Systematic imaging of single transcripts and polysomes reveals a widespread transport mechanism dependent on nascent translation

Adham Safieddine

To cite this version:

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR
DE L’UNIVERSITÉ DE MONTPELLIER

En Biologie - Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé (CBS2)

Unité de recherche Institut de Génétique Moléculaire de Montpellier (IGMM)

L'imagerie systématique de transcrits et de polysomes uniques révèle un mécanisme de transport dépendant de la protéine naissante

Présentée par Adham SAFIEDDINE
Le 12 novembre 2019

Sous la direction de Edouard BERTRAND et Marion PETER

Devant le jury composé de

Mme. Florence BESSE, DR, iBV CNRS Nice
M. Pierre-Emmanuel GLEIZES, DR, CBI CNRS Toulouse
Mme. Dominique WEIL, DR, IBPS CNRS Paris
M. Andreas MERDES, DR, CBI CNRS Toulouse
M. Edouard BERTRAND, DR, IGMM CNRS Montpellier
Mme. Marion PETER, CR, IRCM INSERM Montpellier
M. Kazem ZIBARA, PU, UL Beyrouth

Rapportrice
Rapporteur
Examinatrice
Président
Directeur de thèse
Co-directrice de thèse
Invité
Abstract
Local translation allows a spatial control of gene expression. Here, I participated in two mRNA localization screens imaging more than 1000 transcripts in total: (i) the first was a dual mRNA/protein screen that used a BAComics approach to co-detect mRNAs and the protein they encode; (ii) the second was done using a new high-throughput smFISH approach to screen all genes that encode centrosomal proteins and mitotic regulators. The first screen revealed cases of local translation at various subcellular compartments including cytoplasmic protrusions, centrosomes, Golgi, endosomes and the nuclear pore, which was never described before. Remarkably, translation of the nascent peptide was required for the transport of many localized transcripts. In addition, I showed that several mRNAs (such as ASPM and DYNC1H1) are translated in dedicated structures called translation factories.

The second screen revealed 8 transcripts that are localized and translated at the centrosome. I showed that the localization of these 8 transcripts is regulated by the cell cycle, and that it also requires translation of the nascent polypeptide. Using the endogenous ASPM gene as a model, I imaged single mRNAs and polysomes with the MS2 and SunTag systems, respectively. This revealed a directed transport of ASPM polysomes towards centrosomes at the onset of mitosis, when this mRNA starts localizing. These data provide definitive evidence for a co-translational targeting mechanism dependent on motors as well as the nascent protein. This argues against the current dogma that mRNA transport is an RNA-based process acting on translationally repressed molecules. Instead, it suggests that SRP-like mechanisms are more widespread than previously thought.

Résumé
La traduction locale permet un contrôle spatial de l’expression des gènes. Dans ce travail, j’ai participé à deux crible de localisation d’ARNm concernant plus de 1000 transcrits. Le premier était un crible double ARNm/protéine qui utilisait une approche de BAComics pour co-détecter les ARNm et la protéine pour laquelle ils codent. Le second a été réalisé à l’aide d’une nouvelle approche smFISH à haut-débit et a analysé tous les ARNm codant pour des protéines centrosomales et des régulateurs mitotiques. Le premier crible a révélé des cas de traduction locale dans divers compartiments subcellulaires, et notamment au niveau des protrusions cytoplasmiques, des centrosomes, de l’appareil de Golgi, des endosomes et des pores nucléaires, ce qui n’avait jamais été décrit auparavant. De manière remarquable, la traduction du peptide naissant était nécessaire pour le transport de nombreux transcrits localisés. De plus, j’ai montré que plusieurs ARNm (tels que ASPM et DYNC1H1) sont traduits dans des structures dédiées appelées usines de traduction.

Le deuxième crible a révélé 8 transcrits localisés et traduits au niveau des centrosomes. J’ai montré que la localisation de ces 8 transcrits est régulée par le cycle cellulaire et qu’elle nécessite également la traduction du polypeptide naissant. En utilisant le gène ASPM comme modèle, j’ai visualisé des ARNm et des polysomes uniques avec les systèmes MS2 et SunTag, respectivement. Cela a révélé un transport dirigé des polysomes ASPM vers les centrosomes au début de la mitose, lorsque cet ARNm commence à être localisé. Ces données fournissent des preuves fortes d’un mécanisme de ciblage co-traductionnel dépendant de moteurs moléculaires ainsi que de la protéine naissante. Cela va à l’encontre du dogme actuel selon lequel le transport d’ARNm est un processus basé sur l’ARN et agissant sur des molécules réprimées pour la traduction. En revanche, cela suggère que des mécanismes tels que celui utilisé par le SRP sont plus répandus qu’on ne le pensait auparavant.
Acknowledgements

First, I would like to thank Florence BESSE, Pierre-Emmanuel GLEIZES, Dominque WEIL, and Andreas MERDES for accepting to invest their time and expertise in evaluating my thesis. I hope this manuscript meets their expectations and interest.

I am grateful to Kazem ZIBARA for kindling my passion in molecular genetics, as well as the opportunity to carry out my masters 2 internship abroad. His guidance over many years has been invaluable.

My gratitude also goes to my thesis supervisor, Edouard BERTRAND for entrusting me with many projects in his lab. Thank you for the brilliant science, edifying discussions, and constant challenge. You trained me to take multitasking to a whole new level.

I am in depth to my thesis co-supervisor Marion PETER. Thank you for all the guidance and help that extended well beyond science. Your rigorous readings of my presentations, proposals, and manuscripts have been deeply appreciated. Thank you for all my lab members, especially Xavier PICHON, for the lovely atmosphere particularly in the face of creeping deadlines and rejected manuscripts.

I am candidly fortunate to have made enduring friendships throughout my PhD. Thank you Yousra, 11dem, and Thiziri for the epic adventures and relentless laughter. Thank you Bashar for the stress-relieving tête-à-tête chats. Thank you David (Birdd) for the dawn breaking videogame sessions and cookies. Thank you Fatima (7 years and counting) and Mohammad (صيويل) for the countless gatherings. Thank you Hasan (مجري) for those wonderful dinners. My deepest gratitude goes to my IGMM family: Jacinthe, Amani, Baraah, May, and especially Fabienne. Thank you for the unwavering support and endless laughter you bring; both in and out of work.

Last but certainly not least; I could have never gotten far without my family back home. Thank you Milad, Akram, and especially Ashraf and Mona for the persistent encouragement and unconditional support. Thank you mother and father for everything you allowed me to attain. To you, I dedicate this work.

“Science is not meant to cure us of mystery, but to reinvent and reinvigorate it.”

Robert Sapolsky
Table of contents

Introduction ......................................................................................................................... 15

1. Sending out a message: mRNA localization and local translation ................................................. 16
   1.1 The mRNA lifecycle: a bird’s eye view ............................................................................................ 16
   1.2 Functions of RNA localization related to local translation .......................................................... 18
      1.2.1 Functional advantages of local translation .............................................................................. 18
      1.2.2 Examples of biological processes controlled by local translation ......................................... 19
         1.2.2.1 Local translation and long term potentiation in neurons ..................................................... 20
         1.2.2.2 Local translation and mitochondrial biology .......................................................................... 21
   1.3 Functions of RNA localization not related to local translation .................................................... 23
      1.3.1 P-bodies ...................................................................................................................................... 24
         1.3.1.1 P-body composition ............................................................................................................... 24
         1.3.1.2 P-body functions .................................................................................................................... 25
      1.3.2 Localization of long non-coding RNAs ...................................................................................... 27
         1.3.2.1 General characteristics .......................................................................................................... 27
         1.3.2.2 LncRNAs functions depend on their localization ................................................................. 27
   1.4 Molecular elements controlling mRNA localization ....................................................................... 29
      1.4.1 Cis-acting elements .................................................................................................................. 29
         1.4.1.1 Zipcode characteristics ........................................................................................................... 29
         1.4.1.2 Nuclear events that control cis-acting elements ..................................................................... 31
      1.4.2 Trans-acting elements .............................................................................................................. 32
         1.4.2.1 General characteristics ........................................................................................................... 32
         1.4.2.2 Nuclear events that control trans-acting elements .................................................................. 34
   1.5 Mechanisms of mRNA localization ............................................................................................ 35
      1.5.1 Diffusion and local entrapment ................................................................................................. 35
1.5.2 Local protection from degradation

1.5.3 Motor driven transport

1.6 Translational regulation of mRNAs in transit

1.6.1 The signal recognition particle

1.6.2 Translational silencing of mRNAs during transport

1.6.3 Activating mRNA translation once they reach destination

1.6.4 Redundancy in translational control mechanisms

2. Visualizing mRNA

2.1 Low throughput smFISH techniques

2.1.1 Standard smFISH techniques

2.1.2 Signal amplifying smFISH techniques

2.2 High-throughput smFISH techniques and screens

2.2.1 Systematic bDNA FISH

2.2.1.1 Screen design

2.2.1.2 Analysis of mRNA localization features

2.2.2 Multiplexed error robust FISH (MERFISH)

2.2.2.1 Screen design

2.2.2.2 RNA detection and localization analysis with MERFISH

2.2.3 Sequential FISH (seqFISH and seqFISH+)

2.2.3.1 seqFISH principles and applications

2.2.3.2 seqFISH+ hybridization principle

2.2.3.3 Subcellular localization patterns observed using seqFISH+

2.3 Imaging RNA in living cells

2.3.1 Techniques to visualize RNA in living cells

2.3.1.1 Tag multimerization approaches

2.3.1.2 Other approaches
2.3.2 Applications in imaging RNA metabolism

3. **Translation goes live**

3.1 Biochemical methods for quantifying translation:

3.2 Visualizing translation at the single molecule level in real time

3.2.1 Nascent peptide imaging

3.2.2 Accessory tags for improving nascent peptide imaging

3.2.3 Nascent peptide imaging of endogenous transcripts

3.3 Understanding translation at the single molecule level

3.3.1 Ribosome numbers and kinetics

3.3.1.1 Calculating ribosome density on a single mRNA

3.3.1.2 Calculating translation elongation rates

3.3.1.3 Calculating translation initiation rates

3.3.2 Heterogeneity in translation

4. mRNA localization and translation regulation during the cell cycle

4.1 The eukaryotic cell cycle

4.1.1 Main phases and regulators of the cell cycle

4.1.2 Translational control by cytoplasmic poly-A tail length modulation

4.2 The centrosome

4.2.1 Centrosome structure and composition

4.2.1.1 Centriole structure

4.2.1.2 Pericentriolar material

4.2.2 Centrosome functions

4.2.3 Centrosome biogenesis
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.4 mRNA localization and local translation at centrosomes</td>
<td>85</td>
</tr>
<tr>
<td>4.3 Abnormal spindle microcephaly-associated protein</td>
<td>88</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>91</td>
</tr>
<tr>
<td>1. Visualization of single endogenous polysomes reveals the dynamics</td>
<td>92</td>
</tr>
<tr>
<td>of translation in live human cells</td>
<td></td>
</tr>
<tr>
<td>2. A localization screen reveals translation factories and widespread</td>
<td>114</td>
</tr>
<tr>
<td>co-translational protein targeting</td>
<td></td>
</tr>
<tr>
<td>3. Systematic identification of human centrosomal mRNAs reveals a</td>
<td>188</td>
</tr>
<tr>
<td>localization mechanism dependent on motors and nascent translation</td>
<td></td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td>258</td>
</tr>
<tr>
<td>1. mRNA localization screens: a technical perspective</td>
<td>259</td>
</tr>
<tr>
<td>2. mRNA localization in cell lines: generalities</td>
<td>260</td>
</tr>
<tr>
<td>3. A variety of local translation flavors</td>
<td>262</td>
</tr>
<tr>
<td>4. Programmed translation on centrosomes</td>
<td>263</td>
</tr>
<tr>
<td>5. A new method of protein targeting</td>
<td>264</td>
</tr>
<tr>
<td>6. Visualizing endogenous polysomes: a technical perspective</td>
<td>265</td>
</tr>
<tr>
<td>7. Co-translational targeting captured live</td>
<td>267</td>
</tr>
<tr>
<td>8. ASPM: a transcript with many patterns</td>
<td>268</td>
</tr>
<tr>
<td>9. mRNA localization: the past, present, and future</td>
<td>269</td>
</tr>
<tr>
<td><strong>Annexes</strong></td>
<td>271</td>
</tr>
<tr>
<td>A computational framework to study subcellular mRNA localization</td>
<td>272</td>
</tr>
<tr>
<td>Résumé de la thèse</td>
<td>282</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>290</td>
</tr>
</tbody>
</table>
List of abbreviations

Abnormal Spindle Microtubule Assembly (ASPM)
Actin-related protein 2/3 (Arp2/3)
Amino acids (aa)
Anaphase-promoting complex/cyclosome (APC/C)
Alternative last exon (ALE)
AU-rich elements (AREs)
Azidohomoalanine (AHA)
Bio-orthogonal non-canonical amino acid labeling (BONCAT)
Brain derived neurotrophic factor (BDNF)
Branched DNA (bDNA)
Calmodulin kinase II a (CamKIIa)
Calponin homology (CH)
CDK5 regulatory subunit-associated protein 2 (CDK5RAP2)
Cleavage and polyadenylation specificity factor (CPSF)
Containing blue fluorescent protein (BFP)
CPE binding proteins (CPEBs)
Cyclin-dependent kinases (CDKs)
Cytoplasmic end of an endoplasmic reticulum signal-anchor membrane protein (cytERM)
Cytoplasmic polyadenylation element (CPE)
Dead Cas9 (dCas9)
Degradation machinery (DM)
eIF4E-binding proteins (eIF4E-BPs)
Endoplasmic reticulum (ER)
eukaryotic initiation factor (eIF)
Exon junction complex (EJC)
Fluorescence-activated cell sorting (FACS)
Fluorescent in situ hybridization (FISH)
Gamma tubulin ring complex (γTuRC)
Gap 1 phase (G1 phase)
Gap 2 phase (G2 phase)
Gene ontology (GO)
Glutamate receptors (GluR1 and GluR2)
Green fluorescent protein (GFP)
Guide RNAs (gRNAs)
Heat-shock elements (HSEs)
Homology directed repair (HDR)
Hsp83 degradation element (HDE)
Hsp83 instability element (HIE)
Hsp83 protection element (HPE)
Hybridization chain reaction (HCR)
K homology domain (KH)
Long non-coding RNAs (IncRNA)
Mass spectrometry (MS)
Message transport organizing center (METRO)
messenger ribonucleoprotein (mRNP)
messenger RNA (mRNA)
micro RNAs (miRNAs)
Microtubule associated proteins (MAPs)
Microtubule-organizing center (MTOC)
Microtubules (MTs)
Mitochondrial cloud (MC)
Molecular beacons (MBs)
MS2 coat protein (MCP)
Multiplexed error robust FISH (MERFISH)
Myelin basic protein (MBP)
Nano-bodies (Nbs)
Natural antisense transcripts (NATs)
Nonsense mediated decay pathway (NMD)
Nuclear bodies (NBs)
nucleotides (nts)
open reading frame (ORF)
Pericentrin (PCNT)
Pericentrin-AKAP450 centrosomal targeting domain (PACT)
Pericentriolar material (PCM)
Poly A binding proteins (PABPs)
Polyadenylation hexanucleotide (hex)
Post synaptic dendrites (PSDs)
Post translational modifications (PTMS)

PP7 coat protein (PCP)

processing bodies (p-bodies)

Pseudoknots (PKs)

Pulsed stable isotope labeling by amino acids in cell culture (pSILAC)

Puromycin-associated nascent chain proteomics (PUNch-P)

Quantitative non-canonical amino acid labeling (QuaNCAT)

Reactive oxygen species (ROS)

Restriction point (R point)

Ribonucleoproteins (RNPs)

RNA binding domains (RBDs)

RNA binding proteins (RBPs)

RNA polymerase II (RNA Pol II)

RNA recognition motif (RRM)

RNA-sequencing (RNA-seq)

Rolling circle amplification (RCA)

Sequential FISH (seqFISH)

Sequential Tethered and Intertwined Complexes (STICs)

Short hairpin RNA (shRNA)

Signal recognition particle (SRP)

Signal to noise ratio (SNR)

single chain variable fragment (scFv)

single molecule FISH (smFISH)
single molecule inexpensive FISH (smiFISH)

single stranded RNA (ssRNA)

small interfering RNAs (siRNAs)

small nuclear RNAs (snRNAs)

small nucleolar RNAs (snoRNAs)

Spliced Oskar localization element (SOLE)

superfolder GFP (sfGFP)

Synthesis phase (S phase)

Terminal Oligopyrimidine tract (TOP)

transfer RNA (tRNA)

Translating ribosome affinity purification (TRAP)

Translating RNA imaging by coat-protein knock off (TRICK)

Translocated in liposarcoma (TLS)

Untranslated regions (UTRs)

Vg1 localization element (Vg1LE)
List of figures

Figure 1: A typical mRNA lifecycle ................................................................................................................................. 17
Figure 2: Regulating local translation at synaptic dendrites .......................................................................................... 21
Figure 3: Regulating local translation at mitochondria ................................................................................................. 23
Figure 4: P-body functions ............................................................................................................................................... 26
Figure 5: lncRNA localization and function .................................................................................................................. 28
Figure 6: Examples of different types of zipcodes ......................................................................................................... 30
Figure 7: Alternative splicing can affect mRNA localization ..................................................................................... 31
Figure 8: Trans-acting regulators of localized mRNAs ................................................................................................. 33
Figure 9: Co-transcriptional recruitment of trans-acting regulators ........................................................................... 34
Figure 10: General mechanisms of mRNA localization ............................................................................................... 36
Figure 11: Diffusion-based transport of mRNA ............................................................................................................. 37
Figure 12: Selective protection of mRNA in a specific part of the cell from degradation ............................................ 38
Figure 13: Mechanisms of achieving biased motor transfer ......................................................................................... 39
Figure 14: Mechanisms of translational control of localizing mRNAs ........................................................................ 41
Figure 15: Tight translational control of Ash1 mRNA ................................................................................................. 43
Figure 16: smFISH against Abnormal Spindle Microtubule Assembly mRNA in Hela Kyoto cells 45
Figure 17: Variations in smFISH techniques ................................................................................................................ 46
Figure 18: A bDNA FISH screen .................................................................................................................................... 48
Figure 19: An mRNA localization screen using MERFISH .......................................................................................... 51
Figure 20: seqFISH principle .......................................................................................................................................... 53
Figure 21: seqFISH+ barcoding principle ........................................................................................................................ 54
Figure 22: mRNA localization patterns revealed by seqFISH+ in NIH cells ............................................................... 55
Figure 23: Single-molecule visualization techniques in living cells .............................................................................. 57
Figure 24: Analyzing RNA processing using live imaging approaches ........................................................................ 60
Figure 25: Biochemical approaches for quantifying translation ................................................................................... 63
Figure 26: Visualizing translation in living cells ............................................................................................................... 65
Figure 27: Accessory tags used in nascent peptide imaging .......................................................................................... 67
Figure 28: Imaging the translation of an endogenous gene ............................................................................................ 68
Figure 29: Calculating translation elongation time ....................................................................................................... 71
Figure 30: Spatial variations in translation efficiencies in neurons ................................................................................ 73
Figure 31: A multicolor reporter for visualizing alternative start site selection ........................................................ 75
Figure 32: The cell cycle and its main regulators in animal cells ................................................................. 78
Figure 33: Centriole and centrosome structure and organization ................................................................. 81
Figure 34: Major PCM components during interphase and mitosis in various organisms ...................... 82
Figure 35: Centrosome/centriole biogenesis during the cell cycle ............................................................... 84
Figure 36: Approach for visualizing translating PCNT mRNA at centrosomes ........................................ 87
Figure 37: MT minus-end regulation by an ASPM-katanin complex ......................................................... 89
Introduction
1. Sending out a message: mRNA localization and local translation

1.1 The mRNA lifecycle: a bird’s eye view

Cells are highly dynamic entities hosting millions of chemical reactions that carry out many specialized roles. At the core of proper cellular function lie intricate trafficking mechanisms, in which a cell often targets specific molecules to certain organelles and subcellular locations at precise times. Over the past decade in the literature, RNA trafficking and localization have evolved from a rare occurrence limited to certain transcripts in specialized cells to become a common layer of gene expression regulation in both time and space (Wilk et al., 2016).

The life of a messenger RNA (mRNA) molecule begins in the nucleus where protein-coding genes are transcribed by RNA polymerase II (RNA pol II) with the help of general and specific transcription factors, thereby producing immature nascent pre-mRNA molecules (Figure 1). A wide range of RNA binding proteins (RBPs) and ribonucleoproteins (RNPs) are deposited along the mRNA co-transcriptionally. Such RBPs are crucial for RNA maturation, which encompasses various processes such as 5’ end capping, splicing, cleavage, polyadenylation, and localization (detailed later). Once mature, transcripts leave the nucleus via nuclear pores towards the cytoplasm, where the messenger ribonucleoprotein (mRNP) particle undergoes remodeling to release and/or recruit factors important for regulating its stability, translation, and trafficking. After certain rounds of translation, mRNAs are degraded via various degradation machinery and enzymes such as decapping enzymes and 5’ to 3’ nucleases (Cole, 2001).

Throughout its lifetime, an RNA molecule will travel from the depths of the nucleus, into the vast outreaches of the cytosolic space, cellular organelles, and even beyond the cell itself as in the case of secreted RNA. Therefore, proper transport and localization of mRNA molecules is a process that is crucial for the cell’s life. In the first chapter, I will begin by examining various biological functions achieved by localizing mRNAs. I will then describe cis and trans acting molecular elements that modulate mRNA localization, as well as mechanisms by which mRNA is localized. Finally, I will focus on the translational regulation of mRNAs in transit.
Figure 1: A typical mRNA lifecycle. (1) In the nucleus, RNA Pol II binds to promoter elements with the help of general and specific transcription factors initiating transcription of immature mRNAs. (2) RBPs bind the nascent mRNA co-transcriptionally. (3) RBPs catalyze the addition of a 7-Methylguanosine cap, the removal of introns which leads to the deposition of exon junction complexes (EJCs), as well as the addition of a poly A tail. (4) The now mature mRNP is exported to the cytoplasm via the nuclear pore. (5) Once in the cytoplasm, the mRNP gets remodeled by acquiring and/or releasing specific RBPs which may govern its localization and inhibit its translation. (6) After reaching its target site, the
1.2 Functions of RNA localization related to local translation

In simple terms, mRNA localization can be described as a post-transcriptional process in which mRNA molecules are transported to and enriched in specific subcellular compartments. Generally speaking, mRNA localization serves to restrict protein synthesis to certain areas in the cell, sometimes in response to specific signals. This process is referred to as local translation. (Hilliker, 2014). RNA can be localized within cells for other purposes, and this will be described in following sections. Here, I will highlight the significance of this process from both functional and biological perspectives.

1.2.1 Functional advantages of local translation

Local protein synthesis provides a variety of functional benefits for the cell:

- Given that individual mRNAs serve as templates for multiple translation rounds, it is more efficient to transport and translate a single mRNA locally. This is particularly evident in large cells such as oocytes and neurons. (Martin and Ephrussi, 2009).
- Protein translated from localized mRNAs may harbor distinct post-translational modifications (PTMs) when compared to pre-existing protein molecules. This may allow them to serve different functions. (Medioni et al., 2012).
- mRNA localization ensures compartmentalization of proteins that have harmful effects in certain subcellular areas. An elegant example of this is the mRNA encoding myelin basic protein (MBP). MPP is a sticky protein that can form harmful membrane aggregates. Locally synthesizing MBP around myelin sheaths in oligodendrocytes prevents this from happening. (Smith, 2004).
- Localizing functionally related transcripts encoding subunits of a protein complex may favor co-translational assembly. For example, an actin polymerization nucleator called actin-related protein 2/3 (Arp2/3) is a complex of 7 subunits localizing to the leading edge of migrating cells. Fluorescent in situ hybridization experiments (FISH), revealed that all 7 mRNAs encoding the Arp2/3 subunits co-localize at the leading edge. (Mingle et al., 2005).
1.2.2 Examples of biological processes controlled by local translation

Local translation is involved in diverse biological processes such as cell fate determination, directed migration, embryonic patterning, cell polarization, and synaptic plasticity to name a few. Such processes have been studied across a wide range of organisms and models such as bacteria, budding yeast, Drosophila, Xenopus, and human cells (Buxbaum et al., 2015). Table 1 provides some examples of localized mRNAs and their functions. Detailed examples are provided in the following sections with an emphasis on functional regulation depending on cellular needs.

*Table 1: Summary of some localized mRNAs and the functions they perform across various organisms*

<table>
<thead>
<tr>
<th>mRNA(s)</th>
<th>Organism</th>
<th>Subcellular localization</th>
<th>Function(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanos</td>
<td><em>Drosophila melanogaster</em></td>
<td>Oocyte posterior</td>
<td>Patterning and germ-cell development</td>
<td>(Forrest and Gavis, 2003)</td>
</tr>
<tr>
<td>Bicoid</td>
<td><em>Drosophila melanogaster</em></td>
<td>Oocyte anterior</td>
<td>Anterior cell fate specification</td>
<td>(Berleth et al., 1988)</td>
</tr>
<tr>
<td>Ash1</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>Budding tip</td>
<td>Mating type switching</td>
<td>(Paquin and Chartrand, 2008)</td>
</tr>
<tr>
<td>β-actin</td>
<td><em>Mus musculus</em></td>
<td>The leading edge of fibroblasts</td>
<td>Focal adhesion and migration</td>
<td>(Katz et al., 2012)</td>
</tr>
<tr>
<td>VegT and Vg1</td>
<td><em>Xenopus laevis</em></td>
<td>Oocyte vegetal cortex</td>
<td>Endoderm and mesoderm specification</td>
<td>(Birsoy et al., 2006)</td>
</tr>
<tr>
<td>Pou2</td>
<td><em>Danio rerio</em></td>
<td>Oocyte animal pole</td>
<td>Endoderm specification</td>
<td>(Lunde et al., 2004)</td>
</tr>
<tr>
<td>comE</td>
<td><em>Bacillus subtilis</em></td>
<td>Cell pole</td>
<td>Horizontal gene transfer</td>
<td>(dos Santos et al., 2012)</td>
</tr>
</tbody>
</table>
1.2.2.1 Local translation and long term potentiation in neurons

In neurons, many mRNAs including ones encoding β-actin, Arc, calmodulin kinase Ila (CamKIIa), and the glutamate receptors (GluR1 and GluR2) are transported through neuronal dendrites into dendritic spines (Kacharmina et al., 2000; Sakagami and Kondo, 1996). These mRNAs remain translationally repressed in the absence of synaptic activation. Upon synaptic transmission, local translation of these mRNAs at postsynaptic dendrites (PSDs) reinforces and stabilizes the active synaptic junction over time. This process is called long term potentiation (LTP) and is the cellular basis for learning and memory formation. Indeed, when the localization of endogenous CamKIIa mRNA in mice was disrupted, LTP establishment failed, and mice showed learning defects (Miller et al., 2002).

**Transport:** mRNPs can often be observed moving bi-directionally within dendrites near synapses. This lead to a model called “the sushi belt” in which mRNPs patrol synapses in a multi-directional fashion without irreversible anchoring at a specific location (Doyle and Kiebler, 2011). Experimental evidence for this model was recently obtained by imaging live mRNAs along dendrites (Bauer et al., 2019). Indeed, reporter mRNAs underwent both anterograde and retrograde transport away from and towards the cell body respectively. A bias in anterograde transport towards distal dendrites was observed upon fusing the 3'UTR of Rgs4 mRNA to the reporter, which was dependent on neuronal activity, and an RNA binding protein (RBP) called Stau2.

**Anchoring:** Once reaching an active synapse, mRNAs are anchored in a process referred to as docking. Depending on transcript, docking can require microtubules, actin filaments, and in some cases the molecular motor dynein which serves as an anchoring point on the minus end of microtubules (Doyle and Kiebler, 2011). Interestingly, docking can be regulated both temporally and spatially: (i) β-actin mRNA for instance, transiently docks at un-stimulated spines for about 7 mins. Activating dendritic spines with glutamate increases the docking duration to 2 hours (Yoon et al., 2016). (ii) Adding the 3'UTR Rgs4 mRNA to a reporter causes it to dock closer to activated
synapses than a reporter without the 3’UTR (Bauer et al., 2019). Once delivered to the synapse, mRNAs are released from RNPs and undergo local translation.

**Degradation:** In order to prevent PSD over-stabilization, decay mechanism exist that degrade the mRNAs after certain rounds of translation. One example is Arc mRNA, which undergoes nonsense-mediated decay pathway (NMD) after being translated at active PSDs. Introns in the 3’UTR of Arc mRNA are important for NMD which is mediated through the exon junction complex (EJC) factor eIF4AIII (Giorgi et al., 2007). eIF4AIII can modulate synaptic strength by modulating the local translational output of Arc mRNA.

*Figure 2: Regulating local translation at synaptic dendrites.* In the absence of synaptic transmission, mRNPs travel up and down the dendrite while translationally silent. Upon neuronal stimulation, local signals activate the translation of dormant mRNPs only in the firing dendrite. This spatial translational restriction is achieved via translation dependent anchoring and degradation. This maintains appropriate protein levels at appropriate spines which is crucial for LTP (Kejiou and Palazzo, 2017).

1.2.2.2 Local translation and mitochondrial biology
In *Saccharomyces cerevisiae*, a member of the pumillio family of RBPs, Puf3p, help to localizes some mRNAs to the outer mitochondrial membrane (García-Rodríguez et al., 2007). A simple model
would suggest that such mRNA targeting is meant to facilitate mitochondrial protein incorporation. Paradoxically, Puf3p prevents the translation of its target mRNAs and promotes their degradation via recruiting deadenylation and decapping factors (Houshmandi and Olivas, 2005). This raises the question as to why would Puf3p target mRNAs to the mitochondria, and then degrade them?

Local translation of Puf3p-bound mitochondrial mRNAs is regulated at two levels in yeast (Figure 3):

- **A global level in response to glucose deprivation**: when yeast cells are starved from glucose, Puf3p is regulated by a group of nutrient-sensitive kinases and phosphatases. Puf3 is directly phosphorylated and this phosphorylation stabilizes mRNAs and releases them from translational repression (Lee and Tu, 2015). Translation can then occur either at the membrane via mitochondria-associated ribosomes or in the vicinity of the organelle. This local translation stimulates mitochondrial biogenesis providing the cell with the ability to rapidly respond to drastic changes in its environment. Indeed, yeast strains lacking endogenous Puf3p showed reduced growth rates compared to wild type cells when placed in a low glucose medium, despite normal localization of pre-existing proteins encoded by Puf3p targeted mRNAs (Saint-Georges et al., 2008).

- **A local level in response to mitochondrial stress**: oxidative stress induced by low levels of reactive oxygen species (ROS) relieves Puf3p mRNA targets from their translational silencing, thereby promoting the synthesis of mitochondrial repair proteins. Interestingly, this only occurs in mitochondria that require so in the cell. However, prolonged exposure to high levels of ROS cause the opposite, in which Puf3p maintains translational inhibition. In this case, Puf3p further recruits Mdm12 (Böckler and Westermann, 2014). This interaction causes the destruction of the severely damaged mitochondrion by autophagosomes in a process called mitophagy but leaves healthy mitochondria intact.
Figure 3: Regulating local translation at mitochondria. Under normal growth conditions, the RBP PuF3p localizes almost all of its mRNAs targets to the mitochondria maintaining them in a translationally suppressed state. In low glucose conditions, phosphorylated PuF3p activates target mRNA translation rapidly prompting mitochondrial biogenesis globally in the cell. Another local layer of regulation takes place upon ROS induced mitochondrial stress. Depending on ROS levels and stress duration, PuF3p may act to de-repress the translation of its associated mRNAs to repair a slightly damaged mitochondrion, or sequester and silence its bound mRNAs promoting the mitophagy of heavily damaged mitochondria (Kejiou and Palazzo, 2017).

From these examples, we can begin to appreciate that local translation goes beyond protein targeting. It provides an elegant regulatory layer by which the cell can rapidly and specifically adapt to stimuli and environmental changes by compartmentalizing gene expression to maintain homeostasis.

1.3 Functions of RNA localization not related to local translation

The functions of RNA localization extend well beyond local translation. It can for instance, regulate mRNA metabolism as in the case of processing bodies (P-bodies) and stress granules, where mRNAs are stored in an untranslated from. RNA localization can even exert translation-independent functions. An example of this is the structural role of the VegT mRNA, which organizes cytokeratin networks in the vegetal cortex developing Xenopus oocytes independently of its
protein-coding capacity (Kloc et al., 2011). Similarly, the localization of long non-coding RNAs (IncRNA) to distinct sub-nuclear and cytoplasmic territories can be essential for their proper function in regulating and coordinating gene expression (Batista and Chang, 2013). Here, I discuss the localization of mRNAs in P-bodies and as well as the localization of some IncRNAs.

1.3.1 P-bodies

P-bodies are ~ 500nm cytoplasmic RNP non-membranous organelles found in eukaryotes. They are constitutively present in mammalian cells and induced by stress in yeast. In most mammalian cells, there are less than 10 P-bodies per cell and these bodies show restricted movements with the occasional motor-driven transport (Aizer et al., 2008).

1.3.1.1 P-body composition

Traditionally, immuno-staining and fluorescent-tagging approaches identified in P-bodies proteins related to a few specific aspects of RNA metabolism: 5’–3’ decay (DCP1/2, XRN1), decapping activators (EDC3, EDC4, PAT1B, LSM1-7), deadenylation factors (CCR4, PAN3), miRNA-mediated silencing factors (AGO1-4, GW182), nonsense-mediated decay factors (UPF1, SMG5, SMG7), and translational repression pathway factors (the helicase DDX6, the eIF4E-binding protein 4E-T, and the RBP CPEB1; van Dijk 2002; Andrei et al., 2005; Cougot et al.; 2004 Sen and Blau, 2005; Zheng et al., 2008). In addition, smFISH experiments revealed a few mRNAs localizing to P-bodies, such as : (i) mRNAs silenced by miRNAs (Pillai et al., Science 2005; 2 others refs in 2005), as well as miRNAs themselves (Pillai et al. 2005); (ii) mRNAs subjected to NMD (Durand, JCB 2007; other refs); (iii) mRNAs regulated by AU-rich elements (AREs) (franks TM and lykkeandersen 2007). Interestingly, the translation of the endogenous cationic amino acid transporter 1 (CAT-1) mRNA is reversibly regulated by miRNA in liver cells, and this mRNA localizes in and out P-bodies during translation inhibition and de-repression, suggesting that P-body localization can be a reversible process (Bhattacharyya et al., 2006).

A recent study (Hubstenberger et al., 2017) succeeded in purifying p-bodies by fluorescence-activated cell sorting (FACS). This breakthrough study allowed a genome-wide characterization of P-bodies and identified 110 novel proteins enriched in P-bodies via mass spectrometry, all of which are related to RNA metabolism. This confirmed the role of P-bodies as RNA regulating entities. Interestingly, the molecular motor MYO6 was found there as well suggesting a role in trafficking p-bodies.
In the same study, RNA-seq analysis showed that P-bodies contain a diverse set of mRNAs. Indeed, about one-third of the coding transcriptome was present in p-bodies. Interestingly, another third was found to be excluded from P-bodies demonstrating that mRNA localization to P-bodies is specific. P-body mRNAs were found to be intrinsically poorly translated, with low polysome formation rates and have 3’ untranslated regions (3’UTRs) of variable lengths. Finally, this study showed that some sequence elements are important for inclusion or exclusion of mRNAs from p-bodies: mRNAs containing an ARE element are enriched in P-bodies while those containing a Terminal Oligopyrimidine tract (TOP) motif are excluded.

1.3.1.2 P-body functions

- RNA metabolism: The exact function of P-bodies has been a subject of long debate. Since proteins involved in 5’ to 3’ decay were one of the first P-body proteins to be characterized (Van Dijk 2002; Sheth and Parker, 2003), it was originally assumed that these organelles are directly involved in mRNA decay (Sheth and Parker 2003; Cougot et al. 2004), especially for mRNA subjected to NMD or ARE-mediated regulation. It was later found that mRNAs silenced by miRNA can also localize to P-bodies (Pillai 20015 and others) and that mRNAs that localize to P-bodies can re-enter translation (Bhattacharyya et al., 2006; Aizer et al., 2014). This lead to a cumulative model that views P-bodies as sites of both mRNA decay and storage for future translation.

Recent biochemical and single molecule imaging approaches have challenged this model. First, RNA-seq of purified P-bodies did not show any enrichment of mRNA decay intermediates in P-bodies compared to the cytosol (Hubstenberger et al., 2017). Furthermore, comparing cells with and without P-bodies (via silencing the DDX6 helicase) shows no stabilization of mRNAs when P-bodies are prevented from forming (Hubstenberger et al., 2017). Finally, imaging XRN1 mediated 5’ to 3’ decay with a dedicated reporter in living cells (detailed in chapter 2) did not show any accumulation of decay events in p-bodies (Horvathova et al., 2017). Taken together, these studies challenge the function of P-bodies in RNA degradation, although previous studies did detect RNA degradation intermediates in P-bodies (Cougot et al. 2004).

Such systematic analyses suggest that P-bodies may fulfill an important storage function. The high concentration of RBPs with low complexity sequences and mRNAs form a dense network of multivalent RNA-protein interactions that could then separate P-body from non-P-body mRNAs via a phase separation mechanism (Kroschwitz et al., 2015), while still permitting some exchange
Different combinations of bound RBPs, most of which repress translation, could then ensure a complete repression of the P-body mRNAs. For this model to hold true, several issues need to be addressed. For example, the activity of decay and decapping enzymes, which are plentiful in p-bodies, must be inhibited by a specific local mechanism.

- **Cellular functions:**

  Gene ontology analysis of purified P-bodies showed an interesting dichotomy: p-body excluded mRNAs tend to encode housekeeping genes while P-body localized transcripts encode regulatory proteins (Hubstenberger et al., 2017; Standart and Weil, 2018). For instance: (i) mRNAs encoding histones are excluded from p-bodies while those encoding proteins that control histone methylation localize to p-bodies. (ii) mRNAs encoding proteins that associate with snRNPs and catalyze splicing are not localized in p-bodies, while mRNAs encoding splicing regulators are. (iii) Most mRNAs encoding core proteasome subunits are excluded from p-bodies, while mRNAs that code protein ubiquitination regulators are stored in p-bodies (Figure 4).

![Figure 4: P-body functions](image-url)

Figure 4: **P-body functions.** P-bodies are sites of reversible mRNA storage. mRNAs can move between the cytoplasm and p-bodies encountering different combinations of RBPs in each compartment. P-body associated RBPs such as decapping factors, deadenylation enzymes, and translation suppressors prevent the translation of their targets. Translation and decay can take place once the mRNA leaves
the p-body in the cytoplasm. mRNAs encoding proteins that carry out housekeeping functions tend to be excluded from p-bodies, while those that code for proteins with regulatory roles tend to be localized in p-bodies (Standart and Weil, 2018).

1.3.2 Localization of long non-coding RNAs

1.3.2.1 General characteristics
LncRNAs are transcribed from RNA pol II and have a length of more than 200nts. They are generally transcribed from intergenic regions of the genome, or from opposite strands of protein-coding genes, and are thus sometimes also referred to as natural antisense transcripts (NATs). They are presumably capped, spliced, and polyadenylated, like regular mRNAs. However, lncRNAs fulfill no protein-coding functions. Instead, they have a wide range of regulatory functions that affect gene expression at both the transcriptional and post-transcriptional levels (Chen, 2016).

1.3.2.2 LncRNAs functions depend on their localization
Like many biological molecules, the functions of lncRNAs are influenced by their subcellular localization. A small lncRNA localization screen revealed that many lncRNAs display diverse sub-nuclear patterns potentially important for their functions, while others are transported to the cytoplasm to carry out their putative regulatory role (Cabili et al., 2015). Here, we categorize lncRNAs into classes based on their localization compartment, and whether they act in cis or trans (Figure 5):

- LncRNAs that accumulate at their site of transcription (in cis) and act close to the same site (also in cis). For example, DNA damage induces the expression of a set of lncRNAs from the cyclinD1 (CCND1) gene. The lncRNAs then bind to a protein called translocated in liposarcoma (TLS) modulating its activity. This modulation allows TLS to inhibit histone acetyltransferase in cis, which further inhibits CCND1 transcription (Wang et al., 2008).
- LncRNAs that accumulate in cis but influence gene expression in trans. An example is the lncRNA FIRRE which is transcribed from the X chromosome but forms five inter-chromosomal contacts through 3D chromosome rearrangements (Hacisuleyman et al., 2014).
- LncRNAs that are released from their site of transcription and have global localization in the nucleoplasm. For instance, the BCAR4 lncRNA can bind SNIP1 and PNUTS transcription factors in the nucleoplasm and alter downstream gene expression in response to chemokine signaling in breast cancer (Xing et al., 2014).
- LncRNAs that accumulate in membrane-less sub-nuclear structures called nuclear bodies (NBs) alongside specific proteins. FISH experiments revealed that the lncRNA MALAT1 accumulates in nuclear speckles that contain splicing factors and has been recently linked to regulating mammary cancer biology (Tripathi et al., 2010).
- LncRNAs exported to and function in the cytoplasm. This group of lncRNAs carries out functions that necessitate their presence in the cytosol such as interfering with protein post-translational modifications (PTMs), and acting as decoys for miRNAs and proteins. The highly conserved and abundant lncRNA NORAD can sequester many PUMILIO members activating the translation of certain mRNAs in response to genomic instability (Lee et al., 2016).

It is noteworthy to mention that the exact mechanisms that control whether lncRNAs are retained in specific nuclear compartments or exported to the cytoplasm remain poorly understood.

**Figure 5: LncRNA localization and function.** (A) LncRNAs can both accumulate and act at or close to their transcription site i.e. in cis. (B) LncRNAs can accumulate in cis but still act on distant genes in trans, either on the same or different chromosomes. (C) LncRNAs can be distributed throughout the nucleoplasm in trans, and act in trans as well. (D) Some lncRNAs are enriched in NBs with specific
proteins and act in trans. (E) Certain lncRNAs are exported to the cytoplasm to carry out diverse functions that regulate gene expression. Black arrows indicate where the lncRNA functions, while pink ones indicate where it localizes (Chen, 2016).

1.4 Molecular elements controlling mRNA localization

A basic notion of mRNA localization is that mRNA molecules contain sequences that act in cis called zipcodes that are bound by trans acting factors (RBPs in most cases), which orchestrate various aspects of localization such as translational repression, transport, and anchoring (Jansen, 2001). In this section, we focus on characterizing some cis and trans-acting elements, as well as how nuclear events in the mRNA’s lifecycle can modulate these elements and influence downstream localization in the cytoplasm.

1.4.1 Cis-acting elements

1.4.1.1 Zipcode characteristics

mRNA zipcodes are usually found within the 3’UTR of the transcript, however, they can also be found in the 5’UTR or the coding sequence itself. They range from a few nucleotides (nts), and can go up to thousands of bases. There is no consensus sequence for a zipcode yet. For example, the β-actin mRNA in a 54 nt zipcode found in the 3’UTR of the transcript that forms a bipartite motif recognized by the RBP zipcode-binding protein 1 (ZBP1). ZBP1 is sufficient to localize β-actin mRNA to the leading edge of migrating fibroblasts in a motor-dependent manner (detailed in the next section) (Oleynikov and Singer, 2003). In some cases, the function of a zipcode can be carried out by its primary sequence. However, zipcodes often form secondary and tertiary structures enabling inter-molecular interactions crucial for their proper functioning. Certain mRNAs contain a few zipcodes that are all needed for proper localization probably through a multistep localization process. In other cases, zipcodes on the same transcript have additive effects or carry out redundant functions. Some examples of well-described zipcodes (Jansen, 2001) are presented below (Figure 6):

- Vg1 mRNA contains a 340 nt zipcode in its 3’UTR that is sufficient for its localization to the Xenopus oocyte vegetal cortex called Vg1 localization element (Vg1LE). Vg1LE contain short repetitive functional units (either UAUUUCUAC or UUCAC) necessary for its proper functioning (Gautreau et al., 1997).

- Bicoid mRNA has a 625 nt zipcode in its 3’UTR that forms secondary structures necessary for its transport to the anterior of the Drosophila oocyte in a multistep process. Bicoid
zipcode forms a series of stem-loops, each important for particular stages of oogenesis: (i) for transport from nurse cells and initial accumulation of Bicoid at the anterior margin of the oocyte, a 50 nt localization element called Bicoid localization element 1 (BLE1) is needed and sufficient. (ii) BLE1 and two other stem-loops in the 3’UTR (stems IV and V) are all needed to drive mid and late localization in which the transcript accumulates at the oocyte’s anterior end. (iii) Anchoring requires the presence of an additional stem-loop (stem III) to complete the localization process (Macdonald and Kerr, 1997).

- The Ash1 mRNA in budding yeast is an example of mRNAs that contain zipcodes in their coding regions. This transcript contains four localization elements, three of which (E1, E2a, and E2b) are in the open reading frame (ORF). The fourth element (E3) extends from the end of the coding region into the 3’UTR and forms secondary structures important for its function. All four elements can localize reporter RNAs to the bud tip individually but they do not restrict it to the daughter cell’s tip. Indeed, correct anchoring necessitates the presence and collaboration of all elements (Chartrand et al., 1999).

Figure 6: **Examples of different types of zipcodes.** (a) Vg1 mRNA contains one localization element termed Vg1LE in its 3’UTR that is composed of short repeated functional units. (b) Bicoid mRNA has a zipcode
in its 3’UTR that forms various stem loops each important for specific tasks. Early localization requires the BLE1 loop (light blue) while late localization is achieved with additional help from loops IV and V (dark blue). Anchoring further requires loop III (pink). (c)ASH1 mRNA contains three localization elements (E1, E2a, and E2b) in the coding region, and one extending from the end of the ORF into the 3’UTR. Elements can redundantly localize the transcript but all of them cooperatively anchor it (Jansen, 2001).

1.4.1.2 Nuclear events that control cis-acting elements

The various layers of gene expression in eukaryotes (see figure 1) are coupled and can often affect each other. Events occurring early on in the mRNA’s lifetime in the nucleus can have an impact on the transcript’s fate and function in the cytoplasm. Splicing, for instance, can produce isoforms of a transcript that localize differentially if a zipcode is found in an alternatively spliced exon, a retained intron, or alternative last exon (ALE; Figure 7; Chin and Lécuyer, 2017).

![Figure 7: Alternative splicing can affect mRNA localization](image)

Splicing events can also directly induce the formation of stem-loops that function as zipcodes in the cytoplasm. A well-characterized example is Oskar mRNA in *Drosophila* that localizes to the posterior pole of the oocyte cytoplasm in a way that requires splicing of the first intron (Hachet and Ephrussi, 2004). In fact, splicing mediates two molecular events necessary for Oskar localization: (i) deposition of the *Drosophila* core (EJC) components Mago nashi, Tsunagi, Barentsz, and Eif4AiIII (Mohr et al., 2001). (ii) Formation of a 28 nt stem-loop element called spliced
Oskar localization element (SOLE; Ghosh et al., 2012). Both the EJC and SOLE are required for proper Oskar localization.

### 1.4.2 Trans-acting elements

#### 1.4.2.1 General characteristics

RBPs are the major class of trans-acting localization regulators. RBPs are a deeply conserved family of proteins involved in virtually all aspects of mRNA metabolism including localization (Gerstberger et al., 2014). Mass spectrometry analyses on proteins cross-linked with polyadenylated RNA identified around unique 800 RBPs (Castello et al., 2012) that contain RNA binding domains (RBDs). Canonical RBPs typically contain an RNA-binding domain, which can often be: (i) an RNA recognition motif (RRM); (ii) a K homology domain (KH); (iii) a DEAD-box motif; (iv) a zinc finger domains that can mediate binding to specific single-stranded RNA (ssRNA) sequences (Cook et al., 2011); (v) a double-strand RNA-binding domain (dsRBD). Indeed, some RBPs recognize specific secondary or tertiary structural features in the RNA, in which they bind double-stranded RNA (dsRNA) motifs or specific primary sequences within those motifs (Figure 8-A; Kazan et al., 2010). Non-canonical RBDs also exist and have been systematically identified by UV cross-linking and quantitative proteomics (Castello et al., 2016).

Interestingly, RBPs often contain multiple RBDs. These can bind multiple sequences in the same mRNA, thereby increasing specificity. However, they can also bind multiple RNA targets via complex multivalent interactions (Mitchell and Parker, 2014). This, combined with the fact that mRNAs can bind more than one RBP at a time makes deciphering the contribution of individual RBPs to localizing a given transcript challenging. Here it is worth emphasizing that mRNAs in the cytoplasm often form complexes with proteins of various sizes and complexities called granules. mRNA granules can be defined as transport particles containing multiple transcripts and occur in many systems from yeast to neurons. They form by phase-separation mechanisms.
Non-coding RNA can also function as trans-regulators of localized mRNAs, modulating their localization, stability, and translation (Figure 8-B and C). An example is the miRNA 134 which is localized at postsynaptic dendrites in hippocampal neurons, preventing the translation of Limk1 mRNA which encodes a protein important for dendritic spine morphogenesis. Stimulation by brain-derived neurotrophic factor (BDNF) inhibits the functions of the RNA induced silencing complex (RISC) relieving the inhibition of Limk1 translation promoting synaptic maturation (Schratt et al., 2006). In addition, IncRNAs can hypothetically modulate mRNA localization by acting as a bridge between mRNAs and RBPs or other mRNAs. For instance, the IncRNA BC200 can affect local translation in neuronal projections (Iacoangeli and Tiedge, 2013). However, the mechanism by which IncRNAs can control RNA localization remain generally poorly defined.
1.4.2.2 Nuclear events that control trans-acting elements

The nuclear history of mRNAs can also affect their cytoplasmic localization by recruiting specific RBPs and coordinating the formation of an mRNP complex (Figure 9-A). In budding yeast, Ash1 mRNA is bound by the RBP She2p, which is responsible for transporting the transcript to the daughter cell. She2p is a nucleo-cytoplasmic shuttling protein and it binds co-transcriptionally to Ash1 mRNA. She2p binding is facilitated through direct interactions with the Spt4p-Spt6p elongation factors associated with RNA pol II, ensuring loading of the protein during transcription. Mutating either SPT4 or SPT5 genes delocalizes Ash1 mRNA from the daughter cell due to perturbed She2p binding (Shen et al., 2010).

Figure 9: Co-transcriptional recruitment of trans-acting regulators. (A) RNA localization factors, translational repressors, and adapters bind to the mRNA sequentially and co-transcriptionally forming an mRNP. Once exported to the cytoplasm, maturation of the mRNP allows translation in the right place and time. (B) In the absence of stress, RBPs bind to both normal and stress-response transcripts similarly, orchestrating export and localization. Under stress conditions, specific transcription factors bind to promoter HSEs of stress genes inducing their transcription. This leads to the export of stress response mRNAs to the cytoplasm with minimal RBPs allowing their rapid translation, and the nuclear sequestration of other transcripts. Both models presented here are generic (Chin and Lécuyer, 2017).
Another important concept is that promoter elements, through certain transcription factors, can affect the localization of the transcribed product via gene specific recruitment of trans-acting elements on the nascent transcript (Figure 9-B). A good example of this is the yeast heat-shock elements (HSEs) found in the promoters of heat shock genes (Zander et al., 2016). Under heat stress conditions, the transcription factor Hsf1 binds HSEs and selectively recruits the nuclear export factor Mex67 to stress-response genes. Mex67 is subsequently loaded on specific mRNAs stimulating their nuclear export independently of adaptor proteins. Fusing artificial HSEs to non-stress mRNAs that would normally be retained in the nucleus under stress caused them to be efficiently localized to the cytoplasm confirming this to be a promoter driven process.

### 1.5 Mechanisms of mRNA localization

Now that I describes how a wide variety of cis and trans regulators orchestrate the targeting of mRNA molecules, I will examine the mechanisms in which mRNA localization takes place, and provide a detailed example for each mechanism. Three generic mRNA localization pathways have been described to date (Kloc et al., 2002): (i) random diffusion combined with local entrapment; (ii) general degradation coupled to localized protection; (iii) motor-driven transport along the cytoskeleton, sometimes followed by their anchoring at the target site (Figure 10).

#### 1.5.1 Diffusion and local entrapment

One of the simplest ways to localize an mRNA is for the transcript to diffuse in the cytoplasm and then become trapped at a specific location via binding to anchors (Figure 11-A). An example is the *Xenopus* Xcat-2 mRNA, which encodes a zinc-finger RNA-binding protein. During the early stages of Xenopus oogenesis, Xcat-2 mRNA is restricted to a specific structure that consists of mitochondria and small vesicles in the cytoplasm called the mitochondrial cloud (MC) (Kloc and Etkin, 1995). Localization of Xcat-2 is a two-step process. In stage one, the MC is found next to the nucleus and accumulates RBPs such as Sm proteins, and the XVLG1 DEAD-box RNA helicase, the function of which is thought to be facilitating RNA export from the nucleus (Bilinski et al., 2004). In the second stage, the MC breaks loose from the nucleus and moves upwards the vegetal cortex carrying the mRNA along with it (Figure 11-B).
**Figure 10: General mechanisms of mRNA localization.** (a) Zipcode containing mRNAs are recognized by RBPs and are exported outside the nucleus as mRNPs. mRNAs can then (b) undergo general diffusion, (c) associate directly or via adaptors with motor proteins that mediate their directional transport on cytoskeletal networks, or (d) become localized through selective degradation protection. (e) Localized mRNAs may also carry out noncoding functions by exerting scaffolding activities. (f) Proteins translated from localized mRNAs can undergo retrograde transport to relay signals from the cell periphery back to the nucleus in response to extracellular cues (Neal A.L. Cody et al, 2013).

Xcat-2 accumulates in a region of the MC called the Message transport organizing center (METRO) as revealed by injecting fluorescent Xcat-2 RNA constructs (Chang et al., 2004). This region contains endoplasmic reticulum (ER) aggregates, mitochondria, and proteins including XNOA36 which form a matrix for mRNA entrapment. Xcat-2 mRNA localization is microtubule and motor independent since de-polymerization of the microtubules has no effect on the localization of Xcat-2 to the MC (Chang et al., 2004).
1.5.2 Local protection from degradation

Local stabilization and protection from degradation provide another mechanism by which mRNA can be targeted. In this case, mRNA is protected from degradation at specific subcellular compartments but is rapidly degraded in other parts (Figure 12-A).

An example is the heat shock protein hsp83 mRNA in *Drosophila* embryos. Localization of hsp83 mRNA is regulated via three *cis*-acting elements in the transcript (Figure 12-B): (i) The hsp83 degradation element (HDE) in the 3′UTR. (ii) The hsp83 instability element (HIE) at the 3′ end of the ORF that forms 6 stem-loops structures. (iii) The hsp83 protection element (HPE) in the 3′UTR downstream of the HDE (Sometown et al., 2008). HDE and HIE function together to ensure complete degradation of the transcript via recruiting the maternally encoded RBP Smaug to the HIE stem-loops. Smaug, in turn, recruits the CCR4/POP2/NOT deadenylase complex triggering the degradation of the hsp83 transcript. HPE selectively protects hsp83 mRNA from degradation at the posterior pole by an unclear mechanism (Figure 12-C). HPE is sufficient to protect an unstable transcript in the pole plasm (Sometown et al., 2005).

1.5.3 Motor driven transport

Active directed transport along the cytoskeleton represents the major mechanism by which mRNA is transported. Indeed, an mRNA molecule can be swiftly transported across the cell via binding of the mRNP particle to molecular motors. This process is orchestrated by the combined interactions of zipcodes, RBPs, adapters, translational repressors, and cytoskeletal components (Bullock, 2007). Thus, it is important that these elements work together to achieve a net bias in the transport direction.
Figure 12: Selective protection of mRNA in a specific part of the cell from degradation. (A) Schematic presentation (Medioni et al., 2012). (B) Degradation and protection elements found within the Hsp83 transcript. (C) Smaug degrades Hsp83 throughout the pole plasm except at the posterior pole (Shahbabian and Chartrand, 2012).

Generally speaking, members of the kinesin and dynein families transport mRNAs towards the plus and minus ends of microtubules (MTs) respectively, while myosin transports cargo towards the plus end of actin filaments. The speed and directionality of motor-driven transport depend on the type(s) and numbers of bound motors, as well as microtubule orientation. Motor-driven transport is fine-tuned by various mechanisms (Buxbaum et al., 2015) (Figure 13):

- Binding more motors can increase speed and processivity. For example, Ash1 mRNA has four zipcodes that can bind four molecules of the RBP She3p, which in total can bind four myosins.
- A dense network of similarly-oriented cytoskeletal elements can localize mRNAs more efficiently.
- In the case of mixed-orientation cytoskeleton, mRNAs can experience a “tug of war” in which the direction with the highest amount of net force wins.
- Cargos often bind multiple motors and a “tug of war” can also occur on MTs with similar orientation but with mRNAs that bind both kinesins and dyneins. In this case, there could be advantages in having multiple motors: the cargo moves faster in all directions and may reach its destination quicker if the roads are poorly oriented for example.
- Microtubule-associated proteins (MAPs) can affect motor binding to MTs and modify their dissociation kinetics, as well as act as obstacles causing a motor to change its direction.
• RBPs themselves can alter motor processivity, association with MTs, and speeds.

Figure 13: **Mechanisms of achieving biased motor transfer.** (a) Binding more motors increases transport efficiency. (b) MTs with similar orientation facilitate directional transport of mRNAs. A “tug of war” occurs when mRNAs are transported along MTs with mixed orientation (c) or bind different types of motors (d). MAPs can affect motor dissociation rates and reverse their direction by functioning as obstacles. (f) Cargo binding can affect motor association to MTs and adjust their speeds (Buxbaum et al., 2015).

1.6 Translational regulation of mRNAs in transit

1.6.1 The signal recognition particle

The relationship between the spatial distribution of mRNAs and the proteins they encode has long been the subject of study. One of the first insights into the coupling of mRNA localization and translation came from the signal recognition particle (SRP) system (Wolin and Walter, 1989). The SRP is an abundant, cytosolic, universally conserved RNP. Eukaryotic SRP consists of a 300-nucleotide 7SRNA and six proteins (SRPs 72, 68, 54, 19, 14, and 9). SRP recognizes a peptide sequence called signal peptide of trans membrane and secreted proteins co-translationally as they emerge from the ribosome. SRP then causes translation elongation arrest and targets the
translating mRNA to the ER where translation resumes and the nascent peptide receives distinct PTMs important for its future targeting and/or function.

In contrast to mRNAs encoding transmembrane and secreted proteins, the transport of most localized mRNAs is mechanistically coupled to translational suppression. This restricts protein synthesis in time and space and establishes translational patterns important for cellular functions (discussed in previous sections). Here, I will focus on mechanisms of translational repression of mRNAs during transport, as well as methods of translational de-repression at specific subcellular compartments.

1.6.2 Translational silencing of mRNAs during transport

As mentioned previously, mRNPs contain various trans-acting factors of which translational repressors are important constituents. In theory, various steps of the translation process can be inhibited, but initiation appears to be the most frequently regulated (Besse and Ephrussi, 2008). Three mechanisms regulate translation initiation of localized mRNAs (Figure 14-a):

- Targeting the eukaryotic initiation factor 4F (eIF4F) complex: eIF4F is a cap-binding heterodimeric protein complex essential for translation initiation. It is composed of the scaffolding protein eIF4G, the cap-binding protein eIF4E, and the RNA helicase eIF4A. Translational silencing occurs via recruiting eIF4E-binding proteins (eIF4E-BPs), which compete with eIF4G for binding eIF4E. In Drosophila the Bruno suppressor RBP bound to the 3'UTR of the Oskar mRNA recruits an eIF4E-BP called Cup. Preventing Cup-eIF4E interaction leads to premature translation of Oskar mRNAs, before they are localized (Nakamura et al., 2004).

- Blocking ribosomal subunit joining: for example, the RBP ZBP1 prevents the assembly of the 80S ribosomal machinery on β-actin mRNAs by blocking the incorporation of the S60 subunit (Hüttelmaier et al., 2005).

- Modulating polyA tail length: PolyA tail length is regulated by the opposing actions of poly A polymerases that elongate the tail and recruit translation promoting poly A binding proteins (PABPs), and deadenylation complexes that shorten the tail and repress translation. Some RBPs such a Smaug (discussed in section 1.5.2) shift the balance in favor of deadenylation complexes.
1.6.3 Activating mRNA translation once they reach destination

Once the mRNA reaches its target site, translation inhibition is relieved by two generic mechanisms that decrease the affinity of repressors to the mRNA (Figure 14-b):

- Local kinase-mediated phosphorylation of suppressors: ZBP1, for instance, is a substrate for the Src kinase. ZBP1 phosphorylation reduces its affinity (and translation inhibition) for β-actin mRNA. Interestingly, expressing a mutant ZBP1 that cannot be phosphorylated reduces the amounts of β-actin protein produced locally (Hüttelmaier et al., 2005).
Local competition with pre-existing proteins: In *Drosophila* oocytes; Smaug is a translational inhibitor of Nanos mRNA which restricts its translation at the posterior pole. Interestingly, Smaug also interacts with the Oskar protein that localizes to the posterior pole as well. Ectopic expression of Oskar protein causes ectopic translation of Nanos mRNA via inhibiting the binding of Smaug to Nanos mRNA (Zaessinger et al., 2006). This suggests that at the posterior pole, Oskar protein competes with Nanos mRNA for Smaug binding relieving the mRNA from its translational repression.

1.6.4 Redundancy in translational control mechanisms

In some cases, precise translational control is necessary to achieve the required biological function behind mRNA localization. In such cases, redundant mechanisms act at multiple layers to ensure complete silencing of the transcript during transport. A good example of this is Ash1 mRNA in budding yeast (Figure 15).

Translational inhibition of this mRNA is tightly controlled during its transport on microfilaments via:

- The RBP Khd1 specifically binds Ash1 mRNA and interacts with the C-terminal domain of eIF4G, blocking its function and preventing the formation of the pre-initiation complex required for translation. Khd1 also prevents the incorporation of the 40S ribosomal subunit into the RNP (Paquin et al., 2007).
- The RBP Puf6 also binds Ash1 mRNA and it prevents the formation of the 80S ribosome by blocking joining of the 60S subunit, possibly by competing with eIF5B binding (Deng et al., 2008).

Translational de-repression of Ash1 mRNA occurs in the daughter cell via:

- The type 1 casein kinase Yck1, which phosphorylates Khd1 at the plasma membrane and decreases its affinity for Ash1 mRNA (Paquin et al., 2007).
- The type 2 casein kinase Ck2, which accumulates at the cortex of yeast daughter cells and co-localizes with the translated pool of Ash1 mRNA. Ck2 phosphorylates and releases Puf6 from Ash1 transcripts (Deng et al., 2008).
Figure 15: **The tight translational control of Ash1 mRNA.** (1) In the nucleus, Ash1 mRNA binds She2 and potentially Puf6 co-transcriptionally (see section 1.4.2.2) forming an immature RNP. (2) In the cytoplasm, a mature RNP forms with Khd1 binding. This RNP is translationally inactive due to Khd1 interactions with eIF4G and 40S preventing the formation of a pre-initiation complex and proper ribosome assembly respectively. Puf6 also blocks the 60S ribosomal subunit through eIF5B interactions. (3) She2 links this silent RNP to MTs for transport. (4) Khd1 and Puf6 phosphorylation by the kinases Yck1 and Ck2 respectively reduces their affinity for the transcript causing their release. (5) Localized protein synthesis occurs at the tip of the daughter cell (Besse and Ephrussi, 2008).
2. Visualizing mRNA

Imaging approaches have revolutionized our characterization of various biological processes including mRNA metabolism. Contrary to most biochemical methods, imaging provides insights at the single cell resolution in a population of heterogeneous cells or in tissues. RNA was the first molecule to reach single molecule resolution in fixed and live cells (Femino et al., 1998; Fusco et al. 2003) with techniques called single-molecule FISH (smFISH) and MS2-labelling, respectively. Visualizing single RNA molecules in both live and fixed cells using fluorescent microscopy has since provided unique insights into various aspects of gene expression including stochasticity and RNA localization. Technical improvements, advances in microscopy, and improved fluorophores now allow us to image thousands of mRNA species simultaneously in fixed cells, and to track mRNAs during precise phases of their lifetime. All of this can be done in single intact cells, preserving subcellular environment and allowing the extraction of spatio-temporal information.

In this chapter, I will begin by outlining standard and useful variants of smFISH techniques. I will then focus on high throughput smFISH approaches with an emphasis on mRNA localization screens. Finally, I will describe techniques for imaging RNA in living cells and highlight their strengths and limitations, as well as potential improvements in the field of live RNA imaging.

2.1 Low throughput smFISH techniques

2.1.1 Standard smFISH techniques

In smFISH, multiple fluorescent oligonucleotides (probes or oligos) hybridize to the same target RNA allowing the visualization of an RNA molecule as a diffraction limited spot under a wide field microscope in fixed cells (Figure 16 and Figure 17-A; Femino et al., 1998). Background comes from unbound single probes while single molecules are resolved due to multivalent hybridizations of many probes on the same transcript. The signal to noise ratio (SNR) is, in theory, proportional to the number of fluorophores/probes bound to the RNA. In general, 24 probes provide decent signal to noise ratios and detection rates. Probes range from 20 to 50 nts and can either be directly (Femino et al., 1998; Raj et al., 2008), or indirectly (Tsanov et al., 2016) labeled by one or several fluorophores.

Indirect labeling approaches such as single molecule inexpensive FISH (smiFISH; Tsanov et al., 2016) involve a pre- (or post-) hybridization step that couples a segment of an unlabeled
primary probe called the readout sequence, to a fluorescently labeled secondary oligonucleotide, by base complementarity (Figure 17-B). Advantages of such approaches include flexible design and a few fold decrease in costs.

Figure 16: smiFISH against Abnormal Spindle Microtubule Assembly (ASPM) mRNA in Hela Kyoto cells. smiFISH was performed using 48 primary probes each pre-hybridized to an oligonucleotide labeled with two Cy3 molecules. Single ASPM mRNA appears as dots shown in red indicated by white arrowheads. Transcription sites appear as bright foci in the nucleus and are indicated by white circles. Note the cell-to-cell variability in mRNA levels and active transcription site appearance. DNA was stained with DAPI and is shown in blue (original data from my PhD).

2.1.2 Signal amplifying smFISH techniques

These techniques are similar to smiFISH except that they use the readout sequence as a platform for binding additional labeled oligos in deliberate amplification schemes. They can result in a 10 to 100-fold increase in signal and allow imaging RNA in optically challenging samples such as thick tissues and embryos (Choi et al., 2010). They can also reach single-molecule sensitivity using a small number of primary probes, which makes them suitable for imaging small RNA species including miRNAs (Larsson et al., 2010). These techniques include:

- FISH using Sequential Tethered and Intertwined Complexes (FISH-STICs; Sinnamon and Czaplinski, 2014): primary probes consist of an RNA hybridizing segment and a readout sequence. The readout sequence hybridizes to amplifiers, which in turn bind five oligos, each labeled with a fluorophore (Figure 17-C).
- Branched DNA (bDNA) FISH (Wang et al., 2012): here, primary probes are a pair of contiguous oligos that are both required to bind pre-amplifiers. Pre-amplifiers then bind amplifiers that in turn hybridize to labeled detectors (Figure 17-D).
- Hybridization chain reaction (HCR; Choi et al., 2010): in this variant, the readout sequence acts as a template that initiates the amplification of two complementary labeled oligos that exist as metastable hairpins. This results in the amplification of a fluorescent polymer in situ on the transcript (Figure 17-E).
- Padlock FISH (Larsson et al., 2010): this method employs probes that adopt a circular topology upon binding the RNA or cDNA. This binding allows ligation and closure of the probe which permits amplification via a rolling circle mechanism or recursive padlocking (Figure 17-F).

**Figure 17: Variations in smFISH techniques.** (A) Basic smFISH uses directly labeled oligos. (B) smiFISH adds a pre-hybridization step that couples a 28 nt readout to a labeled oligo forming a duplex. (C) FISH-STICs binds an amplifier to the readout sequence which in turn binds labeled detectors. (D) bDNA FISH is similar to FISH-STICs with the exception that two primary probes form a contig that is required to bind pre-amplifiers. (E) The HCR approach amplifies two complementary labeled hairpins in situ by using the readout sequence as an initiator. (F) FISH with padlock probes employs RCA or recursive padlocking to amplify and detect its target (illustration modified from Pichon et al., 2018).

While such techniques are quite powerful, they are not without disadvantages: (i) protocols tend to be long. (ii) Probes can aggregate leading to false positives. (iii) The use of a low number of
primary probes lowers detection rates since probes may not have access to a particular site on the transcript (due to translating ribosomes for instance). (iv) Large probe structures often cannot enter certain subcellular compartments, impeding RNA detection there (bDNA cannot enter the nucleus for example) (Battich et al., 2013). Nevertheless, these techniques are witnessing rapid improvements in specificity, robustness, multiplexity, and background suppression. Third generation HCR based FISH techniques, for example, allow the simultaneous single-molecule imaging of four mRNAs in whole mount chicken embryos (Choi et al., 2018).

2.2 High-throughput smFISH techniques and screens

The first high-throughput mRNA localization screen was done in Drosophila embryos using whole mount FISH (Lécuyer et al., 2007). This approach did not reach single molecule sensitivity but nevertheless showed that a remarkable 71% of expressed genes out of the 3370 studied genes show a particular localization pattern at certain stages of Drosophila embryogenesis. Various new localization patterns were described, as well as correlations between the transcript's localization and the function of the protein it encodes. This highlighted the wide functional scope of mRNA localization in controlling a wide range of biological processes during development.

In the following years, smFISH improvements and innovations in probe generation strategies allowed imaging thousands of transcripts at the single molecule level in single cells. In particular, sequential hybridization and barcoding strategies (discussed in detail below) succeeded in the simultaneous imaging of single mRNAs transcribed from thousands of genes in cells and tissues. Such multiplexing approaches opened the door to image-based transcriptomics that provide very rich datasets of gene expression and spatial information in situ and in intact single cells. In this section, we highlight FISH techniques used in large scale screens as well as mRNA localization information gleaned from them, with a focus on more recent techniques.

2.2.1 Systematic bDNA FISH

2.2.1.1 Screen design

This screen, carried out by (Battich et al., 2013) was the first to use smFISH in a large-scale format. This was done using bDNA technology (with 15 primary probe pairs; detailed in Figure 18-a to h) combined with an automated liquid handling robot, and high content imaging with low magnification air objectives. Probes were synthesized individually against 928 human genes involved in a variety of basic cellular functions, cancer signaling, and metabolism. Two
independent screen replicates were performed in Hela cells grown in 384 well plates, with stains against mitochondria, nuclei, and cell outlines. However, the probe generation protocol is complex and costly to scale up, which makes this approach difficult to apply in a regular lab.

Figure 18: A bDNA FISH screen. (a) bDNA probes used in this screen consisted of 15 contiguous gene-specific primary probe pairs (each pair around 50nts) that bind 28 nts of a pre-amplifier DNA molecule that in turn bind 15-30 nts of an amplifier labeled with four Alexa dye molecules. (b) Standard o-nuc FISH which uses a set of mono-labeled primary probes. (c) MYC mRNA in Hela cells revealed by bDNA FISH using an epifluorescent microscope at 100x magnification. mRNA is shown in green and the DAPI stained nuclei in blue. Scale bar is 13 microns in the overview subpanel, and 5 microns in the zoomed pannels. The negative control with no primary probe is shown. (d) Same as panel c, but with the o-nuc FISH technique. (e) Intensity profile of pixels in the marked region in panel c after extracellular background subtraction. (f) Same as panel e, but for the region marked in panel d. (g) Mean intensity of 100 spots detected by bDNA and o-nuc FISH after local background subtraction at 1 ms exposure time and 1 camera gain. The spot’s equator is marked by a dashed line. (h) SNR along the spot’s equator estimated in pannel g. (i) Subcellular localisation of mRNAs transcribed from mitochondrial genes and the gene RAN. mRNA is shown in green, and mitochondria in purple. Scale bars represent 5 microns. Yellow circles indicate detected mRNA molecules (illustration modified from Battich et al., 2013).
2.2.1.2 Analysis of mRNA localization features

mRNA localization was assessed based on the distance of each molecule to the cell’s nucleus, the cell’s edges, and to all other molecules. Based on this, five localization classes were described: (i) polarized (mRNA accumulating in one side of the cell), (ii) distal aggregated (clusters of mRNAs far from the nucleus), (iii) distal non-aggregated (exclusion of mRNAs from areas surrounding the nucleus), (iv) proximal (mRNAs around the nucleus), and (v) spread out (random mRNA distribution throughout the cytoplasm). One outcome of this study was that 11 of 13 mRNAs encoded by the mitochondrial genome, co-localized with the mitochondrial stain (Figure 18-i shows some examples with RAN as a negative control).

2.2.2. Multiplexed error robust FISH (MERFISH)

2.2.2.1 Screen design

This screen employs a novel FISH approach named MERFISH (Chen et al., 2015) for detecting hundreds of distinct RNA species at the same time. This multiplexing is based on indirect RNA detection as in smiFISH. All the unlabeled primary probes are hybridized at once on cells, and this is coupled to a combinatorial labeling of the readout sequences sequential rounds of hybridizations, imaging, and bleaching. Each round of hybridization detects only a subset of RNAs that read “1” in a binary code bit (Figure 19-A). Each RNA species is thus coded by an N-bit binary word, where N is the number of hybridizations. In theory, N rounds of hybridization can detect $2^N - 1$ RNA species. However, as N increases, the calling rate (ability to identify a transcript) decreases and misidentification rate increases exponentially. To address this, the authors applied a robust encoding scheme called modified hamming distance 4 (MHD4) that mathematically selects a subset of the $2^N - 1$ binary words that can be correctly called even with one error in the code, and thus allows for error corrections.

The labeling itself was carried out in two steps in an IMR90 human fibroblast cell line. First, an encoding hybridization uses primary probes with an RNA targeting sequence and two flanking readout sequences. Each RNA species is assigned four out of a pool of N readout sequences in an N-bit MHD4 code. Thus, each transcript must be targeted by two types of primary probes (each carrying two different readout probes, for a total of four). To increase the SNR, 192 primary probes were used against each mRNA. In the next phase, N readout hybridizations identify each readout sequence with secondary, fluorescently-labeled FISH probes (Figure 19-B and C). To generate the
probes, a custom oligo pool was amplified by PCR, in vitro transcribed and finally converted to single-strand DNA by reverse transcription. Imaging was done using a wide-field microscope with oblique incidence excitation (HiLo microscopy).

### 2.2.2.2 RNA detection and localization analysis with MERFISH

Two MERFISH measurements were done using two different encoding schemes: (i) detection of 140 mRNAs using a 16-bit MHD4 code that can perform error correction. (ii) Detection of 1001 mRNAs with a 14-bit MHD2 code that cannot correct errors and is prone to misidentification. Spatial distribution analysis was done on the 140 mRNAs detected using the 16-bit MHD4 code. Pairwise correlation of spatial density profiles of each transcript showed two groups with strong correlation after clustering.

Transcripts in group 1 were enriched in the perinuclear region, while those of group 2 were close to the cell periphery (detailed in Figure 19-D). Group 1 had mRNAs encoding extracellular proteins such as FBN1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1. Thus, the enrichment of these transcripts at the perinuclear region probably reflects cotranslational targeting of mRNAs to the ER as part of the SRP pathway. Group 2 contained mRNAs encoding actin-binding proteins such as FLNA, MT-binding proteins such as CKAP5, as well as the motors MYH10 and DYNC1H1. These transcripts had a similar localization pattern as actin mRNA and likely belong to the same localization group.
Figure 19: An mRNA localization screen using MERFISH. (A) A schematic representation depicting the principle of sequential hybridization and coding. mRNAs that fluoresce are assigned a “1” in a binary code in each round of hybridization. (B) MERFISH hybridizations target each RNA with two sets of encoding probes that each contain an RNA targeting sequence and two readout sequences. Each RNA will thus have four total readout sequences out of N, where N is the number of readout hybridizations. During readout hybridizations, a complementary labeled oligo binds to a readout sequence causing a subset of RNAs to fluoresce. Finally, an MHD4 code identifies each RNA species based on the specific rounds in which it fluoresced. (C) All detected single molecules after 16 rounds of hybridizations in a

51
composite false-colored image. White circles in the boxed region indicate potential RNA species, while red circles indicate misidentified RNAs. (D) Two cells showing detected mRNAs that localize around the perinuclear area (group 1, shown in blue) and mRNAs that accumulate close to the cell’s periphery (group 2, shown in red). Histograms quantify the average distances of mRNAs in groups 1 and 2 to the cell edge and nucleus compared to non-localized mRNAs, as well as gene ontology (GO) term enrichment in both groups (illustration compiled and modified from Chen et al., 2015).

2.2.3 Sequential FISH (seqFISH and seqFISH+)

2.2.3.1 seqFISH principles and applications

Sequential Fluorescence in-Situ Hybridization (seqFISH) is similar in concept to MERFISH. It uses successive rounds of hybridization, imaging, and stripping to label RNA with a single specific fluorophore at a certain round. This again establishes unique temporal barcodes that can be used to identify different transcripts. The number of available barcodes scales as \( F^N \), where \( F \) is the number of fluorophores used, and \( N \) is the number of hybridization rounds (Figure 20). In situ sequential barcoding was first performed in yeast cells to barcode 12 genes using four dyes and two rounds of hybridizations (Lubeck et al., 2014).

Since then, a number of seqFISH variants tailor-made to address specific biological questions have been developed. Applications of seqFISH (and its derivatives) included lineage tracking in mouse embryonic stem cells (Frieda et al., 2017), transcription profiling in single cells of the hippocampus (Shah et al., 2016), identifying new neural crest cells in chicken embryos (Lignell et al., 2017), and large scale imaging of transcription in single cells (Shah et al., 2018).

seqFISH-based approaches reached transcriptome-wide scales in vitro by imaging 10,212 transcripts immobilized on an oligo-dT glass surface (Eng et al., 2017). However, global profiling of transcripts in intact cells and tissues was faced with the problem of optical crowding, in which each RNA molecule will occupy a diffraction limited spot in the image. Imaging tens to hundreds of thousands of molecules simultaneously will generate overlapping signals and will thus hinder single molecules from being properly resolved in the image. Efforts were made to combine super-resolution and expansion microscopy with multiplexing schemes, but these approaches were only applicable to optically thin samples and included long imaging times (Lubeck and Cai, 2012).
Figure 20: seqFISH principle. (A) Schematic representation of a generic seqFISH experiment. (B) The first panel shows all RNA molecules detected across five rounds of barcoding, which are shown in subsequent panels. White squares indicate RNAs that are correctly identified, yellow squares indicate transcripts that can be correctly identified after effort correction, while red ones indicated RNAs that produced a pattern that does not correspond to any barcode (illustration modified from www.seqfish.com/technology).

2.2.3.2 seqFISH+ hybridization principle

The problem of optical crowding was solved in the largest in situ super-resolved RNA detection screen performed to date by a technique named seqFISH+ (Eng et al., 2019). This relatively simple, but incredibly powerful approach, replaces the 4-5 colors used in seqFISH by a palette of 60 “pseudocolors” across three channels (Alexa 488, Cy3b, and Alexa 647) effectively diluting the population of different RNA molecules detected in each hybridization round (details in Figure 21). The 60 images are then used to construct a super-resolution image with RNAs resolved below the diffraction limit. This approach allowed the imaging of RNAs transcribed from 10,000 genes in NIH/3T3 fibroblasts, as well as sections of the mouse brain using a standard confocal microscope. However, it is important to mention that cells were embedded in acrylamide, mRNAs attached to a
matrix, and cellular components removed by extensive protease treatments. This reduces background, but also risks loosing some mRNAs and cellular integrity.

**Figure 21: seqFISH+ barcoding principle.** seqFISH+ used a set of 24 primary probes to target each RNA. Primary probes contain an RNA binding segment and four overhang sequences called barcodes (barcodes I to IV). One barcode is targeted by a complementary labeled oligo in three channels during each barcoding round (with round IV used for error correction). Each barcoding round contains a series of twenty readout hybridizations in each color, in which only one-twentieth of RNAs in one channel are detected in each image lowering transcript density. In a way, each channel is subdivided into 20 false colors, each revealed during one readout hybridization. Each RNA is barcoded within only one fluorescent channel and fluoresces during one particular readout hybridization in each barcoding round. This configuration provides $20^3$ barcodes (20 pseudocolours and 3 rounds of barcoding excluding the error correcting round) in each channel for a total of 24,000 barcodes across the three channels. (modified from Eng et al., 2019).

### 2.2.3.3 Subcellular localization patterns observed using seqFISH+

SeqFISH+ was performed against 10,000 genes in NIH/3T3 fibroblasts and this revealed an average of 35,492 transcripts per cell (+/- 12,222). Transcripts were clustered on the basis of co-occurrence in a 10x10 pixel window. Three major localization clusters were observed: (i) the nuclear-perinuclear area, (ii) the cytoplasm, or (iii) cellular protrusions. The nuclear-perinuclear cluster was further divided into three sub-clusters, each enriched with genes that carry out distinct functions (Figure 22). Sub-cluster 1 contains mRNAs encoding extracellular proteins and that are thus translated on the ER, sub-cluster 2 is related in the actin cytoskeleton, while sub-cluster 3 contains proteins that regulate MTs.
Figure 22: **mRNA localization patterns revealed by seqFISH+ in NIH cells.** (A) The image depicts a cell showing color-coded mRNAs on all three localization clusters, and the table lists all genes in each of those clusters. Scale bar is 10 microns. (B) The table shows all genes within each subcluster of the nuclear-perinuclear cluster, and the cell shows an smFISH image against three genes of subcluster 1. Scale bar in 10 microns (illustration modified from Eng et al., 2019).

Finally, I will end this section by noting that multiplexed smFISH approaches have and still are witnessing rapid improvements which makes them an increasingly appealing alternative to single-cell sequencing approaches in gene expression studies (Chen et al., 2018). Such FISH approaches are sensitive enough to detect low abundance transcripts, and provide spatial information related to cell shape, the microenvironment, sub-cellular and tissue level localization. It is possible that they will outperform single cell sequencing in the future.

### 2.3 Imaging RNA in living cells

While smFISH approaches are powerful tools for visualizing many aspects of gene expression, they are limited to fixed cells and thus provide a “snapshot” of an RNA’s life. Here, I will outline techniques used in live imaging of RNA with an emphasis on single molecule detection, as well as briefly illustrate their applications in gene expression and regulation studies.
2.3.1 Techniques to visualize RNA in living cells

2.3.1.1 Tag multimerization approaches

Ash1 mRNA was the first transcript to be visualized in living cells using a multimerization approach (Bertrand et al., 1998). This was achieved by fusig bacteriophage stem-loop repeats called MS2 to a reporter mRNA. MS2 stem loops are recognized and bound by the MS2 coat protein (MCP) with high affinity. Fusing MCP to GFP (or other fluorescent proteins and molecules), allows MS2-tagged transcripts to become fluorescently labeled (Figure 23-A). Since then, the MS2 system has become the standard in live imaging of RNA.

MS2 stem loops are 19 nt long and can bind dimers of fluorescently labeled MCP (MCP-FP) protein with a sub-nanomolar affinity. 24 MS2 repeats can bind enough MCP-FP molecules to resolve single molecules of RNA (Fusco et al., 2003). MS2 tagged mRNAs and MCP-FP have been expressed in a variety of hosts such as bacteria, yeast, insects, plants, mammalian cells, and model organisms using different approaches (homologous recombination, gene editing, retroviral infection, and plasmid transfection) (Tutucci et al., 2018a). Unbound MCP-FP causes background when imaging, but this can be minimized by fine-tuning MCP-FP levels or targeting it to a different subcellular compartment.

When labeling transcripts with MS2-MCP, it is important to control for artifacts resulting from the tagging process:

- A first step is comparing tagged and untagged RNAs (using smFISH for example) with and without MCP-FP expression for steady-state levels and localization.
- Another important aspect is the affinity of the MCP to the MS2-stem loops. High-affinity MCP-MS2 interactions are suitable for studying RNA dynamic by fluorescence recovery after photobleaching (FRAP) experiments for example (Darzacq et al., 2007; Boireau et al. 2007), but tend to prevent the RNA from being properly degraded, especially in yeast. Lower affinity MS2 repeats, however, allow artifact free studies of RNA degradation (Tantale et al., 2016; Tutucci et al., 2018b).
- Optimizing the number of repeated stem-loop is also essential. 24 MS2 repeats (as in the original design) generally allow imaging several hundreds of frames using a wide-field microscope, but longer repeats (up to 128 stem loops) allow the recording of thousands of images. The choice of repeat number depends on the biological question, and the temporal...
Optically channeling samples and transcriptional studies (Tantale et al., 2016) often use higher numbers of compact degenerate loops (that improve RNA folding) since they require improved SNR and higher temporal resolution/longer imaging times.

Figure 23: Single-molecule visualization techniques in living cells. (A) Labeled RBPs bind arrays engineered on the RNA of interest. (B) An array of peptide epitopes called SunTag attached to the protein of interest bind a single chain variable fragment (scFv) of an antibody coupled to GFP. (C) Small Fluorogenic ligands fluoresce upon binding specific RNA aptamers fused to the target of interest. (D) MBs fluoresce only upon target binding that separated the fluorescent molecule from its quencher. (E) GFP fused inactive Cas9/Cas13 enzymes can label RNA with the help of a gRNA (Pichon et al., 2018).

Multicolor single molecule RNA imaging can be performed in live cells using other tags. An orthogonal bacteriophage-derived stem-loop system called PP7 can be used with labeled PP7 coat protein (PCP) in combination with the MS2-MCP system to image two RNAs simultaneously (Chao et al., 2008; Hocine et al., 2013). Other coat-protein and stem-loops based systems include Blg1
stem-loops (Chen et al., 2009), λ boxB RNA (Lange et al., 2008), and the human U1A protein (Takizawa and Vale, 2000). A conceptually similar system called the SunTag (detailed in chapter 3) allows imaging single molecules of protein by attaching an array of 12 to 24 epitopes that are recognized by a GFP-tagged monochain antibody to the protein of interest (Figure 23-B) (Tanenbaum et al., 2014). Remarkably, the SunTag can be used to image translation of single mRNPs (see chapter 3 below).

2.3.1.2 Other approaches

- **Fluorogenic RNAs:** These molecules are engineered to bind to conditionally fluorescent dyes, changing their conformation into a fluorescent state only upon aptamer binding (Figure 23-C). One well-known aptamer called spinach (Paige et al., 2011), binds a DFHBI chemical producing fluorescence comparable to that of GFP. Fluorogenic RNAs allow background-free imaging when fused to the RNA of interest, but have not yet reach single molecule sensitivity. In the near future, it is likely that improvements in brightness, folding, and cell permeability of aptamers will make them the ideal choice for imaging RNA in living cells (Bouhedda et al., 2017).

- **Molecular beacons: (MBs):** MBs are single-stranded DNA probes that contain a fluorescent molecule on one end, and its quencher on the other. MBs naturally adopt a stem-loop structure which suppresses the fluorescent molecule. Binding to a high specificity target abolishes the stem loops, causing the MB to fluoresce (Figure 23-D). An advantage of MBs is that they can target endogenous untagged RNAs. However, the probe often requires microinjection or another perturbing method to be delivered and their design is often complicated (Tyagi and Kramer, 1996).

- **CRISPR-Cas based systems:** A catalytically inactive dead Cas9 (dCas9) has been used to track RNAs in living cells (Nelles et al., 2016). Dead versions of the more recently characterized RNA editing Cas13 family of enzymes can also be repurposed to image RNA when fused to GFP (Figure 23-E; Cox et al., 2017). While these systems do not yet resolve single molecules, they can target endogenous untagged RNAs and are programmable via their guide RNAs (gRNAs).
2.3.2 Applications in imaging RNA metabolism

Live imaging techniques have allowed the imaging of various aspects of RNA metabolism from birth to death in real time. Such approaches allow the extraction of spatio-temporal information and kinetics of such processes. The positioning of tags (MS2 and PP7 stem loops for example) in reporters is of particular importance and depends on the biological phenomenon being investigated. Several processes that have been imaged in living cells include:

- **Transcription elongation rates**: the elongation speed of RNA pol II can be calculated by using two tags along the gene. The time it takes for the second color to appear after the first is a direct readout of the elongation rate (Figure 24-A). Elongation rates are not very different in various model organisms and range from 1 to 4.5 kb/s (Fukaya et al., 2017; Hocine et al., 2013).

- **Splicing**: placing a RNA tag in an intron allows to detect pre-mRNAs and provides a powerful system for visualizing splicing in living cells. For example, the MINX intron was shown to splice co-transcriptionally, in about in 2.5-3 min (Schmidt et al., 2011). A second RNA tag can be placed in an exon, allowing to detect both the spliced and unspliced transcripts in the same cell (Figure 24-B). This allows determining the fraction of spliced RNAs as well as whether splicing occurs during or after transcription (Coulon et al., 2014).

- **Cytoplasmic RNA trafficking**: tracking single RNA molecules in the cytoplasm (by often using 24 MS2 repeats) provides valuable information regarding the mechanisms by which they are localized (rectilinear directed movements reflect motor directed movements for example). One interesting observation is that, regardless of cytoplasmic distributions, mRNAs display complex movements that involve stochastic switching between immobility, diffusive movements, and directional transport (Fusco et al., 2003). This suggests that mRNAs often cycle between anchored, diffusive, and motor bound states.

- **Translation**: mRNA imaging with the coat-protein knock-off sensor (TRICK) allows to image the first round of translation. In this case, one set of stem-loops is placed in a UTR and another different set is placed in the ORF of the mRNA. Translating ribosomes will displace labeled coat proteins bound to the stem loops within the ORF, but not the UTR.
during the pioneer round of translation (Halstead et al., 2015). This causes dual-labeled mRNAs to become mono-labeled after being translated for the first time. More recently, the SunTag system has been combined with MS2/PP7 labeling to image single polysomes (detailed in chapter 3, Figure 24-C).

- **RNA decay**: viral pseudoknots (PKs) block the action of RNA exonucleases. Two-color imaging with a PK element between two RNA tags (Figure 24-C) allows distinguishing full-length RNA (MS2 and PP7 positive) from 3’end intermediate degradation products (MS2 but not PP7 positive). Interestingly, decay intermediates of the mRNA did not accumulate in P-bodies (see section 1.3.1.2; Horvathova et al., 2017).

![Diagram](image.png)

**Figure 24: Analyzing RNA processing using live imaging approaches.** (A) Transcription elongation rates are calculated based on the time delay, indicated by an arrow on the graph, between the two colors appearing in the transcription site. (B) A splicing reporter is shown on top that produces a PP7
tagged intron and an MS2 tagged spliced RNA. Graphs trace the two color intensities at a transcription site if splicing occurred co- or post-transcriptionally. (C) The schematic depicts SunTag repeats (in orange) bound by scFv-GFP (in green) which labels nascent peptides (in grey) in a polysome. Graphs depict polysome intensities after inhibiting translation in a ribosome run-off experiment or FRAP. (D) PKs separate PP7 from MS2 repeats which are protected from Xrn1 mediated degradation which allow imaging these stabilized intermediates (Pichon et al., 2018).
3. Translation goes live

The ability to visualize single transcripts in fixed and live cells has greatly improved our knowledge of RNA biology and metabolism. The last component of the central dogma of molecular biology to be imaged at single-molecule sensitivity was translation. This chapter briefly outlines biochemical approaches for measuring translation. The focus is then shifted towards visualizing translation in living cells at the single molecule level via nascent peptide imaging, using various multimerized tags and their accessories. Finally, spatial and quantitative insights gained from imaging translation are discussed.

3.1 Biochemical methods for quantifying translation

Biochemical methods provide a detailed view of translation at the genome-wide level and in a population of cells. Such approaches can be grouped based on their readout methods:

- Approaches that rely on RNA-sequencing (RNA-seq). These include translating ribosome affinity purification (TRAP) and Ribo-tag techniques (Sanz et al., 2009) which involve tissue-specific stable expression of tagged ribosomal proteins (Figure 25-A). Tagged ribosomes are pulled down and the associated mRNAs are analyzed via microarray or RNA-seq. Ribosome profiling methods also rely on sequencing and are powerful for identifying new ORFs, upstream ORFs (uORFs), and ribosome stalling sites (Ingolia et al., 2009). Ribosome profiling takes advantage of the fact that ribosome-bound fragments of an mRNA are protected from nuclease digestions, and can thus be isolated, sequenced, and identified (Figure 25-B). Partial spatial information can be inferred via localized labeling of ribosome populations within cells. This is done in proximity-specific ribosome profiling (Williams et al., 2014) in which ribosomes are first tagged with a biotin acceptor peptide. A local population of ribosomes is then biotinylated by the activity of a biotin ligase fused to a certain localization element, which allows the ribosomal population to be purified and analyzed (Figure 25-C).

- Approaches that rely on mass spectrometry (MS). These rely on the incorporation of exogenously added molecules and compounds into newly synthesized proteins, followed by MS quantifications. Puromycin-associated nascent chain proteomics (PUNch-P) is based on incorporating biotin-puromycin that can mimic the 3’ end of aminoacylated transfer RNA (tRNA) into nascent peptides (Aviner et al., 2013). Streptavidin beads and MS allow
quantifying newly synthesized proteins (Figure 25-D). **Pulsed stable isotope labeling by amino acids in cell culture** (pSILAC, Schwanhäusser et al., 2009) and **bio-orthogonal/quantitative non-canonical amino acid labeling** (BONCAT and QuaNCAT) respectively rely on short incubations with isotope-labeled amino acids (aa) and the incorporation of a biotin-binding methionine analogue (called azidohomoalanine, AHA) in newly synthesized proteins (Dieterich et al., 2006; Howden et al., 2013). In both cases, MS is used to measure *de novo* protein synthesis (Figure 25-E and F).

![Figure 25: Biochemical approaches for quantifying translation](image)

**Figure 25: Biochemical approaches for quantifying translation.** Techniques are grouped based on the readout method. (A) TRAP/Ribo-tag involves purifying specifically tagged ribosomes. (B) Ribosome profiling isolated ribosome-protected mRNA fragments. (C) Same as in B, but a tagged subpopulation of ribosomes is pulled down. Methods in A-C use RNA sequencing to identify the transcript and measure translation. (D) PUNch-P incorporates biotin-puromycin in nascent peptides. (E) pSILAC and (F) BONCAT/QuaNCAT use heavy aa isotopes and labeled methionine analogs respectively to differentiate *de novo* protein synthesis. Methods in D-E use mass spectrometry to quantify translational output (Chekulaeva and Landthaler, 2016).

While some of these methods (proximity-specific ribosome profiling for example) can provide some form of spatial information regarding translation, they still involve lysing cells in which local translation in its native context cannot be measured. In addition, sequencing and MS-based techniques that pool together cell populations do not take into account the translational heterogeneity between single mRNA molecules and individual cells.
3.2 Visualizing translation at the single molecule level in real time

While GFP is well suited for visualizing the expression, localization, and dynamics of proteins in vivo, it cannot be used for visualizing translation at the single-molecule level because of two reasons: (i) GFP only fluoresces after its fluorophore is fully formed and this takes several minutes, precluding its use as a readout of translation in real time; (ii) even after maturation, GFP molecules in cells are not dispersed enough for single molecule identification and bleach quickly (Chao et al., 2012).

3.2.1 Nascent peptide imaging

Early approaches for imaging translation at the single molecule level in eukaryotes relied on coat protein knock off approaches that can detect the first round of translation (see section 2.3.2) as well as co-localizing mRNAs with labeled ribosomes (Katz et al., 2016).

A breakthrough was made in 2016 in which five independent studies (Morisaki et al., 2016; Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016) succeeded in imaging the translation of single mRNAs in living cells over multiple rounds. Although different tags were used in these studies, they were all based on the principle of multimerization. Each tag consists of a set of epitopes fused to the N-terminal of the protein of interest, which allows to amplify the signal and to break the single molecule barrier when imaging proteins (see also the SunTag above; Tanenbaum et al., 2014). In addition, the great advantage of such systems in imaging translation is that each epitope is detected as soon as it is translated, because it can be bound by a pre-existing fluorescent antibody-like probe. Thus, the nascent peptide, in theory, is labeled as soon as it emerges from the ribosome (in a similar manner as MCP-GFP binds MS2 stem-loops as soon as the mRNA is transcribed). One distinction, however, is that mRNAs are often translated by multiple ribosomes (i.e. forming polysomes), in which multiple tags (each containing an epitope array) are synthesized in close proximity on the same transcript (Figure 26-A and B). This further amplifies the signal and allows detecting translation sites of single mRNPs within living cells. The signal appear as bright foci and can be imaged for minutes to hours in a simple wide-field microscope.

Combining nascent peptide imaging tags with RNA detection tags (or smFISH in fixed samples), permits the visualization of both translated (RNA and nascent peptide double positive) and untranslated mRNAs (only RNA positive). Quantitation of the signals further allows the the calculation of translation efficiency and the number of ribosome on each mRNA (Figure 26-C).
Figure 26: **Visualizing translation in living cells.** (A) The beginning of the gene of interest is tagged by a peptide epitope array to visualize the nascent peptide and the end is tagged by stem loops for mRNA visualization. (B) Labeled antibody fragments bind the epitope array and are further amplified by multiple ribosomes translating the same transcript (Li et al., 2016). (C) Snapshots of U2OS cells transiently expressing a reporter encoding the mRNA for blue fluorescent protein (BFP) fused to the 3’UTR of β-actin, in addition to 24 SunTag repeats at the beginning and 24 MS2 stem loops at the end of the construct. scFv-sfGFP and MCP-halo tag were also expressed at low levels. Polysomes appear bright green foci that colocalise with mRNAs. White arrows indicate untranslated mRNA molecules that appear as red dots, while yellow arrows indicate translated mRNAs that appear as yellow foci. Scale bars indicate 5 and 2 microns in the overview and zoomed panels respectively (illustration modified from Wu et al., 2016).

Three tags were used in the 2016 studies: (i) the SunTag, a 19 aa epitope derived from the yeast GCN4 protein (Tanenbaum et al., 2014) that is repeated 24 times with a linear GSGSG linker, for a total length of 571 aa (Wang et al., 2016; Wu et al., 2016; Yan et al., 2016); smaller (12x) and larger (32x and 56x) SunTag repeats have also been successfully used (Pichon et al., 2016); (ii) 10x repeats of the classical HA (9 aa, Wilson et al., 1984) or FLAG (8 aa, Hopp et al., 1988) tags, which are presented in a structured scaffold called the “spaghetti monster” (Viswanathan et al.,...
The spaghetti monsters consist of a non-fluorescent core with strings of HA or FLAG epitopes hanging off for a total length of 325 aa.

In addition to the tags, fluorescent antibody-like probes are required to label the tagged proteins co-translationally. The SunTag is recognized by a monochain antibody that can be genetically encoded. It consists of a single-chain variable fragment (scFv) fused to a fluorescent protein (superfolder GFP, sfGFP, which works particularly well), and a GB1 solubilization peptide. The scFv-sfGFP fusion has been optimized for solubility in the reducing cytoplasm of living cells (Tanenbaum et al., 2014). Spaghetti monsters are recognized by anti-FLAG or anti-HA antibody fragments (Fabs) conjugated to organic dyes, which can be microinjected into cells. Note that recently, the Stasevitch group also reported a recombinant monochain antibody against the HA tag (Zhao et al., 2019). No matter of the choice of the tagging system, it is important to adjust the amounts and subcellular localization of probes in the cytoplasm since unbound antibody fragments are a source of background.

One major concern of these tags is their large size. When bound by their labeled probes, 24 SunTag repeats weigh 1.4 MDa, while the 10x epitope spaghetti monster is 0.54 MDa. Remarkably, some observations suggest that translation is not greatly affected by the extra mass: (i) elongation rates measured with SunTag and spaghetti monsters are roughly consistent with each other; (ii) tagging a reporter protein with the SunTag only marginally affects its expression levels; (iii) varying the SunTag repeat numbers (from 4 to 56) does not have a large impact on polysomes mobility and translation kinetics (Morisaki and Stasevich, 2018; Pichon et al., 2016; Yan et al., 2016).

### 3.2.2 Accessory tags for improving nascent peptide imaging

In addition to tags mentioned in the previous section, two other useful accessory tags can be used in translation imaging:

- **mRNA immobilization tags**: the rationale behind this is that the high mobility of mRNA necessitates fast acquisition rates for tracking single particles, which leads to faster photobleaching and shorter time-lapses. Artificially tethering mRNA molecules to subcellular compartments via genetically encoded tools restricts mRNA movements and allows tracking single molecules at a lower frame rate, and thus for longer periods. mRNAs
can be tethered to the plasma membrane via cloning a CAAX prenylation domain to the C-terminal of the MCP or PCP (Figure 27-A, Yan et al., 2016). For ER tethering, a cytoplasmic end of an endoplasmic reticulum signal-anchor membrane protein (cytERM) is fused to the N-terminal of the protein of interest, upstream of the nascent peptide imaging tag (Figure 27-B, Wu et al., 2016). The disadvantages of immobilization tags are that they could alter mRNA translational kinetics and impede the study of local translation.

- **Degron tags:** fully translated released proteins can sequester the fluorescent detector antibodies and, depending on subcellular localization, may hinder the detection of polysomes since both mature proteins and nascent peptides will be labeled with the same probe. In such cases, degradation tags such as degrons are often fused to the protein of interest. Degrons (and any protein attached to them) are quickly and efficiently degraded (Wang et al., 2016; Wu et al., 2016). Including a degron in the tag prevents the accumulation of mature proteins so that nascent peptides are imaged with a better signal-to-noise ratio. However, degron induced degradation may interfere with translational processes and its kinetics (Hurley et al., 2016).

![Figure 27: Accessory tags used in nascent peptide imaging.](image)

(A) Immobilization tags include a CytERM peptide fused to the N-terminal of the nascent peptide tag and a CAAX sequence fused to the mRNA stem-loop binding protein that target the transcript to the ER or plasma membrane respectively. (B)
Degron tags upstream of the nascent peptide tags enhance degradation of the mature protein product (illustration modified from Morisaki and Stasevich, 2018).

### 3.2.3 Nascent peptide imaging of endogenous transcripts

Combining the CRISPR-Cas9 technology with the SunTag allows to detect the translation sites of endogenous transcripts. This is done by introducing a repair template containing a start codon, a resistance marker, a self-cleavage sequence, epitope repeats, and a FLAG/HA tag all flanked by homology arms (Pichon et al., 2016) in a cell line that expresses the appropriate fluorescent secondary antibody probe. Homology-directed repair (HDR) causes precise insertion of the tag in the correct frame of the endogenous gene leading to the appearance of bright foci that correspond to polysomes. Resistant clones must be assessed for correct insertion by genotyping. An important control is also to co-localize endogenous transcripts (revealed by smFISH) with the observed bright foci (Figure 28).

![Figure 28: Imaging the translation of an endogenous gene.](image)

The ASPM gene was tagged with 32 SunTag repeats using CRISPR-Cas9 in Hela Kyoto cells stably expressing scFv-sfGFP. SmiFISH with 48 Cy3 labeled probes was done against the endogenous ASPM mRNA. Images of a heterozygous clone were taken with a wide-field microscope. (A) The full field of view. mRNAs are shown in red, and...
polysomes/mature proteins in green. DAPI stained DNA is shown in blue (B) Zoons of the white boxed areas in A. The upper panel shows a cell in mitosis, while the lower one shows a cell in interphase. White arrowheads indicate mRNAs that are either not translated, or untagged, while yellow ones indicate translating mRNAs. White circles indicate regions of mature protein accumulation. Scale bars measure 10 microns (original data from my PhD).

The advantages of tagging endogenous transcripts are that mRNAs are transcribed from their native genomic context, with all their regulatory sequences and are expressed at their physiological levels (provided the tagging process does not affect this). In addition, expression is stable across a population of cells, which facilitates studies that require expression over many generations (cell cycle studies for example).

3.3 Understanding translation at the single molecule level

Here I will discuss some applications of nascent peptide imaging and describe the insights gained from imaging gene expression at the level of translation.

3.3.1 Ribosome numbers and kinetics

3.3.1.1 Calculating ribosome density on a single mRNA

In principle, dividing the fluorescence intensities of polysomes by the one of single molecules of mature protein indicates how many ribosomes are translating the transcript. Mature protein molecules can be identified by the persistence of the signals after puromycin treatment (a translational inhibitor), and by the absence of co-localization with mRNA, and the intensity of their signals can thus be easily measured. One complication for the quantification, however, is that the ribosomes that are still translating the epitope array have only translated half of it on average. Thus, to calculate the ribosome density, the effective length of the ORF becomes \( L' = L_{\text{POI}} + 0.5L_{\text{Tag}} \) where \( L_{\text{POI}} \) is the length of the protein of interest, and \( L_{\text{Tag}} \) is the length of the tag itself (Figure 29-A). Therefore, \( F' = F \times d \times L' \) where \( F' \) is the fluorescence of a polysome, \( F \) is that of a single mature protein, and \( d \) is the ribosome density (Pichon et al., 2016; Wang et al., 2016; Yan et al., 2016). Ribosome densities measured across the five studies indicate densities of one ribosome every 200 to 1000 nt, depending on the transcript.

3.3.1.2 Calculating translation elongation rates

This involves measuring the fluorescence dwell times at a translation site, which is then used to infer elongation rates. This, however, is based on three biological assumptions: (i) polysomes are
in a steady state in which an initiating ribosome replaces the ribosome that finishes translation; (ii) ribosomes that initiate translation always complete it through the entire ORF; (iii) nascent chains that are completely translated immediately leave the polysome. Under these assumptions, the elongation rate \( K_{\text{elong}} \) is the effective ORF length \( L' \), divided by the dwell time. Two approaches that can quantify translation elongation rates are ribosome runoff and FRAP experiments. Fluorescence correlation spectroscopy can also be used (Morisaki et al., 2016) but is more technically demanding and requires many hundreds or even thousands of images.

- **Ribosome runoff experiments**: run-off can be induced using the initiation-specific translational inhibitor harringtonin. The basic concept is that while no new ribosomes will enter the transcripts, the remaining elongating ribosomes will finish translation one after the other, leading to the progressive loss of fluorescence at the translation. Alternatively, stochastic fluctuations in translation initiation rate can be used to find polysome that turn off translation. The decrease in polysome intensity occurs in four stages (I-IV; detailed in Figure 29-B). Fitting stage III of the decay curve into a straight line allows the calculation of the ribosome dwell time (Wang et al., 2016). Alternatively, the entire curve can be fitted with a model describing all four stages (Pichon et al., 2016).

- **FRAP experiments**: the concept here is to bleach the probes that label nascent peptides and measure the rate of signal recovery, which occurs by the synthesis of new nascent peptide (Figure 29-C). This requires the assumption that the probes bind irreversibly during the time scale of the FRAP experiment, ensuring that fluorescent recovery is due to nascent protein synthesis and not probe turnover. An advantage of FRAP, however, is that translation is not inhibited by the use of drugs. Provided that polysomes are at steady-state, FRAP recovery curves can be fitted in an inverse manner as those of runoff experiments (Pichon et al., 2016; Wu et al., 2016).
Figure 29: Calculating translation elongation time. (A) Schematic representation of \( L_{\text{tag}} \) and \( L_{\text{POL}} \). (B) Runoff assays using harringtonin cause loss of fluorescence in four stages. Stage I corresponds to the steady state in which the biological effect of harringtonin did not yet take place. During stage II, initiation is inhibited but the runoff is counter-balanced by elongating ribosomes translating newly labeled epitopes. Stage III begins when all epitopes have been translated in which fluorescence loss is maximal and linear. Dwell times are calculated during this phase when the fitted straight line reaches zero. Stage IV corresponds to heterogeneity in ribosomes leaving the polysome. (C) FRAP assay recovery curves also have four stages and can be fitted inversely with respect to run-off curves (illustration modified from Morisaki and Stasevich, 2018).

Remarkably, the measured elongation rates across the five studies are in general agreement with an average close to 10 aa/s and experimental values ranging from 3 to 18 aa/s. The exact reasons behind these variations are unknown but could be related to codon usage/optimization, experimental interpretation and calculations. Another factor is the expression method, in which elongation rates measured from transiently transfected constructs were slower than measurements performed on endogenous transcripts (Pichon et al., 2016).

3.3.1.3 Calculating translation initiation rates

This parameter can be calculated in a straightforward manner once the number of ribosomes on the transcript and the elongation rates are measured. The polysome steady-state assumption applies here as well, in which terminating ribosomes are replaced by initiating ones at a constant rate, in addition to another assumption that ribosomes move at a constant rate. In this case, initiation time (inverse of the initiation rate) is the total elongation time divided by the number of
ribosomes on the transcript. Interestingly, the five studies agreed on a narrow range of 2-5 ribosomes initiating every minute (Morisaki et al., 2016; Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). This precise agreement could reflect a common factor that limits initiation rate (such as the 60S concentration for instance), or a biological phenomenon that prevents ribosomes from initiating again during a lag period.

3.3.2 Heterogeneity in translation

In this subsection, I will highlight some aspects of gene expression heterogeneity that were uncovered by single-molecule imaging of translation.

3.3.2.1 Heterogeneity in translation efficiency

Single molecule studies of translation have shown that mRNAs can switch between translationally active and inactive states. Indeed, some transcripts are constitutively translated while others cycle between periods of translational bursts and silence (Pichon et al., 2016; Wu et al., 2016). For example, mRNAs encoding the large subunit of the RNA polymerase II, are constitutively translated with 90% of transcripts engaged in translation at any given time. In contrast, mRNAs encoding components of the molecular motor dynein are translated in bursts with around 40-70% of transcripts actively translating at a given time. These observations suggest that burstiness could be genetically encoded, and perhaps relate to the initiation efficiency and the stability of initiation complexes (Pichon et al., 2016).

Another observation is that the same transcript can have different translational efficiencies in different cell lines and subcellular compartments of the same cell. A good example is the 3’UTR of β-actin fused to a reporter, which is constitutively translated in human osteosarcoma cells, but bursty in neurons (Wu et al., 2016). Moreover, this transcript has lower translation efficiency in distal dendrites than in proximal ones and glial cells (detailed in Figure 30). This suggests that post-transcriptional processes affect translational state patterns as well.
Figure 30: **Spatial variations in translation efficiencies in neurons.** (A) smFISH and immunofluorescence experiments detect mRNAs (in red) and translation sites/mature protein (in green) originating from a reporter carrying, among other elements, 24 SunTag repeats and the 3′UTR of β-actin (see figure 26) in neurons. (B and C) Enlarged segments of proximal and distal dendrites. Yellow arrowheads indicate translated mRNAs while white ones indicate untranslated transcripts. All scale bars represent 5 microns. (D) The fraction of translated mRNAs in proximal and distal dendrites compared to glial cells in the same culture. Unpaired t-test, ***p < 0.001. (E) The fraction of translating mRNA in dendrite as a function of distance to the soma (calculated from 53 dendrites, 19 neurons) (Wu et al., 2016).

### 3.3.2.2 Heterogeneity in polysome mobility and localization

In addition to translational kinetics and efficiency, polysome movements can be quantified by these repeated epitope tags. This is done by tracking single polysomes and calculating the mean-squared displacement (MSD) as a function of time. Based on this parameter, four types of movements can be distinguished (Fuscor et al. 2003; Park et al., 2010): (i) confined, in which polysomes are bound to an immobile structure; (ii) corralled, where polysomes can move freely, but within a restricted subcellular area and have MSD values that start linear, but then reach a plateau that corresponds to the boundaries of that area; (iii) purely diffusive, when polysomes can move within the entire cytoplasm and when the MSD plot is linear; (iv) finally, some polysomes display rectilinear directed transport in which they move towards random or specific directions at constant velocities; this is a good indication of active motor-driven transport.

Single mRNAs that are either translationally active or inactive have been shown to undergo active transport in living cells (Pichon et al., 2016; Wu et al., 2016). This was very surprising since the dogma in the RNA localization field is that mRNA undergoing active transport should be translationally silent (Besse and Ephrussi, 2008). However, more studies are needed to determine
how common co-translational targeting of mRNA is in different cellular contexts. Motorized movements of polysomes range from 1 to 3 microns/s and are often interrupted by stochastic switching into diffusive states. Motorized movements are particularly important for establishing local translation in neurons in which dendrites and axons are separated from the nucleus by vast distances.

3.3.2.3 Heterogeneity in mRNA decoding

Cutting edge multi-color imaging of translation has been recently used to address the complexity and heterogeneity in decoding single mRNAs (Boersma et al., 2019). Selection of the correct translation start site by initiating ribosomes is essential for identifying the correct ORF that will, in turn, produce a functional protein. However, identification of the correct start site by scanning ribosomes is not a straight-forward process since: (i) the 5’UTRs of many genes contain more than one AUG (Iacono et al., 2005); (ii) translation can still initiate, though less efficiently, at near-cognate start codons such as GUG or CUG (Lee et al., 2012); (iii) translation at a canonical start site may not be perfectly efficient (Lind and Åqvist, 2016); (iv) after translating an uORF, ribosomes may reinitiate translation at multiple downstream start sites (Hinnebusch et al., 2016).

In order to visualize non-canonical start site selection, a system orthogonal to the SunTag was developed and named MoonTag (Boersma et al., 2019). Moontags consist of repeated 15 aa epitopes from the HIV envelope protein subunit gp41 that are recognized by labeled nano-bodies (Nbs). The SunTag and MoonTag epitopes were then combined in an alternating fashion in a single tag termed MashTag. A MashTag can translate SunTag repeats, MoonTag repeats, or no repeats depending on the reading frame in which translation is taking place. Two MashTags were constructed (details in Figure 31), with start codons in frame for translating either the SunTag (SunStart) or MoonTag (MoonStart). Both start codons were placed in strong initiation sequence contexts (Kozak sequence), and no other start codons were present in the 5’UTR of both reports. As expected, SunStart and MoonStart MashTags showed predominant SunTag and MoonTag polysomes, respectively. However, the MoonStart reporter showed occasional pulses in SunTag, in which the mRNA was double positive for both epitope repeats. Analyzing intensity traces showed that this out-of-frame (OOF) translation is not due to ribosome frameshifting but rather to alternative start site selection. Overall, 7% of the ribosomes showed OOF translation. Alternative start site selection was shown to be stochastic and did not depend on translation initiation rates. Moreover, the usage of a particular start site did not show bursty characteristics and the small
fluctuations of the relative frequency of SunTag and MoonTag translation could be accounted for by chance. However, different mRNA molecules showed different probability of OOF translation, and this property seem to be dynamically regulated over time. Finally, alternative near-cognate start sites could be located either upstream or downstream the AUG start codon.

Figure 31: A multicolor reporter for visualizing alternative start site selection. (A) The MeshTag contains 36 alternating SunTag and MoonTag epitopes. Translation through ORF 0 produces the MoonTag, through ORF -1 the SunTag, while the ORF +1 produces no epitopes. (B) The MashTag was placed upstream of a gene coding for BFP that lacks stop codons in all frames and followed by 24 PCP repeats in the 3’UTR to track the mRNA. The 5’UTR contained only one start codon in-frame with the MoonTag called MoonStart. The construct was expressed in U2OS cells with stable expression of SunTag-scFv-sGFP, MoonTag-Nb-HaloPYG66 dye, and PCP-2xmCherry-CAAX. MoonTag polysomes colocalize with the mRNA and were sensitive to puromycin. Circles with an arrowhead represent start
codons while those with a square represent stop codons. (C) Same as in B, but with a start codon for the SunTag named SunStart, and SunTag polysomes seen. (D) mRNAs that show only canonical translation are marked by an asterisk, while those showing OOF are marked by both an asterisk and an arrowhead. (E) Schematic of ribosomes mediating canonical and OOF translation on the same transcript. All scale bars represent 1 micron (Boersma et al., 2019).
4. mRNA localization and translation regulation during the cell cycle

Cell division is a fundamental process in all forms of life and it requires extensive changes in cell’s morphology and properties. For proper progression through the cell cycle, gene expression must be carefully and precisely regulated at both the transcriptional and post-transcriptional levels. While a large part of this regulation is based on a dedicated set of kinases and phosphatases, the principles of RNA localization and translational regulation discussed in previous chapters also contribute to cell cycle regulation and will be addressed again here in this context. I will begin by outlining generalities concerning the eukaryotic cell cycle, with brief emphasis on translational regulation. I will then focus on the structure and function of centrosomes, as well as mRNA localization on centrosomes. Finally, I will expand on one particular protein that is of particular interest for this thesis.

4.1 The eukaryotic cell cycle

4.1.1 Main phases and regulators of the cell cycle

A important function of the complex machinery controlling progression through the cell cycle is to ensure that one round of DNA replication is followed by an event of cell division and that each event is properly completed before the next one begins.

The eukaryotic cell cycle consists of an interphase and a cell division phase called mitosis. DNA replication canonically occurs during the synthesis (S) phase of interphase, which, in most cases, lies between two gap phases termed G1 and G2 (Pines, 2011). An extra phase, called G0, exists which corresponds to a quiescent state. Cells in G0 often have reduced transcriptional and translational outputs and do not progress into G1 until appropriate mitogenic signals are received. G1 is a period of cellular growth and protein synthesis, especially for many nuclear proteins. A restriction point (R point) is defined in G1, which corresponds to the moment in which the cell becomes "committed" to complete the next cell cycle and after which extracellular proliferation stimulus are no longer required. G2 is the phase that precedes mitosis and involves the synthesizing of proteins necessary for the cell to divide. Mitosis consists of five phases (shown in Figure 32), and ends in cytokinesis (the physical separation of the two daughter cells). Note that in parallel to the cell cycle, a centrosomonal cycle takes place, which involves duplication and maturation (detailed in the next section). The length of the cell cycle varies among different
organisms, different cell types of the same organism, and between identical cell grown in different environmental conditions. Typically, immortalized cell lines grown in culture take about 24 hours to divide with G1 taking the most time and mitosis being completed in about an hour.

![Cell Cycle Diagram](image)

**Figure 32: The cell cycle and its main regulators in animal cells.** Each phase of the cell cycle is controlled by the activity of a certain cyclin-CDK complex. CyclinE-CDK2 has high activity towards the end of G1, past the R point. During S phase, cyclin A-CDK1,2 activity rises and is maintained till early G2. High cyclinB1-CDK1 activity occurs in late G2 and is needed for mitotic entry. Five stages of mitosis are depicted in the top section of the figure: (i) prophase involves nuclear membrane breakdown, chromosome condensation, and centrosome separation. (ii) During prometaphase, chromosomes start attaching to the mitotic spindle at kinetochores and unattached ones generate “wait anaphase” signals. (iii) Metaphase is defined by a characteristic plate of aligned chromosomes in which the wait signal is turned off. (iv) Sister chromatids separate in anaphase. DNA de-condenses and the nuclear membrane reforms during (v) telophase which precedes daughter cell formation by cytokinesis. MTOC: microtubule organizing center (detailed in the next section) (Pines, 2011).

The temporal order is of particular significance in the cell cycle. This is controlled by oscillation in the activity of cyclin-dependent kinases (CDKs, Coudreuse and Nurse, 2010). CDKs are a family of protein kinases that depend on their cyclin partners for activity. As their name indicates, the amount of cyclins vary during the cell cycle, due to specific stabilization/degradation regulations, and the different cyclins are phased differently. Importantly, the degradation of cyclins
causes the rapid inactivation of their associated kinase. Other regulation mechanisms exist on top of this and involve phosphorylations/de-phosphorylations and small peptide inhibitors. Many different types of cyclins are synthesized and degraded throughout the cell cycle, the main one cyclins of the A, B and E family (see Figure 32). The Cyclin-B family, for instance, is important for cells going in and out of mitosis. High levels of cyclin B-CDK activity drives cells into mitosis and prevents formation of replication origins, while its destruction is required for cells to exit mitosis and to assemble replication origins. One regulator of low and high CDK activity states is the anaphase-promoting complex, also known as the cyclosome (APC/C). APC/C is composed of about 20 proteins and it functions as an E3 ubiquitin ligase that is conserved across eukaryotic evolution. It recognizes mitotic specific cyclins and targets them for degradation by the 26S proteasome via ubiquitylation (Noton and Diffley, 2000).

4.1.2 Translational control by cytoplasmic poly-A tail length modulation

Modulation of the poly-A tail length by cytoplasmic polyadenylation is an important regulator of mRNA translation that drives many biological processes such as meiosis and cell cycle progression (Richter, 2007; Richter and Klann, 2009). A cis-acting element in the 3'UTR of such mRNAs called cytoplasmic polyadenylation element (CPE) binds a family of proteins called CPE binding proteins (CPEBs, four members in humans; Hake and Richter, 1994). Depending on the number of CPEs and their position with respect to the polyadenylation hexanucleotide (hex), CPEBs can either mediate translational activation or repression during meiosis. This combinatorial CPE code also controls the timing and extent of translational activation and repression by CPEBs (Piqué et al., 2008).

In Xenopus oocytes and early embryos, CPEB1 has dual functions: (i) maintaining mRNAs in a translationally repressed state in the absence of oocyte stimulation by shortening their poly-A tails and blocking interactions between the mRNA cap and the translation initiation machinery via the recruitment of eIF4E-binding proteins including maskin (Minshall et al., 2007); (ii) activating the translation of mRNAs by lengthening their poly-A tails in response to specific cues such as Aurora A kinase stimulation; this is done by recruiting a polyadenylation complex consisting of cleavage and polyadenylation specificity factor (CPSF), the cytoplasmic poly-A polymerase GLD2, and symplekin (Barnard et al., 2004). For example, during early divisions in Xenopus embryos, CPEB1 promotes polyadenylation-mediated translation of mRNAs encoding cyclin B1 during mitosis, and maskin-mediated repression of translation during S phase (Cao et al., 2006).
Early embryonic mitotic divisions are rapid (S and M phases, without G1 and G2) and often take place without transcription, in which gene expression regulation mainly occurs by post-transcriptional processes including CPEB-mediated translation control. However, it has been shown CPEBs regulate translation during canonical mitosis and cell cycles in which transcription is active (Novoa et al., 2010). Briefly, CPEB1 and CPEB4 control the cytoplasmic polyadenylation, and thus translation, of hundreds of transcripts during specific phases of mitosis. This translational level of regulation is essential for normal cell cycle progression in which short hairpin RNA (shRNA) mediated knock-down CPEB1 and CPEB4 induces mitotic entry defects. This study suggests that translational control by polyA tail length regulation is a general mechanism of gene expression that is essential for cell cycle progression.

4.2 The centrosome

4.2.1 Centrosome structure and composition

The centrosome is composed of two centrioles and their surrounding pericentriolar material (PCM).

4.2.1.1 Centriole structure

Centrioles are cylindrical structures formed by a nine-fold symmetrical array of MT triplets (figure 33a). Centrioles vary in size among different organisms and cell types with mammalian ones measuring around 230nm in diameter and 420nm in length. MT triplets contain 13 complete proto-filament of A-tubules and 10 incomplete proto-filament of B- and C- tubules, with an A-C linker connecting the A-tubule of one triplet to the C-tubule of an adjacent triplet (Greenan et al., 2018; figure 33b).
Figure 33: **Centriole and centrosome structure and organization.** (A) The general architecture of the mammalian centrosome. The centrosome is made up of two orthogonal centrioles surrounded by a protein matrix called PCM which contains gamma-tubulin ring complexes (γTuRC) important for MT nucleation and binding. (B) Schematic representation of a mature parent centriole and an associated procentriole. The proximal lumen of the procentriole contains a “cartwheel” composed of nine spokes emanating from a central hub. Each spoke ends in a pinhead that binds to the A-tubule of MT triplet (Breslow and Holland, 2019).

Two structurally distinct regions have been identified in mammalian centrioles using cryo-electron microscopy along the proximal-distal axis (Greenan et al., 2018): (i) the proximal domain, which shares a common architecture with the *Drosophila* centriole; this end recruits and organizes the PCM; (ii) the distal domain, which has a narrower diameter and distinct A-C linkers. In mature centrioles, this end contains nine distal appendages (detailed in figure 33b), which are required for docking the centrioles at the plasma membrane during ciliogenesis, and a variable number of sub-distal appendages, which are required for MT anchoring to interphase centrosomes. The presence of sub-distal appendages is controlled during the cell cycle.

### 4.2.1.2 Pericentriolar material

One key element in mammalian PCM is pericentrin (PCNT, also known as kendrin). The C-terminal of pericentrin contains a pericentrin-AKAP450 centrosomal targeting domain (PACT), which targets pericentrin to the PCM by a yet unknown mechanism. During interphase, pericentrin clusters and forms elongated fibrils that attach to the centriole through their PACT domain in a pattern that probably follows its nine-fold symmetry (Mennella et al., 2014).

Pericentrin is essential for efficient recruitment and incorporation of other PCM components such as CDK5 regulatory subunit-associated protein 2 (CDK5RAP2, also called
CEP215) and CEP192 with its binding partner NEDD1. These proteins can be seen as toroids of various sizes around the centrosome (Conduit et al., 2015). Another major component of the PCM is CEP152 which does not depend on pericentrin for its centrosomal localization. Notably, CEP152 is not essential for organizing the PCM during interphase, however, it is essential for assembling the mitotic PCM.

Some PCM components form a gamma-tubulin ring complex (γTuRC) that can recruit gamma-tubulin. γTuRC components include pericentrin (the N-terminal part of which binds the γTuRC, Takahashi et al., 2002), CDK5RAP2 (which interacts with pericentrin and γTuRC through its C-terminal CM2 motif and N-terminal CM1 motif, respectively (Wang et al., 2010), and NEDD1 (Haren et al., 2006). Major components of the PCM are well conserved through evolution and presented in Figure 34.

![Image of PCM components]

**Figure 34:** *Major PCM components during interphase and mitosis in various organisms.* (a) PCM composition during interphase around a mother centriole. (b) Major PCM components and their interactions in different organisms. (c) PCM expansion during mitosis. Nomenclature is provided for
PCM expansion occurs during mitosis in which the single orderly layer of PCM changes into an extended amorphous matrix. One major regulator of this process is the mitotic kinase PLK1, which localizes to the centrosome and phosphorylates PCM components enhancing their assembly and multimerization (Conduit et al., 2015). The exact molecular events of mitotic PCM expansion differ among species but a principle is the same: phosphorylation events drive monomeric proteins into assembly-competent states. In mammals, PCM expansion is driven by PLK1 mediated phosphorylation of pericentrin, which augments its ability to recruit other PCM components. It is noteworthy to mention that although CDK5RAP2 depends on pericentrin for centrosomal recruitment during interphase, both proteins depend on one another for centrosomal targeting during mitosis (Lee and Rhee, 2011).

4.2.2 Centrosome functions

Centrosomes are the major microtubule-organizing center (MTOC) in animal cells (Wu and Akhmanova, 2017). They function in MT nucleation, as well as MT stabilization and minus end MT attachment. During mitosis, centrosomes have a dominant role in establishing and correctly positioning the bipolar spindle, in particular via the formation of astral MTs. In interphase, centrosomes establish a radial organization of MTs, which is important for the positioning of membrane organelles and to determine cell polarity.

Pharmacological and genetic perturbation of key factors involved in centrosome duplication such as PLK1 cause centrosomal loss. Phenotype analysis of such cases revealed that centrosomes are not strictly required for cell survival even if centrosomes are physically ablated shortly before mitosis (Khodjakov et al., 2000). However, robust and error-free chromosome separation requires the presence of a functional centrosome (Meraldi, 2016). Centrosomes (centrioles in particular) also serve as a template for cilia formation. Interestingly, mouse embryos devoid of centrosomes produce developmental defects related to lack of cilia (Bazzi and Anderson, 2014).

Centrosomes also carry out MT and cell architecture independent roles. These functions include acting as signaling hubs, organizing actin filaments, and crosstalk with DNA damage response pathways (Conduit et al., 2015; Farina et al., 2016; Mullee and Morrison, 2016).
4.2.3 Centrosome biogenesis

In cycling cells, the centriole duplication cycle is tightly coupled to the cell cycle to ensure a constant number of centrioles in each cell after mitosis (Nigg and Holland, 2018). This imposes spatial, temporal, and numerical control over centriole biogenesis. In brief, a cell contains two centrioles at the start of each cycle connected by a flexible linker located at their base. The younger centriole is referred to as the parent centriole in which it was assembled in the previous cell cycle, and the older one is called mature parent centriole. At the G1-S transition, centriole duplication begins, in which a procentriole emerges and grows orthogonally from a single site at the proximal end of each parent centriole. This perpendicular configuration is called engagement and is maintained during S and G2 during which the procentriole grows into about 80% the size of a parent centriole before mitotic entry. In late G2, the flexible linker connecting the two parent centrioles dissociates allowing them to separate and to establish the bipolar spindle. During mitosis, the procentrioles disengage from their associated parent centriole so that each daughter cell inherits a pair of duplication competent parent centrioles (Figure 35).

![Diagram of centrosome biogenesis](image)

Figure 35: **Centrosome/centriole biogenesis during the cell cycle.** G1 cells contain one mature parent centriole with distal appendages and one parent centriole connected by a linker. At the beginning of S phase, each parent centriole orthogonally assembles one new procentriole. This configuration is termed engagement and prevents reduplication of the parent centriole. Procentrioles elongate as
the cell is progressing through the cycle. Late G2 marks the dissociation of centriole linker and centrosomes separation. Centrosomal maturation and PCM expansion occur in preparation of mitotic spindle formation. The cartwheel is removed from the lumen of the procentriole in human cells during mitosis. The orthogonal arrangement of each centriolar pair is lost towards the end of mitosis in which disengagement allows each daughter cell to inherit a pair of duplication competent centrioles. At the same time, the procentriole matures into a parent centriole allowing PCM recruitment and centriole-to-centrosome conversion. Note that distal and subdistal appendages are transiently disassembled/modified during mitosis. New appendages form on the mature parent centriole that was formed one and a half cycle ago during G1. The mature parent centriole can migrate to the plasma membrane and act as a template for the formation of a cilium in quiescent cells (Breslow and Holland, 2019).

4.2.4 mRNA localization and local translation at centrosomes

Around 20 years ago, mRNAs encoding the cell cycle regulator cyclin B were found to localize on the mitotic spindle in early *Xenopus* embryos (Groisman et al., 2000). This was CPEB dependent, since neutralizing CPEB MT binding ability or mutating the CPEB consensus binding site in the 3’UTR of cyclin B abolished cyclin B mRNA and protein localization at the mitotic apparatus. In *Xenopus*, these mutants show cell division defects and impaired spindle formation, suggesting that local protein synthesis of this important cell cycle regulator has some physiological significance.

Another interesting case of centrosomal mRNA localization was described in mollusk embryos. In this context, mRNAs localizing to centrosomes do not code for centrosomal proteins and are silenced. The purpose of centrosomal mRNA localization is in fact to asymmetrically segregate mRNAs between daughter cells during embryonic cleavages (Lambert and Nagy, 2002).

mRNAs enriched on centrosomes have also been described in *Drosophila* embryos. This was part of a screen that visualized 3370 mRNAs using whole mount in-situ hybridization during different stages of *Drosophila* embryogenesis (Lécuyer et al., 2007). In addition to several localization classes, six mRNAs were found to be enriched on centrosomes across all developmental stages. These mRNAs included ones that encoded centrosomal proteins suggesting local translation at centrosomes. Two more recent studies (Jambor et al., 2015; Wilk et al., 2016) described several others mRNAs as enriched on centrosomes, spindles, or peri-nuclear regions that could contain the centrosome in later stages of *Drosophila* development. Notably, mRNAs encoding Polo-like kinase, involved in regulating mitotic entry and spindle formation, had a peri-nuclear pattern.

More recently, the mRNAs encoding PCNT, a major component of the PCM (see section 4.2.1.2), were shown to be concentrated on the centrosomes of zebrafish embryos and human HeLa
cells during early mitosis (Sepulveda et al., 2018). This localization was sensitive to translation inhibition by puromycin, MT de-polymerization, and dynein activity inhibition. Moreover, the same study combined double immunofluorescence with smFISH to suggest that PCNT mRNA is translated at the centrosome (detailed in Figure 36). This study also suggested that this localized translation enhanced the incorporation of PCNT into mitotic PCM.
Figure 36: **Approach for visualizing translating PCNT mRNA at centrosomes.** (A) Schematic representation of smFISH against PCNT mRNA combined with double immunofluorescence. One antibody recognized epitopes found close to the N-terminal of the PCNT protein, and the other close
to the C-terminal in order to differentiate between nascent and fully formed PCNT protein. (B) Prometaphase HeLa cells subjected to PCNT smFISH and double immunofluorescence. Upper control panels show a magnification of the lower monochrome view in each channel. Nascent PCNT peptides are positive for the N-terminal antibody and the smFISH signal and negative for the C-terminal antibody. The orange boxed area is a highly contrasted view of the dashed orange boxed area containing nascent PCNT peptides. An important control is a puromycin treatment in which co-localization of PCNT mRNAs with the N-terminal antibody was lost. Scale bars represent 5 and 0.5 microns in the full field and the inset panels respectively. (C) PCNT smFISH signals between 1 and 3 millimeter radius from the centrosome center were quantified for the presence of anti-PCNT N-terminal immunofluorescence signals with or without a short puromycin treatment. Data are represented as mean ±95% CI (confidence intervals) from three biological replicates, with the total number of cells analyzed indicated. p-value was obtained with Student’s t-test (Sepulveda et al., 2018).

Although this study provides strong evidence towards co-translational protein targeting on MTs, it did not address the mechanism of transport. In addition, they used various drugs that can potentially disrupt many biological processes. Visualizing and tracking single polysomes in intact cells would be required to confirm this targeting mechanism.

4.3 Abnormal spindle microcephaly-associated protein

Abnormal spindle microcephaly-associated protein (ASPM), also known as abnormal spindle protein homolog or Asp homolog, localizes in the nucleoplasm during interphase and at the mitotic centrosomes during mitosis. It coordinates many mitotic processes including mitotic spindle orientation, microtubule dynamics, and poleward microtubule flux (Jayaraman et al., 2016).

Mutations of ASPM cause microcephaly, an abnormal reduction in brain size accompanied by mental retardation. This defect is likely due to cell division failures as well as improper asymmetric divisions, preventing the normal proliferation of neuronal stem cells (Fujimori et al., 2014). In addition, ASPM mutations and over-expression have been linked to several cancers including glioblastomas, (Bikeye et al., 2010; Hagemann et al., 2008; Horvath et al., 2006), liver cancer (Drozdov et al., 2012; Lin et al., 2008), melanomas (Kabbarah et al., 2010), pancreatic cancer (Wang et al., 2013), prostate cancer (Xie et al., 2017) and epithelial ovarian cancer (Brüning-Richardson et al., 2011). These latter tumors are highly aneuploid (Alsiary et al., 2014), consistent with the fact that ASPM is required for proper cell division. In addition, functional studies indicate that ASPM is a promising target to fight gliomas (Visnyei et al., 2011).

More recently, ASPM has been shown to interact with another protein linked to microcephaly named katanin (Jiang et al., 2017). Indeed, the two proteins localize to the spindle poles in a mutually dependent manner and form a physiological complex. The structural basis of ASPM-katanin interaction is a conserved repeat sequence of ASPM and a heterodimer formed from
the N-terminal of the p60 and the C-terminal of the p80 subunits of katanin. ASPM and katanin cooperate in specifically regulating MT minus-ends and spindle dynamics (Figure 37). Misregulation of this process could lead to microcephaly.

Figure 37: **MT minus-end regulation by an ASPM-katanin complex**. A Katanin heterodimer formed from the N-terminal of the p60 and the C-terminal of the p80 subunits (that can bind and bend MTs) forms an interface that can bind conserved repeats in ASPM. The ASPM-katanin complex regulates MT flux via two mechanisms: 1. MT binding by the calponin homology (CH) domain of ASPM and severing by the ATPase domain of the katanin P60 subunit. 2. Binding MT minus-ends (which promotes ASPM and katanin accumulation near the spindle pole) via katanin end binding activity which enhances the minus-end growth blocking ability of ASPM (illustration modified from Jiang et al., 2017).
mRNA localization and local translation allow fine tuning of gene expression. A few mRNA localization screens have been done in human cells, but none of them directly addressed local translation. Moreover, recent evidences based on polysome imaging, to which I contributed, suggest that mRNAs in transit can be translationally active. This contradicts the dogma that mRNAs are transported in a translationally repressed state. Therefore, my thesis addressed the following questions:

• Is local translation a widespread phenomena in human cells?

• To what extent does co-translational targeting of mRNAs occur in a HeLa cell? Is RNA localization and RNA- or protein-driven process?

To this end, I contributed to two screens and then focused on a family of mRNA localizing to centrosomes, to study in details their localization mechanisms.
Results
1. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells

This paper had the following aims:

- Improve on the original SunTag system and test the new design by imaging the translation of a reporter.

- Combine the improved SunTag system with gene editing to tag genes, enabling the visualization of endogenous polysomes.

- Characterize the translation of two endogenous mRNAs (POLR2A and DYNC1H1) in living cells. This included translational kinetics such as elongation rates, ribosomes densities, and bursting; as well as polysome dynamics and trafficking.

Main conclusions of the paper were:

- The SunTag can be combined with CRISPR-Cas9 gene editing to visualize endogenous polysomes.

- Translation of endogenous transcripts alternates between active and inactive states.

- DYNC1H1 mRNAs are translated in dedicated translation factories.

In this study, I specifically contributed by:

- Performing a SunTag knock-in using CRISPR-Cas9 to tag the endogenous POLR2A and DYNC1H1 genes.
• Screening and characterizing POLR2A and DYNC1H1 SunTag clones by genotyping, smFISH experiments, and visual inspection.

• Analyzing and interpreting my data: showing that POLR2A mRNA is translated in individual polysomes while DYNC1H1 is translated in dedicated translation factories.

• Generating figures and commenting the text.
Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells

Xavier Pichon,1 Amandine Bastide,1 Adham Safedidine,1 Rachael Chouaib,1 Aubin Samacoits,2 Eugenia Basyuk,1 Marion Peter,1 Florian Mueller,2 and Edouard Bertrand1

1Institut de Génétique Moléculaire de Montpellier, UMR5535 CNRS, 34293 Montpellier Cedex 5, France; Université de Montpellier, 34090 Montpellier, France
2Unité Imagerie et Modélisation, Institut Pasteur, UMR3691, Centre National de la Recherche Scientifique, 75015 Paris, France

Translation is an essential step in gene expression. In this study, we used an improved SunTag system to label nascent proteins and image translation of single messenger ribonucleoproteins (mRNPs) in human cells. Using a dedicated reporter RNA, we observe that translation of single mRNPs stochastically turns on and off while they diffuse through the cytoplasm. We further measure a ribosome density of 1.3 per kilobase and an elongation rate of 13–18 amino acids per second. Tagging the endogenous POLR2A gene revealed similar elongation rates and ribosomal densities and that nearly all messenger RNAs (mRNAs) are engaged in translation. Remarkably, tagging of the heavy chain of dynein 1 (DYN1H1) shows this mRNA accumulates in foci containing three to seven RNA molecules. These foci are translation sites and thus represent specialized translation factories. We also observe that DYN1H1 polysomes are actively transported by motors, which may deliver the mature protein at appropriate cellular locations. The SunTag should be broadly applicable to study translational regulation in live single cells.

Introduction

Translation is a fundamental step in gene expression, for which the importance is nicely demonstrated by the thousands of microRNAs that fine-tune expression of a large fraction of mRNAs (Lewis et al., 2005; Baek et al., 2008; Selbach et al., 2008; Hafner et al., 2010). Translational regulation not only provides a quantitative control of protein levels, but it also determines when and where a protein is produced (Sonenberg and Hinnebusch, 2009; Liu et al., 2016). In particular, proteins can be produced locally, in specific subcellular compartments, and this process is intimately linked to subcellular mRNA localization (Martin and Ephrussi, 2009; Jung et al., 2014; Buxbaum et al., 2015). Although many mRNAs localize randomly throughout the cytoplasm, some are highly enriched in particular cytoplasmic areas. Such specific RNA localization occurs in many organisms from bacteria to humans and plays important roles in a variety of cellular processes. Although localized mRNAs are believed to be translated locally, RNA localization can also be the result of other processes. For instance, mRNA can accumulate in processing bodies (P-bodies) for storage or degradation (Cougot et al., 2004; Pillai et al., 2005), and likewise, stress granules are believed to function as mRNA-protective and sorting centers when translation is globally repressed (Kedersha et al., 2000; Mollet et al., 2008; Decker and Parker, 2012). Recent large-scale studies in Drosophila melanogaster embryos and in human cell lines have revealed that a large number of mRNAs can localize in specific subcellular areas (Lécuyer et al., 2007; Battich et al., 2013; Wilk et al., 2016). These studies have also shown many unexpected localization patterns. Whether these patterns are related to local protein synthesis currently remains a mystery.

Although mRNA localization suggests local translation, spatial translational regulation is a distinct process that provides an additional layer of control. A well-known example is that of nanos mRNA in Drosophila oocytes (Gavis and Lehmann, 1994). This mRNA is weakly enriched at the posterior pole of oocytes, with only 4% of the total mRNAs being localized there (Bergsten and Gavis, 1999). However, a tight spatial control of protein synthesis prevents translation of nonlocalized mRNAs, such that the Nanos protein is only produced at the posterior pole. Thus, the spatial control of translation is an important process in itself.

Several microscopy methods have been devised to study translation at the level of single cells (Chao et al., 2012). FRAP and local protein photoconversion can be used to visualize newly translated proteins, but are limited by the time required for chromophore maturation. The translating RNA imaging by coat protein knock-off biosensor identifies single mRNAs that were never engaged in translation (Halstead et al., 2015), but it does not provide information on translation itself. In this study, we developed a strategy to directly image the translational...
activity of single mRNPs in live cells. Our approach makes use of the SunTag system (Tanenbaum et al., 2014), which allows visualization of single molecules of proteins. It is inspired from the MS2 and Laci technology that image single RNA and DNA loci, respectively, using a repeated tag (Robinet et al., 1996; Bertrand et al., 1998). The SunTag uses a recombinant single-chain variable fragment (scFv) antibody that recognizes a peptide from the yeast Gcn4 protein. The protein of interest carries up to 24 tandem repeats of the peptide epitope, whereas the scFv is fused to superfolder GFP (sfGFP). In vivo binding of the scFv-sfGFP to the epitope yields up to 24 molecules of sfGFP per protein, and this is sufficiently bright to visualize single molecules of proteins in wide-field microscopy. We show that the SunTag can also be used to visualize nascent proteins while still being translated. It provides a powerful tool to image translation of endogenous mRNAs in live cells and at the levels of single molecules.

Results

An improved SunTag system to visualize translation of single mRNPs

The SunTag system relies on the binding of a fluorescent antibody to a peptide epitope (Tanenbaum et al., 2014), and this epitope thus becomes immediately detectable upon binding of the antibody. We reasoned that this property could be used to image nascent proteins during their translation, by inserting the SunTag at their N termini. The original SunTag has up to 24 repetitions of the peptide epitope. To improve the sensitivity of the system, we generated a new repeat of 32 epitopes and combined it with the original SunTagx24 to obtain a repeat of 56 copies. We inserted the SunTagx56 at the N terminus of a hygromycin selectable gene and further fused the resulting construct to the mouse Ki67 cDNA, which encodes a 3,177-aa-long nuclear protein (Fig. 1 A). We reasoned that this would increase the number of ribosomes loaded on the reporter mRNA, thereby generating a stronger signal. We additionally introduced the MINX intron into the construct (Zillmann et al., 1988) to allow assembly of the exon junction complex and promote export and translation of the resulting mRNA (Le Hir et al., 2016). The polA signal of the reporter was derived from the herpes simplex virus thymidine kinase gene, and we further added an array of MS2 stem-loops in the 3′ UTR to visualize single mRNAs in live cells with a fluorescent version of the MS2 coat protein (MCP; Fusco et al., 2003). The resulting construct was termed SunTagx56-Ki67.

Imaging single polysomes of SunTagx56-Ki67

We generated stable clones of HeLa cells that expressed both this reporter and a nuclear version of the scFv-sfGFP. We hypothesized that this would decrease background signals, first by reducing the levels of free scFv-sfGFP in the cytoplasm and second by transporting the full-length SunTagx56-Ki67 protein to the nucleus. Clones expressing both the scFv-sfGFP and the SunTagx56-Ki67 reporter displayed numerous small dots corresponding to single proteins, as well as brighter protein foci (Fig. 1 B). In contrast, clones expressing only the scFv-sfGFP displayed a diffuse signal as previously reported (Tanenbaum et al., 2014; Fig. 1 C). Interestingly, the brighter protein foci of the SunTagx56-Ki67 clones disappeared after a brief treatment with the translational inhibitor puromycin, whereas single molecules of proteins were still visible (Fig. 1, B and D). This suggested that the brighter foci were translation sites. In agreement, single-molecule FISH (smFISH; Femino et al., 1998) with SunTag-hygromycin probes revealed that these brighter foci colocalized with single mRNAs (>95% of the time, n > 700; Fig. 1 E and Fig. S1). These brighter protein foci thus corresponded to nascent protein chains translated from a single polysome of the reporter mRNA.

The ability to detect single proteins and single polysomes allowed us to quantify translation at the level of single mRNAs. On average, we detected 8.4 mRNA per cell with 47% of them engaged in translation, but this fraction varied greatly from cell to cell (Fig. 2, A and B). We further measured that the mean nascent protein foci were as bright as 16 single proteins (±2.6 on replicate counts), corresponding to a mean density of 1.3 ribosome per kilobase (see Materials and methods). The nascent protein foci also showed a relatively wide intensity distribution (Fig. 2 C), with the first and last quartiles having means of 8.3 and 27 nascent proteins, respectively. This suggested that different mRNA molecules had different abilities to be translated or that their translation rate varied over time.

Translation of single mRNPs alternates between active and inactive states

To gain more insights into the dynamics of translation, we imaged live cells. We first imaged mRNAs and polysomes at the same time using cells stably expressing an nls-MCP-TagRFP fusion in addition to the scFv-sfGFP (Video 1). This demonstrated the feasibility of the approach and the possibility to simultaneously visualize both translated and untranslated mRNAs. However, the relatively poor quality of the red signal and its rapid bleaching made the experiment technically challenging. Most of the following experiments were thus performed in absence of MCP fusions, by imaging polysomes in a single color with the scFv-sfGFP. We acquired 3D image stacks for 45 min at a rate of six images per minute (Fig. 3 A and Video 2).

Individual nascent protein foci corresponding to single polysomes remained often detectable for tens of minutes, although their intensity could vary (Fig. 3 B). Interestingly, the intensity of a small number of protein foci gradually decreased for ~1 to 2 min before they became undetectable (Fig. 3 B and Video 3). This suggested that translation of these mRNPs had stopped, thereby resulting in the gradual release of the nascent proteins from the mRNP until it no longer contained any ribosomes. In addition, we could also detect particles appearing, as well as particles disappearing and reappearing again (Fig. 3 B and Videos 4 and 5), thus indicating that translation of single mRNPs alternated between active and inactive states. The number of translating mRNPs in the cytoplasm as well as their rapid diffusion made the long-term tracking of single polysomes difficult, and this precluded rigorous quantifications of the on and off time of single mRNPs. However, the gradual decrease of intensity occurring when a single polysome turned off could be used to provide an estimate of the ribosome elongation rate. To analyze these data, we implemented a mathematical model assuming a uniform ribosome distribution, a constant ribosome velocity, and a release of nascent proteins occurring immediately after completion of their synthesis (see Materials and methods). Individual “ribosome runoff” curves are shown in Fig. 3 C, and the fit to the model indicated an elongation rate of 13.2 aa/s (Fig. 3 D). To get additional insights into the dynamics of translation, we
performed FRAP experiments of nascent protein foci (Fig. 3, E and F). It was previously shown that binding of the scFv-sfGFP to its target is stable in vivo, with a $t_{1/2}$ dissociation rate between 5 and 10 min (Tanenbaum et al., 2014). Recovery of single translating mRNPs was complete in 3.5 min. Thus, FRAP recovery indicated synthesis and release of the tagged Ki67 protein rather than exchange of the scFv antibody on its target. To model the data, we used the same model as for ribosome runoff analysis (see Materials and methods). Fitting the FRAP data yielded a ribosome elongation rate of 18 aa/s, in reasonable agreement with the rate measured from the ribosomal runoff experiments. These rates are 2.5–3 times faster than genome-wide estimates based on runoff experiments using the translational inhibitor harringtonine (Ingolia et al., 2011). Given that foci of nascent proteins were as bright as 16 individual proteins, the FRAP data implied that ribosomes initiated a mean of once every 13 s.

Translating SunTagx56-Ki67 mRNPs diffuse through the cytoplasm
The live-cell movies described in the previous section showed that the nascent protein foci moved rapidly throughout the cytoplasm. To obtain quantitative insights, we acquired 3D stacks at a higher frame rate (2.2 images/s for 4 min; Fig. 4 A and Video 6), which allowed tracking of the translating mRNPs.
The histogram of displacements between two consecutive frames revealed a single population with a diffusion coefficient of 0.047 µm²/s (Fig. 4 B). We also calculated the mean square displacement (MSD) as a function of time by averaging individual particles. This indicated that diffusion was normal between 1 and 20 s, with a coefficient of 0.034 µm²/s (Fig. 4 C and Fig. S2 A), in reasonable agreement with the aforementioned estimate. These values were also in range with previous measurements of ribosome movements by spt-PALM (0.1 µm²/s; Katz et al., 2016). We then calculated individual diffusion coefficients for the subset of particles that could be tracked for at least 75 time points (i.e., traces longer than 36 s; n = 75 particles). This yielded a wide distribution of diffusion coefficient (0.042 ± 0.021 µm²/s; Fig. 4 D), indicating substantial differences between individual translating mRNAs (see Materials and methods for details). To test whether this heterogeneity correlated with the ribosome load, we plotted this diffusion coefficient as a function of particle brightness. A weak negative correlation was obtained (Fig. S2 B), indicating that the ribosome load had only a small effect on diffusion. This was further confirmed by measuring the diffusion of mRNAs using the nls-MCP-TagRFPt (Fig. 4 E). We focused on cells in which most of the mRNAs were not translated as assessed by the lack of scFv foci, and we observed that the SunTagx56-Ki67 mRNPs diffused at the same rate as SunTagx56-Ki67 polysomes. Again, this suggested that the ribosome load had little effect on the mRNP diffusion rate.

Visualizing translation of the large subunit of RNA polymerase II

Next, we attempted to visualize translation of endogenous mRNAs and focused first on the housekeeping gene POLR2A, which codes for the large subunit of RNA polymerase II. As in the case of Ki67, we reasoned that the length of its coding region (1,970 aa) would facilitate visualization of the nascent proteins. We generated a CRISPR tagging cassette by fusing a puromycin selectable marker upstream of the SunTagx56 repeat, with a P2A peptide in between them (Fig. 5 A). This leads to a physical separation of the puromycin N-acetyltransferase from the SunTag and thus ensured that the selectable marker was fully functional. The resulting cassette was introduced into a repair construct carrying POLR2A homology arms, such that the PurP2A-SunTagx56 was in frame with both the natural ATG codon and the POLR2A coding sequence. This cassette was then transfected together with guide RNAs and the Cas9 nickase (Ran et
al., 2013) in HeLa cells expressing either a cytoplasmic or nuclear version of the scFv-sfGFP fusion. Heterozygous recombinant clones were identified by genomic PCR and observed by fluorescence microscopy. We obtained the best signals with the cytoplasmic version of scFv-sfGFP and performed subsequent experiments with these clones. As in the case of the Ki67 reporter mRNA, the scFv-sfGFP detected single proteins as well as brighter protein foci (Fig. S3 A). The single proteins localized in the cytoplasm and not the nucleus, indicating that the SunTagx56 cassette disrupted the normal biogenesis pathway of RNA polymerase II (Boulon et al., 2010). Nevertheless, the brighter protein foci labeled by the scFv-sfGFP colocalized with smFISH probes detecting the RNA sequences of POLR2A (Fig. S3 B) and Puro-SunTagx56 (Fig. 5 B). Colocalization analysis revealed that 98% of the protein foci associated with single RNAs detected with the Puro-SunTagx56 probes (±1%, three replicate counts with \( n > 800 \); note that this is not the case for the POLR2A probes, as they also detected the untagged mRNA from the wild-type allele). In addition, the brighter protein foci disappeared after a brief exposure to puromycin, whereas single proteins were still visible (Fig. S3 A). This demonstrated that the bright scFv-sfGFP foci corresponded to nascent proteins being translated from a single mRNA and thus represented single POLR2A polysomes.

Quantification of the tagged POLR2A mRNA indicated that cells contained a mean of 33 mRNAs (±3 on replicate counts), with a unimodal distribution (Fig. 5 C, left). Interestingly, and in contrast to what was observed with the Ki67 reporter RNA, nearly all of the tagged POLR2A mRNAs appeared to be actively engaged in translation (mean of 91%), with little cell-to-cell variation (Fig. 5 C, middle). Image quantification indicated that the nascent protein foci were as bright as 12 individual proteins (Fig. 5 C, right). This corresponded to a density of 1.3 ribosome/kb, similar to the one observed for the SunTagx56-Ki67 mRNA. The frequency of initiation events thus appears to be similar for these two mRNAs, as long as they are engaged in translation. FRAP analysis of translating POLR2A mRNPs indicated that recovery occurred in \( \sim 3.5 \) min (Fig. 5 D), which corresponded to an elongation rate of 13.8 aa/s, similar to what was found for the SunTagx56-Ki67 reporter. Finally, analysis of the movements of translating POLR2A mRNPs revealed that these were diffusing at a rate similar to the one of the SunTagx56-Ki67 reporter (Fig. 5, E and F; Fig. S2 A; and Video 7). Again, individual particles displayed widely varying diffusion constants (Fig. 5 F, right).

Dynein heavy chain mRNAs accumulate in cytoplasmic blobs, where they are translated

We then focused on the heavy chain of dynein 1 (DYNC1H1), the main minus-end–directed motor in the cytoplasm of human cells. smFISH revealed that this mRNA was not randomly distributed throughout the cytoplasm but accumulated in foci (hereafter referred to as “blobs”), although single, isolated mRNA...
molecules were also present (Fig. 6 A). To confirm this localization pattern, we used a HeLa cell line containing a bacterial artificial chromosome (BAC) bearing the entire DYNC1H1 gene with a GFP tag inserted in frame at its C terminus (Poser et al., 2008). smFISH with a probe set recognizing the GFP tag showed that the exogenous mRNA also accumulated in blobs, confirming the specificity of the DYNC1H1 smFISH probes (Fig. 6 B). Interestingly, we could detect similar DYNC1H1 RNA blobs in all of the rodent and human cells that we analyzed: NIH3T3 cells (Fig. 6 C), U2OS, HEK, and primary neurons (not depicted), demonstrating the generality of this observation. Image quantification indicated that, depending on the cell line, cells contained a mean of 1.8 to 9.5 blobs, with three to seven mRNA molecules per blob (Fig. 6 D). Overall, 20–45% of DYNC1H1 mRNA accumulated in blobs. We then tested whether these blobs colocalized with P-bodies or stress-granule markers. P-bodies are small cytoplasmic structures known to degrade and store untranslated mRNAs (Cougot et al., 2004; Decker and Parker, 2012), whereas stress granules form only in stressed cells and are made of untranslated mRNPs (Decker and Parker, 2012). In unstressed cells, the stress-granule marker G3BP did not accumulate in the DYNC1H1 mRNA blobs (Fig. S4 A), and a P-body marker also labeled distinct structures (Fig. S4 B).

In the cell line with the BAC expressing GFP-tagged DYN1CH1, the GFP-tagged dynein subunit accumulated throughout the cytoplasm and was not enriched in the mRNA blobs (Fig. 6 B). However, GFP has a maturation rate of ~15 min, and we cannot determine where the protein is translated. To address this question, we tagged the endogenous dynein subunit at its N terminus with the Puro-P2A-SunTag cassette as done above for POLR2A. We obtained positive clones with both the SunTagx56 and the SunTagx32 cassette and did the following experiment with the SunTagx32. Heterozygous clones were confirmed by genomic PCR, and, in these clones, the scFv-sfGFP labeled single molecules of proteins as well as brighter protein foci that colocalized with mRNAs detected with oligonucleotide probes hybridizing to the Puro-SunTag sequences (Fig. 6 E). The brighter protein foci were no longer detected after a brief treatment with puromycin (Fig. S5 A), confirming that the corresponding nascent proteins still associated to translating mRNPs. Quantifications of the intensities of the nascent protein foci indicated that the DYNC1H1 mRNA associated with a mean of 34
proteins (Fig. S5 B). This corresponded to a mean density of 2.1 ribosomes/kb, higher than for the Ki67 reporter and the endogenous POLR2A mRNA. We then determined whether the tagged allele was properly targeted to the DYNC1H1 mRNA blobs. To this end, we performed a two-color smFISH experiment using probes labeling either the endogenous DYNC1H1 sequences or the Puro-SunTagx32 sequences (Fig. 6E). We counted a total of 27 DYNC1H1 mRNAs per cell, including 7 from the SunTagx32 allele. We found that 20% of the wild-type mRNA and 13% of the tagged DYNC1H1 mRNA (green), and the tagged DYNC1H1 mRNA (green). [Used] A 2.5 x 2.5-μm zoom of the boxed area. Bar, 10 μm. (F) Percentage of translated SunTag32-DYNC1H1 mRNAs, for mRNAs localizing in blobs or being single mRNAs (n = 360 cells). The error bars represent the SD of the mean between experimental replicates (n = 3). (G) Effect of inhibition of translation initiation on the accumulation of DYNC1H1 in blobs. (left) Number of DYNC1H1 polysome per cell, in cells untreated or incubated with Torin1. (right) Fraction of DYNC1H1 mRNAs in blobs, in cells untreated or incubated with Torin1. The error bars represent the SD of the mean between experimental replicates (n = 4).

Figure 6. Dynein 1 heavy chain mRNA accumulates in blobs that are translation sites. (A) Localization of endogenous DYNC1H1 mRNAs in HeLa cells. Image displays the smFISH signal of HeLa cells labeled with probes recognizing DYNC1H1 mRNAs. Blue arrow: single RNA molecules; red arrow: RNA blob. The green dashed outline represents the position of the nucleus. Bar, 6 μm. (B) Localization of BAC-tagged DYNC1H1 mRNAs in HeLa cells. Legend as in A, except that the probes labeled the GFP sequences. Left panel: smFISH images; right panel: GFP signal. Bar, 6 μm. (C) Localization of endogenous DYNC1H1 mRNAs in NIH3T3 cells. Legend as in A, Bar, 10 μm. (D) Quantification of the number of blobs per cell (left graph), the number of DYNC1H1 mRNAs per blob (middle graph), and the fraction of DYNC1H1 mRNAs present in blobs (right graph). 3T3: NIH3T3 mouse cells (n = 60 cells); BAC: HeLa cells expressing the DYNC1H1 BAC (n = 220 cells); HeLa: HeLa cells expressing the SunTag32-DYNC1H1 allele (n = 385 cells). The error bars represent the SD of the mean between experimental replicates (n = 3). (E) Colocalization among the bright protein foci labeled by the scFv-sfGFP, the endogenous DYN CH1 mRNAs, and the tagged DYNC1H1 mRNA. Panels represent microscopy images of cells stably expressing the scFv-sfGFP and containing the SunTag32-DYNC1H1 allele. Left panel: signal of the scFv; middle left panel: signal from smFISH probes recognizing the DYNC1H1 sequences; middle right panel: signal from smFISH probes recognizing the SunTag-Puro sequences; right panel: color overlay of the scFv signal (green), the endogenous DYNC1H1 mRNAs (red), and the tagged DYNC1H1 mRNA (green). [Used] A 2.5 x 2.5-μm zoom of the boxed area. Bar, 10 μm. (F) Percentage of translated SunTag32-DYNC1H1 mRNAs, for mRNAs localizing in blobs or being single mRNAs (n = 260 cells). The error bars represent the SD of the mean between experimental replicates (n = 3). (G) Effect of inhibition of translation initiation on the accumulation of DYNC1H1 in blobs. (left) Number of DYNC1H1 polysome per cell, in cells untreated or incubated with Torin1. (right) Fraction of DYNC1H1 mRNAs in blobs, in cells untreated or incubated with Torin1. The error bars represent the SD of the mean between experimental replicates (n = 4).
be correlated with particular movements or diffusion properties of the corresponding polysomes, we imaged live cells containing the SunTagx32-DYNC1H1 allele. Remarkably, this showed that DYNC1H1 polysomes displayed rapid rectilinear motion, suggestive of motor-dependent movements (Fig. 7, A and B; and Video 8). In a 4-min time window, 19% of the particles displayed a rapid rectilinear motion, with the median speed of the particles being 1 µm/s and the median distance traveled during a run being 2.3 µm. These values are similar to what has been reported for the directional movements of localized mRNPs in cell lines (Fusco et al., 2003). These rapid movements were dependent on an intact microtubule network and not on actin (Fig. 7 C and Video 9). This suggested that they resulted from an active process involving molecular motors.

In contrast to this active transport, diffusive movement of DYNC1H1 polysomes was five times slower than that of Ki67 and POLR2A polysomes (0.01 vs. 0.05 µm²/s; Fig. 7 D and Fig. S2 A). In addition, calculation of diffusion coefficients for single DYNC1H1 polysomes revealed that many were nearly immobile for tens of seconds, suggesting that they may be anchored on some cellular structures (Fig. 7 E). Collectively, these results suggested that translating DYNC1H1 mRNPs alternate between cycles of diffusion, anchoring and motor-dependent transport. Because blobs contained at most one molecule of tagged mRNA, it was not possible to determine whether specific movements were associated with mRNAs localized in blobs or occurring as single molecules. Nevertheless, the peculiar localization of DYNC1H1 mRNAs correlated with specific movement properties of these polysomes.

Discussion

The SunTag as a tool to monitor translation of single endogenous mRNPs

In this study, we show that by labeling a peptide epitope right after its synthesis, the SunTag system allows us to visualize nascent proteins that are still being translated by the ribosome. The SunTag thus provides a convenient assay to monitor translation of single mRNPs in living cells. The use of a larger SunTag repeat, as we did in this study, facilitates visualization of newly translated proteins. Likewise, fusing it to a long reporter protein increases the signal of nascent proteins. However, these improvements are not absolutely necessary, as we show that it is possible to visualize translation sites using a short hygromycin protein reporter (Fig. S5 D), as well as using shorter SunTag repeats (Fig. 6). We demonstrate that combining the SunTag with genome editing allows us to visualize translation of endogenous mRNAs and thus in a context that keeps all the regulatory sequences that control mRNA transport and localization. We thus believe that this technology holds great promise to visualize and measure translation in live cells and at the level of single mRNPs. It will provide a unique view of this fundamental process, and, in particular, it should be very informative to analyze the spatial and temporal regulation of translation, as well as to better understand the cell-to-cell variability of gene expression. Although this manuscript was under consideration, other studies showed the use of the SunTag to monitor translation in live cells (Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). Our work further extends these studies by showing that combining the SunTag with genome editing allows us to image translation of endogenous mRNAs.

Translation of single mRNA alternates between active and inactive states

Translation of the SunTagx56-Ki67 reporter appears discontinuous: single mRNAs appear to switch between translated and untranslated states. The reasons for this phenomenon are at present unclear. It could be that these switches are regulated. Alternatively, they may be related to the stochastic behavior often observed at the level of single molecules (Sanchez and Golding, 2013). Much evidence has shown that during transcription initiation, the stochastic binding of transcription factors on a single promoter DNA molecule often results in stochastic switching of promoters between active and inactive states (Sanchez and Golding, 2013). Likewise, it may be possible that single mRNAs randomly switch between translationally competent and incompetent states. Translation initiation requires the assembly of a large macromolecular complex comprising many initiation factors, and this complex may have a finite lifetime and may spontaneously dissociate within the cell, thereby turning off the mRNPs until the initiation complex reassembles. Interestingly, we observed that nearly all of the POLR2A mRNAs are actively engaged in translation. This mRNA codes for an abundant housekeeping protein and may thus be an optimal substrate for translation. This suggests that the switch between active and inactive states may be somehow regulated by the mRNA sequence. In the future, it will be interesting to analyze mRNAs regulated at the translational level.

Polysomes diffuse rapidly through the cytoplasm of human cells

The ability to visualize polysomes in live cells allowed us to compare the mobility of translated and untranslated mRNAs.
Polysomes were tracked using the scFv, whereas untranslated mRNAs were tracked using nls-MCP-tagRFP. We found that translation had little effect on mRNA mobility, which may be because the diffusion coefficient varies with the volume of the particle and thus with the cubic root of its mass. An eightfold increase in the mass thus translates into only a twofold decrease in diffusion coefficient. If one assumes that ribosomes are twice as compact as a random coil mRNA of similar mass and that the ribosomal density is 1.3/kb, then translation is expected to decrease diffusion by 1.3-fold. In contrast to the small effect of translation, diffusion coefficients of individual polysomes varied over almost an order of magnitude, even when coding for the same protein. The reasons for this heterogeneity are at present unclear. It could be due to the location of the mRNP in the cell, the cellular components that associate with it, or other features of the cellular environment.

A fraction of dynein heavy chain is synthesized in translation factories

The SunTagx56-Ki67 and the endogenous POLR2A mRNAs occur as isolated single molecules that are dispersed throughout the cytoplasm. In contrast, 20–50% of the DYNC1H1 mRNAs concentrate in blobs that contain multiple mRNAs (a mean of three to seven). We show that translation DYNC1H1 mRNAs occur on both isolated molecules and blobs, although it appears slightly more frequent and efficient in blobs. The translation of DYNC1H1 mRNAs in blobs indicate that the blobs represent specialized translation factories. Such structures have been hypothesized before (Chang et al., 2006), but the use of the SunTag allows to provide a direct proof of their existence. The function of such factories is currently elusive. One possibility would be that they help produce the mature protein complex. Dynains are large macromolecular complexes of ~2 MD, which also require dedicated factors for their assembly (Carter et al., 2016). They are composed of two heavy chains, two intermediate chains, two light intermediate chains, and multiple light chains. The dynein heavy chain itself is a large polypeptide that contains several structural domains, including an AAA+ motor domain. Folding of dynein heavy chain and assembly of the dynein complex may thus be favored by the concentration of the nascent chains in dedicated factories that may contain the required chaperone and assembly factors. It is, however, interesting to note that the mRNAs coding for other dynein subunits do not localize in the DYNC1H1 blobs (unpublished data). Future studies will be thus required to understand the functions of the DYNC1H1 translation factories.

Large-scale RNA localization studies in Drosophila have revealed that numerous mRNAs accumulate in cytoplasmic foci (Lécuyer et al., 2007; Wilk et al., 2016). In the future, it will be interesting to determine whether these correspond to translation factories or to sites of accumulation of untranslanted mRNA, as in the case of P-bodies. It would also be interesting to determine whether the accumulation of mRNAs in foci is linked to a specific transport pathway that may help to deliver them when and where they are needed. Motor-dependent movements of foci containing untranslated mRNAs may help to transport them to their future translation site, as proposed for instance in the case of neuronal P-bodies (Cougot et al., 2008; Zeitelhofer et al., 2008). In this study, we show that polysomes can also be actively transported, and an interesting speculation would be that motor-dependent movements of dynein heavy chain polysomes may help to deliver newly synthesized dyneins to their site of action. Another possibility could be that this active transport may be required to counterbalance the particularly slow diffusion of this mRNP (0.008 µm/s, or 50 h to explore a cellular area of 1,600 µm²; Fig. S2 A; Fisher and Cooper, 1967). Although future studies will be needed to explore the function of motor-dependent movements of polysomes, this phenomenon reveals a new facet of RNA metabolism.

Materials and methods

Cells

HeLa cells were maintained in DMEM supplemented with 10% FBS, 10 U/ml penicillin/streptomycin and 2.9 mg/ml glutamine in a humidified CO₂ incubator at 37°C. Cells were transfected with JetPrime (Polyplus) and selected on 150 µg/ml hygromycin or 0.25 µg/ml puromycin. For each stable cell line, several individual clones were picked and screened by smFISH with sets of fluorescent oligonucleotide probes against the integrated sequence. For CRISPR recombination, clones were additionally analyzed by genomic PCR using amplicons specific for either the nonrecombined or the recombined allele.

The scFv-GBI-sfGFP and scFv-GBI-sfGFP-NLS (plasmids 60907 and 60906; Addgene; Tanenbaum et al., 2014) are referred in the text as scFv-sfGFP. These plasmids were introduced into cells by retroviral infection. HEK293T cells were transiently transfected with a cocktail of plasmids coding for retroviral components and producing the genomic scFv-sfGFP retroviral RNAs. Viral particles were collected and used to infect recipient HeLa cells, which were then sorted by FACS. Only lowly expressing cells were selected. The phase-Ubc-nls-MCP-tagRFP was introduced into selected cells by retroviral infection following the same procedure. For translational inhibition, cells were treated with puromycin at 100 µg/ml for 30 min. Cells were treated with nocodazole, cytochalasin D, and Torin1 at 10 µg/ml for 1 h, 5 µM for 1 h, and 250 nM for 2 h, respectively.

Plasmids

Sequences of the plasmids are available upon request. The arrays of MS2 stem-loop consisted of a repeat of 132 MS2 binding sites that was generated by gene-synthesis techniques. It is composed of 33 unique stem-loops that were designed to minimize their similarities, and this sequence was multimerized four times. The size of the tag is 2.9 kb, and its sequence is listed in Table S1.

The repeat of 56 SunTag sequence was composed of the original SunTagx24 (plasmid 60910; Addgene; Tanenbaum et al., 2014), which was fused to a new SunTagx32 sequence. This new repeat was designed to minimize the similarities between individual repeat at the nucleotide level while keeping the same protein sequence. The SunTagx32 sequence was cloned by gene-synthesis techniques. The constructs used in this study were assembled using a mixture of gene synthesis, Gateway reactions, Gibson assembly, and traditional cloning techniques. The sequences are available in Table S1.

The nls-MCP-TagRFP plasmid expresses a nuclear version of MCP that fluoresces in red, and it was generated from pHAGE-Ubc-nls-HA-MCP-YFP (plasmid 31320; Addgene) by replacing GFP with TagRFP by traditional cloning techniques.

Genomic PCR

Genomic DNA was prepared with GenElute Mammalian Genomic DNA Miniprep (Sigma-Aldrich) and analyzed by PCR with Platinum Taq DNA Polymerase (Invitrogen). The sequences of oligonucleotides were as follows: POLR2A-wt forward, 5'-TTTACCACGCATCTGGCTC-3'; and POLR2A-wt reverse, 5'-TGCTCTTGTGAGGATTGGTCC-3'; Puro-forward, 5'-GGTGACCGCTGCTAGTGTG-3';
DYNC1H1-wt forward, 5’-GGTAGCTGTTCTCACATGGT-3’ and DYNC1H1-wt reverse, 5’-CTTACACAGGTAGGTCTGCG-3’; and SunTag reverse, 5’-TACCCCTTCTGACTGGGC-3’.

Heterozygous clones were further characterized by microscopy. The SunTagx56-POLR2A proteins localized in the cytoplasm and were thus nonfunctional. In contrast, live imaging of the SunTags32-DYNC1H1 cell line showed that single molecules of protein displayed rapid rectilinear motions over long distances, suggesting that the tagged dyneine subunit was fully functional.

In situ hybridization

Cells were grown on glass coverslips (0.17 mm), washed with PBS, fixed in 4% PFA for 20 min, and permeabilized in 70% ethanol overnight at 4°C. Cells were hybridized as previously described (Fusco et al., 2003), except that the probes were sets of unlabeled oligonucleotides hybridizing against the target RNA and additionally contained a common supplementary sequence that was preannealed to a fluorescent oligonucleotide probe termed the FLAP (Tsanov et al., 2016). We used sets of 24 oligonucleotides to detect the POLR2A mRNA, 36 for the DYNC1H1 mRNA (60 for the triple-staining experiment), 36 for the SunTag-Puro, and 48 for the SunTag-Hygro sequences. Sequences of the probes are available in Table S1. The GFP probes hybridized to the GFP-ires-neo sequences of the LAP tag present in the BAC (Poser et al., 2008) and were a set of 40 oligonucleotides, each 40 nt long and conjugated to three or four Cy3 molecules (Femino et al., 1998). Slides were mounted in Vectashield with DAPI (Vector Laboratories).

Image acquisition in fixed cells

Fixed cells were imaged at room temperature on an Axioimager Z1 wide-field microscope (63×, NA 1.4; ZEISS) equipped with an sCMOS Zyla 4.0 0.2 camera (Andor Technology) and controlled by MetaMorph (Universal Imaging). 3D image stacks were collected with a Z-spacing of 0.3 μm. Figures were prepared with ImageJ (National Institutes of Health), Photoshop (Adobe Systems), and Illustrator (Adobe Systems), and graphs were generated with R.

P-bodies were labeled with a mouse monoclonal antibody raised against S6K and that is known to recognize GE-1/Helds (sc-8418; Santa Cruz Biotechnology, Inc.; Kedersha and Anderson, 2007). The secondary anti-mouse antibody was coupled to FITC. Slides were mounted in Vectashield with DAPI (Vector Laboratories).

Image acquisition in live cells

Cells were plated on 25-mm-diameter coverslips (0.17-mm thick) in nonfluorescent media (DMEM gfp-2 of Evrogen complemented with rutin and 10% FCS). Coverslips were mounted in a temperature-controlled chamber (37°C) with 5% CO2 and imaged on an inverted OMMX3 Deltavision microscope (GE Healthcare) in time-lapse mode. A 100×, NA 1.4 objective was used, with an intermediate 2× lens and an Evolve 512 × 512 EMCCD camera (Photometrics). For fast imaging, we imaged at a rate of 2.2 stacks/s for 2 min, using stacks of 10 planes with a Z-spacing of 0.6 μm. This spacing allowed accurate point spread function determination without excessive oversampling. For slow imaging, we collected stacks of 14 planes with a Z-spacing of 0.6 μm and at a frame rate of one stack every 10 s. The power of illuminating light and the exposure time were set to the lowest values that still allowed visualization of single molecules of proteins (laser at 1% of full power, exposure of 10 ms/plane). This minimized bleaching and maximized the number of frames that was collected.

FRAP

FRAP was performed on a confocal microscope (Meta LSM780; ZEISS) with a 63×, 1.4 NA objective, at 37°C and in DMEM gfp2 complemented with rutin and 10% FCS. The translation foci labeled by the scFv-sfGFP were bleached at 488 nm in a circle of 1.3-μm diameter at full laser power for 300 ms. Recoveries were measured by making stacks in 3D at a rate of one stack every 2 s (one stack was made of five slices in Z, 0.4 μm apart). Recoveries were analyzed by defining a tracking area in the 4D image stack and by measuring the total intensity of a cube of 3 × 3 pixels centered on the brightest pixel in the tracking area (Boireau et al., 2007). Background was removed, intensities at each time point were corrected for bleaching by dividing them by the total cell fluorescence, and these values were normalized to the fluorescence intensity before the bleach. Finally, we eliminated the first 15 s after recovery to remove the contribution of diffusing scFv-sfGFP molecules, either free or labeling single molecules of tagged proteins.

Ribosome runoff experiments

Live-cell videos were acquired at a rate of one image every 10 s on an OMMX3 Deltavision (GE Healthcare; slow imaging conditions; see Image acquisition in live cells). Single SunTagsx56-Ki67 polysomes were then tracked using the Fiji plugin TrackMate, and their intensity over time was quantitated (see Analysis of live-cell videos). Polysomes that turned off during the video were identified visually, and their tracks were aligned by the last image in which the polysome was visible. The intensities of a polysome over time were then normalized by dividing them by the median intensity of the polysome over the track. The mean normalized intensities of different polysomes over time were then fit to the model described in the section Modeling of the FRAP and ribosome runoff experiments.

Modeling of the FRAP and ribosome runoff experiments

Fit of the recovery curve was done with a linear model that assumes that ribosomes are distributed homogeneously along the mRNA and move at a constant speed, as done previously for modeling transcription (Boireau et al., 2007). The model further assumes that proteins diffuse away immediately after completion of their synthesis. The ribosome runoff and the FRAP experiments yield curves that are symmetric with respect to time, and the calculations are thus presented only for the ribosome runoff experiments. The fit yields three parameters: the intensity before translation turns off (I0), the time at which translation turns off (t0), and the elongation rate (v). The curve is composed of two parts. In the first part, ribosomes are located on both the SunTag repeat and the downstream coding region, whereas in the second part, the last ribosome has passed the SunTag, and ribosomes are all located on the downstream coding region. During this second period, there is no synthesis of SunTag repetition while ribosomes terminate translation at a constant rate. This part thus translates into a linear segment, for which the slope depends on the elongation rate. During the first period, ribosomes similarly terminate transcription, but at the same time, ribosomes located on the SunTag repeat generate a new signal, which is proportional to the number of ribosomes on the repeat. In turn, this number decreases linearly with time because ribosomes move to the 3’ end at a constant rate. Assuming that the length of the SunTag repeat is L1 and that the length of the downstream region is L2, then a formalization of the model yields the following equation for I(t), the intensity at time t, when t < t0, then I(t) = 0, when 0 < t < L1/v, then

$$I(t) = \frac{I_0}{L_1 + 2 \times L_2} \cdot \left(1 - \frac{v \times (t - t_0)}{L_1}\right)^2 \times \left(L_1 + 2 \times L_2 \right);$$

when L1/v < t < (L1+L2)/v, then

$$I(t) = \frac{I_0}{L_1 + 2 \times L_2} \cdot \left(1 - \frac{v \times (t - L_1/v)}{L_1}\right) \times \frac{2 \times L_2}{L_1 + 2 \times L_2};$$

and when (L1+L2)/v < t, then I(t) = 0, with t′ = t - t0.
Analysis of SMFISH images

The numbers of mRNA per cell were counted manually for Figs. 2 and 5. To quantify single mRNA molecules and mRNA blobs in Fig. 6, we used an automated pipeline. Cell segmentation was performed with CellCognition (Held et al., 2010), using DAPI for nuclear segmentation and the SMFISH images for cellular segmentation with a watershed method. Fluorescent spots were detected with FISH-quant (Mueller et al., 2013), with the connected components methods after manual thresholding of Laplacian of Gaussian filtered image. To separate individual mRNA molecules and blob, a threshold on their raw intensity was applied. This threshold was defined as 1.5, the median intensity of all the detected spots. This value was determined by comparing DYNC1H1 images with images obtained from an mRNA that does not make blobs (e.g., CRM1) and by selecting the value that maximized the differences in the number of blobs between the two genes. Blobs inside the nucleus were excluded to avoid transcription sites in the analysis.

To quantify the number of mRNAs per blob, we first calculated the intensities of single mRNA molecules. Images around individual mRNAs were cropped and used to calculate a mean image of all individual mRNAs. The resulting mean image was fit with a 3D Gaussian function and the integrated intensity above background estimated. Each individual mRNA blob was analyzed in a similar fashion (cropping and Gaussian fitting). To infer how many individual mRNA molecules are present in a particular blob, we divided its integrated intensity by the integrated intensity of the mean individual mRNA molecules.

Analysis of SunTag images in fixed cells

The percentage of translated mRNA per cell were counted manually for Figs. 2 and 4, by counting mRNAs molecules that colocalized with a SunTag foci. We used a semiautomated pipeline to quantify the number of nascent protein chains per translating mRNAs. Translation foci and individual proteins spots labeled by the SunTag were defined manually in 3D stacks in a homemade Matlab interface. Images of individual proteins were cropped ±3 pixels in XY and ±1 pixel in Z, with respect to the pixel with maximum intensity. An averaged image of the individual proteins was then calculated using at least 15 individual proteins in the vicinity of a translation foci. The image of each translation foci was cropped to have similar dimensions, and a maximum intensity projection along Z was performed. The resulting images were fit with a 2D Gaussian function, and the integrated intensity above background was calculated. The number of proteins per translation foci was estimated as the ratio of the integrated intensity of the translation foci to the intensity of the averaged image of single proteins.

To calculate the density of ribosomes, we divided the number of nascent protein per mRNA by $L_{norm}$, the normalized cDNA length. If the length of the SunTag coding region is $L_1$, and the length of the coding region after the SunTag is $L_2$, $L_{norm} = L_1 \times 0.5 + L_2$. The reason is that ribosomes located on the SunTag repeat are less bright than those located after the repeat, because only a fraction of the repeat has been translated. If we assume that ribosomes are uniformly distributed along the length of the coding region, then the mean location of the ribosomes that are still on the SunTag repeat is at half the length of the repeat. This means that their mean brightness is half of the brightness of the ribosomes located after the SunTag. A formal calculation has been done in Boireau et al. (2007) in the case of RNA polymerases transcribing an MS2 repeat, but can be equally applied to ribosomes translating a SunTag repeat. The length, in nucleotides, of the various coding sequences is: 2,380 for the SunTagx32; 4,123 for the SunTagx56; 1,023 for hygromycin; 660 for puromycin; 9,531 for Ki67; 5,910 for POLR2A; and 13,938 for DYNC1H1.

Analysis of live-cell videos

Videos were projected along Z using pixels of maximal intensity and corrected for photobleaching using the histogram matching method in ImageJ (National Institutes of Health). Individual spots were identified and tracked with the TrackMate plugin in ImageJ, using the DoG detector, subpixel localization, a blob diameter of 0.8 µm, and a threshold manually adapted for each video. For fast imaging conditions, tracks were reconstructed using the simple LAP tracker option, using a maximal linking distance of 1.5 µm, a gap-closing distance of 1 µm, and a maximal frame gap of 2.

Tracks were imported and analyzed in R. Instant 1D displacements between frames were calculated along the x and y axis, and the resulting histograms were fitted to a Gaussian function, for which variance is directly proportional to the diffusion coefficient ($D$). The values obtained for the x and y displacements were treated independently and within 1% for the SunTagx56-Ki67 and SunTag-POLR2A polysomes and within 7% for the SunTagx32-DYNC1H1. We also calculated a mean MSD as a function of time, by aligning all tracks at their start and averaging the resulting 2D displacements. The segment of the resulting curves comprised between 0 and 18 s was fit to a linear model, and the slope was used to calculate $D$. These values were approximately twofold lower than the ones calculated from instant displacements, which may be a result of errors in pointing accuracy for the instant displacements (mean instant displacement was between 100 and 300 nm) or to a subdiffusive behavior for the particles. We also analyzed single individual tracks, for tracks longer than 36 s. For each track, we calculated all instant 1D displacements in x and y, pooled them, and estimated $D$ for each polysome from the variance of the resulting data (>150 data points). We then calculated the SD of individual diffusion coefficients. If all particles were diffusing identically, they should have the same D, and the measured SD between individual polysomes should be $D_{Dnbp}$, with $D_{nbp}$ being the number of data points used to calculate $D$ for each track (in this study, 150).

For the three polysomes analyzed in this study (Ki67, POLR2A, and DYNC1H1), the measured SD was 5–10-fold higher than expected, indicating that particles are heterogeneous in terms of diffusional properties.

To quantify the fluctuation of polysomes intensity over time, the tracking data were used to locate the particles in the videos. The videos were bleach corrected, projected along Z using the pixels of maximal intensities, and cropped around the spot location at each time point. The resulting images were then fit to 2D Gaussian, and particle intensity was measured from the integrated intensity above background.

Online supplemental material

Fig. S1 shows colocalization of the bright protein foci labeled by the scFv-sfGFP with the SunTagx56-Ki67 mRNAs. Fig. S2 shows diffusion of polysomes and correlation with the number of nascent proteins. Fig. S3 shows characterization of the SunTagx56-POLR2A polysomes. Fig. S4 shows characterization of DYNC1H1 RNA blobs. Fig. S5 shows characterization of DYNC1H1 polysomes. Table S1 contains the sequence of the oligo probes used for the SMFISH experiments, as well as the sequence of the SunTagx32, SunTagx56, and MS2×132 tags. Video 1 shows fast imaging of the SunTagx56-Ki67 reporter (2.2 stacks/s; each Z stack with 10 slices 0.6 µm apart), accelerated four times. Video 2 shows slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 160 times. Video 3 shows slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 160 times. Video 4 shows slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 80 times. Video 5 shows slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 80 times. Video 6 shows fast imaging of the SunTagx56-Ki67 reporter (2.2 stacks/s; each Z stack with 10 slices 0.6 µm apart), accelerated four times. Video 7 shows fast imaging of the SunTagx56-POLR2A gene (2.2 stacks/s), accelerated four times. Video 8 shows fast imaging of the SunTagx32-DYNC1H1 gene (2.2 stacks/s), accelerated four times. Video 9 shows the same as Video 8, except that the cells were treated with cytochalasin D (5 µM for 1 h). Online supplemental material is available at http://www.jcb.org/cgi/content/full/jcb.201605024/D1.
Acknowledgments

We thank D. Fisher and A. Camas for the gift of a Gateway mouse Kι67 donor vector, R. Yole for the gift of the SunTag plasmids, I. Poser and T. Hyman for the gift of the DYNC1H1 BAC-GFP cell line, and the Montpellier RIO Imaging facility for help with the imaging experiments.

A. Safieddine is supported by a Ministère de l’Éducation Nationale de la Recherche et de Technologie fellowship. This work was supported by the Site de Recherche Intégrée sur le Cancer (SIRIC) Montpellier Cancer (grant INCaDGOS-INSEMR 6045) and Fondation de la Recherche Medicale and Agence Nationale de la Recherche Hi-FISH grants to E. Basyuk.

The authors declare no competing financial interests.

Submitted: 7 May 2016
Accepted: 16 August 2016

References


Figure S1. Colocalization of the bright protein foci labeled by the scFv-sfGFP with the SunTagx56-Ki67 mRNAs. Panels represent microscopy images of cells stably expressing the scFv-sfGFP and the SunTagx56-Ki67 mRNA and labeled with smFISH probes recognizing the SunTagHygro (B) sequences. (C) Color overlay of the smFISH image (red, A) and the SunTag signal (green, B). Bars, 45 µm. Signal quality can be appreciated by zooming in the images.
Figure S2. **Diffusion of polysomes and correlation with the number of nascent proteins.** (A) Graph depicts a linear fit to MSD for the three polysomes analyzed (Ki67, POLR2A, and DYNC1H1). Only the first 20 s of the curves are taken into account. (B) Graph depicts the diffusion coefficient of individual DYNC1H1 polysomes as a function of their brightness. The diffusion coefficient of individual polysome is obtained from the variance of the single-step jumps of this particle over time. 
Figure S3. **Characterization of the SunTags56-POLR2A polysomes.** (A) Effect of puromycin on the brighter protein foci labeled by the scFv-sfGFP. Panels represent microscopy images of cells stably expressing the scFv-sfGFP containing the SunTags56-POLR2A allele, untreated (left panel) or treated with puromycin (right panel). The blue arrows indicate a spot corresponding to a single molecule of SunTags56-Ki67 protein, whereas the red arrows point to a brighter protein foci. Inset: a zoom of the boxed area (12 × 12 µm). Bar, 12 µm. (B) Colocalization of the bright protein foci labeled by the scFv-sfGFP with endogenous POLR2A mRNAs. Panels represent microscopy images of cells stably expressing the scFv-sfGFP containing the SunTags56-POLR2A allele and labeled with smFISH probes recognizing the POLR2A sequences. Note that these probes recognize both tagged and untagged mRNA, as cells are heterozygous and thus yield less colocalization than in Fig. 3 B, in which only the tagged mRNA are labeled. (right) Color overlay of the smFISH image (red, left panel) and the SunTag signal (green, middle panel). The red arrows point to a brighter protein foci that colocalizes with the mRNA. [inset] A zoom of the boxed area (4 × 4 µm). Bar, 4 µm.
Figure S4. Characterization of DYNC1H1 RNA blobs. (A) Simultaneous localization of the stress granule marker G3BP and DYNC1H1 RNA blobs in HeLa cells. Image depicts fluorescent microscopy signals of cells stained for the stress granule marker G3BP (green), DYNC1H1 mRNA (red), and DAPI (blue). Red arrow, RNA blob. Bar, 7 µm. (B) Simultaneous localization of P-bodies and DYNC1H1 RNA blobs in HeLa cells expressing the DYNC1H1 BAC. Image depicts fluorescent microscopy signals of cells stained for the P-body marker GE-1/helds (green), DYNC1H1 mRNA (red), and DAPI (blue). Bar, 7 µm.
Figure S5. Characterization of DYNC1H1 polysomes. (A) Effect of puromycin on the brighter protein foci labeled by the scFv-sGFP. Panels represent microscopy images of cells stably expressing the scFv-sGFP and containing the SunTagx32-DYNC1H1 allele, untreated (left) or treated with puromycin (right). The blue arrows indicate a spot corresponding to a single molecule of SunTagx32-DYNC1H1 protein, whereas the red arrows point to a brighter protein foci. Inset: a zoom of the boxed area (8 × 8 µm). Bar, 8 µm. (B) Single-molecule polysome profile of DYNC1H1 mRNAs. Panels represent histograms of the number of nascent protein per translated DYNC1H1 mRNA (n = 120 mRNAs). (C) Quantification of the diffusion rates of SunTagx32-DYNC1H1 polysomes. The graph is a histogram of 1D displacements measured between two consecutive video frames (176 particles total). (D) Colocalization of the bright protein foci labeled by the scFv-sGFP with SunTagx56-Hygro mRNAs. Panels represent microscopy images of cells stably expressing the scFv-sGFP and the SunTagx56-Hygro reporter, labeled with smFISH probes recognizing the SunTagx56-hygro sequences. This reporter is identical to the SunTagx56-Ki67, except that it lacks the Ki67 coding sequence. The green dashed outlines represent the position of the nucleus. The red arrows point to a bright protein foci that colocalize with the mRNA. Bar, 5 µm.
Video 1. Fast imaging of the SunTagx56-Ki67 reporter (2.2 stacks/s; each Z-stack with 10 slices 0.6 µm apart), accelerated four times. HeLa cells stably expressing SunTagx56-Ki67, scFv-sfGFP (green), and nls-MCP-TagRFPt (red) were imaged in two colors by epifluorescence microscopy. Images are maximal intensity projection along Z and corrected for photobleaching using histogram matching. Time is indicated as minutes:seconds.

Video 2. Slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 160 times. HeLa cells stably expressing the SunTagx56-Ki67 reporter and scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video corresponds to the SunTag images of Fig. 2 A. Bar, 5 µm.

Video 3. Slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 160 times. HeLa cells stably expressing the SunTagx56-Ki67 reporter and scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video is a zoom of Video 1 (frames 4–92) and corresponds to the violet trace of Fig. 3 B. The top panel is a maximal intensity projection along z (providing a top view of the cell), and the bottom panel is a maximal intensity projection along y (providing a side view of the cell). Bar, 2 µm.

Video 4. Slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 80 times. HeLa cells stably expressing the SunTagx56-Ki67 reporter and scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video is a zoom of Video 1 and corresponds to the green trace of Fig. 3 B. The top panel is a maximal intensity projection along z (providing a top view of the cell), and the bottom panel is a maximal intensity projection along y (providing a side view of the cell). Bar, 2 µm.

Video 5. Slow imaging of the SunTagx56-Ki67 reporter (1 stack per 10 s), accelerated 80 times. HeLa cells stably expressing the SunTagx56-Ki67 reporter and scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video is a zoom of Video 1 and corresponds to the end of the brown trace of Fig. 3 B. The top panel is a maximal intensity projection along z (providing a top view of the cell), and the bottom panel is a maximal intensity projection along y (providing a side view of the cell). Bar, 2 µm.

Video 6. Fast imaging of the SunTagx56-Ki67 reporter (2.2 stacks/s), accelerated four times. HeLa cells stably expressing the SunTagx56-Ki67 reporter and scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video corresponds to the SunTag images of Fig. 4 A. Bar, 5 µm.
Video 7. **Fast imaging of the SunTagx56-POLR2A gene (2.2 stacks/s), accelerated four times.** Hela cells stably expressing the SunTagx56-POLR2A allele and the scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video corresponds to the SunTag images of Fig. 5 E. Bar, 5 µm.

Video 8. **Fast imaging of the SunTagx32-DYNC1H1 gene (2.2 stacks/s), accelerated four times.** Hela cells stably expressing the SunTagx32-DYNC1H1 allele and the scFv-sfGFP were imaged in a single color by epifluorescence microscopy. The video corresponds to the SunTag images of Fig. 7 A. Bar, 5 µm.

Video 9. **Fast imaging of the SunTagx32-DYNC1H1 gene (2.2 stacks/s), accelerated four times.** Legend as for Video 8, except that the cells were treated with cytochalasin D (5 µM for 1 h). Bar, 5 µm.

Provided online is Table S1, providing the sequence of the oligo probes used for the smFISH experiments, as well as the sequence of the SunTagx32, SunTagx56, and MS2x132 tags. For the smFISH oligonucleotide probes, the pool of unlabeled oligonucleotide is hybridized with the fluorescent FLAP, and the resulting hybrid is used as a probe instead of a traditional fluorescent oligonucleotide.
A localization screen reveals translation factories and widespread co-translational protein targeting

This paper has the following aims:

- Perform a dual mRNA/protein localization screen on randomly selected genes, as well as genes encoding all motors using a BAComics approach in which each cell line contains a GFP-tagged version of a gene.

- Identify, describe, and quantify features of mRNA localization, as well as cases of local translation by analyzing protein/mRNA co-localization and visualizing polysomes.

- Assess whether localized transcripts depend on RNA or protein signals, and to what extent.

- Perform functional studies on one localized candidate to understand the biological function of localizing its mRNA.

Main conclusions of the paper were:

- Co-translational protein targeting is widespread.

- A variety of mRNAs are translated in translation factories.

- In the case of β-catenin, translation factories specialize in co-translational degradation of the nascent protein.

In this study, I specifically contributed by:

- Screening more than 60 mRNAs using smFISH. In particular, I identified an mRNA that localizes to centrosomes (HMMR), which drove my interest to this particular RNA localization class.
- Confirming the localization of a portion of BAC-transcribed mRNAs by inspecting and interpreting the localization of endogenous transcripts using smiFISH.

- Acquiring and quality checking a dataset of images for most localized mRNAs that will be used to quantify mRNA localization features at the single molecule and single cell levels based on existing pipelines (see annex, analysis ongoing at the time of writing this thesis).

- Acquiring and quality checking another dataset that will be used to quantify the effects of various drug treatments on mRNA distributions and co-localization with proteins.

- Describing and interpreting complex patterns of mRNA localization for three genes (ASPM, HMMR, and NUMA1), including localization at the nuclear pore and at centrosomes.

- Designing, cloning and generating an ASPM SunTag clone. Observing and interpreting images obtained with this clone, allowed me to demonstrate that: (i) aggregating ASPM mRNAs correspond to translation factories, and (ii) ASPM mRNAs are locally translated at the nuclear pore and remain there for about 20 mins.

- Sorting and maintaining many of the cell lines used throughout this study.

- Quantifying images, preparing figure, commenting on, and correcting the text.

This paper is currently being revised for *Cell*.

*Supplementary movies and tables for this paper can be downloaded from*

https://filesender.renater.fr/?s=download&token=acd1c19e-ae8f-f5d3-3021-c63b648d9950
A localization screen reveals translation factories and widespread co-translational protein targeting

Racha Chouaib\textsuperscript{1,2,3,11}, Adham Safieddine\textsuperscript{1,2,3}, Xavier Pichon\textsuperscript{1,2}, Oh Sung Kwon\textsuperscript{4}, Aubin Samacoits\textsuperscript{5,6}, Abdel-Meneem Traboulsi\textsuperscript{1,2}, Nikolay Tsanov\textsuperscript{1,2}, Marie-Cécile Robert\textsuperscript{1,2}, Ina Poser\textsuperscript{7}, Christophe Zimmer\textsuperscript{5,6}, Anthony Hyman\textsuperscript{7}, Hervé Le Hir\textsuperscript{6}, Thomas Walter\textsuperscript{8,9,10}, Kazem Zibara\textsuperscript{3,11}, Florian Mueller\textsuperscript{5,6}, Marion Peter\textsuperscript{1,2}, Edouard Bertrand\textsuperscript{1,2,*}

\textsuperscript{1}Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France
\textsuperscript{2}Équipe labélisée Ligue Nationale Contre le Cancer, University of Montpellier, CNRS, Montpellier, France
\textsuperscript{3}ER045, PRASE, DSST, Lebanese University, Beirut, Lebanon
\textsuperscript{4}Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Research University, 46 rue d'Ulm, 75005, Paris, France
\textsuperscript{5}Unité Imagerie et Modélisation, Institut Pasteur and CNRS UMR 3691, 28 rue du Docteur Roux, 75015 Paris, France
\textsuperscript{6}C3BI, USR 3756 IP CNRS – Paris, France
\textsuperscript{7}MPI-CBG, Pfotenhauer Str. 108, 01307 Dresden, Germany
\textsuperscript{8}MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 77300 Fontainebleau, France
\textsuperscript{9}Institut Curie, 75248 Paris Cedex, France
\textsuperscript{10}INSERM, U900, 75248 Paris Cedex, France
\textsuperscript{11}Biology Department, Faculty of Sciences-I, Lebanese University, Beirut, Lebanon

+ Equal contributions

**Running title:** mRNA localization and local translation

**Keywords:** RNA localization, local translation, RNA transport, smFISH, translation factories, cotranslational targeting

*To whom correspondence should be addressed. Edouard Bertrand: Tel. +33 434359647, Email edouard.bertrand@igmm.cnrs.fr; Lead contact*
Summary

Local translation allows a spatial control of gene expression. Here, we performed a dual protein/mRNA localization screen, using smFISH on 521 human cell lines expressing GFP-tagged genes. A total of 32 mRNAs displayed specific cytoplasmic localizations, and we observed local translation at unexpected locations, including cytoplasmic protrusions, cell edges, endosomes, Golgi, the nuclear envelope and centrosomes, the latter being cell-cycle dependent. Surprisingly, mRNA localization frequently required ongoing translation, indicating widespread co-translational targeting. Nevertheless, the kinesin KIF1C localized mRNAs to cytoplasmic protrusions in a translation-independent manner. Interestingly, while P-body accumulation was frequent (15 mRNAs), four mRNAs accumulated in foci that were distinct structures. These foci lacked the mature protein, but nascent polypeptide imaging showed that they were specialized translation factories. For β-catenin, foci formation was regulated by Wnt, relied on APC-dependent polysome aggregation, and led to nascent protein degradation. Thus, translation factories uniquely regulate nascent protein metabolism and create a fine granular compartmentalization of translation.
Introduction

Most mRNAs are distributed randomly throughout the cytoplasm, but some localize to specific subcellular areas (Blower, 2013; Bovaird et al., 2018; Eliscovich and Singer, 2017; Jung et al., 2014) for reviews). This phenomenon is linked to either RNA metabolism, when untranslated mRNAs are stored in P-bodies or other cellular structures (Hubstenberger et al., 2017); or to protein metabolism, when a protein is synthesized locally. Local translation has been observed from bacteria and yeast to humans (Blower, 2013; Bovaird et al., 2018; Eliscovich and Singer, 2017; Jung et al., 2014). It is commonly involved in the delivery of mature proteins to specific cellular compartments, and this is involved in many processes. For instance, it contributes to patterning and cell fate determination during metazoan development, mainly through asymmetric cell division (Driever and Nüsslein-Volhard, 1988; Melton, 1987). In Xenopus embryos, local translation of cyclin B mRNAs at the mitotic spindle is believed to be important for the rapid cell division cycles occurring during early embryogenesis (Groisman et al., 2000). In mammals, RNA localization is involved in cell polarization and motility, mainly through the localization of actin and related mRNAs at the leading edge (Lawrence and Singer, 1986b), and it is also involved in axonal growth and synaptic plasticity of neurons (Van Driesche and Martin, 2018). Importantly, local translation can also be linked to the metabolism of the nascent peptide rather than to directly localize the mature protein. For instance, translation of secreted proteins at the endoplasmic reticulum (ER) allows nascent proteins to translocate through the membrane to reach the ER lumen (Aviram and Schuldiner, 2017). Translation of mRNAs at specific sites may also be important for the assembly of protein complexes (Pichon et al., 2016), and to avoid the deleterious effects of releasing free proteins at inappropriate places (Müller et al., 2013).
RNA localization can be accomplished through several mechanisms. In the case of secreted proteins, the nascent peptide serves as a targeting signal, via the signal recognition particle (SRP) and its receptor on the ER (Aviram and Schuldiner, 2017). In most other cases, targeting is an RNA-driven process (Blower, 2013; Bovaird et al., 2018; Eliscovich and Singer, 2017; Jung et al., 2014) for reviews). Localized mRNAs often contain a zip-code sequence, frequently located in their 3’-UTR, which is necessary and sufficient to transport them to their destination. The zipcode is recognized by one or several RNA-binding proteins (RBP), and it drives the formation of a transport complex sometimes called locasome. This complex can be transported by centrosomes, endosomal vesicles or other cellular structures (Basyuk et al., 2003; Jansen et al., 2014; Lambert and Nagy, 2002). However, direct transport on the cytoskeleton by molecular motors is a frequent mechanism (Blower, 2013; Bovaird et al., 2018; Eliscovich and Singer, 2017; Jung et al., 2014). Once at destination, an anchoring mechanism may limit diffusion away from the targeted site. Alternatives to these transport mechanisms include diffusion and trapping at specific locations, and degradation coupled to local RNA stabilization.

Localized mRNAs are often subjected to a spatial control of translation (Besse and Ephrussi, 2008). In the case of Ash1 mRNA in yeast and β-actin mRNA in neurons, translation is repressed during transport and is activated at their final location by phosphorylation-dependent mechanisms (Hüttelmaier et al., 2005; Paquin et al., 2007). This spatial regulation of translation provides an additional layer of control ensuring that mRNAs are translated only at the desired location, because mis-localized mRNAs are not translated, as for instance in the case of nanos mRNA in Drosophila embryos (Bergsten and Gavis, 1999).

The first locally-translated mRNAs were found by chance or using a candidate approach. Purification of cellular structures and localized RBPs have significantly increased the number of
known localized mRNAs (Blower, 2013; Bovaird et al., 2018; Eliscovich and Singer, 2017; Jung et al., 2014). However, only specific compartments or RBPs were examined and we currently lack a global view of local translation in the entire cellular space or at the genomic level. Few studies attempted to characterize mRNA localization in a systematic manner. A pioneering study in Drosophila used whole-mount fluorescent in situ hybridization to analyze the localization of more than 2000 mRNAs (Lécuyer et al., 2007). As many as 71% of them had a non-random distribution, and a range of new localization patterns were observed. More recent studies expanded these results and confirmed that RNA localization is widespread during Drosophila development (Jambor et al., 2015; Wilk et al., 2016). However, it is not known whether this is also true in other organisms and particularly in humans. Two recent studies addressed this question in cell lines using the more sensitive single molecule FISH technique (smFISH). Several hundred mRNAs were analyzed, which showed a correlation of intracellular mRNA distribution with gene annotation (Battich et al., 2013; Chen et al., 2015). Specifically, these studies identified three groups of localized mRNAs, at the ER, the mitochondria and the cell periphery, the latter being possibly linked to functions related to actin metabolism.

These studies provided information on RNA localization but they did not directly investigate local translation as the encoded proteins were not detected. Thus, we still lack a good understanding of the various functions played by local translation at the cellular level. In this study, we developed an smFISH screen to specifically address this issue. Using a set of 521 GFP-tagged cell lines spanning a variety of cellular functions and an approach that allows simultaneous visualization of mRNA and proteins, we found that local translation occurs at unanticipated and various locations. In particular, we discovered specialized translation factories, where specific mRNAs are translated. These factories are remarkable in that they provide a
unique mean to regulate the metabolism of nascent proteins and they also create a fine granular compartmentalization of translation.
Results

BAC TransgeneOmics allow dual protein/RNA localization screens

In order to simultaneously visualize mRNAs with their encoded proteins, we based our screen on a library of HeLa cell lines, each containing a bacterial artificial chromosome (BAC) stably integrated in their genome (Poser et al., 2008). Each BAC contains a GFP-tagged gene, which possess all its regulatory sequences (promoter, enhancers, introns, 5’ and 3’ UTRs; Figure 1A). The resulting mRNAs are thus identical to the endogenous molecules in terms of sequence and isoform diversity, except for the tag addition. Previous studies showed that such tagged genes are expressed at near endogenous levels and with the proper spatio-temporal pattern (Poser et al., 2008). Since the tagged mRNAs contain all the regulatory sequences, we hypothesized that they should localize like the endogenous ones, provided that the tag does not interfere with localization. Using BACs presents two advantages. First, a single smFISH probe set against the GFP sequence is sufficient to detect all the studied mRNAs. Second, using mild hybridization conditions, the GFP fluorescence can be detected together with the FISH signal (Fusco et al., 2003), and both mRNA and the encoded protein can thus be detected in the same cell. Genes in the BAC collection are tagged at either ends. N-terminally tagged proteins have only the GFP tag since the selectable marker is placed within an intron, and C-terminally tagged proteins carry a GFP-IRES-Neo tag that allows both visualization and selection (Poser et al., 2008). We developed a set of 19 fluorescent oligonucleotide probes against the GFP tag and an additional 25 probes against the IRES-neo sequence (Figure 1A, Table S1). Following smFISH performed with these probes, no signal was detected in the parental HeLa cells, while isolated spots corresponding to single mRNAs were visible in the TERF1-GFP or GFP-VAPA BAC cells (Figure 1B). The signal was more intense for the C-terminally tagged mRNAs because of the
higher number of probes. Importantly, the GFP signal was visible after smFISH, demonstrating the feasibility of the approach.

**Screening transcripts encoding motors reveals six localized mRNAs and three patterns**

Molecular motors transport cargos to different cellular destinations and they accumulate themselves at various cellular locations. The BAC collection contains all the human genes coding for kinesins and myosins (59 and 44, respectively; (Maliga et al., 2013)), as well as 7 out of 11 dynein subunits. These gene families were screened to test whether some mRNAs would localize to specific cellular areas. We hypothesized that this could reveal a local function of motor proteins, or a mean to transport from/to a particular location. Most motor mRNAs localized randomly throughout the cytoplasm (Table S2). However, 6 displayed atypical localizations, with three distinct patterns (Figure 2 and Tables S3). In the first, referred to as "protrusion", mRNAs accumulated in cytoplasmic extensions. This was observed for three kinesins (KIF1C, KIF4A and KIF5B), for the dynein subunit DYNLL2, and for the myosin MYH3 (Figure 2A). Second, the MYH3 mRNA also accumulated inside the nucleus, a pattern referred to as "intranuclear". In the third pattern, referred to as “foci”, the dynein heavy chain mRNA (DYNC1H1) localized throughout the cytoplasm but aggregated in some bright structures containing several RNA molecules (Figure 2B), as reported recently (Pichon et al., 2016). Surprisingly, the proteins encoded by these localized mRNAs did not appear strongly enriched at the site of mRNA accumulation. GFP-KIF4A was mostly nuclear with only a faint staining at the cell periphery, while GFP-tagged KIF5B, DYNLL2 and DYNC1H1 proteins localized throughout the cytoplasm without a specific enrichment in protrusions or foci (Figure 2). Co-localization was only observed for KIF1C-GFP, where both the mRNA and the GFP-tagged protein accumulated in cytoplasmic protrusions. This suggested that the KIF1C-GFP mRNAs were translated locally.
To confirm the BAC results, endogenous mRNAs were analyzed using smiFISH, an inexpensive variant of smFISH (Tsanov et al., 2016). KIF1C mRNAs accumulated in cytoplasmic protrusions in all the examined cell lines, including human HeLa cells as well as mouse 3T3 fibroblasts and C2C12 myoblasts (Figure S1A). Furthermore, in differentiated SH-SY5Y neuronal cells, KIF1C mRNA localized in dendritic-like outgrowths, away from the cell body, while a control CRM1 mRNA remained in the soma (Figure S1B). Likewise, DYNC1H1 mRNAs formed foci in all examined cell lines (Figure S1A). MYH3 mRNA was not expressed in HeLa, 293HEK and 3T3, but localized in cellular protrusions in C2C12 cells, as well as in dendritic-like processes of SH-SY5Y neuronal cells. In contrast, KIF5B mRNA localization varied with the cell line. These mRNA were frequently present in neuronal dendrite-like processes of SH-SY5Y cells, but localized only weakly in HeLa or C2C12 lines (Figure S1, data not shown).

Interestingly, we observed cell-to-cell variations in RNA localization. Manual annotations indicated that 20 to 40% of cells expressing GFP-tagged KIF1C, KIF4A and MYH3 displayed the expected "protrusion" pattern, while the remaining cells were classified as "random" (see below quantification in Figure 7B). Thus, the localization of these mRNAs shows a high degree of plasticity in cultured cells. Next, we determined whether KIF1C, KIF5B, MYH3 and KIF4A mRNAs accumulated in the same protrusions. Two-color smFISH against the endogenous KIF1C mRNAs and the BAC-tagged mRNAs revealed that this was indeed the case (Figure S2A). Taken together, these results validate our BAC approach for simultaneously screening mRNA and protein localization. They identified six localized mRNAs and three localization patterns, with KIF1C likely being locally translated in cytoplasmic protrusions.

The KIF1C motor localizes mRNAs to cytoplasmic protrusions
High-throughput mRNA-protein cross-linking approaches previously showed that KIF1C directly binds mRNAs (Baltz et al., 2012; Castello et al., 2012). To determine the identity of these mRNAs, we immunoprecipitated (IP) KIF1C-GFP with anti-GFP antibodies or uncoated beads as controls, and identified the co-precipitated RNAs using microarrays (Figure 3A and Table S4). Interestingly, RAB13 mRNA is well known to accumulate in cytoplasmic protrusions (Mili et al., 2008), and it was found in the top 30 enriched species (5.5-fold enrichment), along with the KIF1C mRNA itself (2.6-fold enrichment). The IP/microarray data thus suggests that RAB13 mRNA could be transported to protrusions by the KIF1C motor, which could also transport its own mRNA. In agreement with this hypothesis, smFISH experiments performed in the KIF1C-GFP BAC cell line showed that RAB13 mRNAs co-localized with the KIF1C-GFP protein in cytoplasmic protrusions (Figure 3C). In addition, RAB13 mRNAs became less localized when KIF1C expression was reduced with siRNAs (Figure 3B and S2B-C), confirming a role for this kinesin in RAB13 mRNA localization. Next, we performed smiFISH to detect several mRNAs among the most enriched in the KIF1C-GFP IP. TRAK2 and NET1 mRNAs accumulated in cytoplasmic protrusions, while three control mRNAs did not (KIF20B, PAK2, MYO18A; Figure 3B-D). A GO term analysis of the top 200 mRNAs associated with the KIF1C-GFP motor revealed an enrichment for "post-Golgi vesicle-mediated transport" (5.4 fold enriched, pV 3 10^{-2}), "organelle localization by membrane tethering" (4.2 fold enriched, pV 8.10^{-3}), "microtubule-based process" (3.6 fold, pV 9 10^{-7}), and "cilium assembly" (3.5 fold, pV 1.3 10^{-4}). Local translation of KIF1C-associated mRNAs in cell extensions may thus help these processes. Taken together, these data indicate that the kinesin KIF1C is part of a mechanism that localizes mRNAs to cytoplasmic protrusions.
Systematic screening of BAC-tagged mRNAs reveals a total of six cytoplasmic localization patterns

We then analyzed 409 BAC-tagged genes involved in a variety of biological processes (Table S2). The results revealed that 26 mRNAs displayed a specific localization in the cytoplasm (about 6%), which were classified into 6 patterns (Figure 4 and S3 for larger fields of view; Table 1 and S3). In addition to the previous patterns ("protrusion" and "foci"), some mRNAs accumulated at the nuclear envelope ("nuclear envelope"), around the cell periphery ("cell edge"), or were clustered in a cytoplasmic region where individual molecules were still resolved ("polarized"; Figure 4B and S3, Table S3). Interestingly, we also found three mRNAs that accumulated at centrosomes during cell division ("mitotic"; HMMR, ASPM and NUMA1; Figure 5).

Most mRNAs displayed a single localization pattern, however, some showed several (Table 1 and S3). For instance, ASPM mRNAs simultaneously localized at the nuclear envelope and in foci of interphase cells, while it localized at the mitotic apparatus during cell division (Figure 5B). Likewise, HMMR mRNAs localized in both foci and at centrosomes (Figure 5A and see below). These examples highlight the complexity and dynamic nature of RNA localization even in a simple cell culture system.

Identification of potential cases of locally-translated mRNAs

For the localized mRNAs, we then examined the colocalization with their encoded protein. Although accumulation of mRNAs in foci was the most frequent pattern (19/32; see Table 1), only two mRNAs displayed a faint but detectable GFP signal in their foci: CRKL and β-catenin (see Figure 4B and S3 for CRKL, and Figure S8A for CTNNB1). For the other localization
patterns, 9 mRNAs colocalized with their encoded protein (2% of the screened mRNAs; Table 1): AKAP1, AKAP9, AP1S2, ASPM, ATP6A2, FLNA, HMMR, NUMA1 and HSP90B1 (Figure 4B and 5). In some cases, the mRNA/protein colocalization was expected. For instance, both HSP90B1 and ATP6A2 contain a signal peptide leading to translation on the ER, and HSP90B1 is a resident ER protein while ATP6A2 localizes in endo-lysosomes close to this compartment (Figure 4B). Similarly, AKAP1 encodes an RNA-binding protein localizing to the surface of mitochondria, and this transcript thus belongs to the known class of mitochondrion-localized mRNAs (Sylvestre et al., 2003). Finally, Filamin A (FLNA) cross-links actin filaments to membrane proteins and it localizes at the cell periphery together with its mRNA (Figure 4B). It is thus related to the previously described class of peripheral mRNAs involved in actin metabolism (Chen et al., 2015; Lawrence and Singer, 1986a). However, the other proteins localized to cellular structures that were not previously known to use local translation as a targeting mechanism. This included the clathrin adaptor AP1S2 mRNA that localized on endosomes and the AKAP9 mRNA that accumulated at the Golgi. Below, we explore these cases in more details together with mRNAs localizing around centrosomes (ASPM, NUMA1 and HMMR), at the nuclear envelope (ASPM and SPEN), or in cytoplasmic foci (19 mRNAs).

**ASPM, NUMA1 and HMMR mRNAs localize to centrosomes in a cell-cycle dependent manner**

The HMMR protein is known to localize on microtubules and centrosomes, and to have spindle-promoting activities (Groen et al., 2004; Maxwell et al., 2003). The HMMR mRNAs accumulated in the peri-centrosomal region in interphase and much more strongly so in mitosis, and it co-localized there with its protein (Figure 5A and S4A). Moreover, using a RPE1 cell line
stably expressing a Centrin2-GFP protein, we confirmed that the endogenous HMMR mRNAs preferentially localized near centrosomes (Figure S4B).

The other mRNAs localizing at centrosomes were ASPM and NUMA1. Both proteins control spindle function during mitosis (Kouprina et al., 2005; Radulescu and Cleveland, 2010). Interestingly, these proteins had similar localization patterns. In interphase, both localized to the nucleoplasm while during mitosis, they concentrated on the spindle poles and weakly stained microtubules (Figure 5B-C; pink arrow points to cells in prophase). Remarkably, the localization of ASPM and NUMA1 mRNA was dynamic during the cell cycle (Figure 5B-C). In interphase, these mRNAs were dispersed throughout the cytoplasm, with ASPM mRNAs additionally localizing in foci and decorating the nuclear edge. However, both mRNAs relocalized to centrosomes during mitosis, where they became highly concentrated and colocalized with their GFP-tagged proteins (Figure 5B-C, pink arrow indicates cells in early mitosis). Identical localization patterns were observed for the endogenous mRNAs (Figure S4C-D). It is striking that NUMA1 and ASPM mRNAs localized at the centrosomes only during mitosis, while HMMR mRNAs localized to centrosomes during both mitosis and interphase. This indicates a very fine regulation of mRNA localization at centrosomes.

The co-localization of HMMR, ASPM and NUMA1 mRNAs with their proteins suggested that they were translated locally at centrosomes. To confirm this possibility, we analyzed the localization of two translation factors: eIF4E and the phosphorylated form of the ribosomal protein RPS6 (p-RPS6). Immuno-fluorescence showed that the endogenous eIF4E and p-RPS6 proteins were present throughout the cells but also accumulated at mitotic centrosomes (Figure 5D-E, see cells in prophase). Therefore, not only mRNAs but also the translational apparatus was present at the spindle poles. This suggests the presence of a specific translational program occurring on mitotic centrosomes.
The localization of ASPM and SPEN mRNAs at the nuclear envelope is translation-dependent

ASPM-GFP mRNAs accumulated at the nuclear envelope during interphase (Figure 5B; see quantification in 6B). In addition, SPEN-GFP mRNAs, which encode a nuclear protein, were also enriched around the envelope (Figure 4B). Labelling of the nuclear pores using either a CRM1-GFP fusion or an antibody against NUP133 (Figure S5A-B), showed that ASPM and SPEN mRNAs localized close to nuclear pores, rather than between them. This localization could result from two mechanisms. First, mRNAs could transiently localize at the pores on their way out of the nucleus. Second, they could localize at the cytoplasmic side of the pore to facilitate re-entry of newly translated proteins. To distinguish between these possibilities, we inhibited either transcription with actinomycin D or translation with puromycin. After 1h of actinomycin D treatment, SPEN and ASPM mRNAs still localized at the nuclear envelope (left panels in Figure 6A; Figure S5C and 6B for quantifications). On the contrary, both mRNAs became dispersed after 1h exposure to puromycin (Figure 6A and S5C; 6B for quantifications). We then used cycloheximide, which inhibits translation by freezing the ribosomes on the mRNAs, in contrast to puromycin that induces premature termination and releases the nascent peptide. Cycloheximide had no effects on the localization of ASPM mRNAs (Figure 6A; left panels), indicating that their localization did not require ongoing protein synthesis but most likely the presence of nascent proteins on polysomes. Therefore, these ASPM and SPEN mRNAs are not in transit to the cytoplasm but localize on the nuclear envelope by a translation-dependent mechanism requiring the nascent protein.

ASPM mRNAs are translated at the nuclear envelope
We and others recently demonstrated that the SunTag can be used to visualize translation of single mRNPs in intact cells (Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). The SunTag is a repeated epitope, and in cells expressing an scFV-GFP monochain antibody directed against this epitope, the antibody binds the repeated tag as soon as it emerges from the ribosome, allowing the visualization of single molecules of nascent proteins (e.g. monosome and polysomes). To confirm that ASPM mRNAs were translated at the nuclear envelope, we tagged the endogenous ASPM mRNAs using CRISPR genome editing. Homologous recombination allowed us to obtain heterozygous clones where 32 SunTag repeats were fused at the N-terminus of ASPM proteins (Figure S5D). The endogenous ASPM transcripts were then labelled by smiFISH, labelling both tagged and untagged mRNAs. This showed that bright spots of scFv-GFP colocalized with single ASPM mRNAs (Figure 6C; white arrows). These spots disappeared after a 20 minute puromycin treatment, confirming that they were polysomes (Figure S5E). We then performed time-lapse analyses of live cells, recording one image stack every 40 seconds for 50 minutes (see Movie and still images in Figure 6D). Remarkably, some polysomes localized at the nuclear envelope, and while cytoplasmic polysomes moved too rapidly to be tracked, the ones at the envelope remained immobile for extended periods of times (29 minutes in average). Thus, a fraction of ASPM mRNAs was stably anchored and translated at the nuclear envelope.

**Translation inhibition frequently prevents mRNA localization**

Since the localization of ASPM mRNA at the envelope is translation dependent, we tested whether this was also the case for its localization at centrosomes. Puromycin abolished its localization, while actinomycin D and cycloheximide had little effect (Figure 6A, right panels; Figure 6B for quantification). The effect of puromycin was then tested on the other localized
mRNAs. After 1 hour of treatment, KIF1C and MYH3 mRNAs still localized to cytoplasmic protrusions (Figure 7A and 7B for quantifications). In contrast, HSP90B1, HMMR, AP1S2, AKAP1, KIF4A and AKAP9 mRNAs all became delocalized when translation was inhibited, while protein localization was unaffected (Figure 7C-E). This data demonstrated that translation is required to bring these mRNAs to their destination.

**mRNA foci correspond to P-bodies and specialized translation factories**

We next turned to the analysis of mRNAs accumulating in foci, which was the most frequent pattern (19 out of 32 localized mRNAs; Table 1 and S3). We first imaged mRNAs together with a P-body marker. Foci of 15 mRNAs colocalized with P-bodies (Figure S6), while the remaining four were distinct structures (BUB1, DYNC1H1, CTNNB1/β-catenin and ASPM; Figure S7A-B). The dynein heavy chain DYNC1H1 and the ASPM protein were described above. BUB1 is a checkpoint kinase that verifies the attachment of microtubules to kinetochores at the onset of mitosis (Saurin, 2018), and β-catenin is the key transcription factor of the Wnt signaling pathway (Grainger and Willert, 2018). SmiFISH confirmed that the four endogenous mRNAs accumulated in foci (Figure S1A; S4C; S7C and see below S8B). Moreover, dual-color smiFISH showed that these foci did not co-localized together, indicating that they were distinct structures (Figure 8A). We then inhibited translation with puromycin, using an mRNA accumulating in P-bodies as control (AURKA). After 1h of treatment, the four non-P-body mRNA foci virtually disappeared (Figure 8B-C), while the accumulation of AURKA mRNA in P-bodies actually increased. Thus, translation specifically disrupted the mRNA foci that were not P-bodies. We previously used the SunTag to show that DYNC1H1 is translated in mRNA foci (Pichon et al., 2016). Using the CRISPR SunTag clone described above, we found that ASPM mRNAs were similarly translated...
in the foci (Figure 6C, orange arrow; 70% of the foci have a SunTag signal). A 32xSunTag repeat was then introduced at the N-terminus of the BUB1 protein using CRISPR gene editing. BUB1 mRNAs were detected by smiFISH and were found to be frequently translated in the foci (Figure 8D-E; 70% of the foci contained a SunTag signal), in contrast to mRNAs located outside foci that were rarely translated (1% of the time). Taken together, this data demonstrated that ASPM, BUB1 and DYNC1H1 are translated in mRNA foci, which thus correspond to specialized translation factories.

CTNNB1/β-catenin mRNA foci are sites of co-translational protein degradation

We then analyzed in more details CTNNB1/β-catenin. This protein has two roles: it bridges E-cadherin to the actin cytoskeleton at adherens junctions, and it acts as the main transcription factor of the Wnt pathway (for review, see (Grainger and Willert, 2018; MacDonald et al., 2009)). The latter function involves a fast activation of β-catenin expression by Wnt, and this pathway is essential during development and also often a key actor during tumorigenesis. The β-catenin protein localized at adherens junction is stable whereas the one synthesized in the cytoplasm is rapidly degraded (reviewed in (Stamos and Weis, 2013)). However, the cytoplasmic protein is stabilized in presence of a Wnt signal, and it can then accumulate in the nucleus to activate transcription of target genes. In the β-catenin BAC cells, the GFP-tagged protein was weakly expressed and it displayed an accumulation at sites of cell-cell contacts, as expected, as well as a weak staining in the brightest RNA foci (Figure S8A, top panels). Since these foci required ongoing translation for their formation (see Figure 8B-C), this suggested that β-catenin was translated in the foci.
Interestingly, β-catenin mRNAs did not form foci in a small fraction of the cells, where they were instead dispersed as single molecules. Furthermore, these cells expressed high levels of nuclear β-catenin-GFP, suggesting that the Wnt pathway had been activated and was responsible for the disappearance of the foci (Figure S8A, bottom panels). To test this hypothesis, we activated the Wnt pathway by incubating the BAC cell line with the WNT3A protein. Indeed, the mRNA foci disappeared after 30 minutes, concomitant with a higher expression of β-catenin-GFP and its accumulation in the nucleus (Figure 9A and 9D for quantifications). This observation establishes a link between the presence of mRNA foci and the degradation of the β-catenin protein. One hypothesis to explain these results would be that β-catenin degradation takes place co-translationally in the foci. Degradation of cytoplasmic β-catenin requires GSK3, APC and Axin (Stamos and Weis, 2013). In agreement with our hypothesis, labelling with anti-APC antibodies revealed that the β-catenin mRNA foci contained high APC levels (Figure 9B). Moreover, when APC was knocked-down with siRNA, β-catenin levels increased 50-fold and, remarkably, the mRNA foci disappeared with single mRNAs dispersing throughout the cytoplasm (Figure 9C and 9E). Finally, we analyzed the localization of the endogenous β-catenin mRNA by smiFISH in HEK293 cells. Foci of β-catenin mRNAs were visible in 16% of the cells, and this increased to 32% after treatment with LG-007, a Tankyrase inhibitor that reduces Wnt signaling by increasing Axin levels (Huang et al., 2009). Altogether, this indicated that the β-catenin mRNA foci were sites of co-translational protein degradation and that their formation required APC and translation. In addition, Wnt signaling disassembled the foci, thereby allowing a fast post-translational response.

Discussion

Chouaib et al.
In this study, we performed a dual RNA/protein localization screen and analyzed more than 500 genes. We found 32 localized mRNAs, with 15 accumulating in P-bodies and the others being locally translated (Table 1). This screen identifies housekeeping functions for RNA localization and helps defining the entire set of locally translated mRNAs. We also discovered a number of unexpected features, and in particular that: (i) co-translational protein targeting is frequent and occurs at unexpected locations; and (ii) mRNAs can be translated in specialized translation factories, which most likely perform key functions on the metabolism of nascent proteins.

Co-translational protein targeting is widespread

We found a number of mRNAs that co-localize with their encoded proteins, suggesting that these are translated locally. Besides the well-known cases of protein targeted to the ER or the mitochondria (HSP90B1, ATP6A2, AKAP1), we found that this occurred in other locations such as cell edge (FLNA), protrusions (KIF1C), centrosomes (NUMA1, HMMR, ASPM), endosomes (AP1S2), the Golgi apparatus (AKAP9) and the nuclear envelope (ASPM and also likely SPEN).

Translation at the nuclear envelope is surprising and it could have two purposes. First, folding constraints may render import of the mature protein inefficient, for instance if the protein tends to aggregate in absence of its nuclear partners. Second, it could prevent the protein from diffusing in the cytosol and thus interfere with cytoplasmic processes. This might be the case for ASPM because the protein localizes to centrosomes specifically during mitosis and its targeting to centrosomes during interphase may be deleterious.

One remarkable outcome of this study is that mRNA localization frequently requires translation. Indeed, if one excludes mRNAs localized to P-bodies, we found that RNA-driven, translation-independent localization seems the exception rather than the rule, with only two mRNAs out of 12 falling in this category. Moreover, the differential effects of puromycin and
cycloheximide indicate that the nascent peptides appear important for mRNA targeting (see Figure 6). This is reminiscent of the SRP-dependent targeting of proteins to the ER, where the nascent peptide recruits SRP, which halts ribosomes until the entire complex docks on the SRP-receptor on the ER (Aviram and Schuldiner, 2017). In the future, it will be interesting to determine whether co-translational targeting to these other cellular locations also involves ribosome stalling, whether the mRNA contains targeting signals in addition to the nascent proteins, and whether the polysomal complex reaches destination via diffusion or active transport.

Overall, these data indicate that co-translational protein targeting is more widespread than previously appreciated and occurs at diverse intracellular locations. This suggests that some protein may require to be incorporated co-translationally in their relevant macro-molecular complex, directly at their final location.

A specific translational program occurs at mitotic centrosomes

Localization of mRNAs at mitotic centrosomes has been first observed in mollusc embryos, as a way to asymmetrically segregate mRNAs involved in developmental patterning (Lambert and Nagy, 2002). Centrosomal mRNAs have later been found during Drosophila development (Lécuyer et al., 2007), and also very recently in Zebrafish embryos, in the case of the pericentrin mRNAs that encode a centrosomal protein (Sepulveda et al., 2018). Here, we found three mRNAs that localize to mitotic centrosomes in HeLa cells: HMMR, NUMA1 and ASPM. Interestingly, the corresponding proteins are known to play essential roles in the formation of the mitotic spindle, and deregulation in their expression can lead to aberrant cell proliferation and cancer (Mele et al., 2017; Sebestyén et al., 2016; Xie et al., 2017). Inactivating mutations of ASPM in drosophila and mammals yields abnormal neurogenesis, and spontaneous mutations in
human cause microcephaly, most likely due to cell proliferation defects of neuronal progenitors (Barbelanne and Tsang, 2014). HMMR, NUMA1 and ASPM mRNAs and proteins all localize to the spindle poles, and we further show that key components of the translation apparatus also accumulate there. This provides strong support to the idea that these proteins are translated directly at the spindle poles during mitosis.

The localization of HMMR, ASPM and NUMA1 mRNAs appears remarkably complex and each mRNA also has specific features. ASPM and NUMA1 mRNAs do not localize to centrosomes during interphase: NUMA1 mRNAs are randomly distributed while ASPM mRNAs decorate the nuclear envelope and accumulate in foci that function as translation factories. In contrast, HMMR mRNAs localize to centrosomes in interphase and also accumulate in P-bodies in a fraction of the cells (see Figure S6). Therefore, the localization and translation of these mRNAs appear very finely regulated in both time and space, and to also involve multiple cellular compartments. A common and important feature is however that the HMMR, ASPM and NUMA1 proteins accumulate in centrosomes only when their respective mRNAs are present there, thus reinforcing the idea they are locally translated. It is interesting to note that the centrosomal localization of these mRNAs is very high in prophase (Figure 5), as in the case of pericentrin mRNAs (Sepulveda et al., 2018). Since inhibiting translation blocks mitosis entry (Epifanova et al., 1969), it is tempting to speculate that there is a specific translational program that takes place at centrosomes at the onset of mitosis, which would be required for mitosis entry. This program could modify the composition of centrosomes to allow them to drive spindle assembly or other mitotic functions.

**KIF1C localizes mRNAs to cytoplasmic protrusions**

Chouaib et al.
We found 5 mRNAs that localize to cytoplasmic protrusions, among which the kinesin KIF1C is particularly interesting. First, KIF1C mRNA co-localizes there with its protein, suggesting that it is translated locally. Second, this motor appears to be involved in the localization of a number of other mRNAs, including the RAB13 mRNA that was the first described to localize in protrusions (Mili et al., 2008). Interestingly, two types of protrusions can contain mRNAs (Wang et al., 2017), and the KIF1C-associated transcripts mostly correspond to those enriched in high-contractility protrusions, suggesting that KIF1C acts on this pathway. KIF1C also binds its own mRNA, suggesting a possible positive feedback loop in which locally translated KIF1C proteins would help to localize additional mRNAs. This could occur by an anchoring mechanism, or by alternating back-and-forth movements on the cytoskeleton, allowing the motor to explore the cytoplasm and transport back additional mRNAs. The KIF1C motor directly and physically interact with mRNAs (Baltz et al., 2012; Castello et al., 2012), and the localization of KIF1C mRNA to protrusions does not require protein synthesis. This suggests a transport mechanism driven by RNA sequences in association with motor proteins, with the particular feature that the motor physically contact the mRNAs.

**P-bodies and translation factories**

The most frequently observed localization class was "foci", in which mRNAs form cytoplasmic aggregates containing multiple molecules. We identified 19 such mRNAs, and 15 of the corresponding foci were P-bodies, confirming that this is an abundant localization class (Hubstenberger et al., 2017b). Four mRNAs formed foci distinct from P-bodies and also distinct from each other, and we showed that they correspond to specialized translation factories. Except for β-catenin (see below), the function of these factories is not known. However, since the mature
protein does not accumulate there, they likely relate to the metabolism of the nascent protein. They could potentially be involved in the co-translational assembly of protein complexes (Pichon et al., 2016; Shiber et al., 2018), or to help to locally concentrate specific chaperones or modification enzymes.

In the case of β-catenin, we showed that the factories contain APC and that they dissolve upon APC knock-down or Wnt signaling, leading to a large increase in β-catenin expression. APC is well-known to associate with β-catenin and to drive its rapid degradation (Grainger and Willert, 2018; MacDonald et al., 2009). In turn, Wnt signaling sequesters APC at the plasma membrane, thereby stabilizing β-catenin. Interestingly, APC can make multivalent interactions, as one APC molecule can bind more than 30 molecules of β-catenin (Su et al., 1993). Most likely, the nascent β-catenin protein recruits APC, which can then bring multiple β-catenin polysomes together by its ability to bind many β-catenin proteins. This suggests that the formation of the factories could follow general phase-separation mechanisms relying on multivalent interactions made by polymeric molecules. Possibly, the β-catenin mRNAs may also directly contribute to factory formation, since it has been shown to physically interact with APC (Preitner et al., 2014). While the function of APC in β-catenin metabolism was not completely understood, our data indicate a direct role in organizing β-catenin translation factories.

To degrade β-catenin, APC works within a destruction complex that also contains Axin and GSK3 [Stamos, 2013 #39. Axin is also present in factories (data not shown). The high local concentrations of factors within factories may favorize further interactions to ensure that newly-made β-catenin interacts with APC, Axin and the other components of the destruction complex. Our discovery of the β-catenin translation factories provides a function (i.e. to aggregate polysomes) for the previously noted ability of the destruction complex to make large molecular
assemblies (Mendoza-Topaz et al., 2011). This property would otherwise be difficult to rationalize, because concentrating the degradation factors in a mega-complex while the protein is made elsewhere would decrease the chances of binding. Taken together, these data revealed that co-translational protein degradation is a gene regulation mechanism enabled by translation factories.

Overall, we found 4 translation factories while screening 500 mRNAs. Extrapolation to the 20,000 human genes suggests that few hundred such factories may exist in a cell. Translation may thus be compartmentalized with a much finer grain than previously anticipated.

Acknowledgements

We thank Florence Rage and Susanne Schmidt for the preparation of neuronal hippocampus primary cells. We also thank the MRI imaging facilities for their technical support. R.C. was supported by a scholarship from the Lebanese university and the Lebanese National Council for Scientific Research (LNCSR). This project was supported by the Agence Nationale de la Recherche (ANR-11-BSV8-018-02 and ANR-14-CE10-0018-01), the Institut Pasteur, and the Fondation pour la Recherche Médicale (FRM; grant Hi-FISH to EB).

Author contributions statement

EB conceived the study, with help of FM and KZ. Experiments were performed by RC, AS, XP, MP, MCR, OSK, EB and AMT. AS, TW and FM performed the automated image analysis. AH, IP, TW, HLH and CZ provided input to the paper. RC, FM, KZ, MP, HLH and EB analyzed the data. EB prepared the Figures, with help of MP, AS, OSK and RC. EB wrote the manuscript.
Conflict of interest disclosure

The authors declare no competing financial interests.

References


Chouaib et al.


Figure Legends

Figure 1. A dual protein/mRNA localization screen using GFP-tagged BACs.

A-Schematic of the BAC screen by smFISH. The schematic depicts the integrated BAC with tagged gene (top; exons in red), which transcribes a GFP-tagged mRNA with all its regulatory sequences (bottom; 5' and 3' UTR). The location of the probes is indicated (black pins with red head), and X and Z represent neighboring genes.

B-The BACs allow simultaneous detection of mRNAs and their encoded proteins. Panels are micrographs of HeLa cells expressing the indicated BAC, or no BAC as control (top), and hybridized with Cy3-labelled probes against the GFP-IRES-Neo sequence. Red and left: signal from the smFISH probes; green and middle: signal from the GFP channel; blue: DAPI staining. Scale bar: 10 microns. Insets: zoom of the boxed areas.

Figure 2. Localization of the mRNAs encoding molecular motors.

A-Transcripts localizing to the cell periphery and cellular extensions. Legends as in Figure 1B. The BAC tagged gene is indicated on the left. Arrows indicate the mRNA accumulation at cell extensions. Scale bar: 10 microns.


Figure 3. The kinesin KIF1C transports mRNAs to cytoplasmic protrusions.

A-Transcripts associating with the KIF1C-GFP protein. The graph depicts microarray signal intensity of RNAs detected in a KIF1C-GFP pull-down (x-axis), versus the control IP (y-axis).
Each dot represents an mRNA. Red dot: KIF1C mRNA; blue dot: mRNAs enriched in the KIF1C-GFP IP.

B-Quantitative analysis of mRNA localization to the cell periphery. The graph depicts the mean distance of mRNAs to the cell membrane, after normalization for cell size. Stars over blue bars: values significantly different than the control mRNAs (BAC-tagged KIF20, PAK2 and MYO18); star over orange bars: values significantly different between them.

C-RAB13 mRNAs co-localize with the KIF1C-GFP protein in cytoplasmic protrusions. Legends as in Figure 1B, except that the KIF1C-GFP BAC cell line was used and hybridized with probes against the endogenous RAB13 mRNA. Arrows indicate the RAB13 mRNA accumulating in cytoplasmic protrusions. Scale bar: 10 microns. Inset: zoom of the boxed area.

D-The endogenous TRAK2 and NET1 mRNA accumulate in cytoplasmic protrusions. The panels depict micrographs of HeLa cells hybridized with probes against TRAK2 (top), or NET1 (bottom). Red and black, signal from the probes; Blue; DAPI staining. Scale bars: 10 microns.

Figure 4. Examples of localized mRNAs found in the BAC screen.

A-Schematic of the localization classes. The mRNAs are in red; nuclei in grey, and cellular area in blue.

B-Localized mRNAs. Legend as in Figure 1B. The BAC tagged gene is indicated (top of each mRNA panel), as well as the localization class (italics). Scale bar: 10 microns.

Figure 5. Localization of mRNAs and translation factors on centrosomes.

A-C-Localization of tagged transcripts coding for HMMR (A), ASPM (B), and NUMA1 (C). Legend as in Figure 1B. The localization class is indicated on the left. Orange arrows point to
mRNA foci; pink arrows point to mitotic centrosomes of cells in prophase; light blue arrows point to centrosomes of cells in interphase. Scale bar: 10 microns.

D-E-Localization of eIF4E (D) and phospho-RPS6 (E) in early mitotic cells. The α-FOP antibody (green and black) labels centrosomes. The staining of translation factors is in red and black. Scale bar: 10 microns.

Figure 6. Localization of ASPM mRNAs requires active translation.

A-Localization of ASPM mRNA to the nuclear envelope and mitotic centrosomes is abolished by puromycin. Panels depict micrographs of HeLa cells expressing the tagged ASPM BAC, after 1h of treatment with the drugs indicated on the left. The RNA foci in the puromycin panel is a transcription site. Scale bar: 10 microns.

B-Quantification of the localization of the tagged ASPM mRNAs. The graph depicts the percent of cells expressing the tagged ASPM mRNAs and having the transcript localized in the indicated pattern.

C-Foci of ASPM mRNAs contain polysomes. Top: schematic depicting the insertion of the SunTag at the N-terminus of ASPM protein. Bottom: micrograph of HeLa cells with a SunTagged ASPM allele and showing ASPM mRNA (by smiFISH, left and red), and the signal from the SunTag (middle and green). Blue: DAPI staining. White and black arrows: a single mRNA positive for the SunTag; orange arrow: an mRNA foci positive for the SunTag. Scale bar: 10 microns.

D-ASPM mRNAs can be translated at the nuclear envelope. Legend as in C, except that cells are imaged live and white arrow points to polysomes anchored at the nuclear envelope. The DNA (red) is stained with SiR-DNA.
Figure 7. Transcript localization is sensitive to translation inhibition.

A-Transcripts localizing in cytoplasmic protrusions are diversely affected by puromycin. Images are micrographs of HeLa cells expressing the indicated BAC-tagged gene, treated or not with puromycin for 1 hour, and hybridized with Cy3-labelled probes against the GFP-IRES-Neo tag (yellow). Blue: DAPI staining. Scale bar: 10 microns.

B-Quantification of the fraction of cells displaying mRNA accumulation in cytoplasmic protrusions. The graph depicts the percent of cells expressing the indicated tagged mRNA, and displaying mRNA accumulation at the cell periphery, either in absence of treatment (blue bars), or after 1 hour of incubation with puromycin (orange bars). Error bars represent the standard deviation.

C-Puromycin treatment disrupts the localization of AP1S2, AKAP1 and AKAP9 mRNAs. Legend as in Figure 1B; puromycin-treated cells were incubated 1 hour with the drug. Insets: zoom of the boxed areas. Scale bar: 10 microns.

D-Puromycin, but not cycloheximide, disrupts the localization of HMMR mRNA to centrosomes. Legend as in Figure 1B, puromycin treatment was for 1 hour. Scale bar: 10 microns. Dark blue arrows: mRNA in foci that colocalize with P-bodies, see Figure S6. Orange arrows: centrosomes; pink arrow: transcription site. Bottom: the graph depicts the percent of cells expressing the tagged HMMR mRNA and accumulating it at the centrosome during interphase.

E-Quantification of the co-localization between mRNAs and their encoded protein. The graphs display the normalized GFP fluorescence intensities around the indicated BAC-tagged mRNAs, for the genes of panel C. Values are averaged for each cell, and each dot is a cell. Red bars: mean.
Figure 8. Non-P-body mRNA foci are translation factories

A-Foci containing BUB1, ASPM, CTNNB1 and DYNC1H1 mRNAs do not co-localize with each other. Panels are micrographs of HeLa cells expressing the indicated BAC-tagged gene and hybridized with Cy3-labelled probes against the GFP-IRES-Neo sequence (red), together with Cy5-labelled probes against the indicated endogenous mRNA (green). Blue: DAPI staining; scale bar: 10 microns.

B-C-Puromycin dissociates non-P-body foci. B-Quantification of the fraction of cells having mRNAs in foci, for the indicated genes and in presence (orange) or absence (blue) of puromycin for 1h. Error bars represent the standard deviation. C-Micrographs are images of HeLa cells expressing the indicated BAC and hybridized with Cy3-labelled probes against the GFP-IRES-Neo tag. Arrows point to mRNA foci, and the nuclear outline is shown in orange. Scale bar: 10 microns.

D-BUB1 mRNAs are translated in mRNA foci. Top: schematic depicting the insertion of the SunTag at the N-terminus of the BUB1 protein. Bottom: micrograph of HeLa cells with a SunTagged BUB1 allele and showing BUB1 mRNA (by smiFISH; middle and red), and the signal from the SunTag (left and green). Blue: DAPI staining. Arrows: mRNA foci that are positive for the SunTag. Insets: zoom of the boxed areas; scale bar: 10 microns.

E-Quantification of translated SunTagged BUB1 mRNA. The graph represents the percent of BUB1 mRNA foci positive for the SunTag (left bar), and the percent of single BUB1 mRNAs also positive for the SunTag (right bar), in cells with one BUB1 allele tagged with the SunTag. Error: standard deviation of replicates.

Figure 9. CTNNB1 mRNA foci correspond to regulated sites of protein degradation.
A-WNT3A signaling induces β-catenin (CTNNB1) protein expression and dissolve mRNA foci. Images are micrographs of HeLa cells expressing the tagged CTNNB1 BAC, and hybridized in situ with Cy3-labelled probes against the tag (panels CTNNB1 mRNA), and treated or not with WNT3A conditioned media for 30 minutes (+WNT3A and control panels, respectively). CTNNB1-GFP panels: signal from the GFP channel, with the grey levels scaled identically for the WNT3A and control. Scale bar: 10 microns.

B-β-catenin mRNAs contain APC. Images are micrographs of HeLa cells expressing the tagged CTNNB1 BAC and both hybridized in situ with Cy3-labelleled probes against the tag (panel CTNNB1 mRNA and red in the merge), and incubated with Cy5-labelleled anti-APC antibodies (panel IF anti-APC and green in the merge). Blue: DAPI staining. Scale bar: 10 microns.

C-Knock-downs of APC dissolve β-catenin mRNAs foci and induce β-catenin protein expression. Images are micrographs of HeLa cells expressing the tagged CTNNB1 BAC, and hybridized in situ with Cy3-labelled probes against the tag (panels CTNNB1 mRNA), and treated or not with APC siRNAs, as indicated. CTNNB1-GFP panels: signal from the GFP channel, with grey levels scaled identically for the APC siRNA and control. Scale bar: 10 microns.

D-Quantification of the effect of WTN3A. Graph represent the percent of cells expressing β-catenin mRNA and displaying mRNA foci, in control cells or cells treated with WNT3A-conditionned medium for 30 minutes. Error bars represent the standard deviation of replicates.

E-Quantification of the effect of the APC knock-down. The left graph depicts the intensity levels of β-catenin-GFP protein in the nucleus, and the right graph displays the percent of cells expressing β-catenin mRNA and having mRNA foci. Error bars: standard deviation of replicates.
Table 1. Summary of the localized mRNAs found in the screen

<table>
<thead>
<tr>
<th>Compartment where RNA localize</th>
<th>Localization pattern</th>
<th>Gene</th>
<th>Colocalization with the encoded protein</th>
<th>Puromycin sensitivity of RNA localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protrusion</td>
<td>Protrusion</td>
<td>KIF1C, KIF4A, KIF5B, MYH3, DYNLL2</td>
<td>KIF1C</td>
<td>KIF4A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KIF5B and DYNLL2: nd</td>
</tr>
<tr>
<td>P-body</td>
<td>Foci</td>
<td>See Table S3 (15 species)</td>
<td>CRKL</td>
<td>-</td>
</tr>
<tr>
<td>Translation factories</td>
<td>Foci</td>
<td>DYNC1H1, ASPM, BUB1, CTNNB1</td>
<td>-CTNNB1</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-DYNC1H1, ASPM, BUB1 (nascent).</td>
<td></td>
</tr>
<tr>
<td>Spindle poles</td>
<td>Mitotic</td>
<td>ASPM, NUMA1, HMMR</td>
<td>All (during mitosis)</td>
<td>ASPM, HMMR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NUMA1: nd</td>
</tr>
<tr>
<td>Nuclear envelope</td>
<td>Nuclear envelope</td>
<td>ASPM, SPEN</td>
<td>ASPM (nascent)</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SPEN: nd</td>
</tr>
<tr>
<td>Golgi</td>
<td>Polarized</td>
<td>AKAP9</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Endosome</td>
<td>Polarized</td>
<td>AP1S2</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Cell edge</td>
<td>Cell edge</td>
<td>FLNA</td>
<td>yes</td>
<td>nd</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>Polarized</td>
<td>AKAP1</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>RE</td>
<td>Polarized</td>
<td>HSP90B1, ATP6A2</td>
<td>yes</td>
<td>yes, ATP6A2: nd</td>
</tr>
<tr>
<td>Intra-nuclear</td>
<td>Intra-nuclear</td>
<td>See Table S3 (11 species)</td>
<td>nd</td>
<td>nd</td>
</tr>
</tbody>
</table>
STAR Methods

Cell lines and culture conditions

The collection of HeLa-Kyoto stably cells transfected with the GFP-tagged BACs was described previously (Maliga et al., 2013; Poser et al., 2008). Parental HeLa cells and the BAC-GFP clones were maintained in Dulbecco’s modified Eagle’s Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma), 100 U/mL penicillin/streptomycin (Sigma) and 400 µg/ml G418 (Gibco). HeLa, C2C12 and NIH3T3 cell lines were grown in the same medium without G418. RPE1 Centrin2-GFP ((Mikule et al., 2007); a gift of B. Delaval) were cultured in DMEM:F12 medium (Gibco) supplemented with 10% fetal bovine serum (Sigma) and 100 U/mL penicillin/streptomycin (Sigma). SH-SY5Y cells were grown in DMEM with 10% FBS and differentiated for 72h in DMEM containing 3% FBS and 10 µM retinoic acid. Drugs were used at the following final concentrations: 100 µg/ml for puromycin, 200 µg/ml for cycloheximide, 5 µg/ml for actinomycin D and 0.5 µM for LG-007. For puromycin, data are shown for 1h or treatment, but similar results were obtained after 30 minutes only (not shown). For Actinomycin D, data are shown for 1h or treatment, but similar results were obtained for up to 6h of treatment (not shown). Treatment of cells with LG-007 was for 2h.

Insertion of SunTag cassette by CRISPR/Cas9

The recombination cassettes contained 500 bases of homology arms flanking a puromycin resistance gene translated from the endogenous ATG sequence, followed by a P2A sequence, 32 SunTag repeats, and another P2A sequence fused to the protein of interest. Hela Kyoto cells stably expressing the scFv-GFP were transfected using JetPrime (Polyplus) and a cocktail of three plasmids, including the recombination cassette and constructs expressing Cas9-HF1 and
guide RNAs with an optimized scaffold (Chen et al., 2013). Cells were selected on 0.25 µg/ml puromycin for a few weeks. Individual clones were then picked and analyzed by PCR genotyping, fluorescent microscopy and smiFISH with probes against the SunTag sequence. The sequences targeted by the guide RNAs were (PAM sequences are underlined): AAGTGAGCCCGACCAGCGGAGG for ASPM and CGGGGTATTCGAATCGGCAG for BUB1.

**Genotyping**

PCR was done using a Platinum Taq DNA Polymerase (Invitrogen) on genomic DNA prepared with GenElute Mammalian Genomic DNA miniprep (Sigma-Aldrich). The sequences of oligonucleotides were: 5’-TGTTCCTGGAAACCGCAATG (ASPM WT forward); 5’-GTTCCTGGAAACCGCAATG (SunTag reverse).

**Treatments with siRNAs**

HeLa cells were seeded on 0.17 mm glass coverslips deposited in 6-well plates. Cells were transfected at 70% confluency using JetPrime (Polyplus). Double-stranded siRNAs (30 pmoles) were diluted into 200 µl of JetPrime buffer. JetPrime reagent was added (4 µl) and the mixture was vortexed. After 10 at room temperature (RT), it was added to the cells grown in 2 ml of serum-containing medium. After 24 hours, the transfection medium was replaced with fresh growth medium and cells were fixed 24h later. The sequences of the siRNA were: KIF1C : 5’-CCGAUGCCGUCUUUACCAUdTdT-3’; APC-1 : 5’-GCACAAAGCUGUUGAAUUUdTdT-3’; APC-2 : 5’-UGAAAGUGGAGGGGAAUUUdTdT-3’; APC-3 : 5’-
UAAUGAACACUACAGAUAGdTdT. The three APC siRNAs gave similar results, and the data presented correspond to the siRNA APC-1.

**Single molecule fluorescent in situ hybridization**

Cells grown on glass coverslips were fixed for 20 min at RT with 4% paraformaldehyde diluted in PBS, and permeabilized with 70% ethanol overnight at 4°C. For smFISH, we used a set of 44 amino-modified oligonucleotide probes against the GFP-IRES-Neo sequence present in the BAC construction (sequences given in Table S1). Each oligonucleotide probe contained 4 primary amines that were conjugated to Cy3 using the Mono-Reactive Dye Pack (PA23001, GE Healthcare Life Sciences). To this end, the oligos were precipitated with ethanol and resuspended in water. For labelling, 4 µg of each probe was incubated with 6 µl of Cy3 (1/5 of a vial resuspended in 30 µl of DMSO), and 14 µl of carbonate buffer 0.1 M pH 8.8, overnight at RT and in the dark, after extensive vortexing. The next day, 10 µg of yeast tRNAs were added and the probes were precipitated several times with ethanol until the supernatant lost its pink color. For hybridization, fixed cells were washed with PBS and hybridization buffer (15% formamide in 1xSSC), and then incubated overnight at 37°C in the hybridization buffer also containing 130 ng of the probe set for 100 µl of final volume, 0.34 mg/ml tRNA (Sigma), 2 mM VRC (Sigma), 0.2 mg/ml RNAse-free BSA (Roche Diagnostic), and 10% Dextran sulfate. The next day, the samples were washed twice for 30 minutes in the hybridization buffer at 37°C, and rinsed in PBS. Coverslips were then mounted using Vectashield containing DAPI (Vector laboratories, Inc.).

For smiFISH, 24 to 48 unlabeled primary probes were used (sequences given in Table S1). In addition to hybridizing to their targets, these probes contained a FLAP sequence that was hybridized to a secondary fluorescent oligonucleotide. To this end, 40 pmoles of primary probes
were pre-hybridized to 50 pmoles of secondary probe in 10 µl of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl\textsuperscript{2}, pH 7.9. Hybridization was performed at 85°C for 3 min, 65°C for 3 min, and 25°C for 5 min. The final hybridization mixture contained the probe duplexes (2 µl per 100 µl of final volume), with 1X SSC, 0.34 mg/ml tRNA (Sigma), 15% Formamide, 2 mM VRC (Sigma), 0.2 mg/ml RNAse-free BSA, 10% Dextran sulfate. Slides were then processed as above. For AKAP9, TRAK2 and NET1, the probes used were RNA and not DNA (sequence in Table S1). The protocol was similar except that hybridization was performed at 48°C and that 50 ng of primary probe (total amount of the pool of probes) and 30 ng of each of the secondary probes were used per 100 µl of hybridization mix.

**Immunofluorescence**

HeLa cells were seeded and fixed as for smFISH. Cells were permeabilized with 0.1% Triton-X100 in PBS for 10 minutes at RT and washed twice with PBS. For P-body labelling, coverslips were incubated overnight at 4°C with a monoclonal mouse antibody (Santa Cruz sc-8418), diluted 1/250 in PBS containing 0.1% BSA. Coverslips were washed three times with PBS, 10 minutes each time, and incubated with an FITC anti-mouse secondary antibody (Jackson ImmunoResearch 115-096-006) diluted 1/200 in PBS with 0.1% BSA. After 1 hour of incubation at RT, coverslips were washed three times with PBS, 10 minutes each. For APC labelling, coverslips were incubated 1h at RT with an anti-APC mouse monoclonal antibody (F3, Santa Cruz sc-9998), diluted 1/50 in PBS containing 0.1% BSA. After washing as above, coverslips were incubated 1h at RT with a Cy5 anti-mouse secondary antibody (Jackson ImmunoResearch 115-176-003) diluted 1/100 in PBS with 0.1% BSA. Coverslips were washed again and mounted using Vectashield containing DAPI (Vector laboratories, Inc.).
To label the translation machinery, HeLa cells were rinsed with PBS and fixed by 4\% paraformaldehyde (PFA) for 10 minutes in room temperature (RT). After fixation, cells were washed with PBS and permeabilized through 0.1\% Triton X-100 (EUROMEDEX, 2000-A) for 3 minutes in RT. After washing with PBS, cells were incubated with 1\% Bovine Serum Albumin (BSA, Sigma Aldrich, A-7906) in PBS for 30 minutes in RT. Primary antibodies against FOP (Abnova, H00011116-M01), eIF4E (Santa Cruz, SC-13963), Phospho-RPS6 (Santa Cruz, SC-54279) were incubated with cells for 1 hour in RT. HOECHST 33258 (Sigma-Aldrich) was incubated for 5 minutes in RT and secondary antibodies fused with Alexa 488 (Thermofisher, A21141) and Cy3 (Jackson Immuno Research, 111-165-144) were incubated for 1 hour in RT. Cells were mounted with Fluoromount-G\textsuperscript{tm} (Invitrogen, 00-4958-02) on the slide glass. Images were captured with Nikon Ti fluorescence microscope equipped with ORCA-Flash 4.0 digital camera (HAMAMATSU). Images were modified by Image J (NIH).

**Imaging of fixed cells**

Microscopy slides were imaged on a Zeiss AxioimagerZ1 wide-field microscope equipped with a motorized stage, a camera scMOS ZYLA 4.2 MP, using a 63x or 100x objective (Plan Apochromat; 1.4 NA; oil). Images were taken as z-stacks with one plane every 0.3 \( \mu \)m. The microscope was controlled by MetaMorph and Figures were constructed using ImageJ, Adobe Photoshop and Illustrator.

**Imaging of live cells**

Live imaging was done using a spinning disk confocal microscope (Nikon Ti with a Yokogawa CSU-X1 head) operated by the Andor iQ3 software. Acquisitions were performed using a 100X
objective (CF1 PlanApo λ 1.45 NA oil), and an EMCCD iXon897 camera (Andor). Samples were sequentially excited at 488 and 640nm. Cells were maintained in anti-bleaching live cell visualization medium (DMEM<sub>gfp</sub>; Evrogen), supplemented with 10% fetal bovine serum at 37°C in 5% CO2. SiR-DNA (Spirochrome) was kept at 100 nM throughout the experiments to label DNA.

**Deconvolution**

For Figure S5A and S5B, deconvolution was performed using the Huygens Professional software (Scientific Volume Imaging, Hilversum, Netherlands). The point spread function (PSF) was a theoretical one, and background values were manually estimated. The deconvolution was done using the CMLE deconvolution algorithm, with 50 iterations and a quality threshold of 0.01, and without bleaching correction.

**Image analysis and quantifications:**

Automated nuclear and cell segmentation was performed with a custom algorithm based on the U-net deep convolutional network (Ronneberger et al., 2015). Nuclear segmentation was performed with the DAPI channel, cell segmentation was performed with the autofluorescence of the actual smFISH image. For segmentation, 3D images were projected into 2D images with a recently described approach called smiFISH (Tsanov et al., 2016). Messenger RNA were detected with FISH-quant (Mueller et al., 2013) by applying a local maximum detection on LoG filtered images. For Figures 6B, 7B-D, 8B-D, 9D-E, S8C, cell were also manually annotated to assign them a pattern type. For the quantifications of Figure 3B, the distance of every mRNA molecule to the cell membrane was computed as the minimum distance between the mRNA and...
every point of the cell outline polygon. The distances were then normalized by the square root of the cell area to reduce the impact of the cell size. The plot displays the normalized mean distance of mRNA to the cell membrane for different genes and their standard deviation.

To measure the degree of spatial overlap between mRNA (by smFISH) and protein (by the GFP fluorescence) in Figure 7, we calculated an enrichment ratio. We opted for such a quantification since the GFP fluorescence varied between cells, making an object detection in this channel difficult. Cells and nuclei were outlined manually in 2D based on the GFP and DAPI image, respectively. The subsequent analysis was restricted to the cytoplasm. FISH-quant was used to detect mRNAs in 3D and each cell was post-processed separately. First, we calculated the median pixel intensity in the IF image at the identified RNA positions. Second, we estimated a normalization factor as the median IF intensity of the outlined cytoplasm within the z-range of to the detected mRNAs. The enrichment ratio was then estimated as the ratio of the median IF intensity at the RNA positions divided by the mean cytoplasmic intensity. Boxplots of enrichment ratios were generated with the Matlab function notBoxPlot. Each dot corresponds to the estimate of one cell. Horizontal lines are mean values, 95% confidence interval is shown in red, and 1 standard deviation in blue. Statistical comparison between different experimental conditions were performed with two-sample Kolmogorov-Smirnov test (Matlab function kstest2).

**Immuno-precipitation and microarrays**

HeLa cells containing the KIF1C-GFP BAC were grown to near confluence in 10 cm plates, and two plates were used per IP. Cells were rinsed in ice-cold PBS, and all subsequent manipulations were performed at 4°C. Cells were scraped in HTNG buffer (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, 1% Triton X-100, 10% glycerol, 1 mM MgCl₂, 1 mM EGTA), containing an antiprotease cocktail (Roche). Cells were incubated for 20 minutes on a rotating wheel, and
cellular debris were removed by centrifugation of the extracts 10 minutes at 20,000g. Beads coated with GFP-trap antibody (ChromoTek), or uncoated as control, were washed in HNTG (25 μl of beads per IP). Beads were incubated 1h with a control extract to saturate non-specific binding and then incubated with the proper extract. After 4h of incubation on a rotating wheel, beads were washed four times in HNGT with anti-protease, and twice with PBS. Beads were then incubated with Trizol to extract RNAs, and RNA purification was done as recommended by the manufacturer. The resulting RNAs were amplified and converted into cDNAs by the WT PICO kit (Thermo Fisher), and hybridized on HTA 2.0 chip on an Affymetrix platform (Thermo Fisher). Experiments were performed in duplicates, data were normalized and averaged. Data are deposited on GEO with the following accession number: (in progress).

**Supplemental Spreadsheets and Movie**

**Table S1, related to all Figures.** Sequence of the smFISH and smiFISH probes.

**Table S2, related to Figure 2 and 4.** Summary of all the mRNAs screened.

**Table S3, related to Figure 2 and 4.** Summary of all the localized mRNAs.

**Table S4, related to Figure 3.** Microarray identification of mRNAs enriched in the KIF1C IP.

**Movie, related to Figure 6:** SunTagged ASPM polysomes (green) are anchored on the nuclear envelope.

DNA (red) is stained with SiR-DNA. The movie is made from a maximal image projection of z-stacks, acquired every 40 seconds for 50 minutes.
Figure 1

A

integrated BAC

gene X  Tagged gene Y  gene Z

smFISH with 44 Cy3 labeled oligos

mRNA

5' UTR  Y  GFP  IRES-neo  3' UTR

B

mRNA  GFP  Merge

HeLa Kyoto

C-LAP TERF1

N-FLAP VAPA

Chouaib et al.
Figure 2

Chouaib et al.
Figure 3
Figure 5

(A) HMMR

(B) ASPM

(C) NUMA1

(D) α-eIF4E

(E) α-p-RPS6
Figure 6
A

KIF1C

Control | Puromycin

KIF4A

Control | Puromycin

B

% cells with mRNA at the periphery

- KIF1C
- KIF4A
- MYH3

Cont. | Puro

C

AP1S2

Control | Puromycin

AKAP1

Control | Puromycin

AKAP9

Control | Puromycin

D

HMMR

Control | Puromycin

HMMR mRNA

- Centrosome (interphase)

% of cells with pattern

- Control
- Puro
- CHX

E

Normalized GFP signal around mRNAs

- AKAP9
- AKAP1
- AP1S2
- HSP90B1

Control | Puro
Figure 9
Supplemental Information for

A localization screen reveals translation factories and widespread co-translational protein targeting

Racha Chouaib\textsuperscript{1,2,3,11+}, Adham Safieddine\textsuperscript{1,2,3+}, Xavier Pichon\textsuperscript{1,2}, Oh Sung Kwon\textsuperscript{4}, Aubin Samacoits\textsuperscript{5,6}, Abdel-Meneem Traboulsi\textsuperscript{1,2}, Nikolay Tsanov\textsuperscript{1,2}, Marie-Cécile Robert\textsuperscript{1,2}, Ina Poser\textsuperscript{7}, Christophe Zimmer\textsuperscript{5,6}, Anthony Hyman\textsuperscript{7}, Hervé Le Hir\textsuperscript{8}, Thomas Walter\textsuperscript{8,9,10}, Kazem Zibara\textsuperscript{3,11}, Florian Mueller\textsuperscript{5,6}, Marion Peter\textsuperscript{1,2}, Edouard Bertrand\textsuperscript{1,2,*}
Figure S1
Figure S1, related to Figure 2. Several endogenous mRNAs coding for motor proteins localize in diverse cell lines.

A-KIF1C, DYNC1H1, MYH3 and KIF5B mRNAs in HeLa and mouse C2C12 cells. Images are micrographs of the indicated cell line, hybridized with Cy3-labelled oligonucleotide probes against the indicated endogenous mRNAs. Red and black: signal from the probes. Blue: DAPI staining. Arrows point to mRNA localizing in foci (DYNC1H1), or at the cell periphery. Scale bars: 10 microns.

B-CRM1, KIF1C, KIF5B and MYH3 mRNAs in differentiated SH-SY5Y neuroblast cells. Legend as in A; scale bar: 10 microns.
Figure S2
Figure S2, related to Figure 2 and Figure 3. Localization of mRNAs in cytoplasmic protrusions.

A-Co-localization of mRNAs in cytoplasmic protrusions. Images are micrographs of HeLa cells hybridized with Cy3- and Cy5-labelled oligonucleotide probes against the indicated endogenous mRNAs (red and green, respectively). Blue: DAPI staining. Scale bar: 10 microns. Insets: zoom of the boxed areas.

B-RAB13 mRNA is delocalized in KIF1C knock-down cells. Images are micrographs of HeLa cells treated or not with KIF1C siRNAs, and hybridized with Cy3-labelled oligonucleotide probes against RAB13 mRNA (red and black). Blue: DAPI staining. Scale bar: 10 microns. Arrows points to RAB13 mRNAs accumulating in cytoplasmic protrusions.

C-Quantification of the efficacy of KIF1C siRNAs. The graph represents the number of KIF1C mRNA per cell (50 cells in each case), in cells treated (blue) or not (orange) with KIF1C siRNAs. Error bars: standard deviation.
Figure S3
Figure S3, related to Figure 4. Examples of localized mRNAs found in the BAC screen.

A-Schematic of the localization classes. The mRNAs are in red; nuclei in grey, and cellular area in blue.

B-Localized tagged mRNAs. Legend as in Figure 1B. The tagged gene on the BAC is indicated (top of each mRNA panel), as well as the localization class (italics). Scale bar: 10 microns.
Figure S4
Figure S4, related to Figure 5. Localization of endogenous mRNAs at the centrosome.

A-Tagged HMMR mRNA accumulates in the peri-centrosomal region. Legend as in Figure 1B. Arrows point the centrosome. Scale bar: 10 microns.

B-Endogenous HMMR mRNAs accumulate at the centrosome. Images display micrographs of RPE1 cells stably expressing centrin2-GFP (middle and green in the merge), and hybridized with Cy3-labelled probes hybridizing to the endogenous HMMR mRNA. Scale bar: 10 microns. Arrows point the centrosome; inset: zoom of the boxed area.

C-Endogenous ASPM mRNAs accumulate at the nuclear envelope and at the spindle poles. Images display micrographs of HeLa cells hybridized with Cy3-labelled oligonucleotides against endogenous ASPM mRNAs (black and red in the merge). Orange arrows: mRNA foci. Scale bar: 10 microns.

D-Endogenous NUMA1 mRNAs accumulate at the spindle poles. Legend as in C except that probes hybridized to NUMA1 mRNAs.

E-Endogenous AKAP9 mRNAs cluster around the nucleus. Legend as in C except that probes hybridized to AKAP9 mRNAs.
Figure S5
Figure S5, related to Figure 6. Localization of BAC-tagged SPEN and ASPM mRNAs at nuclear envelope.

A-B-Labelling of BAC-tagged SPEN (A) and ASPM (B) mRNAs with the nuclear pores. Panels are deconvolved images of HeLa cells expressing the indicated tagged BAC gene and both hybridized with Cy3-labelled oligonucleotides against the GFP-IRES-Neo tag (left and red in the merge), and labelled for nuclear pores with a transfected CRM1-GFP construct (A; middle and green in the merge) or with an antibody against NUP133 (B; middle and red in the merge). Scale bar: 10 microns.

C-Puromycin, but not Actinomycin D, disrupt localization of tagged SPEN mRNAs at the nuclear envelope. Legend as in S4A, with cells treated for 30 minutes with the indicated drug. In the puromycin experiment, the RNA foci is a transcription site. Scale bar: 10 microns.

D-Genotyping of SunTagged ASPM clones. The image shows a gel loaded with PCR reactions on genomic DNA of SunTagged HeLa clones, with primers detecting the recombined allele. HeLa: parental HeLa DNA; H2O: control without template. The arrow points the expected size (1.3 kb), and a molecular weight marker is shown on the left. Clone 4 was used in subsequent experiments.

E-Puromycin disrupts the ASPM SunTag spots colocalizing with ASPM mRNAs. Images are micrographs of HeLa cells carrying a SunTagged ASPM allele, and hybridized with Cy3-labelled probes against the SunTag sequence (left). Insets: zoom of the boxed area. Scale bar: 10 microns.
Figure S6
Figure S6, related to Figure 8. BAC-tagged mRNAs accumulating in P-bodies.

Images are micrographs of HeLa cells expressing the indicated tagged BAC gene, and both hybridized with Cy3-labelled probes against the tag (left and green in the merge), and labelled by indirect immuno-fluorescence with an antibody against P-bodies (middle panel and green in the merge). Scale bar: 10 microns.
Figure S7
Figure S7, related to Figure 8. Some mRNA foci that are not P-bodies.

A-Legend as in Figure 6. Insets: zoom of the boxed area. Scale bar: 10 microns.

B-Foci containing BUB1, ASPM, or CTNNB1 mRNAs do not co-localize with P-bodies. The experiment was performed with the BAC-tagged cell lines. Legend as in A.

C-Endogenous BUB1 mRNA accumulates in foci. Images are micrographs of HeLa cells hybridized with Cy3-labelled probes against BUB1 mRNA (black and yellow in the merge). Orange arrows: mRNA foci; blue arrow: single mRNA. Scale bar: 10 microns.
Figure S8

A

CTNNB1-GFP

CTNNB1-GFP mRNA

B

CTNNB1 mRNA in HEK293 cells

Merge

C

% cells with foci

DMSO

LG-001

Chouaib et al.
Figure S8, related to Figure 9. BUB1 and beta-catenin translation factories.

A-Expression of β-catenin-GFP protein correlates with a dissolution of β-catenin mRNA foci. Legend as in Figure 1B, except the cells expressed a tagged CTNNB1 BAC. Orange arrow: β-catenin mRNA foci; blue arrow: accumulation of β-catenin-GFP protein in the nucleus. Scale bar: 10 microns.

B-Endogenous β-catenin mRNA accumulated in foci. Legend as in Figure S8A, except that probes bound endogenous β-catenin mRNA in HEK293 cells. Orange arrow points mRNA foci. Scale bar: 10 microns.

C-The graph displays the percent of cells having cytoplasmic β-catenin mRNA foci in HEK293 cells treated with LG-007 for 2 hours, or DMSO as control. Error bars: standard deviation.
3. Systematic identification of human centrosomal mRNAs reveals a localization mechanism dependent on motors and nascent translation

This paper had the following aims:

- Screen mRNAs encoding all centrosomal and spindle proteins using a new cost-effective high-throughput smFISH protocol to identify new centrosomal mRNAs.

- Describe in detail centrosomal mRNA localization across the cell cycle.

- Determine whether mRNA targeting to centrosomes occurs co-translationally and how widespread is this mode of transport.

- Using ASPM as model, image endogenous mRNAs and polysomes in living mitotic cells to gain insights into the mechanisms and dynamics of centrosomal mRNA trafficking.

Main conclusions of the paper were:

- A cell cycle-regulated translational program operates at centrosomes.

- mRNA localization to centrosomes is always translation dependent.

- ASPM mRNAs are co-translationally targeted to centrosomes using motors and microtubules.

In this study, I performed, analyzed and interpreted all experiments (with the notable exception of the screening pipeline). I wrote the manuscript and prepared all figures.

Supplementary movies and tables for this paper can be downloaded from

https://filesender.renater.fr/?s=download&token=29afda4f-2287-b5e5-1634-f93508dbc433
Systematic identification of human centrosomal mRNAs reveals a localization mechanism dependent on motors and nascent translation

Running title: mRNA localization and local translation

Keywords: mRNA localization, local translation, centrosomes, polysome imaging, cotranslational targeting, smFISH, high-throughput screening, ASPM
Abstract (149 words; <150)

Local translation allows for a spatial control of gene expression. Here, we used high-throughput smFISH to screen centrosomal protein-coding genes, and found a total of 8 human mRNAs accumulating at this organelle. These mRNAs localize at different stages during cell division, indicating a cell cycle regulated translational program at centrosomes. Interestingly, RNA localization required translation for all 8 transcripts, and we further show that the untranslated sequences of ASPM mRNA are not required for its localization. Using ASPM as a model, we imaged single mRNAs and polysomes with the MS2 and SunTag systems, respectively. This revealed active movements of polysomes towards the centrosome at the onset of mitosis, when these mRNAs start localizing. ASPM polysomes are anchored to MTs, and they localize by either motor-driven transport or the pulling of entire MTs towards centrosomes. This data indicates widespread targeting of polysomes by active transport mechanisms recognizing nascent proteins.
Introduction:

Messenger RNA localization is a post-transcriptional process by which cells target certain mRNAs to specific subcellular compartments. The trafficking of an mRNA molecule is often intertwined with its metabolism and function (Kejiou and Palazzo, 2017). Indeed, the subcellular localization of a transcript can influence its maturation, translation, and degradation. On one hand, mRNAs can be stored in a translationally repressed state in dedicated structures such as P-bodies (Hubstenberger et al., 2017). On the other hand, some mRNAs can localize to be translated locally. Such a local protein synthesis can be used to localize the mature polypeptide, and in this case it can contribute to a wide range of functions such as cell migration, cell polarity, synaptic plasticity, asymmetric cell divisions, embryonic patterning and others (Ryder and Lerit, 2018; Cody et al., 2013; Chin and Lécuyer, 2017 for reviews). More recently, local translation has also been linked to the metabolism of the nascent protein, rather than to localize the mature polypeptide. This is for instance the case for mRNAs translated in distinct foci termed translation factories, which correspond to small cytoplasmic aggregates containing multiple mRNA molecules of a given gene (Pichon et al., 2016; Chouaib, 2019). For instance in the case of CTNNB1 mRNAs, coding for β-catenin, these translation factories function in the co-translational degradation of the nascent protein (Chouaib et al., 2019).

Specific sub-cellular localization of mRNA molecules can be achieved by several mechanisms. Passive diffusion coupled with local entrapment and/or selective local protection from degradation are two strategies that can establish specific distributions of mRNA molecules (reviewed in Chin and Lécuyer, 2017). In most cases however, mRNA transport and localization occurs via motor-driven transport on the cytoskeleton (Bullock, 2007; Bertrand et al., 1998; Buxbaum et al., 2015). Molecular elements that regulate and control mRNA localization include cis- and trans- acting elements. Cis-acting...
elements are referred to as zip-codes and are often found within the 3’UTR of the transcript (Kim et al., 2015; Trcek and Singer, 2010; Jansen, 2001). Many types of zip-code have been described based on primary sequence, number, redundancy, and secondary structure. What these zip-codes have in common is that they carry sufficient information for localizing the transcript. Zip-codes bind one or several trans-acting RNA-binding proteins (RBPs), which mediate diverse aspects of RNA metabolism such as motor binding and translational regulation (reviewed in Chin and Lécuyer, 2017). Indeed, RNAs in transit are often subjected to a spatial control of translation (reviewed in Besse and Ephrussi, 2008). A long-standing notion in the field is that the transport of localized mRNAs occurs in a translational repressed state and that this serves to spatially restrict protein synthesis. Local translational de-repression occurs once the transcript has reached its destination, for instance by phosphorylation events and/or competition with pre-existing local proteins (Zaessinger et al., 2006).

While active transport of silent transcripts through RNA zipcodes appears a frequent mechanism, mRNA localization can also involve the nascent peptide, as in the case of secreted proteins. Here, the signal recognition particle (SRP) binds the nascent signal peptide, inhibits translation elongation, and mediates anchoring of the nascent polysome to the SRP receptor on the endoplasmic reticulum, where translation elongation resumes (Gilmore et al., 1982; Walter and Blobel, 1980, 1982). Interestingly, a few hints recently suggested that the translation of nascent peptides may play a role in the localization of other kinds of mRNAs. First, inhibiting translation with puromycin delocalizes certain mRNAs (Chouaib et al, 2019; Sepulveda et al., 2018) Second, nascent peptide imaging techniques revealed that some polysomes could be seen undergoing rectilinear motions characteristic of motor-driven transport. This was the case for a reporter carrying the 3’UTR of Arc mRNA (Wang et al., 2016) and for the endogenous
dynein heavy chain mRNA that form translation factories (Pichon et al., 2016). This showed that mRNPs can be translated during motor-mediated transport. Whether nascent peptides mediate active transport and localization of some classes of polysomes remain however to be demonstrated.

Centrosomes are ancient and evolutionary conserved organelles that are the major microtubule (MT) organizing centers in most animal cells. They play key roles in cell division, signaling, polarity and motility (Breslow and Holland, 2019; Wu and Akhmanova, 2017 for reviews). A centrosome is composed of two centrioles and their surrounding pericentriolar material (PCM). In cycling cells, the centriole duplication cycle is tightly coupled to the cell cycle to ensure a constant number of centrioles in each cell after mitosis (reviewed in Nigg and Holland, 2018). Briefly, G1 cells contain one parent and one mature parent centriole connected by a linker. At the beginning of S phase, each parent centriole orthogonally assembles one new procentriole. This configuration is termed engagement and prevents reduplication of the parent centriole. Procentrioles elongate as the cell is progressing through the cycle. Late G2 marks the dissociation of centriole linker and centrosomes separation. Centrosomal maturation and PCM expansion occur in G2 in preparation of mitotic spindle formation (Breslow and Holland, 2019). The first clues suggesting the importance of mRNA localization and local translation at the centrosomes were discovered almost 20 year ago in *Xenopus* early embryos (Groisman et al., 2000). It was found that cyclin B mRNA concentrated on the mitotic spindle, and that this localization was dependent on the ability of CPEB to associate with microtubules and centrosomes. A more global view was obtained in *Drosophila* embryos where a systematic analysis of RNA localization was performed (Lécuyer et al., 2007). Although this study did not reach single molecule sensitivity, it revealed that 6 mRNAs localized at centrosomes across different stages of early *Drosophila* development. Two more recent studies
(Jambor et al., 2015; Wilk et al., 2016) described several other mRNAs as enriched on centrosomes, spindles, or peri-nuclear regions that could contain the centrosome in later stages of *Drosophila* development. Notably, mRNAs encoding Polo-like kinase, involved in regulating mitotic entry and spindle formation, had a peri-nuclear pattern. In humans, four mRNAs were recently found to localize at centrosomes (PCNT, ASPM, NUMA1 and HMMR; Sepulveda et al., 2018; Chouaib et al., 2019). These mRNAs all code for centrosomal proteins, suggesting that they are locally translated. Interestingly, their localization was abolished by puromycin, raising the possibility that the nascent protein may be involved in the localization process. Here, we performed a systematic smFISH screen of mRNAs coding for centrosomal proteins. We found several new centrosomal mRNAs and show that their localization requires on-going translation. By imaging single ASPM and NUMA1 mRNAs and polysomes, we demonstrate that RNA localization at centrosomes occurs by active transport of polysomes.
Results

Screening genes encoding centrosomal proteins reveals a total of 8 human mRNAs localizing at the centrosome

Few human mRNAs were recently found to localize at the centrosome (Sepulveda et al., 2018; Chouaib et al., 2019). In order to acquire a global view of centrosomal mRNA localization, we developed a high-throughput smFISH technique working in a 96-well plate format and we screened genes encoding centrosomal and mitotic spindle proteins. We designed 50 to 100 individual probes against each mRNA of the screen. The probes were then generated from complex pools of oligonucleotides (12,000 to 92,000), using gene-specific barcodes (Xu et al., 2009), PCR and in vitro transcription (Figure 1A; see Materials and Methods). This procedure generated single-stranded RNA probes, each containing a gene specific sequence flanked by two overhangs common to all probes. A pre-hybridization step then labeled the overhangs with TYE-563 labeled LNA oligonucleotides, and the heteroduplexes were hybridized on cells as in the smiFISH technique (Tsanov et al., 2016). The advantages of this approach include cost-effectiveness because the probes are generated from an oligonucleotide pool. In addition, the probes can be used individually or combined together in different colors, allowing a flexible experimental design.

We screened a total of >400 genes using HeLa cells stably expressing a Centrin B1-GFP fusion to label the centrosome. High-throughput spinning-disk microscopy was used to acquire full 3D images at high resolution (200 nm lateral and 600 nm axial), and two sets of images were recorded: one to image interphase cells and another for mitosis (see Material and Methods). RNA localization during interphase and mitosis was assessed visually. These analyses yielded 6 candidate mRNAs and their localization to centrosomes was then confirmed by performing low-throughput smFISH. The results confirmed that
these mRNAs accumulated at centrosomes during interphase and/or mitosis. They included PCNT and NUMA1 mRNA that also were recently identified by us and others (Sepulveda et al., 2018; Chouaib et al., 2019), as well as several new ones: NIN, BICD2, CDC88C and CEP350 (Figure 1B, and see below). With ASPM and HMMR that we also identified previously (Chouaib et al., 2019), a total 8 mRNAs are thus known to localize at centrosomes in human cells. These transcripts encode proteins that regulate various aspects of centrosome maturation, spindle positioning, and MT dynamics. Interestingly, the localization of these mRNAs varied during the cell cycle. PCNT, NIN, BICD2, CDC88C, HMMR and CEP350 mRNAs localized to centrosomes during interphase and early mitosis, and delocalized at later mitotic stages (Figure 1C, and see below). In contrast, NUMA1 and ASPM mRNAs only localized during mitosis (Chouaib et al., 2019; see below). Centrosomes thus have a dedicated spatio-temporal translational program that is regulated during the cell cycle.

Expression and localization of ASPM, NUMA1 and HMMR proteins during the cell cycle

We then concentrated on ASPM, NUMA1 and HMMR and analyzed in more detail their expression and localization pattern. We first analyzed the expression of the respective proteins during the cell cycle. For this, we took advantage of HeLa Kyoto cell lines that stably expressed bacterial artificial chromosomes (BAC) containing the entire genomic sequences of the genes of interest, and also carrying a C-terminal GFP tag (Chouaib et al., 2019). These BACs contain all the gene regulatory sequences. Consequently, they are expressed at near-endogenous levels and with the proper spatio-temporal pattern (Poser et al., 2008). Time-lapse microscopy of single cells revealed that ASPM-GFP expression raised continuously during the cell cycle to culminate just before mitosis, while that of
NUMA1-GFP and HMMR-GFP was rather constant during the cell cycle. Interestingly, ASPM-GFP and NUMA1-GFP had similar localization patterns. Both proteins were nucleoplasmic during interphase and precisely initiated centrosomal localization in prophase. During cell division, they accumulated at the spindle pole with a weak labeling of the proximal spindle fibers (Figure S1A-B, Movies 1-2). In contrast, HMMR-GFP accumulated on the entire spindle throughout mitosis and additionally concentrated on the cytokinetic bridge in telophase. During interphase, it labeled microtubules (MTs) and localized to centrosomes (Figure S1C, Movie 3).

**ASPM, NUMA1, and HMMR mRNAs have unique spatio-temporal localization dynamics at mitotic centrosomes**

We next determined the localization of ASPM, NUMA1 and HMMR mRNAs during the different mitotic phases. We again used the GFP-tagged BAC HeLa cell lines to correlate protein and mRNA localization. Single molecule FISH (smFISH) was performed against the GFP RNA sequence using a set of 44 Cy3 labeled oligonucleotide probes. In interphase, ASPM-GFP and NUMA1-GFP mRNAs did not localize to centrosomes as previously reported, while HMMR-GFP mRNAs localized to centrosomes when its protein accumulated there. During mitosis, ASPM-GFP mRNAs were enriched on mitotic centrosomes across all phases of cell division, where it co-localized with its protein (Figure 2A and 2B). In contrast, NUMA1 and HMMR mRNAs only accumulated at centrosomes during the early stages of cell division, prophase and prometaphase, and a random mRNA distribution was seen during metaphase and anaphase, although both proteins still remained on the mitotic spindle (Figure 2C and 2D). Unlike NUMA1, the centrosomal localization of HMMR mRNA was re-established in telophase. It also accumulated at the cytokinetic bridges also labelled by HMMR-GFP (Figure 2E and 2F).
To detail these findings, we performed two-color smFISH experiments detecting one BAC-GFP mRNA in Cy3 and an endogenous mRNA in Cy5. We analyzed all pairwise combinations of ASPM, NUMA1 and HMMR mRNAs. To gain more precision, we divided prophase, prometaphase and telophase into two sub-phases, early and late (see Materials and Methods). During early prophase, NUMA1 and HMMR mRNAs could be seen on centrosomes but not ASPM mRNAs for which the transcript only joined during late prophase (Figure S2). During early prometaphase, all three mRNAs were enriched on centrosomes. However, the centrosomal localization of NUMA1 and HMMR mRNAs then became much less frequent, starting at late prometaphase, while that of ASPM mRNA could still be observed in metaphase and anaphase (Figure S2 and S3). Finally, ASPM but not HMMR mRNAs accumulated on centrosomes during early telophase, while the opposite was observed at late telophase. Interestingly, the three mRNAs also never perfectly co-localized on centrosomes at any of the cell-cycle stages: certain peri-centrosomal regions were occupied by one transcript, and others by the other mRNAs (see Figures S2 and S3). Taken together, these data demonstrate a finely regulated spatio-temporal dynamics for the centrosomal localization of ASPM, NUMA1, and HMMR mRNAs, with each mRNA localizing at a specific stage and place during cell division.

The localization of all 8 centrosomal mRNAs is inhibited by puromycin but not by cycloheximide

Next, we analyzed the localization mechanism of these mRNAs and first questioned whether localization requires translation. To this end, we used a HeLa cell line expressing CentrinB1-GFP to label centrosomes and treated it for 10 minutes with either cycloheximide, which blocks ribosome elongation, or puromycin that induces chain termination. We first analyzed the mRNA localizing during interphase (NIN, BICD2,
CCDC88C, CEP350, HMMR and PNCT) and remarkably, these six mRNAs became delocalized after puromycin treatment while cycloheximide had no effect (Figure 3A and 3B). Likewise, a 5-minute puromycin treatment inhibited the centrosomal localization of ASPM, NUMA1 and HMMR mRNAs at all the mitotic phases in which these mRNA normally localize (Figures S4, S5, and S6). Again, cycloheximide had no effect. Since this inhibitor inhibits translation but leaves the nascent peptide chain on ribosomes, while puromycin removes it, our data suggests that mRNA localization to centrosomes requires the nascent peptide. RNA localization is thus expected to occur co-translationally for all the 8 mRNAs, pointing toward a common localization mechanism.

The 5’ and 3’ UTRs of ASPM mRNA are not required for its centrosomal localization
To investigate how mRNAs localize to centrosomes in more details, we focused on ASPM and asked whether the 5’ and 3’ UTRs were necessary for its localization. To this end, we fused a full-length mouse cDNA to GFP and expressed the construct via transient transfection in HeLa Kyoto cells. To detect mRNAs produced from this reporter only, we performed smFISH with probes directed against the GFP RNA sequence. Mitotic cells expressing the plasmid could be identified by the accumulation of GFP-ASPM, which localized on centrosomes and the mitotic spindle. Interestingly, we could detect ASPM-GFP mRNAs on mitotic centrosomes in most of the transfected cells (Figure 4). This demonstrated that the 5’ and 3’ UTRs of ASPM mRNA are not required for its centrosomal enrichment, further supporting a role for the nascent protein in the localization mechanism.

Visualizing ASPM mRNA reveals directed transport toward centrosomes and anchoring on the mitotic spindle
To gain more insights into the localization mechanism, we imaged the endogenous ASPM mRNAs in living mitotic cells. To this end, we tagged the 3’UTR of ASPM mRNA with 24 MS2 repeats in HeLa Kyoto cells, using CRISPR/Cas9 mediated homologous recombination (Figure 5A). Heterozygous clones were confirmed by genotyping (Figure S7A). Moreover, two-color smFISH performed with MS2 and ASPM probes showed that tag-containing single molecules co-localized with endogenous ASPM mRNA molecules. The tagged mRNA accumulated at centrosomes in mitosis (Figure 5B), indicating that the MS2 sequences did not interfere with localization. Next, we stably expressed low levels of the MS2-coat protein fused to GFP and a nuclear localization signal (MCP-GFP-NLS). This fusion protein binds the MS2 repeat and allows to visualize the tagged RNA in live cells (Bertrand et al., 1998). Indeed, mitotic cells expressing MCP-GFP-NLS displayed diffraction limited fluorescent spots that localized near the centrosomes. These spots co-localized with single RNA molecules revealed with probes against either endogenous ASPM mRNA or the MS2 tag, indicating that binding of MCP-GFP-NLS to the tagged mRNA did not abolish its localization to the centrosome (Figure S7B). This demonstrated the feasibility of this approach for live imaging of ASPM mRNA at the centrosome.

Next, we labeled DNA using SiR-DNA to identify the mitotic phase and we imaged live cells in 3D at a rate of 2-4 fps using spinning disk microscopy. During prophase and prometaphase, three populations of ASPM mRNA molecules were observed: (i) immobile molecules that corresponded to mRNAs at the centrosome, (ii) mRNAs diffusing in the cytosolic space, and (iii) mRNAs undergoing directed movements towards the centrosome (Figure 5C, Movie 4). Cell thickness yielded low signal-to-noise ratios and, combined with the rapid movements of the mRNAs, made single particle tracking with automated software difficult. We thus manually tracked mRNA molecules undergoing directed movements and this showed that they move at speeds ranging from 0.5 to 1
m/sec (Figure 5D), which is compatible with dynein-mediated transport (Schiavo et al., 2013).

In metaphase and anaphase where the mRNA is at centrosomes, we expected several possibilities for the movement of mRNA molecules: (i) stable anchoring to the centrosome; (ii) diffusion within a confined space around the centrosome; or (iii) diffusion away from centrosomes and re-localization by a motor-dependent mechanism. Live imaging revealed that ASPM mRNA localizing at mitotic centrosome did not diffuse and were immobile (Figure 5E, Movie 5). Interestingly, we also observed that some ASPM mRNAs were attached on the spindle fibers rather than on the spindle poles (Movie 6). Taken together, these live imaging experiments demonstrated that ASPM mRNAs are actively transported to the mitotic centrosomes and are then anchored on the spindle poles and fibers.

**ASPM polysomes are actively transported towards centrosomes during prophase and prometaphase**

Since the localization of ASPM mRNAs is abolished by puromycin and might thus involve the transport of translated mRNAs, we next aimed at imaging ASPM polysomes. To this end, we used the SunTag system that allows to image nascent polypeptide chains (Pichon et al., 2016). It is composed of a repeated epitope tag inserted in the protein of interest and a monochain antibody that is fused to GFP. Binding of the fluorescent antibody to the epitope occurs when it emerges from the ribosome and allows to visualize nascent chains and polysomes in live cells. We engineered a HeLa Kyoto cell line with 32 SunTag repeats fused to the 5’ end of the ASPM gene using CRISPR/Cas9 and homologous recombination (Figure 6A). The cells were then transduced with a lentivirus expressing the scFv monochain antibody fused to GFP (scFv-sfGFP). Bright GFP foci were observed and
confirmed to be polysomes based on both their sensitivity to puromycin (Figure S8A) and co-localization with endogenous ASPM mRNA by smiFISH (Figure 6B). Heterozygous clones were confirmed by genotyping (Figure S8B).

We first imaged SunTag-ASPM polysomes and mRNA in fixed cells. In early prophase, most mRNAs and polysomes did not localize to centrosome, in agreement with the smFISH data (Figure S2). In metaphase and anaphase, the accumulation of the SunTag-ASPM mature protein at the spindle poles prevented visualizing ASPM polysomes at this location. However, some ASPM polysomes were observed outside the spindle area indicating that translation was pursued during entire the mitosis (Figure 8B). At the end of telophase, ASPM mRNAs could be seen translated on the nuclear pores as previously reported for interphase (Chouaib et al., 2019). We then performed live imaging in 3D and with high acquisition rates (around 1-1.3 3D stacks per second), using spinning disk microscopy. We first imaged cells in prophase. Remarkably, while many ASPM polysomes did not localize at centrosomes at the beginning of the movie, they displayed rapid directed motions towards this organelle, leading to their accumulation at this location (Movie 7). Single particle tracking (SPT) confirmed that an average of 53% of ASPM polysomes displayed such directed movements in early prophase (Figure 6C-E). Directed movements were also detected in prometaphase but to a much lesser extent (Figure S8C, S8D, and Movie 8). Calculating the average velocity of polysomes undergoing directed movements showed that their speed ranged from 0.25 to 1 μm/s, which is compatible with motor-dependent transport (Figure 6F). Taken together, this data proved that ASPM polysomes were actively transported to the centrosome at the onset of mitosis.
ASPM mRNAs are translated on MTs during interphase

Similar to MS2-tagged ASPM mRNAs, some SunTag-ASPM polysomes did not co-localized with the spindle poles but rather with spindle fibers that were weakly labeled by the mature SunTag-ASPM protein (Figure 6B, arrows). This prompted us to investigate in more details the role of MTs in the metabolism of ASPM mRNAs. We labeled MTs in living cells using a far-red dye (SiR-Tubulin) and performed sequential two-color live imaging in 3D using a spinning disk microscope. We first imaged interphase cells and remarkably, we observed that many ASPM polysomes remained stably anchored to MTs during the course of the movies (66%; Figure 7A-C, Movie 9). In addition, we also observed directed motion of single polysomes, albeit at a low frequency (around 4%). We then characterized in more details the movements of ASPM polysomes and for this we first classified them in four categories: (i) polysomes localizing on MTs; (ii) not localizing on MTs; (iii) anchored at the nuclear envelope (as previously reported; Chouaib et al., 2019); (iv) showing directed transport (Figure 7C). The histogram of displacements between consecutive frames revealed a diffusion coefficient of 0.011µm²/s for polysomes on MTs, 0.041 µm²/s for those not on MTs, 0.004 µm²/s for the ones on nuclear pores and 0.019 µm²/s for the ones displaying directed motion (note that in this latter case the directed motion is only a fraction of the entire trajectory; Figure S9). We also calculated the mean square displacement (MSD) as a function of time (Figure 8D). This confirmed that ASPM polysomes not bound to MTs diffused faster than those bound to MTs or to the nuclear pore, respectively. As a control, we depolymerized MTs with a 10-minute nocodazole treatment before starting imaging. SiR-tubulin labeling confirmed the absence of MTs in treated cells (Figure S10A, Movie 10), and we then tracked ASPM polysomes excluding the ones attached to nuclear pores (Figure S10A and S10B). The histogram of displacements and MSD curve revealed a single population with a diffusion coefficient of
0.035µm²/s, similar to polysomes not on MT in untreated cells (Figure S10C and S10D). Overall, this showed that a large fraction of ASPM mRNAs are locally translated on MTs during interphase, and that these polysomes are stably anchored to MTs, thereby limiting their diffusion.

**ASPM polysomes are transported to mitotic centrosomes by either sliding on MTs or being pulled with entire MTs**

To assess whether MTs are necessary for ASPM mRNA localization to centrosomes, we combined a brief nocodazole treatment (10 minutes) with smFISH, using the ASPM SunTag clone. A 10-minute treatment depolymerized MTs but centrosomes were still visible (Figure S10A). ASPM mRNAs no longer accumulated at mitotic centrosomes after depolymerizing MTs, despite the fact that the mRNAs were still translated (Figure S10E-G). This indicated that intact MTs are required for ASPM mRNA localization.

We then performed dual color imaging of MTs and ASPM polysomes during mitosis. Tracks were shorter than in the mono-color movies because the Z-stacks only had 3 planes in the two-color experiments (as opposed to 15-20 planes in the single color movies). Nevertheless, this allowed us to distinguish two types of movements towards centrosomes. In the first, an ASPM polysome rapidly slide along an immobile MT (Figure 7E, Movie 11). This likely corresponded to a motor driven movement along a MT cable. In the second type of movements, both the MT and the attached ASPM polysome moved together and are pulled towards the centrosome (Figure 7F, Movie 12). In this case, the MT appears being hauled towards the centrosome and drags a tethered polysome with it. Also note that MTs pulling forces have been described to affect centrosome and spindle positioning and function during mitosis (Enos et al., 2018; Forth and Kapoor, 2017). This
demonstrated that ASPM polysomes are trafficked to the mitotic centrosome via two mechanisms: sliding on MT, and MT remodeling.
Discussion

Here, we studied centrosomal mRNA localization in human cells. We uncovered a tight temporal and spatial control of mRNA trafficking to centrosomes, particularly during mitosis, and we provide definitive evidence for a co-translational transport of polysomes by motors and MTs.

A targeted smFISH screen reveals that centrosomal mRNAs are conserved from human to Drosophila

We screened >400 genes encoding centrosomal proteins by smFISH in HeLa cells and identified four new cases of mRNA localization at the centrosome (Supplementary Table S2). In total, eight mRNAs now belong to this class: NIN, PCNT, CEP350, BICD2, CCDC88C, ASPM, NUMA1 and HMMR. All the corresponding proteins localize to the centrosome suggesting that their mRNAs are locally translated. Most of them also perform important centrosomal functions. NIN is localized to the sub-distal appendage of mother centrosomes and it functions in microtubule nucleation as well as centrosome maturation (Ou et al., 2002; Stillwell et al., 2004). CEP350 is also localized to sub-distal appendages and it is important for centriole assembly and MT anchoring to centrosomes (Yan et al., 2006; Le Clech, 2008). Pericentrin (PCNT) is a major component of the pericentriolar material (PCM) and it plays a structural role by bridging the centrioles to the PCM (Lee and Rhee, 2011). BICD2 contributes to centrosomal positioning and to centrosomal separation at the onset of mitosis (Raaijmakers et al., 2012; Splinter et al., 2010). It also has non-centrosomal functions by acting as an adaptor linking the dynein motor complex to its cargo (Splinter et al., 2012). ASPM and NUMA1 are two MT minus end binding proteins that accumulate at centrosomes during mitosis, and they control several aspects of spindle assembly and function (Jiang et al., 2017; Seldin et al., 2016). HMMR acts to...
separate centrosome and to nucleate MT during spindle assembly. It also modulate the cortical localization of NUMA-dynein complexes to correct mispositioned spindles (Connell et al., 2017). The diversity of functions performed by these proteins suggests that RNA localization and local translation play an important role for the centrosome.

Centrosomes may contain additional mRNAs. Multiplexed smFISH recently reported the detection of 10,000 mRNAs in mouse NIH3T3 cells (Eng et al., 2019). A spatial correlation analysis defined a set of 9 perinuclear mRNAs, called subcluster3, which contained a few mRNAs shown here to localize at centrosomes. We therefore tested whether the other mRNAs of this cluster also localized and found that only TRIP11 mRNA is slightly enriched at the centrosomes of HeLa cells (data not shown). We also analyzed the localization of Cyclin B1 mRNA, which localizes to the mitotic apparatus in *Xenopus* embryos (Groisman et al., 2000). This mRNA was however randomly distributed in HeLa cells during both interphase and mitosis.

Remarkably however, a comparison with previous screens done in Drosophila embryos revealed that mRNA localization at centrosomes is well conserved cross species (Supplementary Table S3; Lécuyer et al., 2007; Wilk et al., 2016). Indeed, The orthologs of CCDC88C and Pericentrin are Girdin and Plp respectively, and both the mRNAs and proteins localize to *Drosophila* centrosomes. Furthermore, the orthologs of NIN and BICD2 are BsgD25 and BICD, whose mRNAs are annotated as "cell division apparatus" in the *Drosophila* screens. The orthologs of NUMA1 and ASPM (Mud and Asp, respectively), were not tested before. However, smFISH in Drosophila S2 cells revealed that while Asp mRNA is randomly distributed, Mud mRNA is indeed localized to centrosomes during mitosis (data not shown). Therefore, five out of the eight human centrosomal mRNAs are also localized in *Drosophila*, pointing to a striking and unique degree of evolutionary
conservation in RNA localization. This likely underlies conserved features in mRNA localization mechanism and/or function.

**A cell-cycle dependent translational program operates at centrosomes**

Our data reveal that centrosomal mRNA localization vary with the cell cycle. Six mRNAs localize in interphase (HMMR, BICD2, CEP350, PCNT, NIN, CCDC88C), while all eight localize in prophase. Expansion of the pericentriolar material occurs during G2 and early mitosis and this is essential for the subsequent steps of cell division. Pericentrin is required for PCM expansion. Locally producing Pericentrin at centrosomes specifically during late G2/early mitosis could provide a platform for the rapid recruitment of other PCM components, such as the γ-TuRC complex that binds the N-terminal domain of Pericentrin. Two mRNAs only localized during mitosis. ASPM mRNA localized to centrosomes across all mitotic phases while NUMA1 only during prophase and prometaphase. This is in agreement with the function of these proteins, which are MT minus-end binding proteins that control spindle positioning and MT dynamics during mitosis (Jiang et al., 2017; Seldin et al., 2016). Finally, HMMR was the only transcript that localized at the cytokinetic bridge at the end of cell division, together with its protein. This shows the variety, complexity and precision of centrosomal mRNA localization, as well as its potential role during the centrosome cycle. These data demonstrate the existence of a unique translational program at centrosomes, which is cell cycle regulated.

It is interesting to speculate why these eight proteins and not others are locally translated. Since they function in centrosome/spindle maturation and that this occurs over short time periods, having optimal amounts at centrosomes during specific time points of the cell cycle is crucial. Interestingly, these proteins have relatively large sizes.
(more than 2000 aa, with the exception of HMMR and BICD2), and it would thus take some
time to synthesize them. Therefore, a local translational regulation may provide an
efficient and rapid method for targeting them to the centrosome when needed. Another
possibility is that mRNA accumulation at centrosomes plays a structural role. An
emerging model is that phase separation helps the formation of centrosomes (Woodruff
et al., 2018). Since RNAs are often critical components of phase separated condensates
(Navarro et al., 2019), their accumulation at centrosomes could contribute to their
formation. Finally, a likely and not exclusive possibility is that these eight proteins need
to be assembled co-translationally with their partners at the centrosome. In this regard,
we noted that they share some sequence similarities. For instance, NUMA1 has some
homologies with Pericentrin and CCDC88C, while its Drosophila homolog Mud
additionally displays homology with HMMR. Moreover, this link extends beyond
centrosomal localization: NUMA1 also shares homology with AKAP9 and GOLGA4, two
Golgi proteins, and we observed that these two mRNAs localize to this compartment
(Chouaib et al. 2019; see the present screen for GOLGA4). Conversely, AKAP9 has
sequence homologies with Pericentrin, CCDC88C and Ninein, whose mRNAs all localize to
centrosomes. These remarkable sequence homologies appear to arise in so-called "SMC
domains", suggesting a link of this domain with local translation. One hypothesis would
be that this domain requires to fold and assemble with its partners in a co-translational
manner.

**Centrosomal mRNA localization occurs by active transport of polysomes and
requires the nascent protein**

For all the eight centrosomal transcripts studied here, premature termination of the
ribosomes delocalized the mRNAs while freezing the ribosome and the nascent protein
chain on the mRNA had no effect. In the case of ASPM mRNA, we further observed that its untranslated regions were not needed for its centrosomal localization. Moreover, ASPM polysomes were both transported to centrosomes by molecular motors. Together, this shows that centrosomal mRNA localization is an active mechanism driven by the nascent peptide. While RNA localization is often conceptualized as an RNA-driven transport-process, our data contradict this dogma and shows that for centrosomal mRNAs, polysome transport dependent on the nascent protein is the rule.

This suggests the nascent peptide contains a localization signal that would drive the polysome toward centrosomes. ASPM polysomes are actively transported towards centrosomes via two mechanisms. The first is motorized transport whereby a polysome slides on MTs. The second involves the pulling of an entire MT with an-ASPM polysome attached to it. The N-terminal part of ASPM contains two calponin-homology domains that can bind MTs. It is likely that once these domains are translated, they bind MTs and cause the entire ASPM polysome to attach. Interestingly, local translation of ASPM at MTs can also be seen during interphase. The purpose of this is unclear, but one can speculate that ASPM mRNAs preemptively attach to MTs in anticipation for mitotic entry, to then maximize the efficiency of centrosomal targeting. Interestingly, all eight localized mRNAs encode proteins that either bind MTs directly (ASPM, NUMA1, HMMR, CCDC88C, CEP350) or contribute to MT anchoring (NIN, PCNT, BICD2). This property could thus be part of the transport mechanism.

The paradigm for protein-dependent RNA localization is that of secreted proteins. In this case, translation starts and leads to the synthesis of the signal peptide, which is recognized by SRP. This halts the ribosome until the entire complex docks on the SRP receptor on the ER, where translation resumes. It is thus tempting to envision a scenario...
where ribosomes translating centrosomal mRNAs enter a pause and only resume translation after reaching the centrosome. It has recently been shown that unfolded domains can halt ribosomes (Liu et al., 2013; Shalgi et al., 2013). Moreover, in at least one case of co-translational assembly, the ribosome enters a pause at a specific location and this pause is relieved upon interaction with the partner of the nascent protein (Panasenko et al., 2019). If indeed the proteins encoded by the centrosomal mRNAs need to be co-translationally assembled, it is likely that the domain responsible for this would remain unfolded before the polysome reaches the centrosome. It could thus halt the ribosomes, while the nascent polypeptide located upstream of this unfolded domain could connect the polysome to transport systems and drive it to the centrosome. If true, this could be an elegant and general mechanism that ensures RNA localization and local translation, as it could work at any place in the cell. This could explain why co-translational mRNA targeting appears to be a widespread mechanism in cell lines (Chouaib et al., 2019).
References


Materials and methods

Cell lines, culture conditions and treatments
The collection of HeLa-Kyoto stably transfected cell lines with the GFP-tagged BACs was described previously (Poser et al., 2008, Maliga et al., 2013). They were maintained in Dulbecco’s modified Eagle’s Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich), 100 U/mL penicillin/streptomycin (Sigma-Aldrich) and 400 µg/ml G418 (Gibco). Parental HeLa-Kyoto cells, WT HeLa cells and HeLa cells expressing centrin B1-GFP (a gift from Dr. B. Delaval) were grown in the same medium without G418. Drugs were used at the following final concentrations: 100 µg/ml for puromycin, 200 µg/ml for cycloheximide, and 5 µg/ml for nocodazole. Treatment of cells with translation inhibitors was for 20 mins (reduced to 5 min when mitotic cells were needed). Treatment of cells with nocodazole was for 10 minutes. Transfection of the GFP-ASPM cDNA was done using JetPrime (Polyplus) and 2 µgs of DNA were transfected overnight in a 6 well plate containing 2ml of medium.

Insertion of the MS2 cassette by CRISPR/Cas9
The recombination cassettes contained 500 bases of homology arms flanking a 3xHA tag, a stop codon, followed by 24 MS2 repeats. A start codon was placed after the MS2 repeats followed by an IRES, a neomycin resistance gene, and a stop codon. The IRES-Neo segment was flanked by two LoxP sites having the same orientation. HeLa Kyoto cells were transfected using JetPrime (Polyplus) and a cocktail of four plasmids, including the recombination cassette and constructs expressing Cas9-nickase and two guide RNAs with an optimized scaffold. Insertion was targeted at the stop codon of the ASPM gene. Cells were selected on 400 µg/ml G418 neomycin for a few weeks. Individual clones were then
picked and analyzed by PCR genotyping, fluorescent microscopy and smFISH/smiFISH with probes against both the endogenous ASPM mRNA and MS2 sequences. Stable MCP-GFP-NLS expression was then set up via retroviral infection. The sequences targeted by the guide RNAs were (PAM sequences are underlined): TCTCTTCAAAAACCAAATCTGG for ASPM guide 1 and GCAAGCTATTCAAATGGTGATGG for ASPM guide 2.

**Insertion of SunTag cassette by CRISPR/Cas9**

The recombination cassettes contained 500 bases of homology arms flanking a puromycin resistance gene translated from the endogenous ATG sequence, followed by a P2A sequence, 32 SunTag repeats, and a P2A-T2A sequence fused to the protein of interest. Hela Kyoto cells stably expressing the scFv-GFP were transfected using JetPrime (Polyplus) and a cocktail of three plasmids, including the recombination cassette, and constructs expressing Cas9-HF1 and guide RNAs with an optimized scaffold. Cells were selected on 0.25 µg/ml puromycin for a few weeks. Individual clones were then picked and analyzed by PCR genotyping, fluorescent microscopy and smiFISH with probes against the SunTag sequence. The sequences targeted by the guide RNAs were (PAM sequences are underlined): AAGTGAGCCCGACCGAGGGAGG for ASPM

**Genotyping**

PCR was done using a Platinum Taq DNA Polymerase (Invitrogen) on genomic DNA prepared with GenElute Mammalian Genomic DNA miniprep (Sigma-Aldrich). The sequences of oligonucleotides were: 5’-TCAGAGGGTATGGAGGGGAA-3’ (ASPM gene end WT forward), 5’-GACATCTGTGGCCCTGAAAC-3’ (ASPM gene end WT reverse), and 5’-GCCCTCACATTGCCAAAAGA-3’ (IRES reverse) for genotyping ASPM-MS2x24 clones; and 5’-TGTTCCTGGAAACCGCAATG-3’ (ASPM gene start WT forward); 5’-
GTTTATGTGTTGCTCCCGCC-3’ (ASPM gene start WT reverse); 5’-
GGTGACCCGCTCGATGTG-3’ (puro\textsuperscript{r} reverse) for genotyping SunTag-ASPM clones.

**Single molecule fluorescent in situ hybridization**

Cells grown on glass coverslips or 96-well glass bottom plates (SensoPlates, Greiner) were fixed for 20 min at RT with 4% paraformaldehyde (Electron Microscopy Sciences) diluted in PBS (Invitrogen), and permeabilized with 70% ethanol overnight at 4°C.

For smFISH performed on BAC cells, we used a set of 44 amino-modified oligonucleotide probes against the GFP-IRES-Neo sequence present in the BAC construction (sequences given in Table S1). Each oligonucleotide probe contained 4 primary amines that were conjugated to Cy3 using the Mono-Reactive Dye Pack (PA23001, GE Healthcare Life Sciences). To this end, the oligos were precipitated with ethanol and resuspended in water. For labelling, 4 µg of each probe was incubated with 6 µl of Cy3 (1/5 of a vial resuspended in 30 µl of DMSO), and 14 µl of carbonate buffer 0.1 M pH 8.8, overnight at RT and in the dark, after extensive vortexing. The next day, 10 µg of yeast tRNAs (Sigma-Aldrich) were added and the probes were precipitated several times with ethanol until the supernatant lost its pink color. For hybridization, fixed cells were washed with PBS and hybridization buffer (15% formamide (Sigma-Aldrich) in 1X SSC), and then incubated overnight at 37°C in the hybridization buffer also containing 130 ng of the probe set for 100 µl of final volume, 0.34 mg/ml tRNA, 2 mM VRC (Sigma-Aldrich), 0.2 mg/ml RNase-free BSA (Roche Diagnostics), and 10% Dextran sulfate (MP Biomedicals). The next day, the samples were washed twice for 30 minutes in the hybridization buffer at 37°C, and rinsed in PBS. Coverslips were then mounted using Vectashield containing DAPI (Vector laboratories, Inc.). For smFISH against the MS2 tag,
25 ng of an oligonucleotide labeled by two Cy3 molecules at the first and last thymidine (sequence in table S1) was used per 100 µl of hybridization mix.

For smiFISH using DNA probes (Tsanov et al., 2016), 48 unlabeled primary probes were used (sequences given in Table S1). In addition to hybridizing to their targets, these probes contained a FLAP sequence that was pre-hybridized to a secondary fluorescent oligonucleotide. To this end, 40 pmol of primary probes were pre-hybridized to 50 pmol of secondary probe in 10 µl of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, pH 7.9. Pre-hybridization was performed on a thermocycler with the following program: 85°C for 3 min, 65°C for 3 min, and 25°C for 5 min. The final hybridization mixture contained the probe duplexes (2 µl per 100 µl of final volume), with 1X SSC, 0.34 mg/ml tRNA, 15% Formamide, 2 mM VRC, 0.2 mg/ml RNase-free BSA, 10% Dextran sulfate. Slides were then processed as above.

For smiFISH using RNA probes, a pool of DNA oligonucleotides (GenScript) was used to generate the primary probes. The oligonucleotide design was based on the Oligostan script (Tsanov et al., 2016). A first PCR (25 cycles) was performed on the oligopool with primers corresponding to the couple of barcodes (Xu et al., 2009) specific of each target mRNA. A second PCR (13 cycles) was achieved using the following primers: FLAP Y sequence with the addition of the T7 RNA polymerase promoter sequence at its 5’ end (5’-TAATACGACTCACTATAGGGTTACACTCGGACCTCGTCGACATGCATT-3’), and reverse complement sequence of FLAP X (5’-CACTGAGTCACCAGCTCGAAACTTAGGAGG-3’). All PCR reactions were carried out with Phusion DNA Polymerase (Thermo Fisher Scientific), in 96-well plates, with a Freedom EVO 200 (Tecan) robotic platform. PCR products were checked by capillary electrophoresis on a Caliper LabChip GX analyzer (PerkinElmer). The products of the second PCR were then purified with a NucleoSpin 96 PCR Clean-up kit (Macherey-Nagel), lyophilized, and resuspended in DNase/RNase-free
distilled water (Invitrogen). In vitro transcription was subsequently performed with T7 RNA Polymerase (New England BioLabs) and the obtained primary probes were analyzed by capillary electrophoresis using a Fragment Analyzer instrument (Advanced Analytical). 50 ng of primary probes (total amount of the pool of probes) and 25 ng of each of the secondary probes (TYE 563 labeled LNA oligonucleotides targeting FLAP X and FLAP Y, Qiagen) were pre-hybridized in 100 μL of the following buffer: 1X SSC, 7.5 M urea (Sigma-Aldrich), 0.34 mg/mL tRNA, 10% Dextran sulfate. Pre-hybridization was performed on a thermocycler with the following program: 90°C for 3 min, 53°C for 15 min. Plates with fixed cells were washed with PBS and hybridization buffer (1X SSC, 7.5 M urea). Hybridization was then carried out overnight at 48°C. The next day, plates were washed with hybridization buffer at 48°C. Cells were rinsed with PBS at RT, stained with 1 μg/mL Dapi diluted in PBS, and washed with PBS. The following mounting medium was used: 90% glycerol (VWR), 1 mg/mL p-Phenylenediamine (Sigma-Aldrich), PBS pH 8.

**Imaging of fixed cells**

Microscopy slides were imaged on a Zeiss Axioimager Z1 wide-field microscope equipped with a motorized stage, a camera scMOS ZYLA 4.2 MP, using a 63x or 100x objective (Plan Apochromat; 1.4 NA; oil). Images were taken as z-stacks with one plane every 0.3 μm. The microscope was controlled by MetaMorph and figures were constructed using ImageJ, Adobe Photoshop and Illustrator.

96-well plates were imaged on an Opera Phenix High-Content Screening System (PerkinElmer), with a 63x water-immersion objective (NA 1.15). Three-dimensional images were acquired, consisting of 35 slices with a spacing of 0.3 μm.
Deconvolution

For Figures 6A, S2, S3, and S8A; deconvolution was performed using the Huygens Professional software (Scientific Volume Imaging, Hilversum, Netherlands). The point spread function (PSF) was a theoretical one, and background values were manually estimated. The deconvolution was done using the CMLE deconvolution algorithm, with 50 iterations and a quality threshold of 0.01, and without bleaching correction.

Image analysis and quantifications

Mitotic phases were identified based on visual inspection of DNA condensation and cell shape. Early prophase was defined with low DNA compaction that increases in late prophase. Early prometaphase was marked by the rupture of the nuclear envelope, while late prometaphase additionally displayed cell rounding. For late mitosis, we subdivided cells into early telophase (without cytokinesis), and late telophase (with cytokinesis marked by the accumulation of HMMR-GFP at cytokinetic bridges). Centrosomal localization was assessed by visual inspection of individual cells.

Imaging of live cells

Live imaging was done using a spinning disk confocal microscope (Nikon Ti with a Yokogawa CSU-X1 head) operated by the Andor iQ3 software. Acquisitions were performed using a 100X objective (CF1 PlanApo λ 1.45 NA oil), and an EMCCD iXon897 camera (Andor). For fast imaging, we imaged at a rate of at least 1 stack/s for 1-3 mins, using stacks with a Z-spacing of 0.5-0.6 µm. This spacing allowed accurate point spread function determination without excessive oversampling. For slow imaging, we collected stacks of 19 planes with a Z-spacing of 0.6 µm and at a frame rate of one stack every 5 mins for 62 hours. The power of illuminating light and the exposure time were set to the lowest
values that still allowed visualization of single molecules of proteins. This minimized bleaching and maximized the number of frames that was collected. Samples were sequentially excited at 488 and 640nm in the case of dual-color imaging. For mono-color SunTag ASPM movies, Z-stacks covering the entire cell were imaged. For dual color imaging of SunTag-ASPM and MTs, a Z-stack covering a section of the cell was imaged to maintain frame rates and compensate for the extra color. Cells were maintained in anti-bleaching live cell visualization medium (DMEMgfp; Evrogen), supplemented with 10% fetal bovine serum at 37°C in 5% CO₂. SiR-DNA (Spirochrome) was kept at 100 nM throughout the experiments to label DNA. SiR-tubulin (Spirochrome) was kept at 100 nM throughout the experiments to label MTs.

**Single particle tracking and analysis**

Single particle tracking of ASPM polysomes was performed using the TrackMate plugin in ImageJ. The DoG detector was used. Blob diameter was set to 0.7-0.8 microns and the detection threshold was between 100-120. Median filtering and sub-pixel localization were additionally used. The simple LAP tracker option was used to construct tracks. Both linking and gap closing distances were assigned a maximum value of 1.5 microns. Three frame gaps were allowed when constructing tracks. Tracks were displayed color-coded according to displacement (red corresponds to highest values while blue corresponds to lowest). For SunTag-ASPM mono-color movies, the top 20 tracks with highest displacements were chosen. The velocity was calculated by averaging displacement over 5 frames of directed motion. Directionality towards the centrosome was visually determined. For dual color movies of SunTag-ASPM and MTs, tracks shorter than 15 seconds were filtered out. Tracks were annotated across four groups based on their localization and dynamics after manual visual inspection.
Tracks were imported and analyzed in R. Instant 1D displacements between frames were calculated along the x and y axis, and the resulting histograms were fitted to a Gaussian function, for which variance is directly proportional to the diffusion coefficient ($D$). We also calculated a mean MSD as a function of time, by aligning all tracks at their start and averaging the resulting 2D displacements.
Figure legends:

**Figure 1: A screen identifies new mRNAs localizing to centrosomes.**

(A) Schematic representation of probe generation. BC: target-specific barcode, FLAP X and Y: shared sequences hybridized with TYE-563 labeled LNA oligonucleotides (shown in red), Hyb: Hybridization sequence specific of the target mRNA, T7 pro: T7 RNA polymerase promoter. (B) Images are micrographs of HeLa cells stably expressing centrin B1-GFP, and captured in interphase and mitosis. Left and red: Cy3 fluorescent signals corresponding to PCNT, CEP350, BICD2, NIN, or CCDC88C mRNA labeled by smFISH; middle and green: GFP signals corresponding to the centrin B1 protein. Blue: DNA stained with DAPI. Scale bars: 10 microns. (C) Histogram depicting the percentage of cells in each phase showing centrosomal localization of each mRNA (n=50 for interphase, and 25 for each mitotic phase). Data represent mean and standard deviation.

**Figure 2: Differential centrosomal localization of ASPM, NUMA1 and HMMR mRNAs across mitosis.**

(A) Images are micrographs of HeLa cells containing an ASPM-GFP BAC and captured at different phases of mitosis, as well as interphase. Top and red: Cy3 fluorescent signals corresponding to ASPM mRNAs labeled by smFISH; middle and green: GFP signals corresponding to the ASPM protein. Blue: DNA stained with DAPI. Scale bar: 10 microns. (B) Histogram depicting the percentage of cells in each phase showing centrosomal localization of ASPM mRNAs (n=25). Data represent mean and standard deviation. (C and D) Legend as in A and B, but for HeLa cells containing a NUMA1-GFP BAC. (E and F) Legend as in A and B, but for HeLa cells containing an HMMR-GFP BAC.
Figure 3: Translation initiation is required for the localization of all centrosomal mRNAs during interphase.

(A) Images are micrographs of HeLa cells during interphase expressing either centrin B1-GFP or an HMMR-GFP BAC and treated with either cycloheximide or puromycin. Left and red: Cy3 fluorescent signals corresponding to PCNT, CEP350, BICD2, NIN, CCDC88C, or HMMR mRNAs labeled by smFISH; middle and green: GFP signals corresponding to either centrin B1 or the HMMR protein. Blue: DNA stained with DAPI. Scale bars: 10 microns.

(B) Histogram depicting the percentage of cells showing centrosomal localization of each mRNA after both treatments and in untreated cells (n=50). Data represent mean and standard deviation.

Figure 4: ASPM untranslated regions are not required for centrosomal localization.

(A) Images are micrographs of HeLa cells transiently expressing GFP-ASPM cDNA captured at early mitosis. Left and red: Cy3 fluorescent signals corresponding to exogenous ASPM mRNAs labeled by smFISH; middle and green: GFP signals corresponding to the exogenous ASPM protein. Blue: DNA stained with DAPI. Scale bar: 10 microns. (B) Histogram depicting the percentage of mitotic cells showing centrosomal localization of exogenous ASPM mRNAs (n=20). Data represent mean and standard deviation.

Figure 5: ASPM mRNAs transition between diffusion, directed movements, and anchoring to spindle poles and fibers across mitosis.

(A) Schematic representation showing the insertion of 24 MS2 repeats at the end of the ASPM gene using CRISPR-Cas9 generating a heterozygous HeLa cell line. HA: human influenza hemagglutinin tag. Stop: stop codon. IRES: internal ribosome entry site. Neo^r:
neomycin resistance. UTR: untranslated region. PolyA: poly A tail. (B) Images are micrographs of a HeLa clone expressing endogenous ASPM-MS2x24 captured in mitosis. Upper and red: Cy3 fluorescent signals corresponding to MS2-tagged mRNAs labeled by smFISH; middle and green: Cy5 signals corresponding to the endogenous ASPM mRNA. Blue: DNA stained with DAPI. Scale bar: 10 microns. (C) Snapshots of the ASPM-MS2x24 clone expressing MCP-GFP-NLS captured live during prometaphase. In the main panels, the GFP signal is shown in green and corresponds to ASPM mRNAs labeled by the MS2-MCP-GFP-NLS. The Cy5 signal is shown in red and corresponds to DNA. Scale bar: 10 microns. Time is in seconds. The white circle indicates ASPM mRNAs with limited diffusion. Insets are zooms of the white-boxed region and show ASPM mRNA in black. White arrowheads indicate the starting position of two mRNA molecules, while blue and yellow ones follow their current position. Scale bar in the inset: 1 micron. (D) Dot plot showing the average speed of ASPM mRNA molecules over 5 frames or 2.5 seconds of directed motion. The horizontal line represents the mean. (E) Legend same as in E, but captured during metaphase.

Figure 6: ASPM polysomes show directed movements towards the centrosome in early mitosis.

(A) Schematic representation showing a cassette containing 32 SunTag repeats that was inserted at the beginning of the endogenous ASPM gene. Puro\textsuperscript{r}: puromycin resistance. T2A, P2A: self cleavage signals. FLAG: octapeptide FLAG tag. (B) Images are micrographs of a HeLa clone expressing endogenous SunTagx32-ASPM and scFv-sfGFP captured at interphase and mitosis. Upper and red: Cy3 fluorescent signals corresponding to ASPM mRNAs labeled by smFISH; middle and green: GFP signals corresponding to ASPM polysomes and mature protein. Blue: DNA stained with DAPI. Scale bar: 10 microns. Red
and green arrowheads indicate ASPM mRNA and polysomes respectively. White arrows indicate the overlay of red and green arrows. Insets represent zooms of the white-boxed areas and contain a scale bar of 1 micron. (C) Snapshots of the SunTagx32-ASPM and scFv-sfGFP clone captured live during prophase. The GFP/SunTag signal is shown in black and corresponds to ASPM polysomes and mature proteins. Scale bar: 10 microns. Time is in seconds. The yellow circle indicates ASPM mature protein. Red arrowheads indicate the starting position of an mRNA molecule, while the blue one follows its current position. (D) A TrackMate overlay of the same cell in C showing polysomes tracks. Color code represents displacement (dark blue lowest, red highest). The outer yellow outline represents the cell border, while the inner one represents centrosomes marked by mature ASPM protein. Scale bar: 10 microns. (E) A histogram showing the percentage of polysomes showing directed movements towards the centrosome during prophase. Last column represents the mean and standard deviation. (F) Dot plot showing the average speed of ASPM polysomes over 5 frames or 4.5 seconds of directed motion in prophase cells. The horizontal line represents the mean.

**Figure 7:** ASPM polysomes are anchored to microtubules in interphase and dragged towards the centrosome in early mitosis.

(A) Snapshots of the SunTagx32-ASPM clone expressing scFv-sfGFP captured live during interphase with labeled MTs. The GFP/SunTag signal is shown in green and corresponds to ASPM polysomes and mature proteins; the Cy5 signal is shown in red and corresponds to a tubulin staining. Scale bar: 10 microns. Time is in seconds. Upper and lower insets represent one-plane zooms of white- and blue-boxed areas respectively, and contain a scale bar indicating 1 micron. White and blue arrowheads follow two ASPM polysomes. (B) A TrackMate overlay of the same cell in A showing polysomes tracks. Color code
represents displacement (dark blue lowest, red highest). Scale bar: 10 microns. The red arrow indicates a track showing directed movement, the green one, a track that does not localize with MTs, and the orange one a track that localizes with MTs. (C) Stacked histogram showing the percentage of ASPM tracks grouped in four categories depending on subcellular localization and behavior in interphase. (D) Graph showing the mean squared displacement (MSD) in micron$^2$ of polysomes in each of four categories as a function of time in seconds. (E) Snapshots of the SunTagx32-ASPM and scFv-sfGFP clone captured live during prophase with labeled MTs. The GFP/SunTag signal is shown in green and corresponds to ASPM polysomes and mature proteins; the Cy5 signal is shown in red and corresponds to a tubulin staining. Scale bar: 10 microns. Time is in seconds. Insets represent one-plane zooms of the white-boxed areas, and contains a scale bar indicating 1 micron. White and blue arrowheads follow the initial and current position of a polysome respectively. (F) Same legend as in E, but during prometaphase.
Figure 1
Figure 5

A

B

C

D

E

Safieddine et al.
Supplementary information

Systematic identification of human centrosomal mRNAs reveals a localization mechanism dependent on motors and nascent translation.
Supplementary figure legends:

**Figure S1: Expression and localization of ASPM, NUMA1, and HMMR proteins across a full cell cycle.**

(A) Images are snapshots of living HeLa cells containing an ASPM-GFP BAC imaged across a full cell cycle. Signal is in green and corresponds to the ASPM protein. Scale bar is 10 microns and time is in hours : minutes. Red and blue arrows follow two dividing cells. (B) Legend as in A, but for HeLa cells containing a NUMA1-GFP BAC. (C) Legend as in A, but for HeLa cells containing a HMMR-GFP BAC.

**Figure S2 (related to figure 2): mRNAs localize to distinct peri-centrosomal regions at precise times of the same mitotic phase during early cell division.**

(A) Images are micrographs of HeLa cells containing a NUMA1-GFP BAC and captured at different phases of early mitosis. Middle left and red: Cy3 fluorescent signals corresponding to NUMA1 mRNAs labeled by smFISH; middle right and green: Cy5 signals corresponding to ASPM mRNA labeled by smiFISH. DNA stained with DAPI. Scale bar: 10 microns. Insets represent zooms of the boxed areas and contain a scale bar of 1 micron. (B) Stacked histogram showing the percentage of cells having one, both, or neither transcript localized to centrosomes in each phase (n=10 cells). Data represent mean and standard deviation. (C and D) Legend as in A and B, but for HeLa cells containing an HMMR-GFP BAC with HMMR mRNA labeled in Cy3 shown in red, and ASPM mRNA labeled in Cy5 and shown in green. (E and F) Legend as in A and B, but for HeLa cells containing an HMMR-GFP BAC with HMMR mRNA labeled in Cy3 shown in red, and NUMA1 mRNA labeled in Cy5 and shown in green.
Figure S3 (related to figure 2): mRNAs differentially localize to centrosomes during late cell division.

(A) Images are micrographs of HeLa cells containing a NUMA1-GFP BAC and captured at different phases of late mitosis. Middle left and red: Cy3 fluorescent signals corresponding to NUMA1 mRNAs labeled by smFISH; middle right and green: Cy5 signals corresponding to ASPM mRNA labeled by smiFISH. DNA stained with DAPI. Scale bar: 10 microns. (B) Stacked histogram showing the percentage of cells having one, both, or neither transcript localized to centrosomes in each phase (n=10 cells). Data represent mean and standard deviation. (C and D) Legend as in A and B, but for HeLa cells containing an HMMR-GFP BAC with HMMR mRNA labeled in Cy3 shown in red, and ASPM mRNA labeled in Cy5 and shown in green. (E and F) Legend as in A and B, but for HeLa cells containing an HMMR-GFP BAC with HMMR mRNA labeled in Cy3 shown in red, and NUMA1 mRNA labeled in Cy5 and shown in green.

Figure S4 (related to figure 3): Translation initiation is required for the localization of ASPM mRNAs during all phases of mitosis.

(A) Images are micrographs of HeLa cells expressing an ASPM-GFP BAC and treated with puromycin captured during all phases of mitosis. Up and red: Cy3 fluorescent signals corresponding to ASPM mRNA labeled by smFISH; middle and green: GFP signals corresponding to the ASPM protein. Blue: DNA stained with DAPI. Scale bar: 10 microns. (B) Legend same as in A, but for cells treated with cycloheximide. (C) Histogram depicting the percentage of cells showing centrosomal localization of ASPM mRNA after both treatments (n=25). Data represent mean and standard deviation.
Figure S5 (related to figure 3): Translation initiation is required for the localization of NUMA1 mRNAs during all phases of mitosis.

(A) Images are micrographs of HeLa cells expressing a NUMA1-GFP BAC and treated with puromycin captured during all phases of mitosis. Up and red: Cy3 fluorescent signals corresponding to NUMA1 mRNA labeled by smFISH; middle and green: GFP signals corresponding to the NUMA1 protein. Blue: DNA stained with DAPI. Scale bar: 10 microns.

(B) Legend same as in A, but for cells treated with cycloheximide. (C) Histogram depicting the percentage of cells showing centrosomal localization of NUMA1 mRNA after both treatments (n=25). Data represent mean and standard deviation.

Figure S6 (related to figure 3): Translation initiation is required for the localization of HMMR mRNAs during all phases of mitosis.

(A) Images are micrographs of HeLa cells expressing a HMMR-GFP BAC and treated with puromycin captured during all phases of mitosis. Up and red: Cy3 fluorescent signals corresponding to HMMR mRNA labeled by smFISH; middle and green: GFP signals corresponding to the HMMR protein. Blue: DNA stained with DAPI. Scale bar: 10 microns.

(B) Legend same as in A, but for cells treated with cycloheximide. (C) Histogram depicting the percentage of cells showing centrosomal localization of HMMR mRNA after both treatments (n=25). Data represent mean and standard deviation.

Figure S7 (related to figure 5): Characterization of the ASPM-MS2x24 clone.

(A) Image is a scan of a gel loaded with the product of a PCR performed on genomic DNA extracted from various ASPM-MS2x24 clones. Wild-type and edited alleles are differentially amplified and give a product size of 1.6 and 2.6 kbs respectively. A ladder is placed on the left with the corresponding size markers. WT HeLa K.: PCR performed on
the parental unedited cell line. Water control: PCR performed without any DNA. Ta: annealing temperature. (B) Images are micrographs of a HeLa clone expressing endogenous ASPM-MS2x24 and MCP-GFP-NLS captured at interphase and mitosis. Left and blue: Cy5 fluorescent signals corresponding to endogenous ASPM mRNAs labeled by smiFISH; middle and red: Cy3 signals corresponding to MS2-tagged mRNAs labeled by smFISH; right and green: GFP signals corresponding to ASPM mRNAs labeled by the MCP-GFP-NLS. Blue: DNA stained with DAPI. Scale bar: 10 microns.

Figure S8 (related to figure 6): Characterization of the SunTagx32-ASPM clone

(A) Images are micrographs of a HeLa clone expressing endogenous SunTagx32-ASPM and scFv-sfGFP, treated with puromycin and captured in interphase and mitosis. Upper and red: Cy3 fluorescent signals corresponding to ASPM mRNAs labeled by smFISH; middle and green: GFP signals corresponding to the ASPM mature protein. Blue: DNA stained with DAPI. Scale bar: 10 microns. (B) Image is a scan of a gel loaded with the product of a PCR performed on genomic DNA extracted from various SunTagx32-ASPM clones. Wild-type and edited alleles are differentially amplified and give a product size of 1.2 and 0.7 kbs respectively. A ladder is placed on the left with the corresponding size markers. WT HeLa K.: PCR performed on the parental unedited cell line. Water control: PCR performed without any DNA. Ta: annealing temperature. (C) Snapshots of the SunTagx32-ASPM and scFv-sfGFP clone captured live during prometaphase. The GFP/SunTag signal is shown in black and corresponds to ASPM polysomes and mature proteins. Scale bar: 10 microns. Time is in seconds. Red arrowheads indicate the starting position of an mRNA molecule, while the blue one follows its current position. (D) A TrackMate overlay of the same cell in C showing polysomes tracks. Color code represents displacement (dark blue lowest, red highest). The outer yellow outline represents the cell
border, while the inner one represents centrosomes marked by mature ASPM protein. Scale bar: 10 microns.

**Figure S9 (related to figure 7): 1D displacements of ASPM polysome tracks grouped based on sub-cellular localization and dynamics.**

(A) Histogram showing 1D displacements measured between two consecutive time frames of ASPM polysome tracks localizing on MTs. D is the average diffusion coefficient. (B) Legend same as in A, but for ASPM polysome tracks not on MTs. (C) Legend same as in A, but for ASPM polysome tracks on the nuclear pore. (D) Legend same as in A, but for ASPM polysome tracks showing directed movements.

**Figure S10 (related to figure 7): The effects of nocodazole on ASPM polysome dynamics and ASPM mRNA localization.**

(A) Micrograph represents a temporal projection of the SunTagx32-ASPM clone expressing scFv-sfGFP imaged live during interphase with labeled MT. The GFP/SunTag signal is shown in green and corresponds to ASPM polysomes and mature proteins; the Cy5 signal is shown in red and corresponds to a tubulin staining. Scale bar: 10 microns. (C) Histogram showing 1D displacements measured between two consecutive time frames of ASPM polysome tracks not on the nuclear pore after a nocodazole treatment. D is the average diffusion coefficient. (D) Graph showing the mean squared displacement (MSD) in micron$^2$ of ASPM polysomes not localizing to the nuclear pore as a function of time in seconds after a nocodazole treatment. (E) Images are micrographs of the SunTagx32-ASPM clone expressing scFv-sfGFP captured at early mitosis. Left and red: Cy3 fluorescent signals corresponding to ASPM mRNAs labeled by smFISH; middle and green: GFP signals corresponding to ASPM polysomes and mature protein. Blue: DNA stained.
with DAPI. Scale bar: 10 microns. (F) Same legend as E, but with cells treated with nocodazole. (G) Histogram depicting the percentage of cells showing centrosomal localization of ASPM mRNA in early mitosis before and after a nocodazole treatment (n=25). Data represent mean and standard deviation.
Movie legends:

**Movie 1**: HeLa cells expressing the ASPM-GFP BAC imaged every 5 minutes for about 62 hours. GFP signal is shown in black and corresponds to ASPM mature protein.

**Movie 2**: HeLa cells expressing the NUMA1-GFP BAC imaged every 5 minutes for about 62 hours. GFP signal is shown in black and corresponds to NUMA1 mature protein.

**Movie 3**: HeLa cells expressing the HMMR-GFP BAC imaged every 5 minutes for about 62 hours. GFP signal is shown in black and corresponds to HMMR mature protein.

**Movie 4**: HeLa cells expressing ASPM-MS2x24 and MCP-GFP-NLS were imaged every 0.5 secs for 90 secs during prophase. GFP signal is shown in green and corresponds to ASPM mRNA; Cy5 signal was imaged in one frame, is shown in red, and corresponds to DNA.

**Movie 5**: HeLa cells expressing ASPM-MS2x24 and MCP-GFP-NLS were imaged every 0.26 secs for 45 secs during metaphase. GFP signal is shown in green and corresponds to ASPM mRNA; Cy5 signal was imaged in one frame, is shown in red, and corresponds to DNA.

**Movie 6**: HeLa cells expressing ASPM-MS2x24 and MCP-GFP-NLS were imaged every 0.625 secs for 112 secs during prometaphase. GFP signal is shown in green and corresponds to ASPM mRNA; Cy5 signal was imaged in one frame, is shown in red, and corresponds to DNA.
**Movie 7:** HeLa cells expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.9 secs for 180 secs during prophase. GFP signal is shown in black and corresponds to ASPM polysomes and mature protein.

**Movie 8:** HeLa cells expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.9 secs for 180 secs during prometaphase. GFP signal is shown in black and corresponds to ASPM polysomes and mature protein.

**Movie 9:** HeLa cells expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.66 secs for 130 secs during interphase with labeled MTs. GFP signal is shown in green and corresponds to ASPM polysomes and mature protein; Cy5 signal is shown in red and corresponds to MTs.

**Movie 10:** HeLa cells treated with nocodazole, expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.53 secs for 105 secs during interphase with labeled MTs. GFP signal is shown in green and corresponds to ASPM polysomes and mature protein; Cy5 signal is shown in red and corresponds to MTs and centrosomes.

**Movie 11:** HeLa cells expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.66 secs for 260 secs during prophase with labeled MTs. GFP signal is shown in green and corresponds to ASPM polysomes and mature protein; Cy5 signal is shown in red and corresponds to MTs.

**Movie 12:** HeLa cells expressing SunTagx32-ASPM and scFv-sfGFP were imaged every 0.66 secs for 130 secs during prometaphase with labeled MTs. GFP signal is shown in
green and corresponds to ASPM polysomes and mature protein; Cy5 signal is shown in red and corresponds to MTs.
Figure S6

A

Prophase    Prometaphase    Metaphase    Anaphase    Telophase

Puromycin

HMMR mRNA

HMMR protein

Merge

B

Prophase    Prometaphase    Metaphase    Anaphase    Telophase

Cycloheximide

HMMR mRNA

HMMR protein

Merge

C

% of HMMR DAC cells with the mRNA on centrosomes

Prophase    Prometaphase    Metaphase    Anaphase    Telophase

Puromycin

Cycloheximide

Safieddine et al.
ASPM polysomes on MTs

\[ D = 0.011 \mu m^2/s \]

ASPM polysomes not on MTs

\[ D = 0.041 \mu m^2/s \]

ASPM polysomes on nuclear pores

\[ D = 0.004 \mu m^2/s \]

ASPM polysomes with directed movements

\[ D = 0.019 \mu m^2/s \]
Discussion
1. mRNA localization screens: a technical perspective

One approach for studying subcellular mRNA localization is to carry out localization screens. Imaging based screens involve visualizing single molecules of mRNAs in single cells using fluorescent microscopy. A bottleneck for such studies is probe generation. Indeed, breaking the single molecule barrier in RNA imaging requires designing many labeled oligonucleotides that hybridize on the same transcript (generally speaking at least 24). The cost of synthesizing and labeling oligonucleotides quickly becomes a limiting step in high-throughput RNA localization studies.

To circumvent this, we and others developed indirect labeling schemes (Femino et al., 1998; Tsanov et al., 2016) in which the primary probes are unlabeled but carry a common extra sequence (the FLAP or the readout), which can be detected by hybridizing a secondary, fluorescently labelled oligonucleotide. High-throughput can then be achieved by either parallelization or multiplexing. In this latter case, the primary probes for many genes are hybridized all at once, but they carry different FLAP sequences that are detected separately by performing sequential hybridization with different secondary oligonucleotides. To increase multiplexing density, barcoding schemes that rely on combinatorial identification have been used, where each FLAP sequence is analogous to a bit (Chen et al., 2015; Lubeck et al., 2014). In seqFISH+ for instance (Eng et al., 2019), samples are subjected to 80 rounds of hybridizations and secondary probe removal. A series of transcripts light up in specific hybridization rounds, which produces a code that can then be used to identify each mRNA (see Figures 20 and 21 in the introduction). This allows detecting thousands of transcripts simultaneously in the same cell. The drawbacks however, include the need to embed cells in acrylamide, attach mRNAs to a matrix, and remove cellular components by extensive protease treatments. This reduces background, but also increases the risks of losing some mRNAs. A comparison performed in the lab suggests that a regular smFISH experiment detects 10 times more molecules than seqFISH+.

In my thesis, I contributed to two mRNA localization screens, each using a different smFISH approach:

- The first was a dual mRNA/protein screen carried out using a BAComics approach. This screen used HeLa cell lines, each stably expressing a BAC containing the gene of interest fused to a GFP tag, and flanked by all its regulatory genomic sequences. Since the BAC library
contains the same tag, the same set of labeled oligonucleotides that hybridize to the tag sequence was used to detect all mRNAs in the collection. The simultaneous detection of the encoded protein makes this approach particularly well suited for detecting cases of local translation. The major disadvantage of this approach is however that one mRNA corresponds to one cell line, which necessitates time for handling and maintaining cells in culture.

- The second screen involved generating large number of RNA probes using PCR amplification and in vitro transcription. Probes were designed and synthesized as a pool of oligonucleotides containing up to 92,000 of sequences. The probes contained two common FLAP sequences that were pre-hybridized with labeled oligos in an additional step similar to the smiFISH technique (Tsanov et al., 2016). This strategy drastically reduces probe generation costs and was successfully used in the lab to generate probes against thousands of genes.

Some studies use biochemical approaches to study subcellular mRNA localization. A recent study combined RNA proximity labeling by the peroxidase enzyme APEX2 with RNA-seq to describe spatial patterns of the human transcriptome in HEK cells (Fazal et al., 2019). More than 3200 RNAs were partitioned across four general categories using unbiased clustering: (i) nuclear, (ii) mitochondrial membrane and the ER, (iii) cytosolic, and (iv) ER lumen/mitochondrial matrix/nuclear pore.

Compared to standard imaging approaches, biochemical methods can provide a higher spatial resolution (nanometer-resolution in the case of APEX2-seq) and are more readily applied on a larger scale. However, such methods address mRNA localization in a population of cells and do not provide information on cell-to-cell heterogeneity. Moreover, despite the fact that some biochemical approaches (including APEX-seq) can be applied in living samples, they involve lysing cells in which dynamic spatial information in intact cells cannot be obtained. Finally, biochemical screen require expressing some form of recombinant tag. RNAs that are part of macromolecular complexes may not be properly tagged due to steric hindrance and would thus escape detection.

2. mRNA localization in cell lines: generalities

The two mRNA localization screens provided insights into general mRNA localization features in cell lines:
• First, even in the case of mRNAs that produce a strong pattern (DYNC1H1 mRNAs localizing in foci for example), a non-negligible fraction of mRNA molecules always seems to have random distribution. This is not the case in other systems such as developing embryos or oocytes in which most molecules of a localized transcript actually display the pattern. The reasons behind this difference are unclear; but (i) they could be a general aspect of mRNA localization in cell lines, in which they are much less stereotyped than embryos, especially cancer ones like HeLa; (ii) they could arise due to the relatively smaller size of HeLa cells (although similar observations are made in neurons which argues against this); (iii) localizing only a fraction of mRNA molecules is sufficient to carry out the intended biological function (discussed in detail below), in which there is no need to localize all transcripts for the cell to function properly.

• mRNA localization occurs in a wide range of subcellular locations in cell lines. These include cellular protrusions, the Golgi apparatus, endosomes, the nuclear envelope, cell edges, centrosomes, and even the cytosolic space itself (as in the case of mRNA foci and polarized distributions). This demonstrates the variety of mRNA trafficking and the diverse functions of mRNA localization.

• Most localized mRNAs display “simple” localization patterns in which they belong to one of the described localization classes (intranuclear, nuclear envelope, cell edge, cell extension, polarized, foci, centrosomes). However, some transcripts show “complex” patterns in which they localize to multiple compartments either simultaneously or at different times. An example of this is HMMR mRNA that localizes to both centrosomes and P-bodies during interphase. This difference is probably related to the function of each transcript but it implies that mRNA localization can be complex and can involve multiple sub-cellular compartments at once (see ASPM below for a detailed example).

• mRNA localization often exhibits cellular heterogeneity. Indeed, not all cells exhibit the pattern, and the strength of the pattern could vary between cells that display it. The is probably related to: (i) cells having different physiological states and thus different gene expression needs; (ii) cells being in different stages of the cell cycle; (iii) variations in mRNA
expression levels; (iv) variation in motors protein availability, RBP expression, cytoskeletal arrangements; or (v) a stochastic process yet unidentified.

3. A variety of local translation flavors

In many cases, mRNAs co-localized with the protein they encode suggesting local translation. This was observed in a variety of regions such as cell extensions, cell edges, centrosomes, endosomes, and the Golgi network. Local translation could function in:

- **Protein targeting:** ASPM mRNA for instance was the first case of local translation at the nuclear pore during interphase and this could facilitate nucleoplasmic targeting of the mature ASPM protein.

- **Preventing aberrant dispersal of proteins that could have harmful effects.** For example, NUMA1 mRNA and protein are only localized on centrosomes during mitosis. Targeting NUMA1 protein to centrosomes during interphase could have deleterious effects.

- **Enhancing translation efficiencies:** this applies to mRNAs that are localized and translated in cytoplasmic foci referred to as translation factories. These structures could host an environment more suitable for translation (a more favorable concentration of translation factors for instance). In fact, DYNC1H1 mRNAs are translated as both single molecules and mRNA foci. However, mRNAs in the foci are more often found engaged in translation than ones that exist as single molecules. Here, it is tempting to speculate that translation factories harbor specialized ribosomes or components of the translation machinery that are dedicated for localized cases of protein synthesis.

- **Nascent peptide metabolism:** synthesizing components of a protein complex locally can enhance the efficiency of complex assembly for example. Another possibility is co-translational assembly at certain organelles. For instance, translating mRNAs encoding centrosomal proteins such as PCNT, CEP350, NIN, and HMMR could help to incorporate the mature protein into centrosomes during the expansion of the pericentriolar material that occurs during the late G2 phase of the cell cycle. Finally, local translation could allow certain
proteins to acquire distinct post-translational modifications that are only catalyzed in a specific subcellular compartment.

It is also important to note that some localized mRNAs do not co-localize with their encoded proteins, whether nascent or mature, suggesting functions beyond local protein synthesis. A notable localization class in such a case is made of mRNA accumulating in P-bodies. These structures mediate several processes related to RNA metabolism such as translational repression, storage, and degradation. The exact functions carried out by P-bodies remain a subject of debate with recent evidence pointing towards a storage role (Hubstenberger et al., 2017).

4. Programmed translation on centrosomes

Examining centrosomal mRNA localization in detail revealed that mRNA localization and local translation were cell cycle dependent, and that each mRNA localized with its own spatio-temporal pattern:

- Some transcripts (ASPM and NUMA1) exclusively localized on **mitotic centrosomes**. This coincided with the localization of ASPM and NUMA1 proteins on centrosomes, which also occurs exclusively during mitosis, suggesting local translation there.

- Moreover, ASPM, NUMA1, and HMMR mRNAs localized to mitotic centrosomes at different phases of the cell division process, which further differed for each transcript. They also localized to **distinct peri-centrosomal regions**.

- Several others localized to centrosomes at **interphase** (HMMR, NIN, BICD2, CCDC88C, CEP350, and PCNT) and additionally, during **prophase**.

This localization class demonstrates well how precisely cells can target mRNAs and can add a temporal layer in regulating mRNA localization. Different combinations of transcripts are translated on centrosomes at different times. This is likely important for centrosome biogenesis, pericentriolar expansion, and spindle functioning.
5. A new method of protein targeting

For all localized mRNAs found in both screens, the dependency on active translation was examined. This was done by treating cells with either the translation inhibitor **puromycin**: a translation terminator that induces release of the ribosomal subunits and causes premature chain termination, or **cycloheximide**: an inhibitor that stops elongating ribosomes but stably tethers the ribosome and the nascent protein chains to the mRNAs. These experiments segregated localized mRNAs into two groups:

- mRNAs that still localized after inhibiting translation with puromycin. These included KIF1C, MYH3, and all 15 mRNAs localizing to P-bodies. In such cases, mRNA localization is an RNA-driven process that depends on Cis-acting elements often found in untranslated regions of the transcript. mRNA transport may occur while the mRNA is translationally repressed. This is in agreement with the current dogma (Besse and Ephrussi, 2008).

- mRNAs that de-localized after inhibiting translation with puromycin. This group included all other mRNAs, and notably all 8 mRNAs localizing to centrosomes. This was surprising since it indicated that mRNA localization is a protein-driven process that depends on the nascent peptide. Moreover, all transcripts in this group were cycloheximide insensitive, which indicated that the nascent protein chain was specifically required for mRNA localization. Imaging of ASPM polysomes directly proved that they are actively transported by motors to the centrosome (see below).

Collectively, both screens showed that co-translational targeting of mRNA is widespread and is in fact the only mechanism currently known to induce mRNA trafficking to centrosomes. This clearly contradicts the long-standing notion that mRNAs in transit are translationally silent and that transport is an RNA-driven process.

Co-translational targeting of mRNAs could resemble the signal recognition particle (SRP) targeting mechanism (Milstein et al., 1972; Walter et al., 1981). In eukaryotes, the SRP binds a signal sequence found in the nascent peptide as they emerge from the ribosome. This binding causes elongating ribosomes to arrest and translocates the whole mRNA-nascent peptide complex to the ER, where translation resumes. An "SRP-like" mechanism could exist and likely involves synthesizing a peptide-based signal that could be (i) the nascent chain itself, (ii) a binding partner
of the nascent chain, or (iii) ribosomal pausing when translating a section of the nascent chain. Regardless of the signal’s nature, it carries the information needed for the translating mRNA to reach its target site. The localization method would probably be motor-dependent (see below).

Ribosomal pausing in particular provides an attractive mechanism especially when considering co-translational protein folding. Co-translational protein folding has been documented in both bacteria and eukaryotes (Buskirk and Green, 2017; Kudva et al., 2018). Indeed, many nascent chains can begin to fold and assemble their quaternary structure during biosynthesis on the ribosome. This folding is encoded in the amino acid sequence, but also influenced by the ribosome exit tunnel shape and chaperons (Waudby et al., 2019). In certain cases, ribosomes slow down or even stop to provide time for the nascent chain to properly fold (Kim et al., 2015).

This fine-tuning of translation rates could occur during co-translational targeting of polysomes towards the centrosome. In this case, some part of the nascent protein remains unfolded and this drives ribosomal arrest. Proper folding only occurs with help of a nascent chain partner. If this partner is centrosomal, translation will resume on centrosomes. This mechanism would agree well with the concept of co-translational assembly at centrosomes.

6. Visualizing endogenous polysomes: a technical perspective

In 2016, a series of papers described a technique to visualize the translation of single mRNAs (Morisaki et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016; Pichon et al., 2016). Although various constructs were used in these studies, the principle was the same: multimerization of peptide epitopes (12 to 56 repeats depending on design and target) to the N-terminal of the protein of interest. One of these epitopes (which was used in my thesis) is derived from the yeast GCN4 peptide and is called the SunTag (Tanenbaum et al., 2014). The SunTag epitopes are recognized by single chain variable fragments of an antibody fused to GFP (scFv-sfGFP). As soon as SunTag epitopes are translated, they are become bound by scFv-sfGFP, thereby labeling the nascent peptide when it emerges from ribosomes. The signal is further amplified since multiple ribosomes are translating the tagged mRNA, leading to the appearance of bright foci visible by fluorescence microscopy, which correspond to polysomes.

Throughout my PhD, I tagged multiple genes with the SunTag. Some technical insights gleaned from such experiments include:
**Epitope number:** In simple terms, the more epitopes used, the brighter polysomes will appear. This, however, is affected by several parameters: steric hindrance on binding sites, unavailability of scFv-sfGFP molecules, variable number of translating ribosomes, and length of the ORF, among other things. I tried both 32 (for ASPM, NUMA1, and HMMR) and 56 (for POLR2A, and DYNC1H1) SunTag repeats and both arrays produced polysomes bright enough for both fixed and live imaging. Yet, the larger 56 repeat array might be better suited for tagging small proteins. However, the bigger the tag, the higher risk of altering the function, localization, and/or expression of the gene. Note that adding 32 SunTag repeats for instance, did not prevent ASPM and NUMA1 mature proteins from localizing to the nucleoplasm during interphase and centrosomes during mitosis. On the other hand, 56 SunTag repeats de-localized POLR2A from the nucleoplasm, but did not abolish motorized movements and function of DYNC1H1 proteins. Thus, (i) the tagging effect seem to be gene dependent, (ii) it is better to minimize repeat number when possible (with long ORFs for example), but this depends on the details of the experiments and in particular on how long one wants to image polysomes.

**Expression levels and localization of scFv-sfGFP:** having too little scFv-sfGFP can either reduce polysome brightness or leave a population of polysomes unlabeled. Inversely, expressing too much scFv-sfGFP causes background due to unbound scFv-sfGFP and this reduces the signal to noise (SNR) ratio. Having optimal and stable amounts of scFv-GFP is thus crucial. One possibility is to sequester unbound scFv-sfGFP (or scFv-sfGFP bound to mature proteins) to a different subcellular compartment when possible. This can be done by fusing a localization signal (such as a nuclear localization signal, NLS) to the scFv-sfGFP. However in application, expressing scFv-sfGFP or scFv-sfGFP-NLS in a SunTagx32-POLR2A clone did not result in an increase in SNR. Localization tags could also affect the spatial distributions and dynamics of polysomes themselves and their application will thus depend on the biological question.

**Differentiating polysomes from the mature protein product:** in general, polysomes can be easily identified due to (i) their sensitivity to puromycin, and (ii) co-localization with mRNAs revealed by smFISH or MS2 labeling. However, in the cases where mRNAs are locally translated and accumulate at the same site where their mature protein product resides,
identifying individual polysomes becomes problematic. This is because the SunTagged mature protein will mask the signal from polysomes. This issue was encountered when attempting to image ASPM and NUMA1 translation on centrosomes in mid to late mitosis. Indeed, both mature and nascent ASPM and NUMA1 proteins co-localize and accumulate on centrosomes with makes assessing translation difficult. To circumvent this, I attempted to separate the SunTag array from the mature protein chain using a variety of genetic tools. These included self-cleavage signals (P2A, T2A, E2A, F2A, and combinations of them in tandem) and internal ribosome entry sites (IRES) to produce a polycistronic mRNA. However, none of the approaches worked in preventing mature ASPM and NUMA1 protein from accumulating on centrosomes. This is likely due to the fact that these signals are not 100% efficient in which and small leaks quickly cause a build-up of mature proteins on centrosomes due to their high expression during mitosis. Another possibility is that the ORF resulting from excluding the main protein sequence is too short for sufficient signal amplification. The resulting polysomes are thus too dim to be detected above background levels.

In the future, it is likely that “background-free” imaging systems will be developed. A good example is a system that uses split GFP (Kaddoum et al., 2010). In this case, scFv (or other nanobodies) is fused to a GFP fragment that only fluoresces after binding the target epitope that carries the complimentary GFP portion. This strategy is of particular interest in the field of live RNA imaging using the MCP-MS2 system in which SNRs in the cytoplasm are often low.

7. Co-translational targeting captured live

The technology for live imaging of endogenous polysomes and mRNAs unlocked the ability to directly inspect the translational state of transcripts in transit. To this end, I tagged the endogenous ASPM gene with either 32xSunTag repeats or 24xMS2 stem-loops to visualize single ASPM polysomes and mRNAs respectively. Remarkably, a large portion of ASPM polysomes could be seen undergoing directed movements in prophase towards the centrosome after nuclear envelope breakdown. This was in contrast to interphase in which ASPM polysomes showed mainly diffusive movements. This provided definitive evidence that (i) protein-dependent mRNA localization occurs co-translationally, (ii) and that the localization mechanism involves motor driven transport.
along the cytoskeleton. Whether or not this applies to other localized mRNAs (particularly at the centrosome) and to what extent remains to be discovered.

One interesting aspect of directed co-translational targeting of ASPM mRNAs is that it is cell cycle regulated. Indeed, this was exclusively seen during prophase (and slightly during prometaphase), which is about 5-10 minutes of a 24-hour cell cycle. How cells orchestrate this remains a mystery, but one can speculate that it is related to (i) cell cycle or mitotic entry specific phosphorylation events related to the peptide signal discussed in section 5, (ii) expression or modification of certain binding partners of the peptide signal at specific times, (iii) cytoskeletal rearrangements, or (iv) nuclear envelope breakdown.

8. ASPM: a transcript with many patterns

In both mRNA localization screens, ASPM displayed the most complex cell-cycle dependent localization pattern. Combining smFISH with live polysome and mRNA imaging revealed remarkable diversity in ASPM mRNA localization, dynamics, and translation, which involved multiple sub-cellular compartments:

- During interphase, ASPM mRNAs are translated on (i) nuclear pores, (ii) in translation factories, and (iii) on microtubules. The exact function of anchoring ASPM polysomes on microtubules is unknown especially since the mature protein does not localize there during interphase. However, it could serve as preparation for the onset of prophase in which translating mRNAs are preemptively bound to microtubules.

- During early mitosis ASPM polysomes undergo directed movements through (i) sliding on microtubules, and (ii) being pulled by microtubules towards the centrosome. Sliding involves molecular motors, while pulling is probably mediated by two calponin homology domains (CH1 and CH2) found towards the N-terminal of the ASPM proteins that can directly bind microtubules (Jiang et al., 2017). Whether these two methods are complementary or redundant remains to be seen.

- Once they reach their target site, ASPM mRNAs remain stably anchored on both centrosomes and spindle fibers themselves. This suggests that mRNAs there always have
a nascent peptide chain associated to them since they do not diffuse away. This likely means that they are constantly initiating translation on centrosomes.

Collectively, ASPM translation appears to be extremely fine-tuned in space and time. This demonstrates the extent to which mRNA localization and local translation can be elaborate.

9. mRNA localization: the past, present, and future

In the past decade, the view of mRNA localization has shifted from a rare occurrence that concerns a select set of transcripts, into a bona fide aspect of gene expression regulation. Messenger localization was first discovered in *Xenopus* oocytes (Weeks et al., 1985). Since then, it has been characterized with great precision in various embryos, cell lines, tissues, and even unicellular organisms such as yeast and bacteria. Such feats were possible thanks to advances in biochemical and imaging techniques. Recent technological leaps can be illustrated by two examples:

- The development of smFISH variants tailor-made for targeting thousands of transcripts: these include multiplexing (MERFISH) and sequential hybridizations (seqFISH+) that can detect up to tens of thousands of RNAs at the single molecule level. Almost 20 years after the invention of single molecule FISH (Femino et al., 1998), these approaches (and others) are paving the way for imaging-based in situ transcriptomics. These techniques can now compete with single cell sequencing and additionally provide spatial information.

- The advent of nascent peptide imaging: the ability to inspect the translational status of mRNAs in real time has proved to be invaluable in understanding mechanisms of mRNA trafficking and translational regulation (a main objectives of this thesis). Although polysome imaging is relatively recent, newer tags have already been developed (Boersma et al., 2019; Zhao et al., 2019). These allow the inspection of many aspects of translational regulation and even multiplexing at the level of polysomes.

In the future, it is likely that better microscopes, brighter fluorophores, and robust tags will continue to reshape the field of single molecule imaging. One can already envision systems with background free imaging (ex: split GFP), tag free imaging (ex: with Cas13 or RNA aptamers), and also efficient smFISH variants that could cover entire transcriptomes. One important challenge will be to develop high-throughout and multiplexed methods for live cell imaging. Such powerful tools
will address demanding questions in the field of gene expression in general, and various spatiotemporal aspects of RNA metabolism in particular.
Annexes
A computational framework to study sub-cellular RNA localization

Aubin Samacoïts1,2, Racha Chouaïb3,4, Adham Safieddine3,4, Abdel-Meneem Traboulsi3,4, Wei Ouyang1,2, Christophe Zimmer1,2, Marion Peter3,4, Edouard Bertrand3,4, Thomas Walter5,6,7 & Florian Mueller1,2

RNA localization is a crucial process for cellular function and can be quantitatively studied by single molecule FISH (smFISH). Here, we present an integrated analysis framework to analyze sub-cellular RNA localization. Using simulated images, we design and validate a set of features describing different RNA localization patterns including polarized distribution, accumulation in cell extensions or foci, at the cell membrane or nuclear envelope. These features are largely invariant to RNA levels, work in multiple cell lines, and can measure localization strength in perturbation experiments. Most importantly, they allow classification by supervised and unsupervised learning at unprecedented accuracy. We successfully validate our approach on representative experimental data. This analysis reveals a surprisingly high degree of localization heterogeneity at the single cell level, indicating a dynamic and plastic nature of RNA localization.
on-random sub-cellular RNA localization is important for cellular function and its misregulation is linked to a number of diseases\textsuperscript{12,13}. Initially observed in highly polarized cells such as oocytes or embryonic fibroblasts, more recent studies revealed diverse and wide-spread RNA localization in other systems\textsuperscript{4}, including bacteria\textsuperscript{5}, yeast\textsuperscript{6}, and developing embryos of fruitfly, ascidians and zebrafish\textsuperscript{7,14}. RNA localization also occurs in cultured mammalian cell\textsuperscript{7}. Besides the particular case of neurons where a large number of mRNAs localize in cellular processes, mRNA localization also occurs in regular cell lines to regulate gene expression at the spatial level. Secreted and mitochondrial proteins are often translated at the endoplasmic reticulum and mitochondria, respectively, while mRNA repressed for translation can accumulate in P-bodies or stress granules. More specific examples of localization include mRNAs that accumulate at the tip of cellular extensions\textsuperscript{8}, localize at the cell periphery\textsuperscript{9}, or DYNCHI mRNAs that accumulates in foci representing dedicated translation factories\textsuperscript{10}. With the rapid development of high-throughput techniques, it is likely that many more localized RNAs will be discovered. However, validated analysis tools to identify and classify such RNA localization patterns are currently lacking.

Imaging technologies, especially single molecule FISH\textsuperscript{7-12,13} (smFISH), allow to observe single RNA molecules in their native cellular environment. This technique is now easy to implement and can be performed at large scale in cell lines and embryos\textsuperscript{7,10,12,14,15}. Image analysis then allows to discover genes displaying non-random localization patterns. While many localization patterns are distinguishable by visual inspection\textsuperscript{4,8}, manual annotation can be biased, is often not quantitative and influenced by confounding factors such as RNA expression level. In addition, comprehensive manual annotation at the single cell level hardly seems an option for larger scale studies where thousands of cells are imaged in a single experiment. Indeed, the benefits of automatic analysis of smFISH data\textsuperscript{7,16} include scalability and reproducibility, allowing for an accurate and quantitative description of the spatial aspects of gene expression.

In smFISH images, individual RNA molecules appear as bright diffraction-limited spots, which can be localized in 3D with published image analysis tools\textsuperscript{12,14}. In contrast to the analysis of cellular phenotypes\textsuperscript{17} and protein localization\textsuperscript{18}, smFISH data can be treated as point clouds. The smFISH signal inside a cell can thus be represented by features describing this spatial distribution of points, such as the mean nearest neighbor distance between spots or their average distance to the nuclear envelope. These features can then be used to group cells based on similarity in their RNA localization patterns, using supervised or unsupervised machine learning methods\textsuperscript{19}. However, one of the main difficulties in this approach is the absence of a ground truth for RNA localization in smFISH data, making it impossible to assess usefulness of features and performance of the classificationWorkflow. Hence, as of today, there is no rigorously validated method to analyze smFISH data at the cellular level.

Here, we present a simulation framework to create a synthetic ground-truth data set to perform this validation. Such simulated ground-truth data provide a number of key advantages to the traditional strategy relying exclusively on manual annotation\textsuperscript{17-23}. Manual annotation of 3D point clouds irrespective of their number and reference volume is time consuming, difficult, error prone and tends to be subjective, in particular for subtle differences. In addition, we can only annotate already observed patterns from already identified example genes. This encouraged us to build a simulation framework in order to complement or replace manual annotation. We generated point patterns from known localization rules to create large amounts of ground-truth data. This allowed us to also control the parameters of the generative model in order to study robustness and limitations of the automatic algorithms. We show that the simulation of a large set of images enables designing and validating workflows for unsupervised and supervised analysis of smFISH data, which are capable of detecting a large variety of localization classes. We applied this approach to experimental data and successfully detected the different manually annotated localization classes. We also implemented a metric to quantitatively analyze heterogeneity of localization patterns. Application of this metric to our experimental data set revealed a surprisingly high degree of localization heterogeneity.

Results

Simulating realistic ground-truth data. We designed a simulation framework capable of generating ground-truth data closely mimicking experimental smFISH images. First, we acquired fluorescence microscopy data to determine the 3D volume of cells and nuclei, as well as the typical smFISH background signal. We then placed individual RNAs with realistic signals according to pre-defined localization rules inside these cellular volumes. This set of simulated images allowed us to validate the entire analysis workflow from RNA detection to the identification of localization classes. Our approach to generate such synthetic smFISH images consists of four main steps (Supplementary Note 1):

First, we inferred accurate 3D cellular shapes from experimental data\textsuperscript{22}. To do so, we performed an smFISH experiment in HEla cells using Cy5-labeled probes against the highly expressed GAPDH mRNA. We defined the 3D cellular outline as the conforming boundary containing all detected mRNAs (Fig. 1a, b). We further used our observation that GAPDH mRNA was largely absent from the nucleus (Fig. 1a) to determine the average height and position of nuclei in the cells (Fig. 1b). This provided us with a collection of cellular and nuclear volumes, where RNAs can be placed. For each cell, we also acquired another channel with realistic background from a mock smFISH experiment, using Cy3-labeled probes against a not-expressed reporter gene. In addition, we acquired images for 2D segmentation of cells and nuclei (CellMaskTM and DAPI, respectively), as they are used in standard screening applications. The open design of our workflow makes it possible to add additional marker channels, e.g., P-bodies, for more specific screening applications.

Second, we defined different RNA levels. It is well known that expression levels can vary greatly between genes and even between clonal cells for a single gene. Ongoing efforts identify stochastic noise in transcription and extrinsic factors such as cellular microenvironment or cell volume as the molecular and environmental origins of these cell-to-cell variations\textsuperscript{22,23}. In contrast, an analysis workflow of RNA localization has to be independent of RNA levels, because different cells should be grouped together based on similarity in their RNA localization patterns, and not in their expression levels. We defined RNA density as a free parameter, and the absolute number of RNAs was assumed to be proportional to the cell volume, as shown in recent studies\textsuperscript{22,23} (Supplementary Note 1). We simulated four regimes of expression, each regime with constant RNA density modulated with an additional Poisson noise term, which when pooled cover a large range of expressions levels in agreement with a recently observed large-scale screen (Fig. 1c)\textsuperscript{2}. Lastly, we also considered a scenario with very high expression levels to test the potential limitations of the classification approaches.

Third, we simulated realistic images of individual RNA molecules. Each RNA was simulated as a point-spread-function (PSF) with sub-pixel localization and intensities from an
experimentally measured distribution (Fig. 1d, Supplementary Note 1). Together with the realistic smFISH background, this simulation strategy also allows testing the performance of RNA detection algorithms 12,14 (Supplementary Note 2).

Fourth, we placed RNAs within the 3D cellular volume either randomly (Fig. 1e) or according to one of seven specific localization patterns (Fig. 1f–g). We designed these patterns to either mimic experimental data or to generate so far unobserved, yet plausible localizations. The patterns correspond to RNAs localizing to the tip of cellular extensions, around the cell membrane in 2D or 3D, with a polarized pattern, in foci, and around the nuclear envelope in 2D or 3D. Each pattern corresponds to a set of rules defining the spatial distribution of RNA molecules. For each pattern, there is a set of fixed parameters that define the nature of the localization pattern. In addition, we define one parameter (pattern strength) to control how extreme a pattern is (e.g., for the nuclear envelope localization, the pattern strength corresponds to the fraction of RNA in proximity of the envelope). For low pattern strength, the distribution approaches spatial randomness. For more details we refer to Supplementary Note 1. For this study, we defined three strengths, with the intermediate corresponding to a typically observed experimental pattern as assessed by a human observer (for an example see Fig. 1g, where the localization of RAB13 mRNA) (left), simulated data with low mRNA density and moderate pattern strength (right).

Analysis of simulated data with existing approaches. We first generated data with little heterogeneity for both the pattern strength and expression level. We used simulated images with moderate pattern strength and an average of 200 RNAs per cell (100 cells per condition), to assess the performance of previously published analysis methods aiming at identifying RNA localization patterns. Our first analysis was based on a recent publication.
presenting a very comprehensive analysis workflow\(^7\,^{24}\). Here, 32 localization features are calculated from a mRNA detection performed on a maximum-intensity projection of the smFISH images (Supplementary Note 4).

In order to visualize the feature distributions for the simulated patterns, we projected these features onto two dimensions by t-Distributed Stochastic Neighbor Embedding (t-SNE)\(^2\). While half of the patterns were reasonably well separated, the other half was inseparable (Fig. 2a). The inseparable patterns were RNA localizing in cellular extensions, in foci, at the cell membrane in 3D or randomly. k-means clustering yielded the same set of indistinguishable patterns (50% accuracy, Fig. 2b). This encouraged us to improve RNA detection in 3D and to design new features describing their spatial distribution.

**RNA detection and localization features.** Large-scale studies have shown that accumulation of RNA in cytoplasmic foci is a predominant localization pattern\(^3\). Despite their biological relevance, previous analysis approaches were not designed to correctly identify these foci. We hypothesized that this might be due to problems during RNA detection in existing methods, where an accumulation of very close RNAs is detected as a single molecule. We hence designed an analysis method based on Gaussian Mixture Models (GMM), where these foci are decomposed into individual RNA molecules. We detail the validation of this approach on simulations and experimental data in Supplementary Note 2.

We then designed new feature families based on established concepts from spatial statistics and image analysis. First, we used Ripley’s L-function, which provides information about homogeneity of spatial density. We assumed that this would be particularly useful for detection of patterns like foci and polarized RNAs. Second, we used morphological operators to extract cellular extensions from the 2D mask of the cells and used the enrichment ratio of RNA counts in these extensions as a feature. Third, we developed features capturing RNA localization with respect to the cell membrane. This is challenging, since usually no information about the 3D cellular shape is available in standard smFISH experiments. By using our mock smFISH experiment, we could show that the estimated background of wide-field smFISH images was correlated with the cell height (Supplementary Note 3). We thus defined as a feature the correlation of the measured z-positions of RNAs and the background intensity (approximation for cell height). Lastly, we normalized the features describing distances between RNAs and cellular structures by the expectation of those distances under complete spatial randomness. Using our simulated data, we found that this normalization reduced the impact of cellular shape on the feature distributions, as compared to previously proposed normalization schemes\(^1\) (Supplementary Note 3). We also added two previously published features for the description of polarized mRNA localization (Polarization and dispersion index)\(^1\). Altogether, we defined a new set of features (14 new and 9 previously published, Table 1).

We then analyzed the simulated data with this new feature set and the new RNA detection scheme in 3D including the GMM. This greatly improved the identification of localization patterns: all localization patterns were distinguished in an unsupervised setting with an overall accuracy of 90% (Fig. 2c, d, Supplementary Note 4).

**Robustness of classification toward heterogeneity.** We next investigated the robustness of this classification toward increased heterogeneity. First, we pooled the three pattern strengths defined above and found that the classification results remain good (accuracy 79%; Supplement Note 4). We then pooled data simulated with RNA densities from an average of 50–400 RNAs per cell (Supplement Note 4), corresponding to the range recently measured in a large-scale smFISH study\(^7\). This test is crucial, since the analysis should identify cells with similar RNA localization patterns independently of variations in their expression level. We found that in a t-SNE plot, cells remained separated based on their localization pattern (Fig. 2e), showing that the features are stable with respect to expression heterogeneity. We also analyzed simulated cells with a very high expression level (average of 800 RNAs per cell). Here, we found that out of 8 patterns could still be correctly identified (Supplementary Note 4). Only RNA foci were confounded with random, which can be explained by the increased local density of genes expressed at very high levels. However, we expect such extreme levels to be rare.

We then compared various clustering strategies (Supplementary Note 4) and found that k-means applied to a six-dimensional t-SNE analysis gave the best results (88% accuracy, Fig. 2f). If the number of clusters is not known, it can be inferred from the data with traditional methods such as the silhouette score or an analysis of the within-class variability (Supplementary Note 4).

Lastly, we tested whether our simulations could also be used to analyze the extent by which two spatial distributions must differ to be still detectable as being different. This is particularly interesting in the context of drug perturbation experiments. Specifically, we can use our simulation framework to perform a sensitivity analysis. We simulated data with a large range of pattern strength and investigate the impact of the pattern strength on the accuracy of the analysis (Supplementary Note 4).

Overall, we designed a workflow for the unsupervised analysis of localization patterns, that we validated on simulated data with varying degrees of both pattern strength and expression heterogeneity. Our results indicate that, while some localization patterns are easier to detect than others, all these localization patterns could be detected with good accuracy.

**Detection of localized mRNAs in experimental data.** Motivated by these results, we analyzed experimental data from 10 genes in HeLa cells (150–400 cells per gene, 2600 cells in total; Supplementary Note 5). Three mRNAs were manually annotated as random (KIF20B, MYO18A, PAK2), and seven were chosen because of their non-random localization. RAB13\(^9\) and KIF1C mRNA accumulate in cell extensions. DYNC1H1\(^11\) and BUB1 mRNA concentrate in cytoplasmic foci. The last three mRNAs display a localization associated with the nucleus: the ATP6A2 protein is synthesized on the endoplasmic reticulum and its mRNA concentrate in the perinuclear region; SPEN mRNA form a rim that decorates the nuclear edge; while CEP192 mRNA concentrates inside the nucleus. Some of these patterns were not explicitly included in our simulated classes (e.g., perinuclear and intranuclear), but they were nonetheless included to test whether our feature set would be general enough to enable classification of novel patterns. It is also important to note that, for any of these genes, we typically observe different cellular sub-populations with different spatial distributions of the encoded transcripts. Gene-level annotation therefore describes only a tendency and cannot be considered as a single-cell annotation.

We applied the workflow that we had benchmarked on the simulated data to the experimental data and extracted localization features for each cell. We represented the extracted localization features as a t-SNE plot (Fig. 3a, Supplementary Note 5). We also created a more detailed version of this plot where each point was replaced by a thumbnail representation of the cell, showing outlines of cells and nuclei and the detected RNAs. The plot can...
We found that cells from different genes with the same or similar manually annotated localization pattern showed an important overlap in the t-SNE plot. Conversely, we found that cells from genes with different manual annotations tended to populate distinct regions (Fig. 3a). This shows that the localization features are well suited to describe mRNA localization patterns in different regions of the feature space look like.

be viewed at https://muellerflorian.github.io/locFISH_deepzoom/results/tsne_exp (for examples of the thumbnails see Fig. 3a). The tools to generate such plots are part of the provided Matlab source code. This zoomable t-SNE plot gives a clear picture of what the localization patterns in different regions of the feature space look like.
localization from experimental data. This result was confirmed by the analysis of a different cell line (C2C12) (Supplementary Note 5). These cells are larger than HeLa cells (600 vs. 200 μm²) and also more elongated. We found that C2C12 cells with three different manually annotated localization patterns (DYNC1H1 - cellular extension, KIF1C - cellular extension, ACTN1 - random and polarized) were well separated.

Interestingly, the t-SNE plot also captures the heterogeneity we had observed during visual inspection of the data. First, for genes with non-random localization, not all cells have a distinct mRNA localization. For instance, most CEP192 cells form a clearly separated cluster with strong intra-nuclear localization (Fig. 3a, region 2), while other CEP192 cells are random and grouped together with genes also showing a random localization pattern (Fig. 3a, region 1). Similar observations can be made for all other genes annotated as having non-random RNA localization. Interestingly, the intra-nuclear localization pattern of CEP192 transcripts was not among the patterns used for simulation. This shows that the features we have designed are informative beyond the simulated patterns and in principle capable of adequately describing other patterns as well. Second, some spatial RNA distributions appear to be pattern mixtures. For instance, DYNC1H1 mRNAs form foci1, which are displayed in the upper part of the t-SNE plot (Fig. 3a, region 3). However, in some cells these foci are located toward the nuclear envelope, and these cells are positioned in close proximity with other cells displaying nuclear localization in the t-SNE plot (Fig. 3a, region 4). We conclude that our features are capable of capturing the complex structure in the data, and that t-SNE visualization allows to explore the heterogeneous localization patterns displayed by different cell populations.

Supervised classification of RNA localization. Next, we investigated whether we can infer different clusters of subcellular localization at the single-cell level. k-means with 4 classes correctly separates the manually annotated localization classes (random, cellular extension, foci, and nuclear-associated) but failed to isolate a small cluster of cells with strong intra-nuclear localization (Fig. 3b). In contrast, spectral clustering was able to find this cluster but fused larger localization patterns (Fig. 3b, Supplementary Note 5). Thus, unsupervised learning with fixed number of clusters can be used in order to identify localization patterns. These results further suggest that it is important to explore the data structure in detail and in particular the large clusters. This then allows to tailor the level of detail in the analysis to the biological question.

Next, we applied hierarchical clustering (Supplementary Note 5). This approach arranges the data based on similarity and does not require a pre-definition or inference of the number of classes. Importantly, this allows to visually explore the substructure of large clusters. It also permits to inspect which features are similar among sub-groups, providing a basis for a more informed bio-physical interpretation of clusters and sub-clusters. Hierarchical clustering revealed diverse and subtle localization patterns. For instance, we observed highly specific sub-clusters of nuclear-associated patterns corresponding to intra-nuclear localization (Fig. 3c). We also observed distinct pattern mixtures for RNA foci, with localization either towards the cell membrane or the nuclear envelope (Fig. 3d).

In summary, we show that the developed workflow is capable of describing the different manually annotated localization patterns in real experiments. We also illustrate how different non-supervised clustering methods can be used to explore the data.

Supervised classification of experimental data. We next addressed the question of whether the simulation framework can be used in a supervised setting. Unlike unsupervised methods, supervised learning allows to impose prior knowledge in the form of a training set, i.e., to give the algorithm the opportunity to learn which feature combinations are relevant for a biologically meaningful distinction between patterns. We found that due to the realistic nature of our simulations (Supplementary Note 5), we could train a classifier (here: Random Forests26) on simulated data, and successfully detected the correct localization patterns in experimental data (Fig. 4a, Supplementary Note 5). As a further validation, we compared the performance of this classifier and a classifier trained on manually annotated data and found nearly identical performance (Supplementary Note 5). These results show that the feature distributions of simulated and real data are close enough to allow to train a classifier with simulated data without a notable loss in performance.

Heterogeneity in mRNA localization. Interestingly, with this analysis we can study heterogeneity in RNA localization in detail. As described before, such heterogeneity can correspond to the co-existence of several cellular subpopulations with pure patterns, or a mixture of patterns within individual cells. To distinguish between these two scenarios, we turned to the posterior probabilities for single cells, i.e., the probabilities of a cell to belong to each of the patterns (Fig. 4b, c, Supplementary Note 5). From these posterior probabilities we can calculate the Gini impurity at the single-cell level, which gives an indication of the purity of the pattern for that particular cell (low values indicating a very pure pattern). The Gini impurity can also be calculated on the population level, i.e., on the average posterior probabilities, indicating the heterogeneity at the population level. By plotting the Gini impurity for the population against the intracellular Gini impurity (Fig. 4d), we can further investigate the nature of the heterogeneity. This reveals that some genes are characterized by rather low heterogeneity both at the cellular and the population level, such as RAB13 and KIF1C, the two mRNAs accumulating in cell extensions. For other genes, we observed low heterogeneity at the cellular level and high heterogeneity at the population level. This is the case for CEP192, and it indicates the existence of pure subpopulations and suggests the co-existence of different localization states (Fig. 4c). Other genes are characterized by a high intra-cellular heterogeneity, i.e., a mixture of patterns, such as DYNC1H1 (Fig. 4c). Both population and intra-cellular

### Table 1 List of localization features used in Figs. 3 and 4

<table>
<thead>
<tr>
<th>Feature ID</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ripley: maximum</td>
</tr>
<tr>
<td>2</td>
<td>Ripley: max gradient [0, max]</td>
</tr>
<tr>
<td>3</td>
<td>Ripley: min gradient [max, end]</td>
</tr>
<tr>
<td>4</td>
<td>Ripley: value at mid-point between center and boundary</td>
</tr>
<tr>
<td>5</td>
<td>Ripley: Spearman correlation between Ripley and radius</td>
</tr>
<tr>
<td>6</td>
<td>Ripley: radius of max value</td>
</tr>
<tr>
<td>7</td>
<td>Polarization index</td>
</tr>
<tr>
<td>8</td>
<td>Dispersion index</td>
</tr>
<tr>
<td>9-12</td>
<td>Morph opening-enrichment ratio: 15, 30, 45, 60 pixels</td>
</tr>
<tr>
<td>13</td>
<td>Cell height: Spearman correlation with Z_mRNA</td>
</tr>
<tr>
<td>14</td>
<td>Cell height: R² with Z_mRNA</td>
</tr>
<tr>
<td>15</td>
<td>Cell membrane: distance - mean</td>
</tr>
<tr>
<td>16-19</td>
<td>Distance membrane: quantile 5%, 10%, 20%, 50%</td>
</tr>
<tr>
<td>20</td>
<td>Nucleus: distance - mean</td>
</tr>
<tr>
<td>21</td>
<td>Cell centroid: distance - mean</td>
</tr>
<tr>
<td>22</td>
<td>Nucleus centroid: distance - mean</td>
</tr>
<tr>
<td>23</td>
<td>Ratio: mRNAs inside nucleus/Outside nucleus</td>
</tr>
</tbody>
</table>
heterogeneity could imply a high degree of plasticity in mRNA localization mechanisms, possibly because of the need to rapidly adapt to changes in cell shape and cellular micro-environment. This also shows that cell-to-cell variations in gene expression occur not only on the expression level as often noted \cite{44}, but also extends to the spatial dimension.

**mRNA localization in different experimental conditions.** We finally tested how our workflow could be used to investigate changes in RNA localization. Such an analysis can allow to infer the biological relevance of an observed pattern, for instance by performing perturbation experiments. In a recent study, we reported that DYNC1H1 displays RNA foci \cite{45}. We further showed that these foci do not overlap with P-bodies or stress granules, but are translation factories. To test the role of translation in the formation of foci, we treated cells with Puromycin, a terminator of translation elongation. Visual examination indicates that the foci disappeared after a brief puromycin treatment (30 min). When analyzing these data with our GMM RNA detection approach, we could successfully quantify this change in localization pattern, and also show a change in the localization classes with both unsupervised and supervised approaches (Supplementary Note 5).

**Discussion**

In this study, we present a framework for the simulation of smFISH images with non-random mRNA localization patterns and methods for their computational analysis. The realistic simulations allow to: (i) benchmark computational workflows aiming at studying RNA localization; (ii) assess the descriptive power of features; and (iii) evaluate the overall performance of clustering and classification methods. We developed a set of 23 localization features that could group cells based on seven simulated localization patterns with unprecedented accuracy. We further analyzed experimental data and correctly identified manually annotated localization classes. Importantly, we also demonstrate that simulated smFISH data can be used to train classifiers in a supervised setting, which when applied on real smFISH data allow to observe and quantify heterogeneity in RNA localization.

Manual annotation of smFISH images is complicated, time-consuming and subjective. When manually annotating RNA localization patterns, one typically would like to label cells with respect to the rule according to which RNAs are placed in the cell (e.g., localization close to a membrane or polarized localization). While visually solving this problem can be challenging, it is straightforward to generate images using this rule by simulation. With this, we do not only overcome the shortcomings of manual...
annotation, but we can also use these simulations to identify limitations of the methods and to evaluate the robustness with respect to confounders, such as the density of transcripts or the shape of cells. While we cannot exclude differences between synthetic and real images, we demonstrate that synthetic and real images are similar enough to be used for benchmarking and training.

We explore several different ways of analyzing simulated and experimental data and illustrate how the richness of RNA localization can be explored and quantified by different established statistical approaches. First, we used a qualitative exploration with t-SNE. This analysis not only provides important insights about the existence of different localization patterns but is also informative about the subtleties in these patterns. For instance, it can show the existence of small sub-groups of cells with a very specific localization pattern. Second, we used different unsupervised classification approaches (k-means, hierarchical clustering, and spectral clustering). k-means provides better sensitivity. Complementary to this, hierarchical clustering allows performing a detailed inspection of the data. In this analysis, the number of classes is not pre-defined or inferred and the structure of the data can be investigated to discover sub-clusters. An additional benefit is the interpretability of the features, allowing attribution of biophysical properties to the identified clusters. Lastly, we applied supervised classification with Random Forests. We used the high quality of the simulated data to train a classifier, which could then be applied to experimental data. Supervised classification allows imposing prior biological knowledge and an advantage is to find rare localization classes that can be missed by unsupervised approaches. We also propose a way to quantitatively analyze the localization heterogeneity at the population and intra-cellular level. Which approach is used for a particular data set depends on the underlying biological question and the required level of detail necessary to answer this question.

Automatic analysis of an experimental data set consisting of 10 genes revealed different levels of heterogeneity in terms of RNA localization: heterogeneity at the population and intra-cellular level have different origins and have to be interpreted accordingly. The existence of different localization patterns in different cellular subpopulations for a given gene could be indicative of different biological states, and the localization pattern could be the approximation of these cellular states. Interpretation of pattern mixtures within a single cell is more complicated. If the patterns are mutually exclusive (for instance, if a fraction of RNA localizes at the nuclear envelope and another subset of RNA at the cell membrane), this may point to a double function of the encoded protein or a very high localization dynamic. If the patterns are not mutually exclusive (for instance organization into foci, where foci can be at different locations in the cell), this rather indicates the existence of subclasses, i.e., subtleties inside the defined classes, that have not been defined as separate classes. These subtleties could potentially bare important information, as

Fig. 4 Supervised analysis of experimental data reveals localization heterogeneity. a Results of supervised random forest classification trained on 5 classes (mRNA foci, nuclear envelope in 2D or 3D, cell extension and random) from simulated data. Heatmap shows the majority voting results for all cells of each gene. The simulation classes, “Nuclear envelope 3D” is closest to the experimental class intranuclear. b Posterior probabilities for single cells of the indicated genes. c Example of individual DYNC1H1 cells and their posterior probability. d Scatter plot of Gini impurity calculated on average probability of a gene (population) against the average Gini impurity of individual cells of a gene (intra-cellular)

ARTICLE
they might be indicative of different biological functions, e.g., by interaction with different cellular components. An important case corresponds to mRNA accumulating in foci. This generic localization class represents one of the most frequently observed patterns in Drosophila, but occurs in other model systems as well. Such foci are likely implicated in different biological processes, such as storage in P-bodies, vesicles, or specialized translation factories. Their function can then be investigated by perturbation experiments, additional markers, or secondary criteria such as sub-cellular position.

While we validated and applied the workflow mainly in HeLa cells, we also successfully identified RNA localization patterns in a different cell line (C2C12 cells, Supplementary Note 5). The presented methodology should thus be applicable to other model systems. Recent studies show wide-spread RNA localization in organisms such as Drosophila embryos. While patterns occurring at the scale of the entire embryo (or tissue) would require an extension of the proposed workflow, namely the addition of other landmarks and probably other localization features, there are numerous intra-cellular localization patterns where our methodology could essentially complement these studies.

In this work, we integrated two generic cellular landmarks, namely the cytoplasm and the nucleus, which we believe are important for defining scale in the observed screen. Systematic screening could then reveal interesting localization patterns that are investigated in targeted follow-up experiments employing other markers tailored to the null hypothesis inferred from the large-scale screen. For instance, RNA foci have been observed for a number of genes, but their functional role still needs to be elucidated. Such an analysis requires perturbation experiments and/or the inclusion of additional markers (Supplementary Note 5). We showed previously that DYNCHI foci are neither P-bodies nor stress granules, but act as specialized translation factories. The open design of our approach allows integration of additional markers to discriminate cellular compartments.

In summary, we provide a framework for simulation of synthetic smFISH images and validated methods and tools for the analysis of intracellular RNA localization. We validated this workflow on simulated data and demonstrated how different statistical approaches can be used to investigate the complexity and richness of smFISH images and therefore provide methods and tools to explore the spatial dimension of gene expression inside cells.

Methods

Cell culture. HeLa and C2C12 cell lines were obtained from ATCC and grown in DMEM medium (Gibco) supplemented with 10% FCS (Sigma). HeLa cell lines stably transfected with GFP-tagged BACs were a kind gift from Anthony Hyman (MPI-CBG, Dresden, Germany) and were grown in DMEM medium supplemented with 10% FCS and 0.2 mg/ml BSA (Roche Diagnostics). The next day, cells were washed with PBS, and mounted with Vectashield mounting medium with DAPI (Vector Laboratories).

Imaging to infer cell shape. Three-dimensional image stacks were captured on a wide-field microscope (Zeiss Axiosimager Z1) equipped with a 63 x 1.4 NA objective and a sCMOS camera (Zyla 4.2 MP, Andor Technology) and controlled with Metamorph (Version 7.8.8.0; Molecular Devices).

Image analysis to infer cell shape. 2D segmentation of nuclei and cells was performed with the DAPI and CellMask channels with the open-source software CellCognition using a standard segmentation workflow: Otsu and watershed separation for nuclei in the DAPI channel. Each nucleus then serves as a seed for a watershed segmentation to obtain the cells in the CellMask channel. Individual GAPDH mRNA molecules were localized with FISH-quant in 3D. The Matlab function boundary was used to calculate the containing 3D convex boundary around the mRNAs, corresponding to the 3D cellular outline. GAPDH14 is predominately excluded from the nuclei. To infer the positioning and height of the nuclei, the z-positions of mRNAs localized in small region in the center of the nuclei were analyzed. Therefore, mRNAs were automatically separated into two groups above and below the nucleus by k-means clustering. This allowed inference of the average 3D height and position of the nuclei. For more information, see Supplementary Note 1.

mRNA detection and localization in simulations. mRNA detection in 2D was performed with the MATLAB script provided in Battich et al. Although FISH-quant was performed in 3D, we used it to detect the mRNA distribution in 2D. The function kmeans using 50 replicates; spectral clustering with the function SpectralClustering from Matlab file exchange; hierarchical clustering with the Matlab function clustergram with linkage method ward. Random forest (RF) classification was performed with the Matlab function treebagger. 100 trees were trained on simulated cells with moderate and high pattern strength and mRNA densities corresponding to an average of 100 and 200 mRNA/cell. Gini impurity was calculated from the posterior probabilities for the RF classification: either from the mean probability per mRNA (population impurity) or the mean impurity for all cells of a mRNA (intracellular impurity).

Cell segmentation for mRNA localization. Automated nuclear and cell segmentation was performed with a custom algorithm based on the U-net13,26 deep convolutional network. Nuclear segmentation was performed with the DAPI channel, cell segmentation was performed either with CellMask (if present) or with the actual smFISH image. For segmentation, 3D images were extracted into 2D images with a recently described approach13.

mRNA detection and feature calculation. Analysis of experimental smFISH data was performed as described for simulated data. Except that the detection intensity threshold was determined for each experiment (mRNA) separately. We implemented a simple Matlab user interface FQ_detect, which is now part of FISH-quant.
quant, to facilitate the determination of this threshold. It proposes an automatically calculated threshold based on Otsu thresholding, based on empirically determined criteria. We refer to the FISH-quant documentation for more details.

**Code availability.** We provide the entire Matlab code for the analysis described in this paper, which we integrated into our smFISH analysis package FISH-quant<sup>14</sup> available at https://bitbucket.org/muellertonian/fish_quant

**Data availability**

Data used to simulate images are available on Zenodo https://doi.org/10.5281/zenodo.1413488. Additional experimental data supporting the findings of this study are available from the corresponding authors upon reasonable request.

**Acknowledgements**

We thank Anthony Hyman for providing the HeLa cell lines transfected with BAC. We thank Maria Vera for critical reading of the manuscript. This research was supported by the Agence Nationale de la Recherche (ANR-11-BSV8-018-02 and ANR-14-CE10-0018-01), Institut Pasteur, Fondation pour la Recherche Médicale (FRM DB2013228556), and La Ligue Nationale Contre le Cancer.

**Author contributions**

F.M., T.W., and E.B. conceived the project. F.M. and T.W. supervised A.S. and S.U. and F.M. developed the Matlab package to simulate and analyze RNA localization data. R.C., A.M., M.P., and F.M. performed the smFISH experiments and analyzed the localization patterns with E.B. A.S., F.M., and T.W. analyzed the data. C.Z. provided suggestions on data analysis. W.O. implemented the cell segmentation with neural networks. F.M. and T.W. wrote the manuscript.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06868-w.

**Competing interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**References**


Résumé de la thèse

A. Aperçu de l'introduction

L’introduction commence par discuter des principes, des mécanismes et de la régulation de la localisation et de la traduction locale des ARNm. En effet, le trafic des ARNm donne une dimension supplémentaire à l’expression génique en permettant une production de protéines à des sites précis dans la cellule. Des éléments *cis* et *trans*-actifs sont impliqués dans la régulation du trafic des ARNm. Les éléments agissant en *cis* sont appelés zipcodes et se trouvent souvent dans le 3’UTR des ARNm. Les codes postaux existent dans une variété de format et ils contiennent les informations nécessaires et suffisantes pour localiser la traduction. Dans ce cas, la localisation de l’ARNm est considérée comme un processus dépendant de l’ARN. Les éléments agissant en *trans* constituent en premier lieu des protéines de liaison à l’ARN et capables de se lier aux zipcodes. Elles interviennent dans de nombreux aspects de la vie de l’ARNm et notamment le transit vers des compartiments subcellulaires spécifiques. Le principal mécanisme de trafic utilise le cytosquelette et des moteurs moléculaires pour transporter les mRNPs de manière active. Pendant le transit, des mécanismes garantissent que la traduction des molécules d’ARNm est réprimée, de sorte que la synthèse protéique naissante ne se produise que sur le site cible, où elle est nécessaire.

La deuxième partie expose les techniques d’imagerie de l’ARN et leurs applications dans les cellules vivantes et fixées. Les techniques de FISH à molécule unique (smFISH) sont décrites, y compris des variants utilisant une amplification du signal pour les échantillons présentant un défi optique particulier. Trois cribles de localisation d’ARNm réalisés dans des cellules humaines sont ensuite décrits. Ces cribles utilisent différentes techniques de smFISH, de schémas de marquage et de pipelines d’analyse: (i) bDNA-FISH; (ii) MERFISH; et (iii) seqFISH+. Dans ces cribles, la localisation de l’ARNm a souvent été évaluée en regroupant des transcrits sur la base de leur co-occurrence dans une certaine région sub-cellulaire, telle que les extensions cellulaires et le pourtour du noyau. Bien que ces cribles aient fourni des informations inestimables sur la localisation des ARNm, aucun d’entre eux n’a directement abordé la traduction des ARNm localisés. En outre, ils utilisaient des protocoles d’hybridation complexes. Ces protocoles étaient de plus soit coûteux à mettre en œuvre (comme dans le cas du bDNA-FISH), soit utilisant des méthodes de clarification qui demandent une utilisation intensive de protéases pour réduire le bruit de fond.
(comme dans seqFISH+). Ces méthodes pourraient entraîner une perte d’ARN et/ou de l’intégrité cellulaire. La section suivante est consacrée à la description des techniques d’imagerie des ARNm dans les cellules vivantes. L’accent a été mis sur l’utilisation et l’application d’étiquettes multimérisées telles que le MS2 pour l’imagerie des molécules uniques d’ARNm.

La thématique suivante était toujours dédiée à l’imagerie, mais portait sur les dernières techniques permettant de visualiser la traduction dans des cellules vivantes au niveau d’une seule molécule d’ARN. Plusieurs études, dont une que je présente, ont mis en évidence des polysomes uniques mais toutes ont utilisé le même concept: fusionner des épitopes peptidiques répétés (souvent 24) avec le début de l’ORF. Chaque épitope peptidique est reconnu dès sa sortie du ribosome par un anticorps recombinant fluorescent. Étant donné que les ARNm sont souvent traduits par plusieurs ribosomes, le signal sera encore amplifié, permettant ainsi la détection de polysomes uniques apparaissant comme des foyers en microscopie. Un système qui fonctionne bien est appelé SunTag. Il contient des épitopes reconnus par le fragment variable d’un anticorps monochaine (scFv). Il est intéressant de noter que deux études, dont une des nôtres, a montré que certains polysomes présentaient un mouvement rectiligne rapide indiquant un transport dépendant de moteurs moléculaires. Ainsi, le transport de ces ARNm pourrait se produire de manière co-traductionnel.

Enfin, j’ai présenté un bref aperçu de la structure et de la fonction du centrosome, ainsi que quelques études portant sur la localisation des ARNm aux centrosomes. Un criblage chez des embryons de Drosophile a identifié 6 ARNm enrichis au niveau des centrosomes. Une étude plus récente sur des lignées cellulaires humaines a identifié un ARNm traduit localement au niveau des centrosomes. Cependant, aucune étude dédiée à la caractérisation systématique des ARNm centrosomaux n’a jamais été réalisée.
B. Objectif de ma thèse

Sur la base de ce qui précède, mon doctorat visait à :

• Développer un système permettant de visualiser et de quantifier la traduction de molécules uniques d'ARNm.

• Caractériser de manière systématique la traduction locale des ARNm dans les cellules humaines, en mettant l’accent sur les centrosomes.

• Comprendre si la localisation des ARNm est un processus contrôlé par l'ARN ou les protéines naissantes, et déterminer l’ampleur du ciblage co-traductionnel des ARNm.

• Étudier les mécanismes du ciblage de l’ARNm à l'aide d'un transcrit modèle.
C. Aperçu des résultats

Les résultats de ma thèse sont présentés sous la forme de trois manuscrits:

**Manuscrit 1: la visualisation de polysomes uniques dans des cellules vivantes révèle la dynamique de la traduction**

**Ce travail avait les objectifs suivants:**

- Développer un système basé sur le SunTag pour visualiser la traduction de molécules uniques d'ARNm

- Combiner le système SunTag avec l'édition de gènes pour visualiser des polysomes endogènes.

- Caractériser la traduction de plusieurs ARNm dans des cellules vivantes. Ceci incluait la cinétique de traduction comme les vitesses d’elongation, la densité des ribosomes et leur espacement.

**Les principales conclusions du manuscrit sont les suivantes:**

- Le SunTag permet de visualiser la traduction de molécules uniques d'ARNm.

- Le SunTag peut être combiné avec l'édition de gène pour visualiser les polysomes endogènes.

- La traduction des transcrits endogènes alterne de manière stochastique entre des états actifs et inactifs, sur une échelle de la dizaine de minutes.

- Les ARNm DYNC1H1 sont traduits dans des usines de traduction dédiées.

**Dans cette étude, j'ai spécifiquement :**

- Effectué un knock-in SunTag en utilisant CRISPR-Cas9 pour marquer les gènes endogènes POLR2A et DYNC1H1.
• Criblé et caractérisé des clones SunTag POLR2A et DYNC1H1 par génotypage, des expériences de smFISH et par inspection visuelle des résultats.

• Analysé et interprété mes données, notamment en concluant que les ARNm de POLR2A sont traduits sous la forme de polysomes individuels, tandis que ceux de DYNC1H1 sont traduits dans des usines de traduction dédiées.

• Généré des figures et commenté le texte.
Manuscrit 2: Un crible de localisation révèle des usines de traduction et un ciblage fréquent des protéines de manière co-traductionnelle

Ce document a les objectifs suivants:

• Réaliser un double crible de localisation ARNm/protéine sur des gènes sélectionnés de manière aléatoire, ainsi que sur des gènes codant pour les moteurs moléculaires. La stratégie était d’utiliser une approche BAComics dans laquelle chaque lignée cellulaire contient une version d’un gène marquée par la GFP.

• Identifier, décrire et quantifier les caractéristiques de la localisation de certains ARNm, ainsi que des cas de traduction locale, en analysant la co-localisation protéine/ARNm et en visualisant des polysomes.

• Déterminer si et dans quelle mesure les transcrits localisés dépendent de signaux présents sur l’ARN ou les protéines naissantes.

• Réaliser des études fonctionnelles sur un candidat localisé pour comprendre la fonction biologique de la localisation de cet ARNm.

Les principales conclusions du document sont les suivantes:

• Le ciblage des protéines de manière co-traductionnelle est répandu.

• Certains ARNm sont traduits dans des usines de traduction spécifiques.

• Dans le cas de la β-caténine, les usines de traduction sont spécialisées dans la dégradation co-traductionnelle de la protéine naissante.

A cette étude, j’ai spécifiquement contribué en:
• Criblant plus de 60 ARNm à l’aide du smFISH. En particulier, j’ai identifié un ARNm qui se localise dans les centrosomes (HMMR), ce qui m’a amené à m’intéresser à cette classe de localisation particulière.

• Confirmant la localisation obtenue à partir des BAC, en inspectant et en interprétant la localisation des transcrits endogènes à l’aide du smiFISH.

• Acquérant et contrôlant la qualité d’un large ensemble d’images qui sont utilisées pour quantifier les profils de localisation des ARNm.

• Acquérant et vérifiant la qualité d’un autre jeu de données qui est utilisé pour quantifier les effets de divers traitements sur la distribution d’ARNm et la co-localisation avec leurs protéines.

• Décrivant et interprétant le détail des localisations de trois ARNm (ASPM, HMMR et NUMA1), y compris leur localisation au niveau des pores nucléaires et des centrosomes.

• Concevant, clonant et générant un clone SunTag-ASPM. L’interprétation des images obtenues avec ce clone m’a permis de démontrer que: (i) cet ARNm est parfois traduit au niveau d’usines de traduction, et (ii) cet ARNm est traduit localement au niveau des pores nucléaires, où il reste en moyenne 20 minutes.

• Triant et maintenant de nombreuses lignées cellulaires utilisées tout au long de cette étude.

• Quantifiant les images, préparant les figures, commentant et corrigéant le texte.

Ce manuscrit est en cours de révision à Cell.
Manuscrit 3: L'identification systématique des ARNm centrosomaux révèle un mécanisme de localisation des polysomes dépendant de moteurs moléculaires et de la protéine naissante

Ce travail avait les objectifs suivants:

- Cribler les ARNm codant pour toutes les protéines centrosomales en utilisant un nouveau protocole smFISH à haut-débit, de manière à identifier de nouveaux ARNm centrosomaux.

- Décrire en détail la localisation des ARNm centrosomaux au cours du cycle cellulaire.

- Déterminer si le ciblage des ARNm vers les centrosomes dépend de leur traduction.

- En utilisant ASPM comme modèle, visualiser des ARNm et des polysomes endogènes dans des cellules vivantes, afin de mieux comprendre les mécanismes présidant au transport de cet ARNm.

Les principales conclusions de cette étude sont :

- La mise en évidence d’un programme de traduction régulé par le cycle cellulaire au niveau des centrosomes.

- La localisation des ARNm au niveau les centrosomes dépend toujours de leur traduction.

- Les ARNm ASPM sont transportés de manière co-traductionnelle vers les centrosomes, à l'aide de moteurs moléculaires et de microtubules.

Dans cette étude, j’ai effectué, analysé et interprété toutes les expériences (à l’exception notable du pipeline de criblage). J’ai écrit le manuscrit et préparé toutes les figures.


