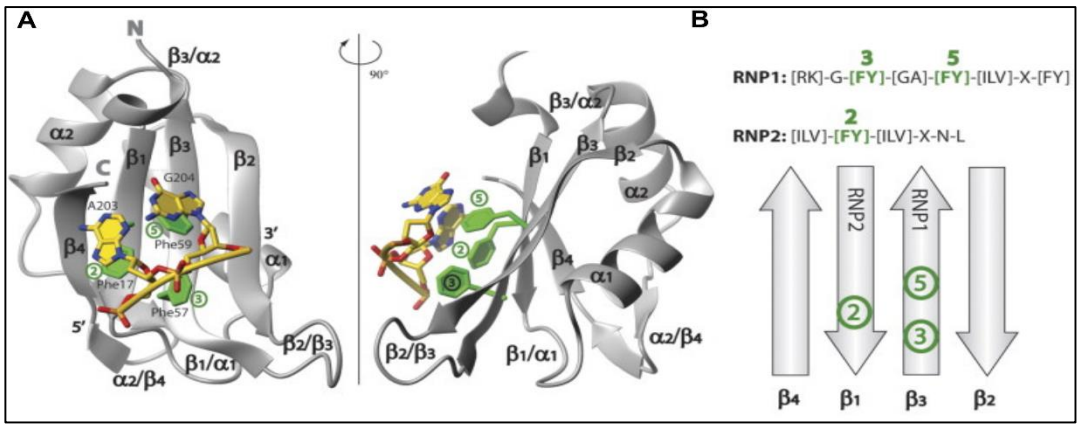


Fig 10. RRM domain structure and topology.

(A) Represents a nucleic acid binding model of an RRM (hnRNP A1 and telomeric DNA). **(B)** Shows the structural conservation of RNP1 and RNP2 and (in green) the position of conserved aromatic residues in the four-stranded beta-sheet. Taken from (Cléry et al., 2008) with permission by the authors.

The KH domain (heterogeneous nuclear ribonucleoprotein K or hnRNP K homology domain) was first identified in the human protein hnRNP K (Siomi et al., 1993). This domain is evolutionarily conserved, 70 aa long, and formed from three α-helices that are packed onto an antiparallel β-sheet. The KH domain binds to RNA or ssDNA (Grishin, 2001). Though the minimal core motif is the same (βααβ), depending on the N- and C- terminal extensions, the KH domains can be subcategorized into two groups. The type I has an additional α and β elements at the C-terminal and is found only in eukaryotes whereas the type II KH domain is only found in prokaryotes and has the additional α and β elements at the N- terminal end, leading to a completely different fold (Nicastro et al., 2015). The characteristic feature found in both KH types is the conserved G-X-X-G motif that links two α helices with the core.

Two helices (α1 and α2) together with the G-X-X-G motif and the β-sheet and its attached variable loop form a hydrophobic groove that interacts with the bases of the ssDNA or RNA molecule (Musco et al., 1996). In general, the phosphate backbone of the first two nucleotides of the RNA molecule interacts with the conserved G-X-X-G motif via H-bonding or electrostatic interaction depending on the complex (Fig. 11A, and B) (Backe et al., 2005). Deletion of (Nakel et al., 2010) or mutation (G-X-X-G to G-D-D-G) (Hollingworth et al., 2012) within this motif results in loss of RNA binding of the entire domain and increases non-specific protein-protein interactions (Oddone et al., 2007).

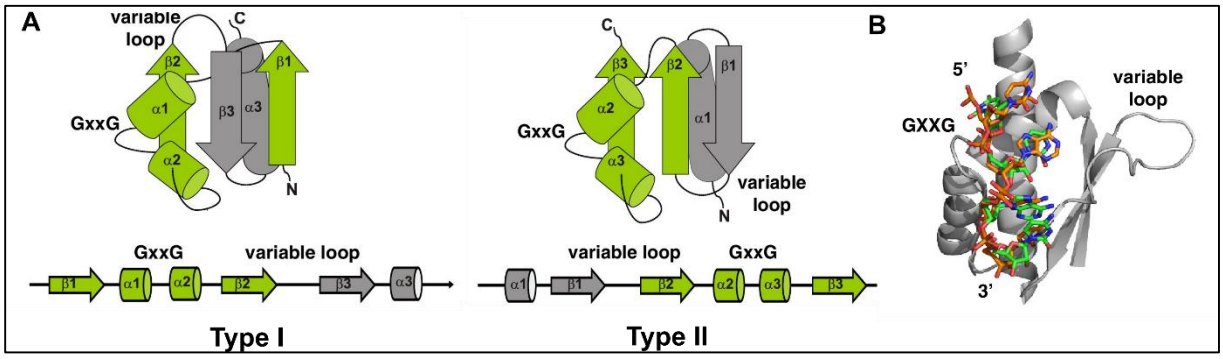


Fig 11. KH domain topology and structure.

(A) Left and middle: representative picture of the secondary structure and folding of the type I and type II KH domains with green represents the core secondary structure, and the other elements are grey. **(B)** Ribbon structure represents the Nova-2 KH3-RNA complex. The sticks represent the four nucleotides in yellow where the carbon atoms are in orange. Modified from (Nicastro et al., 2015).

The **DSRM (dsRNA binding motif)** is a 70 aa long RNA binding domain with a conserved αββββα topology (Fig. 12). It has a binding preference towards the double-stranded stretches in the RNA and also helps in shaping the secondary structure of the RNA (Stefl et al., 2005). Some of the extensively studied RBPs that bind to the double-stranded RNAs are Staufen and ADAR. The two helices in DSRMs are packed towards one surface of a three-stranded antiparallel sheet. The α1 helix and the GPxH motif in the β1β2 sheet and the intermediate loop bind to the minor groove of the dsRNA and the positively charged motif on the N-terminal of the α2 helix (KKxAK) interact with the phosphate backbone of the major groove (Banerjee and Barraud, 2014).

The DSRB is also known as a non-sequence specific RNA binding domain, which adopts the shape of the A-form of the helix via the direct interaction of the 2'OH group of the ribose sugar backbone with the phosphodiester bond of the RNA via non-bridging oxygen (SAUNDERS and BARBER, 2003). Because of this interaction pattern, the DSRM prefers stem-loops as a substrate for interaction (Banerjee and Barraud, 2014).

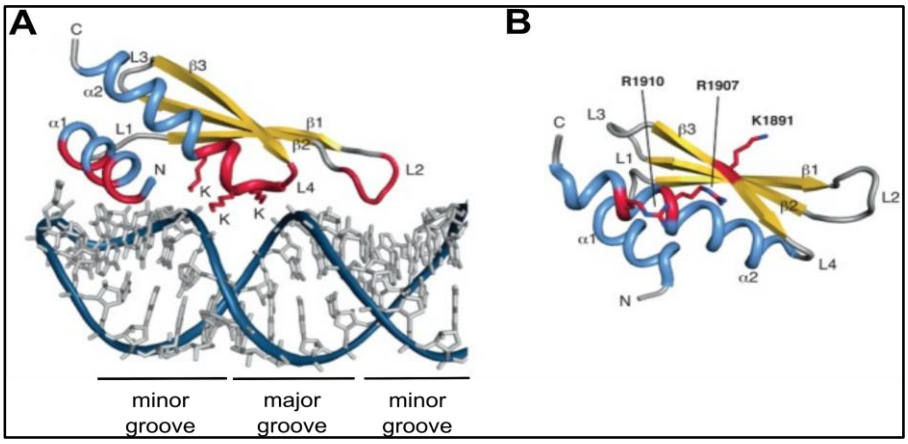


Fig 12. dsRBD domain structure.

(A) Pictorial representation of dsRNA structure and interaction motif of dsRBP ADAR2. The reds indicate three regions of the RBP interacts with the dsRNA. The a2 helix and the L2 region interacts with the minor grooves of the RNA while the three conserved lysines showed as sticks interact with the major groove. **(B)** Molecular model of the C-terminal domain of human Dicer a dsRNA binding protein with atypical dsRBD NLS composition with positively charged amino acid residues designated as red sticks in a noncanonical region of interaction with dsRNA. Taken from (Banerjee and Barraud, 2014) with permission by the authors.

3.1.4.1.2 Role of localization elements in mRNA localization

Eukaryotic 3'UTR contains multiple signatures of RNA regulation, including nuclear export, alternative polyadenylation, cytoplasmic localization, translation efficiency, and mRNA stability (Moore, 2005).Though the 3'UTR is full with regulatory regions, the motif required for asymmetric distribution of RNA, known as localization element (LE) or zipcodes could be only a primary sequence or secondary structures like stems, hairpin loops, bulges, and internal loops which could help in tethering the trans-actin factors or linker proteins (Hamilton and Davis, 2011).

Apart from the 3'UTR, localization signals for several mRNAs can also be found in the coding region, for example, the mRNA encoding the *Drosophila* synaptotagmin-like protein, yeast *ASH1* mRNA. The localization elements can be present in the 5'UTR, as well. Examples are found in the *Drosophila* yemanuclein-alpha transcript (Aït-Ahmed et al., 1992) or XNIF mRNA in *Xenopus* (Claussen, 2004). Sometimes mRNAs have multiple localization elements that are spread over the entire transcript. For example, the localization signals in *grk* mRNA in *Drosophila* are found in the coding region, 5'UTR, and 3'UTR. However, they serve different functions as the 5'UTR signal is responsible for localization during late oogenesis (Saunders and Cohen, 1999), the signal within the ORF is essential for later localization (Thio et al., 2000), and the LE in the 3'UTR is responsible for dorso-anterior localization of this mRNA (Thio et al., 2000). Identifying the zipcodes is problematic (reviewed in (Jambhekar and DeRisi, 2007)), mostly because they are highly variable in length, secondary structure, sequence, and complexity (Herve, 2006). In addition, in many cases, the mRNA localizes in a multistep process where individual zipcodes within an mRNA have to work together to achieve the final localization. Furthermore, RBPs can associate with the zipcodes in a spatiotemporal manner and can alter the structure to allow binding of the next proteins (Edelmann et al., 2017), so that binding of certain RBPs to a zipcode only occurs in the context of the assembling mRNP. Another problem is that the secondary structure of the zipcodes is also important as the binding sites for the cognate protein(s) Localization of MBP (Myelin basic protein) mRNA is an ideal example for the requirement of both conserved sequence and secondary structure of the zipcode for localization of the mRNA. In oligodendrocytes of the central nervous system, a 21nt long region at the 3'UTR of MBP mRNA is known as RNA transport sequence (RTS). This signal is also present in several other mRNAs in different cell types and might thus represent a general transport sequence. The sequence is required for the initial granular assembly in the perikaryon. An additional essential sequence at the 3'UTR of this mRNA known as RNA localization region (RLR) makes a stable secondary structure which is then localized this mRNA to the myelin compartments (Ainger et al., 1997). The cases of the localized *bicoid, K10* and *hairy* mRNAs in *Drosophila*, and

ASH1 mRNA in budding yeast also showcase the importance not only of a specific sequence of the LEs but on its secondary structure.
In other instances, for some mRNA, it has been identified that specific sequences act as LEs and is sufficient for localizing a reporter RNA to the target region. For example, in the case of β actin mRNA, a 54nt long zipcode sequence next to the stop codon is important for localization of this mRNA. Two subsequences within this 54nt long region are essential for binding of ZBP1 protein, which is essential for the localization of the mRNA to the protrusion of chicken or mouse fibroblasts. It recognizes CGGACT with its KH4 domain and ACACCC with the KH3 domain (Nicastro et al., 2017). Sometimes repetitive sequences act as LEs and are enough to localize a reporter mRNA to the target region like in the case of the localized Vg1 and VegT mRNAs in *Xenopus*, the repetition of a CAC triplet (also called E2 motif) is essential for binding of Vera/VgRBP1 (the *Xenopus* homolog of ZBP1) to the localization element.

3.1.4.2 Dissecting the behavior of β-actin mRNA in polarized cells

β- actin mRNA was for the first time shown to be localized in ascidian eggs and embryos in 1983 (Jeffery et al., 1983). However, it is localized in a large variety of cell types and organisms. It localizes to the protrusions of chicken embryonic fibroblasts (Sundell and Singer, 1991), 3T3 mouse fibroblasts (Hill et al., 1994), endothelial cells (Hill and Gunning, 1993), myoblasts (Hill and Gunning, 1993), and non- metastatic (but not metastatic) adenocarcinoma cells (Shestakova et al., 1999). In addition, it also localizes to the growth cones and dendrites of neurons.

3.1.4.2.1 β-actin mRNA localization in neurons: Role in local memory formation

β-actin mRNA is highly abundant in adult neurons (Cajigas et al., 2012). In the last decade, studies on the postsynaptic signaling pathways showed that, being the primary cytoskeletal component, actin plays a significant role in morphological plasticity, stability, formation and elimination of dendritic spines (Fischer et al., 2000; Schubert and Dotti, 2007), memory formation and learning at the neuritis of excitatory synapses (Bassell et al., 1998; Kasai et al., 2003). A significant role of β-actin mRNA can be found in growth cone motility. In response to external cues, neurites grow towards a specific direction to form neural networks. This guidance response of the growth cone is majorly coordinated by localized protein synthesis and degradation of β-actin (Bunnell et al., 2011; Ming et al., 2002). It has been shown that delocalization of the β-actin mRNA results in growth cone retraction and non-directionality of growth cone guidance (Zhang et al., 2001). The local translation of β-actin mRNA can be stimulated

by neurotransmitters at the neuronal growth cones or neurites. For example, local uncaging of glutamate at neurites by a focused laser beak showed that within 10 mins, the β-actin mRNA localizes near the activated growth cone, which was followed by its local translation (Yoon et al., 2016). Similar kinds of experiments demonstrated that β-actin mRNA travels as a single mRNA copy in transport granules that also contain ribosomes. The RNA apparently remains masked by other proteins and is released only upon external stimulation (Buxbaum et al., 2014). Actin network remodeling at the neurites is facilitated by local β-actin synthesis, which in turn helps in strengthening or weakening of synaptic connections. This plasticity is essential for learning and memory formation in the brain (Buxbaum et al., 2014).

3.1.4.2.2 β-actin mRNA localization in fibroblasts

Why does β-actin mRNA localize to protrusions?

In polarized mouse embryonic fibroblasts, migration is a well-organized cyclic process. Starting with the formation of the protrusions, focal adhesions are formed at the leading edge (also known as lamellipodium). Then the nucleus starts to move, the fibroblast detaches at the rear and finally, the whole body of the cell translocate. The localization of several proteins is crucial for the migrating behavior of these cells. The actin polymerization nucleator Arp2/3 complex is found to be localized at the lamellipodia (Insall et al., 2001; Welch et al., 1997). The myosin heavy chain X is also present at this site as it induces the formation of filopodia (Kerber and Cheney, 2011) and interacts with β-1 integrin and PIP3 (Phosphatidylinositol (3,4,5)-trisphosphate) to induce localization of these proteins at the protrusions. Another protein that is enriched at lamellipodia is actin itself. *In situ* hybridization techniques showed that β-actin mRNAs are localized at the lamellipods of crawling cells, and this localization is correlated with the localization of the β-actin protein to the apical structures like the leading edge of lamellipodia and filopodia (Hill and Gunning, 1993; Shestakova et al., 2001; Sundell and Singer, 1991). Delocalization of β-actin mRNA from the protrusions results in the suppression of cell motility in chicken embryonic fibroblasts (CEFs). The mechanism behind this affected cellular polarity and crawling depends on several interdependent events. (i) Local synthesis of β-actin mRNA helps to maintain intact cytoskeleton, which in turn drives protrusions at the lamellipodium. (ii) Local protein synthesis of β-actin mRNA augments localization and local translation of other mRNAs as well as a nucleating complex containing Arp 3 mRNA, and together they form the active filopodia (Shestakova et al., 2001).

Regulators of β-actin mRNA localization at the fibroblasts protrusions

For β-actin mRNA localization, a cis-acting 54nt long zipcode sequence at the 3'UTR, next to the stop codon is required. Binding of 68kDa protein, ZBP1or IGF2BP1, to the zipcode is essential for the localization of β-actin mRNA (Ross et al., 1997). Either deletion of the IGF2BP1 protein or mutating the zipcode results in delocalization of β-actin mRNA resulted in inhibition of β-actin protein synthesis and slower cell motility (Katz et al., 2012; Shestakova et al., 2001). IGF2BP1 interacts with the zipcode via two K-homology (KH) RNA-binding domains and is required for RNA localization in migrating cells, including fibroblasts and neurons. In addition, in fibroblasts, it controls the translation of β-actin by blocking the assembly of ribosomes at the start codon and only when the mRNA is at the protrusions, the tyrosine 396 (Tyr396) residue on IGF2BP1 phosphorylates by Src and releases β-actin mRNA for translation. In neurons, an additional mTORC2 dependent phosphorylation on serine 181 (Ser181) of IGF2BP1 is needed for proper IGF2BP1 mobility and dendritic distribution (Hüttelmaier et al., 2005; Urbanska et al., 2017). The IGF2BP1 has two RRM domains and 4 KH domains. It has shown that the RRM di-domain of IGF2BP1 interacts with KIF11 motor protein and helps to transport the RNA-protein complex along the microtubules in MDA231 cells (Song et al., 2015). On the other hand, the KH3 and KH4 domains form an intra-molecular pseudo-dimer with two RNA binding grooves of the zipcode. The KH4 domain of IGF2BP1 binds with high specificity to the CGGACT sequence of the zipcode, whereas, the KH3 recognizes a short RNA sequence with lower affinity. This sequence contains CC and CA in the central positions (Farina et al., 2003; Nicastro et al., 2017).

Apart from IGF2BP1, several other proteins have been shown directly or indirectly binds to β-actin mRNA, but their functionality is not highly characterized like IGF2BP1. For example, another isoform of IGF2BP1, IGF2BP2 binds to β-actin mRNA but how they regulate the localization of β-actin mRNA or whether it is essential for localization or not has not been identified (Wächter et al., 2013). Another protein, the receptor-activated C kinase 1 (RACK1) is a ribosomal associated protein and a substrate of Src. The Y246 amino acid of RACK1 is an Src binding and phosphorylation site and acts as a docking site on ribosome for the β-actin mRNA/IGF2BP1 complex. Phosphorylation of this Y246 site is critical to the release and translation of β-actin mRNA (Ceci et al., 2012). Immunoprecipitation followed the identification of bound RNA revealed that another nuclear RNA-binding protein Sam68 (Src-associated in mitosis, 68 kDa, mouse homolog KHDRBS1) binds to the UUUUUU sequence of β-actin mRNA. How this protein affects or regulates β-actin mRNA localization has not been described (Itoh et al., 2002). By using trimolecular fluorescence complementation (TriFC) assay, it was shown microscopically that the fragile X mental retardation protein (FMRP) isoform 18 and a human ortholog of ZBP1, IMP1 associate at the 3'UTR of β-

actin mRNA. Tethering of FMRP on β-actin mRNA recruits IMP1 and facilitates granule formation (Rackham and Brown, 2004). In another study (Dormoy-Raclet et al., 2007) it was shown that HuR (Human antigen R, aka ELAV like RNA binding protein 1(ELAV1)) stabilizes β-actin mRNA by associating with uridine rich element at the 3'UTR of β-actin mRNA. Deletion of HuR in Hela cells does not affect the nuclear or cytoplasmic distribution of the β-actin mRNA but the effects the expression of β-actin protein, which in turn affects the actin stress fiber formation in the cell.

Effect of growth factor on β-actin mRNA regulation

β-actin mRNA localizes to the leading edge of chicken or mouse fibroblasts, where actin polymerization drives the motility of the cell. In the absence of serum or growth factors, the cells enter a quiescent phase characterized by a decrease in the β-actin protein level. Less protein expression results in a reduction in focal adhesions, decrease actin stress fibers, and an overall decrease in actin polymerization in the cellular periphery (Ridley and Hall, 1992). At this stage, the β-actin mRNAs remains in a diffuse and nonlocalized stage. Addition of growth factors like serum or chemotactic factors (like Platelet-derived growth factor (PDGF)) or lysophosphatidic acid results in a rapid redistribution of β-actin mRNA in the leading edges (Latham et al., 1994). The PDGF induced redistribution of β-actin mRNA can be inhibited by inhibiting tyrosine kinase activity, which indicates that the localization process also depends on tyrosine kinase activity (Latham et al., 1994).

3.2 Methods to visualize mRNA in cells

The problem of biochemical assays is that they cannot provide information about the high stochastic variations amongst the cells with the same genotype which can be found in diseases (Eldar and Elowitz, 2010). Both mRNA and protein are highly compartmentalized in the cell, methods help in providing the spatial information and kinetics of the RNA.

3.2.1 Methods to visualize mRNA in fixed cells

The first visual evidence of DNA in agar embedded fixed cells came from Pardue et al. using tritium labeled rRNA probes against extra-chromosomal rDNA in Xenopus oocyte (Gall and Pardue, 1969). The whole method used to take around three months, since it depended on the half-life of the radioactive molecule and its emission of electrons during the decay. An improvement in the signal to noise ratio was made by using a fluorescently labeled in situ hybridization method to visualize DNA (Cheung et al.,

1977) and by developing a biotinylated analog of dUTP containing an allylamine linker arm between biotin molecule and pyrimidine ring. The incorporation of this molecule in the DNA was first achieved by nick translation (Langer et al., 1981). A significant technological advance was made in 1998 by the group of Robert Singer during their studies on β-actin mRNA localization. The Singer group used several in vitro synthesized antisense oligonucleotide probes against the target RNA that had been chemically conjugated with several fluorophores. This massive labeling allowed proper hybridization of the probes as well as sufficient fluorescent signal to acquire images from different focal planes. In combination with the restoration of the three-dimensional information by deconvolution software (for example Deconv; (Sun et al., 2009)), this allowed for a first complete spatiotemporal information about a specific mRNA.

3.2.1.1 smFISH (Single-molecule fluorescence in-situ hybridization): In this method, the tissue or cells are fixed and permeabilized. Afterward, they are hybridized with a set short (~20bp) fluorescently labeled antisense DNA oligonucleotides, which are complementary to the target mRNA and regularly interspaced along its sequence (see Fig. 13A). The multiple probes increase the signal to noise ratio and allow it to detect single mRNA molecules by standard fluorescence microscopy as diffraction limited spots. This method can be used to gain information on the spatial distribution of the target RNA as well as quantify the number of mRNAs (Fig. 13A). smFISH related methods can not only be applied to mRNAs but can also detect and quantify ribosomal RNA (Buxbaum et al., 2014), viral RNA (Chou et al., 2013) as well as non-coding RNAs (Cabili et al., 2015). The probes can also be quantum dots (QD) which are basically nanocrystals with fluorescent properties. The advantage of QD over standard fluorescent dyes is the anti-bleaching property since they are stable for up to 12 min of exposure time (Liu et al., 2018).

3.2.1.2 RollFISH: ROLLFISH is a modified smFISH protocol for parafilm embedded tissue samples. It can be used with non-formaldehyde fixed samples and is useful for increasing the signal to noise ratio to such a high level that resolution can be achieved even at 20x magnification. In short, DNA oligonucleotides are used that, in addition to a 46 nt long sequence complementary to a given RNA contain a 30 nt docking sequence to human transcriptome at the 3'end. Then the signal from one RNA complementary probe is amplified by using padlock probes against the 46nt long docking sequence and amplifying by rolling circle amplification (RCA) with Phi29 polymerase after circularizing the probes. Then a secondary fluorescent-labeled secondary oligonucleotide is used that binds in multiple copies to the amplified signal (see Fig. 13B; also (Wu et al., 2018)).

3.2.1.3 Rolling circle amplification (RCA) of padlock probe-based techniques: Amplification and sequencing of mRNAs *in situ* have been challenging for a long period. There are three major methods based upon reverse transcription and annealing, that revolutionized the exploration of RNA detection *in situ*. However, all these methods have their advantages and disadvantages. The first one is the padlock probe method in which a padlock probe, which is an oligonucleotide binds complementary to two adjacent sequences on its target inside the cell by its two flanking end called arms (Fig. 13C). The two arms of a padlock probe are linked by a generic backbone sequence and the arm hybridizing downstream at the 5'end allows ligation of both arms to form an ssDNA circle using a dsDNA ligase. This circularized ssDNA acts as a template for rolling circle amplification to generate a <1 µm nano ball of DNA containing thousands of copies of the original sequence. This method has an advantage with high sensitivity and about ~30% RNA detection capacity. However, the major disadvantage of this approach is that it is not capable of amplifying satisfactory length, and the protocols are tedious and complicated, limiting the reproducibility.

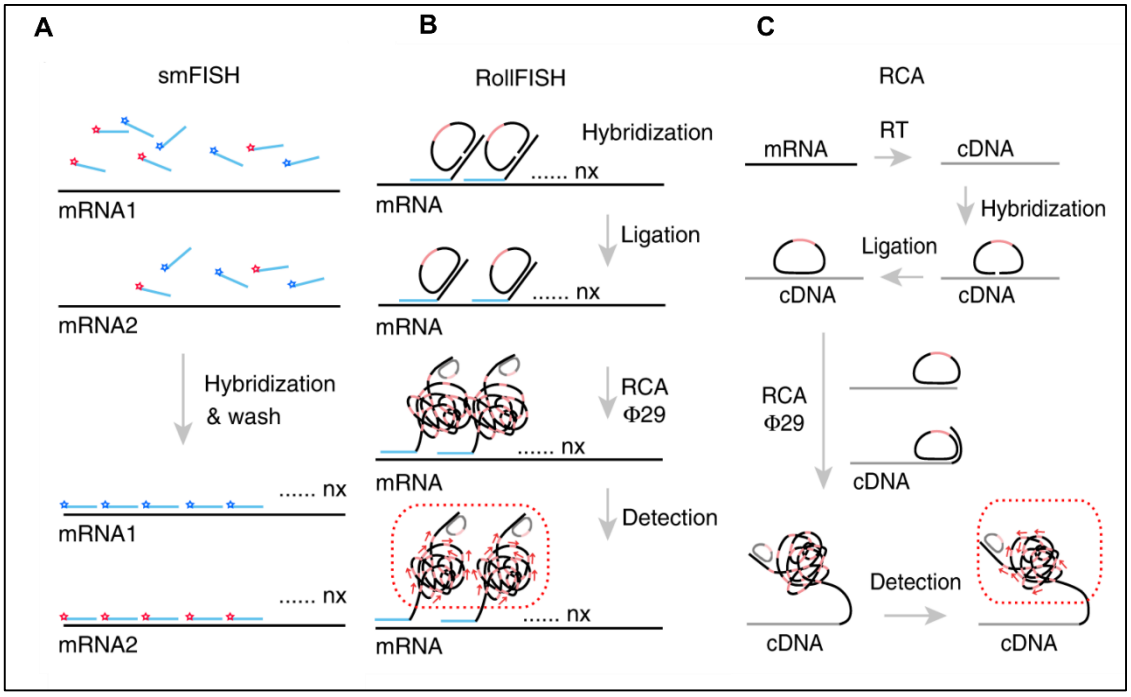


Figure 13. The workflow of RollFISH in comparison to smFISH and standard RCA.

Each Oligodeoxynucleotides in a RollFISH probe consists of 30nt complementary to the target RNA sequence and 46nt orthogonal to the human transcriptome, serving as docking sequence for a padlock probe. A hinge sequence of four thymidines (T) is included between the two sequences to facilitate recognition and binding of the padlock probe to the docking sequence.Taken from (Wu et al., 2018) with permission from the authors.

3.2.1.4 RNAscope: RNA scope is an RNA FISH method that can be used on parafilm embedded tissue or formaldehyde-fixed cells. A pair of RNA complementary probes (each 18-25 nucleotide long) with a spacer sequence and different 14bp tail sequences, is used which conceptualized as Z shape, as described in Fig. 14. The tail sequence (28 bases together) acts as a hybridization site for the preamplifier which contains in total 20 binding sites for the next amplifier, which in turn, contains 20 binding sites for the labeled probes. So finally, from 2 probes (~50 bases in length), the signal from one mRNA can be amplified up to 8000 times which could be easily visible by epifluorescent microscope (for a fluorescent label) or standard bright-field microscope (for enzyme label) (Wang et al., 2012).

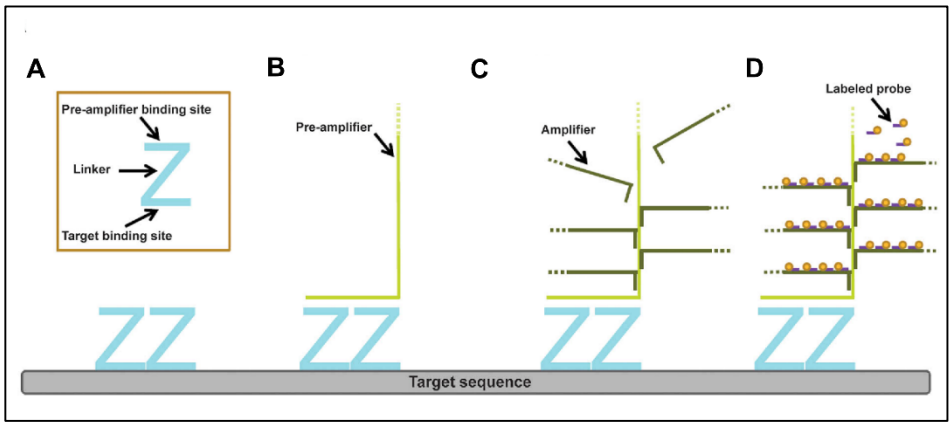


Fig. 14. Overview of RNAscope technology.

(A) A standard target probe consists of a pool of 20 double Z probes targeting a region of 1000 bases. Each Z target probe contains three elements: the lower region is complementary to the target RNA and is selected for target-specific hybridization and uniform hybridization properties; a spacer sequence links the lower region to an upper region; the two adjacent upper regions from a double Z target probe form a 28 base binding site for the pre-amplifier. **(B)** Once the Z probe pairs hybridize to the RNA target, the pre-amplifier binds to the upper regions of the Z probe pairs. **(C)** Hybridization of multiple amplifiers per pre-amplifier. **(D)** Hybridization of multiple labeled probes per amplifier. Labeled probes contain a chromogenic enzyme to generate one punctate dot per RNA target. The size of the dot is directly proportional to the number of Z probe pairs hybridized onto the RNA target. Hybridization of only three Z probe pairs is sufficient to obtain a detectable signal by brightfield microscopy. Taken from (Anderson et al., 2016)with permission from the authors.

3.2.2 Methods to visualize mRNAs in live cells

Live cell imaging of mRNA is of paramount importance in biological sciences. It provides comprehensive information about its expression, kinetics, localization, decay, and storage. RNA imaging is also critically important in clinical biology, especially for

detecting spatial drug delivery in RNA therapeutics (Niu and Chen, 2009). The different live-cell imaging techniques that are mainly used in basic and clinical biological research will be briefly discussed below. The initial approaches used to visualize the mRNA was to fluorescently tag an RBP that is known to bind to the mRNA of interest to indirectly follow mRNA dynamics (Wang and Hazelrigg, 1994). Later, this was improved by the introduction of MS2 based systems (Bertrand et al., 1998). RNA loops from the RNA bacteriophage MS2 were introduced at the 3'UTR of budding yeast *ASH1* mRNA and co-expressed with a fusion of MCP (MS2 coat protein) and GFP. The MCP part recognizes the MS2 RNA loop, directing the GFP to the RNA. This system has been adapted to various fluorescent proteins and allows kinetics analysis of mRNAs by live-cell imaging in the cell. Together with the MS2-MCP system, the other methods to visualize mRNAs in live cells will be mentioned below.

3.2.2.1 Rhodamine labeled mRNA injection

In this method, the RNA of interest is transcribed by *in vitro* transcription in the presence of 5-(3-aminoallyl)-UTP and conjugated with tetramethylrhodamine-6-isothiocyanate (Wang et al., 1991). The modified RNA molecules are then injected into the cell or organism of interest and can be visualized under the fluorescence microscope (Anhäuser et al., 2019).
The drawbacks of this method are several. The labeling efficiency is rarely 100%, so there is always a population of the mRNA, which remains unlabeled. As the overexpression of a system, there is always an introduction of more mRNA injected into the cell compared to the endogenous RNA. Also, there is always competition with endogenous RNA. As these mRNAs are mostly injected directly into the movement inside the nucleus cannot be seen.

3.2.2.2 Molecular Beacons (MB)

The molecular beacons method is based upon engineered antisense oligonucleotide labeled with fluorophore against its mRNA of interest. An oligonucleotide is attached with a fluorescence reporter at one end and a fluorescence quencher at the other end. In principle, upon binding to the respective complementary sequence on the mRNA of interest, the MB linearizes, which leads to an increase in distance between the fluorophore and the quencher (≥ 10nm) and restoration of fluorescence (Tyagi and Kramer, 1996). There are two types of quenching occurs in the MB methods

A. Dynamic quenching: the oligonucleotide sequence of the MB can form a stem-loop like a hairpin structure. When the MB is in the stem-loop form, then the quenching occurs through energy transfer or electron transfer. Due to the long-range dipole-dipole interaction between the reporter and the quencher, the photon is not released from the reporter. The energy transfer efficiency depends on the spectral overlap between the emission spectrum of the reporter and absorption spectrum of the quencher (Schatz, 1993), quantum yield of the donor, dipole moment and distance between the two groups.

B. Static quenching: principally in this contact quenching when the reporter and the quencher are in proximity, there is a subtle level of heat energy is transferred, which helps in determining the basal fluorescence value. Static quenching depends on the stem sequence, the linker, and the reporter quencher pair. The most frequently used quenchers are Dabcyl (4-([4′-(dimethylamino) phenyl]azo)benzoic acid, BHQ1, BHQ2, Iowa Black FQ, and RQ. The static quenching ration for these molecules is in the range of 85-97%. These quenchers can be paired with various reporters like,

Reporter	**Compatible quenchers**
Cy3	BHQ-2 and Dabcyl
Cy5	BHQ-3 and Dabcyl
CR-6G	Dabcyl
6-FAM	Iowa Black-FQ, BHQ-1, and Dabcyl
HEX	BHQ-1, Iowa Black-FQ, and Dabcyl
TAMRA	Dabcyl
TET	BHQ-1, Iowa Black-FQ, and Dabcyl

3.2.2.3 Aptamer based imaging system

Aptamer based systems rely on the expression of an exogenous transcript together with a fluorescent protein or organic dye of interest fused to an aptamer sequence. The aptamer sequence is usually fused at the 3'UTR of the target mRNA (chimeric transcript). The exogenous transcript or the aptamer itself is not fluorescent, but when it binds to the exogenous transcript fused to a fluorescent protein, the entire RNA

becomes visible. The chimeric transcript can also be imaged using a suitable fluorescent which activated only after binding to the aptamer. All the aptamers that are used for microscopy-based RNA visualization techniques have been mentioned in Fig. 15.

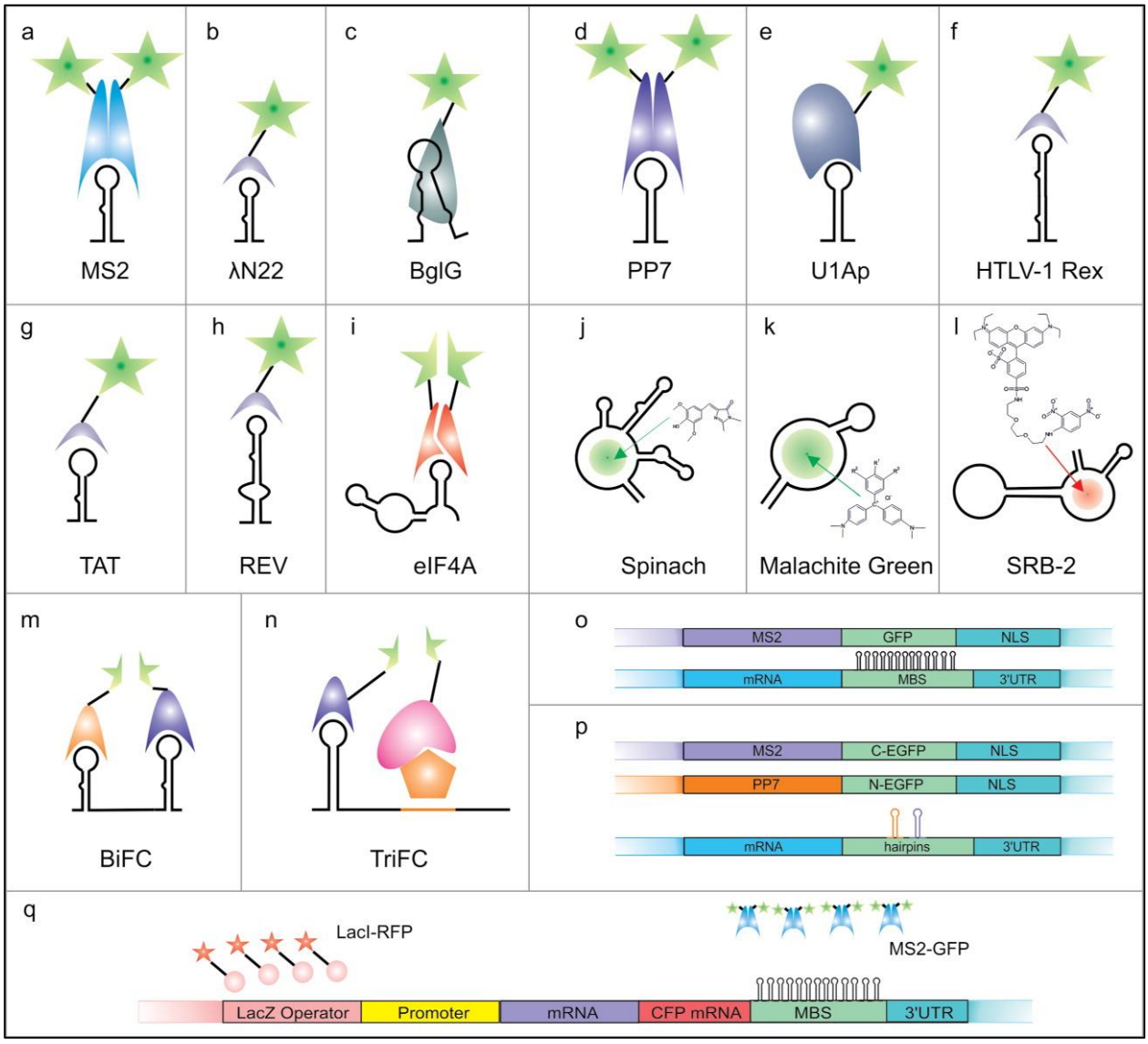


Fig 15. Vector-based systems for RNA live imaging.

Schematic structures of RNAs with protein partners or fluorescent dyes are presented (a–l). Additionally, examples of genetic constructs used for imaging experiments are depicted (o–r). (a) MS2 systems, (b) λN22 system, (c) BglG systems, (d) PP7 system, (e) U1Ap system, (f) HTLV-1 Rex system, (g) TAT system, (h) REV system, (i) eIF4A system, BiFC with the use of 2 domains, (j) Spinach system, (k) Malachite green system, (l) SRB-2 system, (m) BiFC with the use of 2 systems, (n) TriFC, (o) DNA construct for MS2 system, (p) DNA constructs for BiFC with 2 systems and (r) system for gene locus, mRNA and protein product imaging. Taken from (Urbanek et al., 2014), with permission by the authors.

3.2.2.4 The MS2 RNA labeling system

For MS2 loops, the RNA hairpin contains a four-nucleotide loop and a seven bp steam and has a single adenine bulge. Mutations at specific sites on the loops can increase or decrease the binding efficiency with the coat protein. A single substitution of U to C in the stem-loop can increase the coat protein binding efficiency to 50-fold over unmodified loops (Lowary and Uhlenbeck, 1987). The MS2 bacteriophage derived hairpin loop (MS2 loop) (Peabody, 1993) bound to its complementary protein MCP (Johansson et al., 1997) has been extensively used as a standard tool to visualize RNA in live-cell imaging (Tutucci et al., 2018). In short, a gene of interest, tagged with MS2 loops (24 copies in general) and the MCP fused to a fluorescent protein (MCP-FP), and NLS (nuclear localization sequence) are co-expressed in the same cell. The MCP-FP binds to the MS2 loop as a dimer such that via the fluorescent protein the MS2 tagged RNAs can be visualized throughout their lifetime. When used in combination with orthogonal, MS2 like systems, such as PP7 (Lim and Peabody, 2002), Bgl loops (Bann and Parent, 2012; Chen et al., 2009), lambda boxB RNA (Legault et al., 1998), or U1A (Brodsky and Silver, 2002) this type of indirect RNA labeling can be used for simultaneous tracking of multiple RNAs in single cells (Lange et al., 2008). Due to the NLS, the MCP-FP fusion protein binds to the RNA co-transcriptionally in the nucleus and is co-exported to the cytoplasm together with its cognate RNA. Unbound MCP-FP proteins freely move in the nucleus, resulting in a nuclear (background) fluorescence. To eliminate this problem, a combination of MS2 and PP7 loops or MS2 and lambda boxB have been used instead of MS2 loops (Kd = 1 nM; (Carey and Uhlenbeck, 1983)) alone, When combined with split fluorescent proteins (e.g., MCP-split FP and PCP-split FP), it is possible to track mRNAs in background-free live cells (Wu et al., 2014).

The disadvantage of the MS2 system is that, due to genetic modification of the mRNA, it can lead to alteration in the stability of the mRNA itself. To identify the site of integration of the loops on the mRNA is a critical decision because the 3’UTR of any mRNA consists of regulatory regions, critical determinants of mRNA stability, localization, and function. If different isoform exists for this mRNA, it difficult to visualize all the isoforms of the same mRNA.

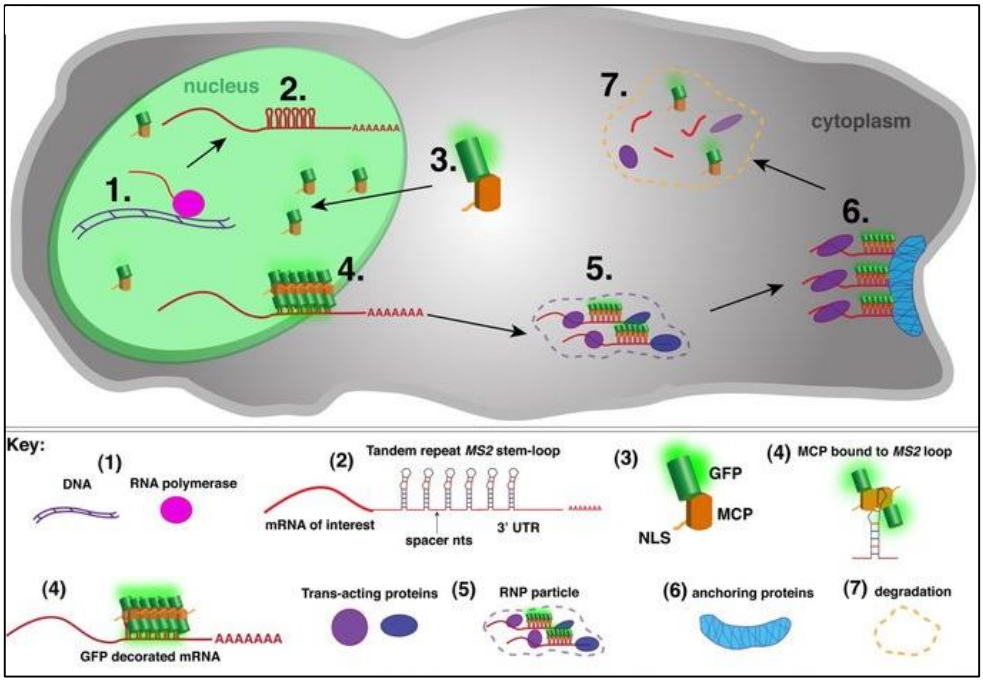


Fig. 16 MS2 labeling the life cycle of endogenous mRNA.

The picture is self-explanatory but still in short, it represents the use of MS2-MCP system to visualize any endogenous MS2 tagged mRNA (here β-actin tagged with 24MS2 loops) in a cell starting after transcription of the mRNA (2), during export (4), localization (5) tethering at the protrusion, and (6) until degraded. Taken from (Weil et al., 2010) with permission.

3.2. Methods to detect proteomes by proximity biotinylation

3.2.1 BioID (proximity-dependent biotin identification)

Biotin (vitamin H or Vitamin B7) is an essential coenzyme that is mainly synthesized in plants, bacteria, and in some fungi. Biotin (MW = 244.31 g/mol) is essential for humans and used as a cofactor for enzymes like biotin carboxylases and decarboxylases that are essential for cellular processes like amino acid metabolism, lipogenesis, or gluconeogenesis. A well-characterized carboxylase in *E. coli* is known as BirA. This enzyme uses CO_2 from bicarbonate and catalyzes the transfer of a carboxyl group to organic acids by using biotin as carboxyl carrier, the whole process is known as biotinylation (Chapman-Smith and Cronan Jr, 1999). In protein biotinylation, the number of biotins attached is not uniform, and the modification at the lysine residue of the target protein may lead to inactivation of the binding site(s). By screening combinatorial peptide libraries, a minimal biotinylation sequence (13 or 15 amino acid

long) termed as AviTag was identified as an effective target for BirA both *in vivo* and *in vitro* (Beckett et al., 2008; Schatz, 1993). Target specific biotinylation can be done by tagging a target protein with an AviTag and co-expressing BirA in the cell (Reid and Nicchitta, 2015). BirA biotinylates in a two-step process. First, it generates reactive biotinyl-AMP (bioAMP) from biotin and ATP. In the second step, the bioAMP reacts with a nearby lysine residue by acylation. A mutant (R118G) version of BirA known as BirA* prematurely releases the reactive bioAMP, which then covalently reacts with the adjacent primary amines (mostly lysine) (Fig. 17A) (Kwon and Beckett, 2000). Biotin has a strong affinity for streptavidin, avidin or neutravidin. Biotin and streptavidin are one of the strongest known interactors in nature (Kd =10^{-15} M) that is why biotinylated proteins are isolated by streptavidin or avidin tagged beads.

3.2.2 APEX2

APEX (ascorbate peroxidase) is a 28 kDa monomeric enzyme, derived from dimeric pea enzyme (Martell et al., 2012) or its soybean homolog (APEX2) (Rhee et al., 2013). APEX2 lacks a calcium-binding site and thus can be expressed in a variety of cellular environments without loss of activity (Martell et al., 2012). In the presence of oxidant, H_2O_2, APEX enzyme catalyzes the conversion of biotin-phenol into reactive radicals (active for <1 ms), which are then attached covalently to electron-rich amino acids like tyrosine (Fig. 17B). The labelling reaction can be terminated by removing the H_2O_2, by quenching solution. The biotinylated proteins can be subsequently isolated by streptavidin beads (Han et al., 2017).

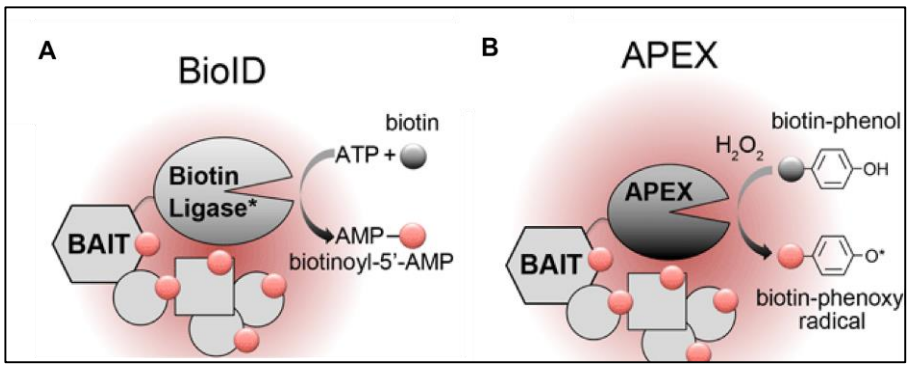


Fig 17. Proximity biotinylation methods.

(A) BioID is based on the expression of a bait protein fused to a mutant biotin ligase that catalyzes the conversion of biotin to biotinoyl-5′-adenosine monophosphate (AMP). This highly reactive form of biotin attaches covalently to accessible lysine residues in neighboring proteins. **(B)** APEX is based on the

expression of a bait protein fused to a peroxidase that, in the presence of H_2O_2, catalyzes the oxidation of biotin-phenol to a biotin-phenoxyl radical. This activated biotin can attach covalently to electron-rich amino acids (tyrosine and possibly tryptophan, cysteine, and histidine) in neighboring proteins. Taken from (Trinkle-Mulcahy, 2019) with permission.

4.Method

4.1 Plasmids and cloning

The lentiviral vector pHAGE UbiC carrying NLS-2X MCP-eGFP (Mukherjee et al., 2019) was used as a backbone to generate NLS-2X MCP- eGFP-APEX2 plasmid. To generate the pHAGE-NLS-2XMCP-eGFP-APEX2 plasmid, the APEX2 fragment was amplified from pcDNA3.1 APEX2 plasmid using forward primer XhoI-Apex2-F (CCGCTCGAGGGAAAGTCTTACCCAACTGTGAGTGCTGATTACCA) and reverse primer ClaI-Apex2-R (CCATCGATTTAGGCATCAGCAAACCCAAGCTCGGAAAGCTTTTG) and cloned into the lentiviral vector in XhoI and ClaI sites.

4.2 Stable cell line generation with lentivirus and (Fluorescent activated cell sorting) FACS analysis

A detailed description of generating stable cell lines expressing proximity labeling enzymes can be found in MS1 and MS2. In short, 20 µg of MCP-GFP-APEX2 (backbone plasmid) together with accessory plasmids (1 µg pCEP4-tat, 1 µg pRSV-Rev, 1 µg pMDLg/pRRE, 2 µg pMD2.G) were transfected in Hek293FT using Fugene HD reagent (Promega). The culture supernatant was harvested during the following three days and, after passing through 0.45 µm filter, concentrated with a LentiXconcentrator (Clontech) according to manufacturer's recommendation (for details see Mukherjee et al., 2019). The concentrated supernatant was used for transduction of MEF cells at a dilution of 1:100 and in the presence of 8 mg/ml protamine sulphate. Three days after the transfection, MEFs were washed 3x with serum containing media and sorted in FACS buffer (1x DPBS without calcium and magnesium, 0.2% BSA, 0.5 mM EDTA, 5 mM MgCl2). FACS Cells expressing low levels of GFP were isolated with a FACS Aria cell sorter (Becton-Dickinson) before further culturing.

4.3 Cell Culture and serum starvation assay

WT mouse embryonic fibroblasts (MEFs) from C57BL/6J mice, MEFs expressing NLS-2X MCP-eGFP or 24MBS-β-actin together with NLS-2X MCP-eGFP (Lionette et a., 2014), MEFs expressing NLS-2X-MCP-eGFP-BirA* or 24MBS-β-actin together with NLS-2X-MCP-eGFP-BirA* (Mukherjee et al., 2019), Hek293T cells were cultured in DMEM (with 4.5 g glucose and L-glutamine containing 10% fetal bovine serum (FBS) and 1% pen-strep).

For serum starvation of MEFs, cells were starved in DMEM medium without serum (including 1% pen-strep) for at least 24 hrs and induced with DMEM media containing 10% FBS and 1% pen-strep for different time points as mentioned in the results section

(for BirA* and APEX2 MEFs cells were induced for 6 hrs. For GFP containing cells the cells were induced for 1hr only). A detailed description of the starvation assay can be found in MS1.

4.4 Immunoprecipitation and western blot

For immunoprecipitation (IP), the protocol is described in detail in Mukherjee et al., (2019). In short, 200 µg of total protein was used per pull down experiment. 100 µl of protein G coupled magnetic beads (dynabeads) were used for IP. For antibody coupling, the beads were washed 3x in NT2 buffer (Keene et al., 2006) and blocked with 5% BSA and 0.5 mg/ml ssDNA. After washing the beads once again with NT2 buffer, the beads were incubated with either 20 µg of FUBP3 antibody for overnight at 4°C with end to end rotation. After preclearing the lysates with 100 µl of Protein G magnetic beads, antibody coupled beads were added to the lysates and incubated for 2 hrs at 4°C with end to end rotation. After the incubation, beads were separated from the lysates by a magnetic stand and washed five times with ice cold NT2 buffer, before the supernatant was removed. Beads were resuspended in 100 µl of NT2 buffer. For isolation of proteins from the beads, 40 µl of beads were boiled in 100 µl 1x Laemmli buffer for 10 min at 95°C and the elute was separated from the beads on a magnetic strand. For western blots 40 µl from the eluted sample were used (Mukherjee et al., 2019 see also MS2). After transfer the membrane were blotted against rabbit anti GFP (Invitrogen-A-6455 / 1:5000 dilution), mouse anti-GAPDH (Proteintech. 60004-1-Ig.Clone no. 1E6D9 / 1:2000 dilution), rabbit anti FUBP3 (Abcam: ab181122, 1:2000 dilution) or rabbit anti STAU2 (a gift from M. Kiebler, LMU Munich, 1:2000 dilution). Probing the blot for IGF2BP1 or STAU1 has been described in detail in MS1.

4.5 smFISH and microscopy and data analysis

MEFS expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP-APEX2 were seeded on cover glass in a 6-well cell culture plate and grown for 24 hrs. in serum-free medium followed by induction with serum for 1hr. To visualize β-actin mRNA, the protocol was followed as described previously (Eliscovich et al,. 2017 ; Mukherjee et al., 2019). In short, cells were washed three times with PBS, fixed for 10 min with 4% paraformaldehyde in PBS, washed three times in PBS and then quenched in 50 mM glycine, and finally permeabilized with 0.1% Triton X-100 (28314; Thermo Fisher Scientific) and 0.5% Ultrapure BSA (AM2616; Life Technologies) in 1× PBS-M for 10 min. After washing with PBS, cells were treated with 10% formamide, 2X SSC, and 0.5% Ultrapure BSA in RNase-free water for 1 hr at room temperature, followed by incubation for 3 hrs at 37°C with 10 ng custom made ATTO488 labeled MBS probes (MS2_LK20 TTTCTAGAGTCGACCTGCAG, MS2_LK51_1 CTAGGCAATTAGGTACCTTAG, MS2_LK51-2 CTAATGAACCCGGGAATACTG). After incubation and quick washing, cells were washed with 2X SSC containing formamide and 2 times with 2XSSC without formamide., DNA was counterstained with

DAPI (0.1 µg/mL in 2XSSC; Sigma-Aldrich), and after a final wash with 2X SSC, cells were mounted using Antifade Reagent (Life Technologies).

Imaging was performed using Zeiss Cell Observer wide-field fluorescence microscope, operated illuminated with a xenon arc lamp, and detected with a CCD camera with 100×/1.45 α-Plan fluor oil immersion objectives (Zeiss). The images were analyzed in FISH-QUANT as described in Mukherjee et al. (2019).

4.6 RNA-BioID

A step by step protocol for RNA-BioID is described in MS2 (see also Mukherjee et al., 2019). In short, 500 µg of total protein were used for each streptavidin pulldown experiment. 50 µM biotin were added for 6 hrs to unstarved or 24 hrs starved MEFs expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP-BirA*. MEFs expressing only NLS-2X-MCP-eGFP-BirA* were not treated with biotin. After lysis 200 µL of streptavidin magnetic bead suspension (GE Healthcare) were added, and the lysate was incubated overnight at 4°C with end-to-end rotation. The next day, the beads were collected (by keeping the beads on the magnetic stand for 2 min) and washed as described in Mukherjee et al. (2019).

4.7 RNA-APEX2

For RNA-APEX2, four 90% confluent 100 mm dishes containing either MEFs expressing NLS-2X-MCP-eGFP-APEX2 or MEFS expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP-APEX2 were used (for each round of mass spectroscopy). Before starting the experiments, all the reagents were freshly prepared. Cells were washed once with growth media and treated with biotin-phenol (final concentration 500 µM) for 30 min in the incubator. For the labeling reaction, cells were treated with H_2O_2 for exact 1 min and the reaction was stopped by adding 5 ml of quencher solution (Han et al., 2017) per 100 mm dish for another 2 mins. Cells were washed with 5 ml of quencher solutions for three more times.

For lysis, 3 ml of fresh quencher solution was added on the plates and the cells were scraped off from the dish. After a spinning down the cells (500xg for 5 min at 4°C), two packed cell volumes of quencher solution and two packed cell volumes of 2X RIPA buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.5 % sodium deoxycholate; 0.8 % Triton) was added. Cells were lysed with a 21G needle for 15 times. The lysate was centrifuged (15,000 g, 10 min, 4°C) and the supernatant fraction was isolated.

For streptavidin pull down, 2 mg of total protein (protein concentration was measured by a Pierce 660 assay) were used. For pull down of biotinylated protein 200 µl of streptavidin magnetic beads were and incubated similar to like biotinylation pull down.

Next day, the beads were washed for 5 - 10 minutes at RT with end to end rotation, twice with 1 ml RIPA lysis buffer, once with 1 ml 1 M KCl, once with 1 ml 0.1 M Na2CO3, once with 1 ml 2M Urea in 10 mM Tris-HCl at pH 8 and again twice with 1 ml RIPA lysis buffer to remove nonspecific binders. The beads were washed twice with 50 mM ammonium bicarbonate buffer before on-bead digestion.

For biotinylation after serum induction, MEFs expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP-APEX2, were starved for 24 hrs and then induced with 10% serum-containing medium for 6 hrs followed by labeling of biotinylated proteins for 1 min with biotin-phenol as described above.

4.8 MS2-based RNA-pulldown

For affinity purification of β-actin mRNA with anti-GFP beads, 12 90% confluent 100 mm dishes containing either MEFs expressing NLS-2X-MCP-eGFP or MEFS expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP were used for each round of mass spectroscopy. Cells were washed twice with ice-cod 1X PBS and lysed in ice-cold 10 mM Hepes-KOH, pH 7.0, 100 mM KCl, 5 mM MgCl2, 0.5% Nonidet P-40 , supplemented with protease Inhibitor Mixture (Roche), and 100 U/mL Ribolock (Thermo Scientific) by keeping 20 min on ice followed by a quick vortexing. Lysates were centrifuged (3000 g for 5min at 4°C) and the lysates were kept on ice.

For affinity purification of β-actin mRNA with anti-GFP beads, 12, 90% confluent 100 mm (for each round of mass spectroscopy) dishes containing either MEFs expressing NLS-2X-MCP-eGFP or MEFS expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP were used. Cells were washed twice with ice-cod 1XPBS and lysed in ice-cold 10 mM Hepes-KOH, pH 7.0, 100 mM KCl, 5 mM MgCl2, 0.5% Nonidet P-40 , supplemented with protease Inhibitor Mixture (Roche), and 100 U/mL Ribolock (Thermo Scientific) by keeping 20 min on ice followed by a quick vortexing. Lysates were centrifuged (3000 g for 5min at 4°C). 10 mg of total protein were used for each pull down experiments. Cell lysates were precleared with 20 ul of magnetic agarose beads (chromotek) for 20 min at 4 C with end to end rotation. For pulling down the RNA, 100 ul of GFP-magnetic-agarose nanotrap beads (chromotek) were added to the 10 mg precleared lysate and incubated for 1 hr at 4°C with end to end rotation. The beads were then washed five times with cold NT2 buffer (1ml each time) supplemented with protease and Rnase inhibitors. After washing the beads two times with 50 mM ammonium bicarbonate, mass spectroscopy analysis was performed by on bead digestion.

For biotinylation after serum induction, MEFs expressing 24MBS-β-actin together with NLS-2X-MCP-eGFP were starved for 24 hrs and then induced with 10% serum-containing medium for 1 hr.

5. Aims of this thesis and significance

Due to the controlling functions of RNA-associated proteins on an mRNA, it would be key to know the proteome associated with a specific mRNA at any time of its life. This is true not only for the RBPs directly bound to the mRNA but also for those proteins that transiently interact but play a critical role in the localization of the mRNA of interest.

In this study, I developed a method to identify the interacting proteome of β-actin mRNA before and after stimulating the localization and compared the proteomic difference under these two conditions. This was achieved by tethering the biotin ligase BirA* to the 3'UTR of 24MS2 tagged endogenous β-actin mRNA and induced biotinylation of RNA-associated proteins. I could identify all previously known β-actin mRNA interactors like IGF2BP1, ZBP2, KHDRBS1, FMR1, RACK1. Also, I identified new interactors of this mRNA, including FUBP3, STAU1, STAU2, MATR3, ELAVL1 with high confidence via mass spectroscopy. A similar protein set was identified via tethering of the APEX2 enzyme at the 3'UTR of β-actin mRNA. I also studied the role of FUBP3 in β-actin mRNA localization. By super-registration microscopy and *in vivo* pull-down of FUBP3, I showed that FUBP3 binds to β-actin mRNA comparable to the level of IGF2BP1 (the mouse homolog of the β-actin mRNA zipcode binding protein ZBP1). By *in vitro* binding assays, I could show that the FUBP3 protein binds to a region in 3'UTR downstream of the cognate zipcode. shRNA mediated knockdown of FUBP3 revealed that though FUBP3 does not bind to the known zipcode sequence on β-actin mRNA, its binding is essential for proper localization of the β-actin mRNA to the protrusion of mouse embryonic fibroblasts. The RNA-proximity biotinylation methods and the downstream analysis experiments have brought us a step closer to understanding the entire proteome of a specific mRNA and how to characterize the function of the identified proteome.

6. Summary of the results

6.1 Methods to identify β-actin interacting proteome

To identify the β-actin mRNA interacting proteome I tethered either the proximity labeling enzymes BirA* or APEX2, or only eGFP to the 3'UTR of β-actin mRNA (Fig. 18A) by stably expressing a fusion protein containing a nuclear-localized signal (NLS), MS2 coat protein (MCP), GFP, and or BirA*/ APEX2 in either WT (Fig. 18A or 18B, left) or in immortalized mouse embryonic fibroblasts (MEFs), isolated from transgenic β-actin-24 MBS heterozygous mice (Fig. 18A or 18B, right). WT MEFs only expressing the fusion protein(s) were generated as control cell lines to estimate endogenous biotinylation of proteins binding non-specifically to the streptavidin affinity matrix. All cell lines expressed a fusion of two copies of the MCP protein together with one copy of eGFP or one copy of eGFP attached to BirA* or APEX2. Thus, in an ideal scenario, there will be 24 GFP, 24 GFP plus 24 BirA*, or 24 GFP plus 24 APEX2 on the mRNA. To check for the proper expression of the heterologous protein, I performed western blot to evaluate the degree of (over)expression of the protein (Fig. 18D also Fig. 1D and S2 of (manuscript 1) MS1). To validate whether the overexpression of heterologous protein has any effect on the β-actin mRNA localization, I performed smFISH against β-actin mRNA to visualize its targeting to the protrusion. To induce the β-actin mRNA localization in MEFs, cells were serum-starved for 24 hrs. followed by stimulation with serum addition for 1hr. As a negative control for localization of β-actin mRNA at the protrusion, I used IGF2BP1 KO cells expressing endogenous β-actin-24 MBS and expressing 2XMCP-eGFP. As expected, the control cells showed less mRNA at the protrusion after serum starvation (Fig. 18E also Fig. 1B and S1 of MS1).

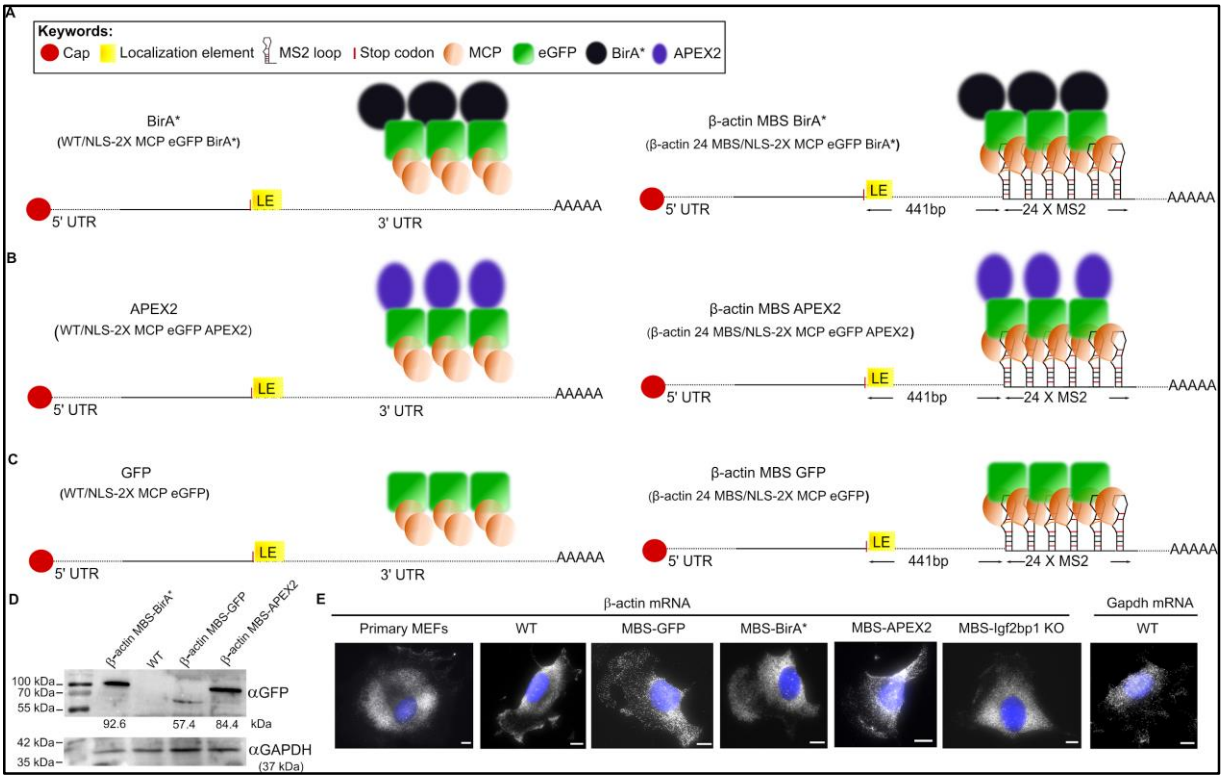


Fig. 18. Design of RNA BioID.

(A) RNA-APEX2 (B) and RNA-GFP pull-down to detect proteins interacting with localized β-actin RNA. (A) Schematic of the β-actin–MBS/GFP-BirA*. (Left) Control construct (BirA*) used to detect background biotinylation due to overexpression of the NLS-MCP-GFP-BirA* construct. Control cells expressing only NLS-2xMCP-eGFP-BirA* lack the MBS cassette in the β-actin mRNA. (Right) Construct used to detect β-actin mRNA-associated proteins (β-actin–MBS-BirA*). A 24xMS2 aptamer array (24MBS) was integrated into the 3′ UTR of the endogenous β-actin gene 441 bp downstream of the stop codon. BirA* is targeted to 24MBS by its fusion to an MS2 coat protein dimer (2xMCP). **(B)** Schematic of the β-actin–MBS/GFP-APEX2. (Left) Control construct (APEX2) used to detect background biotinylation due to overexpression of the NLS-MCP-GFP-APEX2. **(C)** Schematic of the β-actin–MBS/GFP (Left) Control construct (GFP) used as a control for the pull-down due to overexpression of the NLS-MCP-GFP. **(D)** Expression of the heterogeneous proteins was detected against the anti-GFP antibody, expression of GAPDH was detected with an anti-GAPDH antibody which was used as a loading control. **(E)** Representative β-actin smFISH images of (from left to right) primary MEFs, immortalized MEFs (WT), β-actin–MBS-GFP, β-actin–MBS BirA*, β-actin–MBS APEX2 and β-actin–MBS Igf2bp1 KO MEFs, as well as Gapdh smFISH (rightmost image) images in immortalized (WT) serum-induced MEFs.

6.2 Identification of proteins in proximity biotinylation and β-actin pull-down methods

The localization of β-actin mRNA to the fibroblast protrusion can increase after serum starvation. To differentiate the change in the proteomic interaction of the β-actin mRNA before and after stimulating localization with serum, I performed a serum starvation assay (Mukherjee et al., 2019) in all six different kinds of generated MEFs (WT/ β-actin MBS GFP, WT/ β-actin MBS BirA* and WT/ β-actin MBS APEX2). All cells were starved for 24 hrs. in cell culture media without FBS. After 24 hrs. fresh serum-containing growth media was added to the cell. The serum induction time was different for different cell lines. As BirA* required at-least 6 hrs. for proper labeling (Mukherjee et al., 2019; Roux et al., 2012), β-actin-MBS BirA* cells were induced with serum for 6 hrs. in 50 µM biotin and FBS containing growth media. As the labeling time for APEX2 is only 1 min, β-actin-MBS APEX2 cells were induced with serum for 1hr followed by treatment with biotin-phenol. The β-actin-MBS GFP cells were also induced for 1hr with serum-containing growth media. In all cases, the WT cells were used as a control to eliminate the background and were treated only growth media similar to before. For identification of proteomes from different cell lines under unstarved condition, β-actin-MBS BirA* or β-actin-MBS APEX2 cells were treated with either biotin or biotin-phenol for 6 hrs. or 1min respectively without starvation while β-actin-MBS GFP cells were kept in normal growth media. Biotinylated, streptavidin-captured proteins were identified and quantified by mass spectrometry using label-free quantification, followed by an enrichment analysis (Fig. 19). In summary, more proteins were identified by RNA-BioID in induced: 341 or (uninduced: 324) conditions (Fig. 19A left), compared to RNA-APEX2 234, (243) (Fig. 19B left) or β-actin-pull down 72, (64) (Fig. 19C left). Gene ontology (GO) database description for the cellular components (CC) in RNA-BioID or RNA-APEX2 samples showed enrichment for RNP complex components (Fig. 19A and 19B middle). In contrast, the β-actin-pull down samples showed an enrichment of ribosomal subunits (Fig. 19C middle). These results suggest that RNA-proximity biotinylation labels functional or regulatory proteins close to the site where the enzymes were tethered, in this case, the β-actin 3'UTR. In the β-actin-pull down method, however, it seems that the total mRNA bound proteome can be captured, indicated by the enrichment of ribosomal subunits that are predicted to be bound to the 5'UTR during scanning and the coding region during translation. Gene ontology (GO) description for the biological process (BP) in either RNA-BioID (Fig. 19A, left) or RNA-APEX2 (Fig. 19B, left) samples showed enrichment of proteins related to RNA processing activity whilst for β-actin-pull down (Fig. 19C left) samples the enriched terms are related to intracellular transport and protein localization. There are only 16

proteins that were present in all three methods including RBPs like IGF2BP1, IGF2BP2, STAU1, STAU2, FUBP3, YBOX1, SAM68, and PABP1. Amongst these candidates, most of the proteins were identified previously as β-actin 3'UTR binders (Eliscovich et al., 2017; Mukherjee et al., 2019). FUBP3 was also characterized as an RBP, which binds to β-actin mRNA and affects localization (Mukherjee et al., 2019).

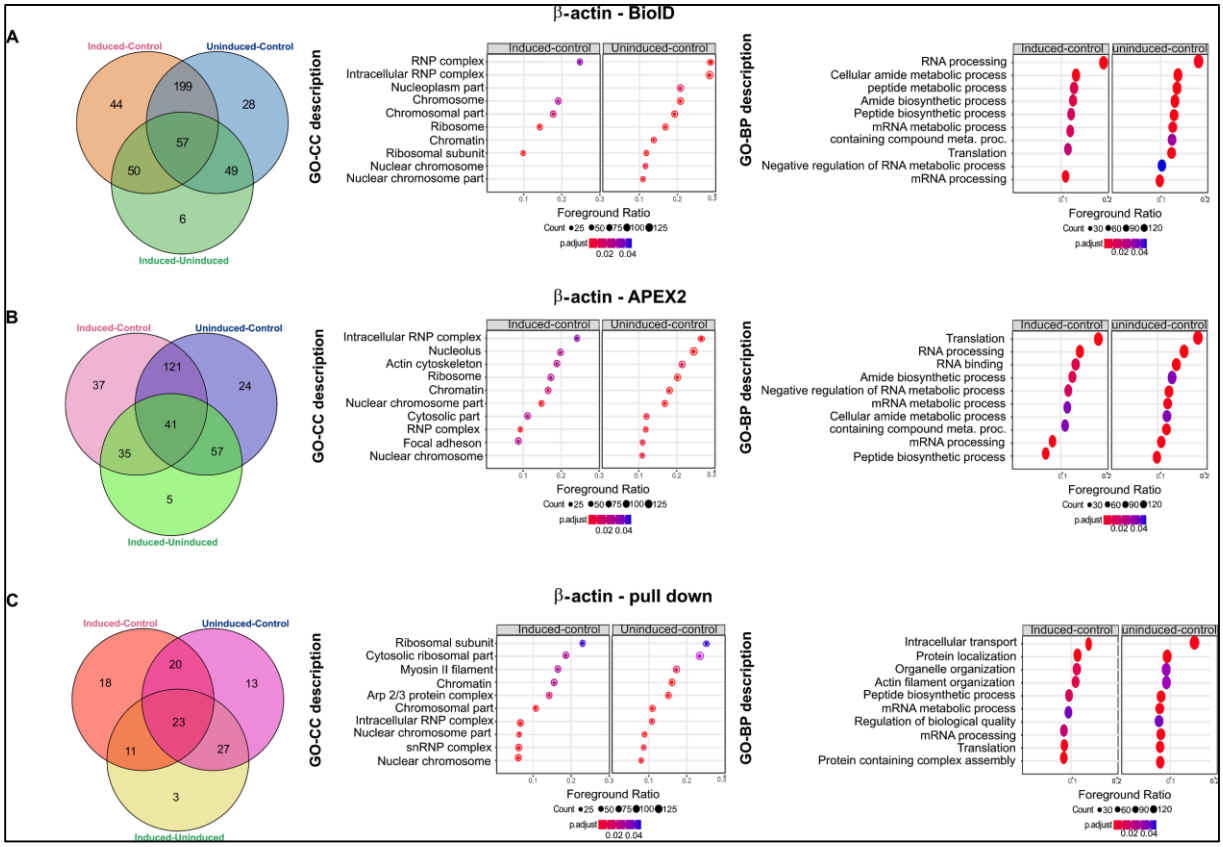


Fig. 19. Comparison of the identified RNA-interacting proteomes in RNA-proximity biotinylation and RNA pull-down.

(A) Left: Venn diagram representation of significantly enriched proteins in the β-actin-BioID method in MEFs (control/induced/uninduced). In total 341 protein was enriched in induced compared to control, 324 in uninduced compared to control and 116 in induced compared to uninduced. Middle to right: Gene Ontology (GO). Middle: Cellular Component. Right: Biological process. Both are displayed as over-representation analysis of the proteins enriched in uninduced and induced conditions compared to the control.

(B) Left: Venn diagram representation of significantly enriched proteins in the β-actin-APEX2 method in MEFs (control/induced/uninduced). In total 234 protein was enriched in induced compared to control, 243 in uninduced compared to control and 138 in induced compared to uninduced. Middle to right: Gene Ontology (GO). Middle: Cellular Component. Right: Biological process. Both are displayed as over-representation analysis of the proteins enriched in uninduced and induced conditions compared to the control.

(C) Left: Venn diagram representation of significantly enriched proteins in the β-actin-pull-down method in MEFs (control/induced/uninduced). In total 72 protein enriched in induced compared to control, 83 in uninduced compared to control and 64 in induced compared to uninduced. Middle to right: Gene Ontology (GO). Middle: Cellular Component. Right: Biological process. Both are displayed as over-representation analysis of the proteins enriched in uninduced and induced conditions compared to the control.
For all GO term analysis: The foreground ratio represents the number of significantly enriched proteins divided by the number of proteins within each GO term. The color gradient represents the adjusted p-value (threshold adj. p-value ≤ 0.05).

6.3 Characterization of FUBP3 as novel β-actin-associated RBP

Amongst other proteins, FUBP3, STAU1, and STAU2 are enriched in both RNA-proximity labeling and β-actin-pull down mass spec data. FUBP3 is a member of the far-upstream element (fuse) binding protein family (FBP) (Braddock et al., 2002), comprising of FBP1 (FUBP1), FBP2 (FUBP2/MARTA1/KHSRP/KSRP) and FBP3 (FUBP3/MARTA2) (Fig. 20A). All three FBP proteins contain four KH domains. FUBP1 and FUBP2 are predominantly nuclear, consistent with their feature of having a nuclear localization sequence (NLS) (shown as green, Fig. 20A). It has previously been shown that both FUBP2 and FUBP3 bind to the dendritic targeting element (DTE) of neuronal MAP2 mRNA (Rehbein et al., 2000). Interestingly, in the case of β-actin, it was also identified that chicken homolog of FUBP2, ZBP2 facilitates binding of ZBP1 to the β-actin localization element in the nucleus. In both RNA-BioID and RNA-APEX2, all three FBP isoforms were identified. The string interaction network analysis showed that all the three proteins might interact with each other while the interaction can be extended with PCBP1 (also known as hnRNP E1) and PCBP2 (also known as hnRNP E2) (Fig. 19B) that were also enriched in the proximity biotinylation mass spec data. Interestingly, FUBP3 and FUBP2 were also identified via β-actin-pull down, suggesting a stable interaction with β-actin mRNA. By RNA-dependent or -independent pull down assays, I also demonstrated that FUBP3 binds to STAU2 and IGF2BP1 binds to STAU1 in an RNA independent manner (Fig. 20C, see also Fig. 5B of MS1).

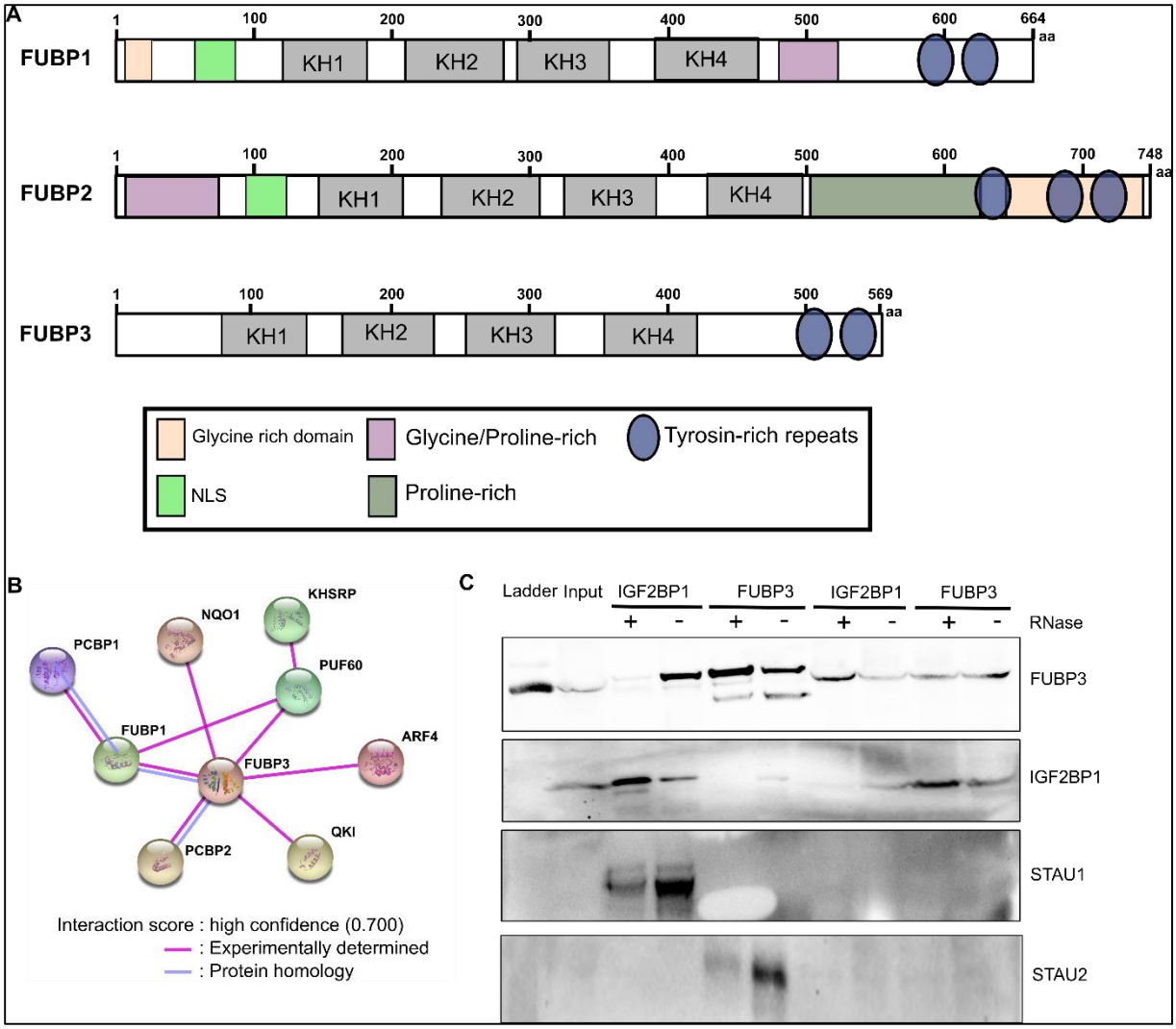


Fig. 20. Characterization of FUBP3.

(A) FUBP1 (top), FUBP2 (middle), and FUBP3 (bottom) all contain four KH domains (gray squares) along with the tyrosine rich C terminus (blue ovals)). In contrast to FUBP3, FUBP1, and FUBP2 contains a nuclear localization sequence (NLS) shown as green boxes. In addition to the NLS, both FUBP1 and 2 have glycine-rich (yellow boxes) and glycine/proline-rich stretches. Furthermore, FUBP2 has an additional proline-rich region before the tyrosine repeats.
(B) String interaction network of FUBP3 in mouse showed that experimentally it has determined that FUBP3 binds to several proteins (connected in pink lines) including FUBP1 and FUBP2. By protein homology modeling (connected with blue lines), FUBP3 has structural similarity with PCBP1 and PCBP2. **(C)** Co-Immunoprecipitation of FUBP3, IGF2BP1, and staufens. Immunoprecipitation was performed from WT MEFs with either anti-FUBP3 or anti-IGF2BP1 antibodies in the presence and absence of RNase A. IGF2BP1 coprecipitates with FUBP3 only in the absence of RNase A, while binding of STAU1 to IGF2BP1 and STAU2 to FUBP3 is RNA-independent.

6.4 FUBP3 binds to β-actin mRNA and affects β-actin mRNA localization

The association of FUBP3 with β-actin mRNA was confirmed both biochemically and microscopically. By *in situ* hybridization combined with immunofluorescence (smFISH-IF), followed by super registration microscopy (Eliscovich et al., 2017), it was identified that co-association between FUBP3 and β-actin mRNA is 29% (Fig. 4C, E, F of MS1). The association between the chosen positive control IGF2BP1 and β-actin mRNA was found to be 37% (Fig. 4B, E, and F of MS1). This result suggests that both IGF2BP1 and FUBP3 physically interact with β-actin mRNA. To determine the physical contact site of FUBP3 on the β-actin mRNA, a pull down of either FUBP3 or IGF2BP1 protein was performed. After isolating the associated mRNAs followed by qPCR analysis showed that, FUBP3 binds to 23% of total β-actin mRNA whereas IGF2BP1 binds to 37% of total β-actin mRNA (Fig. 5A of MS1). Secondly, to identify the potential FUBP3 binding site (UAUG, (Dominguez et al., 2018)) on β-actin mRNA His-tagged FUBP3 and IGF2BP1 were expressed in *E. coli*. *In vitro* transcribed RNA covering the full-length 643 bp β-actin UTR, the IGF2BP1 binding 49 bp zipcode, or a 79 bp long sequence spanning the potential FUBP3 binding site (Fig. 5D of MS1) containing RNA was added to the lysates containing His-tagged proteins. RNAs bound to affinity purified proteins were analyzed by qPCR, which confirmed that FUBP3 binds either the whole 3'UTR or the 79 bp long sequence containing potential FUBP3 binding motif. Deletion of the UAUG motif in the 79 bp long sequence resulted in the loss of the FUBP3 binding activity (Fig. 5C, and D of MS1). To validate the function of FUBP3 during β-actin localization to fibroblast protrusions after serum starvation, inducible FUBP3 knockdown MEFS were generated. Knockdown of FUBP3 (Fig. 6A, and B of MS1) resulted in a decrease in β-actin mRNA localization (measured by calculating the polarization index from the smFISH-IF analysis against β-actin and FUBP3, respectively) at the protrusion compared to the polarization found in WT MEFs (Fig. 6C, D, and G of MS1). Overexpression of shRNA resistant FUBP3 in FUBP3 knockdown cells led to the restoration of β-actin mRNA localization (Fig. 6E, and G of MS1). These results indicate that FUBP3 is important for β-actin mRNA localization.

7. Discussion

7.1 Identifying RNA-associated proteins

β-actin mRNA localization has been extensively studied in the last three decades starting from the first evidence of β-actin mRNA localization in ascidian eggs (Jeffery et al., 1983) to mammalian neurons (Turner-Bridger et al., 2018), revealing its evolutionarily conserved localization pattern in polarized cells. Multiple models have been created that depict why this mRNA localizes in most of the polarized cells, but how it localizes to the target region is not entirely understood. Biochemical and high-throughput studies in different model organisms have revealed only a few proteins that associate with this mRNA, but the complete interacting proteome has never been determined. One reason was that tools to identify the interacting proteome of a specific mRNA were missing. A widely used method in RNA interaction biochemistry is the direct pull down of the mRNA of interest via antisense mRNA oligonucleotides or via a protein which is bound to the mRNA of interest (Doron-Mandel et al., 2016; Slobodin and Gerst, 2010). These methods are limited by resolution as they mainly give information about the proteins which are strictly bound to the mRNA. Co-purification of even the strictly bound proteins with an mRNA can also be problematic because specific buffer conditions during lysis and purification can influence the RNA-protein interaction and also allow otherwise non-cognate proteins to bind the RNA (Mili and Steitz, 2004). These problems can be overcome by cross-linking the mRNP complexes in vivo with formaldehyde or UV light. Formaldehyde can link different nucleophilic molecular functional groups that are ~2 Å apart from each other, making it a widely used substrate to capture interactions between proximal proteins. The drawback of it is that the macromolecules have to be very close to each other, and the cross-linking is largely dependent on reaction conditions like pH and temperature (Hoffman et al., 2015). A UV light of approximately 260 nm wavelength allows the cross-linking of macromolecules with aromatic residues. The problem with UV cross-linking is that the cross-linked residues have to be in direct contact with each other. Since both crosslinking methods induce covalent bonds between RNA and protein, the cross-linking has to be reversed before identifying the proteins via mass spectroscopy.

To improve the limitations of studies concerning RNA-RBP interactions and to achieve a global view on the mRNA-protein interaction, proximity biotinylation tools provide a good option. BirA* or APEX2 mediated proximity biotinylation using an mRNA as bait can enhance the characterization of the dynamic RNA interacting protein complexes. Tethering a biotin ligase to an mRNA via MS2 tagging of the target mRNA should lead to capturing of the entire proteome of this mRNA during the time scale of the experiment. BirA* and APEX2 mediated biotinylation should allow labeling in the

vicinity of a 10 to 20 nm radius of the enzyme (Che and Khavari, 2017). This method allows us to explore the information of all the stable and transiently interacting factors and expands the information of the molecular network. The main difference between BirA* and APEX2 lies in their enzymatic features and substrate usage. The BirA* enzyme uses both biotin and ATP as substrates to generate reactive biotinoyl-5'-AMP (bioAMP). This reactive bioAMP is then released into the vicinity and can tag nearby lysine residues on proximal proteins (Roux et al., 2012). The reactive bio-AMP has a half-life of 1minute in water and likely less inside the cell due to the presence of a high-density of intracellular nucleophiles (Xu and Beckett, 1994). Depending on the labeling time, many copies of the proteins get biotinylated as many times they come close to the labeling enzyme throughout the labeling period (usually 6-24 hrs.). Therefore in the quantitative mass spectroscopy, the proteins are ranked based on proximity to the enzyme, although enzyme-proximal proteins with inaccessible lysine residue will remain unlabeled by this approach. APEX2 mediated labeling requires H2O2 and the substrate molecule, biotin-phenol, or its variants like BxxP, alkyne-phenol, or desthiobiotin phenol(Han et al., 2017). The peroxidase oxidizes the biotin phenol to a biotin radical that reacts with electron-rich side chains of nearby proteins. The resulting 10 - 20 nm labeling range depends on the half-life of the phenoxyl radicals, which usually is in the range of 1 ms. So within 1ms, the labeling intensity falls off a from the peroxidase active site generating a gradient of labeled proteins which is then analyzed by quantitative mass spectroscopy, which gives an idea about the proximity of the molecule to the enzyme (Fazal et al., 2019).

In principle, both enzymes can be used depending on the biological question being asked. BirA* can be used in vivo and in organoid models, as biotin can be supplemented via food or culture media. Being a small molecule, attachment of biotin to a protein should not change the functionality of the protein. However since APEX2 requires phenol and peroxidase reagents which are toxic to the cells, this is less suitable for in vivo or organoid models. In addition, biotin phenol has higher hydrophobicity than bio-AMP, which can affect the bioavailability of the substrate in specific cellular regions Due to the long labeling time for BirA* the dynamic of the enzyme tagged molecule cannot be analyzed. In the case of APEX2, the labeling time is much shorter although this might mean that it has to enrich labeled proteins for mass spec analysis, on the other hand, quick labeling time allows to study dynamic changes of the enzyme attached molecules. The sample processing stages are more cumbersome for APEX2 compared to BirA* especially during the required quenching reaction to inactivate the biotin phenol.

To use of proximity biotinylating enzymes with a (localized) mRNA requires their tethering, best-using RNA aptamers like MS2 tags in their 3'UTR. For β-actin mRNA, a mouse was generated (Lionnet et al., 2011) where both copies of the endogenous β-

actin mRNA were tagged with 24MS2 loops at 441 bp downstream of the stop codon in the β-actin gene loci (β-actin-MBS). The mouse was then cross-bred with MCP transgenic mouse (MCP-GFP express in every cell) to generate a double homozygous β-actin-MBS X MCP mouse where, due to binding of two copies of MCP to each MS2 hairpin, under an ideal scenario all the β-actin mRNAs are labeled with 48 MCP-GFP protein molecules (Park et al., 2014). The benefit of the system is that it keeps an endogenous level of β-actin mRNA expression. Via qRT-PCR and northern blots it was shown that the addition of the loops does not have any deleterious effect on the β-actin mRNA (Kim et al., 2019; Lionnet et al., 2011). I used mouse embryonic fibroblasts (MEFs) from a β-actin-MBS mouse and expressed a fusion of two MCPs, eGFP and BirA or APEX2 (2MCP-eGFP-BirA*/APEX2) in these cells. So, in an ideal scenario, there are 24 BirA* or APEX2 tethered per mRNA molecule. If this represents the minimal number of enzymes that needs to be tethered has not been tested though I found no difference in the labeling efficiency and the mass-spectroscopy data with 24 (by expressing a fusion of two copies of MCP-one GFP and one BirA*) and 48 (by expressing a fusion of one copy of MCP-one GFP) and one BirA*) copy of BirA* on β-actin mRNA.

In my thesis work, I could identify more beta-actin mRNA interacting proteins in the proximity labeling assays (BioID or APEX2) than in MS2-pull down affinity purification assay. A likely reason is the enrichment of transiently interacting proteins in proximity labeling.

7.2 RNA proximity biotinylation and RNA pull down capture both stable and transient binders

The number of identified potential β-actin interactors is higher in the BioID or APEX2 method than the affinity co-purification approach, and this is due to the higher sensitivity of proximity labeling (Varnaitė and MacNeill, 2016). Another reason why fewer proteins were identified by the β-actin -pull down might be that this method preferentially identifies direct β-actin mRNA binding proteins. Interestingly, no matter what method I used, I always identified more proteins under induced than uninduced conditions, which likely represents an increase in the number of proteins assembled after serum induction of the mRNA. A detailed description of previously identified β-actin mRNA interacting proteins has been made in the introduction section under the β-actin mRNA localization in the fibroblast section. Both β-actin proximity biotinylation and MS2 pull-down resulted in a common set of proteins, IGF2BP1, STAU1, FUBP3, and FUBP1 as well as additional RBPs. This suggests that these RBPs are stable binders of the mRNA and important trans-acting factors of the β-actin mRNA localization pathway. In contrast, FMR1, SMN1, KHDRBS1, and ELAVl1 were identified only using proximity labeling tools, which indicates that either these are rather

transient or less stable interactors of the mRNA or these proteins are a part of a bigger complex which is difficult to isolate via affinity purification. Interestingly, I could not identify KIF11 protein as mentioned before (Song et al., 2015) to be associated with IGF2BP1 and a potential role in mRNP delivery.

Interestingly another motor protein, the myosin heavy chain MYH9 was identified in the proximity labeling methods but MYH10 was present in the β-actin pull down.

7.3 Future improvement of RNA-based proximity labeling system

RNA-proximity biotinylation is currently a state of the art tool to identify an mRNA interactome. However, there are certain limitations to this technique. Firstly, unless targeted by dCAS9-BioID fusion (Cas-ID) (Schmidtmann et al., 2016), the endogenous mRNA has to be tagged with MS2 loops. Instead of generating an MS2 knock-in mouse, one can utilize the benefit of CRISPR like editing tools to insert MS2 loops at a specific location of the 3'UTR of target mRNA in any cells. Secondly, the choice of the biotinylating enzyme can influence the identified proteome. As mentioned before (see chapter 3.3.1 in the introduction), BirA* is easy to use, can be used in living cells, but the labeling time is long, which is a drawback when trying to understand the dynamics of an mRNA. APEX2 labeling is fast enough to allow dynamic studies but the sample handling is not as user-friendly as for BirA*. To overcome these limitations, one could use proximity labeling enzymes with faster activity. New proximity labeling enzymes have been generated that have many benefits. For example, BioID2 (27 kDa) is a smaller version of BirA* (35.4 kDa) that requires less biotin (1-3.2 µM instead of 50 µM in the culture medium) and has an enhanced labeling efficiency (Kim et al., 2016). BASU (a C-terminal mutated version of *Bacillus subtilis* biotin ligase) needs 1 hr labeling time and 200 µM biotin for labeling proximal proteins (Ramanathan et al., 2018). However, 1 hr labeling timing is still a bit long for identifying dynamic RNPs. In order to decrease the labeling time, an alternative is TurboID or miniTurboID (Branon et al., 2018). TurboID (35.4 kDa) is another version of the BirA* enzyme with 14 additional mutations and requires only 10 min of labeling time. miniTurboID (28 kDa) is an N-terminal deleted version of BirA* with 13 similar mutations like TurboID. The miniTurboID is 1.5-2 fold less active than TurboID though they show less background labeling before the addition of biotin which makes it useful for precise temporal control of the labeling time (Branon et al., 2018). So based on the purpose and localization of the mRNA, one can tether other proximity labeling enzyme via MCP on any RNA for mRNA interactome analysis

8. References

Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E. and Carson, J. H. (1997). Transport and localization elements in myelin basic protein mRNA. *J. Cell Biol.* **138**, 1077–1087.

Aït-Ahmed, O., Bellon, B., Capri, M., Joblet, C. and Thomas-Delaage, M. (1992). The yemanuclein-alpha: a new Drosophila DNA binding protein specific for the oocyte nucleus. *Mech. Dev.* **37**, 69–80.

Allain, F. H., Bouvet, P., Dieckmann, T. and Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870–81.

Anderson, C. M., Zhang, B., Miller, M., Butko, E., Wu, X., Laver, T., Kernag, C., Kim, J., Luo, Y., Lamparski, H., et al. (2016). Fully Automated RNAscope In Situ Hybridization Assays for Formalin-Fixed Paraffin-Embedded Cells and Tissues. *J. Cell. Biochem.* **117**, 2201–2208.

Anhäuser, L., Hüwel, S., Zobel, T. and Rentmeister, A. (2019). Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly(A) tail enhances translation and can be performed in living cells. *Nucleic Acids Res.* **47**, e42.

Backe, P. H., Messias, A. C., Ravelli, R. B. G., Sattler, M. and Cusack, S. (2005). X-Ray Crystallographic and NMR Studies of the Third KH Domain of hnRNP K in Complex with Single-Stranded Nucleic Acids. *Structure* **13**, 1055–1067.

Bandziulis, R. J., Swanson, M. S. and Dreyfuss, G. (1989). RNA-binding proteins as developmental regulators. *Genes Dev.* **3**, 431–437.

Banerjee, S. and Barraud, P. (2014). Functions of double-stranded RNA-binding domains in nucleocytoplasmic transport. *RNA Biol.* **11**, 1226–1232.

Bann, D. V and Parent, L. J. (2012). Application of live-cell RNA imaging techniques to the study of retroviral RNA trafficking. *Viruses* **4**, 963–979.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–33.

Bashirullah, A., Cooperstock, R. L. and Lipshitz, H. D. (2002). RNA LOCALIZATION IN DEVELOPMENT. *Annu. Rev. Biochem.* **67**, 335–394.

Bassell, G. J., Zhang, H., Byrd, a L., Femino, a M., Singer, R. H., Taneja, K. L., Lifshitz, L. M., Herman, I. M. and Kosik, K. S. (1998). Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. *J. Neurosci.* **18**, 251–265.

Beckett, D., Kovaleva, E. and Schatz, P. J. (2008). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929.

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y. and Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10.

Berkovits, B. D. and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. and Long, R. M. (1998). Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445.

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., de Crécy-Lagard, V., Ross, R., Limbach, P. A., Kotter, A., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307.

Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D. and Clore, G. M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**, 1051–1056.

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N. and Ting, A. Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–898.

Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269.

Brodsky, A. S. and Silver, P. A. (2002). Identifying proteins that affect mRNA localization in living cells. *Methods* **26**, 151–155.

Bullock, S. L. (2012). Messengers, motors and mysteries: sorting of eukaryotic mRNAs by cytoskeletal transport. *Biochem. Soc. Trans.* **39**, 1161–1165.

Bunnell, T. M., Burbach, B. J., Shimizu, Y. and Ervasti, J. M. (2011). Beta-Actin specifically controls cell growth, migration, and the G-actin pool. *Mol. Biol. Cell* **22**, 4047–4058.

Burns, L. T. and Wente, S. R. (2014). From hypothesis to mechanism: uncovering nuclear pore complex links to gene expression. *Mol. Cell. Biol.* **34**, 2114–20.

Buxbaum, A. R., Wu, B. and Singer, R. H. (2014). Single β-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **343**, 419–22.

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J. L. and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20.

Cajigas, I. J., Tushev, G., Will, T. J., Tom Dieck, S., Fuerst, N. and Schuman, E. M. (2012). The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. *Neuron* **74**, 453–466.

Carey, J. and Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry* **22**, 2610–2615.

Carmody, S. R. and Wente, S. R. (2009). mRNA nuclear export at a glance. *J. Cell Sci.* **122**, 1933–7.

Ceci, M., Welshhans, K., Ciotti, M. T., Brandi, R., Parisi, C., Paoletti, F., Pistillo, L., Bassell, G. J. and Cattaneo, A. (2012). RACK1 is a ribosome scaffold protein for β-actin mRNA/ZBP1 complex. *PLoS One* **7**, e35034.

Chapman-Smith, A. and Cronan Jr, J. E. (1999). Molecular Biology of Biotin Attachment to Proteins. *J. Nutr.* **129**, 477S-484S.

Chapman, K. B. and Boeke, J. D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* **65**, 483–492.

Che, Y. and Khavari, P. A. (2017). Research Techniques Made Simple: Emerging Methods to Elucidate Protein Interactions through Spatial Proximity. *J. Invest. Dermatol.* **137**, e197–e203.

Chen, J., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C. E., Coffin, J. M., Ni, N., Lockett, S., Pathak, V. K. and Hu, W. S. (2009). High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13535–13540.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J. L., Dias, A. and Reed, R. (2006). Human mRNA Export Machinery Recruited to the 5′ End of mRNA. *Cell* **127**, 1389–1400.

Cheung, S. W., Tishler, P. V, Atkins, L., Sengupta, S. K., Modest, E. J. and Forget, B. G. (1977). Gene mapping by fluorescent in situ hybridization. *Cell Biol. Int. Rep.* **1**, 255–62.

Chou, Y. ying, Heaton, N. S., Gao, Q., Palese, P., Singer, R. and Lionnet, T. (2013). Colocalization of Different Influenza Viral RNA Segments in the Cytoplasm before Viral Budding as Shown by Single-molecule Sensitivity FISH Analysis. *PLoS Pathog.* **9**, e1003358.

Claussen, M. (2004). Evidence for overlapping, but not identical, protein machineries operating in vegetal RNA localization along early and late pathways in Xenopus oocytes.

Development **131**, 4263–4273.

Clery, A. and Allain, F. H. T. (2015). From Structure to Function of RNA Binding Domains BT - RNA Binding Proteins. In *RNA Binding Proteins*, pp. 137–158. Landes Bioscience.

Cléry, A., Blatter, M. and Allain, F. H.-T. (2008). RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298.

Cui, X. A. and Palazzo, A. F. (2014). Localization of mRNAs to the endoplasmic reticulum. *Wiley Interdiscip. Rev. RNA* **5**, 481–492.

Cullen, B. R. (2000). Nuclear RNA export pathways. *Mol. Cell. Biol.* **20**, 4181–7.

Desrosiers, R., Friderici, K. and Rottman, F. (1974). Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci.* **71**, 3971–3975.

Di Liegro, C. M., Schiera, G. and Di Liegro, I. (2014). Regulation of mRNA transport, localizationand translation in the nervous system of mammals (Review). *Int. J. Mol. Med.* **33**,.

Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Arcelus, M. G., Sekowska, M., et al. (2009). Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. *Science (80-.).* **325**, 1246–1250.

Ding, D., Parkhurst, S. M., Halsell, S. R. and Lipshitz, H. D. (1993). Dynamic Hsp83 RNA localization during Drosophila oogenesis and embryogenesis. *Mol. Cell. Biol.* **13**, 3773–81.

Doma, M. K. and Parker, R. (2007). RNA quality control in eukaryotes. *Cell* **131**, 660–8.

Dominguez, D., Freese, P., Alexis, M. S., Yeo, G. W., Graveley, B. R. and Burge, C. B. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70**, 854–867.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206.

Dormoy-Raclet, V., Ménard, I., Clair, E., Kurban, G., Mazroui, R., Di Marco, S., von Roretz, C., Pause, A. and Gallouzi, I.-E. (2007). The RNA-binding protein HuR promotes cell migration and cell invasion by stabilizing the beta-actin mRNA in a U-rich-element-dependent manner. *Mol. Cell. Biol.* **27**, 5365–5380.

Doron-Mandel, E., Alber, S., Oses, J. A., Medzihradszky, K. F., Burlingame, A. L., Fainzilber, M., Twiss, J. L. and Lee, S. J. (2016). Isolation and analyses of axonal ribonucleoprotein complexes. *Methods Cell Biol.* **131**, 467–486.

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.

Edelmann, F. T., Schlundt, A., Heym, R. G., Jenner, A., Niedner-Boblenz, A., Syed, M. I., Paillart, J. C., Stehle, R., Janowski, R., Sattler, M., et al. (2017). Molecular architecture and dynamics of ASH1 mRNA recognition by its mRNA-transport complex. *Nat. Struct. Mol. Biol.* **24**, 152–161.

Eldar, A. and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature* **467**, 167–173.

Eliscovich, C., Buxbaum, A. R., Katz, Z. B. and Singer, R. H. (2013). mRNA on the Move: The Road to Its Biological Destiny. *J. Biol. Chem.* **288**, 20361–20368.

Eliscovich, C., Shenoy, S. M. and Singer, R. H. (2017). Imaging mRNA and protein interactions within neurons. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1875–E1884.

Elkon, R., Ugalde, A. P. and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506.

Engel, C., Neyer, S. and Cramer, P. (2018). Distinct Mechanisms of Transcription Initiation

by RNA Polymerases I and II. *Annu. Rev. Biophys.* **47**, 425–446.
Eom, T., Antar, L. N., Singer, R. H. and Bassell, G. J. (2003). Localization of a beta-actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses. *J. Neurosci.* **23**, 10433–44.
Farina, K. L., Hüttelmaier, S., Musunuru, K., Darnell, R. and Singer, R. H. (2003). Two ZBP1 KH domains facilitate β-actin mRNA localization, granule formation, and cytoskeletal attachment. *J. Cell Biol.* **160**, 77–87.
Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., Chang, H. Y. and Ting, A. Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26.
Fischer, M., Kaech, S., Wagner, U., Brinkhaus, H. and Matus, A. (2000). Glutamate receptors regulate actin-based plasticity in dendritic spines. *Nat. Neurosci.* **3**, 887–894.
Friedman, R. C., Farh, K. K.-H., Burge, C. B. and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105.
Galbraith, M. D., Allen, M. A., Bensard, C. L., Wang, X., Schwinn, M. K., Qin, B., Long, H. W., Daniels, D. L., Hahn, W. C., Dowell, R. D., et al. (2013). HIF1A Employs CDK8-Mediator to Stimulate RNAPII Elongation in Response to Hypoxia. *Cell* **153**, 1327–1339.
Gall, J. G. and Pardue, M. L. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. U. S. A.* **63**, 378–383.
Galloway, A. and Cowling, V. H. (2019). mRNA cap regulation in mammalian cell function and fate. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862**, 270–279.
Garneau, N. L., Wilusz, J. and Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126.
Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–6.
Gilbert, W. (1978). Why genes in pieces. *Nature* **271**, 501–501.
Grishin, N. V. (2001). KH domain: one motif, two folds. *Nucleic Acids Res.* **29**, 638–643.
Gudipati, R. K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L. M., Jacquier, A. and Libri, D. (2012). Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Mol. Cell* **48**, 409–421.
Haag, J. R. and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**, 483–492.
Hamilton, R. S. and Davis, I. (2011). Identifying and searching for conserved RNA localisation signals. *Methods Mol. Biol.* **714**, 447–466.
Han, S., Udeshi, N. D., Deerinck, T. J., Svinkina, T., Ellisman, M. H., Carr, S. A. and Ting, A. Y. (2017). Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem. Biol.* **24**, 404–414.
Herve, C. (2006). Zipcodes and postage stamps: mRNA localisation signals and their trans-acting binding proteins. *Briefings Funct. Genomics Proteomics* **3**, 240–256.
Hill, M. A. and Gunning, P. (1993). Beta and gamma actin mRNAs are differentially located within myoblasts. *J. Cell Biol.* **122**, 825–832.
Hill, M. A., Schedlich, L. and Gunning, P. (1994). Serum-induced signal transduction determines the peripheral location of β-actin mRNA within the cell. *J. Cell Biol.* **126**, 1221–1230.
Hocine, S., Singer, R. H. and Grünwald, D. (2010). RNA processing and export. *Cold Spring Harb. Perspect. Biol.* **2**, a000752.
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**, 26404–11.
Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R. and

Ramos, A. (2012). KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886.

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

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. and Zhao, Y. (2014). SnapShot: Histone Modifications. *Cell* **159**, 458-458.e1.

Hüttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G. J., Condeelis, J. and Singer, R. H. (2005). Spatial regulation of β-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438**, 512–515.

Inose, H., Mukai, K., Ito, M. and Masuda, S. (2015). Gene Regulation through mRNA Expression. *Adv. Biol. Chem.* **05**, 45–57.

Insall, R., Müller-Taubenberger, A., Machesky, L., Köhler, J., Simmeth, E., Atkinson, S. J., Weber, I. and Gerisch, G. (2001). Dynamics of the *Dictyostelium* Arp2/3 complex in endocytosis, cytokinesis, and chemotaxis. *Cell Motil. Cytoskeleton* **50**, 115–128.

Itoh, M., Haga, I., Li, Q. H. and Fujisawa, J. I. (2002). Identification of cellular mRNA targets for RNA-binding protein Sam68. *Nucleic Acids Res.* **30**, 5452–5464.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

Jambhekar, A. and DeRisi, J. L. (2007). Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA* **13**, 625–642.

Jeffery, W. R., Tomlinson, C. R. and Brodeur, R. D. (1983). Localization of actin messenger RNA during early ascidian development. *Dev. Biol.* **99**, 408–417.

Johansson, H. E., Liljas, L. and Uhlenbeck, O. C. (1997). RNA recognition by the MS2 phage coat protein. *Semin. Virol.* **8**, 176–185.

Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N. and Nakahara, H. (2003). Structure-stability-function relationships of dendritic spines. *Trends Neurosci.* **26**, 360–8.

Kastner, B., Will, C. L., Stark, H. and Lührmann, R. (2019). Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol.* a032417.

Katahira, J. (2015). Nuclear export of messenger RNA. *Genes (Basel).* **6**, 163–184.

Katz, Z. B., Wells, A. L., Park, H. Y., Wu, B., Shenoy, S. M. and Singer, R. H. (2012). β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration. *Genes Dev.* **26**, 1885–1890.

Ke, S., Alemu, E. A., Mertens, C., Gantman, E. C., Fak, J. J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M. J., Park, C. Y., et al. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* **29**, 2037–53.

Kerber, M. L. and Cheney, R. E. (2011). Myosin-X: a MyTH-FERM myosin at the tips of filopodia. *J. Cell Sci.* **124**, 3733–41.

Kim, J. H. and Richter, J. D. (2006). Opposing Polymerase-Deadenylase Activities Regulate Cytoplasmic Polyadenylation. *Mol. Cell* **24**, 173–183.

Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K. and Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196.

Kim, S. H., Vieira, M., Kim, H. J., Kesawat, M. S. and Park, H. Y. (2019). MS2 Labeling of Endogenous Beta-Actin mRNA Does Not Result in Stabilization of Degradation Intermediates. *Mol. Cells* **42**, 356–362.

King, M. Lou, Messitt, T. J. and Mowry, K. L. (2005). Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biol. Cell* **97**, 19–33.

Kloc, M., Zearfoss, N. R. and Etkin, L. D. (2002). Mechanisms of subcellular mRNA localization. *Cell* **108**, 533–544.

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA.

Science (80-.). **184**, 868–871.

Kung, C.-P., Maggi, L. B. and Weber, J. D. (2018). The Role of RNA Editing in Cancer Development and Metabolic Disorders. *Front. Endocrinol. (Lausanne).* **9**, 762.

Kuss, S. K., Mata, M. A., Zhang, L. and Fontoura, B. M. A. (2013). Nuclear imprisonment: Viral strategies to arrest host mRNA nuclear export. *Viruses* **5**, 1824–1849.

Kwon, K. and Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci.* **9**, 1530–1539.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Lange, S., Katayama, Y., Schmid, M., Burkacky, O., Brauchle, C., Lamb, D. C. and Jansen, R. P. (2008). Simultaneous transport of different localized mRNA species revealed by live-cell imaging. *Traffic* **9**, 1256–1267.

Langer, P. R., Waldrop, A. A. and Ward, D. C. (1981). Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6633–6637.

Latham, V. M. J., Kislauskis, E. H., Singer, R. H. and Ross, A. F. (1994). Beta-actin mRNA localization is regulated by signal transduction mechanisms. *J. Cell Biol.* **126**, 1211–1219.

Lee, Y. and Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84**, 291–323.

Legault, P., Li, J., Mogridge, J., Kay, L. E. and Greenblatt, J. (1998). NMR structure of the bacteriophage λ N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289–299.

Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L. E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* **21**, 3536–45.

Lerner, R. S. and Nicchitta, C. V (2006). mRNA translation is compartmentalized to the endoplasmic reticulum following physiological inhibition of cap-dependent translation. *RNA* **12**, 775–789.

Lewis, C. J. T., Pan, T. and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* **18**, 202–210.

Lim, F. and Peabody, D. S. (2002). RNA recognition site of PP7 coat protein. **30**, 4138–4144.

Lima, S. A., Chipman, L. B., Nicholson, A. L., Chen, Y.-H., Yee, B. A., Yeo, G. W., Coller, J. and Pasquinelli, A. E. (2017). Short poly(A) tails are a conserved feature of highly expressed genes. *Nat. Struct. Mol. Biol.* **24**, 1057–1063.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A. L., Chao, J. A., Park, H. Y., De Turris, V., Lopez-Jones, M. and Singer, R. H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8**, 165–170.

Lipshitz, H. D. and Smibert, C. A. (2000). Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genet. Dev.* **10**, 476–488.

Liu, Y., Le, P., Lim, S. J., Ma, L., Sarkar, S., Han, Z., Murphy, S. J., Kosari, F., Vasmatzis, G., Cheville, J. C., et al. (2018). Enhanced mRNA FISH with compact quantum dots. *Nat. Commun.* **9**, 4461.

Lowary, P. T. and Uhlenbeck, O. C. (1987). An RNA mutation that increases the affinity of an RNA-protein interaction. *Nucleic Acids Res.* **15**, 10483–93.

Luse, D. S. (2013). The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? *Transcription* **5**, e27050.

Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H. and Ting, A. Y. (2012). Engineered ascorbate peroxidase as a

genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148.

Martin, K. C. and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–30.

Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. and Cusack, S. (2001). Crystal structure of the human nuclear cap binding complex. *Mol. Cell* **8**, 383–96.

Medioni, C., Mowry, K. and Besse, F. (2012). Principles and roles of mRNA localization in animal development. *Development* **139**, 3263–76.

Meyer, K. D. and Jaffrey, S. R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326.

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E. and Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149**, 1635–1646.

Mili, S. and Steitz, J. A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694.

Ming, G. L., Wong, S. T., Henley, J., Yuan, X. B., Song, H. J., Spitzer, N. C. and Poo, M. M. (2002). Adaptation in the chemotactic guidance of nerve growth cones. *Nature* **417**, 411–418.

Moore, M. J. (2005). From Birth to Death: The Complex Lives of Eukaryotic mRNAs. *Science (80-.).* **309**, 1514–1518.

Moore, L. D., Le, T. and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38.

Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B. and Jansen, R. P. (2019). β-Actin mRNA interactome mapping by proximity biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12863–12872.

Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J. and Pastore, A. (1996). Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85**, 237–45.

Mutka, S. C. and Walter, P. (2001). Multifaceted Physiological Response Allows Yeast to Adapt to the Loss of the Signal Recognition Particle-dependent Protein-targeting Pathway. *Mol. Biol. Cell* **12**, 577–588.

Nakel, K., Hartung, S. A., Bonneau, F., Eckmann, C. R. and Conti, E. (2010). Four KH domains of the C. elegans Bicaudal-C ortholog GLD-3 form a globular structural platform. *RNA* **16**, 2058–2067.

Nicastro, G., Taylor, I. A. and Ramos, A. (2015). KH–RNA interactions: back in the groove. *Curr. Opin. Struct. Biol.* **30**, 63–70.

Nicastro, G., Candel, A. M., Uhl, M., Oregioni, A., Hollingworth, D., Backofen, R., Martin, S. R. and Ramos, A. (2017). Mechanism of β-actin mRNA Recognition by ZBP1. *Cell Rep.* **18**, 1187–1199.

Nilsen, T. W. and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–63.

Nishikura, K. (2010). Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79**, 321–349.

Niu, G. and Chen, X. (2009). The Role of Molecular Imaging in Drug Delivery. *Drug Deliv. (London, England. 2007)* **3**, 109–113.

Oddone, A., Lorentzen, E., Basquin, J., Gasch, A., Rybin, V., Conti, E. and Sattler, M. (2007). Structural and biochemical characterization of the yeast exosome component Rrp40. *EMBO Rep.* **8**, 63–9.

Palacios, I. M. (2007). How does an mRNA find its way? Intracellular localisation of transcripts. *Semin. Cell Dev. Biol.* **18**, 163–70.

Park, N. J., Tsao, D. C. and Martinson, H. G. (2004). The Two Steps of Poly(A)-Dependent

Termination, Pausing and Release, Can Be Uncoupled by Truncation of the RNA Polymerase II Carboxyl-Terminal Repeat Domain. *Mol. Cell. Biol.* **24**, 4092–4103.

Park, H. Y., Lim, H., Yoon, Y. J., Follenzi, A., Nwokafor, C., Lopez-Jones, M., Meng, X. and Singer, R. H. (2014). Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **343**, 422–4.

Park, J.-E., Yi, H., Kim, Y., Chang, H., Narry, V. and Correspondence, K. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle Genes with the TOP element escape translational suppression in M phase. *Mol. Cell* **62**, 462–471.

Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *Embo J* **12**, 595–600.

Plet, A., Eick, D. and Blanchard, J. M. (1995). Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene* **10**, 319–28.

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (80-.).* **352**, aad9926.

Rach, E. A., Winter, D. R., Benjamin, A. M., Corcoran, D. L., Ni, T., Zhu, J. and Ohler, U. (2011). Transcription Initiation Patterns Indicate Divergent Strategies for Gene Regulation at the Chromatin Level. *PLoS Genet.* **7**, e1001274.

Rackham, O. and Brown, C. M. (2004). Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J.* **23**, 3346–3355.

Raghavan, A., Ogilvie, R. L., Reilly, C., Abelson, M. L., Raghavan, S., Vasdewani, J., Krathwohl, M. and Bohjanen, P. R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Res.* **30**, 5529–5538.

Ramanathan, A., Robb, G. B. and Chan, S. H. (2016). mRNA capping: Biological functions and applications. *Nucleic Acids Res.* **44**, 7511–7526.

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., et al. (2018). RNA–protein interaction detection in living cells. *Nat. Methods*.

Reed, R. and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523–31.

Rehbein, M., Kindler, S., Horke, S. and Richter, D. (2000). Two trans-acting rat-brain proteins, MARTA1 and MARTA2, interact specifically with the dendritic targeting element in MAP2 mRNAs. *Brain Res. Mol. Brain Res.* **79**, 192–201.

Reid, D. W. and Nicchitta, C. V. (2015). Comment on “principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling.” *Science (80-.).* **348**, 1217-a.

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A. and Ting, A. Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331.

Ridley, A. J. and Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell*.

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L. and Singer, R. H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. *Mol. Cell. Biol.* **17**, 2158–65.

Rossetto, D., Avvakumov, N. and Côté, J. (2012). Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098–1108.

Rougemaille, M., Villa, T., Gudipati, R. K. and Libri, D. (2008). mRNA journey to the cytoplasm: attire required. *Biol. Cell* **100**, 327–342.

Roux, K. J., Kim, D. I., Raida, M. and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810.

Saunders, C. and Cohen, R. S. (1999). The role of oocyte transcription, the 5'UTR, and translation repression and derepression in Drosophila gurken mRNA and protein localization. *Mol. Cell* **3**, 43–54.

SAUNDERS, L. R. and BARBER, G. N. (2003). The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J.*

Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: A 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Bio/Technology* **11**, 1138–1143.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. and Leonhardt, H. (2016). Determination of local chromatin composition by CasID. *Nucleus* **7**, 476–484.

Schubert, V. and Dotti, C. G. (2007). Transmitting on actin: synaptic control of dendritic architecture. *J. Cell Sci.* **120**, 205–212.

Shav-Tal, Y. and Singer, R. H. (2005). RNA localization. *J. Cell Sci.* **118**, 4077–4081.

Shearwin, K. E., Callen, B. P. and Egan, J. B. (2005). Transcriptional interference--a crash course. *Trends Genet.* **21**, 339–45.

Shestakova, E. A., Wyckoff, J., Jones, J., Singer, R. H. and Condeelis, J. (1999). Correlation of β-Actin Messenger RNA Localization with Metastatic Potential in Rat Adenocarcinoma Cell Lines. *Cancer Res.* **59**, 1202–1205.

Shestakova, E. A., Singer, R. H. and Condeelis, J. (2001). The physiological significance of beta -actin mRNA localization in determining cell polarity and directional motility. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7045–50.

Shilatifard, A. (2006). Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annu. Rev. Biochem.* **75**, 243–269.

Shlyueva, D., Stampfel, G. and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286.

Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Res.* **21**, 1193–1198.

Slobodin, B. and Gerst, J. E. (2010). A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* **16**, 2277–2290.

Song, M.-G., Li, Y. and Kiledjian, M. (2010). Multiple mRNA Decapping Enzymes in Mammalian Cells. *Mol. Cell* **40**, 423–432.

Song, T., Zheng, Y., Wang, Y., Katz, Z., Liu, X., Chen, S., Singer, R. H. and Gu, W. (2015). Specific interaction of KIF11 with ZBP1 regulates the transport of β-actin mRNA and cell motility. *J. Cell Sci.* **128**, 1001–1010.

Stefl, R., Skrisovska, L. and Allain, F. H. T. (2005). RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.* **6**, 33–38.

Sun, Y., Davis, P., Kosmacek, E. A., Ianzini, F. and MacKey, M. A. (2009). An open-source deconvolution software package for 3-D quantitative fluorescence microscopy imaging. *J. Microsc.* **236**, 180–193.

Sundell, C. L. and Singer, R. H. (1991). Requirement of microfilaments in sorting of actin messenger RNA. *Science* **253**, 1275–7.

Thio, G. L., Ray, R. P., Barcelo, G. and Schüpbach, T. (2000). Localization of gurken RNA in Drosophila Oogenesis Requires Elements in the 5′ and 3′ Regions of the Transcript. *Dev. Biol.* **221**, 435–446.

Tian, B. and Manley, J. L. (2013). Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.* **38**, 312–320.

Topisirovic, I., Svitkin, Y. V., Sonenberg, N. and Shatkin, A. J. (2011). Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* **2**, 277–298.

Trinkle-Mulcahy, L. (2019). Recent advances in proximity-based labeling methods for interactome mapping. *F1000Research* **8**, 135.

Turner-Bridger, B., Jakobs, M., Muresan, L., Wong, H. H.-W., Franze, K., Harris, W. A. and Holt, C. E. (2018). Single-molecule analysis of endogenous β-actin mRNA trafficking reveals a mechanism for compartmentalized mRNA localization in axons. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9697–E9706.

Tutucci, E., Livingston, N. M., Singer, R. H. and Wu, B. (2018). Imaging mRNA In Vivo, from Birth to Death. *Annu. Rev. Biophys.* **47**, 85–106.

Tyagi, S. and Kramer, F. R. (1996). Molecular beacon probes that fluoresce on hybridiztion. *Nat. Publ. Gr.* **14**, 303–308.

Urbanek, M. O., Galka-Marciniak, P., Olejniczak, M. and Krzyzosiak, W. J. (2014). 1083-1095 RNA imaging in living cells - Methods and applications. *RNA Biol.* **11**, 1083–1095.

Urbanska, A. S., Janusz-Kaminska, A., Switon, K., Hawthorne, A. L., Perycz, M., Urbanska, M., Bassell, G. J. and Jaworski, J. (2017). ZBP1 phosphorylation at serine 181 regulates its dendritic transport and the development of dendritic trees of hippocampal neurons. *Sci. Rep.* **7**, 1876.

Valverde, R., Edwards, L. and Regan, L. (2008). Structure and function of KH domains. *FEBS J.* **275**, 2712–2726.

Varnaitė, R. and MacNeill, S. A. (2016). Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics* **16**, 2503–2518.

Wächter, K., Köhn, M., Stöhr, N. and Hüttelmaier, S. (2013). Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. *Biol. Chem.* **394**, 1077–1090.

Walter, P. and Blobel, G. (1983). Subcellular distribution of signal recognition particle and 7SL-RNA determined with polypeptide-specific antibodies and complementary DNA probe. *J. Cell Biol.* **97**, 1693–9.

Wang, S. and Hazelrigg, T. (1994). Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature* **369**, 400–403.

Wang, J. I. N., Cao, L.-G., Wang, Y.-L. and Pederson, T. (1991). Localization of pre-messenger RNA at. *Cell* **88**, 7391–7395.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J. and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–9.

Wang, I. X., So, E., Devlin, J. L., Zhao, Y., Wu, M. and Cheung, V. G. (2013). ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. *Cell Rep.* **5**, 849–860.

Weil, T. T., Parton, R. M. and Davis, I. (2010). Making the message clear: Visualizing mRNA localization. *Trends Cell Biol.* **20**, 380–390.

Welch, M. D., DePace, A. H., Verma, S., Iwamatsu, A. and Mitchison, T. J. (1997). The Human Arp2/3 Complex Is Composed of Evolutionarily Conserved Subunits and Is Localized to Cellular Regions of Dynamic Actin Filament Assembly. *J. Cell Biol.* **138**, 375–384.

Wenzel, S., Schweimer, K., Rösch, P. and Wöhrl, B. M. (2008). The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with α/β type topology. *Biochem. Biophys. Res. Commun.* **370**, 414–418.

West, S., Gromak, N. and Proudfoot, N. J. (2004). Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525.

West, S., Proudfoot, N. J. and Dye, M. J. (2008). Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination. *Mol. Cell* **29**, 600–610.

Wilson, R. C. and Doudna, J. A. (2013). Molecular Mechanisms of RNA Interference. *Annu.*

Rev. Biophys. **42**, 217–239.

Wu, B., Chen, J. and Singer, R. H. (2014). Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **4**,.

Wu, C., Simonetti, M., Rossell, C., Mignardi, M., Mirzazadeh, R., Annaratone, L., Marchiò, C., Sapino, A., Bienko, M., Crosetto, N., et al. (2018). RollFISH achieves robust quantification of single-molecule RNA biomarkers in paraffin-embedded tumor tissue samples. *Commun. Biol.* **1**, 209.

Xu, Y. and Beckett, D. (1994). Kinetics of Biotinyl-5′-adenylate Synthesis Catalyzed by the Escherichia coli Repressor of Biotin Biosynthesis and the Stability of the Enzyme-Product Complex. *Biochemistry* **33**, 7354–7360.

Yoon, Y. J., Wu, B., Buxbaum, A. R., Das, S., Tsai, A., English, B. P., Grimm, J. B., Lavis, L. D. and Singer, R. H. (2016). Glutamate-induced RNA localization and translation in neurons. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6877–E6886.

Yu, J. and Russell, J. E. (2001). Structural and functional analysis of an mRNP complex that mediates the high stability of human beta-globin mRNA. *Mol. Cell. Biol.* **21**, 5879–88.

Yudkovsky, N., Ranish, J. A. and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229.

Yue, Y., Liu, J. and He, C. (2015). RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.* **29**, 1343–55.

Zabidi, M. A. and Stark, A. (2016). Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet.* **32**, 801–814.

Zhang, H. (2004). PolyA_DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res.* **33**, D116–D120.

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

Zhang, C., Fu, J. and Zhou, Y. (2019). A Review in Research Progress Concerning m6A Methylation and Immunoregulation. *Front. Immunol.* **10**, 922.

Zheng, D. and Tian, B. (2014). Sizing up the poly(A) tail: insights from deep sequencing. *Trends Biochem. Sci.* **39**, 255–257.

Zheng, C. L., Xiang-Dong, F. U. and Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* **11**, 1777–1787.

Zhou, Z., Luo, M., Straesser, K., Katahira, J., Hurt, E. and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401–405.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A. and Oliviero, S. (2009). Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. *Cell* **138**, 1122–1136.

9. List of publications

1. Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B., and Jansen, R.P. (2019).
   β-Actin mRNA interactome mapping by proximity biotinylation.
   Proc. Natl. Acad. Sci. U. S. A. 116, 12863–12872.

2. Mukherjee, J., Franz-Wachtel, M., Maček, B., and Jansen, R.P. (2019).
   RNA interactome identification via RNA BioID in mouse embryonic fibroblasts.
   Bio-protocols (accepted).

3. Mukherjee, J., Schweikert, A., Franz-Wachtel, M., Maček, B., and Jansen, R.P.
   Comparison between proximity labelling methods and RNA pull down to analyze β-Actin interacting proteome.
   Manuscript under preparation

10. Personal contribution to the publications contained in this thesis

1. Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B., and Jansen, R.P. (2019).
   β-Actin mRNA interactome mapping by proximity biotinylation. Proc. Natl. Acad. Sci. U. S. A. 116, 12863–12872.

   I performed all the experiments with the exception of experiment shown in Figure 5E.
   I prepared the samples for the on-bead digestion proteomics but the samples were processed by the Boris Maček group.
   I analyzed all the data after mass spec processing and made the representative graphs except S5, S6, S7, and S8.
   I wrote the manuscript together with Ralf-Peter Jansen.

2. Mukherjee, J., Franz-Wachtel, M., Maček, B., and Jansen, R.P. (2019).
   RNA interactome identification via RNA BioID in mouse embryonic fibroblasts. Bio-protocols (accepted).

   I performed the experiments and wrote the RNA-BioID portion of the manuscript together with Ralf-Peter Jansen.
   The rest of the protocol containing on beads digestion methods was written by Mirita Franz-Wachtel.

3. Mukherjee, J., Schweikert, A., Franz-Wachtel, M., Maček, B., and Jansen, R.P.
   Comparison between proximity labelling methods and RNA pull down to analyze β-Actin interacting proteome . Manuscript under preparation.

   I performed all the experiments.
   I prepared the samples for the on-bead digestion proteomics but the samples were processed by the Boris Maček group.
   I analyzed all the data after mass spec processing.

11. Acknowledgments

First of all, I want to thank my supervisor Prof. Dr. Ralf-Peter Jansen for giving me the opportunity to work in his laboratory, in this amazing project under his outstanding scientific guidance. Thank you for believing in me and giving me encouragement the freedom to venture around on risky experiments and motivating me to develop the best scientific outlook, it helped me a lot to build my scientific mind and attitude. Your doors were always open for scientific discussion which has always helped me to discuss problems personal or professional. I cannot thank you enough for the international exposure you have given me. Your guidance, support and motivation has taught me a lot and helped me to grow personally and professionally.

I am also very much thankful to my second supervisor, Prof. Dr. Gabriele Dott for being such an amazing mentor with her help and suggestions. Throughout the interactions during TAC meetings or on several other occasions you have always given me a good suggestion, moral support and a feeling that you are always there whenever I need. I am thankful to Prof. Dr. Dirk Schwarzer and Prof. Dr. Ana Garcia Saez for their time and effort as members of my doctoral examination committee.

I am immensely grateful to Dr. Frank Essmann for his time and efforts to guide me through the initial times during my PhD.
I am thankful to the Elisabeth Knoop stiftung for providing me the travel grant which helped me a lot to learn a great microscopy-based technique.

Furthermore, I want to thank all the past and present members of the Jansen lab to maintain a helpful, friendly, and joyful work environment. For all the silly jokes and also meaningful scientific conversations. Thank you, Claudia, for your help throughout my journey in this lab. Thanks, Ingrid, Ulrike, Ruth for your

helping nature and providing constant technical help in the lab, I learned a lot about lab management from you guys.

Thanks a lot, Ruth for your help in my project and all the friendly discussions we had.

Orit, I will always remember our coffee time discussions, all the scientific-nonscientific talks. Thank you so much for your help in my project.

Thank you Ibrahim, Srinivas, Phany (bhai), Kiran, Karthika (sis), Swetha for being such amazing friends and all the amazing time and support you guys have always showered on me.
Together with my Bengali group friends, Debapriya, Dipda and Modhumitadi and off course Kuttu you guys are my extended family in Tuebingen, and we will always remain as a family.

I am also thankful to Rob and all the singer lab people, especially Caro and Jeet for your help in this project. I had one of the best times of my life with you all especially with Manisha, Evelina, Young, Sulagna.

I am really grateful to my amazing family. My life, mistiburi, Dadamoni, Tipu, boudi, Maa, Baba, babin, I am blessed to have you all in my life. Thanks a lot for your constant support.

And finally, everything is for you MAA…

β-Actin mRNA interactome mapping by proximity biotinylation

Joyita Mukherjee[a], Orit Hermesh[a], Carolina Eliscovich[b], Nicolas Nalpas[c], Mirita Franz-Wachtel[c], Boris Maček[c], and Ralf-Peter Jansen[a,1]

[a]Interfaculty Institute of Biochemistry, University of Tübingen, 72074 Tübingen, Germany; [b]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461; and [c]Proteome Center Tübingen, University of Tübingen, 72074 Tübingen, Germany

Edited by Michael Rosbash, Howard Hughes Medical Institute, Brandeis University, Waltham, MA, and approved May 20, 2019 (received for review December 12, 2018)

The molecular function and fate of mRNAs are controlled by RNA-binding proteins (RBPs). Identification of the interacting proteome of a specific mRNA in vivo remains very challenging, however. Based on the widely used technique of RNA tagging with MS2 aptamers for RNA visualization, we developed a RNA proximity biotinylation (RNA-BioID) technique by tethering biotin ligase (BirA*) via MS2 coat protein at the 3′ UTR of endogenous MS2-tagged β-actin mRNA in mouse embryonic fibroblasts. We demonstrate the dynamics of the β-actin mRNA interactome by characterizing its changes on serum-induced localization of the mRNA. Apart from the previously known interactors, we identified more than 60 additional β-actin–associated RBPs by RNA-BioID. Among these, the KH domain-containing protein FUBP3/MARTA2 has been shown to be required for β-actin mRNA localization. We found that FUBP3 binds to the 3′ UTR of β-actin mRNA and is essential for β-actin mRNA localization, but does not interact with the characterized β-actin zipcode element. RNA-BioID provides a tool for identifying new mRNA interactors and studying the dynamic view of the interacting proteome of endogenous mRNAs in space and time.

RNA-BioID | mRNA localization | RNA-binding protein | FUBP3

The spatial distribution of mRNAs contributes to the compartmentalized organization of the cell and is required for maintaining cellular asymmetry, proper embryonic development, and neuronal function (1). Localized mRNAs contain cis-acting sequences, termed zipcodes or localization elements, that constitute binding sites for RNA-binding proteins (RBPs) (1). Together with these RBPs, localized mRNAs form transport complexes containing molecular motors, such as kinesin, dynein, and myosin (2, 3). These ribonucleoprotein complexes (RNPs) usually include accessory factors, such as helicases, translational repressors, RNA stability factors, and ribosomal proteins (3). Thus, mRNPs as functional units not only contain the information for an encoded polypeptide, but also determine the precise spatiotemporal regulation of the polypeptide's translation and stability, thereby facilitating proper subcellular localization of the translation product (4).

One of the best-studied localized mRNAs is β-actin, which encodes the β isoform of the cytoskeleton protein actin (5). β-Actin mRNA is localized to the protrusion of migrating fibroblasts (6), where its local translation critically contributes to the migrating behavior of this cell type (7–11). In the developing mouse (12) and *Xenopus* (13, 14) neurons, β-actin mRNA is transported to the growth cone during axonal extension, and its deposition and local translation are highly regulated by external cues. In addition, translation of this mRNA in dendritic spines is involved in reshaping the postsynaptic site of synapses (14). A well-defined localization element is present in the proximal region of the β-actin 3′-untranslated region (UTR) (15). This cis-acting element is recognized and bound by the zipcode-binding protein ZBP1 (16), the founding member of the conserved VICKZ RBP family (17). ZBP1 (also called IGF2BP1 or IMP1) interacts with the β-actin zipcode via the third and fourth KH (hnRNP K homology) domains (16) and is required for RNA localization in fibroblasts and neurons (18). It has also been suggested that IGF2BP1 controls the translation of β-actin mRNA by blocking the assembly of ribosomes at the start codon (11). IGF2BP1 appears to act as a key RBP in β-actin mRNA distribution, but other proteins, including IGF2BP2 (19), RACK1 (20), KHSRP/FUBP2 (21), KHDRBS1/SAM68 (22), FMR1 (23), and HuR (24), also have been suggested to be involved in β-actin mRNA localization, although their molecular function is less clear.

To fully understand the mechanism(s) of mRNA localization, it is important to identify and study the mRNA-binding factors. Major technological advances, such as cross-linking and immunoprecipitation (CLIP) combined with next-generation sequencing, have allowed the identification of RNAs bound to specific RBPs (25) and the system-wide identification of RBPs bound to polyA RNA (26, 27). However, the major techniques for determining which proteins associate with a specific RNA include affinity purification of modified or tagged RNAs together with their bound proteins, along with coimmunoprecipitation (co-IP) of RNP components with the aid of known RBPs (28). In addition, affinity capturing of specific RNPs with hybridizing antisense probes or via integrated aptamers has been successful (29–31). A limitation of these techniques is the potential loss of low-affinity binders during purification, which so far has been addressed by in vivo UV cross-linking before cell lysis (25, 26). However, cross-linking enhances only the recovery

Significance

Transport of specific mRNAs to defined sites in the cytoplasm allows local protein production and contributes to cell polarity, embryogenesis, and neuronal function. These localized mRNAs contain signals (i.e., zipcodes) that help direct them to their destination site. Zipcodes are recognized by RNA-binding proteins that, with the help of molecular motor proteins and supplementary factors, mediate mRNA trafficking. To identify all proteins assembling with a localized mRNA, we advanced a proximity labeling method, BioID, by tethering a biotin ligase to the 3′ UTR of mRNA encoding the conserved β-actin protein. We demonstrate that this method allows the identification of functionally important proteins required for mRNA localization.

Author contributions: J.M. and R.-P.J. designed research; J.M., O.H., and M.F.-W. performed research; J.M. contributed new reagents/analytic tools; J.M., O.H., C.E., N.N., M.F.-W., B.M., and R.-P.J. analyzed data; and J.M. and R.-P.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Proteomic data supporting this study have been deposited in the PRIDE archive, www.ebi.ac.uk/pride/archive/ (accession no. PXD010694).

[1]To whom correspondence may be addressed. Email: ralf.jansen@uni-tuebingen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820737116/-/DCSupplemental.

Published online June 12, 2019.

of RBPs directly contacting nucleobases and thus does not overcome the loss of other physiologically important RNA interactors (e.g., motor or adapter proteins). These limitations could be overcome by in vivo labeling of proteins while they are associated with the target RNA.

Proximity-dependent biotin identification, or BioID (32–34), has been successfully used to detect subunits of large or dynamic protein complexes, such as the nuclear pore complex (32) and centrosome (34). In BioID, a protein of interest is fused to a mutant version of the *Escherichia coli* biotin ligase BirA (BirA*) that generates AMP biotin ("activated biotin"), which reacts with accessible lysine residues in its vicinity (33). After cell lysis, biotinylated proteins can be isolated via streptavidin affinity purification and identified using standard mass spectrometry techniques. Recently, BioID has also been applied to identify proteins associated with the genomic RNA of Zika virus (35).

In this study, we used BioID to characterize the proteome of endogenous β-actin mRNPs. We found that tethering of BirA* to an endogenous transcript not only allows identification of its associated proteins, but also can be used to probe the environment of this mRNA. We identified FUBP3/MARTA2, an RBP from the conserved FUBP family of proteins (36–38), which was previously shown to mediate dendritic targeting of MAP2 mRNA in neurons (39, 40). We found that FUBP3 binds to and facilitates localization of β-actin mRNA to the fibroblast leading edge. FUBP3 does not bind to the zipcode or IGF2BP1, but mediates β-actin RNA localization by binding to a distal site in its 3′ UTR. Therefore, the RNA-BioID approach allows the identification of novel functional mRNA interactors within the cell with high confidence.

Results

Tethering Biotin Ligases to the 3′ UTR of β-Actin mRNA. To tether BirA* to the 3′ UTR of β-actin mRNA (Fig. 1*A*), we stably expressed a fusion of the nuclear localized signal (NLS), MS2 coat protein (MCP) (41), GFP, and BirA* (MCP-GFP-BirA*) in immortalized mouse embryonic fibroblasts (MEFs) from transgenic β-actin-24 MBS mice (Fig. 1 *A*, *Right*) (8). These mice have both β-actin gene copies replaced by β-actin with 24 MS2 binding sites (MBS) in their distal 3′-UTR. In parallel, NLS-MCP-GFP-BirA* was stably expressed in WT (wildtype) MEFs with untagged β-actin mRNA, to generate a control cell line to eliminate background biotinylation due to the presence of constitutive expression of BirA* (Fig. 1 *A*, second left panel). Both constructs contain two copies of the MCP protein leading to a maximum of 24 GFP and 24 BirA* that can potentially bind to an mRNA. Since biotinylation or the expression of the MCP-GFP-BirA* might affect localization of the β-actin mRNA, we checked for the proper targeting of β-actin mRNA to the leading edge of the cell by single molecule fluorescent in situ hybridization (smFISH) (42) and analyzed RNA localization by polarization index calculation (9) (Fig. 1 *B* and *C* and *SI Appendix*, Fig. S1 *A–F*). The distribution of mRNAs within cells was assessed using probes against the β-actin ORF (for primary and immortalized MEFs) and β-actin–MBS (for the genetically modified immortalized MEFs: β-actin–MBS, or β-actin–MBS IGF2BP1 KO) (10). To account for random distribution of an mRNA within the cell, we used probes against Gapdh as a control. Gapdh mRNA is a highly abundant and uniformly distributed mRNA. To induce β-actin mRNA localization, cells were serum starved for 24 h followed by stimulation with serum addition for 1 h. The median of the polarization index of β-actin mRNA distribution was significantly lower in immortalized (WT) or genetically modified immortalized MEFs compared with primary MEFs (Fig. 1*C*). Stimulation of polarization by serum was observed for all of the cell types used in a similar manner (Fig. 1 *C*, gray bars). Also, as shown before (10) knockout of IGF2BP1 reduces significantly β-actin–MBS mRNA polarization (Fig. 1*C*). We observed that β-actin and Igf2bp1 mRNA or protein levels were not affected (Fig. 1*D* and *SI Appendix*, Fig. S2 *A* and *B*). Altogether these results suggest that biotinylation and/or the expression of the MCP-GFP-BirA* does not affect regulation of β-actin mRNA in MEFs. Furthermore, cells with similar expression levels of MCP-GFP-BirA* were sorted by FACS (fluorescence activated cell sorting). As shown before (43), we also found no differences in the biotinylation efficiency at labeling conditions of 50 μM to 300 μM of biotin for 6–48 h. For optimal biotinylation, we decided to perform proximity labeling by addition of 50 μM biotin to the medium for 24 h. To test if proximity labeling can identify known β-actin mRNA-associated proteins, we affinity purified biotinylated proteins followed by Western blot detection of IGF2BP1 (mouse ZBP1). IGF2BP1 was biotinylated in MEFs expressing β-actin–MBS/MCP-GFP-BirA* but not in those expressing only GFP-BirA* (Fig. 1*E*), which demonstrates that our tool can successfully biotinylate zipcode-interacting proteins. To differentiate between endogenously biotinylated proteins and RNA-dependently biotinylated proteins, we performed streptavidin pulldown in cells expressing β-actin–MBS/MCP-GFP-BirA* and in cells expressing only MCP-GFP and observed biotinylation of numerous additional proteins (*SI Appendix*, Fig. S3). We expected that MCP-GFP-BirA* represents a major fraction of these biotinylated proteins and therefore aimed at depleting the fusion protein from the lysate by GFP pulldown before streptavidin affinity purification. To our surprise, most of the biotinylated proteins were enriched in the GFP pulldown fraction (*SI Appendix*, Fig. S3), likely due to copurification of MCP-GFP-BirA*, β-actin mRNA, and biotinylated proteins via binding to the mRNA or the fusion protein. RNA degradation with RNase A (*SI Appendix*, Fig. S4) shifted a large part of the biotinylated proteins into the streptavidin fraction (*SI Appendix*, Fig. S3), supporting the idea that most of the biotinylated proteins are associated with β-actin mRNA. Additional treatment with high salt and 0.5% SDS further optimized the streptavidin affinity purification and decreased the background binding of the magnetic beads used in this purification (*SI Appendix*, Fig. S3).

β-Actin mRNA Interactors Under Serum-Induced and Uninduced Conditions. β-Actin mRNA localization to the lamellipodia of chicken and mouse fibroblasts is known to increase after serum induction (6, 44). It also has been shown that cells enter a quiescent phase of the cell cycle during serum starvation (6), involving an overall reduction in actin stress fibers or focal adhesions (44). Since efficient biotinylation requires at least 6 h of incubation with biotin, we next applied smFISH to verify the persistence of β-actin mRNA localization during our labeling period. As has been shown previously (5), MEFs induced β-actin mRNA localization after serum addition (Fig. 1 *B* and *C*), and the fraction of MEFs with β-actin localized to lamellipodia increased within 1 h but then remained constant over the next 6 h.

To identify and compare the β-actin–associated proteomes in uninduced and serum-induced MEFs, we performed RNA-BioID under both conditions (three replicate experiments each). Unspecific as well as endogenous biotinylation was assessed by performing BioID in MEFs expressing MCP-GFP-BirA* in the absence of MS2 aptamers in β-actin mRNA. Affinity-captured biotinylated proteins were identified and quantified by mass spectrometry using label-free quantification. Principal component analysis of the datasets revealed that the different conditions cluster apart from each other in dimensions 1 and 2 (explaining 33.8% and 15.5% of the variance, respectively), while the replicates with the same conditions cluster together, demonstrating biological reproducibility (*SI Appendix*, Fig. S5). Calculating the Spearman correlation between all sample types and replicates (*SI Appendix*, Fig. S6) supports the high reproducibility between biological replicates (correlation ≥0.97). In

addition, it showed better correlation between uninduced and induced samples (average 0.95) compared with controls. In total, we found 169 (or 156) significantly enriched proteins in induced (or uninduced) MEFs compared with control cells (*SI Appendix*, Figs. S7 and S8*A*). Of these, 47 were enriched only under induced conditions (*SI Appendix*, Table S5). To assess the differential enrichment of the proteins under each condition, a Tukey post hoc test was performed after ANOVA, and significance was set to an adjusted P value of 0.05 following Benjamini–Hochberg multiple correction testing (*Materials and Methods*). Large fractions of the enriched proteins under induced conditions (30%) or uninduced conditions (34%) over control represent RBPs (Fig. 2, red solid circles); among these are RBPs (IGF2BP1, IGF2BP2, KHSRP, KHDRBS1, FMR1, HuR, RACK1, named in red) already known to control specific aspects of β-actin mRNA physiology. Other enriched RBPs have been associated with the localization of mRNAs in other cell types or organisms, including STAU1 and STAU2 (45–47), SYNCRIP (48), and FUBP3 (38). Furthermore, 85 proteins were significantly more enriched under serum-induced conditions than under uninduced conditions (*SI Appendix*, Fig. S8). However, the majority of the aforementioned RBPs (including IGF2BP1) become biotinylated under both induced and uninduced conditions, indicating that they are associated with β-actin mRNA under both conditions (Fig. 2*C*).

A cluster analysis (Fig. 3) reveals at least five different patterns of biotinylated proteins in induced, noninduced, and control MEFs (Fig. 3 *B* and *C*). In control MEFs, we see enrichment of mainly nuclear proteins (cluster 1). This is expected, since the unbound MCP-GFP-BirA* is enriched in the nucleus due to an N-terminal nuclear localization sequence (8). Cluster 1 also

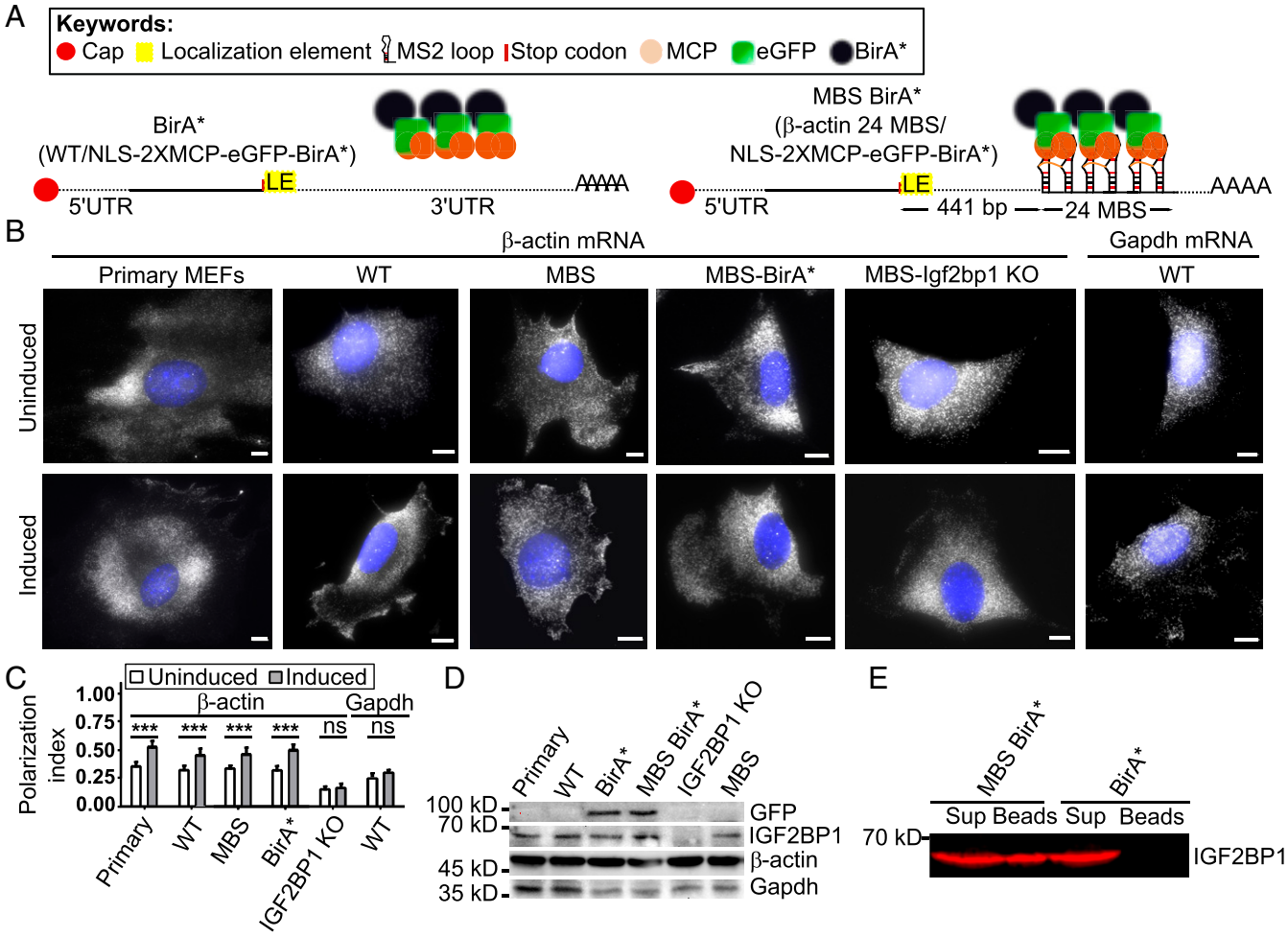


Fig. 1. RNA BioID to detect proteins interacting with localized β-actin RNA. (*A*) Schematic of the β-actin–MBS/GFP-BirA*. (*Left*) Control construct (BirA*) used to detect background biotinylation due to overexpression of the NLS-MCP-GFP-BirA* construct. Control cells expressing only NLS-2xMCP-eGFP-BirA* lack the MBS cassette in the β-actin mRNA. (*Right*) Construct used to detect β-actin mRNA-associated proteins (β-actin–MBS-BirA*). A 24xMS2 aptamer array (24MBS) was integrated in the 3′ UTR of the endogenous β-actin gene 441 bp downstream of the stop codon. BirA* is targeted to 24MBS by its fusion to a MS2 coat protein dimer (2xMCP). (*B*) Representative β-actin smFISH images of (from left to right) primary MEFs, immortalized MEFs (WT), β-actin–MBS, β-actin–MBS BirA*, and β-actin–MBS Igf2bp1 KO MEFs, as well as Gapdh smFISH images in immortalized (WT) MEFs (rightmost images). These and similar images were used to calculate the polarization index (C) of mRNA localization under serum-uninduced (*Top*) and serum-induced (*Bottom*) conditions. β-Actin mRNA was detected by probes against the β-actin ORF or MBS region, and Gapdh mRNA was detected by probes against its ORF (gray). (Scale bar: 10 μm.) (*C*) Bar graphs of the polarization index for Gapdh mRNA and β-actin mRNA in different MEFs [from left to right: primary, immortalized (WT), β-actin–MBS, β-actin–MBS BirA*, β-actin–MBS Igf2bp1 KO]. The polarization index was calculated in a total 100 of cells from three biological replicates. The line represents the median values. ***$P < 0.005$; not significant (ns), $P > 0.05$. (*D*) Protein levels of endogenous β-ACTIN, IGF2BP1, and heterologous MCP-GFP-BirA* detected by anti-GFP antibody. Quantification of Western blot analysis is provided in *SI Appendix*, Fig. S2. (*E*) Biotinylation of IGF2BP1 depends on MBS sites in β-actin. Following RNase A treatment, biotinylated proteins were affinity-purified with streptavidin-coated beads from cells expressing 2xMCP-eGFP-BirA* in the presence (β-actin-24MBS) or absence (β-actin) of MBS. The presence of IGF2BP1 was probed by a specific antibody in bead fractions (Beads) and supernatant (Sup).

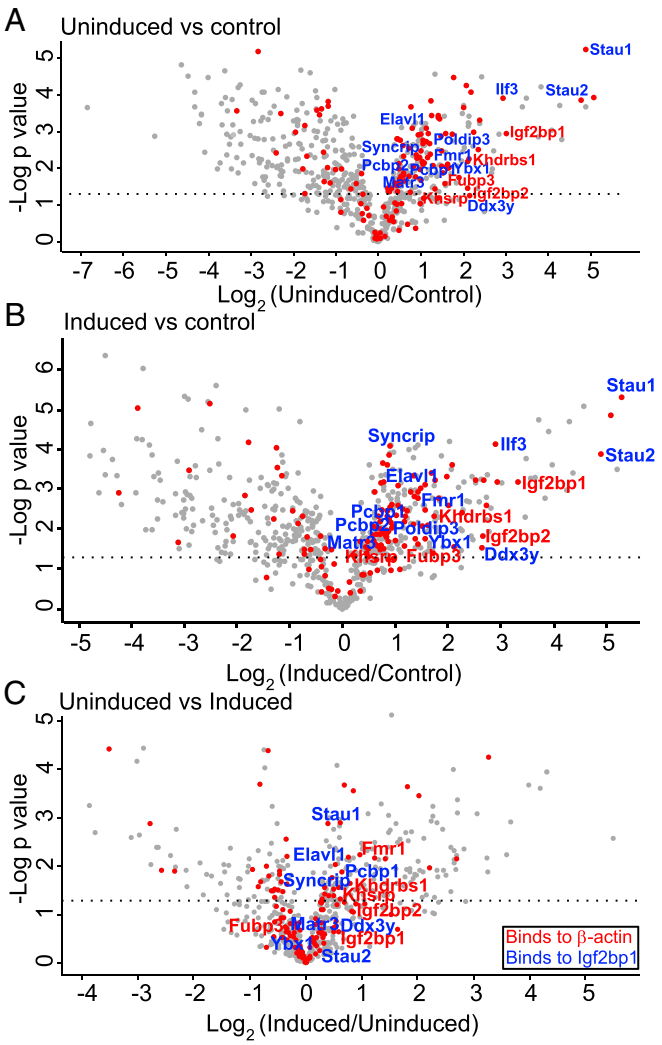


Fig. 2. Enrichment of biotinylated proteins in control MEFs, or MEFs expressing β-actin–MBS-BirA* under serum-induced or uninduced conditions. Volcano plot representation of biotinylated proteins in uninduced MEFs compared with control MEFs (*A*), serum-induced MEFs compared with control MEFs (*B*), and serum-induced MEFs compared with uninduced MEFs (*C*). In the volcano plots, the *x*-axis represents log2 fold change in protein abundance and the *y*-axis represents the −log10 *P* value. Red circles are known RBPs identified by Gene Ontology (GO) molecular function analysis. Proteins in red represent known β-actin mRNA interactors, and proteins in blue are RBPs known to bind to IGF2BP1. The dotted line indicates $P = 0.05$.

contains abundant cytoplasmic proteins, including glycerol aldehyde phosphate dehydrogenase (GAPDH). Cluster 3 represents proteins found equally in MEFs under all conditions and contains ribosomal proteins, among others. Proteins allocated to the other three clusters (clusters 2, 4, and 5) are overrepresented in the biotinylated proteome of MEFs expressing β-actin–MBS/GFP-BirA*. Of specific interest are clusters 4 and 5. In cluster 4, with proteins that are more biotinylated under serum-induced conditions, we find RBPs, including FMR1 and KHSRP, that have been reported to function in β-actin mRNA localization or to bind to IGF2BP1.

Another group of proteins that are enriched in this cluster comprises proteins of the actin cytoskeleton (e.g., Filamin B, Cofilin-1, Myh9, Tpm4, Plastin-3). Their enrichment likely reflects deposition of the β-actin mRNA in the actin-rich cortical environment of the leading edge of MEFs. Finally, cluster 5 contains proteins found in β-actin–MBS MEFs under both induced and uninduced conditions but not in control MEFs. This cluster shows enrichment for proteins involved in mRNA-binding, RNP constituents, and ribosomal proteins. Since this cluster contains the RBP IGF2BP1, we hypothesized that other proteins in this cluster, such as FUBP3, are likely candidates for β-actin mRNA regulatory factors.

FUBP3 Is a Component of the β-Actin mRNP. To confirm the association of the identified proteins and MS2-tagged β-actin mRNA, we combined single-molecule fluorescence in situ hybridization with immunofluorescence (smFISH-IF) using Cy3-labeled probes against either the ORF or the MBS of β-actin mRNA and antibodies against GFP, FUBP3, or IGF2BP1 in WT or β-actin–MBS MEFs (Fig. 4 *A*–*C*). While ORF probes were used to detect β-actin mRNA in WT MEFs, MBS probes against the MS2 loop sequences were used to detect the β-actin mRNA in β-actin–MBS MEFs. The association between β-actin mRNA and the proteins was determined by super-registration microscopy (47). In brief, we corrected the images for chromatic aberration and mechanical shifts in Cy3 and Cy5 channels using broad spectra fluorescent microsphere beads (*SI Appendix*, Fig. S9) and found that colocalization of smFISH and IF signals did not occur by chance within the cell using a positive control (MBS-GFP; Fig. 4*A*) and a negative control (Gapdh-GFP; Fig. 4*D*) for RNA–protein interaction. We calculated the association between the RNA and protein molecules as a function of their distances apart for positive and negative controls (Fig. 4*E*). For the positive control, 91% of the observed distances from the labeled probes to the MBS and from the antibodies to the GFP were within 60 nm (the optimal distance). In contrast, only 10% of the observed associations in the negative control (using Gapdh probes and MCP-GFP) were within 60 nm (Fig. 4 *E* and *F*). When combining smFISH of Gapdh with IF against MCP-GFP, fewer overlapping events were observed at a distance of <150 nm compared with MBS-GFP (Fig. 4 *A*, *D*, and *E*). At greater distances (>150 nm), the fluorescence signals in both channels were more likely to overlap by chance and thus are considered a random event. We found that at the optimal distance of 60 nm, β-actin mRNA was associated with IGF2BP1 and FUBP3 in MEFs. The RNA–protein associations were 37% for IGF2BP1 with β-actin and 29% for FUBP3 with β-actin in MEFs (Fig. 4 *B*, *C*, and *F*). These associations were significantly higher than the nonspecific interaction between Gapdh and MCP-GFP (10%), suggesting the physical contact between the molecules.

FUBP3 and IGF2BP1 Bind on Different Regions of β-Actin mRNA and Interact with Each Other in an RNA-Dependent Manner. To validate the data demonstrating the RNA–protein association by super-registration microscopy, we performed co-IP of β-actin mRNA with FUBP3 and IGF2BP1 (Fig. 5*A*). Co-IP was tested with four mRNAs: β-actin, Cofilin1, Igf2bp1, and Fubp3 (Fig. 5*A*). IGF2BP1 bound to all the mRNAs tested, reflecting previous observations in HeLa cells, where almost 3% of the transcriptome was shown to bind to IGF2BP1 (49). Coprecipitation of β-actin with FUBP3 (23% of input bound to FUBP3) was similar to that with IGF2BP1 (37%). These values are consistent with the degree of RNA–protein association seen on colocalization (Fig. 4). In contrast, β-actin mRNA was not efficiently bound by the RBP VIGILIN, indicating that this mRNA does not associate with every RBP (*SI Appendix*, Fig. S10). The localized Cof1 mRNA (50) was bound by both FUBP3 and IGF2BP1 to a similar extent (48%).

To further substantiate our finding that FUBP3 can bind independently of IGF2BP1 to β-actin mRNA, we performed co-IP experiments with IGF2BP1 and FUBP3 (Fig. 5*B*) in the presence and absence of RNase A. The RBP STAU1 served as a positive control since it has been shown to bind to IGF2BP1 (51). Co-IP of IGF2BP1 and FUBP3 vanished on RNase treatment, indicating an RNA-dependent interaction between these two proteins. We conclude that FUBP3 does not bind to β-actin

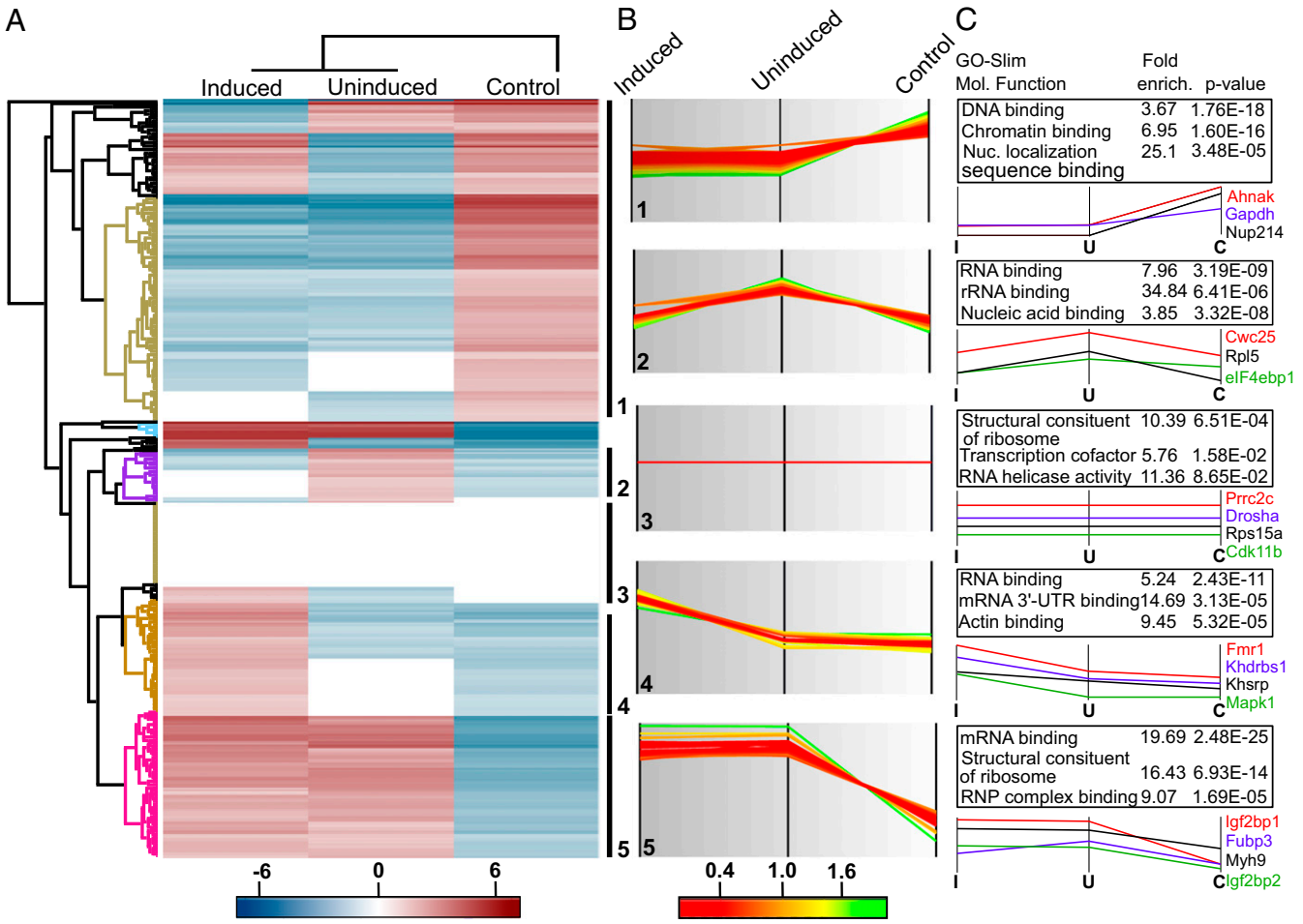


Fig. 3. Cluster analysis of biotinylated proteins in control MEFs or MEFs expressing β-actin–MBS-BirA* under serum-induced and uninduced conditions. (*A*) Hierarchical clustering of biotinylated proteins in serum-induced and uninduced β-actin–MBS-BirA* MEFs and control MEFs (lacking β-actin–MBS). Enrichment is indicated in red; depletion, in blue. Various clusters of protein groups are highlighted in the dendrogram. (*B*) Profile plots of five selected clusters showing distinct enrichment patterns of biotinylated proteins: 1, strongly enriched in control MEFs; 2, enriched in β-actin–MBS-BirA* MEFs under uninduced conditions; 3, similar enrichment in all MEFs; 4, enriched in β-actin–MBS-BirA* MEFs under serum-induced conditions; and 5, enriched in β-actin–MBS-BirA* MEFs under serum-induced and uninduced conditions compared with control MEFs. Color-coding shows the degree of enrichment in each specific cluster: green, more enriched; red, less enriched). (*C*) Functional analysis of protein annotation terms results in multiple categories that are enriched in the selected clusters. GO-slim molecular function terms, the corresponding enrichment factors, and *P* values are shown in the table. Selected examples of proteins found in each cluster are shown below the tables.

mRNA via IGF2BP1, but that both proteins may bind β-actin independently at different sites.

We next used recombinant histidine-tagged proteins (FUBP3-HIS and IGF2BP1-HIS) in pulldown assays (Fig. 5 *C* and *D*) to test binding to in vitro transcribed RNA fragments of β-actin mRNA. We selected the complete 643-bp-long β-actin 3′ UTR and the 54-nt localization zipcode element of β-actin. As negative control for IGF2BP1 binding, we used a mutant version of the zipcode region (16). In addition, we used a 49-nt region adjacent to the zipcode (proximal zipcode; ref. 16). A 79-nt region in the 3′ UTR at 460 nt downstream to the stop codon of β-actin mRNA, which spans a potential FUBP3-binding motif UAUG (52), along with a 75-nt fragment of the same region but carrying a deleted UAUG motif were used to specifically probe FUBP3 binding. The capturing assay was performed in total bacterial lysates to allow bacterial RBPs to compete for RNA binding. RNA captured by the His-tagged fusion proteins was detected by quantitative RT-PCR and normalized to the input (Fig. 5*D*). We found that IGF2BP1 and FUBP3 were bound to the 3′ UTR of β-actin mRNA, while neither could interact with the mutated zipcode or zipcode proximal region. Only FUBP3 was bound to the 79-nt region containing the UAUG motif on the 3′ UTR of β-actin mRNA, and the binding was abolished in absence of this motif (Fig. 5*D*). This is highly suggestive of direct binding of FUBP3 to the UAUG motif in the 3′ UTR of β-actin.

To identify the KH domain(s) of FUBP3 responsible for binding β-actin mRNA, we introduced mutations in the conserved KH domains of the protein. Each functionally important G-X-X-G motif in the four KH domains was changed to the inactive G-D-D-G (53), and individual mutant proteins were transiently expressed in MEFs as C-terminally tagged mCherry fusion protein. The G-D-D-G mutation in KH domain 2 resulted in loss of the cytoplasmic punctate signal seen in WT FUBP3, reminiscent of the punctate pattern observed for mRNPs (Fig. 5*E*). We conclude that KH2 in FUBP3 is important for its integration into RNP particles and likely constitutes the critical domain for RNA binding.

Loss of FUBP3 Affects β-Actin mRNA Localization. To validate that proteins identified by RNA-BioID are functionally significant for the mRNA used as bait, we performed shRNA-mediated knockdown experiments for FUBP3. The effectiveness of the knockdown was validated by quantitative RT-PCR and Western blot

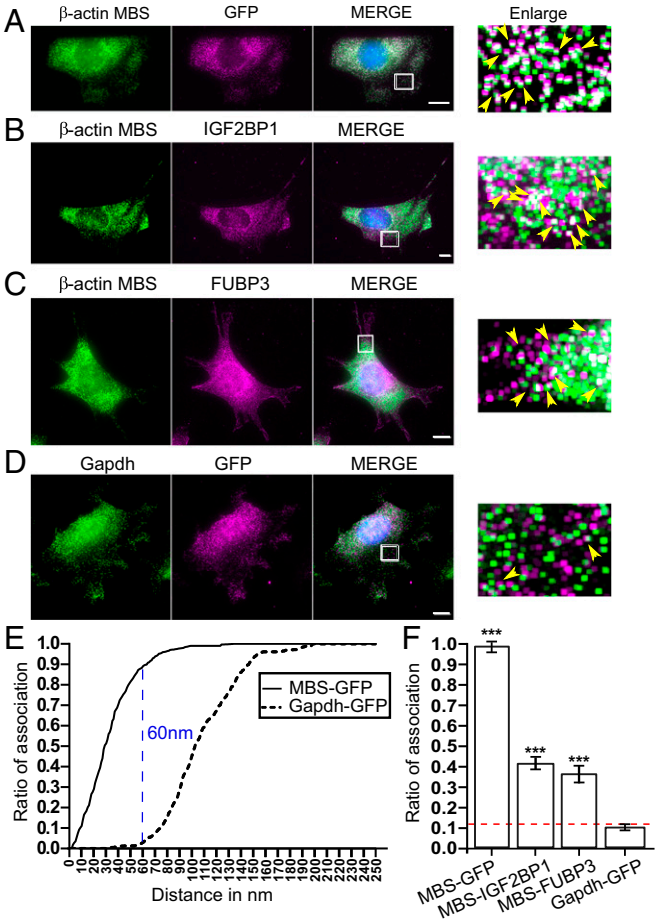


Fig. 4. Association analysis of IGF2BP1 and FUBP3 with β-actin–MBS mRNA by super-registration microscopy. (*A–D*) Representative smFISH-IF images of MEFs expressing β-actin–MBS and MCP-GFP. Shown are MEFs stained for β-actin mRNA (MBS FISH probes, Cy3; green) and MCP-GFP (*A*), IGF2BP1 (*B*), and FUBP3 (*C*). Immunofluorescence staining is shown in magenta. MCP-GFP served as a positive control to determine the optimum distance between mRNA and protein. *D* represents staining of Gapdh mRNA (probes from Biosearch, Cy3; Green) together with MCP-GFP and served as a negative control to determine the distance for association between two signals occurring by chance. A 1-pixel dilated, enlarged version is shown on the right side of each panel (47). (*E*) Association curves between an mRNA (black, β-actin–MBS; dotted, Gapdh) and MCP-GFP protein. The curve of association is calculated as the cumulative ratio of association for intermolecular distances (in the range of 0–250 nm) that were less than a given observed distance, as described previously (48). The blue line represents the distance where the mRNA–protein association for MCP-MBS and MCP-Gapdh is maximally separated, the optimal distance (OD = 60 nm). (*F*) Summary of association analysis of β-actin mRNA and indicated proteins by smFISH-IF and super-registration. The dotted red line indicates background association defined by MCP-Gapdh. The error bar represents SD. $P > 0.05$; $*P < 0.05$; $***P < 0.001$, unpaired *t* test.

analysis (Fig. 6 *A* and *B*) using GAPDH as a control since it does not interact with β-actin mRNA, as shown by RNA BioID (Fig. 3*C*). FUBP3 knockdown only mildly reduced mRNA levels of β-actin or IGF2BP1 mRNAs (Fig. 6*A*). Similarly, IGF2BP1 protein levels did not significantly change on FUBP3 knockdown (Fig. 6*B*), ruling out an indirect effect of FUBP3 on β-actin mRNA by limiting IGF2BP1 levels. However, we observed a slight increase in β-ACTIN protein level, indicating that FUBP3 might coregulate β-actin mRNA translation or β-ACTIN protein stability.

We assessed the effect of the FUBP3 knockdown or overexpression of mutant FUBP3 on β-actin mRNA localization by smFISH-IF (Fig. 6 *C–F* and *SI Appendix*, Fig. S11) and calculated the polarization index (Fig. 6*G*). In control cells (immortalized MEFs; Fig. 6*C*), FUBP3 and β-actin mRNA were expressed, and the polarization index of β-actin mRNA was 0.37 (Fig. 6*G*). In FUBP3 knockdown cells (Fig. 6*D*), almost no FUBP3 signal was detectable, and the β-actin mRNA polarization index dropped to 0.25 (Fig. 6*G*). To test whether the reduction in polarization is due to a loss of FUBP3, we expressed a knockdown-insensitive mCherry-tagged FUBP3 in these MEFs. Expression of this fusion protein was accessed by indirect immunofluorescence against mCherry, and β-actin mRNA was visualized by smFISH (Fig. 6*E*). The polarization index was determined using only MEFs positive for mCherry. Although full rescue was not observed, the polarization index was increased, to 0.33 (Fig. 6*G*). This indicates that FUBP3 is important for β-actin mRNA localization.

We also analyzed the effect on β-actin mRNA distribution when overexpressing a mCherry-tagged FUBP3mt2 mutant lacking a functional KH2 domain (Fig. 6*F*). As before, we selected MEFs with an mCherry signal for determination of the polarization index. We found a polarization index of 0.31 (Fig. 6*G*), which is not significantly different from that of β-actin mRNA in WT MEFs. These data suggest that although KH2 is important for the formation of FUBP3-containing RNP particle-like structures in the cytoplasm, it does not act as dominant negative mutation, probably because a mutant with this mutation does not compete with endogenous FUBP3.

Discussion

Proximity biotinylation has facilitated the characterization of dynamic protein complexes by in vivo labeling of interaction partners. Here we exploit this approach and demonstrate its utility for identifying functionally relevant RBPs of a specific mRNA, mammalian β-actin. This is achieved by combining MS2 tagging of the mRNA of choice and coexpression of a fusion protein of the MS2 coat protein (MCP) and the biotin ligase (BirA*).

The primary goal of RNA-based BioID is to identify novel RNA interactors. As seen in several proximity labeling (BioID or APEX-driven) approaches (43, 54, 55), the number of identified potential interactors for β-actin is far higher than the number of proteins identified by classical co-IP or coaffinity purification approaches. This might be due to proximity labeling's greater sensitivity or its propensity to allow the capture of transient interactors (56). Although this can result in a skewed view of the actual components of a complex due to the rapid change in the composition of mRNP, it is beneficial to identify all mRNP components during the life stages of an mRNA. The most highly represented class of proteins was RBPs (Fig. 3 and *SI Appendix*, Fig. S8*B*), among them all RBPs previously associated with localization, translational control, or (de)stabilization of β-actin mRNA. Other RBPs, such as survival of motor neuron 1 (SMN1), which supports the association of IGF2BP1 with β-actin mRNA (57), were also found to be enriched in MEFs expressing β-actin–MBS compared with control MEFs, although with lower significance ($P < 0.1$).

We also analyzed our dataset for motor proteins involved in mRNA transport. Neither MYH10 (58) nor KIF11 (59), which have been suggested to work as β-actin mRNA transport motors, were found as biotinylated proteins. The only motor that we identified was MYH9, the heavy chain of an MYH10-related class II-A myosin, although it was not significantly enriched ($P = 0.08$). The lack of motor proteins is compatible with a recent observation that β-actin localization in fibroblasts works primarily by diffusion to and trapping in the microfilament-rich cortex (60). This is also corroborated by our finding that components of the actin-rich cell protrusion (Fig. 3, cluster 4) are heavily biotinylated in MEFs after serum-induced localization of β-actin.

Overall, our cluster analysis shows that the majority of previously identified β-actin RBPs behave similarly under the two test conditions (serum-induced and uninduced MEFs). This not

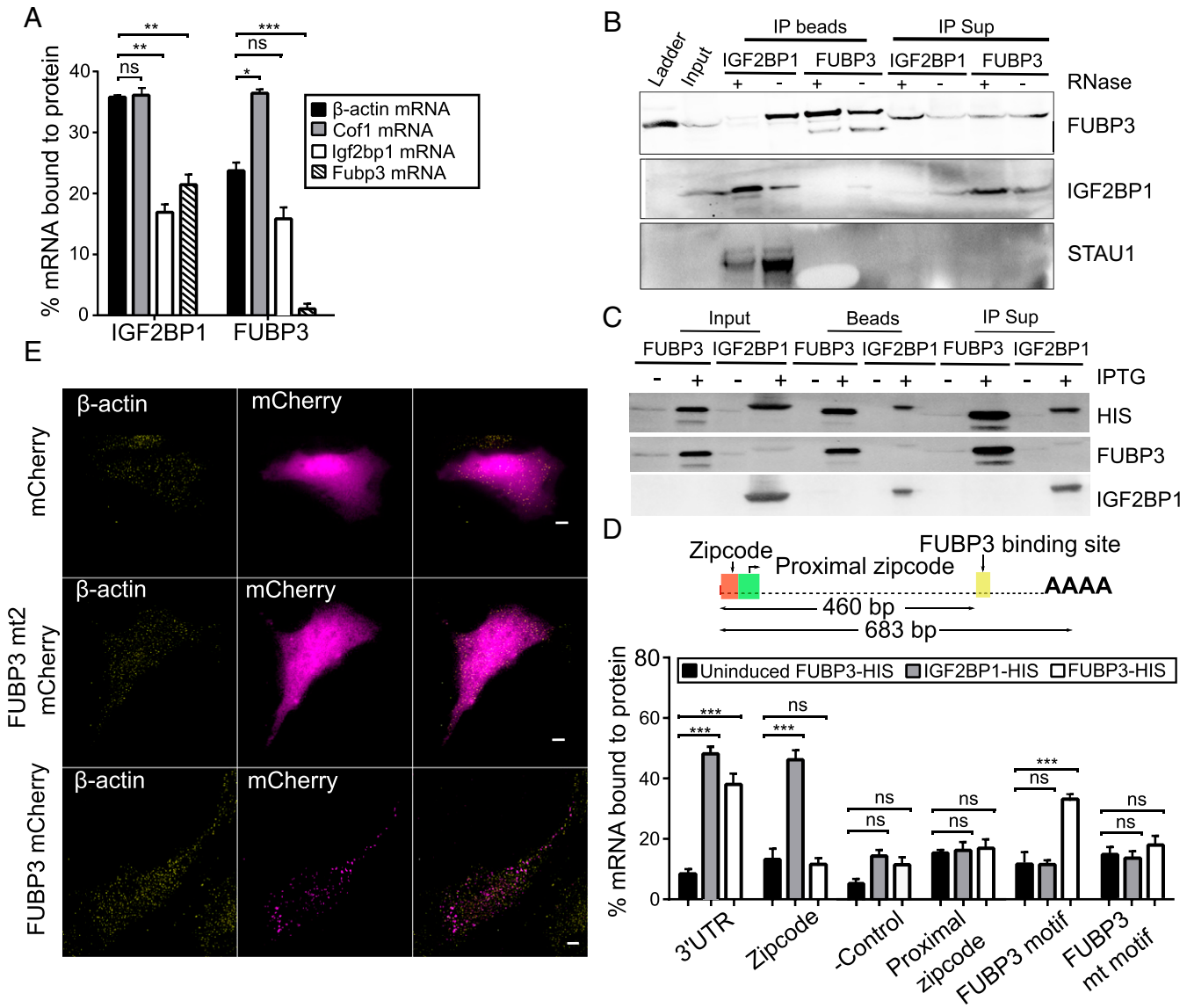


Fig. 5. FUBP3 binds to β-actin 3′ UTR. (*A*) Co-IP of selected mRNAs with IGF2BP1 and FUBP3. Bars represent percentage of input mRNA copurifying with the indicated protein. IGF2BP1 binds to several endogenous mRNAs, including Cofilin1, Igf2bp1, and Fubp3. FUBP3 binds to 23% of endogenous β-actin mRNA, while IGF2BP1 was associated with 37% of endogenous β-actin mRNA. Error bars represent mean ± SEM from three independent experiments. (*B*) Co-IP of STAU1, FUBP3, and IGF2BP1. Immunoprecipitation was performed from WT MEFs with either anti-FUBP3 or anti-IGF2BP1 antibodies in the presence and absence of RNase A. IGF2BP1 coprecipitates with FUBP3 only in the absence of RNase A, while binding of STAU1 to IGF2BP1 is RNA-independent. (*C*) Pulldown of His-tagged fusion proteins of IGF2BP1 and FUBP3 from bacterial lysates of *E. coli* grown under isopropyl β-D-1-thiogalactopyranoside (IPTG)-induced or IPTG-uninduced conditions. Magnetic beads were used to precipitate either IGF2BP1-HIS or FUBP3-HIS. (*D*, *Top*) Schematic representation of the 3′ UTR of β-actin mRNA. The 683-bp-long 3′ UTR contains the 54-nt zipcode sequence (after the stop codon), the proximal zipcode sequence (49 bp following the zipcode), and a potential FUBP3-binding sequence (460 bp downstream of the stop codon) with a consensus UAUG motif. (*D*, *Bottom*) Binding of in vitro transcribed RNA fragments of β-actin (complete 3′ UTR, zipcode, proximal zipcode, zipcode mutant, FUBP3-binding motif region, region with mutated FUBP3-binding motif) to IGF2BP1 or FUBP3. RNAs were added to *E. coli* lysates with or without (IPTG-uninduced) expressed His-tagged fusion protein. After affinity purification, bound RNAs were detected by quantitative RT-PCR. Bars represent percentage of input RNA. In contrast to IGF2BP1, FUBP3 shows little affinity for the zipcode sequence but binds to the 3′ UTR and a region containing the UAUG motif in the 3′ UTR. Error bars represent mean ± SEM from three independent experiments. Statistical significance of each dataset was determined using Student's *t* test. *$P < 0.05$; ***$P < 0.001$; not significant (ns), $P > 0.05$. (*E*) RNA-binding domain KH2 is required for FUBP3 cytoplasmic granule formation. The conserved G-X-X-G motif of FUBP3 KH domains were individually mutated into G-D-D-G and WT and mutant proteins expressed in MEFs as mCherry fusion. Live cell imaging shows that WT FUBP3-mCherry forms cytoplasmic granules, whereas a KH2 mutant (FUBP3 mt2) is evenly distributed in the cytoplasm like the control mCherry protein. (Scale bars: 5 μm.)

only indicates that they interact with β-actin mRNA in MEFs even under steady-state conditions, but also makes it likely that other proteins, especially RBPs, found in this cluster might represent as-yet-unknown β-actin mRNA interactors. By choosing the far-upstream binding protein FUBP3 as a potential candidate, we demonstrate that this assumption holds true for at least this protein. Not only does FUBP3 bind to β-actin mRNA, but its knockdown also results in a similar decrease of β-actin localization to the leading edge as is seen with loss of IGF2BP1.

FUBP3 (also known as MARTA2) has been reported to bind to the 3′ UTR of the localized MAP2 mRNA in rat neurons (39) to regulate its dendritic targeting (40). Although the binding site of FUBP3 in MAP2 mRNA is not known, its preferred binding motif (UAUA/UAUG) was recently identified (52). This motif is

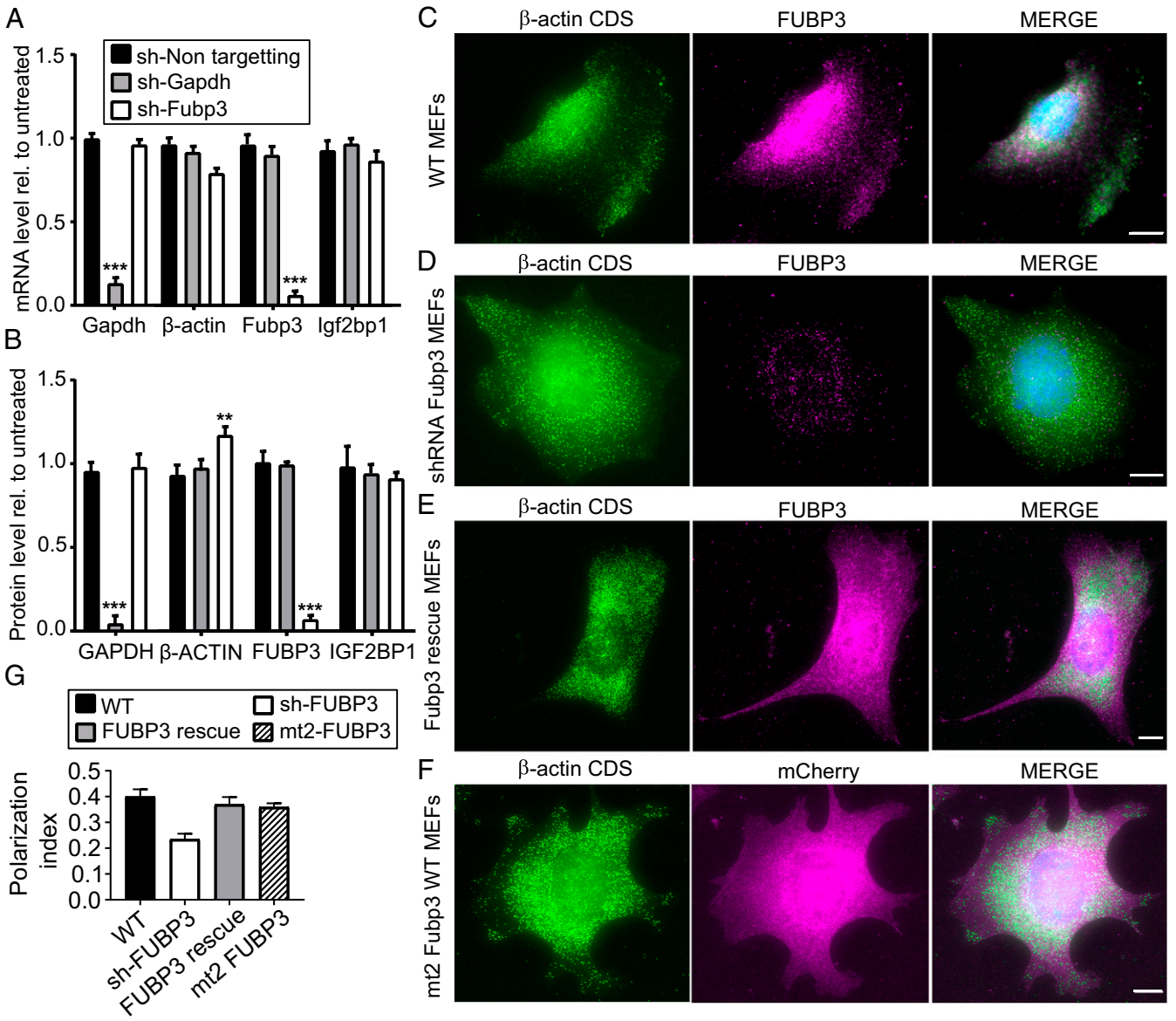


Fig. 6. Down-regulation of Fubp3 affects β-actin mRNA localization. (*A*) Western blot analysis to monitor shRNA-mediated knockdown of FUBP3 in non-targeted shRNA stably expressing in WT (second lane) compared with unmodified WT (lane 1) and WT stably expressing shRNA against Gapdh (lane 3) and stably expressing shRNA against FUBP3 (lane 4). Blot was reprobed against GAPDH and β-ACTIN and IGF2BP1 to access the changes in these proteins due to knockdown of FUBP3 or GAPDH. (*B*) Quantitative RT-PCR analysis of Gapdh, β-actin, Igf2bp1, and Fubp3 levels in knockdown cells. Corresponding mRNA levels in untreated cells were used for normalization. The statistical significance of each dataset was determined using Student's *t* test. $*P < 0.05$; $***P < 0.001$. (*C*) Western blot quantification of GAPDH, β-ACTIN, IGF2BP1, and FUBP3 protein levels in knockdown cells. Protein levels in untreated cells served as a normalization control. The statistical significance of each dataset was determined using Student's *t* test. $*P < 0.05$; $***P < 0.001$. (*D–G*) Representative smFISH-IF images of immortalized MEFs before or after Fubp3 knockdown. (*D–F*) β-Actin mRNA (Cy3; green) and FUBP3 (magenta). Shown are MEFs before knockdown (WT) (*D*), MEFs after shRNA treatment (*E*), and shRNA-treated MEFs expressing a FUBP3 rescue construct (*F*) (*Materials and Methods*). (*H*) Representative smFISH-IF image of MEFs expressing the FUBP3 KH mutant mt2 (Fig. 5). (Scale bars: 10 μm) (*I*) Polarization index for β-actin mRNA in MEFs from experiments shown in *D–G*. The polarization index was calculated from a total of 65 cells from three biological replicates. Bars represent the median values. Statistical significance of each dataset was determined using Student's *t* test. $*P < 0.05$; $***P < 0.001$.

present in the 3′ UTR of β-actin 460 nt downstream of the zip-code, and a 79-nt region containing this motif is bound by FUBP3. FUBP proteins might play a more substantial role in RNA localization, since homologs of a second member of the FUBP family, FUBP2, not only are reportedly involved in MAP2 or β-actin mRNA localization, but also are present among the biotinylated proteins that we identified. However, FUBP2 is mainly nuclear, and its role in β-actin mRNA localization might be indirect (61). In contrast, FUBP3 seems to have a direct function in localizing β-actin, as it binds to the 3′ UTR and its loss reduces β-actin mRNA localization independently of IGF2BP1. This independent function is supported by the observation that both proteins do not directly bind to each other but do bind to different regions of β-actin mRNA. A potential additional function could be translational regulation. Although less dramatic than seen for loss of IGF2BP1, knockdown of FUBP3 results in increased amounts of β-ACTIN protein, while β-actin mRNA levels are similar or even lower than those in untreated MEFs. This could be due to a loss of translational inhibition, as has been shown for IGF2BP1 (11).

Its role in β-actin and MAP2 mRNA localization suggests that FUBP3/MARTA2 is a component of several localizing mRNPs. Of note, RNA-BioID on β-actin mRNA has identified even more RBPs involved in the localization of other mRNAs, including SYNCRIP (48) and Staufen (45). Several of these RBPs (e.g., STAU1, STAU2) are highly enriched in our β-actin biotinylated proteome. This finding might reflect the participation of multiple RBPs in β-actin localization or regulation. It also shows that a common set of RBPs is used to control the fate of several different localized mRNAs in different cell types. Although RNA-BioID does not currently allow us to determine whether all these RBPs are constituents of the same β-actin mRNP, belong to different states of an mRNP, or belong to different populations, their identification allows us to address these questions to achieve a more detailed understanding of the common function of RBPs on diverse mRNAs.

Materials and Methods

RNA-BioID. For RNA-BioID, cells were incubated with 50 μM biotin for at least 6 h. Following incubation, cells were washed twice with 1× PBS, lysed in IP lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2.5 mM $MgCl_2$, 1 mM DTT, 1% Tween-20, and 1× protease inhibitor) and passed 10–12 times through a 21-gauge needle. The lysate was cleared by centrifugation at 12,000 × *g* for 10 min at 4 °C to remove cell debris. Protein from the supernatant (total cell lysate; 10 μg) was used to check for protein biotinylation. In the remaining lysate, NaCl was added to a final concentration of 500 mM. Then 200 μL of streptavidin magnetic bead suspension (GE Healthcare) were added, and the high salt lysate was incubated overnight at 4 °C with end-to-end rotation. The next day, the beads were collected (by keeping the beads on the magnetic stand for 2 min) and washed as described previously (43). The beads were washed twice for 5 min with 0.3 mL of wash buffer 1 (2% SDS), once with wash buffer 2 (0.1% wt/vol deoxycholate, 1% wt/vol Tween-20, 350 mM NaCl, 1 mM EDTA pH 8.0), once with wash buffer 3 (0.5% wt/vol deoxycholate, 0.5% wt/vol Tween-20, 1 mM EDTA, 250 mM LiCl, 10 mM Tris·HCl pH 7.4) and 50 mM Tris·HCl pH 7.5, once with wash buffer 4 (50 mM NaCl, 50 mM Tris·HCl pH 7.4), and finally twice with 500 μL of 50 mM ammonium bicarbonate. Then 20 μL of the beads were used for Western blot and silver staining, and 180 μL were subjected to mass spectrometry analysis. To release captured proteins from streptavidin beads for Western blot analysis, the beads were incubated in 2× Laemmli buffer containing 2 mM saturated biotin and 20 mM DTT for 10 min at 95 °C.

For biotinylation after serum induction, cells were starved for 24 h and then induced with 10% serum-containing medium containing 50 μM biotin for 6–24 h. Samples were processed for mass spectrometry analysis as described in *SI Appendix, Materials and Methods*.

Microscopy and Super-Registration Microscopy. For live cell imaging, cells were imaged with a Zeiss Cell Observer wide-field fluorescence microscope, operated by ZEN software, illuminated with a xenon arc lamp, and detected with a CCD camera (Axiocam 506) with 100×/1.45 α-Plan fluor oil immersion objectives (Zeiss). Live cell imaging was done using a dual-band GFP/mCherry filter set (F56-319; AHF). For imaging of fixed cells, the microscope setup was the same as described by Eliscovich et al. (47).

Imaging Analysis. Single-molecule localization was determined with FISH-QUANT (62), and super-registration analysis was performed as described by Eliscovich et al. (47) with existing software packages and custom algorithm programs written in MATLAB (MathWorks). For polarization index calculation, after taking the maximum projections from all of the Z-stacks, polarization and dispersion indices were measured as described previously (9) with an existing software package written in MATLAB.

smFISH-IF. Immortalized WT MEFs or MEFs containing MS2-tagged β-actin but no MCP-GFP were seeded on a fibronectin-coated cover glass in a 12-well cell culture plate and grown for 24 h in serum-free medium, followed by the addition of serum-containing medium to the cells for 1–2 h. The protocol for smFISH-IF has been described previously (47). In brief, cells were washed three times with PBS, fixed for 10 min with 4% paraformaldehyde in PBS, washed three times in PBS and then quenched in 50 mM glycine, and finally permeabilized with 0.1% Triton X-100 (28314; Thermo Fisher Scientific) and 0.5% Ultrapure BSA (AM2616; Life Technologies) in 1× PBS-M for 10 min. After washing with PBS, cells were exposed to 10% (vol/vol) formamide, 2× SSC, and 0.5% Ultrapure BSA in RNase-free water for 1 h at room temperature, followed by incubation for 3 h at 37 °C with either 10-ng custom-labeled probes or 50-nM Stellaris RNA FISH probes (Biosearch Technologies) (*SI Appendix*, Table S4). Primary antibodies against GFP (GFP-1010; Aves Labs), IGF2BP1 (RN001M; MBL), or FUBP3 (Abcam) were diluted (*SI Appendix*, Table S3) in hybridization buffer containing 10% formamide, 1 mg/mL *E. coli* tRNA, 10% dextran sulfate, 20 mg/mL BSA, 2× SSC, 2 mM vanadyl ribonucleoside complex, and 10 U/mL SUPERase-In (Ambion) in RNase-free water. After incubation and quick washing, cells were further incubated twice with an Alexa Fluor 647-conjugated secondary antibody (Life Technologies) in 10% formamide and 2× SSC in RNase-free water for 20 min at 37 °C. After four washes in 2× SSC, DNA was counterstained with DAPI (0.1 μg/mL in 2× SSC; Sigma-Aldrich), and after a final wash, cells were mounted using ProLong Diamond Antifade Reagent (Life Technologies).

Data Availability. Proteomic data supporting this study have been deposited in the PRIDE database, www.ebi.ac.uk/pride/archive/ (accession no. PXD010694).

ACKNOWLEDGMENTS. We thank Jeff Chao (Friedrich Miescher Institute for Biomedical Research, Basel), Imre Gaspar (European Molecular Biology Laboratory, Heidelberg), Julién Bethune (Heidelberg University Biochemistry Center, Heidelberg), Dierk Niessing (University of Ulm), Michael Kiebler (University of Munich), Stefan Kindler (University of Hamburg), Stefan Hüttelmaier (University of Halle), and Ibrahim Muhammad Syed (Interfaculty Institute of Biochemistry, Tübingen), for plasmids, cell lines, antibodies, or spike RNA. We are grateful to Robert H. Singer for hosting J.M. during an imaging internship. We also thank Frank Essmann, Ruth Schmid (both at Interfaculty Institute of Biochemistry, Tübingen), and Silke Wahle (Proteome Center Tübingen) for technical support; Jeetayu Biswas (Albert Einstein College) for help with the polarization index scripts and the IGF2BP1 KO cell line; and Matthew Cheng (Interfaculty Institute of Biochemistry, Tübingen) for suggestions on the manuscript. The project was funded as a project of the Deutsche Forschungsgemeinschaft (DFG) Research Unit FOR2333 by a grant from the DFG (DFG JA696/11-1). C.E. was supported by an NIH grant (NS083085).

1. K. C. Martin, A. Ephrussi, mRNA localization: Gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
2. C. Eliscovich, A. R. Buxbaum, Z. B. Katz, R. H. Singer, mRNA on the move: The road to its biological destiny. *J. Biol. Chem.* **288**, 20361–20368 (2013).
3. V. Marchand, I. Gaspar, A. Ephrussi, An intracellular transmission control protocol: Assembly and transport of ribonucleoprotein complexes. *Curr. Opin. Cell Biol.* **24**, 202–210 (2012).
4. G. Dreyfuss, V. N. Kim, N. Kataoka, Messenger RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195–205 (2002).
5. E. H. Kislauskis, X. Zhu, R. H. Singer, β-Actin messenger RNA localization and protein synthesis augment cell motility. *J. Cell Biol.* **136**, 1263–1270 (1997).
6. J. B. Lawrence, R. H. Singer, Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **45**, 407–415 (1986).
7. A. F. Ross, Y. Oleynikov, E. H. Kislauskis, K. L. Taneja, R. H. Singer, Characterization of a beta-actin mRNA zipcode-binding protein. *Mol. Cell. Biol.* **17**, 2158–2165 (1997).
8. T. Lionnet *et al.*, A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8**, 165–170 (2011).
9. H. Y. Park, T. Trcek, A. L. Wells, J. A. Chao, R. H. Singer, An unbiased analysis method to quantify mRNA localization reveals its correlation with cell motility. *Cell Rep.* **1**, 179–184 (2012).
10. Z. B. Katz *et al.*, β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration. *Genes Dev.* **26**, 1885–1890 (2012).
11. S. Hüttelmaier *et al.*, Spatial regulation of β-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438**, 512–515 (2005).
12. G. J. Bassell *et al.*, Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. *J. Neurosci.* **18**, 251–265 (1998).
13. J. Yao, Y. Sasaki, Z. Wen, G. J. Bassell, J. Q. Zheng, An essential role for beta-actin mRNA localization and translation in $Ca2^{+}$-dependent growth cone guidance. *Nat. Neurosci.* **9**, 1265–1273 (2006).
14. B. Turner-Bridger *et al.*, Single-molecule analysis of endogenous β-actin mRNA trafficking reveals a mechanism for compartmentalized mRNA localization in axons. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9697–E9706 (2018).
15. E. H. Kislauskis, X. Zhu, R. H. Singer, Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *J. Cell Biol.* **127**, 441–451 (1994).
16. J. A. Chao *et al.*, ZBP1 recognition of beta-actin zipcode induces RNA looping. *Genes Dev.* **24**, 148–158 (2010).
17. J. K. Yisraeli, VICKZ proteins: A multi-talented family of regulatory RNA-binding proteins. *Biol. Cell* **97**, 87–96 (2005).
18. Y. J. Yoon *et al.*, Glutamate-induced RNA localization and translation in neurons. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6877–E6886 (2016).
19. K. Wächter, M. Köhn, N. Stöhr, S. Hüttelmaier, Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. *Biol. Chem.* **394**, 1077–1090 (2013).

20. M. Ceci *et al*., RACK1 is a ribosome scaffold protein for β-actin mRNA/ZBP1 complex. *PLoS One* **7**, e35034 (2012).
21. F. Pan, S. Hüttelmaier, R. H. Singer, W. Gu, ZBP2 facilitates binding of ZBP1 to beta-actin mRNA during transcription. *Mol. Cell. Biol.* **27**, 8340–8351 (2007).
22. M. Itoh, I. Haga, Q.-H. Li, J. Fujisawa, Identification of cellular mRNA targets for RNA-binding protein Sam68. *Nucleic Acids Res.* **30**, 5452–5464 (2002).
23. O. Rackham, C. M. Brown, Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J.* **23**, 3346–3355 (2004).
24. V. Dormoy-Raclet *et al*., The RNA-binding protein HuR promotes cell migration and cell invasion by stabilizing the beta-actin mRNA in a U-rich element-dependent manner. *Mol. Cell. Biol.* **27**, 5365–5380 (2007).
25. F. C. Y. Lee, J. Ule, Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369 (2018).
26. M. Hafner *et al*., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
27. A. Castello *et al*., Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
28. J. Zielinski *et al*., In vivo identification of ribonucleoprotein-RNA interactions. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1557–1562 (2006).
29. B. Rogell *et al*., Specific RNP capture with antisense LNA/DNA mixmers. *RNA* **23**, 1290–1302 (2017).
30. I. Gaspar, F. Wippich, A. Ephrussi, Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).
31. B. Slobodin, J. E. Gerst, A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* **16**, 2277–2290 (2010).
32. D. I. Kim *et al*., Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2453–E2461 (2014).
33. K. J. Roux, D. I. Kim, B. Burke, D. G. May, BioID: A screen for protein-protein interactions. *Curr. Protoc. Protein Sci.* **91**, 19.23.1–19.23.15 (2018).
34. E. N. Firat-Karalar, T. Stearns, Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. *Methods Cell Biol.* **129**, 153–170 (2015).
35. M. Ramanathan *et al*., RNA-protein interaction detection in living cells. *Nat. Methods* **15**, 207–212 (2018).
36. H.-J. Chung *et al*., FBPs are calibrated molecular tools to adjust gene expression. *Mol. Cell. Biol.* **26**, 6584–6597 (2006).
37. L. M. Quinn, FUBP/KH domain proteins in transcription: Back to the future. *Transcription* **8**, 185–192 (2017).
38. K. H. Zivraj *et al*., The RNA-binding protein MARTA2 regulates dendritic targeting of MAP2 mRNAs in rat neurons. *J. Neurochem.* **124**, 670–684 (2013).
39. A. Blichenberg *et al*., Identification of a cis-acting dendritic targeting element in MAP2 mRNAs. *J. Neurosci.* **19**, 8818–8829 (1999).
40. M. Rehbein, S. Kindler, S. Horke, D. Richter, Two trans-acting rat-brain proteins, MARTA1 and MARTA2, interact specifically with the dendritic targeting element in MAP2 mRNAs. *Brain Res. Mol. Brain Res.* **79**, 192–201 (2000).
41. D. S. Peabody, The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).
42. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
43. K. J. Roux, D. I. Kim, M. Raida, B. Burke, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
44. S. Tyagi, O. Alsmadi, Imaging native β-actin mRNA in motile fibroblasts. *Biophys. J.* **87**, 4153–4162 (2004).
45. J. E. Heraud-Farlow, M. A. Kiebler, The multifunctional Staufen proteins: Conserved roles from neurogenesis to synaptic plasticity. *Trends Neurosci.* **37**, 470–479 (2014).
46. V. Balasanyan, D. B. Arnold, Actin and myosin-dependent localization of mRNA to dendrites. *PLoS One* **9**, e92349 (2014).
47. C. Eliscovich, S. M. Shenoy, R. H. Singer, Imaging mRNA and protein interactions within neurons. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1875–E1884 (2017).
48. S. M. McDermott, C. Meignin, J. Rappsilber, I. Davis, *Drosophila* Syncrip binds the gurken mRNA localisation signal and regulates localised transcripts during axis specification. *Biol. Open* **1**, 488–497 (2012).
49. L. Jønson *et al*., Molecular composition of IMP1 ribonucleoprotein granules. *Mol. Cell. Proteomics* **6**, 798–811 (2007).
50. Y. Maizels *et al*., Localization of cofilin mRNA to the leading edge of migrating cells promotes directed cell migration. *J. Cell Sci.* **128**, 1922–1933 (2015).
51. D. Weidensdorfer *et al*., Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15**, 104–115 (2009).
52. D. Dominguez *et al*., Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854–867.e9 (2018).
53. D. Hollingworth *et al*., KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886 (2012).
54. I. M. Schopp *et al*., Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat. Commun.* **8**, 15690 (2017).
55. H.-W. Rhee *et al*., Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
56. R. Varnaitė, S. A. MacNeill, Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics* **16**, 2503–2518 (2016).
57. C. Fallini *et al*., Dynamics of survival of motor neuron (SMN) protein interaction with the mRNA-binding protein IMP1 facilitates its trafficking into motor neuron axons. *Dev. Neurobiol.* **74**, 319–332 (2014).
58. V. M. Latham, E. H. Yu, A. N. Tullio, R. S. Adelstein, R. H. Singer, A Rho-dependent signaling pathway operating through myosin localizes beta-actin mRNA in fibroblasts. *Curr. Biol.* **11**, 1010–1016 (2001).
59. T. Song *et al*., Specific interaction of KIF11 with ZBP1 regulates the transport of β-actin mRNA and cell motility. *J. Cell Sci.* **128**, 1001–1010 (2015).
60. H. Y. Park *et al*., Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **343**, 422–424 (2014).
61. W. Gu, F. Pan, H. Zhang, G. J. Bassell, R. H. Singer, A predominantly nuclear protein affecting cytoplasmic localization of β-actin mRNA in fibroblasts and neurons. *J. Cell Biol.* **156**, 41–51 (2002).
62. F. Mueller *et al*., FISH-quant: Automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).

Supplementary Information (Mukherjee et al., 2019)

Supplementary Figures

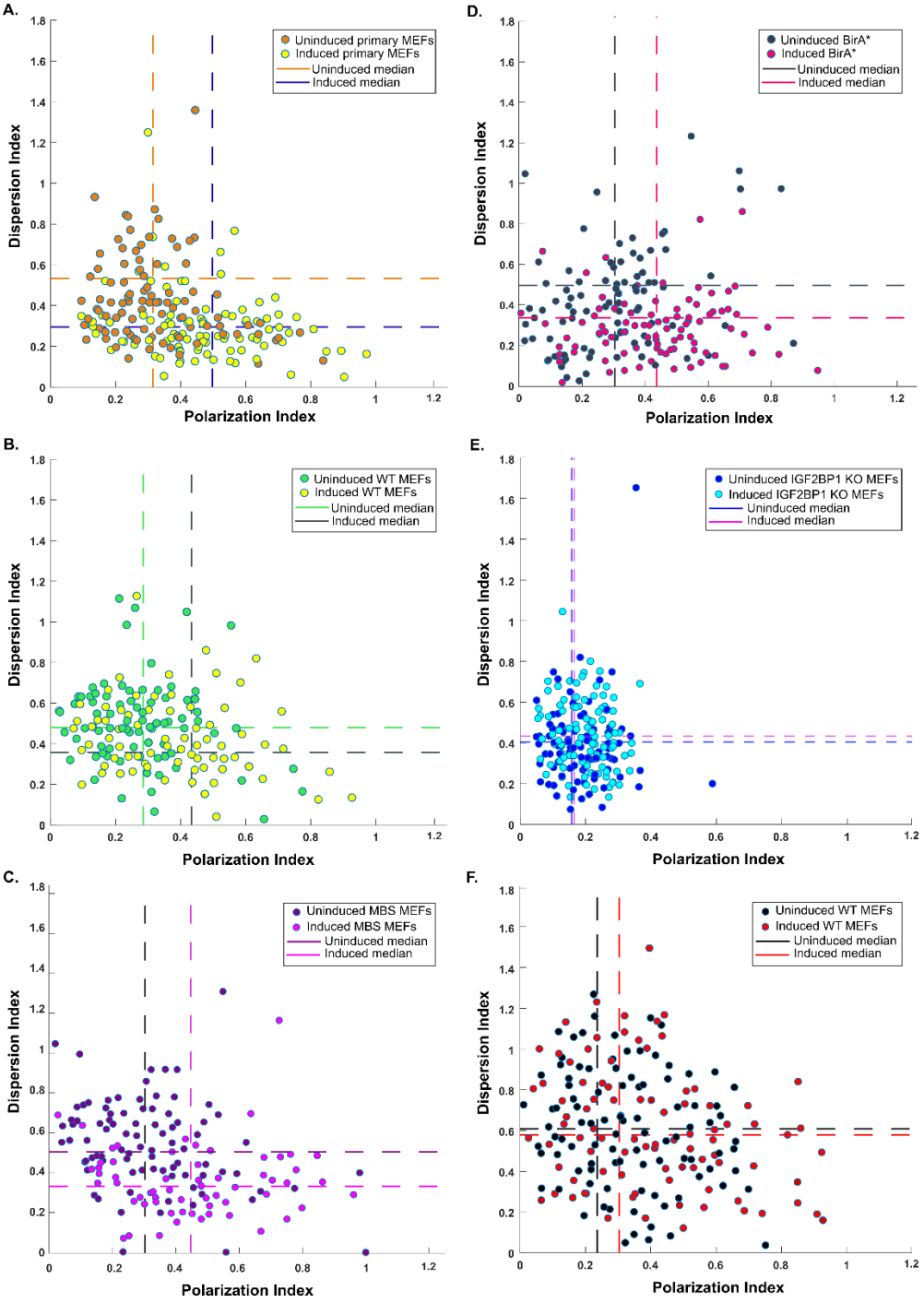


Fig.S1

Scatter plot to compare changes in β-actin or Gapdh mRNA distribution. Scatter plot displays polarization index for β-actin mRNA (A-E) or Gapdh (F) mRNA on X axis against dispersion index on y axis. Data in A - E are derived from smFISH images detecting β-actin mRNA in primary MEFs, immortalized (WT) MEFs, MBS MEFs, MBS-BirA* MEFs, or IGF2BP1 knockout MEFs. Data in F are derived from smFISH images detecting Gapdh mRNA in immortalized (WT) MEFs. 100 cells were counted for each growth condition (serum induced or uninduced).

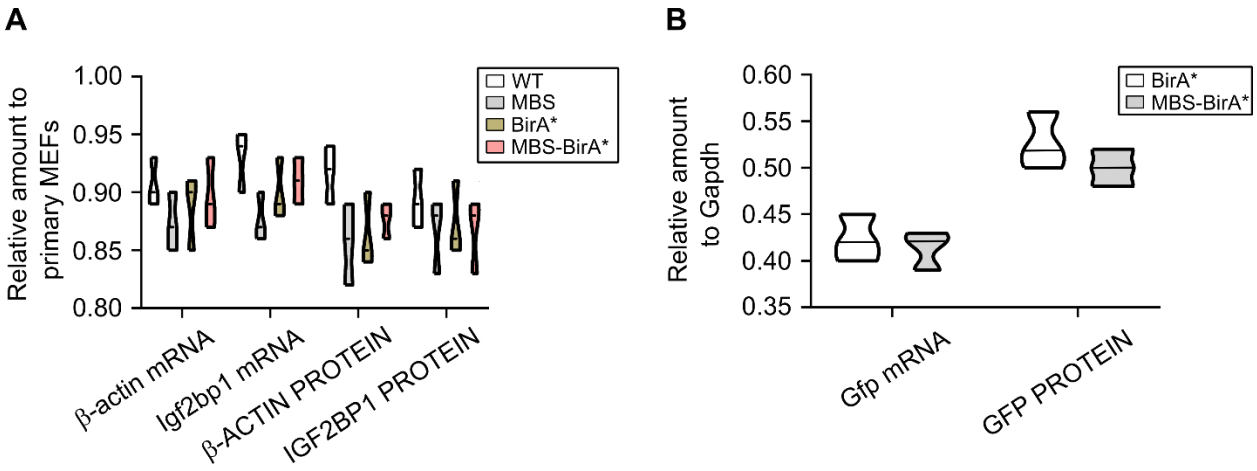


Fig.S2

Violin plot to compare changes in β-actin or Igf2bp1 or heterogenous Gfp mRNA and protein expression in various cell types.

(A) Violin plot displaying amounts of endogenous β-actin or Igf2bp1 mRNA and protein (compared to primary MEFs). (B) Violin plot displaying amounts of heterogenous GFP mRNA (normalized to endogenous Gapdh mRNA or protein levels) Data relates to Fig. 1D and shows measurements in primary MEFs, immortalized (WT) MEFs, MBS MEFs, MBS-BirA* MEFs, or BirA* MEFs. The black lines represent median values derived from 3 biological replicates.

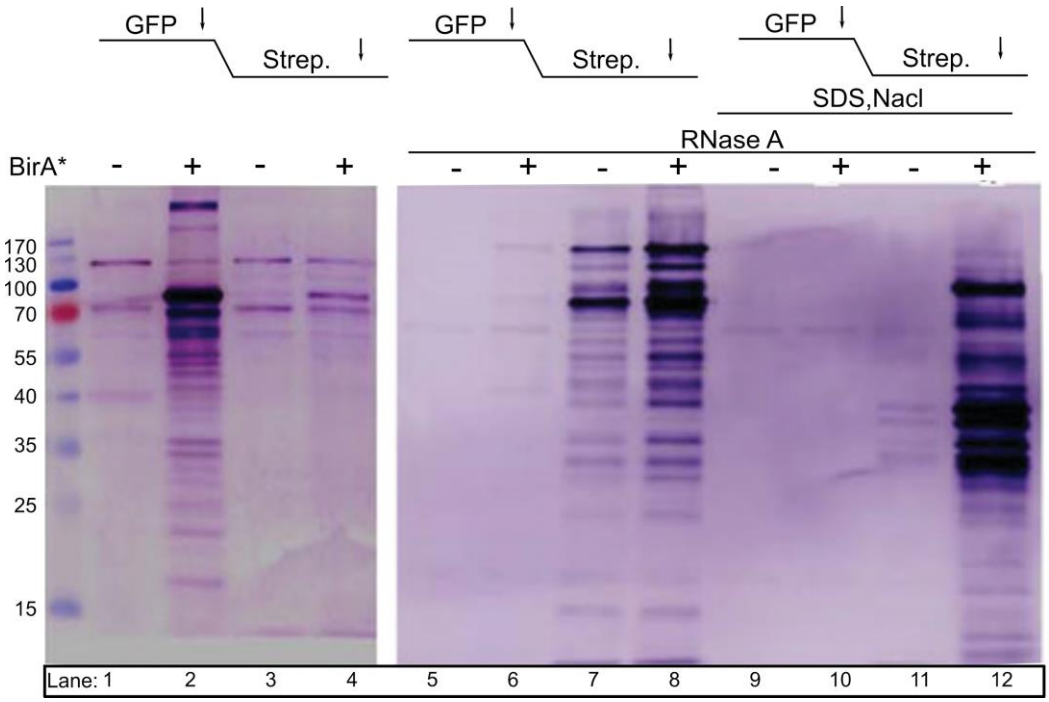


Fig.S3

Optimization of purification schedule for biotinylated proteins labeled by RNA-BioID. Specific enrichment of β-actin-MBS associated, biotinylated proteins is achieved by stringent conditions during purification. Two consecutive affinity purifications (anti-GFP followed by streptavidin pulldown) were performed. Western blots were stained for biotinylated proteins by streptavidin-alkaline peroxidase. Left panel: The majority of the biotinylated proteins remain associated with 2xMCP-eGFP-BirA* in the GFP pulldown fraction under low-stringency purification conditions (lane 2). Combination of treatment with RNase A (lane 6 versus 8), or 0.5% SDS and 500 mM NaCl (lane 10 versus 12) leads to quantitative enrichment of biotinylated proteins by streptavidin pulldown.

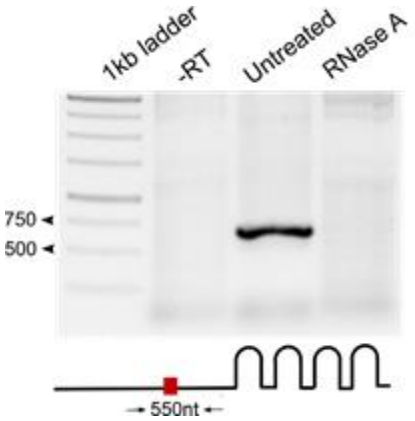


Fig.S4

Validation of β-actin mRNA degradation. β-actin-MBS-BirA* and β-actin-MBS-eGFP expressing cells were treated with 100 µg/ml RNase A for 30 min at 37°C before affinity purification (Fig. S1). Efficient RNA removal was checked by RT-PCR using a primer set that spans the indicated 550 nucleotide region upstream of the MBS cassette.

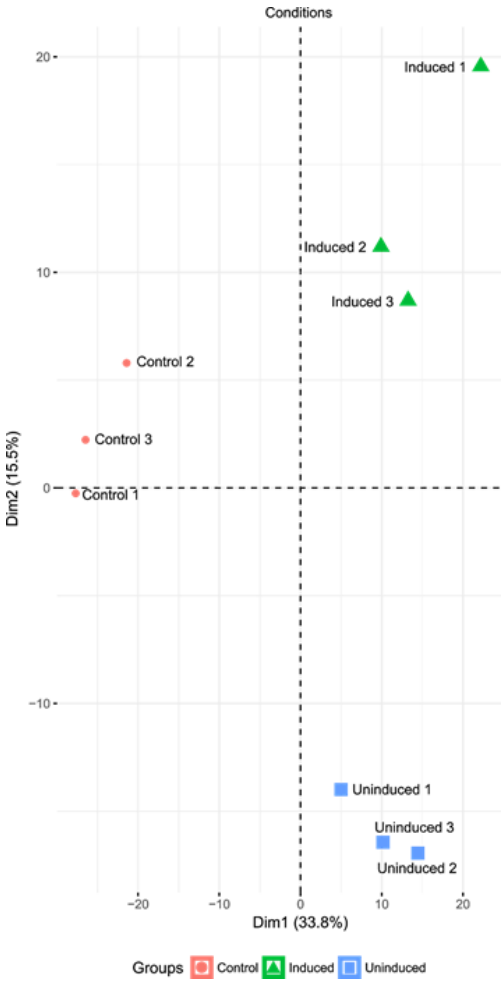


Fig.S5

Visualization of sample correlation by principal component analysis (PCA). PCA analysis reveals the largest variance between control and induced / uninduced samples on dimension 1 (variation explained 33.8%) and further separates control, uninduced and induced samples on dimension 2 (variation explained 15.5%).

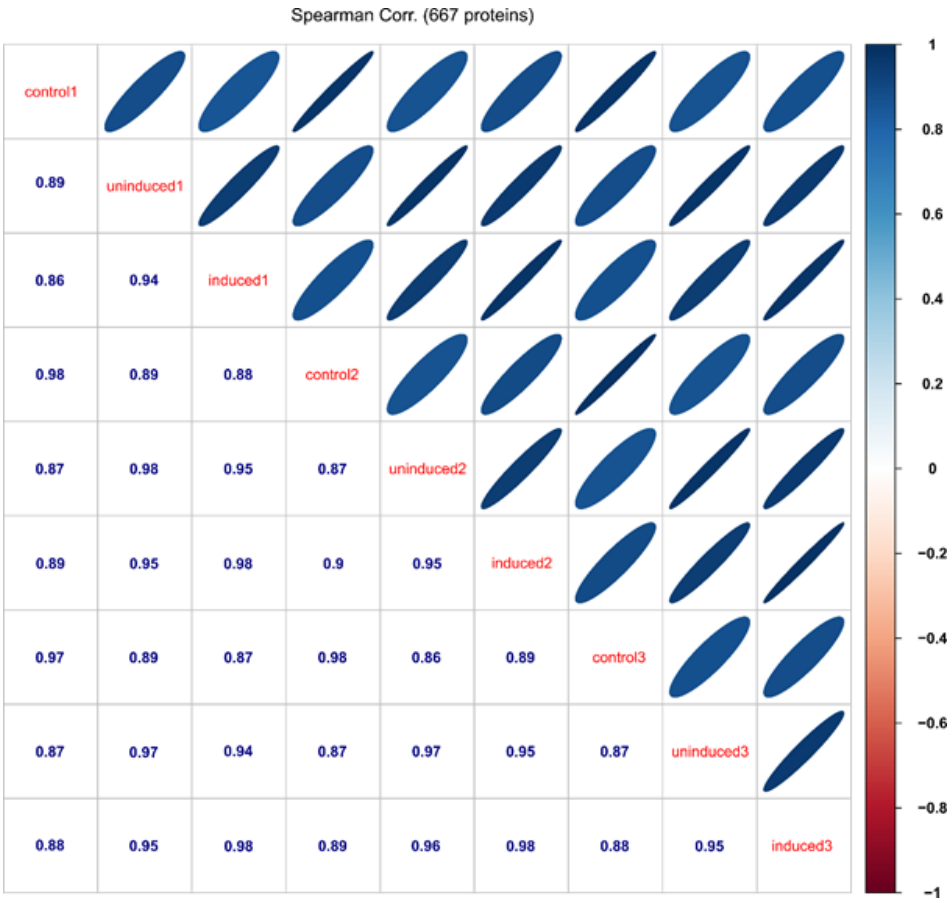


Fig.S6

Spearman correlation calculation between all possible combinations of control (MCP-eGFP-BirA* only), induced and uninduced MEF samples (Control, induced, uninduced). Correlation calculation was done based on LFQ values for common proteins. The color gradient represents high correlation (dark blue) to high anti-correlation (dark red). The analysis shows that there is a very high correlation within biological replicates per condition (>0.97), which shows the high reproducibility of the datasets.

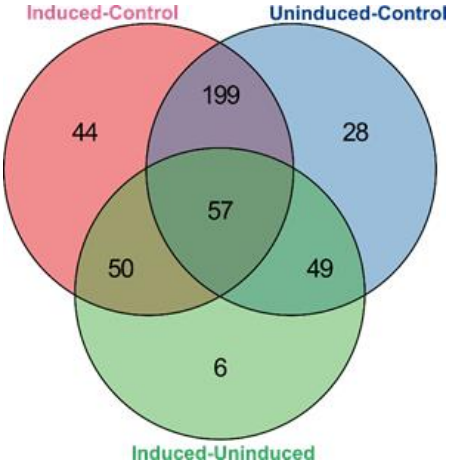


Fig.S7

Venn-diagram representation of significantly enriched proteins in each population of MEFs (control/induced/uninduced). In total, 350 proteins were significantly enriched in the dataset of induced MEFs versus control, 333 proteins in the dataset of uninduced MEFs versus control, and 162 proteins in the dataset of induced versus uninduced MEFs. 57 proteins were common between all three datasets. 44 unique proteins were identified in induced MEFs versus control, 28 proteins in uninduced MEFs versus control and six proteins were unique to the dataset of induced MEFs versus uninduced.

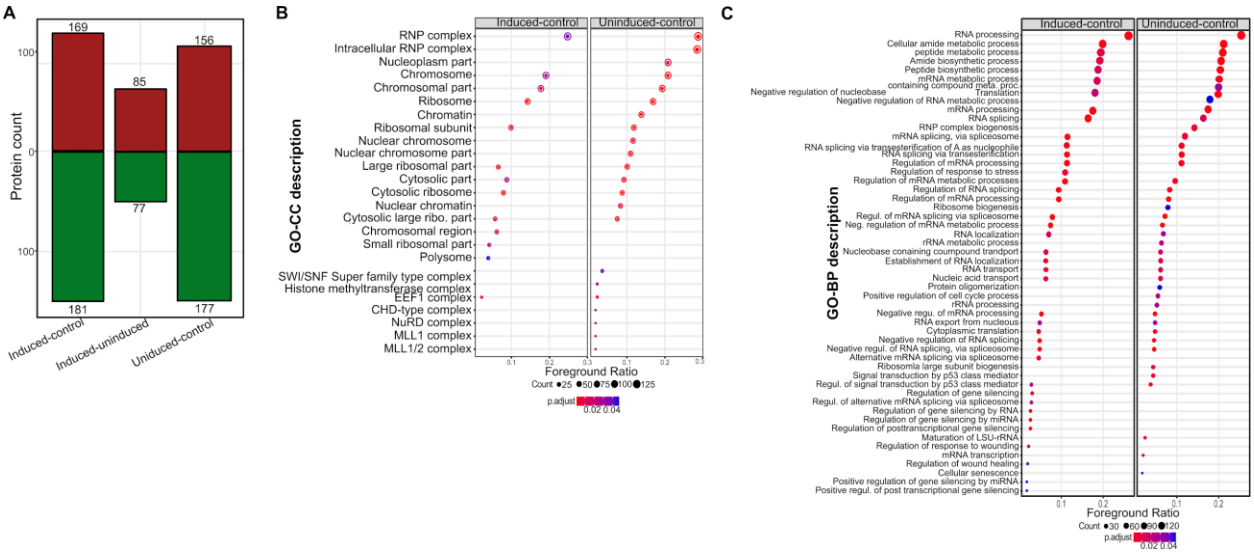


Fig.S8

GO term analysis of the proteins found in MEFs expressing β-actin-24MBS and MCP-eGFP-BirA* under two different conditions (uninduced, and serum-induced MEFs) versus control MEFs.

(A) Comparison of the number of significantly enriched proteins between all conditions. 169 proteins were significantly enriched in the induced sample (left bar) over control, while 156 proteins were overrepresented in uninduced over control (right bar). 85 proteins were enriched in the serum induced fractions over samples from uninduced MEFs (middle bar).

(B) Gene Ontology (GO) over-representation analysis of the proteins enriched in uninduced and induced conditions compared to the control. The foreground ratio represents the number of significantly enriched proteins divided by the number of proteins within each GO term. Color gradient represents the adjusted p-

value (threshold adj. p-value ≤ 0.05). Ranking of gene ontology term for cellular components (GO-CC) based on foreground ratio revealed ‘RNP complex” as one of the most over-represented function.
(C) Similar analysis as in (b) for gene ontology molecular function terms (GO-MF).

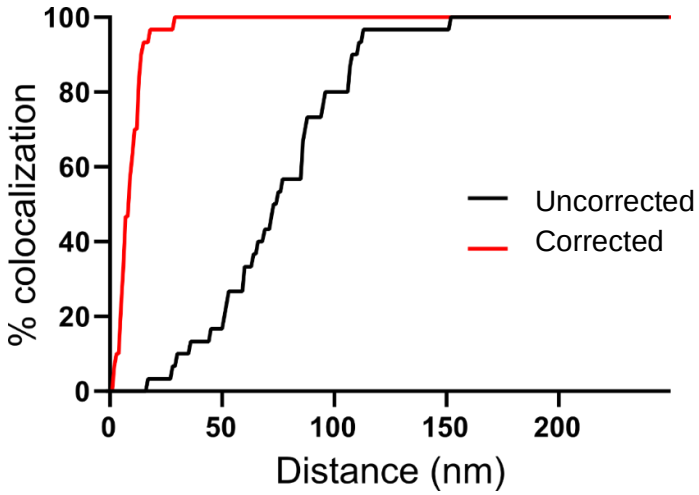


Fig S9.
Correction for chromatic and mechanical shifts by microsphere fluorescent beads Percentages of co-localization within spectrally separated centroids (collected from the beads) before (black line) and after (red line) correction was applied to the entire field of view.

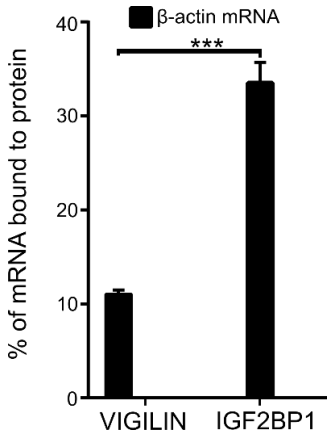


Fig S10.
Co-immunoprecipitation of β-actin mRNA with IGF2BP1 and VIGILIN.
Bars represent percentage of input mRNA co-purifying with the indicated protein. IGF2BP1 binds to around 33% of endogenous β-actin mRNA while VIGILIN was associated with 10% of endogenous β-actin mRNA which is non-significant compared to the binding efficiency of IGF2BP1 with the same mRNA. Error bars represents mean ± sem from three independent experiments. Statistical significance of each dataset was determined by Student's t-test; ***P < 0.001

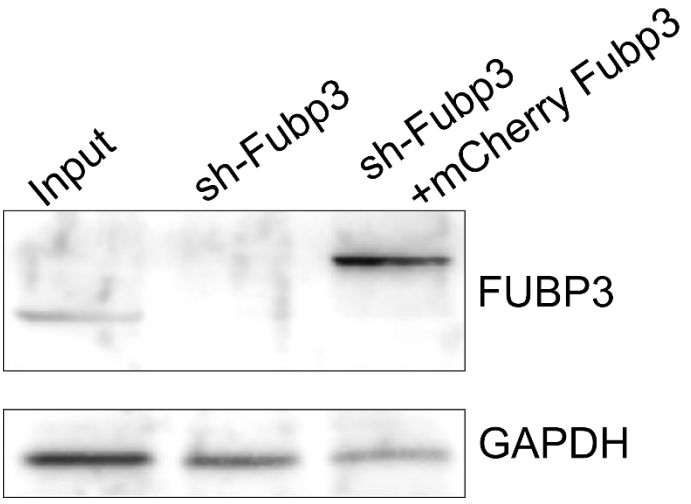
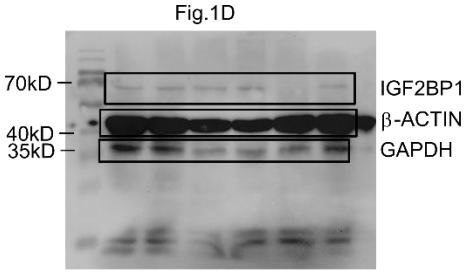
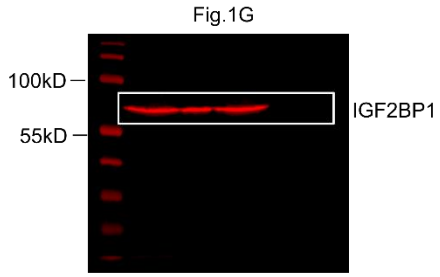
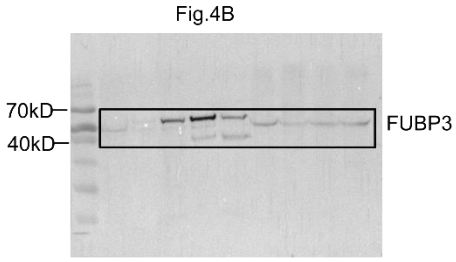
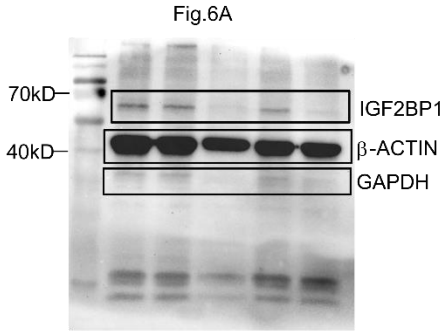


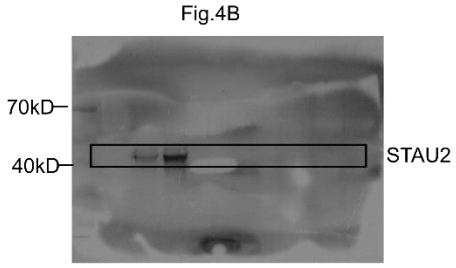
Fig S11.
FUBP3 expression in rescue experiment (related to Fig.6). Compared to input (lane 1: untreated), expression of FUBP3 or mCherry-FUBP3 was checked by western blot in MEFs with a stably transfected shFubp3 knockdown construct (lane 2) or in MEFs carrying this construct and overexpressing a knockdown-resistant mCherry-FUBP3.

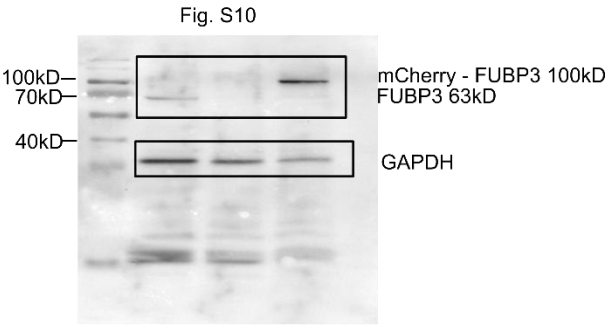


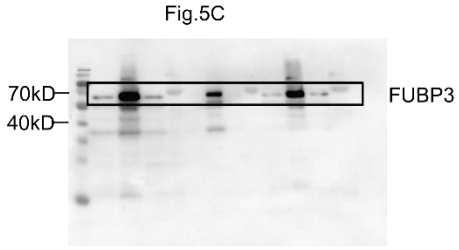












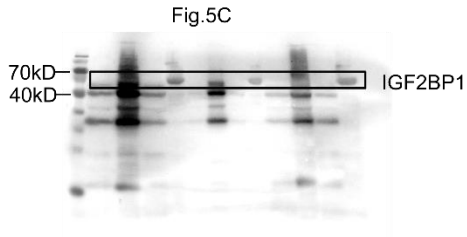


Fig S12.
Uncropped western blots related to indicated figures in the manuscript.

Supplementary Tables

Supplementary table 1: Primers and oligos used to generate plasmids

Plasmid	Forward primers/oligos	Reverse primers /oligos
pHAGE-NLS-2XMCP-eGFP	GCCTCTAGAATGGTGAGCAAGGGCGAGGAGCTGTT	CCGGATCGATTTACTTGTACAGCTCGTCCATG
pHAGE-NLS-2XMCP-eGFP-BirA*	GCCTCTAGAATGGTGAGCAAGGGCGAGGAGCTGTT	CCGATCGATTTACTTCTCTGCGCTTCTCAGGGAGATTT
N-terminal mCherry pcDNA3.1(+)	CTAGCTAGCATGGTGAGCAAGGGCGAGGAGGATAAC	CCCCTTAAGCTTGTACAGCTCGTCCATGCCGCCGG
mCherry Fubp3 pcDNA3.1(+)	CGGGGTACCATGGCGGAGCTGGTGCAGGGGCAGAGCGCTCCGGTGGG	ATAAGAATGCGGCCGCCTGCTCCTGGCTGTGGGCCTGAGCCTGCCC
mCherry Fubp3 KH1mt pcDNA3.1(+)	GTTGGATTTATTATTGGCGACGACGGTGAGCAGATTTCACGA	TCGTGAAATCTGCTCACCGTCGTCGCCAATAATAAATCCAAC
mCherry Fubp3 KH2 pcDNA3.1(+)	GTGGGGCTGGTCATCGGCGACGACGGGGAAACGATCAAGCAG	CTGCTTGATCGTTTCCCCGTCGTCGCCGATGACCAGCCCCAC
mCherry Fubp3 KH3 pcDNA3.1(+)	GTTGGGATTGTCATAGGAGACGACGGAGAAATGATTAAGAAG	CTTCTTAATCATTTCTCCGTCGTCTCCTATGACAATCCCAA
mCherry Fubp3 KH4 pcDNA3.1(+)	GTGCGGCCTTGTGATAGGCGACGACGGCGAGAACATCAAAAGCA	TGCTTTTGATGTTCTCGCCGTCGTCGCCTATCACAAGGCCGCAC
mCherry Fubp3 4KH mt pcDNA3.1(+)	CGGGGTACCATGGCGGAGCTGGTGCAGGGGCAGAGCGCTCCGGTGGG	ATAAGAATGCGGCCGCCTGCTCCTGGCTGTGGGCCTGAGCCTGCCC
Mod-pcDNA3.1(+)	CGGGCTAGCGTTTAAACTTAAGCTTGGTACCGAGCTCGG	CCATCGATATTTCGATAAGCCAGTAAGCAGTGGGTTCTC
β-actin-3'UTR mod-pcDNA3.1(+)	CCATCGATTAATACGACTCACTATAGGGCGGACTGTTACTGAGCTGCG	CGGGCTAGCTGTTTGTGTAAGGTAAGGTGTGCACTTTTATTG
Fubp3 pETM14	CATGCCATGGATGGCGGAGCTGGTGCAGGGGC	ATAAGAATGCGGCCGCCTGCTCCTGGCTGTGGGCCTGAGCCTGCCC

Igf2bp1 pETM14	CATGCCATGGACAAGCTTTACA TCGGCAACCTCAACGAGAGTG	GCCGAATTCCTTCCTCCGAGCCTGGGCCA GGTTGCT
pHAGE-mCherry FUBP3 shRNA rescue	F1: AGATGCCAACACGCCGCCAGA ATCATCAATGAGCTCATTCTC F2: : CGGGGTACCATGGCGGAGCTG GTGCAGGGG	R1: GAGAATGAGCTCATTGATGATTCTGGCGG CGTGTTGGCATCT R2: AAGGAAAAAAGCGGCCGCCTGCTCCTGGC TGTG

Supplementary table 2: Primers used for PCR or qRT-PCR

Gene/Insert	**Forward primers/oligos**	**Purpose**
eGFP F	GCCTCTAGAATGGTGAGCAAGGGCGAGGAGCTGTT	PCR
eGFP R	GGCCTCGAGCTTGTACAGCTCGTCCATGCCGAGAG T	PCR
BirA* F	GCCCTCGAGGACAAGGACAACACCGTGCCCCTGAA	PCR
BirA* R	CCGATCGATTTACTTCTCTGCGCTTCTCAGGGAGAT TT	PCR
Igf2bp1 F	TACAAGTGTTCATCCCCGCC	q-RT
Igf2bp1 R	AGGTGTTTCTGGTGGTGCAA	q-RT
β-actin R	GCCTTCACCGTTCCAGTTTTT	q-RT
β-actin F	CTGAGCTGCGTTTTACACCC	q-RT
β-actin F	CTGAGAGGGAAATCGTGCGT	RT
β-actin R	AGGGTGTAAAACGCAGCTCAG	RT
Gapdh F	GAGGGATGCTGCCCTTACC	q-RT
Gapdh R	AAATCCGTTCACACCGACCT	q-RT
Fubp3 F	TAGTACTCAGCCCAGGCCAT	q-RT
Fubp3 R	GTAGTGACTGGAGTGTGGGC	q-RT
β-actin zip code F	CGGACTGTTACTGAGCTG	q-RT
β-actin zip code R	CGCAAGTTAGGTTTTGTC	q-RT
β-actin proximal zip code F	GCAGAAAAAAAAAAAAT	q-RT

β-actin proximal zip code R	ACAAAGCCATGCC	q-RT
β-actin mutant zip code F	GTTACTGAGCTGCGT	q-RT
β-actin mutant zip code R	CGCAAGTTAGGTTTTGTC	q-RT
β-actin 3’ UTR F	CGGACTGTTACTGAGCTG	q-RT
β-actin 3’ UTR R	GCCTTCACCGTTCCAGTTTTT	q-RT
Sense-T7 Fubp3 bind-F	CGGCTAATACGACTCACTATAGGCCCCGGGGAAGGTGACAGCATTGCTTCTGTGTAAATTATGTACTGCAAAAATTTTTTTAAATCTTCCGCCTTAATAC	In vitro transcription
Sense-Fubp3 bind T7-R	GTATTAAGGCGGAAGATTTAAAAAAATTTTTGCAGTACATAATTTACACAGAAGCAATGCTGTCACCTTCCCCGGGGCCTATAGTGAGTCGTATTAGCCG	In vitro transcription
Sense-T7 Fubp3 mt bind-F	CGGCTAATACGACTCACTATAGGCCCCGGGGAAGGTGACAGCATTGCTTCTGTGTAAATACTGCAAAAATTTTTTTAAATCTTCCGCCTTAATAC	In vitro transcription
Sense-Fubp3 mt bind T7-R	GTATTAAGGCGGAAGATTTAAAAAAATTTTTGCAGTATTTACACAGAAGCAATGCTGTCACCTTCCCCGGGGCCTATAGTGAGTCGTATTAGCCG	In vitro transcription
T7-Zipcode-F	CCTAATACGACTCACTATAGGCGGACTGTTACTGAGCTGCGTTTTACACCCTTTCTTTGACAAAACCTAACTTGC	In vitro transcription
T7-Zipcode-R	GCAAGTTAGGTTTTGTCAAAGAAAGGGTGTAAAACGCAGCTCAGTAACAGTCCGCCTATAGTGAGTCGTATTAGG	In vitro transcription
T7-Zipmt-F	CCTAATACGACTCACTATAGGGTTACTGAGCTGCGTTTTTTTCTTTGACAAAACCTAACTTGC	In vitro transcription
T7-Zipmt-R	GCAAGTTAGGTTTTGTCAAAGAAAAAAACGCAGCTCAGTAACCCTATAGTGAGTCGTATTAGG	In vitro transcription
T7-zipcode proximal-F	CCTAATACGACTCACTATAGGGCAGAAAAAAAAAAAATAAGAGACAACATTGGCATGGCTTTGT	In vitro transcription
T7-zipcode proximal-R	ACAAAGCCATGCCAATGTTGTCTCTTATTTTTTTTTTTTCTGCCCTATAGTGAGTCGTATTAGG	In vitro transcription

Supplementary table 3: Antibodies used in this study

Antibody	Host	Dilution	Buffer (for Western blot)	Supplier
anti-β-ACTIN	Mouse-monoclonal	1:4000 - western	0.3 % BSA/TBST	Clone AC15 [Sigma-Aldrich (A1978)]
anti-GAPDH	Mouse-monoclonal	1:1000 - western	0.3 % BSA/TBST	Proteintech. 60004-1-Ig. Clone no. 1E6D9
anti-IGF2BP1	Mouse	1:4000 - Western 1:2000 - IF	0.3 % BSA/TBST	[MBL (RN001M)]
anti-FUBP3	Rabbit monoclonal	1:4000 - Western 1:1000 - IF	0.3 % BSA/TBST	Abcam: ab181122
anti-GFP	Chicken	1:5000 - IF		Aves Labs, Inc. (GFP-1010)]
anti-GFP	Rabbit polyclonal	1:4000 - Western	0.3 % BSA/TBST	Invitrogen- A-6455
anti-AP streptavidin		1:4000	100 mM NaCl, 50 mM MgCl2, 100 mM Tris-Cl pH 9.5	
anti-Mouse IgG HRP	Sheep	1:10.000	1 % BSA/TBST	Jackson-ImmunoResearch: 515-035-003
anti-Rabbit IgG HRP	Goat	1:10.000	1 % BSA/TBST	Jackson-ImmunoResearch:111-055-144
Alexa Fluor-labeled anti-chicken, -mouse, or -rabbit secondary antibodies	Goat	1:1000 - IF		Life Technologies

IF: Immunofluorescence

Supplementary table 4: Probes used for smFISH

smFISH probes used to detect the MS2 binding site (labeled on both 5' and 3' with ATTO488)

MS2_LK20	TTTCTAGAGTCGACCTGCAG
MS2_LK51_1	CTAGGCAATTAGGTACCTTAG
MS2_LK51-2	CTAATGAACCCGGGAATACTG

smFISH probes used to detect the β-actin ORF (labeled on 3' with ATTO633)

Actb807	ATAGTGATGACCTGGCCGTCAG
Actb830	CATCGGAACCGCTCGTTGCC
Actb863	ACCCAAGAAGGAAGGCTGGAA
Actb886	TCATGGATGCCACAGGATTCC
Actb927	CGGATGTCAACGTCACACTTCA
Actb960	CCAGACAGCACTGTGTTGGCAT
Actb984	ATGCCTGGGTACATGGTGGTAC
Actb1007	TCTCCTTCTGCATCCTGTCAGC
Actb1034	CATGGTGCTAGGAGCCAGAGC
Actb1073	CAGAGTACTTGCGCTCAGGAGG
Actb1099	CCAGGATGGAGCCACCGATC
Actb1122	ATCTGCTGGAAGGTGGACAGTG
Actb1145	CGTACTCCTGCTTGCTGATCCA
Actb1178	CTTGCGGTGCACGATGGAGGG
Actb1208	CGCAGCTCAGTAACAGTCCGC
ACTB1228	TCAAAGAAAGGGTGTAAAAC
ACTB1254	TTTTTTTTTTTTCTGCGCAAGTTAG
ACTB1284	AAAGCCATGCCAATGTTGTC
ACTB1355	GCGCCAAAACAAAACAAAAAAACTTA
ACTB1394	TCACCGTTCCAGTTTTTAAA

ACTB1423	ATGTTTGCTCCAACCAACTG
ACTB1446	CCACATTTGTAGAACTTTGG
ACTB1470	CAAAACAATGTACAAAGTCC
ACTB1511	GGAATGACTATTAAAAAAAGAC
ACTB1535	ACCACTTATTTCATGGATAC
ACTB1577	AGGAGTGGGGGTGGCTTTTG
ACTB1598	GGACGCGACCATCCTCCTCT
ACTB1626	ACCTTCCCCGGGGTGGACT
ACTB1661	ATTTTTGCAGTACATAATTTACAC
ACTB1683	TTAAGGCGGAAGATTTAAAAAAA
ACTB1724	ACCTGGGCCATTCAGAAATT
ACTB1749	GGGACAAAAAAAAGGGAGGC
ACTB1787	CTCCCAGGGAGACCAAAGCC
ACTB1810	GGCTGCCTCAACACCTCAAC
ACTB1831	GTCAGTGTACAGGCCAGCCC
ACTB1855	GGTGTGCACTTTTATTGGTC

Supplementary Methods

Plasmids and cloning: The lentivirus vector pHAGE-UbiC carrying NLS-2X MCP-tagRFPT (1) was used as backbone to generate all lentiviral vectors used in this study. In order to generate pHAGE-NLS-2XMCP-GFP, the GFP fragment was amplified from pEGFPC1 and cloned into the lentiviral vector within XbaI and ClaI restriction sites, thereby replacing the tagRFPT sequence. To generate pHAGE-NLS-2XMCP-eGFP-BirA*, a BirA* fragment was amplified from plasmid pSF3-TGN38-cMyc-BirA* (2) with XhoI and ClaI sites, eGFP fragment was amplified with XbaI and XhoI sites, after digestion both the fragments were ligated and the ligated product was PCR amplified using forward XbaI eGFP primer and reverse ClaI BirA* primer and the amplified product was integrated into pHAGE-NLS-2XMCP-eGFP plasmid under XbaI and ClaI sites.
Plasmids expressing mCherry fusion proteins were generated on pcDNA3.1(+) backbone. The mCherry CDS was amplified from plasmid pmCherry-C1 (Clontech), introducing NotI and XhoI sites at the 5' or 3' ends and the PCR product ligated into pcDNA3.1(+) to generate a C-terminal tag mCherry pcDNA3.1(+) entry plasmid. To create a N-terminal tag mCherry pcDNA3.1(+) entry plasmid, mCherry was amplified with NheI-mCherry forward and AflIII-mCherry reverse primer and cloned at the beginning of the MCS of pcDNA3.1(+).
Coding sequences of Igf2bp1 (Acc.No. NM_009951.4), and Fubp3 (Acc.No. NM_001290548.1), were amplified from Mouse (C57BL/6J) cDNA and cloned into C-terminal mCherry using restriction sites NheI and EcoRI (for Igf2bp1), KpnI and NotI (Fubp3).
To create plasmids expressing KH mutants of Fubp3-mCherry, Fubp3-mCherry was chosen as the template and different primer sets (Supplementary table 2) were used for inverse PCR to change the G-X-X-G motifs to G-D-D-G (3). After treating the PCR reaction with DpnI for overnight at 37°C, the linear amplification product was transformed into E.coli Dh5α for re-ligation and amplification.
To create a backbone vector for in vitro transcription, plasmid pcDNA3.1 was modified (mod-pcDNA3.1(+)) to remove unwanted restriction sites and to introduce a new restriction site before the T7 promoter. Two primers were designed for inverse PCR of pcDNA3.1(+), resulting in an additional ClaI site in front of the T7 promoter while at the same time removing a part of the multiple cloning site. DNA sequences encoding the β-actin zipcode, the proximal zipcode or a mutant zipcode were generated by annealing two complementary oligos containing NheI-ClaI overhangs at the ends and after phosphorylation ligated into the mod-pcDNA3.1(+) vector within NheI and ClaI sites. To generate pcDNA3.1(+) β-actin-3'UTR, the complete β-actin 3'UTR was amplified and cloned into the same sites.

Cell culture and serum starvation assay: Mouse embryonic fibroblasts (MEFs) from C57BL/6J mice or Hek293T cells were cultured in DMEM (with 4.5 g glucose and L-glutamine containing 10% fetal bovine serum (FBS) and 1% pen-strep). For serum starvation of MEFs, cells were starved in DMEM medium without serum (including 1% pen-strep) for at least 24 hrs and induced with DMEM media containing 10% FBS and 1% pen- strep for different time points as mentioned in the results section.

Lentivirus generation: Lentiviral particles were produced by transfecting the expression vector along with plasmids (Addgene plasmid numbers 12259, 12251, and 12253) for ENV (pMD2.VSVG), packaging (pMDLg/pRRE), and REV (pRSV-Rev) into HEK293T cells using Fugene HD reagent (Promega). The virus-containing supernatant was harvested on day 2 and day 3 and centrifuged at 500 x g for 10 min at 4˚ C and passed through 0.45 µm filter. The filtered supernatant was used for transduction of MEFs cells at a dilution of 1:100 and in presence of 8 mg/ml protamine sulphate. 3 days after the transfection, MEFs were washed 3x with serum containing media, trypsinized and washed twice again. Cells were washed once in FACS buffer (1x DPBS without calcium and magnesium, 0.2% BSA, 0.5 mM EDTA, 5 mM $MgCl_2$) and resuspended in FACS Cells expressing low levels of GFP were isolated with a FACS Aria cell sorter (Becton-Dickinson) before further culturing.

shRNA knockdown of Fubp3 and generation of FUBP3 rescue cell line. For generating stable cell lines expressing an shRNA construct against FUBP3 (Dharmacon clone ID. V3IMMMCG_11752124), GAPDH (Dharmacon clone ID. VSM11618), or a non-targeting control (Dharmacon clone ID. VSM11618), SMARTvector™ plasmids expressing inducible lentiviral shRNAs and shMIMIC™ inducible lentiviral microRNA were stably integrated in WT MEFs. For selection of positive cell lines 2µg/ml concentration of puromycin was chosen to establish stably integrated cell lines. For induction, dose and time response curves for doxycycline were generated and an incubation for three days with a doxycycline concentration of 200 ng/ml was chosen for the knockdown.
For the FUBP3 rescue cell line, a plasmid was generated using the PHAGE-UBC lentiviral backbone. The rescue construct contained an N-terminal mCherry sequence and silent mutations (CGG TGT CAG CAT GCA GCT CGC) in the shRNA target sequence (CGG TGC CAA CAC GCC GCC AGA). The target region is starting at base pair 926 n of the Fubp3 CDS. The resulting plasmid was stably integrated in FUBP3-shRNA MEFs and positive cells were collected by FACS sorting against mCherry.

Immunoprecipitation, Western blot, and qRT-PCR: For immunoprecipitation, cells were lysed in IP lysis buffer polysome extraction buffer (PEB) (without cycloheximide), atleast 200 µg of total protein was used per pull down experiment. For the rest of the experiment including western and qPCR, the protocol was followed as mentioned before (4) with the following modifications. A total of 200 µg of proteins were taken after lysing the cells in IP lysis buffer. 100 µl of protein A (for anti-mouse antibodies) or protein G (for anti-rabbit antibodies) coupled magnetic beads were used for IP. After washing them 3x in NT2 buffer (4) the beads were blocked with 5% BSA and 0.5 mg/ml ssDNA. After washing the beads once again with NT2 buffer, the beads were incubated with either 20 µg of FUBP3 antibody or 10 µg of IGF2BP1 in NT2 buffer at a total volume of 200µl, for overnight at 4°C with end to end rotation. After preclearing the lysates with 100µl of protein A or Protein G mag beads, antibody coupled beads were added in the lysates and incubated for 4 hrs at 4°C with end to end rotation. After the incubation, beads were separated from the lysates by a magnetic stand and washed 5 times with ice cold NT2 buffer and the supernatant was removed. Beads were resuspended in 100 µl of NT2 buffer. For isolation of proteins from the beads, 40 µl of beads were boiled in 100 µl 1x Laemmli buffer for 10 min at 95 C and the elute was separated from the beads on a magnetic strand. For western blots 40 µl from the eluted sample were used. For RNA isolation, in the 60 µl of the remaining beads, 5 µl proteinase K (10 mg/ml) and 1 µl of 10% SDS were added and incubated at 55°C water bath for 30 min. 100 µl of buffer NT2 was added in the sample and 200 µl of acidic phenol-chloroform mix was added and vortexed for 10 sec. After centrifugation at 16,000g at room temperature for 5 min, the upper aqueous layer was isolated, sodium acetate (pH 5.2, final concentration 0.3 M), 5 µl glycoblue (Invitrogen), and 600 µl of ethanol was added to precipitate the RNA. The RNA pellet was resuspended in 50 µl of RNase free water. For qPCR analysis, 500 ng of RNA was used after treating with 10U of DNase I.

NanoLC-MS/MS analysis and data processing: Beads were resuspended in denaturation buffer (6 M urea, 2 M thiourea, 10 mM Tris buffer, pH 8.0), and proteins were reduced by incubation in 1 mM dithiothreitol (DTT) for 1 h at room temperature. Alkylation of reduced cysteines was performed in 5.5 mM iodoacetamide (IAA) in 50 mM ammonium bicarbonate (ABC) buffer for 1 h at room temperature in the dark. On beads digestion of proteins was started with endoproteinase LysC (2 µg per 100 µg protein) for 3 h of incubation at pH 8.0 and room temperature. Tryptic digestion (2 µg per 100 µg protein) was performed overnight at room temperature after diluting the sample with four volumes of 20 mM ABC and adjusting the pH to 8.0. Acidified peptides were purified via PHOENIX Peptide Clean-up Kit (PreOmics) according to user manual, and separated on an EasyLC nano-HPLC (Thermo Scientific) coupled to an LTQ Orbitrap Elite (Thermo Scientific) as described elsewhere (5) with slight modifications: The peptide mixtures were injected onto the column in HPLC solvent A (0.1% formic acid) at a flow rate of 500 nl/min and subsequently eluted

with an 87 minutes segmented gradient of 5–33-50-90% of HPLC solvent B (80% acetonitrile in 0.1% formic acid) at a flow rate of 200 nl/min.
Precursor ions were acquired in the mass range from m/z 300 to 2000 in the Orbitrap mass analyzer at a resolution of 120,000. Accumulation target value of 106 charges was set. The 15 most intense ions were sequentially isolated and fragmented in the linear ion trap using collision-induced dissociation (CID) at the ion accumulation target value of 5000 and default CID settings. Sequenced precursor masses were excluded from further selection for 60 s. Acquired MS spectra were processed with MaxQuant software package version 1.5.2.861 with integrated Andromeda search engine (6). Database search was performed against a target-decoy Mus musculus database obtained from Uniprot, containing 60,752 protein entries, the sequence of MCP-eGFP-BirA and 284 commonly observed contaminants. Endoprotease trypsin was defined as protease with a maximum of two missed cleavages. Oxidation of methionine and N-terminal acetylation were specified as variable modifications, whereas carbamidomethylation on cysteine was set as fixed modification. Initial maximum allowed mass tolerance was set to 4.5 ppm (for the survey scan) and 0.5 Da for CID fragment ions. Peptide, protein and modification site identifications were reported at a false discovery rate (FDR) of 0.01, estimated by the target/decoy approach (7). The label-free algorithm was enabled, as was the “match between runs” option (8).

Downstream analysis and functional interpretation: Downstream analyses were performed in the R environment. The resultant proteome profiles obtained were quality checked for replicate correlation using principal component analysis and hierarchical clustering. The data was then filtered for low abundant proteins via a two-step process. Firstly, all proteins were ranked (in descending order) by LFQ intensity and the 3rd quartile LFQ value set as the minimum intensity threshold. Secondly, to pass the filtering, all samples within a condition must have an LFQ intensity above this threshold. Contaminants and reverse hits, as well as proteins only identified by sites, were also removed from downstream analyses. Following low abundance filtering the LFQ values were quantile normalized using the MSnbase package (10). In order to statistically compare protein abundances across conditions (even when the protein was not detected in one of the conditions) imputation was used on the data using a mixed model of nearest neighbor averaging and left-censored missing data from a truncated distribution (10). Comparisons in which both conditions contained imputed values were completely discounted. Proteins which were significantly different between conditions were identified using ANOVA followed by the Tukey post-hoc test. Significance was set at an adjusted p-value of 0.05 following Benjamini-Hochberg multiple correction testing. Functional information, namely gene ontology (GO) and KEGG ID, for Mus musculus proteins were retrieved using the UniProt.ws package (11). The over-representation testing for GO and KEGG pathways were done for each comparison via the cluster Profiler package67 based on hypergeometric distribution (p-adj. ≤ 0.05). Further analysis were performed using the Perseus software (11, 12).

Expression of recombinant proteins and *in vitro* binding assay with *in vitro* transcribed RNA: Full length Igf2bp1, Fubp3, were cloned into vector pETM 14 containing a HIS tag. For pETM 14-Fubp3, a 1704 bp long fragment of Fubp3 was cloned into NcoI and NotI. For pETM 41-Igf2bp1, a 1724 bp long fragment of Igf2bp1 was cloned into NcoI and EcoRI sites. The proteins were expressed in Rossetta-gami™ 2(DE3) cells according to manufacturer’s protocol (Novagen). To generate constructs for *in vitro* transcription, the 683 bp long full length β-actin 3’UTR was cloned between NheI and ClaI sites into a modified pCDNA3.1(+) (described above in Plasmids and Cloning) that contains a T7 instead of the CMV promoter. For the 54 nt long localization element of β-actin, the 49 bp long region proximal to the zipcode, the mutated zipcode (13), the 79 nucleotide long region containing UAUG motif, and the 75 nucleotide long region lacking the UAUG regions, two oligonucleotides containing the corresponding sequences and a T7 promoter on the forward strand were annealed. For *in vitro* transcription, either 1 µg of the linearized plasmid (digested with NheI), or the annealed oligonucleotides were used. Transcription was done for 1 h at 37°C in 10 µl of 40 mM Tris-HCl, pH 7.5, 6 mM Mg-Acetate, 10 mM DTT, 1 mM spermidine, 0.5 mM each of ATP, GTP and

CTP, 10 µM UTP, 100 units/ml RNase inhibitor, and 500 units/ml of T7 RNA polymerase. RNA was recovered by ethanol precipitation at -20°.

For the *in vitro* binding assay, 40 µg of total protein lysates were pre-cleared with empty magnetic beads before incubation with 50 pmol of in vitro transcribed RNAs at 4°C for at least 4 hrs with end to end rotation. His-tagged proteins and bound RNAs were captured with 30µl of His60 Ni Magnetic Beads (Takara, cat no. 635693), washed twice with 1xPBS, once with buffer containing 250 mM NaCl and 10 mM Tris-HCl pH7.5 and finally once with 10 mM Tris-HCl pH7.5. Half of the beads were used for protein isolation and western blot, the rest for RNA isolation. Bound RNA was determined using qRT PCR and primers according to supplementary table 1 and 2.

References

1. Halstead JM, et al. (2015) An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science (80-)* 347(6228):1367–1370.
2. Béthune J, Artus-Revel CG, Filipowicz W (2012) Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* 13(8):716–723.
3. Hollingworth D, et al. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res* 40(14):6873–86.
4. Keene JD, Komisarow JM, Friedersdorf MB (2006) RIP-Chip: The isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 1(1):302–307.
5. Franz-Wachtel M, et al. (2012) Global Detection of Protein Kinase D-dependent Phosphorylation Events in Nocodazole-treated Human Cells. *Mol Cell Proteomics* 11(5):160–170.
6. Cox J, et al. (2011) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10(4):1794–1805.
7. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4(3):207–214.
8. Luber CA, et al. (2010) Quantitative Proteomics Reveals Subset-Specific Viral Recognition in Dendritic Cells. *Immunity* 32(2):279–289.
9. R Foundation for Statistical Computing (2016) R: A Language and Environment for Statistical Computing. *R Found Stat Comput* 1(3.3.2). doi:10.1007/978-3-540-74686-7.
10. Gatto L, Lilley KS (2012) Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* 28(2):288–289.
11. Carlson M (2018) *UniProt.ws: A package for retrieving data from the UniProt web service* Available at: https://www.bioconductor.org/packages/devel/bioc/vignettes/UniProt.ws/inst/doc/UniProt.ws.pdf [Accessed August 2, 2018].
12. Tyanova S, et al. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13(9):731–740.
13. Nicastro G, et al. (2017) Mechanism of β-actin mRNA Recognition by ZBP1. *Cell Rep* 18(5):1187–1199.

bio-protocol

RNA interactome identification via RNA-BioID in mouse embryonic fibroblasts

Joyita Mukherjee[1], Mirita Franz-Wachtel[2], Boris Maček[2], Ralf-Peter Jansen[1]*

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany.

[2]Proteome Center Tübingen, University of Tübingen, Tübingen, Germany.

*For correspondence: ralf.jansen@uni-tuebingen.de

[Abstract] Cytoplasmic localization of mRNAs is common to all organsims and serves the spatial expression of genes. Cis-acting RNA signals (mostly found in the mRNA's 3'-UTR), called zipcodes recruit trans acting RNA-binding proteins that faciliate the localization of the mRNA. UV-crosslinking or affinity purification have been applied to identify such proteins but suffer from the need of stable RNA-protein binding or direct contact of protein and RNA. To identify stably or transiently interacting proteins that directly or indirectly associate with the localization elements and the body of the mRNA, we developed an in vivo proximity labeling method, RNA-BioID (*Mukherjee et al.,2019*). In RNA-BioID, we tether a fusion of the BirA* biotin ligase and the MS2 coat protein (MCP) at the 3'-UTR of MS2 tagged β-actin mRNA in vivo. Exposing BirA* expressing cells to biotin in the media and induces biotinylation of β-actin mRNA associated proteins that can be isolated with streptavidin beads. This technique allowed us to identify by mass spec analysis the β-actin mRNA 3'UTR interacting proteome in fibroblasts. The protocol can be useful to identify the interacting proteome of any mRNA in mammalian cells.

Keywords: RNA-BioID, proximity labeling, biotinylation, RNA binding protein, RNA tagging

[Background]

Localization of mRNAs to specific subcellular sites is a widespread phenomenon and has been observed from bacteria to humans (*Fei et al., 2018; Bovaird et al.,2018*). Diverse human pathologies of the neural system are linked to defects in mRNA localization (*Tolino et al., 2012; Bovaird et al., 2018*). Additionally, this mechanism is essential for developmental processes including mesoderm formation in the clawfrog *Xenopus laevis* or the determination of the embryonic body axes of the fruit fly *Drosophila melanogaster* (*Jansen et al., 2001*). To achieve intracellular asymmetry of transcripts, mRNAs are incorporated into motor-protein containing particles. that can move along the cytoskeleton to distinct cellular sites, where the transcripts are locally translated (*Marchand et al., 2012; Buxbaum et al., 2015).* Such particles containing mRNA and RNA-binding proteins (RBPs) are beginning to assemble during transcription and undergo several changes upon maturation in the cytoplasm (*Martin et al., 2019*). Localized mRNAs differ from non-localized ones by containing cis acting motifs termed localization elements or zipcodes that are recognized by specific RBPs (*Bovaird et al., 2018*). Zipcodes can be positioned in the coding sequence but are mainly found in the 3' untranslated region (3'-UTR; *Bovaird et al., 2018*).

A well studied example is the 54nt long zipcode in the 3'-UTR of β-actin mRNA that is essential for

targeting the mRNA to the leading edge of fibroblasts as well as to growth cones and dendrites in neurons (*Ross et al., 1997; Eom et al., 2003; Buxbaum et al., 2015*). The zipcode is recognized by the RBP ZBP1 (aka IGF1 or IGF2BP1) that is not only essential for the localization of β-actin mRNA (*Ross et al., 1997; Oleynikov et al., 2003*) but also represses translation the mRNA until it has arrived at the target site (*Hüttelmaier et al., 2005*). However, additional RBPs and supplementary factors are involved in splicing, translational regulation, stabilization, transport an decay of β-actin mRNA.

Obtaining a complete set of the proteins that associate with any mRNA remains challenging, despite a large range of high throughput methods available so far that include diverse pulldown approaches as well as cross-linking and immunoprecipitation(CLIP) (for an overview see *Ramanathan et al., 2019*). A specific challenge remains the identification of transient interactors that are not directly coantacting the RNA. To solve this, we recently introduced a proximity biotinylation method based on tethering the BioID biotin ligase BirA (BirA*) via MCP (MS2 coat binding protein) at the 3'-UTR of MS2 tagged β-actin mRNA (*Park et al., 2014*). In presence of biotin, BirA* generates AMP-biotin ('activated biotin'), which reacts with accessible lysine residues in its vicinity (10 - 20nm). After lysis, biotinylated proteins can be isolated via streptavidin affinity purification and identified using standard mass spectrometry techniques. The RNA-BioID method can be applied to any MS2 tagged specific localized mRNA which will allow not only the identification of its associated proteins but can also be used to visualize and probe the environment of this mRNA.

Materials and Reagents

1. phage – ubc – nls – 2xmcp – egfp – BirA* (Addgene Plasmid # 131132) or phage – ubc – nls – 2xmcp – egfp – cMyc BirA* (Addgene Plasmid # 131133) or phage – ubc – nls – 2xmcp – mCherry – BirA* (Addgene Plasmid # 131136)
2. pCEP4-tat (Addgene Plasmid # 22502)
3. pRSV-Rev (Addgene Plasmid #12253)
4. pMDLg/pRRE (Addgene Plasmid #12251)
5. pMD2.G (Addgene Plasmid #12259)
6. Hek293 FT cells (Thermo scientific, Cat. No. R70007)
7. Wildtype mouse embryonic fibroblasts (MEFs; e.g. *Park et al., 2014*)
8. MBS MEFs (MEFs with 24 MS2 binding sites (MBS) inserted in the 3'-UTR of β-actin gene loci; see *Park et al., 2014*)
9. T175 flasks (Corning, Cat. No. CLS431080)
10. 10cm culture dishes (Corning, Cat. No. 430167)
11. 6 well dishes (Corning, Cat. No. CLS3516)
12. FACS tubes (Falcon, Cat. No. 352235)
13. 1.5 ml safe seal pyrogen free tubes (Sarstedt, Cat. No. 72.706.201)
14. 0.45 µM filter (Sartorius, Cat. No. 10109180)
15. OptiMEM (Thermo scientific, Cat. No. 31985062, store at 4°C)

16. DMEM high glucose media (Sigma, Cat. No. D6429, store at 4°C)
17. Fugene 6 (Promega, Cat. No. E2691, store at 4°C)
18. Trypsin (Sigma, Cat. No. T4049, store at -20°C)
19. Foetal calf serum (FBS; Sigma, Cat. No. F4135, store at -20°C)
20. 1x PBS (Sigma, Cat. No. D6429, store at 4°C)
21. Penicilin – Streptomycin solution (Sigma, Cat. No. P4333, store at -20°C)
22. Bleach powder (Schülke perform® disinfectant, Art. No. 107912, store at room temperature)
23. Bench cleaner (terralin ® PAA, Art. No. 126203, store at room temperature)
24. Hand disinfectant (STERILLIUM Virugard solution, Art. No.: 126064, store at room temperature)
25. LentiX concentrator (Takara, Cat. No. 631231, store at 4°C)
26. Biotin (Sigma, Cat. No. B4501, store at 4°C)
27. Magntic sepharose streptavidin Beads (GE, Cat. No. 28985799, store at 4°C)
28. Complete protease inhibitor, EDTA free (Roche, Cat. No. 11697498001, store at 4°C)
29. 5X Roti Quant Bradford reagent (Roth, Art. No. K015.3, store at 4°C)
30. SDS (sodium dodecyl sulfate) (Roth, Art. No. CN30.1, store at room temperature)
31. Bromophenol blue (Roth, Art. No. A512.1, store at room temperature)
32. Glycerol (Sigma, Cat. No. G6279, store at room temperature)
33. Nu-PAGE SDS page gel 4 – 12% (Invitrogen, Cat. No. WG1402BOX, store at 4°C)
34. NuPAGE MOPS SDS Running Buffer (20X) (Invitrogen, Cat. No. NP0001, store at 4°C)
35. PVDF membrane (GE Healthcare Amersham™ Hybond™ -P Membranes, Product Code.10471085, store at room temperature)
36. Alkaline phosphatase conjugated streptavidin.
37. β-mercaptoethanol (Roth, Cat. No. 42273, store at room temperature)
38. p-Nitrotetrazoliumblauchlorid (NBT) (Roth, Cat. No. 4421.2, store at -20°C)
39. 5-Bromo-4-chloro-3-indolyl-phosphate Disodium Salt (BCIP) (Roth, Cat. No. A155.2, store at -20°C)
40. Ammonium Bicarbonate (Sigma, Cat. No. 09830, store at room temperature)
41. NuPage SDS Page gels (Thermo scientific, Cat. No. NP0315BOX, store at 4°C)
42. Trypsin, sequencing grade, modified, 0.5 µg/µl (Promega, Cat.No. V5113)
43. PHOENIX Peptide Cleanup Kit (PreOmics, Cat. No. P.O.00023)
44. Dithiothreitol (VWR, Cat. No. 0281-5G, store at room temperature)
45. Urea (Merck, Cat. No. 1.08488.1000, store at room temperature)
46. Thiourea (Merck, Cat. No. 1.07979.0250, store at room temperature)
47. Iodoacetamide (Merck, Cat. No. 8.04744.0025, store at room temperature)
48. Acetonitrile (Merck, Cat. No. 1.00030.2500)
49. Formic acid (Sigma-Aldrich, Cat. No. 33015-500ML)
50. Acetic acid (Roth, Art. No. 3738.5)
51. Trifluoracetic acid (Merck, 1.08262.0100)
52. Lysyl endopeptidase LysC (Waco, Cat. No. 129-02541), stocks: 0.5 µg/µl in H2O

53. Reprosil-Pur C18-AQ, 1.9 µm resin (Dr. Maisch GmbH)
54. 25 cm fused silica emitters with an inner diameter of 75 µm (DNU-MS GbR)

Equipment

1. 37˚ C waterbath (Fisher Scientific Isotemp 205 Water Bath)
2. Cell culture incubator, set to 5% CO_2 (Thermo Scientific Heracell™ 150i CO2 Incubators)
3. Centrifuge (Eppendorf 5430)
4. End to end rotator
5. Thermomixer (Eppendorf, Cat. No. 5350 000.013)
6. Magnetic stand (BioRad, Cat. No. 1614916)
7. Centrifuge (Heraeus Fresco 17, Thermo Fisher Scientific, Cat. No. 75002420)
8. Concentrator plus vacuum centrifuge (Eppendorf, Cat. No. 5305 000.304)
9. EASY-nLC 1200 (Thermo Fisher Scientific) or any (U)HPLC system compatible with MS system of choice
10. LTQ Orbitrap or Q Exactive mass spectrometer (Thermo Fisher Scientific) or any state-of-the-art high resolution mass spectrometer.

Software

1. MaxQuant software suite (www.maxquant.org, Max-Planck-Institute of Biochemistry, Martinsried), or other programs that can be used for downstream analysis
2. Perseus (http://coxdocs.org/doku.php?id=perseus:start, Max-Planck-Institute of Biochemistry, Martinsried)

Procedure

A. Generating stable cell lines expressing a MCP-GFP-BirA* fusion protein

1. The MCP-GFP-BirA* gene fusion under control of the UBC romoter was integrated into a third generation lentiviral vector and is available via Addgene (Addgene no. 131132 / 131133 / 131136). The three different versions behave the similar ways.
2. For transduction, use the MCP-GFP-BirA* plasmid Gag/Pol in combination with Env and VSV-g plasmids to assemble and generate the complete viral particles in Hek293FT cells.
3. On the day before transfection, seed 4 x 10^6 cells Hek293FT cells in complete growth media in a 10 cm dish.
4. On the day of transfection, set up two 1.5 ml centrifuge tubes, each containing 500 µl of OptiMEM media.
5. To one tube add the corresponding amounts of the following plasmids:

MCP-GFP-BirA* backbone plasmid: 20 µg
pCEP4-tat: 1 µg
pRSV-Rev: 1 µg
pMDLg/pRRE: 1 µg
pMD2.G (Insert name VSV G): 2 µg
In total 25 µg of plasmid DNA.

6. To the other OptiMEM-containing centrifuge tube add 3 total plasmid concentration volume of Fugene HD transfection reagent, e.g. in case of 25 ug add 75ul.
7. Incubate both tubes separately for 5 min at room temperature.
8. Add the OptiMEM media containing plasmid DNA to the tube containing the OptiMEM-Fugene HD mix, mix briefly, and incubate for at least 25 min (up to 45 min) at room temperature. This results in formation of the DNA-Fugene complex.
9. At the end of this incubation, exchange the media of the 10 cm dish containing Hek293FT cells with fresh 10 ml complete growth media.
10. Add the DNA-fugene complex on top of the cells in a dropwise manner. Swirl the dish gently to mix the fugene-DNA mix with the medium and incubate the dishes at 37°C for at least 16 hours.

    Critical: For biosafety reasons, keep the dishes on a tray inside the incubator to avoid spilling of the media in the incubator due to handling errors.
    Critical: At this stage, the lentivirus is being produced, so strictly follow the biosafety class II regulations. After the 16 hr incubation period, all dishes and pipettes that have and will come in contact with virus-containing solutions have to be treated with bleach prior to disposal. Always wear a labcoat and wipe the entire hood with ethanol. When you leave the working space treat your gloves with bleach as well.
11. Before the end of the 16 hr incubation, prepare 500 ml of 10% bleach and keep it in a large beaker in the tissue culture hood for treating surfaces (prepare fresh 19% bleach every day, it remains effective only for 24hrs).
12. Once the bleach solution is ready under the hood, carefully remove 8ml of the media from the 10cm dish and transfer it to a 50 ml conical tube. Keep it at 4˚ C for overnight.
13. Add 8ml fresh complete growth media to the dish and return it to incubator for another 24 hours.
14. At this time, seed wildtype (WT) MEFs and MBS MEFs (3 x 10^6/10 cm dish) and incubate at 37°C in a cell culture incubator (5% CO_2).
15. Repeat step 11 to 13 twice more during the consecutive days. Collect the media from each day and pool it in the same 50 ml conical tube from step 12, keep at 4˚ C. After collecting all supernatants, spin the conical tube for 10 min at 500 xg and 4°C to pellet large debris.
16. Filter the supernatant from this centrifugation through a 0.45 µm filter.
17. Concentrate the lentiviral particles in the supernatant with a LentiX concentrator according to manufacturer's instructions.
18. After concentrating the viral particles, add 1ml of serum free media and carefully resuspend the

pellet.

19. Remove the dishes with WT MEFs and MBS-MEFs from the incubator for transduction, replace the old media with fresh complete medium and add 100 µl of the viral particle solution per 10cm dish with 10ml of media. Store the rest of the virus particles at -80˚ C. Place the MEFs back into the cell culture incubator and leave for 24 hrs.
20. After 24 hrs, check the viral transduction efficiency under a fluorescence microscope (filter setting for GFP). If no GFP fluorescence is visible, repeat step 19 with another aliquot of the stored viral particles.
21. Fluorescent cells should stably express the MCP-GFP-BirA* fusion protein.
22. On last day of collection, bleach all the dishes that came in contact with the viral particles.

B. Isolating cells expressing a functional MCP-GFP-BirA* fusion protein

1. After generation of stable cell lines (WT-BirA*: WT MEFs expressing MCP-GFP-BirA*; MBS-BirA*: MEFs with 24xMBS in the 3'-UTR of β-actin and expressing MCP-GFP-BirA*), sort positive, GFP expressing cells by fluorescence activated cell sorting against GFP. This can be done by a core facility at your institution.
2. After sorting the cells into FACS tubes containing 1ml of complete media, spin down the cells at 500 xg for 5 min at room temperature. Resuspend and, depending on the number of positive cells identified in cell sorting, culture them either on 10cm dish or bigger dishes (10,000 cells / each 6 well).
3. When atleast 10,000 cells have been isolated, test for the GFP expression levels in both transduced cell types (WT-BirA* / MBS-BirA*). For checking GFP expression, prepare cell lysates in the lysis buffer as mentioned in (*Mukherjee et al., 2019*) and use at least 10 µg of total protein for a western blot against GFP and a house keeping gene (e.g. GAPDH). Use untransfected WT MEFs as controls for GFP (should be negative) and GAPDH (should be positive). This step is important to check whether the full NLS-HA-2XMCP-eGFP-BirA* construct is expressing giving a band at 91.4 kDa (Figure 1A).
4. To check for the activity of BirA* and biotinylation efficiency, use a 6 well dish. Seed duplicates of 3 x 10^4 cells of WT untransfected, WT-BirA* and MBS-BirA* and culture them in complete media for 24 hrs. On the next day, exchange the media. Use complete media containing 50uM of biotin with only on one well of each duplicates. To the second set of duplicates just exchange the media with new complete media without biotin.
5. Culture the cells with the new media (with or without biotin) for at least 6 hrs up to 24 hrs. Biotinylation efficiency might not increase after 6 hrs, depending of the cell line.
6. At the end of the incubation, wash the wells of the 6 well dishes two times with cold PBS and finally isolate the cells by adding trypsine.
7. For trypsinization, add 500ul/well of 6 well plate of warm trypsin solution and incubate the dish at 37˚ C for 2 - 3 minutes. After 2 min, check under a cell culture microscope (at 10X magnification) if all cells are completely detached. If not, incubate the cells in the incubator a

bit longer but not more than for 5 min.

8. After complete detachment of the cells from the wells, remove the solution with the cells and transfer them to a 15 ml conical tube. Add 6 ml of complete media (the FBS in the media deactivates the typsine).
9. Spin down the cells at room temperature for 5 min at 500 xg.
10. After centrifugation, remove the media and add 1 ml of cold 1X PBS, resuspend the pellet with a 1 ml tip and pellet the cells again by centrifugation at room temperature for 4 min at 500 xg at 4°C.
11. Repeat step 10 again to completely remove the media.
12. Add 3 PCV (packed cell volume) lysis buffer containing 1X complete protease inhibitors to each pellet, resuspend cells and keep them on ice for 10 min. After 10 min, lyse the cells by passing the pellet 10 - 15 times through a 21G needle. After the lysis, spin the lysates at 12,000 xg for 10 min at 4°C to remove the unbroken cells.
13. Determine the protein concentration (e.g. by a Bradford assay following the manufacturer's protocol).
14. Mix 10 µg of total protein with SDS PAGE loading buffer.
15. Boil the mixture for 5 min at 95°C and afterwards spin down debris at 12,000 xg at room temperature for 3 min.
16. Load the supernatant from step 15 on a Nu-PAGE SDS page gel in 1X MOPS running buffer and run at 200 V until the front dye exits the gel.
17. After the run is complete, transfer the protein from the gel onto an activated PVDF membrane by semi-dry western blotting.
18. After the transfer is complete, wash the PVDF membranein a small box with TBST twice for 5 min each at room temperature with rotation or tilting.
19. In order to probe for biotinylated proteins, incubate with alkaline phosphatase conjugated streptavidin.
20. Incubate the membrane with AP substrate buffer for 5 min at room temperature.
21. Remove AP substrate buffer. In a 15 ml conical tube, mix 5 ml of freshly made AP substrate buffer together with NBT (final concentration 50ug/ml) and BCIP (final concentration 50ug/ml). Pour the mixture on top of the membrane.
22. Incubate in dark on a shaker and check at 1 min intervalls until you can see purple bands from the biotinlyted proteins (Figure 1B).
23. Once you tested the biotinlation is working properly, continue with large scale biotinylation experiment for purification of bioyinylated proteins and mass spec.

C. RNA-BioID

1. To capture and analyze biotinylated proteins in RNA-BioID, you need a confluent T175 flask of cells (2-3 x 10^7) for each round of mass spectrometric analysis.

2. At the day before the streptavidin pulldown, confluency should have reached 80%. Exchange media of one T175 flask, containing BirA* or MBS-BirA* cells biotin-containing medium and culture for 16 - 24hrs . For control of biotinylation, also take a T175 flask containing MBS-BirA* cells but donot add biotin. This control ‘ll provide information about background biotinylated proteins in the cell, also about the proteins which ‘ll bind to the beads unspecifically.
3. Next day, wash the cells twice with PBS at room temperature to dispose dead cells and serum-containing medium. Trypsinize for 2 - 3 mins in the incubator. After checking the detachment under a microscope, transfer cells into conical tube and add 5 ml of serum containing media to inactivate the trypsin.
4. Collect cells by centrifugation at 500 xg for 5 min at 4˚ C.
5. Wash cells twice with ice-cold PBS to completely remove the media.
6. Add 3 pcv (packed cell volume) of lysis buffer (including Complete™ protease inhibitor mix) to each pellet, and resuspend the pellet.
7. Lyse cells, collect lysate and determine protein concentration as mentioned above (section B, step 12 - 13)
8. For the isolation of biotinylated proteins, use at least 1 mg of total protein per lysate and 300 µl of magnetic streptavidin beads for each 1mg of protein.
9. To prepare the beads, mix the bead solution in the orginal vendor’s tube by a quick vortexing, remove 200 µl of bead slurry to a fresh 1.5 ml tube and separate the beads from its storage buffer by capturing the beads on a magnetic stand for 5min. Remove storage solution and wash the beads three times with 1 ml of lysis buffer. Rotate tubes with the bead slurry on a end to end rotator for 5 min.
10. After the final wash, the beads are collected on a magnetic stand. Add 1 mg of total cell lysate onto the beads, remove from magnetic stand and leave mixture overnight with end to end rotation at 4˚ C.
11. On the next day, collect the beads on a magnetic strand and wash twice with wash buffer 1, once with wash buffer 2, once with wash buffer 3 and once with wash buffer 4, each time by end to end rotation for 5 min at room temperature.
12. During the final washing with 1 ml of wash buffer 4, remove 100 µl and collect the beads which ‘ll leter be for checking with silver staining. Collect the beads form the remaining 900 µl on a magnetic stand.
13. Wash the beads 3 times with 50 mM of ammonium bicarbonate buffer for 5 min at room temperature with rotation. After the final wash, collect the beads. Either immediately proceed with section D or store beads at -80 until further processed.
24. From the remaining beads from step C-12, add 10µl of 1X SDS sample buffer, boil the mixture for 5 min at 95°C and afterwards spin at 12,000 xg at room temperature for 1 min.
14. Keep the centrifuge tube on a magnetic stand for 2 min at room temperature and take out the supernatant. Load 20 µl of the supernatant on a Nu-PAGE SDS page gel in 1X MOPS running buffer and run at 200 V until the front dye exits the gel.

15. Perform silver staining on this gel to compare proteins from MBS-BirA* cells with or without treated with biotin (*Aboulaich et al., 2011*) (Figure 1C).

D. On-bead tryptic digestion

1. Add an appropriate volume of denaturation buffer to the beads (usually 30 - 80 µl), such that they are well covered and can float in the suspension during incubation.
2. Add reduction buffer to a final concentration of 1 mM, incubate 1 hr at room temperature in a Thermomixer at 1000 rpm.
3. Add alkylation buffer to a final concentration of 5.5 mM, incubate 1hr at room temperature in a Thermomixer at 1000 rpm.
4. Add 1 µg LysC for predigestion and check pH on a pH 6-10 pH strip. The pH needs to be around pH 8, otherwise adjust with 1M Tris-base pH 8. Incubate 3 hrs at room temperature in the Thermomixer at 1000 rpm.
5. Add 4 volumes of 20 mM ammonium bicarbonate buffer to dilute the urea in the buffer.
6. Add 1 µg trypsin for digestion and check pH on a pH 6-10 pH strip (pH should be 8), incubate over night at room temperature in a Thermomixer at 1000 rpm.
7. Separate beads and supernatant by using a magnetic stand and transfer the supernatant to a new tube.
8. Acidify samples with TFA to $\leqslant$ pH 2.

E. Peptide cleanup

1. Load sample onto a PHOENIX Peptide Cleanup Kit device.
2. Perform cleanup procedure by following the kit protocol (washing and elution steps).
3. After elution, reduce the sample volume to ≤ 9 µl in a vacuum concentrator (30 – 40 min at 30° C).
4. Add 1 µl solvent A* and adjust the sample volume to 10 µl with solvent A.

F. Mass spectrometry analysis

1. Peptide samples are analysed using online nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS). Note that this is a routine analysis in proteomics that can be done on any high performance liquid chromatography system coupled to a state-of-the-art high resolution mass spectrometer.
2. Here, nanoLC-MS/MS analyses are done on an EASY-nLC 1200 system connected to an LTQ-Orbitrap Elite through an electrospray Ionization source:
3. Pack a 25 cm long 75 µm-inner diameter analytical column with reversed-phase C18-AQ Reprosil-Pur 1.9 µm particles at a constant pressure of 600 bar, and cut the column to a final length of 20 cm.
4. Load peptides directly onto the column at a flow rate of 500 nL/min.
5. Reduce flow rate to 200 nL/min after loading, and separate peptides with a segmented linear gradient of HPLC solvent B from 5-33-50-90% for 87 minutes.

6. Operate the mass spectrometer in the positive ion mode, with the following acquisition cycle: a full scan is measured using the Orbitrap cell at a resolution R 120,000 followed by MS/MS of the top 15 most intense peptide ions in the linear ion trap using collision-induced dissociation (CID. Acquire precursor ions in the mass range from m/z 300 to 2000 and use the target values 1E6 charges for the full scans (Orbitrap analyser), and 5E3 charges for the collision-induced dissociation (CID) in the Linear Ion Trap. Sequenced precursor masses need to be excluded from further selection for 60 s.

Figure 1.

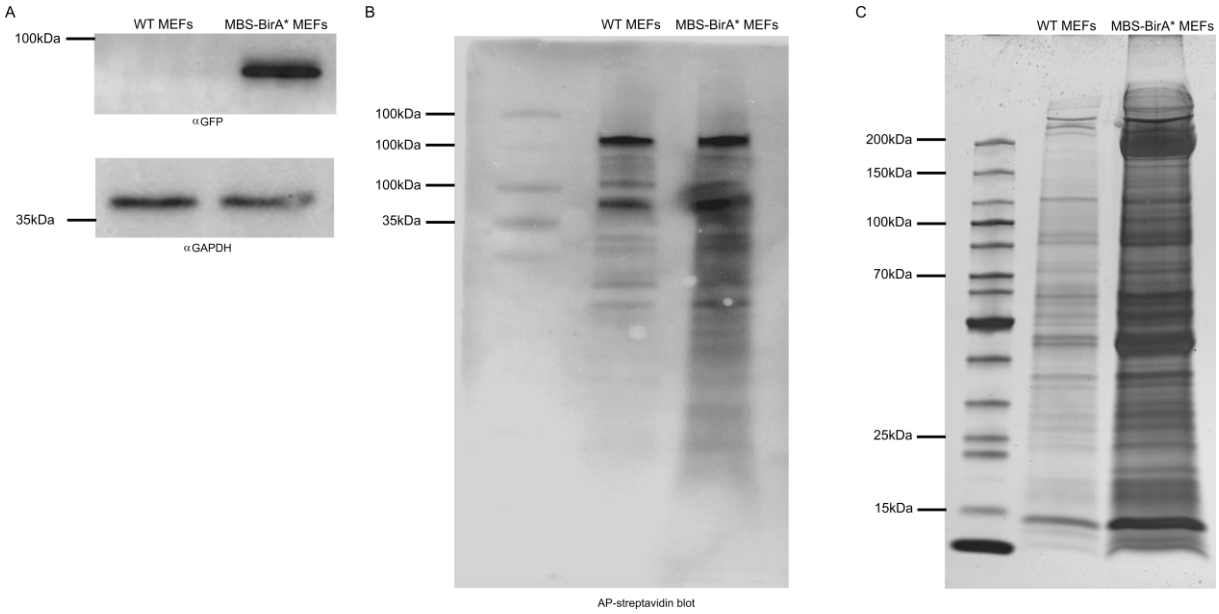


Figure legend : Functional validation of the integrated BirA* gene. A. Full expression check of BirA* construct via westernblot. Full expression of nls –HA - 2xmcp – egfp - BirA* transgene, in total 91.4 KDa protein, was checked by western blot after stable integration in the MBS MEF cells taking WT MEFs as a negative control for GFP expression. GAPDH was used as a loading control. B. Check for BirA* activity in the cell by streptavidin blot. 10µg of total proteins were loaded from WT and MBS-BirA* MEF cells, after incubating the cells with 50µM biotin for 24hrs. The biotinylated bands were ditected by streptavidin blot. after stable integration in the MBS MEF cells which taking WT MEFs as a negative control for GFP expression. C. Detection of biotinylated proteins after purifiacation via magnetic streptavidin beads. 1/10th of the magnetic beads used for the mass spec analysis were boiled in 100µl of 1X SDS sample buffer and 20µl from WT-MEFs and MBS-BirA* fractions were loaded on the gel followed by silver staining.

Recipes

1. Complete media (500ml): 445ml DMEM media, 50ml FBS, 5ml pen-strep.
2. Serum free media (100ml): 99ml DMEM media, 1ml pen-strep.

3. Biotin Media (100ml): 89ml DMEM media, 10ml FBS, 1ml pen-strep, 50uM biotin.
4. 50mM Biotin solution: Dissolve 2.44 mg biotin in10ml of serum-free DMEM. Pipetting is required to dissolve biotin completely. Sterilize by passing through a 0.22-µmsyringe-driven filter unit. Active up to 8 weeks at 4°C.
5. BCIP : 50 mg/mL in H2O.
6. NBT : 10 mg/mL in H20.
7. AP subatrate buffer: 100mM NaCl, 5mM MgCl2, 100mM Tris-Cl-pH 9.5.
8. 1X TBS : 50 mM Tris-Cl, pH 7.5, 150 mM NaCl final pH 7.5.
9. 1X TBST : 1X TBS, 0.1% Tween 20.
10. Lysis Buffer: 50 mM Tris (pH 7.5), 150 mM NaCl, 2.5 mM MgCl2, 1 mM DTT, 1% Triton X-100, and 1X proteinase inhibitor (Roche).
11. Wash Buffer 1: 2% SDS in dH2O.
12. Wash Buffer 2: 0.1% deoxycholate, 1% Tween 20, 500 mM NaCl, 1 mM EDTA, and 50 mM Hepes, pH 7.5.
13. Wash Buffer 3: 250 mM LiCl, 0.5% tween 20, 0.5% deoxycholate, 1 mM EDTA, and 10 mM Tris, pH 8.1.
14. Wash Buffer 4: 50 mM Tris, pH 7.4, and 50 mM NaCl.
15. 20mM ABC buffer: 79.06 mg ammonium bicarbonate was dissolved in 50 ml of RNAse, DNAse free water.
16. 2X SDS Sample buffer : 100 mM Tris-Cl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanol. Make 100ul aliquots and store at -20°C.
17. Denaturation buffer: 6M urea, 2M thiourea in 10 mM Tris buffer, pH 8.0.
18. Reduction buffer: 1M dithiothreitol in 50 mM ammonium bicarbonate.
19. Alkylation buffer: 550 mM iodoacetamide in 50 mM ammonium bicarbonate.
20. Solvent A*: 2% acetonitrile, 1% formic acid.
21. Solvent A: 0.5% acetic acid.
22. HPLC solvent A: 0.1% formic acid in Milli-Q water.
23. HPLC solvent B: 0.1% formic acid, 80% ACN in Milli-Q water.

Acknowledgments

We thank Jeff Chao (FMI, Basel) and Julién Bethune (BZH, Heidelberg), for plasmids and cell lines. Silke Wahle (PCT Tübingen) is acknowledged for technical support. The project was funded as project of the DFG Research Unit FOR2333 by a grant of the Deutsche Forschungsgemeinschaft (DFG JA696/11-1).

Competing interests

The authors declare no competing interests.

References

1. Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B. and Jansen, R-P. (2019). β-Actin mRNA interactome mapping by proximity biotinylation. *Proc Natl Acad Sci U S A* 116 (26): 12863-12872.
2. Fei. J., and Sharma, C.M. (2018). RNA localization in bacteria. *Microbiology Spectrum* 6(5): 10.1128.
3. Bovaird, S., Patel, D., Padila, J-C.A. and Lécuyer, E. (2018). Biological functions, regulatory mechanisms, and diseaserelevance of RNA localization pathways. *Febs letters* 592: 948–972.
4. Tolino, M., Köhrmann, M. and Kiebler, M.A. (2012). RNA-binding proteins involved in RNA localization and their implications in neuronal diseases. *Eur J Neurosci.* 35(12): 1818-36.
5. Jansen, R-P. (2001). mRNA localization: message on the move. *Nat Rev Mol Cell Biol* 2(4): 247-256.
6. Marchand, V., Gaspar, I. and Ephrussi, A. (2012). An intracellular transmission control protocol: assembly and transport of ribonucleoprotein complexes. *Curr Opin Cell Biol.* 24(2): 202-10.
7. Buxbaum, A.R., Haimovich, G. and Singer, R.H. (2015). In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol.* 16(2): 95-109.
8. Holt, C.E., Martin, K.C. and Schuman, E.M. (2019). Local translation in neurons: visualization and function. *Nat Struct Mol Biol.* 26(7):557-566.
9. Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L. and Singer RH. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. *Mol Cell Biol.* 17(4): 2158-65.
10. Eom, T. Antar, L.N., Singer, R.H. and Bassell, G.J. (2003). Localization of a beta-actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses. *J Neurosci*. 23(32): 10433-44.
11. Oleynikov, Y. and Singer, R.H. (2003). Real-time visualization of ZBP1 association with beta-actin mRNA during transcription and localization. *Curr Biol.* 13(3):199-207.
12. Hüttelmaier, S. Zenklusen, D. Lederer, M., Dictenberg, J., Lorenz, M., Meng, Z., Bassell, G.J., Condeelis, J. and Singer, R.H. (2005). Spatial regulation of β-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* 438(7067): 512–515.
13. Ramanathan, M., Porter, D.F., and Khavari, P.A. (2019). Methods to study RNA-protein interactions. *Nat Methods* 16(3): 225-234.
14. Park, H.Y., Lim, H., Yoon, Y. J., Follenzi, A., Nwokafor, C., Lopez-Jones, M., Meng, X., Singer, R.H. (2014). Visualization of Dynamics of Single Endogenous mRNA Labeled in Live Mouse. *Science* 343(6169): 422-424.
15. Aboulaich, N. (2011). Silver Staining. Bio-101: e26. DOI: 10.21769/BioProtoc.26.