Université de Montréal

CHARACTERISING (PRE-)MRNP ORGANISATION AT DIFFERENT STAGES OF GENE REGULATION USING SINGLE-MOLECULE MICROSCOPY

par

Srivathsan Adivarahan

Département de biochimie

Faculté de médecine

Thèse présentée à la Faculté de médecine en vue de
l'obtention du grade de doctorat en biochimie

July 2021

Université de Montréal

Département de biochimie et médecine moléculaire, Faculté de médecine

---

Cette thèse est intitulée

CHARACTERISING (PRE-)MRNP ORGANISATION AT DIFFERENT STAGES OF GENE REGULATION USING SINGLE-MOLECULE MICROSCOPY

Présentée par

Srivathsan Adivarahan

A été évaluée par un jury composé des personnes suivantes

James Omichinski
Président-rapporteur

Daniel Zenklusen
Directeur de recherche

Martin Sauvageau
Membre du jury

Timothy Stasevich
Examinateur externe

Résumé

Les ARNm sont des molécules centrales pour la régulation des gènes, aidant à convertir l'information génétique stockée dans l'ADN en protéines fonctionnelles. En tant que polymère simple brin, mesurant des centaines à des milliers de nucléotides, les ARNm peuvent former des structures secondaires et tertiaires étendues formant des particules appelés ribonucléoprotéines messagères (RNPm) en s'assemblant avec des protéines. L'organisation 3D des (pré-)RNPm influence de nombreux aspects de leur métabolisme, incluant la régulation de leur maturation, de leur export et de leur traduction dans le cytoplasme. Malgré leur importance, notre compréhension de l'organisation structurelle des (pré-)RNPm in vivo, et des principes qui la régissent est minime.

Au cours de ma thèse, j'ai analysé l'organisation des (pré-)mRNP en développant une vision centrée sur l'ARN. Pour cela, j'ai mis en place une approche combinant l'hybridation in situ d'ARN monomoléculaire (smFISH) avec la microscopie à illumination structurée (SIM) et l'ai utilisée pour étudier l'organisation des mRNP dans le noyau et le cytoplasme. Nos résultats suggèrent que l'organisation (pré-)mRNP varie à différents stades de sa vie. Nous montrons que l'empaquetage (pré-)mRNP commence de manière co-transcriptionnelle, avec des introns organisés en conformations compactes. Cette organisation est modifiée au cours de la transcription au fur et à mesure que la polymérase se déplace le long du gène, assemblant finalement un intron avec les extrémités à proximité l'une de l'autre, d'une manière dépendante du spliceosome, suggérant que l'organisation co-transcriptionnelle des introns pourrait être critique pour déterminer son excision. Une fois libérés, les mRNP ont une organisation linéaire compacte dans le nucléoplasme et éventuellement une conformation en tige. L'organisation d'un mRNP dans le cytoplasme est influencée par sa traduction. Alors que la traduction ouvre les mRNP, la séparation des extrémités de l'ARNm, l'inhibition de la traduction et la libération de ribosomes, ou le recrutement dans les granules de stress, donnent aux mRNP une structure très compacte. Fait intéressant, nous trouvons rarement des ARNm avec les extrémités 5' et 3' à proximité, ce qui suggère que la traduction en boucle fermée n'est pas un état universel pour tous les ARNm en cours de traduction. Ensemble, nos résultats fournissent une image essentielle de l'organisation du mRNP dans les cellules et souligne le rôle important de la conformation du RNPm dans la régulation de la traduction et de la maturation d'une RNPm.

Mots clés: organisation des mRNPs, structure de l'ARNm, modèle en boucle fermée, communication 5'-3', traduction des ARNm, épissage des ARNm, compactage des ARN, imagerie des ARNs, microscopie à super-résolution, smFISH

Abstract

mRNAs act as the central molecules in gene regulation, helping convert the genetic information stored in the DNA to functional proteins. As a single-stranded polymer, hundreds to thousands of nucleotides in length, mRNAs can form extensive secondary and tertiary structures and, together with proteins, are packaged into assemblies called messenger ribonucleoproteins (mRNPs). The 3D organisation of (pre-)mRNPs influences many aspects of what happens to them, including regulating their processing, export and translation in the cytoplasm. Despite their significance, our understanding of the structural organisation of (pre-)mRNPs *in vivo* is minimal, as is our comprehension of the principles that govern it.

During my PhD, I have developed an RNA-centric view on (pre-)mRNP organisation. For this, I have established an approach combining single-molecule RNA *in situ* hybridisation (smFISH) with structured illumination microscopy (SIM) and used it to study mRNP organisation in the nucleus and cytoplasm. Our results suggest that (pre-)mRNP organisation is altered at various stages during its lifetime. We show that (pre-)mRNP packaging starts co-transcriptionally, with introns organised into compact conformations. This organisation is altered during the course of transcription as the polymerase travels along the gene, finally assembling an intron with the ends in proximity in a spliceosome dependent manner, suggesting that co-transcriptional intron organisation could be critical in determining its excision. Once released, mRNPs have a compact linear organisation in the nucleoplasm and possibly a rod-like conformation. mRNP organisation in the cytoplasm is influenced by its translational status. While translation opens up mRNPs, separating the ends of the mRNA, translation inhibition and release of ribosomes, or recruitment to stress granules result in mRNPs having a highly compact structure. Interestingly, we rarely find mRNAs with the 5' and 3' ends in proximity, suggesting that closed-looped translation is not a universal state for all translating mRNAs. Together, our results provide a unique and essential view of mRNP organisation in cells and reveal important insight into the role of mRNP conformation in regulating translation and mRNP processing.

Keywords: mRNP organisation**,** mRNA structure, closed-loop model, 5'-3' communication, mRNA translation, mRNA splicing, RNA compaction, RNA imaging, super-resolution microscopy, smFISH

Table of Contents

Résumé ... ii
Abstract ... iv
Table of Contents ... v
List of Figures ... ix
List of Tables ... xi
List of abbreviations ... xii
Acknowledgements ... xv
1. Introduction ... 2
1.1 Foreword ... 3
1.2 The general importance of mRNP packaging and organisation ... 3
1.2.1 Preventing the formation of R-loops and maintaining genomic stability ... 4
1.2.2 Regulating the size of the mRNP complex ... 4
1.2.3 Regulating exposure of certain regions of the mRNA ... 4
1.2.4 Facilitating interaction between different regions of the mRNA ... 5
1.3 mRNP composition and its role in mRNP packaging and organisation ... 5
1.3.1 Discovery of RBPs and their target mRNAs sequences ... 5
1.3.2 The role of RBPs in the metabolism of Pol II transcripts, RNP compaction and organisation ... 7
1.3.2.1 Heterogeneous nuclear ribonucleoproteins (hnRNPs) ... 7
1.3.2.2 SR proteins and the exon junction complexes (EJCs) ... 10
1.3.2.3 Transport and Export (TREX) complex ... 13
1.3.2.4 Y-box proteins ... 15
1.3.3 The role of mRNA in mRNP metabolism and RNP organisation ... 18
1.3.3.1 Short and long-range interaction within mRNAs can bridge different regions and compact mRNPs ... 18
1.3.3.2 RNA Modifications can alter mRNP composition and organisation ... 21
1.4 The synergy between mRNA structure and RBP binding ... 22
1.4.1 mRNA structure and its effect on RBP composition ... 22
1.4.2 RBPs can alter local and global mRNA folding ... 23
1.4.3 The interplay between RBPs affects mRNP composition ... 24
1.5 mRNP assembly and its importance in mRNA biogenesis ... 24
1.5.1 Co-transcriptional assembly of mRNPs ... 24
1.5.1.1 Nascent Balbiani ring mRNPs are assembled into compact structures ... 25

1.5.1.2 mRNP composition and gene structure could alter pre-mRNP packaging 29
1.5.2 mRNP processing – spliceosome assembly, intron compaction and excision..... 29
1.5.2.1 Assembly of the spliceosome and the splicing reaction..... 30
1.5.2.2 Role of pre-mRNP packaging in splice site selection 32
1.5.2.3 Intron packaging and compaction 34
1.5.2.4 Bridging the 5' and 3' ends of introns..... 35
1.5.3 mRNP organisation in the nucleoplasm and during export 36
1.5.3.1 Nuclear mRNPs are possibly linearly organised..... 37
1.5.3.2 mRNP remodelling during export..... 37
1.5.4 mRNP organisation in the cytoplasm and its role in regulating translation and degradation..... 38
1.5.4.1 Local mRNA structure and its role in regulating translation and stability..... 39
1.5.4.2 5'-3' communication tunes translation and mRNA stability 39
1.5.5 mRNP compaction and organisation in the cytoplasm..... 43
1.6 Research objectives of this work..... 46
2. Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway..... 48
2.1 Aims of Article 1 49
2.2 Article 1 50
2.2.1 Summary..... 51
2.2.2 Introduction..... 52
2.2.3 Results..... 54
2.2.3.1 Visualising different regions within mRNAs using smFISH and SIM..... 54
2.2.3.2 Open mRNP conformation corresponds to translating mRNA..... 57
2.2.3.3 Inhibiting eIF4G1-PABPC1 interaction does not alter open mRNP conformations 59
2.2.3.4 Ribosome occupancy determines mRNP compaction 61
2.2.3.5 Compaction state of lncRNAs and mRNA sequestered to stress-granules..... 63
2.2.3.6 Organisation of nuclear mRNAs..... 65
2.2.4 Discussion..... 67
2.2.4.1 Nuclear mRNPs show a linear organisation..... 67
2.2.4.2 Variable levels of RNP compaction in cells..... 68
2.2.4.3 Closed-loop translation and regulation of gene expression 69
2.2.5 Acknowledgement 71
2.2.6 Author contributions 71

2.2.7 References 72
2.2.8 Materials and Methods 77
2.2.9 Supplementary Figures 85
2.2.10 Supplementary Tables 92
3. Single-molecule imaging suggests compact and spliceosome dependent organisation of long introns 111
3.1 Aims of Article 2 112
3.2 Article 2 113
3.2.1 Summary 114
3.2.2 Introduction 115
3.2.3 Results 116
3.2.3.1 Introns are organised as compact particles with the ends in proximity 116
3.2.3.2 Intron organisation is altered during transcription 120
3.2.3.3 Spliceosome assembly on pre-mRNAs determines the final organisation of introns 123
3.2.4 Discussion 127
3.2.4.1 Introns are co-transcriptionally packaged into compact particles 127
3.2.4.2 5'-3' ends are possibly bridged through U1snRNP-Pol II tethering 128
3.2.5 Materials and Methods 129
3.2.6 Acknowledgement 132
3.2.7 Author contributions 132
3.2.8 References 132
3.2.9 Supplementary Figures 135
3.2.10 Supplementary Tables 144
4. Discussion 157
4.2 General goals of the thesis 158
4.3 Article 1: Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway 158
4.3.4 Objectives and summary of results 158
4.3.5 Nuclear mRNP organisation 159
4.3.5.1 Many nuclear mRNPs are linearly organised into rod-like conformations 159
4.3.5.2 Expanding our understanding of nuclear mRNP organisation 161
4.3.5.3 Is there heterogeneity in nuclear mRNPs of the same transcript? 164
4.3.5.4 mRNP transit through the nuclear pore complex 165
4.3.6 Cytoplasmic mRNP organisation 166

4.3.6.1 Ribosomes regulate local mRNP compaction 167
4.3.6.2 Non-translating mRNPs and stress granule dynamics 169
4.3.6.3 Implication for the closed-loop model and 5'-3' communication 170
4.3.6.4 The compaction of 3' UTRs and its effect on mRNP metabolism 176
4.4 Article 2: Single-molecule imaging suggests compact and spliceosome dependent organisation of long introns 177
4.4.1 Objectives and summary of results 177
4.4.2 Introns are compact assemblies *in vivo* 178
4.4.3 Spliceosome assembly and co-transcriptional intron organisation 180
4.4.4 An alternate model for splicing of long introns – recursive splicing 183
4.4.5 Determining the model for spliceosome deposition – intron vs exon definition 184
4.4.6 Effect of intron looping on (pre-)mRNP organisation 186
4.5 Concluding remarks 187
4.6 The 3D organisation of mRNPs and perspectives for future research 188
A. Annex 192
a. Additional Figures 192
b. Lessons from (pre-) mRNA imaging 197
c. Probing the conformational state of mRNPs using smFISH and SIM 246
Bibliography 269

List of Figures

Figure 1-1: hnRNP C structures observed in vitro 10

Figure 1-2: mRNP compaction through EJC-EJC and EJC-SR protein collaboration 13

Figure 1-3: Structure of human TREX complex and putative role of ALYREF 15

Figure 1-4: Domain organisation of YB-1 17

Figure 1-5: Exon-intron organisation of BR genes 27

Figure 1-6: Assembly of BR mRNPs 28

Figure 1-7: pre-mRNA splicing cycling 31

Figure 1-8: Models for hnRNPA1 mediated exon skipping 34

Figure 1-9: Schematic presentation of BR mRNP export 38

Figure 1-10: Factors associating with 5' and 3' ends of mRNAs in the cytoplasm 40

Figure 1-11: Polysome conformations observed in vivo 46

Figure 2-1: Visualising single mRNA reveals open conformations of cytoplasmic mRNAs .56

Figure 2-2: Open mRNP conformation corresponds to translating mRNA. 58

Figure 2-3: Inhibiting eIF4G1-PABC1 interactions does not alter 5'-3' distances. 60

Figure 2-4: Ribosome occupancy determines mRNP compaction. 62

Figure 2-5: lncRNAs in the cytoplasm and mRNAs sequestered to stress granules show compact conformations. 64

Figure 2-6: Organisation of nuclear MDN1 mRNAs. 66

Figure 3-1: Organisation of intron 36 of the POLA1 gene 120

Figure 3-2: Co-transcriptional assembly of intron 35 of the POLA1 gene 123

Figure 3-3: Intron Organisation is altered upon inhibition of U2 assembly 126

Figure 4-1: EJC multimerisation can result in mRNP organisation into a linear-rod 163

Figure 4-2: The effect of ribosome occupancy and ribosome distribution on local secondary structures 168

Figure 4-3: Translation status and length of mRNAs associating with closed-loop factors in S. cerevisiae 175

Figure 4-4: Cotranscriptional assembly and compaction of intron and the possible establishment of the spliceosome 186

Figure A-1: Comparing 5'-3' distances for nuclear and cytoplasmic non-translating mRNPs 193

Figure A-2: Distances for MDN1 and AHNAK nuclear mRNAs 195

Figure A-3: Cytoplasmic mRNP conformations – shorter mRNAs 196

Supplementary Figure 2-1: Positions of smFISH probes used in this study. 85

Supplementary Figure 2-2: Visualising mRNP conformations for MDN1, POLA1 and PRPF8 and measurement of 5'-3' distances for MDN1 in 2D, 3D and in cells permeabilized using either TritonX or Ethanol 86

Supplementary Figure 2-3: 5'-3' distance measurements for MDN1 mRNA upon treatment with homoharringtonine and visualising mRNP conformation of single POLA1, PRPF8 and MDN1 mRNAs when treated with Puromycin. 87

Supplementary Figure 2-4: Visualising mRNP conformations in mutant cell lines of PABPC1 and eIF4G1. 88

Supplementary Figure 2-5: Compaction of the 5' end is altered upon a pulsed homoharringtonine treatment for 10min. 89

Supplementary Figure 2-6: mRNA and lncRNA compaction and accumulation in stress granules. 90

Supplementary Figure 2-7: Compaction of nuclear MDN1 mRNA upon puromycin or homoharringtonine treatment. 91

Supplementary Figure 3-1: Positions of smFISH probes used in this study 135

Supplementary Figure 3-2: Intron 36 smFISH distance distributions 137

Supplementary Figure 3-3: Splicing inhibition upon Pladienolide B treatment 138

Supplementary Figure 3-4: Pladienolide B treatment alter intron organisation 140

Supplementary Figure 3-5: Organisation of long AHNAK exon 141

Supplementary Figure 3-6: mRNA organisation upon Pladienolide B treatment 142

List of Tables

Table S2-1: Primers used for making CRISPR/Cas9 cells lines 92

Table S2-2: List of smFISH probes used 94

Table S2-3: Probe and antibody combinations used 109

Table S 3-1: List of smFISH probes used 144

Table S 3-2: Probe combinations used 155

List of abbreviations

4E-T	*eIF4E-transporter protein*
AFM	*atomic force microscope*
AS	*antisense*
ATP	*adenosine tri-phosphate*
BR	*Balbiani ring*
Cas9	*CRISPR associated protein 9*
CBC	*cap-binding complex*
CBP20	*cap binding protein 20*
CBP80	*cap binding protein 80*
CDS	*coding sequence*
ChIP	*chromatin immunoprecipitation*
CHX	*cycloheximide*
CPF	*cleavage and polyadenylation factor*
CRISPR	*clustered regularly interspaced short palindromic repeats*
DAPI	*4',6-diamidino-2-phenylindole*
DMS	*dimethyl sulfate*
DNA	*Deoxyribonucleic acid*
DRB	*5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole*
dSTORM	*direct stochastic optical reconstruction microscopy*
eCLIP	*enhanced crosslinking and immunoprecipitation*
eIF4A1	*eukaryotic translation initiation factor A1*
eIF4A3	*eukaryotic translation initiation factor A3*
eIF4E	*eukaryotic translation initiation factor 4E*
eIF4EBP1	*eukaryotic translation initiation factor 4E binding protein 1*
eIF4F	*eukaryotic translation initiation factor 4F*
eIF4G1	*eukaryotic translation initiation factor 4G1*
EJC	*exon junction complex*
EM	*electron microscopy*
ENCODE	*encyclopedia of DNA elements*
ER	*endoplasmic reticulum*
ESE	*Exonic splicing enhancer*
ESS	*Exonic splicing silencer*
ET	*electron tomography*
FCS	*fluorescence correlation spectroscopy*
FISH	*fluorescent in situ hybridisation*
FRAP	*fluorescence recovery after photobleaching*
FRET	*Förster resonance energy transfer*
GFP	*green fluorescent protein*
HEK293	*human embryonic kidney 293*

HEK293T	*human embryonic kidney 293 T*
hnRNP	*heterogeneous nuclear ribonucleoprotein*
hnRNPA1	*heterogeneous nuclear ribonucleoprotein A1*
hnRNPA2B1	*heterogeneous nuclear ribonucleoprotein A2/B1*
hnRNPC	*heterogeneous nuclear ribonucleoprotein C*
iCLIP	*individual-nucleotide resolution UV crosslinking and immunoprecipitation*
icSHAPE	*in cell Selective 2′ Hydroxyl Acylation analysed by Primer Extension*
ISE	*intronic splicing enhancer*
ISS	*intronic splicing silencer*
lncRNA	*long non-coding RNA*
m1A	*N1-methyladenosine*
m5C	*5-Methylcytosine*
m6A	*N6-methyladenosine*
m7Gpp	*7-methylguanosine cap*
MCP	*MS2 coat protein*
MDN1	*Midasin AAA ATPase 1*
miRNA	*micro RNA*
mRNA	*messenger RNA*
mRNP	*messenger ribonucleoprotein*
NAB2	*nuclear polyadenylated RNA-binding 2*
NCBP1	*nuclear cap binding protein 1*
NCBP2	*nuclear cap binding protein 2*
NPC	*nuclear pore complex*
ncRNA	*non-coding RNA*
ORF	*open reading frame*
PABPC1	*poly(A) binding protein cytoplasmic 1*
PALM	*photoactivated localisation microscopy*
PAP1	*poly(A) polymerase 1*
PAR-CLIP	*photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation*
PARIS	*psoralen analysis of RNA interactions*
PB	*Pladienolide B*
PCP	*PP7 coat protein*
PlaB	*Pladienolide B*
POLA1	*polymerase (DNA) alpha 1, catalytic subunit*
Pre-mRNA	*Precursor messenger ribonucleic acid*
Pre-mRNP	*Precursor messenger ribonucleoprotein*
PRPF8	*pre-mRNA processing factor 8*
qPCR	*quantitative polymerase chain reaction*
RBD	*RNA binding domain*
RBP	*RNA binding protein*
RIC	*RNA interaction capture*
RIP	*RNA immunoprecipitation*
RIPiT	*RNA: protein immunoprecipitation in tandem*
RIPPLiT	*RNA immunoprecipitation and proximity ligation in tandem*

RNA	*Ribonucleic acid*
RNAPII	*RNA polymerase II*
RRM	*RNA recognition motif*
rRNA	*ribosomal RNA*
RS	*recursive splicing*
RT-qPCR	*reverse transcriptase qPCR*
SF1	*Splicing Factor 1*
sgRNA	*single-guide RNA*
SHAPE	*selective 2'-hydroxyl acylation analysed by primer extension*
SIM	*structured illumination microscopy*
siRNA	*small interfering RNA*
SLBP	*stem loop binding protein*
smFISH	*single molecule FISH*
smFRET	*single molecule Förster resonance energy transfer*
snoRNA	*small nucleolar RNA*
snRNA	*small nuclear RNA*
snRNP	*small nuclear RNP*
SR	*Serine and arginine-rich*
ss	*splice site*
SUNtag	*SUperNova tag*
TCERG1	*Transcription Elongation Regulator 1*
TOP	*Terminal oligopyrimidine*
TREX	*transcription and export*
tRNA	*transfer RNA*
TS	*transcription site*
TSS	*transcription start site*
TTS	*transcription termination site*
TUG1	*taurine up-regulated 1*
U2AF	*U2 Small Nuclear RNA Auxiliary Factor 2*
U2OS	*U2- osteosarcoma*
UTR	*untranslated region*
XRN1	*exoribonuclease 1*
YB-1	*Y-box binding protein 1*
YBX1	*Y-box binding protein 1*

Acknowledgements

The years that I have spent during my PhD have been an immense learning experience, both scientifically and personally. I would like to show my appreciation towards those who have helped me reach here, without whom I would not be writing this at all. First of all, I would like to express my gratitude towards my PhD advisor, Dr Daniel Zenklusen, for his guidance, unwavering support, and for being an amazing human being. He gave me the freedom to explore different scientific ideas and supported me at times when even I didn't believe in myself. No matter the topic, he has always welcomed people knocking at his door for insightful and sometimes crazy discussions. During my time in this lab, I have improved my scientific communication skills, both in writing and presentations, and have become an overall independent and better scientist capable of developing and pursuing my own ideas. We might disagree on many things, but I appreciate his inputs and the effort he puts in to support the students in his lab.

I am grateful to the people in the Zenklusen laboratory and at the university, both past and present, whom I have worked or interacted with during my time here. First, Pascal, for being the bedrock of this laboratory, taking care of all our scientific needs, sometimes even during his vacation days (sorry, Pascal!). I might have troubled him more often than perhaps anyone else in the lab before or since, and I appreciate his patience towards me. During my time in the lab, I also have had the privilege of mentoring several summer students – Xuedong, Rushikesh, Axel and Zhengyang. Not only has it been a huge learning experience for me, having mentored for the first time in my life, but it was also immensely fulfilling to see many of them pursuing graduate studies in their fields of interest. I wish them the best in their careers and hope that our paths cross again in the future. I owe special thanks to Samir, with whom I worked when I joined the lab and learned the tricks of the trade, Kalo, with whom I worked at the end of my PhD and others, including Om, Bineta, Kevins, Justina, Catherine and Cornelia, for making the lab a lovely environment to work in. I also owe a special thanks to Nicolas for tolerating me whenever I bothered him with both scientific and random topics. Finally, I would like to thank Pierre, my fellow graduate student and thesis writing companion, for helping me throughout my time here and being the wonderful person he is. As he is also submitting his thesis alongside me, I wish him the best for his defence and future. Those crazy evening chats on Star Wars, Rick and Morty, Bladerunner, LOTR, space (and NASA),

football and sometimes science will be missed. The Zenklusen and Oeffinger labs will not be the same without you.

My time in Montreal has had its tough moments. Being away from most of my family and friends, especially during the pandemic, has been a weird experience. So I would like to thank my friends, who have supported me during this time. Jean, my roommate throughout my time in Montreal, has been the pillar of support, helping me whenever I needed any. Also, those chocolate cakes and the times we boiled and ate blé d'Inde are going to be missed. And heartful thanks to all my friends either in Montreal or throughout the world, including, Kousik, Kami, Shurthy, Nishant, Purval, DAS, Tejas, Bonda, Patel, Mods, and Smarjeet, for supporting and motivating me at times when I needed it.

I am where I am because of my parents and my brother. They have been the pillars throughout my entire life, and their love, support and encouragement have been crucial for me to reach this point in my PhD. I owe a big thanks to them for everything they have done for me and hope they will forgive me for the endless times I might not have picked up their calls. I would also like to thank my family, including my Patti, aunts, uncles, and cousins, Shilpa, Sathya Kutti, Hayagreev, and Lekshmi's parents. We might not communicate as often as we did when I was in India, but it is always a great feeling to meet everyone and spend hours and days catching up on things whenever I visit India. As I finish my PhD, Sathya Kutti is starting hers, and I wish her all the best for her adventure and tell her that we should talk more often.

Finally, I would like to thank Lekshmi for being the ever-present emotional support during my PhD and beyond. I could not have asked for a better partner and have to count myself extremely lucky for getting to spend my life with her (as I am rightfully reminded once in a while). These past few years have been challenging for us being away on different continents, and now that it is drawing to a close, I am eagerly looking forward to not having to worry about time zones, travel restrictions and visas and just the boring things like "what are we going to have for dinner tonight?". Maybe someday we can look back at this period and laugh at the insanity of it all.

1. Introduction

1.1 Foreword

The regulation of gene expression is an essential step for cells and organisms to develop and maintain homeostasis. mRNAs play a central and critical role in this process, acting as intermediary molecules helping convert the genetic information stored within an organism's DNA into protein molecules that perform much of the cellular function. Despite the often short life of mRNAs, controlling their metabolism is among the most complex cellular processes composed of numerous steps, many subjected to regulation and quality control involving hundreds to thousands of proteins. The regulation of mRNA metabolism in cells is mediated through the binding of RNA-binding proteins (RBPs), which together with the mRNA assemble an mRNA-protein complex called messenger ribonucleoprotein (mRNP). While the primary sequence of the mRNA coding for a protein remains the same, the composition of an mRNP is dynamic and is altered throughout its lifetime, with several distinct proteins associating with and regulating different processes essential for the function and stability of the mRNA (Gehring et al., 2017; Müller-McNicoll and Neugebauer, 2013; Singh et al., 2015). Through such dynamic associations, the mRNP is altered not only in its compositions but also in its 3D spatial organisation. Though RNA-binding proteins and mRNA sequences and structures have been studied extensively to determine their role in regulating different steps of mRNA metabolism, very little is understood about how the mRNA and the bound RBPs assemble and organise an mRNP in 3D. mRNP organisation is critical for many aspects of mRNA metabolism, from transcription to mRNA processing, export, translation, storage, and degradation. Therefore, a better comprehension of the biophysical properties of mRNPs and the principles and mechanisms that govern mRNP packaging and organisation in cells is essential towards obtaining a fundamental view on how gene expression is regulated in cells and how, as a result, cells can regulate and maintain homeostasis.

1.2 The general importance of mRNP packaging and organisation

Eukaryotic mRNAs need to be processed in the nucleus and exported to the cytoplasm, where they are translated into proteins. To ensure that these steps are carried out correctly, cells have developed several mechanisms to verify and regulate different steps of mRNA biogenesis. One such regulatory step involves the packaging and organisation of mRNPs, which has relevance for several reasons. Besides preventing RNAs from being randomly attacked by nucleases, mRNP

packaging and organisation play a significant role in numerous processes, including prevention of R-loop formation, regulating and determining the size of the mRNP complex, regulating the regions of mRNAs that are available for recognition, and facilitating the interaction between different regions of the mRNA.

1.2.1 Preventing the formation of R-loops and maintaining genomic stability

On average, human pre-mRNAs are much longer than the fully processed transcript, with introns making up to 95% of the total pre-mRNA sequence in humans (Venter et al., 2001). These long transcripts must be packaged and compacted to ensure that the newly synthesised transcripts do not bind to the DNA forming DNA-RNA hybrids called R loops. The formation of R-loops is detrimental to the survival of cells, as it hinders transcription and can also serve as a hot spot for recombination events resulting in genomic instability (Crossley et al., 2019). Co-transcriptional packaging and assembly of an mRNP can help separate the nascent transcript from the DNA and prevent the annealing of the nascent strand at the transcription bubble (see 1.5.1).

1.2.2 Regulating the size of the mRNP complex

The final processed eukaryotic mRNAs can vary in size ranging from a few hundred to more than a hundred thousand nucleotides. Compaction of mRNPs can help diffusion through the crowded interchromatin and nucleoplasmic environment, especially for mRNPs containing long mRNAs. Additionally, the size of an assembled mRNP could determine its passage through the nuclear pore complex (NPC). The pore's central channel has a fixed size, estimated to be~30 nm in yeast to ~50 nm in humans (Lin and Hoelz, 2019). The compaction and organisation of mRNPs could aid in its transit through the NPC and help it reach the cytoplasm (see 1.5.3.2).

1.2.3 Regulating exposure of certain regions of the mRNA

As mRNP’s regulation requires the specific recognition of certain sequence elements, mRNP packaging could be an approach cells use to control the accessibility of the mRNA to certain regulatory factors. This is potentially important in regulating processes like alternative splicing or association of miRNAs to silence or degrade mRNAs (see 1.5.2.2).

1.2.4 Facilitating interaction between different regions of the mRNA

mRNP compaction and organisation are a means to bridge the gap and bring together regions of RNAs separated by hundreds to thousands of nucleotides. This reduction in spatial separation and the general biophysical properties of assembled mRNPs can aid in the interaction between proteins/regulatory complexes bound to these regions, which are essential to help modulate the regulatory stages of mRNA lifecycle like mRNA splicing, translation and degradation (see 1.5.2.4 & 1.5.4.2).

1.3 mRNP composition and its role in mRNP packaging and organisation

To understand the principles that determine mRNP organisation in cells, we need to understand how its two components, the primary RNA sequence and the RBPs, contribute to organising an mRNP. Various biochemical studies have provided an extensive overview of the proteins capable of associating with mRNAs and the role of specific RBPs in regulating mRNP organisation and metabolism. Furthermore, structural studies of either in-vitro assembled or purified RNA-protein complexes have shed some light on the contribution of specific proteins in mRNP organisation. Extensive effort has also been made to identify the binding targets for RBPs to help identify the mechanisms through which these proteins act and possible physiological roles resulting from mutations within the binding targets.

The contribution of mRNAs, which as single-stranded molecules can form extensive secondary and tertiary structures in solutions, has also been studied extensively. Combining bioinformatics and biochemical approaches have helped identify the mRNA regions that fold, forming long and short-range interactions in cells and *in vitro*. Together, these studies have highlighted the importance of RBPs, RNA-protein complexes, mRNA folding, and processes that participate in the assembly of mRNPs and have helped understand how this assembly can help regulate mRNAs' metabolism in cells.

1.3.1 Discovery of RBPs and their target mRNAs sequences

Evidence for proteins with RNA-binding activity was first discovered when nascent RNAs emerging from the chromosome were found to be condensed into compact particles, known as hnRNP particles (Gall, 1956). When purified from cells, these nascent particles were found to be

composed of RNA and proteins, suggesting the co-transcriptional binding of proteins to the nascent RNA (Samarina et al., 1966). Since this initial discovery, the landscape of proteins capable of associating with mRNAs has been vastly expanded. This has been aided by developing methodologies, summarised together as RNA interaction capture (RIC) approaches. RIC relies on crosslinking the RNA-binding proteins to mRNAs, purifying the mRNAs of interest using the poly(A) tail as an anchor, dissociating the bound RBPs, and identifying them using mass-spectrometry (Baltz et al., 2012; Castello et al., 2012, 2016; Kramer et al., 2014). More recently, alternate organic phase-separation-based approaches have been developed that exploit the inherent physicochemical properties of RNA-protein complexes (Queiroz et al., 2019; Trendel et al., 2018; Urdaneta et al., 2019). These biochemical approaches have been complemented with computational methodologies that have been used to predict RNA-binding activity of proteins using the database of known RNA-binding domains, or other domain features characteristic of proteins known to associate with RNAs (Gerstberger et al., 2014; Puton et al., 2012; Si et al., 2015; Tuszynska et al., 2014). Together, ~1900 RBPs have been identified in humans using the ensemble of approaches, many of which lack known RNA-binding domains previously thought to be essential to bind RNAs, raising questions in general about our understanding of factors involved in RNA-protein interactions (Hentze et al., 2018).

Beyond their identification, the binding of RBPs has also been characterised using approaches to determine footprints of specific RBPs within target mRNAs. UV crosslinking and RNA immunoprecipitation using antibodies against the protein of interest, followed by high throughput sequencing, have helped generate transcriptome-wide maps of RNA-crosslink sites at a single-nucleotide resolution (PAR-CLIP, iCLIP, and its improved version eCLIP) (Castello et al., 2016; Hafner et al., 2010; König et al., 2010; Van Nostrand et al., 2016, 2017, 2020). Additionally, several *in vitro* approaches like SELEX and RNA Bind-n-Seq have helped identify the sequence motifs for specific RNA-binding proteins (Dominguez et al., 2018; Jolma et al., 2020; Lambert et al., 2014). These studies have helped characterise the binding of several RNA-binding proteins, identify the role of various new RBPs, and provided insights into the possible stoichiometry and composition of individual mRNPs.

1.3.2 The role of RBPs in the metabolism of Pol II transcripts, RNP compaction and organisation

Since the discovery of RNA packaging in nascent transcripts emerging from the lampbrush chromosomes, our understanding of the role of RNA-binding proteins and protein complexes in the packaging and organisation of mRNPs has improved significantly (Gall, 1956). Various biochemical and structural studies have been conducted with specific RBPs and protein complexes, both nuclear and cytoplasmic, highlighting the importance of these classes of RNA-binding proteins in the assembly of mRNPs. In eukaryotes, a few different families of proteins and protein complexes have been implicated in mRNP organisation. Among these, five families of proteins are understood to be involved in the packaging of mRNPs at different stages of its lifetime – heterogeneous nuclear ribonucleoproteins (hnRNPs), SR proteins, exon junction complexes (EJCs), the transport and export complex (TREX), and Y-box proteins.

1.3.2.1 Heterogeneous nuclear ribonucleoproteins (hnRNPs)

The hnRNP family of proteins are some of the most abundant nuclear proteins and some of the earliest known RNA-binding proteins. Their identification as RNA-binding proteins involved in packaging stems from electron microscopy images of the nascent hnRNP particles (Gall, 1956). This observation led to the hypothesis that the proteins comprising these particles, or hnRNP proteins, could be the principal regulators of mRNA biogenesis by acting as scaffolds for nascent pre-mRNAs during transcription, in a manner analogous to DNA packaging by histones (Beyer et al., 1977). To characterise their composition, hnRNP particles were isolated from nuclear lysates from HeLa cells. Purified hnRNP particles were shown to contain a 40S core made up of 6 different proteins belonging to the hnRNP A/B and C families (Beyer et al., 1977; Weighardt et al., 1996). Since this discovery, however, hnRNP particles have been discovered to contain other proteins, with more than 20 different proteins currently identified as belonging to the hnRNP family of proteins (designated in groups A1-U). All hnRNP proteins are nuclear and mostly associate with pre-mRNA, though some are known to shuttle with mRNAs to the cytoplasm. Due to the diversity in both the structure and number of proteins that comprise this family and their generally high abundance in cells, they have been associated with a diverse range of functions ranging from pre-mRNA maturation to nucleo-cytoplasmic transport and mRNA translation (Piñol-Roma et al.,

1988; Singh et al., 2015). However, the mechanistic role of hnRNPs in RNA packaging has generally been poorly characterised, with only proteins in the hnRNP A/B and C families observed to compact RNAs *in vitro.*

hnRNP C tetramers are involved in the packaging of introns

Among the different hnRNP proteins, hnRNP C is believed to have a predominant role in the packaging of pre-mRNAs, especially of intronic regions within the pre-mRNA (König et al., 2010). Two isoforms of hnRNP C - C1 and C2 have been identified, each differing by 13 amino acids. The hnRNP C proteins have just a single RNA-binding domain and are known to multimerise to form strong and specific interactions with mRNAs. Mutltimierised heterotetramers of hnRNPC have been isolated from nuclear extracts containing a fixed ratio of the two isoforms C1 and C2 (present in 3:1 ratio). Purified as part of the core 40S hnRNP complex, the tetramers were later characterised using electron microscopy and gel filtration assays to determine their structure, composition, and role in RNA compaction (Huang et al., 1994; Rech et al., 1995). These studies showed that individual hnRNP C tetramers could fold 150-240nt of RNA into compact rounded, but slightly ellipsoidal particles with an average diameter of ~9.7 nm. Longer RNAs associated with an increased number of tetramers - two tetramers bound RNAs ~456nt in length, and three of them were found to associate with an ~700nt RNA. The assembly of multiple tetramers on RNAs contributed to an alteration in the overall organisation of the RNA-tetramer complex. Though the binding of two heterotetramers led to the formation of two morphologically similar but unique complexes, the assembly of three tetramers in adjacent regions of the RNA resulted in the folding of the RNA into a compact but unique 19S triangular complex with a base of ~18 nm and the sides ~23 nm (Figure 1-1). This assembly was found to repeat over longer stretches of RNAs, with up to nine tetramers (or three 19S triangular complexes) observed on a single RNA (Huang et al., 1994; McAfee et al., 1996; Rech et al., 1995).

Evidence for the existence of such structure in cells was lacking for a long time; however, a recent study revealing the transcriptome-wide binding sites of hnRNP C has identified possible biological implications of these observations. This iCLIP dataset from HeLa cell lysates showed hnRNPC was bound to stretches of uridine-rich tracts within intronic regions, demonstrating the direct contact between RNAs and hnRNPC. Interestingly, individual hnRNP C peaks were found to be separated by regular intervals of ~165 nt and ~300nt, a spacing similar to the length of the RNAs

assembled by single hnRNP C tetramers *in vitro* (König et al., 2010). Additionally, such a regular binding of hnRNPC could also result in possible oligomerisation of hnRNPC tetramers resulting in structures similar to those observed *in vitro*. Curiously, hnRNP C binding patterns were observed in ~55% of all introns, implying a significant role of hnRNP C in forming hnRNP assemblies, either by themselves or through interaction with other proteins in the hnRNP family to form more diverse and sequence-specific structures. Furthermore, a preference of polyuridine stretches, generally found within intronic regions, could implicate hnRNP C in the packaging of long intron stretches, potentially compacting them and assisting in their splicing (König et al., 2010; Van Nostrand et al., 2020) (see 1.5.1 and 1.5.2).

hnRNP A/B family of proteins potentially assist hnRNP C in the formation of higher-order structures

The hnRNP A/B family includes four proteins - A1, A2/B1, A3, and A0, of which A1 and A2/B1 were initially identified as a part of the 40S hnRNP core complex. hnRNP A1 and A2/B1 both contain two RNA-binding domains, and similar to hnRNP C, they have been found to bind predominantly to intronic regions (Beyer et al., 1977; Van Nostrand et al., 2020). Though initially believed to be one of the essential players involved in RNA packaging in cells, with tetramers observed for A1 and A2/B1, similar to hnRNP C, their role in mRNA packaging has since been classified to be dependent on and secondary to hnRNP C in the co-transcriptional packaging of nascent pre-mRNAs. *In vitro* studies have shown that hnRNPA2/B1 can assist in the formation of higher-order 35S and 40S hnRNP intermediary complexes, with properties very similar to higher-order complexes purified from cells (Huang et al., 1994; Rech et al., 1995). Interestingly, the hnRNP C 19S complexes were a requirement for assembly of the 35S and 40S complexes, suggesting that hnRNPA1 and A2/B1 could assist in the formation of higher-order structures once hnRNPC has assembled onto the RNA. The formation of such assemblies could assist hnRNPC in the packaging of nascent introns. Additionally, hnRNPA1 has also been suggested to multimerise upon binding to different regions of mRNAs. This multimerisation could provide a mechanistic explanation to its function as a regulator of alternative splicing (see 1.5.2) (Chen and Manley, 2009; Jean-Philippe et al., 2013).

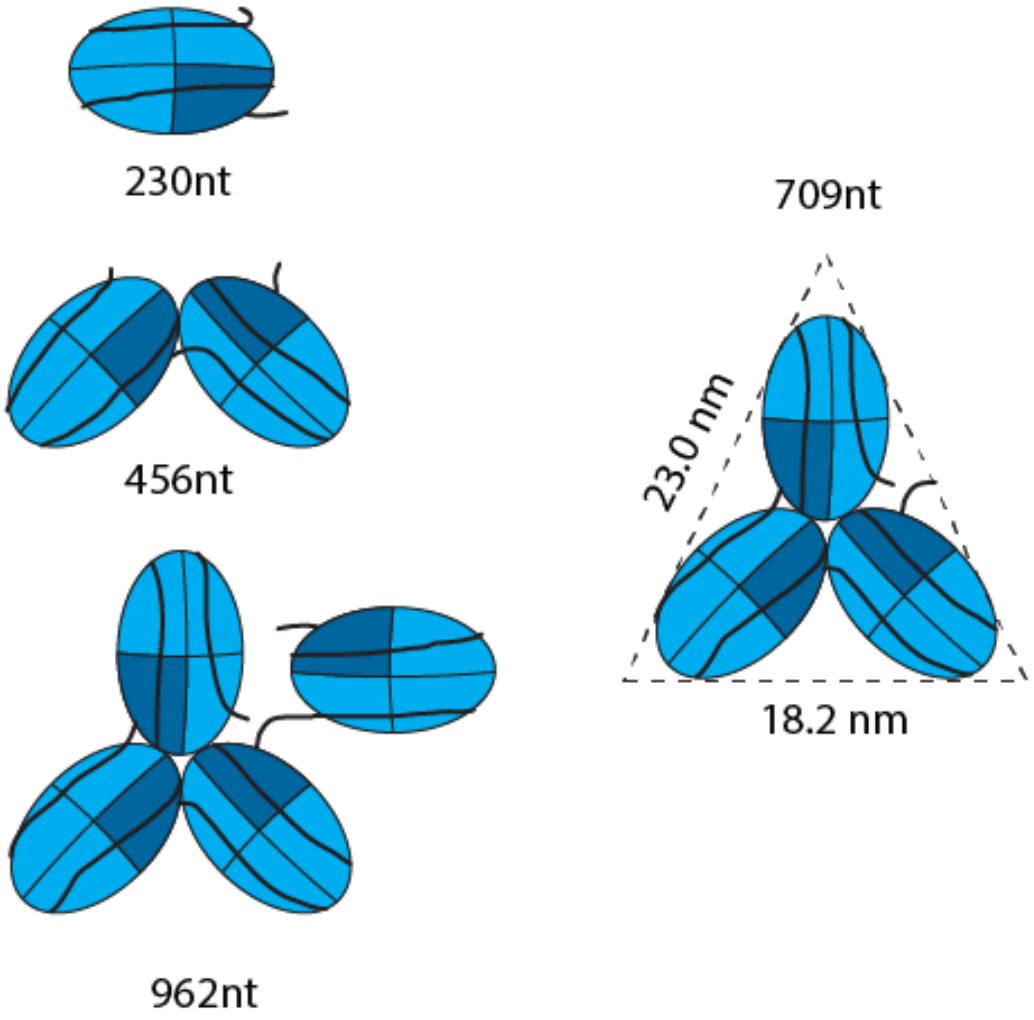


Figure 1-1: hnRNP C structures observed in vitro

This figure shows cartoons of the shapes of hnRNP C tetramers observed in vitro when assembled with RNAs of varying lengths. The dimensions of the 19S triangular complex are shown on the right. Adapted from (Huang et al., 1994)

1.3.2.2 SR proteins and the exon junction complexes (EJCs)

SR proteins are the second major family of nuclear RBPs that are often loaded co-transcriptionally to nascent pre-mRNAs and have been implicated in diverse functions from splicing to RNA

degradation (Howard and Sanford, 2015; Long and Caceres, 2009). Known for their serine-arginine dipeptide-rich domains (RS domain), their function can be dynamically regulated by phosphorylation of the serine and arginine residues. There are 12 canonical members of the SR protein family, all of which share the characteristic RS domain structure. They can have either one or two RNA-binding domains and are known to bind to specific recognition motifs (Howard and Sanford, 2015; Long and Caceres, 2009). On the other hand, EJCs are multiprotein complexes that are generally deposited 24 nucleotides upstream of the exon-exon junction in a splicing-dependent manner (Le Hir et al., 2000). However, recent studies have shown that EJC occupancy on an mRNA can stretch beyond its canonical deposition site (Metkar et al., 2018; Saulière et al., 2012; Singh et al., 2012). The EJC consists of four core proteins and several peripheral components (Boehm and Gehring, 2016). The composition of an EJC is altered at various points in time through alterations in the peripheral factors, which can influence the function of the EJC (Mabin et al., 2018; Woodward et al., 2017). They have been shown to play several different roles ranging from splicing regulation to export, translation, and decay (Blazquez et al., 2018; Boehm and Gehring, 2016; Woodward et al., 2017).

SR proteins and EJCs can multimerise to compact mRNAs

The function of SR proteins and EJCs has been expanded beyond their canonical role in regulating mRNA metabolism, and they have been hypothesised to be essential modulators of mRNP compaction in the nucleus. The first evidence for such a role came when EJCs purified from HEK293 cells using RNA-protein immunoprecipitation in tandem (RIPiT) were found to exist as megadalton-sized complexes. These massive complexes formed through multimerisation between EJC proteins and SR and SR-like proteins (Singh et al., 2012). Surprisingly, the entire family of SR proteins was highly enriched, with some SR proteins even found in super-stoichiometric amounts relative to the EJCs (Singh et al., 2012). Furthermore, the formation of these EJC-SR protein complexes was also associated with unusually long EJC footprints extending beyond their canonical binding site, 24nt upstream of the exon-exon junction. These footprints were primarily found in exonic regions, especially in regions predicted to contain exonic splicing enhancers (ESEs), typically occupied by SR proteins. This led to the model where mRNAs were proposed to be compacted and packaged through short and long-range interactions between EJCs and SR proteins bound to different regions of the mRNA (Figure 1-2).

To validate this model, Metkar et al. developed an elegant approach termed "RNA immunoprecipitation and proximity ligation in tandem" (RIPPLiT), which aimed to probe the transcriptome-wide 3D conformations of EJC bound RNAs in cells (Metkar et al., 2018). In short, EJC containing RNPs were purified in tandem using two core EJC proteins, partially digested with RNase T1 and proximity ligated to create chimeric junctions representing sequences from different parts of the mRNA. RIPPLiT revealed that the chimeric junction arose not from RNA secondary structures but rather as a result of higher-order EJC-SR protein complexes. Interestingly, the abundance of these chimeric junctions decreased with increased nucleotide span, suggesting that EJCs and SR proteins were more likely to interact when bound to neighbouring regions of the mRNA. Furthermore, computational polymer analysis of the data from multiple mRNAs indicated that pre-translated mRNPs could be compacted and organised into a linear and flexible rod-like structure (Figure 1-2). However, the effect of this compaction on the size of the mRNP remained unclear (see 1.5.1 and 1.5.3).

This EJC mediated compaction, however, can be modulated through a change in its protein composition. Affinity purification using two peripheral components of the EJC (RNPS1 and CASC3) showed an independent association with the core components, indicating the presence of two mutually exclusive EJCs in cells (Mabin et al., 2018; Singh et al., 2012). Furthermore, these two EJC complexes were found to form different levels of higher-order structures. These observations suggest that EJC mediated mRNP compaction could further be dynamically modulated through an alteration in EJC components within the cell, providing an additional layer of control for cells to regulate mRNP organisation and metabolism.

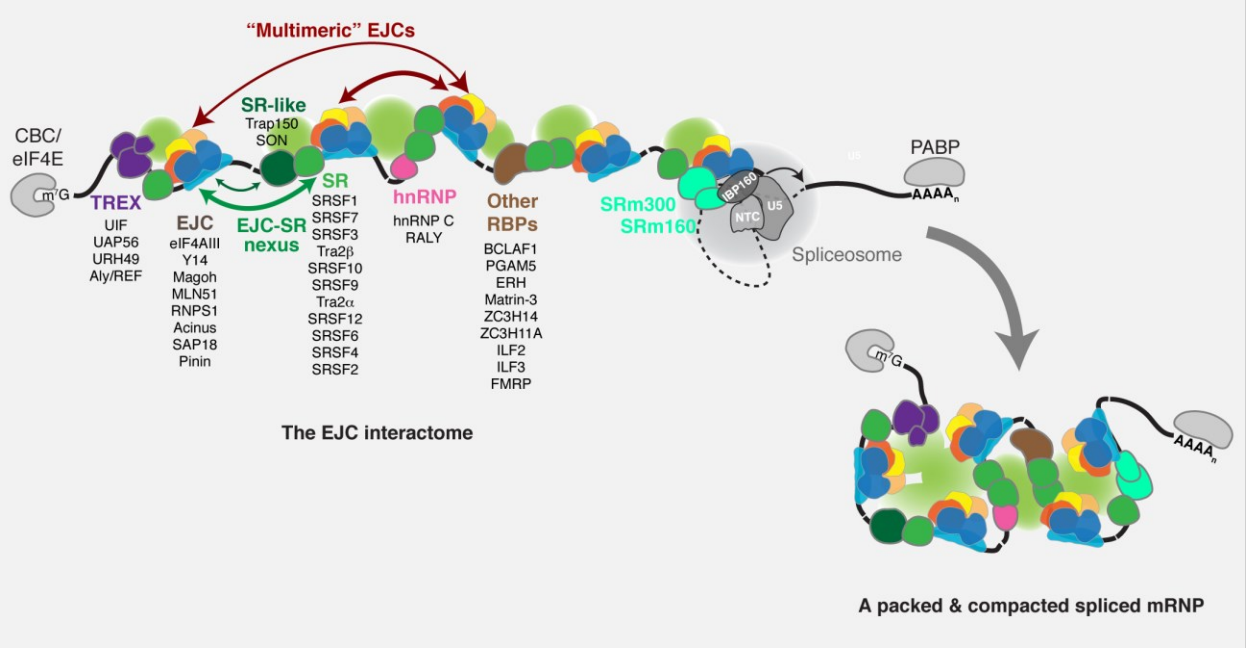


Figure 1-2: mRNP compaction through EJC-EJC and EJC-SR protein collaboration

This figure illustrates how EJCs can help in the compaction of mRNPs through interactions with each other and SR proteins. Adapted from (Singh et al., 2012)

1.3.2.3 Transport and Export (TREX) complex

TREX (or THO/TREX) is a multisubunit complex that has been shown to play a pivotal role in linking transcription, mRNA process, and export. The TREX complex consists of the core THO proteins (THOC 1-3 and 5-7) and some additional peripheral components. Similar to hnRNPs and SR proteins, the TREX components are recruited co-transcriptionally at different stages of mRNA transcription and are known to associate with mRNAs till their export to the cytoplasm (Heath et al., 2016; Köhler and Hurt, 2007; Meinel and Sträßer, 2015). While in *S. cerevisiae*, this recruitment happens through interaction with the transcriptional machinery, particularly the C-terminal domain (CTD) of RNA pol II, mammalian TREX components have been known to associate with (pre-)mRNAs through interaction with the splicing machinery and splicing dependent components like the EJCs (Wende et al., 2019). Additionally, components of the TREX can be recruited through interaction with the cap-binding complex or through proteins associating with the 3' end of mRNAs (Cheng et al., 2006; Shi et al., 2017). Beyond its role in mRNA processing and export, the TREX complex has been suggested to be involved in co-transcriptional

mRNA packaging. This became evident when depletion of TREX not only caused a decrease in mRNA transcription and export but additionally resulted in genomic instability due to increased R-loop formation (Domínguez-Sánchez et al., 2011; Heath et al., 2016; Wende et al., 2019).

The case for ALYREF as a potential mRNP compactor

Among the TREX components, ALYREF, an mRNA export adapter, is believed to have a critical role in the co-transcriptional compaction of mRNPs. The origins for this hypothesis come from the studies of its yeast homolog, Yra1, which was shown to be essential in maintaining genomic stability (Heath et al., 2016). Additionally, Yra1 was revealed to possess an RNA annealing activity, indicating its possible role in mRNP compaction (Portman et al., 1997). ALYREF, like Yra1, contains a central RNA-recognition motif (RRM) flanked by two unstructured arginine-rich regions that act as RNA-binding domains (Heath et al., 2016). This similarity in structure between ALYREF and Yra1 has led to the belief that ALYREF could perhaps be involved in RNA packaging in higher eukaryotes. However, to date, no biochemical or structural data exist to support such a function.

ALYREF can also be recruited to mRNAs through other core THO components (Heath et al., 2016; Shi et al., 2017). Though such recruitment was believed to act as a quality control measure for mRNP export, a recent cryoEM study of *in vitro* reconstituted TREX complex suggests a possible alternate function (Pühringer et al., 2020). This study found that THO could form individual assemblies on mRNAs capable of oligomerising, suggestive of TREX acting as an mRNP packaging and compacting complex. Interestingly, overlaying ALYREF onto this structure indicated its putative role in bridging these assemblies and potentially regulating TREX-mRNA interactions (Figure 1-3). If such structures are assembled in cells, it could provide a TREX-mediated mechanism for mRNP compaction.

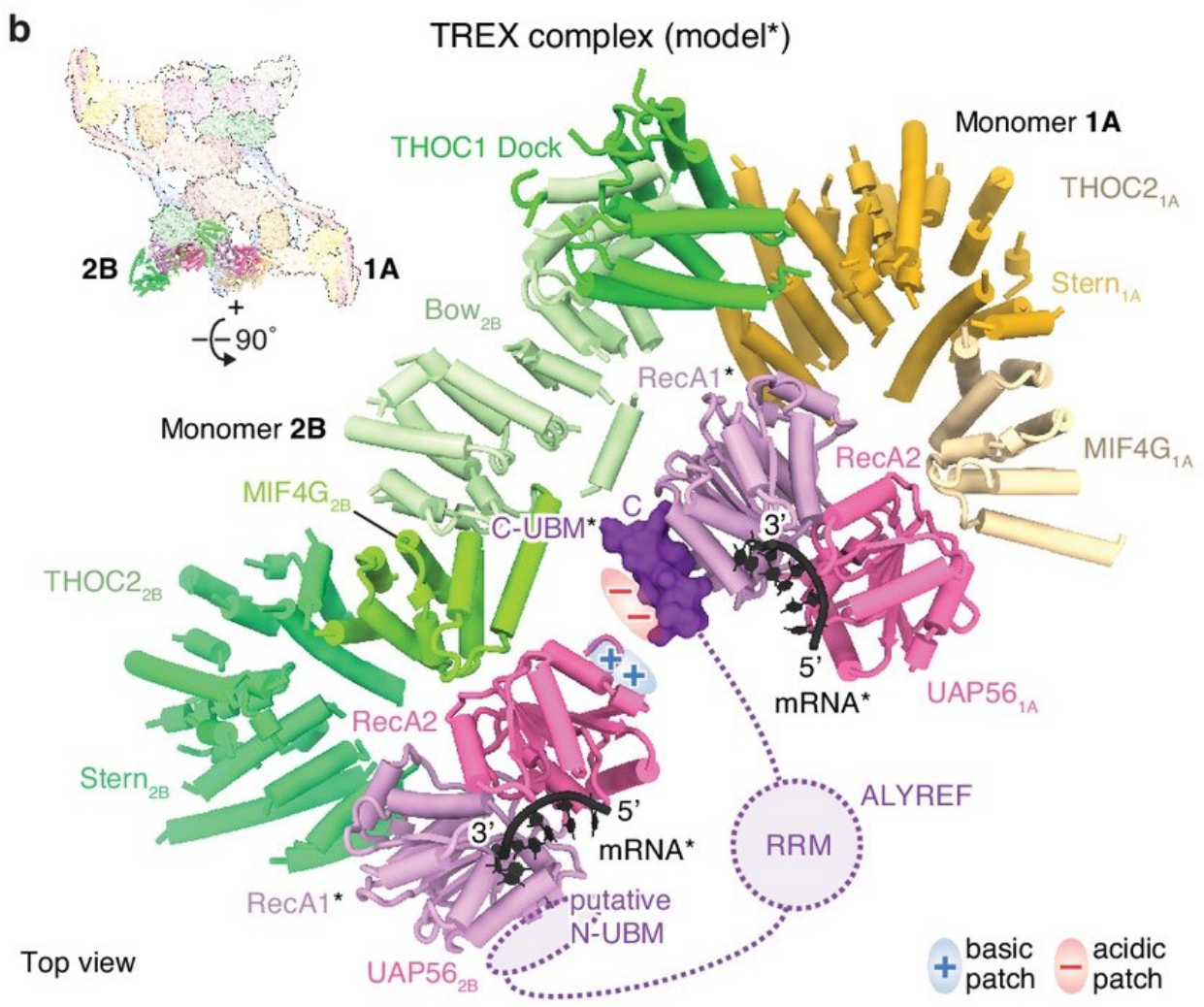


Figure 1-3: Structure of human TREX complex and putative role of ALYREF

The recent cryo-EM structure of the human TREX complex showing how ALYREF could bridge adjacent helicases of the TREX complex and regulate its interaction with mRNAs. Adapted from (Pühringer et al., 2020)

1.3.2.4 Y-box proteins

Y-box proteins are a family of cold-shock nucleic acid-binding proteins that were first discovered to bind with a DNA nucleotide sequence called the Y-box, from which their names are derived. In humans, there are three Y-box proteins (YB-1/DbpB, DbpA, and Contrin), two of which are expressed in somatic and germ cells (Matsumoto and Bay, 2005). Among the three Y-box proteins, YB-1 has been proposed to assist in the packaging and compaction of mRNPs.

YBX1/YB-1 was identified as one of the three abundant proteins associating with purified cytoplasmic mRNPs from cell lysates (Eliseeva et al., 2011; Evdokimova et al., 1995; Lyabin et al., 2014). Due to its high abundance in cells and its localisation in both the nucleus and cytoplasm, YB-1 has been associated with a variety of different functions. Its role as a DNA-binding protein has been linked with transcriptional regulation and DNA damage repair. Its RNA-binding activity is associated with pre-mRNA splicing and mRNA packaging and translation regulation (Lyabin et al., 2014). In the cytoplasm, YB-1 predominantly associates with non-translating mRNPs, and YB-1 mediated packaging of mRNPs has been suggested to play a role in translational silencing (Lyabin et al., 2014).

YB-1 can form homomultimers

YB-1 has three domains – an N-terminal domain (A/P domain), a cold-shock domain (CSD), and a large C-terminal domain (CTD) with alternating clusters of positively and negatively charged residues (Lyabin et al., 2014) (Figure 1-4). The CSD is mainly responsible for the nucleic acid-binding activity of YB-1. The N-terminal and C-terminal domains are intrinsically disordered, with the majority of the YB-1 interactions with its protein partners and to other YB-1 proteins believed to be mediated through the C-terminal domain (Lyabin et al., 2014). Similar to other RNA-binding proteins implicated in mRNA packaging and compaction, YB-1 has been observed to form multimers. YB-1 extracted from rabbit reticulocyte mRNPs was shown to form very large homomultimeric complexes with a molecular weight of 800 kDa (Evdokimova et al., 1995). When visualised using electron microscopy (EM) and atomic force microscopy (AFM), these complexes were found to have a 30-40 nm diameter and a height of 8-10 nm (Skabkin et al., 2004). Taken together, these observations suggest that multiple RNA-bound YB-1 proteins, like EJCs and SR proteins, can potentially be assembled into multimers to compact the mRNP.

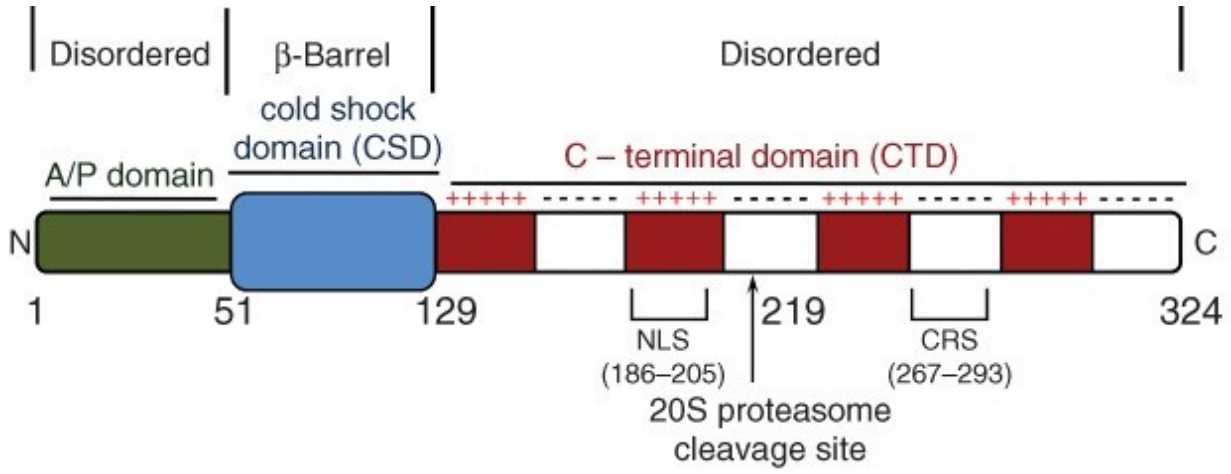


Figure 1-4: Domain organisation of YB-1

The YB-1 protein contains three domains – the disordered Ala/Pro-rich N-terminus domain, cold shock domain and a disordered C-terminal domain containing clusters of positively and negatively charged amino acid residues. Adapted from (Lyabin et al., 2014)

YB-1 homomultimers compact mRNPs *in vitro*

To assess the role of YB-1 in mRNP compaction, several *in vitro* studies have been conducted using reconstituted YB1-mRNA complexes (Kretov et al., 2019; Skabkin et al., 2004). These studies showed that YB-1 associates with mRNAs as a monomer at low YB-1/mRNA ratios, suggesting that binding of YB-1 to mRNAs can lead to the dissociation of YB-1 multimers otherwise found in solution. When visualised with atomic force microscopy, these YB-1-mRNA complexes were found to be relatively unfolded. Increasing the YB1-mRNA ratio caused YB-1 to multimerise, and this multimerisation was linked with the compaction of the RNA. Each multimer, containing 15-18 YB-1 proteins, was found to bind a region of 600-700nt. For shorter RNAs, like the α-globin mRNAs, this resulted in a single multimer bound to it, resulting in a compact conformation similar in size to a free YB-1 multimer. Longer mRNAs had several YB-1 globules on them, each linked by the RNA giving it a beads-on-string type structure. Regardless of the length of the mRNA, the formation of YB-1 multimers resulted in mRNA compaction. In cells, YB-1 is known to bind to non-translating mRNAs predominantly. Compaction of non-translating mRNAs through the multimerisation of YB-1 could prevent access to the translation initiation factors and degradation factors and keep the mRNP in a translationally repressed but stable state (see the section titled “Non-translating mRNPs and stress granule dynamics” in the discussion).

1.3.3 The role of mRNA in mRNP metabolism and RNP organisation

Apart from its principal role in helping convert the DNA message to proteins, the primary sequence of an mRNA is important for regulating other aspects of mRNA behaviour and function. As single-stranded nucleic acid molecules, mRNAs can adopt a wide variety of secondary and tertiary structures, many of which have implications in the biological function of an mRNA. An mRNA sequence can also influence its RBP composition, as many of the known RBPs are recruited to mRNAs through binding with specific sequence motifs or through binding to single or double-stranded regions of the RNA. Finally, single nucleotides within an mRNA can be modified, and RNA modifications in cells provide an additional avenue to modulate the structure and composition of the mRNP.

1.3.3.1 Short and long-range interaction within mRNAs can bridge different regions and compact mRNPs

RNAs, unlike most proteins, can adopt an ensemble of different conformations in solution. Though the degree of conformational flexibility is relatively small for some RNAs like tRNAs and ribosomes, the same is not true for mRNAs. The presence of transient structures within specific regions of the mRNA is critical for regulating various steps of mRNA metabolism (Kudla et al., 2020).

Unlike in solution, where secondary structure formation and folding of RNA occurs in microseconds, the folding of RNAs in cells is coupled to other cellular processes (Gralla and Crothers, 1973; Pörschke, 1974). For example, the rate of elongation by RNA polymerase II ragnes from 17-~800 nt/sec, and the rate of translation is ~3-60nt/sec and alterations in the rate of transcription has been linked with changes in RNA folding (Liu et al., 2016; Maiuri et al., 2011; Morisaki and Stasevich, 2018; Saldi et al., 2021; Stasevich et al., 2014). Furthermore, mRNA structure in cells is influenced by the presence of RNA-binding proteins, which may mask regions of RNAs capable of annealing with each other. Therefore, RNA structures found either computationally or *in vitro* must be complemented with methodologies used to identify RNA folding *in vivo* to determine not only the structures that RNAs are capable of forming in cells but also the biological function of such structures (Ding et al., 2014; Kwok et al., 2013; Lu et al., 2016; Mahen et al., 2010; Rouskin et al., 2014; Sharma et al., 2016; Spitale et al., 2015).

mRNAs predominantly fold *in vitro* with ends in proximity

Long-range interactions, especially between the ends of an mRNA, have essential roles in various cellular processes. Therefore, the organisation of mRNPs in a manner that facilitates these interactions is important in gene expression regulation. Several theoretical studies have predicted that long RNA sequences can fold into structures that keep the 5' and 3' of the mRNA in proximity regardless of the length of the mRNA (Clote et al., 2012; Fang, 2011; Han and Reidys, 2012; Yoffe et al., 2008, 2011). Such a folding has been verified *in vitro* for a select group of yeast, fungal and human mRNAs and lncRNAs by fluorescently labelling the ends of the RNAs and measuring their relative distances using single-molecule Förster resonance energy transfer (smFRET) (Lai et al., 2018; Leija-Martínez et al., 2014). smFRET has a sensitivity of ~1-10 nm, allowing for accurate measurement of the separation between the two ends. The measurements showed that these *in vitro* transcribed mRNAs could form multiple different structures, as observed due to the presence of several FRET states, many with their 5' and 3' ends just a few nm apart. These observations were made for several different RNAs with widely different sequences and lengths, suggesting that RNAs might have an intrinsic propensity to fold into structures with their ends in proximity.

Chemical probing suggests mRNAs are far less structured in cells than in solution

To characterise the *in vivo* folding of mRNAs, experimental methods have been developed that differentially label RNA nucleotides based on whether they are base-paired or not. Detection of these modifications using high-throughput sequencing can help build a profile of the secondary structures adopted by RNAs. The development of cell-permeable reagents capable of adding these modifications has enhanced the use of such methods in detecting RNA folding *in vivo*. A few different chemicals have been used for such probing (NAI-N_3 in case of icSHAPE and dimethyl sulphate (DMS) in DMS-Seq). These approaches have revealed that mRNAs are far less structured in cells than when they are assembled *in vitro*, suggesting that cellular processes and RBP binding can modulate mRNA secondary structures in cells (Rouskin et al., 2014; Spitale et al., 2015; Sun et al., 2019). Interestingly, some structure elements, like ones near the translational start sites and ribosomal pause sites, were highly similar both *in vitro* and *in vivo.* In addition, some structural elements were also found to be conserved between chromatin-associated, nuclear and cytoplasmic mRNAs. Though it is unclear whether these structural elements have any functional importance, their presence suggests that despite cellular processes and RNA-binding proteins altering mRNA

structure *in vivo,* some regions of RNAs can still fold in their most thermodynamically stable form (Spitale et al., 2015; Sun et al., 2019).

Cellular mRNAs can have structures with long-range interactions

While chemical probing methods can identify and map profiles of RNA structures within cells, they are limited in their capability in capturing long-range interactions. Lu et al. developed a crosslinking-based approach called psoralen analysis of RNA interactions (PARIS) to get a transcriptome-wide identification of based-paired sequences in cells. The approach employed a reversible crosslinker to selectively crosslink base-paired RNA sequences, followed by selective RNAse digestion, proximity ligation, and identification of the chimeric reads linking using high throughput sequencing to determine short and long-range interactions that exist within mRNAs (Lu et al., 2016). Though most structures identified using this approach were local, multiple independent long-range interactions were also observed, many linking the 5' UTR and 3' UTRs (Lu et al., 2016). Similar long-range interactions were also observed for dsRNAs pulled using specific RNA-binding proteins (Sugimoto et al., 2015). Interestingly, many of the long-range interactions and secondary structures were found to be evolutionarily conserved between mice and humans, suggesting possible functional importance for secondary structures and long-range interactions within mRNAs.

RNA folding can help compact mRNPs

The presence of such long and short-range interactions can, in addition to bridging different regions, help in local and global compaction of the mRNA. Studies with structured RNAs like viral RNA and ribosomal RNAs have shown that folding could potentially drive the compaction of RNAs (Borodavka et al., 2013; Russell et al., 2002; Takamoto et al., 2004). Similar observations have since been made *in vitro* for a subset of mRNAs whose size was determined using fluorescence correlation spectroscopy (FCS) and cryo-EM. When compared to a linear stretched molecule, mRNAs in solutions were found to be compacted by more than 70 fold, assuming a spacing of 0.59 nm between two nucleotides (Borodavka et al., 2016; Gopal et al., 2012). Combined with the presence of mRNA structure *in vivo* and the resemblance of several secondary structural elements between *in vitro* and *in vivo* mRNAs, this could suggest a critical role of RNA folding in either local or global mRNP compaction (Rouskin et al., 2014).

1.3.3.2 RNA Modifications can alter mRNP composition and organisation

Though the primary nucleotide sequence of an mRNA remains the same, individual nucleotides can be chemically modified in cells, and these modifications can alter the behaviour and properties of mRNAs. Several post-transcriptional RNA modifications have been identified, most of which are ubiquitous and evolutionarily conserved (Boo and Kim, 2020; Nachtergaele and He, 2018; Zhao et al., 2017). While, for tRNAs and rRNA, extensive RNA modification is required for their processing and maturation, the presence and importance of these chemical modifications within mRNA are only started to be understood. To date, out of more than 150 different RNA modifications discovered, only a handful have been identified on mRNAs. These include Adenosine to Inosine (A to I) editing, N^6-methyladenosine (m^6A), N^1-methyladenosine (m^1A), 5-Methylcytosine (m^5C), Pseudouridine (ψ), 2′-O-methylnucleosides and N^6, 2'O-methyladenosine. While several of these modifications have been implicated in modulating different steps of mRNP metabolism, recent studies have also shown that they can influence the secondary structure of mRNAs and the association of RNA-binding proteins.

mRNA modifications impact RNA duplex formation

Several RNA nucleotides modifications have been associated with a change in the secondary structure of mRNAs. N^6-methyladenosine (m^6A), one of the most abundant and well-studied RNA modifications, is known to destabilise RNA duplex formation (Kierzek and Kierzek, 2003; Roost et al., 2015). As a result, it is predominantly found to occur in sites within mRNAs that contain unpaired motifs (Spitale et al., 2015; Zhao et al., 2017). Similar duplex melting properties have also been associated with N^1-methyladenosine (m^1A) and Pseudouridine (ψ), while A to I editing has been linked with increased RNA duplex formation through stabilisation of secondary structures (Brümmer et al., 2017; Solomon et al., 2017; Zhao et al., 2017). Interestingly, transcriptome-wide mapping of these modifications has found that the occurrence and distribution of modified nucleotides can vary based on the sequence and position of the nucleotide within the RNA. While m^6A is predominantly found near the stop codon and in the 3' UTR, m^1A modifications were found to occur near the start codon, suggesting a region-specific modification could influence the local and global folding of mRNAs and subsequently their function (Li et al., 2016; Meyer et al., 2012).

mRNA modifications can alter RBP binding

The presence of RNA modifications has been observed to result in the altered binding of specific packaging RNA-binding proteins, either directly through recognition of the modified nucleotide or indirectly through a change in mRNA structure. mRNA sequences containing m^6A modification can selectively bind with hnRNP proteins. While the binding of hnRNP C is known to occur due to a change in local secondary structure, hnRNPA2/B1 is thought to recognise the m6A modification to associate with the mRNA. (Nachtergaele and He, 2018; Zhao et al., 2017). Similar observations have also been made for mRNA containing 5-Methylcytosine (m^5C) modification enriched with ALYREF in humans and YB-1 in zebrafish. The binding of ALYREF and YB-1 to these mRNAs is known to occur through specific recognition of the m^5C moiety (Yang et al., 2017, 2019). Thus, in addition to modulating mRNA structure, RNA modification could alter the mRNP organisation of specific regions of the mRNP by recruiting RNA-binding proteins, some of which are protein components involved in the packaging and compaction of mRNPs.

1.4 The synergy between mRNA structure and RBP binding

Though mRNP organisation has been studied in the context of specific mRNAs and RNA-binding proteins, several of these processes overlap in cells, and mRNAs and proteins work together and in opposition to each other to organise an mRNP. The formation of local secondary structures is known to alter RBP binding, either increasing or decreasing the binding of several proteins. At the same time, RBP binding can prevent or stabilise specific secondary structures or long-range interactions. Furthermore, there is also an interplay between numerous RNA-binding proteins, each vying for the same stretch of RNA. The dynamics between these different processes and the constant remodelling of the mRNP can together continuously change the organisation of an mRNP, influencing its processing and function.

1.4.1 mRNA structure and its effect on RBP composition

RNA-protein interactions can be driven by the affinity of RBPs for their target sequence motifs. However, the presence of a motif alone is not sufficient for *in vivo* binding of RNA-binding proteins, where only a small subset of sequence motifs are known to be bound by their RBP targets (Van Nostrand et al., 2020). A possible explanation for such an observation is the presence of local secondary structures around regions with RNA-binding motifs, which can mask the RNA motifs,

making them inaccessible for RBPs. As mentioned earlier, such an observation has been made for hnRNP C, which is known to preferentially associate with m^6A modified mRNAs due to a change in secondary structure (Zhao et al., 2017). Similarly, an *in vitro* binding assay using 12,000 mouse RNA sequences and two RBPs - MBNL1 and RBFOX2 found binding of these RBPs to be determined by the presence of local secondary structures (Taliaferro et al., 2016).

Interestingly, RNA structure can also increase the association of proteins, as observed in a study that combined RNA structure measurements (DMS-Seq) with theoretical and experimental RNA-protein interaction datasets (Sanchez de Groot et al., 2019). Such an observation could result from increased stability in regions with structures, providing the RBPs with more time to bind stably or an increase in binding of RBPs to these structured regions. While the former needs to be verified, the evidence for the latter comes from a recent high throughput RNA-SELEX analysis which identified multiple RBPs that have a preference for structured RNA motifs (Jolma et al., 2020).

1.4.2 RBPs can alter local and global mRNA folding

mRNA structure is also regulated through its association with RBPs on a local and global level. While the binding of RBPs can coat the RNA and prevent it from forming RNA-RNA interactions, RBPs can also function by annealing or dissolving RNA secondary structures. The most abundant class of proteins capable of altering the secondary structures of RNA belong to the DEAD-box (DDX) and DEAH box (DHX) families (15 and 37 different proteins in each family). DDX and DHX proteins are ATP-dependent RBPs that have shown a wide range of functions, including RNA annealing and RNA unwinding. Their role in RNA structure alteration is essential in regulating splicing, export, translation, and decay (Bourgeois et al., 2016).

RNA unwinding activity is not limited to DDX and DHX proteins, and other RBPs have been shown to contain activity capable of modulating the RNA secondary structure. YB-1 and proteins from the hnRNP family have been shown to possess RNA chaperone-like activity, capable of annealing and melting different strands of the RNA (Rajkowitsch et al., 2007; Skabkin et al., 2004). Similarly, Yra1, the yeast homolog of ALYREF, has been shown to anneal RNAs *in vitro* (Portman et al., 1997). In addition to RNA-binding proteins, RNA structures are also altered during translation, when a passing ribosome can dissolve local secondary structures allowing for the binding of RBPs and changes in mRNP composition and organisation.

1.4.3 The interplay between RBPs affects mRNP composition

An mRNA sequence could be a target for several RNA-binding proteins. While some RBPs might bind in redundant and cooperative manners, the association of others could be considered to be antagonistic. Some of the best-studied RBPs with antagonistic binding are SR proteins and hnRNP proteins. SRSF1 and SRSF2 are known to compete with hnRNPA1 in binding to splicing enhancers and regulate splicing (Zhu et al., 2001). hnRNPs can also compete against one another, with many hnRNP proteins known to bind to polypyrimidine tracts within mRNAs (Geuens et al., 2016). Similarly, hnRNP C has been shown to bind to Alu elements and prevent the binding of U2 snRNP for their subsequent exclusion from mRNAs (Zarnack et al., 2013). As many of these proteins are also involved in RNA packaging, the competition between their binding could alter the local packaging of mRNPs, providing a mechanistic explanation of their functions.

1.5 mRNP assembly and its importance in mRNA biogenesis

The assembly of an mRNP is a complex process involving the addition and removal of several individual factors on the mRNA. During transcription, protein factors loaded early onto an mRNA can influence further downstream steps coupling mRNA transcription with splicing, 3' end processing, and even the final organisation of the export-competent mRNP. Once released, mRNPs go through compositional and/or conformational changes when diffusing through the nucleoplasm, during export and upon translation once they reach the cytoplasm. The interlink between several of these processes has made it hard to identify the exact sequence of events that drive mRNP assembly and its subsequent organisation. Regardless, some aspects of mRNP composition and organisation at different stages of its life have been revealed through biochemical, microscopy, and structural studies in cells.

1.5.1 Co-transcriptional assembly of mRNPs

mRNP assembly is believed to start as soon as the nascent transcript emerges from the elongating RNA pol II. The first step of mRNP maturation is the formation of a 7-methylguanosine cap at the 5' end of the mRNA. This step coincides with the promoter-proximal pausing of RNA Pol II and the recruitment and assembly of the nuclear cap-binding complex (CBC) on the newly synthesised cap. The release of the polymerase from the pause site requires the phosphorylation of its carboxy-

terminal domain (CTD) and the negative elongation factor (NELF) (Bentley, 2014). Together, the CBC and RNA Pol II CTD play prominent roles in the steps following the pause release of the polymerase, recruiting several factors and influencing mRNA processing and packaging.

In eukaryotes, the cap-binding complex consists of two proteins CBP20 (NCBP2) and CBP80 (NCBP1). Mammalian CBP80-ALYREF interaction helps the subsequent association of ALYREF and the TREX complex to the 5' end of mRNAs (Cheng et al., 2006; Shi et al., 2017). Additionally, several members of the hnRNP family and the spliceosome have also been shown to interact with the CBC suggesting potential cap-dependent recruitment of these proteins onto the mRNA (Gamberi et al., 1997; Topisirovic et al., 2011). Similarly, Pol II is also known to play a role in recruiting numerous processing and packaging factors to the nascent mRNA. The Pol II CTD contains numerous heptad repeats capable of undergoing dynamic phosphorylations, and the phosphorylation of these residues can create a hub for several RNA biogenesis factors and mRNP components to bind the nascent transcript, including protein components of the TREX complex (in yeast), U1 and U2 snRNPs and members of SR protein family (Das et al., 2006; David et al., 2011; Harlen et al., 2016; Meinel and Sträßer, 2015; Meinel et al., 2013; Nojima et al., 2018; Sapra et al., 2009). However, to date, very little is known about when proteins are recruited to the nascent RNA and how they contribute to the packaging of the pre-mRNPs.

From the initial visualisation of lampbrush nascent transcripts, some progress has been made towards understanding the co-transcriptional packaging of pre-mRNPs (Gall, 1956). Our current view of the co-transcriptional assembly of mRNPs is primarily based on electron microscopy (EM) and electron tomography (ET) visualisations of Balbiani ring (BR) mRNPs.

1.5.1.1 Nascent Balbiani ring mRNPs are assembled into compact structures

Balbiani ring (BR) genes are sites of active mRNA synthesis on polytene chromosomes within the salivary glands of some families of flies, including Chironomids, Simuliidaes and Nematoceras. The mRNPs coding for BR genes are sufficiently electron-dense to be visualised using EM and have been an extremely valuable model system to study different aspects of mRNP metabolism. Five different BR genes exist in *Chironomus tentans*, of which four – BR1, BR2.1, BR 2.2, and BR6 are between 35-40kb in length with a similar intron-exon structure. The BR3 gene is shorter and has a different gene organisation with evenly spaced introns and exons (Figure 1-5). The

existing structures of nuclear mRNPs are derived from the assembly of the BR1, BR2.1, BR2.2, and BR6 mRNPs, all of which have similar morphology and undergo structural changes in a reproducible manner along the gene (Björk and Wieslander, 2015).

The nascent BR mRNPs are initially assembled into thin, 7-10 nm fibres as they emerge from the site of transcription. This packaging of the 7-10 nm fibre is further altered as the transcript gets longer. First, 19-20 nm thick fibres are formed, possibly due to recoiling this basic 7-10 nm fibre, and this repackaged mRNP grows until a length of ~90 nm. After that, a conformational change at the tip of the 20 nm fibre causes further compaction of the mRNP and results in the formation of a granular ribbon-like structure with a diameter of 26 nm. As more of the mRNP emerges, the assembly continues with the initial 7-10 nm fibre formation followed by 19-20 nm fibre and the 26 nm granular ribbon. Finally, at the end of the gene, the BR mRNP is folded into a ring-like shape with a diameter of 50 nm (Figure 1-6) (Björk and Wieslander, 2015; Lönnroth et al., 1992; Skoglund et al., 1983). The remarkable consistency in the folding of BR mRNP suggests a directed mechanism through which this group of mRNPs are packaged in cells, though it is unclear how this assembly takes place and which protein components are involved in this process. Detailed studies with the BR mRNP have identified several proteins and protein complexes that bind to the nascent pre-mRNA, including hnRNP-like proteins, Y-box proteins, TREX components, including ALYREF, SR proteins and EJCs (Björk and Wieslander, 2015). However, how these proteins contribute to forming the BR mRNP structure observed in cells is still unclear.

Some inferences can be made about co-transcriptional mRNP packaging based on the observations from the EM images of BR mRNPs. 1) BR mRNPs are linearly organised during transcription 2) the BR mRNP undergoes multiple morphological changes during its assembly, and each newly transcribed region goes through the same series of changes as the region before or after, suggesting that there could be a directed assembly of mRNPs, possibly with a serial order of recruitment of protein factors involved in its packaging. The initial 7-10 nm fibre could be due to CTD dependent assembly of SR proteins and TREX components and other hnRNP and Y-box proteins. 3) the first significant morphological change that results in the formation of 19-20 nm thick fibre is thought to occur around the time when intron 3 (and also possibly intron 1 and intron 2) of the BR genes is spliced. EJC deposition and the subsequent role EJCs and possibly EJC-SR protein multimers could contribute to this compaction of the nascent transcript. However, the stoichiometry of EJCs on BR mRNPs has been observed to be similar to the number of exon-exon junctions (Björk et al.,

2015), and BR mRNPs contain three introns at the 5' end and one at the 3' end separated by a long exon. Though the EJC alone could trigger the first morphological change, the likelihood that it by itself influences subsequent changes downstream of the 3rd intron is low. It is, however, possible that the formation of EJCs could trigger multimerisation with SR proteins bound within the long exon causing it to fold, resulting in its compaction. 4) Interestingly, purified BR granules were found to respond differently to RNAse treatment than vertebrate hnRNP particles. This observation suggests that either the mechanism for the organisation of mRNPs is vastly different in flies and mammals or, as evidenced from iCLIP studies, vertebrate hnRNP particles, mainly composed of hnRNP C proteins, could be involved in the packaging of intronic regions largely absent from BR mRNPs (Wurtz et al., 1990).

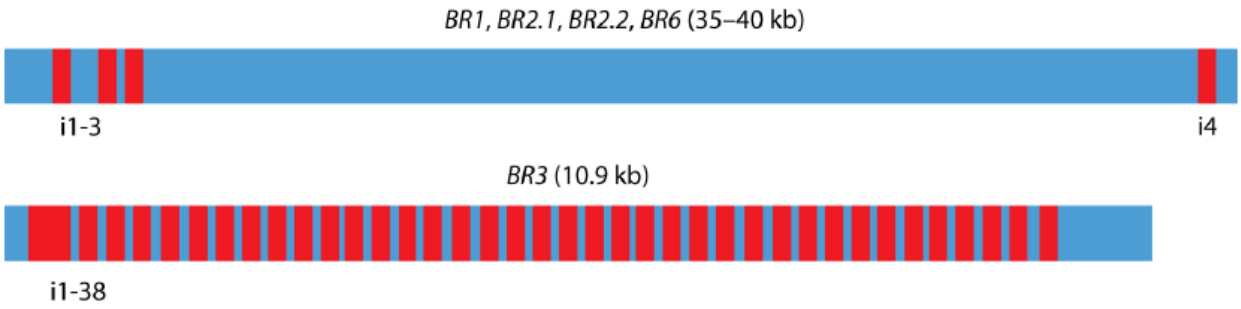


Figure 1-5: Exon-intron organisation of BR genes

The BR1, BR2.2, BR2.2 and BR6 genes share a similar exon-intron organisation – with three introns present within the first 3.2kb of the gene. Exon 5 is ~35kb in length, and intron 4 is present within ~600 from the end of the gene. BR3 gene is 10.9kb long, containing 38 introns spread throughout the gene. The exons are shown in blue and the introns in red. Adapted from (Björk and Wieslander, 2015)

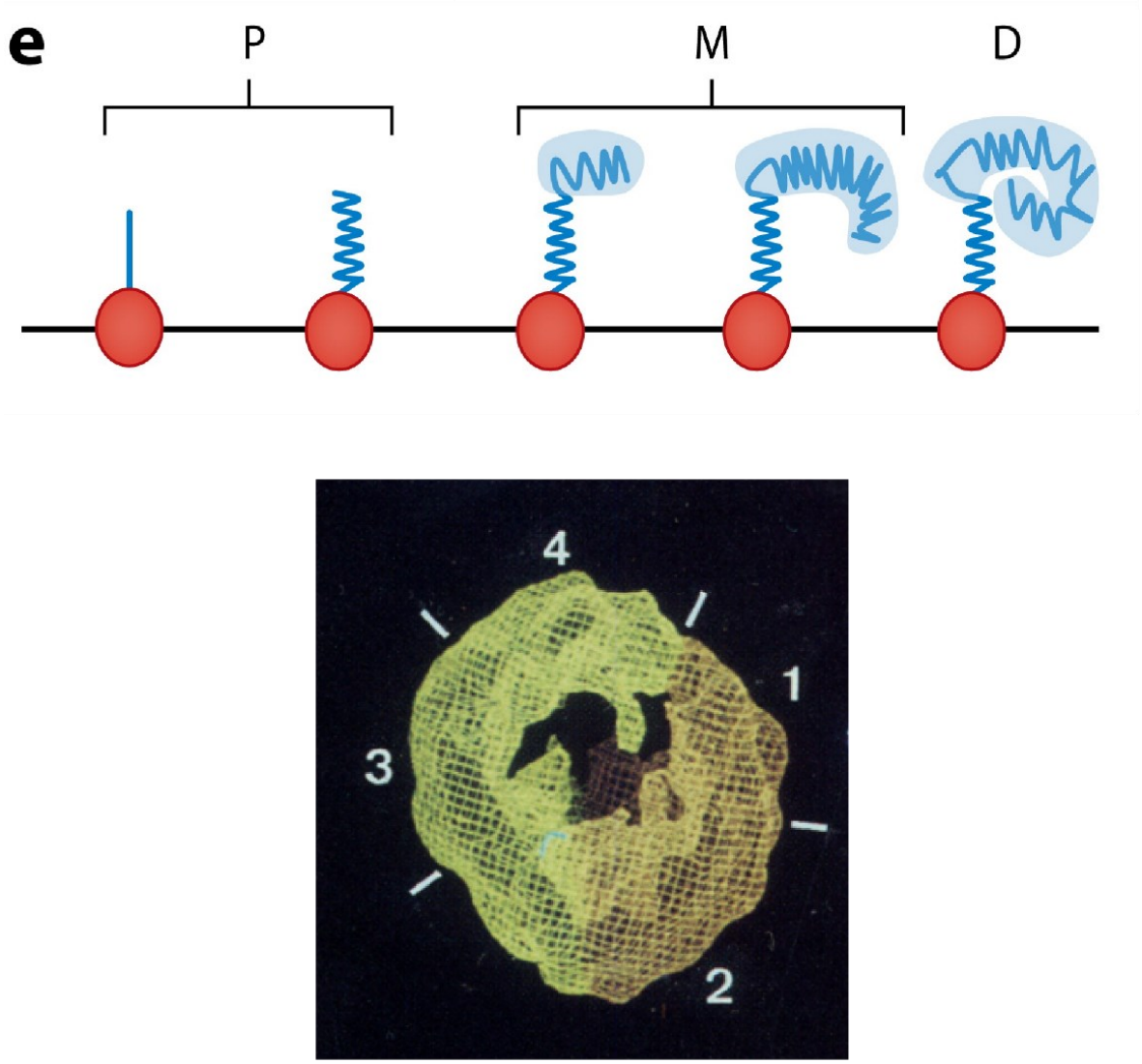


Figure 1-6: Assembly of BR mRNPs

(Top) A cartoon representing the co-transcriptional folding and assembly of the Balbiani Ring mRNPs. The P, M and D represent the proximal, medial and distal segments of the gene. Cartoon adapted from (Björk and Wieslander, 2015) *(Bottom) Three-dimensional reconstruction of BR mRNPs in the interchromatin visualized using cryo-ET. The outer numbers represent four domains. The 5' end of the transcript is present in domain 1, and the 3' end in domain 4. Adapted from* (Mehlin et al., 1995)

1.5.1.2 mRNP composition and gene structure could alter pre-mRNP packaging

The different Balbiani ring (BR) genes – BR1, BR2.1, BR2.2, and BR6 have a unique and highly similar gene structure, with long exonic regions and very few short introns. The introns are distributed at the extremities of the gene, leaving the middle region of the mRNP devoid of the assembly of spliceosome or splicing dependent factors. However, this is not true for most genes in higher eukaryotes, where the number and size of introns and exons can vary vastly. Even within *Chironomus tentans,* the BR3 gene is 10.9kb in length and has 38 introns excised at various stages during transcription. This variability in splicing has been seen to affect the packaging of mRNPs, with BR3 mRNPs known to have varied structures at the site of transcription (Wetterberg et al., 1996). The final assembled BR3 mRNPs have not been observed in cells, so it is unclear whether the presence of introns and their eventual splicing results in an altered mRNP conformation. Additionally, mRNP organisation could be modulated by differential recruitment of protein factors to genes, as is observed for SR proteins and other RBPs, implying a possible effect of either the gene structure and possibly the mRNA sequence in the final packaging of mRNPs (Björk et al., 2009; Sapra et al., 2009).

1.5.2 mRNP processing – spliceosome assembly, intron compaction and excision

The nascent pre-mRNA contains introns that need to be excised. Splicing is carried out by the megadalton ribonucleoprotein complex called the spliceosome that assembles at the distal ends of an intron. In eukaryotes, spliceosome assembly and mRNA processing are tightly coupled with transcription. As the RNA polymerase proceeds along the gene, the emerging RNA sequences act as targets for regulatory proteins and spliceosomal components (Herzel et al., 2017). However, pre-mRNAs, and introns, in particular, can contain multiple cryptic sites capable of binding these proteins and spliceosomal components, which could influence the ‘correct’ association of these factors.

Furthermore, the length of introns steadily increases with the increased complexity of an organism. For example, in humans, introns are ~10x longer than exons, with more ~10% of introns having a length greater than 11,000 nucleotides and the longest known intron being more than 1.2 million

nucleotides in length (Sakharkar et al., 2004). Not only does the increased length provide more significant avenues for cryptic binding of the spliceosome, but it adds additional complexity to intron assembly, organisation and splicing. Therefore, introns within pre-mRNAs have to be packaged in ways that suppress cryptic RNA elements and, in addition, facilitate the interaction between the intron ends, required for spliceosome assembly and intron excision.

1.5.2.1 Assembly of the spliceosome and the splicing reaction

The principal components of the spliceosome include five different U-rich small nuclear ribonucleoprotein (U snRNPs), named after their small nuclear RNA (snRNA) components – U1, U2, U4, U5 and U6 (Shi, 2017). Spliceosome assembly occurs through the dynamic base pairing among snRNAs and between the snRNAs and the pre-mRNA. Our view of spliceosome assembly and the architecture that the spliceosome takes at different steps in the splicing reaction have been deduced from the decades of genetic studies and the milieu of *in vitro* structures of the spliceosome at different stages of its assembly (Herzel et al., 2017; Lee and Rio, 2015).

According to our current understanding, the spliceosome assembly begins by recognising the 5' and 3' splice sites (ss) and the branch-point sequence (BPS) at the ends of an intron. This results in the formation of the E complex through the binding of the U1 snRNP to the 5' ss, splicing factor 1 (SF1) to the branch point and the U2AF to the 3' splice site. After that, SF1 is displaced by the U2 snRNP, forming the A complex and helping bridge the two ends of the intron. The A complex can interact with the tri snRNP complex (U4/U6 and U5) and assemble the pre-catalytic spliceosome (B complex). The B complex represents the fully assembled spliceosome, capable of carrying out the different catalytic steps of splicing through the addition and removal of protein factors and the rearrangement of the snRNPs and pre-mRNAs within the complex. Once spliced, the intron is released in the form of an intron-lariat, which is believed to be debranched by DBR1 and degraded by the nuclear exonucleases (Figure 1-7) (Shi, 2017).

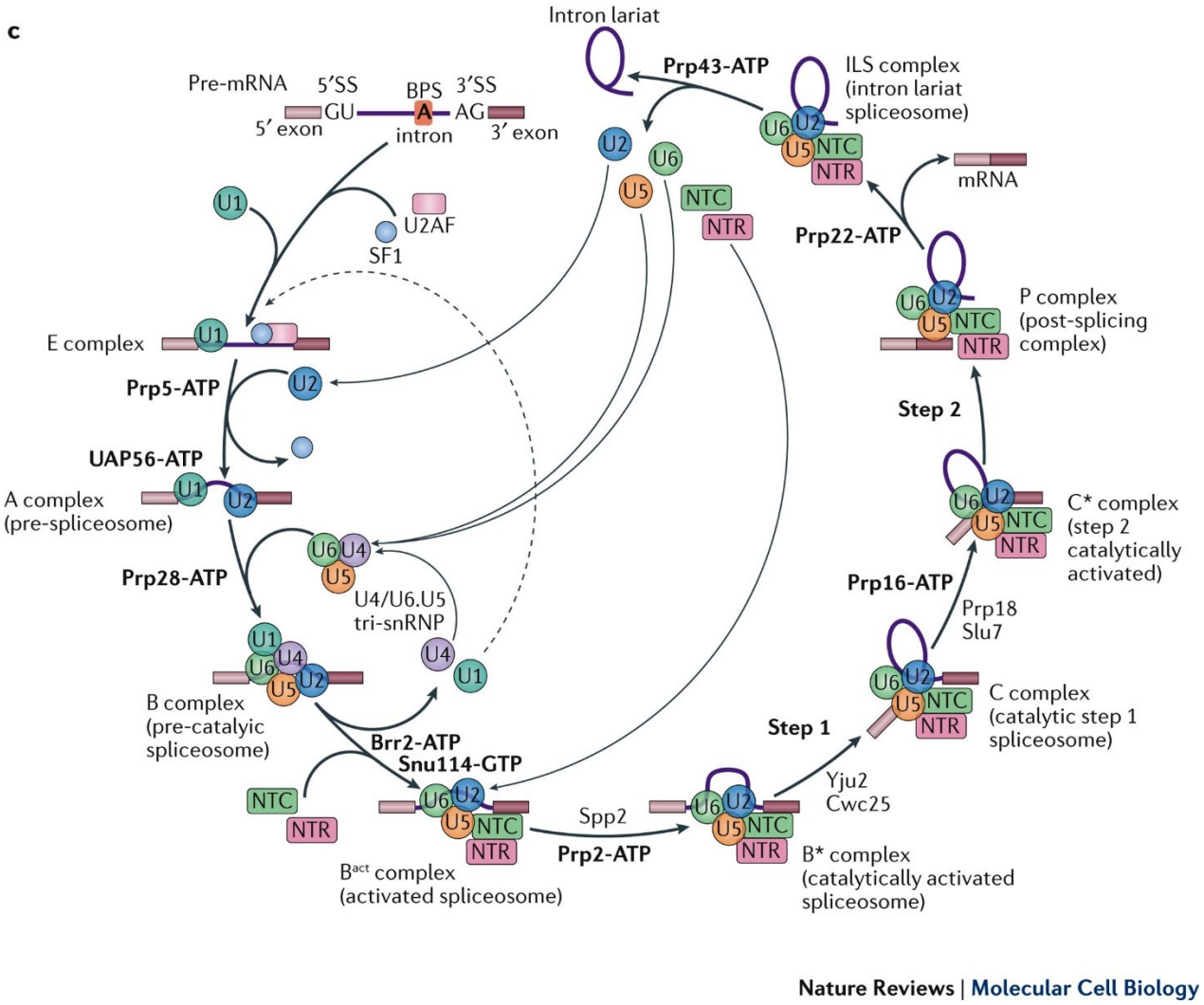


Figure 1-7: pre-mRNA splicing cycling

This figure illustrates the assembly and activation of the yeast spliceosome and the complete splicing cycle. Splicing begins through the binding of U1, SF1 and U2AF to form the E complex. Assembly of U2 and displacement of SF1 results in the formation of the A complex, which then requires the tri-snRNP complex (U4/U6 and U5) to form the pre-catalytic spliceosome (pre-B complex). Additional components are assembled on the spliceosome to complete the catalytic steps and the release of the intron lariat. The assembly of the snRNPs in humans follows the same steps as in budding yeast. Adapted from (Shi, 2017)

1.5.2.2 Role of pre-mRNP packaging in splice site selection

Though the primary pre-mRNA sequence dictates where the spliceosome assembles, the choice of splice sites is determined by several additional factors. These include proteins that bind to cis-elements near the splice sites, RNA structure, gene architecture, nucleosome positioning within the gene, histone modifications and the transcription rate of RNA polymerase (De Conti et al., 2013). How these different factors and processes add up to determine splice site selection is still a mystery; nevertheless, studies analysing the splicing of individual genes and transcripts have helped in shedding some light on it. I will be focusing on how some of these processes influence splice site selection by altering the local and possibly global packaging of the nascent transcript.

Nascent RNA folding affects splicing

The nascent RNA transcript has the propensity to fold into secondary structures in a manner that is dictated by the sequence, nucleotide modifications, transcription elongation rates and association of RNA-binding proteins. As mentioned earlier, numerous structural elements have been identified for mRNAs in cells, and the function of several structures in regulating splice site selection has been characterised. Secondary structures alter splicing by either concealing or exposing 5' ss, BPS, 3' ss and cis-acting elements, required to recruit and assemble the spliceosome (De Conti et al., 2013; Herzel et al., 2017; Lai et al., 2013; Lewis et al., 2017; Warf and Berglund, 2010). Additionally, secondary structures can also facilitate splicing or affect alternative splicing by compacting the mRNA and bringing together regions of RNAs, like the BPS and 3' ss, otherwise separated in space (Buratti and Baralle, 2004; Warf and Berglund, 2010). Concurrent with the importance of RNA folding, mutations within regions of mRNAs that form secondary structures have been associated with altered splicing of mRNA. For example, a mutation within the *tau* gene affects exon 10 inclusion through alteration in the secondary structure (Xu et al., 2021). Many such mutations near the 5' and 3' ss resulting in altered splicing have been linked to disease phenotypes (Buratti and Baralle, 2004).

RBPs can alter local packaging to determine splice site selection

Several RBPs are known to modulate the splicing of mRNPs through alterations in the local packaging. hnRNPA1 is one such protein that has been heavily implicated with altered splice site selection and alternative splicing (Jean-Philippe et al., 2013). Several models have been proposed

to explain its role as a splicing modulator, including “exon looping” and cooperative binding of multiple hnRNP A1 proteins (Figure 1-8) (Cartegni et al., 2002; Jean-Philippe et al., 2013). The cooperative binding model proposes that the first hnRNPA1 protein binding to the RNA promotes the accumulation of multiple hnRNPA1 proteins on the exon. This aggregation of hnRNPA1 is thought to prevent the recruitment of SR protein, hence preventing exon inclusion (Jean-Philippe et al., 2013; Okunola and Krainer, 2009). Though the mechanism behind this accumulation is unknown, it is possible that the assembly of multiple hnRNPA1 proteins on neighbouring regions of the RNA could be mediated by its oligomerisation resulting in the compaction of the exon, limiting the access to SR proteins. Similarly, oligomerisation provides an explanation for the “looping out” model. hnRNPA1 proteins bound to either side of the exons have been proposed to interact and loop out the exon, promoting its skipping, though evidence for such a loop has not been observed *in vivo*.

In addition to exon skipping, hnRNPA1 and other hnRNP proteins could also be involved in hiding cryptic sites within the body of introns. CLIP-Seq data for hnRNPA1 and hnRNP C from HeLa cells suggests a prominent binding throughout the body of the intron (Bruun et al., 2016; König et al., 2010; Van Nostrand et al., 2020). In agreement with this hypothesis, the knockdown of hnRNP C has been associated with an increase in alternative splicing of introns (König et al., 2010). Similar roles could also be played by other hnRNP and SR proteins, binding of which could potentially hide or expose regions of RNAs that the spliceosome could otherwise recognise.

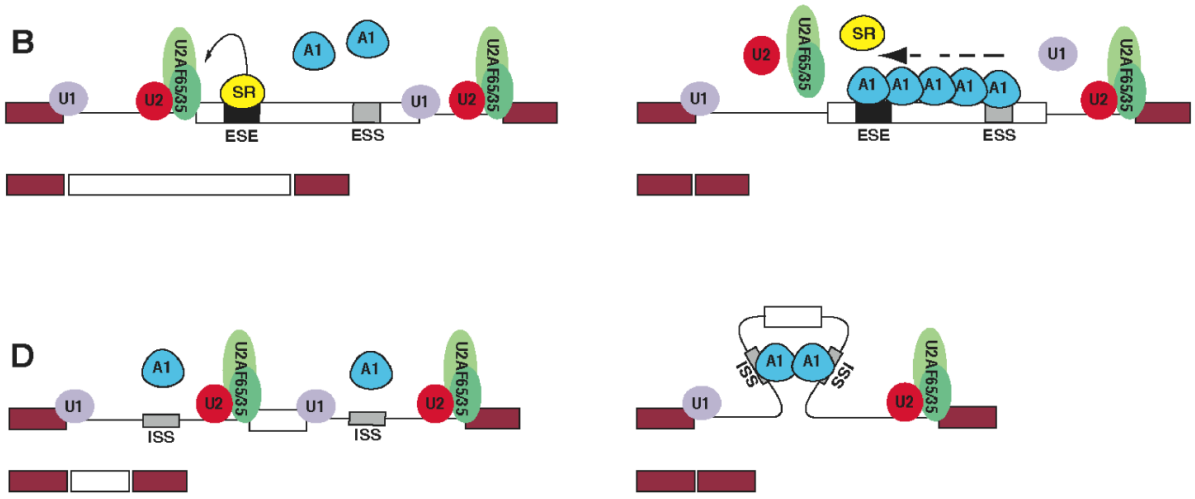


Figure 1-8: Models for hnRNPA1 mediated exon skipping

Top: Cooperative binding model – when binding of one hnRNPA1 to an exonic silencing silencer promotes the cooperative binding of other hnRNPA1 proteins along the length of the transcript. Bottom: Exon-looping model - when binding of hnRNPA1 on the introns, either side of the exon, promotes the looping-out of the exon and its exclusion. Figure adapted from (Jean-Philippe et al., 2013).

1.5.2.3 Intron packaging and compaction

In addition to its role in preventing R-loop formation, packaging of introns can help in intron compaction and its eventual organisation. The organisation of introns is essential to ensure the ends of the introns can be bridged to facilitate the assembly of the spliceosome. Despite this critical role in mRNA splicing, very little is understood about how introns are packaged and organised in cells. Our view of introns organisation comes from the Miller spread of chromosomes extracted from *Drosophila* embryos (Beyer and Osheim, 1988). These electron microscopy images showed nascent transcripts containing multiple unspliced introns organised in a looped conformation, with the ends bridged by the spliceosome. Though their importance in validating co-transcriptional splicing cannot be questioned, little can be deciphered from these images about intron compaction and organisation in cells. In cells, introns are possibly packaged, likely through a combination of secondary structures and RNA-binding proteins. Biochemical assays and *in vitro* structural data have highlighted the possible roles of hnRNP C and other hnRNP proteins in this process and the

abundance of RNA folding in intronic regions (Bruun et al., 2016; Huang et al., 1994; König et al., 2010; Sun et al., 2019; Van Nostrand et al., 2020). However, to date, intron conformations have not been visualised *in vivo*, making it unclear as to how packaging could influence the overall size and organisation of an intron and affect its splicing.

1.5.2.4 Bridging the 5' and 3' ends of introns

To be spliced, introns must bring their 5' and 3' ends together and subsequently assemble the spliceosome. Though packaging could help compact introns and keep different regions in proximity, this alone might be insufficient to allow for stochastic interactions between the ends, sometimes separated by thousands of nucleotides. In mammalian cells, introns can be spliced within minutes after the transcription of the 3' end, in a manner that is independent of the length of the intron, suggesting that a diffusion-based model that allows the ends to communicate, even in compacted and packaged introns, might not be the sole mechanism for splicing (Coulon et al., 2014; Hollander et al., 2016; Martin et al., 2013; Singh and Padgett, 2009). Therefore, alternative mechanisms are thought to exist in cells that ensure the proximity of introns' ends as soon as the polymerase has finished transcribing the 3' end of an intron (Hollander et al., 2016; Sibley et al., 2016).

Tethering of U1 snRNP to RNA Pol II could facilitate spliceosome assembly for long introns

One prominent model for splicing long introns suggests that U1 snRNP bound to the 5' splice site of an intron could be tethered to the elongating polymerase as it travels along the gene (Hollander et al., 2016). If the 5' ss is tethered to the polymerase, it will always be in proximity to the newest transcribed intronic region. Consequently, when the polymerase reaches the 3' end of the intron and U2 snRNP is recruited to the 3' splice site, this proximity will help assemble the A complex. Several studies initially supported this model showing a direct interaction between RNA Pol II and U1 snRNP (David et al., 2011; Harlen et al., 2016; Nojima et al., 2018; Robert et al., 2002). However, such interactions could also represent the role of the Pol II CTD in the recruitment of the spliceosome factors, as evidenced by the presence of U1 in intron-containing and intronless genes, and structural or functional insight into these interactions was needed to provide evidence for the U1-Pol II tethering model (Brody et al., 2011; Spiluttini et al., 2010).

A recently obtained cryo-EM structure of the RNA-Pol II – U1snRNP complex provides direct evidence for this model. The Pol II -U1 snRNP complex was assembled *in vitro* using Pol II extracted from the domestic pig, *Sus scrofa domesticus*, U1 snRNP purified from HeLa cells and a DNA-RNA scaffold replicating the transcription bubble and the nascent transcript (Zhang et al., 2021). This assembly was imaged with single-particle cryo-electron microscopy to obtain a structure with an overall resolution of 3.8 Å. Analysis of this structure revealed a direct interaction between the positively charged residues in U1-70K, a protein within the U1 snRNP complex, and the negatively charged pocket formed by RNA Pol II subunits RBP2 and RBP12. Interestingly, the residues in U1-70K and RBP2 that establish the contact between the proteins are highly conserved among the different metazoa, suggesting that U1 snRNP- Pol II tethering could be a universal model for splicing across higher eukaryotes. However, as tethering has not been observed *in vivo*, it is not yet known whether these structures can be formed by transcribing polymerase. Furthermore, the mechanisms that lead to the formation of the U1 snRNP-Pol II complex and its subsequent role in the spliceosome assembly still need to be determined.

1.5.3 mRNP organisation in the nucleoplasm and during export

The co-transcriptional packaging and compaction of mRNPs could be important after its release from the site of transcription, helping mRNPs diffuse through the nucleoplasm to reach the nuclear pore complex (NPC) for their eventual export to the cytoplasm. Studies determining mRNP architecture in the nucleus have observed that mRNPs are organised into compact particles in the nucleus. For example, the BR mRNP, once released from the site of transcription, adopts a compact croissant-shaped conformation in the nucleoplasm with a diameter of ~50 nm and a thickness of ~15 nm (Skoglund and Daneholt, 1986; Skoglund et al., 1983). Similar compact conformations have also been observed for mRNPs purified from S. cerevisiae (Batisse et al., 2009). However, very little is otherwise known about the structural organisation of nuclear mRNPs. Single-molecule imaging MS2 tagged mRNAs ranging from 1,700 to 17,000 nt suggested that mRNA in the nucleus have a diffusion rate reminiscent of compact particles. Interestingly, though the diffusion coefficient of longer mRNAs decreased, the magnitude of this decrease did not scale with the increase in length, suggesting that longer mRNA could undergo multiple levels of compaction similar to what is observed for BR mRNPs (Mor et al., 2010).

1.5.3.1 Nuclear mRNPs are possibly linearly organised

Unless mRNPs undergo massive conformation change in the nucleoplasm, the co-transcriptional assembly of mRNP would determine the final assembly of the mRNP particle. Pre-translational mRNPs purified using the exon junction complex (EJC) were shown to fold into linear rod-like structures. Similarly, Balbiani Ring mRNPs are packaged linearly into compact particles, and though their composition has been known to change within the nucleoplasm, there is no evidence to suggest that this change in composition results in an alteration in its conformation. A linear rod-like organisation has also been observed for mRNPs in simple eukaryotes like *S. cerevisiae.* Nuclear mRNPs purified using the nuclear poly(A) binding protein Nab2 were shown to have compact rod-like conformations with individual mRNPs varying in length (15-30 nm) but having a fixed width (~5 nm) (Batisse et al., 2009). *S. cerevisiae* does not contain either EJCs and lack the diversity in SR and SR-like proteins, so the packaging and the resulting linear organisation of mRNPs are possibly mediated by TREX, hnRNP-like or SR-like proteins. A rod-shaped particle presents multiple advantages to mRNPs, which could be a potential reason for the evolutionary conservation of this architecture (advantages discussed in the discussion section "Nuclear mRNP organisation").

1.5.3.2 mRNP remodelling during export

Diffusion takes mRNPs to the nuclear periphery, where they interact with the nuclear pore complex to get exported. The export of mRNPs is initiated first through its docking at the nuclear basket, followed by its transit through the NPC central channel and eventual release from the cytoplasmic side of the pore (Oeffinger and Zenklusen, 2012). To date, BR mRNPs remain the only known mRNAs whose conformations have been visualised when transiting through the pore. EM images showed that these mRNPs are unfolded at the distal ring of the nuclear basket before entering the basket with the 5’ of the mRNA first (Figure 1-9) (Mehlin et al., 1992). However, whether such remodelling is required for all mRNPs is unknown, as most mRNPs are at least one order magnitude smaller than BR mRNPs. In addition, a linear rod-like organisation of mRNP, as observed for yeast mRNPs and purified EJC containing mRNPs, could negate the need for such remodelling at the nuclear basket and facilitate faster export of mRNPs. After their transit through

the central channel, mRNPs undergo extensive remodelling at the cytoplasmic side of the NPC. DDX19B (or Dbp5 in yeast) is associated with this ATP-dependent remodelling and the subsequent release of several nuclear RBPs from the mRNPs. However, whether the release of these RBPs alters the conformation of the mRNP, preparing it for translation in the cytoplasm is not yet determined (Björk and Wieslander, 2017; Bourgeois et al., 2016; Oeffinger and Zenklusen, 2012).

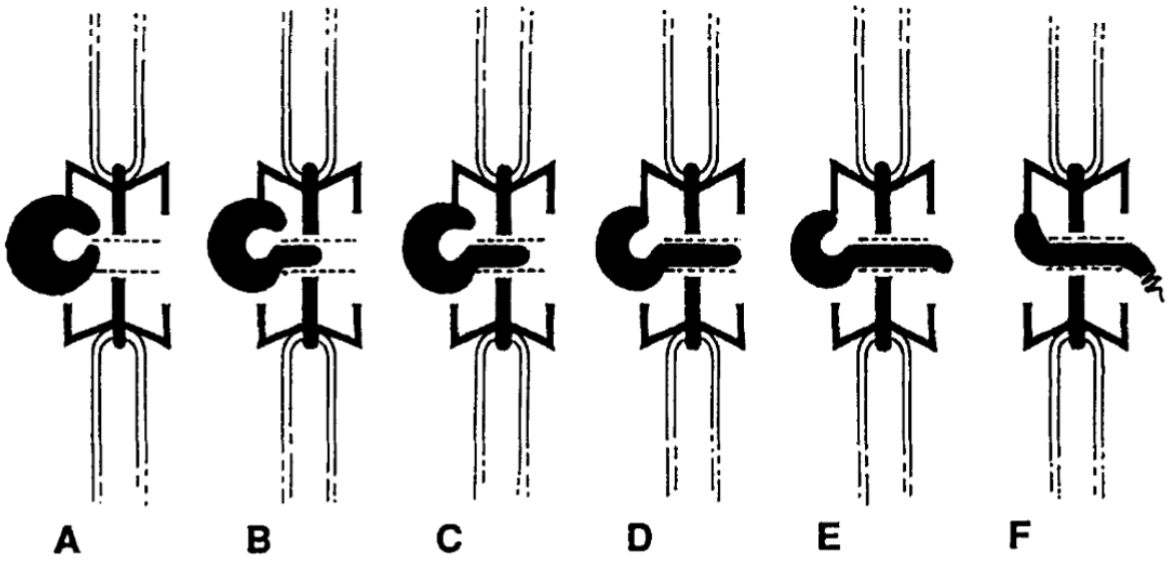


Figure 1-9: Schematic presentation of BR mRNP export

Cartoon representation of the organisation of BR mRNP during export through the nuclear pore complex as observed using electron microscopy. The BR mRNP is unfolded at the nuclear basket before its transit as a linear rod. Figure adapted from (Mehlin et al., 1995).

1.5.4 mRNP organisation in the cytoplasm and its role in regulating translation and degradation

Once mRNPs are exported to the cytoplasm, they need to get translated into polypeptides and eventually degraded. EM studies of BR mRNP have shown that mRNPs can initiate translation even before the entire mRNP has transited through the pore. However, translation of most mRNPs is believed to be coordinated by regulatory factors binding to either end of the mRNA, thus requiring the export of the entire mRNP (Mehlin et al., 1992). During their lifetime, mRNAs can undergo multiple rounds of translation before they are degraded. The first round of translation,

observed to occur within minutes after export, influences the properties of an mRNP by drastically altering its composition and organisation (Halstead et al., 2015). After that, the regulation of translation and mRNA decay is mediated by a host of protein factors that are tightly linked to one another, to the local structure of the mRNP and the global organisation, involving interaction between elements located at different regions of the mRNA.

1.5.4.1 Local mRNA structure and its role in regulating translation and stability

Several mRNA structural features have been linked with changes in translational rates. In many eukaryotes, the length and secondary structure of the 5' UTRs can negatively correlate with the translation of several mRNAs (Leppek et al., 2017; Yan et al., 2016). Secondary structures in the 5' UTR can inhibit the 43S preinitiation complex from associating with the mRNA. Consistent with this observation, the 5' UTR and the region around the start codon are generally devoid of secondary structures, with the amount of secondary structure inversely correlated with the translation efficiency of mRNAs (Ding et al., 2014; Mauger et al., 2019; Spitale et al., 2015). The presence of secondary structures in the 5' UTR is not always decisive in translational regulation, with RNA helicases like the DEAD-box helicase eIF4A thought to play a critical role in unwinding them (Leppek et al., 2017).

Secondary structures within other regions of the mRNA can also be important in regulating translation and stability. An increased amount of secondary structures in the coding sequence was linked with higher translational output through changes in the half-life of mRNA, though the mechanism for this regulation is still unclear (Mauger et al., 2019). Similarly, structures in the 3'UTR have been linked with mRNA degradation, with some structures shown to increase mRNA stability while others increase the decay (Lewis et al., 2017).

1.5.4.2 5'-3' communication tunes translation and mRNA stability

In addition to modulating the local packaging, mRNP organisation is critical in enabling interaction between different regions of mRNAs, especially the 5' and 3' ends of cytoplasmic mRNAs. The 5' cap and the 3' poly(A) tail are present in most cytoplasmic mRNA, and their role in the cytoplasm extends beyond acting as blockades to protect the mRNA from attack by

exonucleases. They play a fundamental role in gene expression by dynamically associating with protein factors involved in translation and mRNA decay (Figure 1-10). Interestingly, the function of the cap and the poly(A) tail is highly intertwined, with both structures known to stimulate translation of mRNAs and mRNAs lacking either of them known to have very short lives in cells (Rissland, 2017).

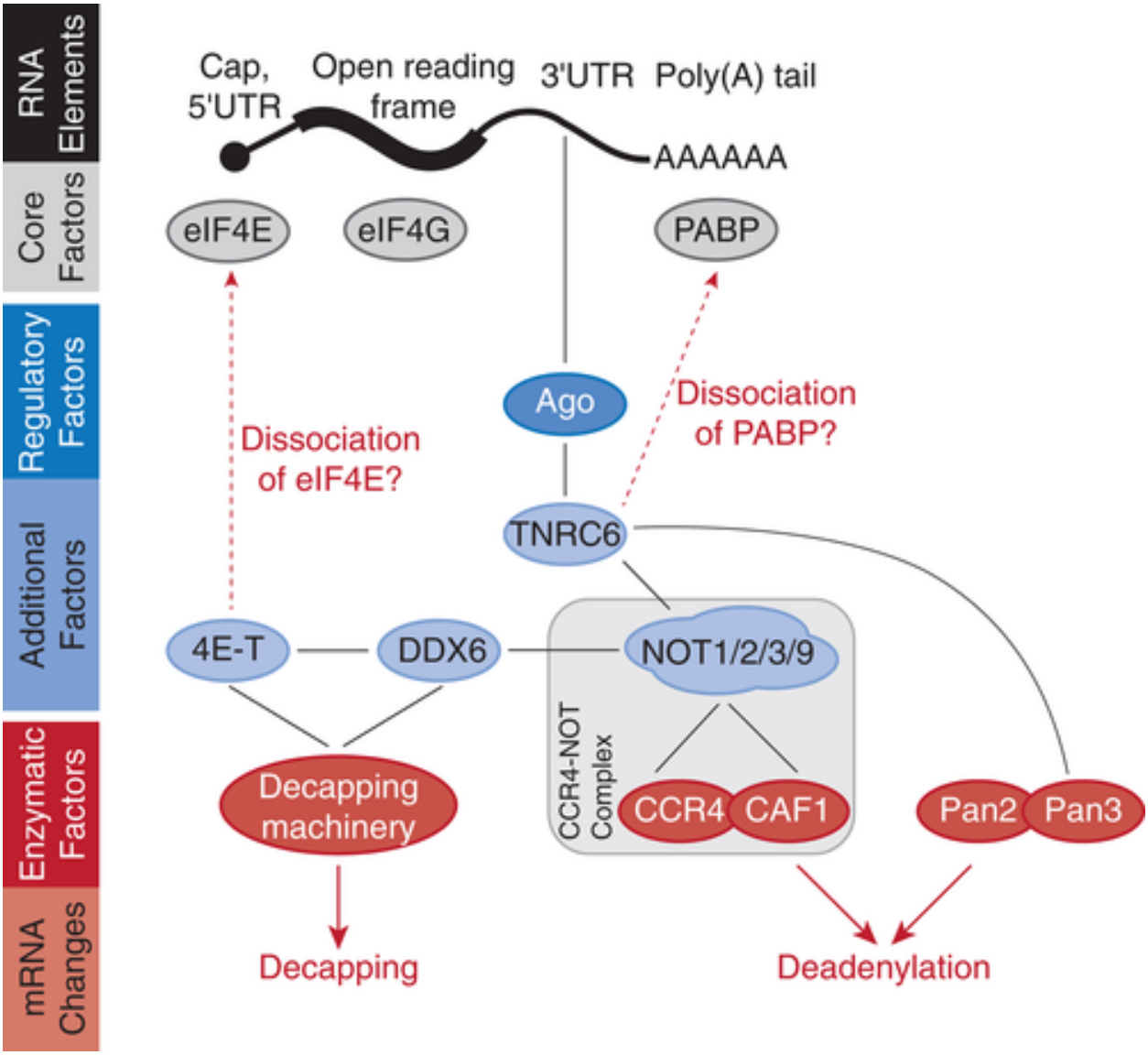


Figure 1-10: Factors associating with 5' and 3' ends of mRNAs in the cytoplasm

The figure illustrates the core and regulatory factors that associate with the 5' and 3' end of mRNAs during translation, translational silencing and mRNA decay. Figure adapted from (Rissland, 2017)

The closed-loop model and mRNA translation

Translation of mRNAs is stimulated by proteins associating directly or indirectly to the 5' end of the mRNA. These include components of the eIF4F complex - the cap-binding protein eIF4E, the scaffold protein eIF4G and the DEAD-box helicase eIF4A, which are responsible for the recruitment of the 43S pre-initiation complex to the mRNA (Hinnebusch, 2014). Interestingly, a similar stimulatory effect on translation is also associated with the cytoplasmic poly(A) binding proteins (PABPC1). Studies characterising this role of PABPC1 have identified a possible explanation involving the interaction between PABC1 with eIF4G1 (Imataka et al., 1998; Tarun and Sachs, 1996; Tarun et al., 1997; Wakiyama et al., 2000). As PABPC1 is predominantly bound to the 3' end of mRNAs, this resulted in suggestions that mRNAs could form a closed-loop structure through eIF4E-eIF4G-PABPC1 interactions (also known as the "closed-loop" model). Similar structures could also be formed through other 3' end binding proteins, like the stem-loop binding protein (SLBP) associated with the 3' ends of histone mRNAs, which have also been shown to interact with eIF4G (Ling et al., 2002). Several studies characterising the molecular interactions between eIF4E, eIF4G, and PABPC1 further strengthened the argument for this model. PABPC1-poly(A) interaction is known to strengthen the binding of PABPC1 to eIF4G. Similarly, the binding of PABPC1 to eIF4F and eIF4G to eIF4E increased the affinity of eIF4E for the cap (Gross et al., 2003; Rissland, 2017; Safaee et al., 2012; Wei et al., 1998). Interestingly, interactions between eIF4G and PABP are evolutionarily conserved from yeast to humans, indicating potential importance for such interaction in cellular function.

Despite the biological characterisation of PABPC1-eIF4G1 interaction, single mRNAs showing a closed-loop form in cells mediated through eIF4E-eIF4G-PABPC1 have not yet been observed. *In vitro* preparation of capped and polyadenylated mRNAs in the presence of eIF4E, eIF4G, and PABPC1 presents the only visual validation for these interactions on mRNAs. When visualized through AFM, these mRNAs formed a closed-loop structure only in the presence of all closed-loop components (Wells et al., 1998). Though this observation provides a theoretical possibility for such structures to exist in cells, whether the cellular environment allows for such structures is unknown. Furthermore, if such structures exist, the fraction of mRNAs in cells with a circular conformation is still not determined.

The closed-loop state - translation silencing and degradation

Protein components dynamically associated with the cap and poly(A) tail can also modulate translation repression and mRNA stability. The deadenylation and decapping complexes which compete with PABPC1 and eIF4E, respectively, have functions that are intertwined with one other. For example, depletion of Pab1 in *S. cerevisiae* can result in increased decapping of mRNAs with longer poly(A) tails, possibly resulting from increased recruitment of the deadenylation complex to unoccupied poly(A) tails (Caponigro and Parker, 1995). Similarly, an increase in deadenylation of mRNAs leads to their decapping and degradation in the 5'-> 3' direction (Muhlrad et al., 1994).

Several proteins linking the decapping and deadenylation complexes have also been identified through recent structural and biochemical studies, providing a mechanism through which the message could be transferred between the two ends of the mRNA. DDX6, a decapping activator and a translational repressor, has been shown to interact with CNOT1, a scaffold protein of the CCR4-NOT deadenylase complex (Mathys et al., 2014; Rouya et al., 2014). Furthermore, a recent study has proposed 4E-T, a translational repressor and known interactor of eIF4E and DDX6, to act as the bridge to link the 5' and 3' ends (Nishimura et al., 2015; Ozgur et al., 2015). However, similar to translating mRNPs, the interactions between the decapping and deadenylation complexes have only been observed through biochemical and *in vitro* studies, and single mRNAs forming a closed-loop structure through these interactions have not yet been observed in cells.

The interaction between the deadenylation and decapping complexes is vital for mRNA regulation through other pathways. miRNAs, which generally bind to the 3' end of the mRNA, are known to associate with Argonature (Ago) and GW182 to form the miRNA-induced silencing complex (miRISC) complex, which is loaded onto the mRNA. GW182 interacts with the deadenylase complexes CCR4-Not and Pan2-Pan3 (Braun et al., 2011; Chekulaeva et al., 2011). Similarly, m^6A modification, predominantly found within the 3'UTRs of mRNAs, can induce degradation of mRNAs. The degradation is mediated through m^6A binding protein YTHDF2, which interacts with and recruits the CCR4-Not complex to the mRNA (Nachtergaele and He, 2018).

Is a closed-loop state a default organisation of cytoplasmic mRNPs?

A closed-loop organisation could be considered a preferable state for cytoplasmic mRNPs, allowing for easy communication between the ends, modulating mRNA translation and stability.

Several observations support a model where mRNAs, through evolution, could have maintained spatial proximity between their ends. For example, RNAs, including several mRNAs, have been observed to fold into structures *in vitro* with their ends in proximity. This general property of RNAs might have played a determinant role in the evolution of complexes and might also provide a means for maintaining a closed-loop state *in vivo* (Lai et al., 2018; Leija-Martínez et al., 2014). Similarly, the interaction between 4E-4G-PABPC1 is conserved across evolution and increases the affinity for closed-loop factors on the mRNA. The presence of such interactions could have further stabilised a closed-loop structure.

Moreover, a closed-loop structure could have numerous functional benefits. Keeping the 5' and 3' ends in proximity could aid in recycling ribosomes once they have finished translating the mRNA, increasing the efficiency of translation initiation. Additionally, the increased association on mRNAs of the closed-loop factors upon interacting with each other could increase the stability of the mRNAs as a result of decreased accessibility of the cap and the poly(A) tail for degradation factors. Furthermore, the interaction between the deadenylation and decapping complexes could be a quality control measure to detect if the entire mRNA is intact.

Despite the arguments for the closed-loop model, the prevalence of a closed-loop structure is still questionable. For instance, though abrogation or inhibition of eIF4G-PABPC1 interaction results in a decreased translation level, yeast strains lacking these interactions are still viable (Park et al., 2011). Furthermore, mammalian cells express multiple isoforms of eIF4G1, some lacking the N-terminus required for the interaction with PABPC1, and multiple eIF4E, eIF4G and PABPC1 proteins, some not capable of forming interactions required for a closed-loop state (de la Parra et al., 2018). eIF4G could also interact with PABC1 bound to A-tract stretches within the body of mRNAs, though the nature of such interactions has not yet been characterized (Rissland, 2017).

1.5.5 mRNP compaction and organisation in the cytoplasm

The assembly of ribosomes drastically alters the composition and properties of mRNPs. Ribosomes being large macromolecular structures not only influence the mass of the mRNPs, but their distribution within the ORF could dictate the biophysical properties, compaction and flexibility of the polysomes, influencing the communication between different regions. Our current understanding of mRNP organisation and compaction in the cytoplasm comes from a milieu of

studies visualising polysomes *in vitro* and *in vivo* using electron microscopy, cryo-EM, live-cell single mRNA tracking and atomic force microscopy.

EM and super-resolution images of polysomes show polysomes with different conformations

Though polysomes had been visualised earlier, direct evidence determining their spatial conformations was only obtained in the 1980s from electron micrographs of Endoplasmic reticulum (ER) bound polysomes. These membrane-bound polysomes were observed to have either a circular or a G-spiral shape (Figure 1-11B) (Christensen et al., 1987). The possible conformations of polysomes were expanded to include hairpin and spiral shapes found in the ER of cultured fibroblasts and thyroid epithelial cells (Figure 1-11C) (Christensen and Bourne, 1999). The abundance of hairpin and circular polysome conformations indicated that cytoplasmic mRNPs could have stable closed-loop conformations in cells. However, recent cryo-ET visualisation of polysomes has yielded vastly different results suggesting that such conformations might not be universal (Brandt et al., 2010). Though the authors found polysomes to adopt numerous conformations, the predominant one had a helical shape, with the ends far from each other (Figure 1-11A). Surprisingly, the organisation of ribosomes within these polysomes was highly similar to polysomes visualised in bacteria, suggesting a possible universal evolutionary pressure that determines the polysome conformations (Brandt et al., 2009). Similar variability in polysome conformations has also been observed for purified polysomes imaged using atomic force microscopy or mRNAs translated using an *in vitro* wheat germ translation system visualised through cryo-EM (Afonina et al., 2013, 2014; Viero et al., 2015). The vast difference in these observations points to the lack of clarity about polysome conformations in cells. Though there is an abundance of biochemical evidence to suggest a closed-loop state in cells, whether it results from the spatial proximity of the 5' and 3' ends and a stable closed-loop form is yet to be determined.

The advantage of electron microscopy in resolution is coupled with several drawbacks that have limited its use in visualising cytoplasmic mRNPs. 1) mRNPs are dynamic molecules that can go through active translation cycles mixed with intermittent periods where the mRNA is not translated (Morisaki et al., 2016; Wu et al., 2016; Yan et al., 2016). A single transcript could have mRNAs in these different stages, mRNAs with a varying number of ribosomes bound to them or with a varied distribution of ribosomes along the length of the CDS. As electron microscopy only allows

for visualising the entire pool of polysomes, the variation within a single transcript and its effect on mRNP conformations cannot be captured. 2) In addition to variability between different transcripts, there is inherent variability within an mRNA over time. The dynamic changes in mRNP composition and its effect on mRNA conformations cannot be visualised due to the fixed view of polysomes *in situ*. 3) The inability to directly visualize the mRNA within the polysome makes it harder to ascertain where different regions of the mRNA are. As ribosome distribution can vary within a transcript, and mRNAs can have UTRs with varying lengths, this could influence the relative localisation of the ends of the mRNA not captured through the visualisation of polysomes (Wu et al., 2016)..

Ribosome-bound mRNPs are potentially bulkier and less compact than non-translating mRNPs

EM images suggest that polysomes potentially occupy a volume more significant than observed for naked mRNAs *in vitro*. Ribosomes are megadalton complexes with a helicase activity capable of disassembling secondary structures (Takyar et al., 2005). Therefore, the assembly of polysomes could result in the decompaction of regions with active translation. Consistent with this hypothesis, single-molecule tracking of single mRNAs found that the presence of ribosomes on mRNAs drastically slowed its diffusion, suggestive of a larger and less compact particle (Katz et al., 2016; Wu et al., 2015, 2016). However, the change in diffusion could also be attributed to an increase in mass as an mRNA assembles multiple ribosomes. The role of ribosomes in mRNA decompaction and the effect of decompaction on cytoplasmic mRNP organisation is not yet known.

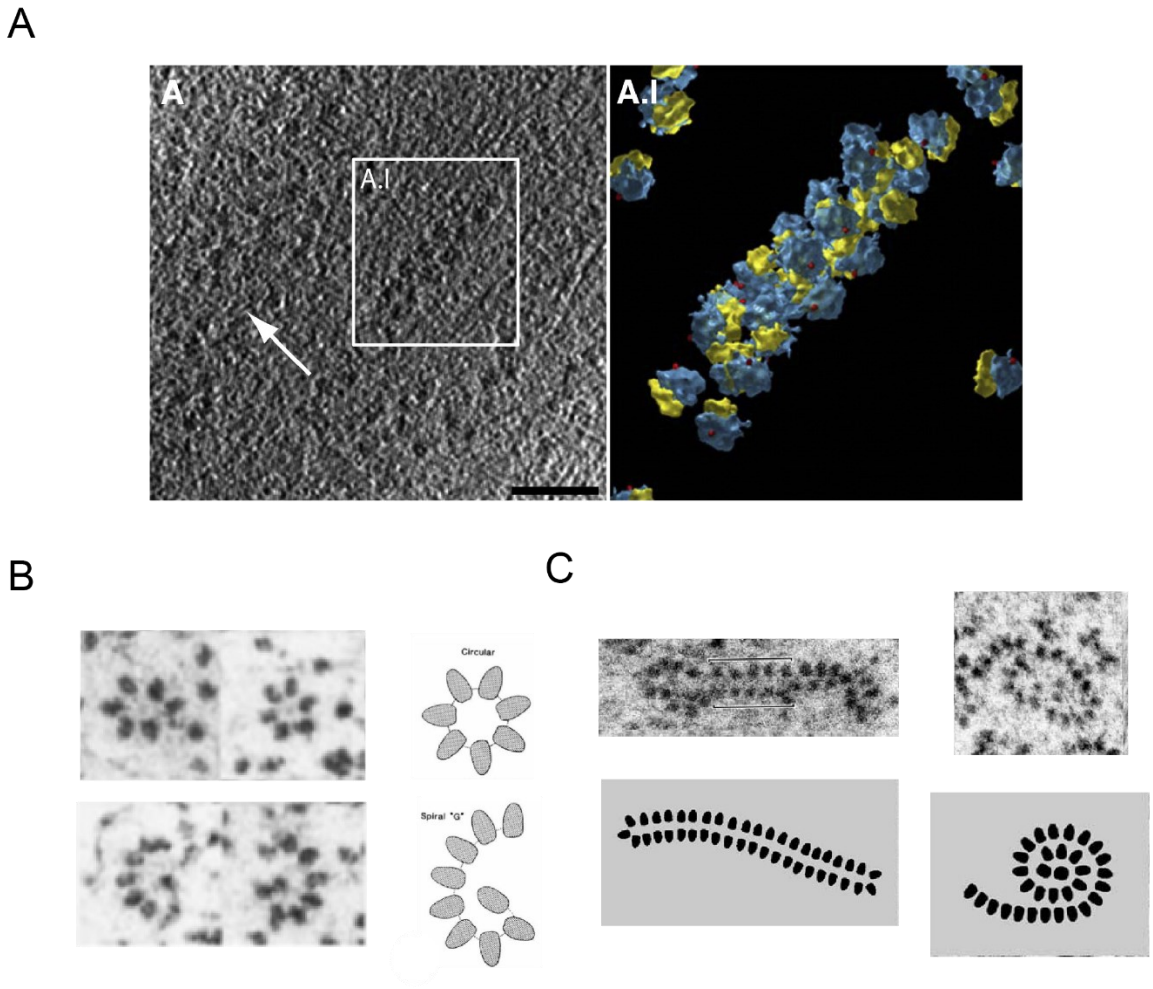


Figure 1-11: Polysome conformations observed *in vivo*

A) Polysomes observed in human glioblastoma cells using cryo-ET. Adapted from (Brandt et al., 2010). *B and C) Polysome conformation found on the ER of B)Rat anterior pituitary and C) Rat cultured fibroblasts visualised using electron microscopy. Adapted from* (Christensen et al., 1987) *and* (Christensen and Bourne, 1999).

1.6 Research objectives of this work

The role of the mRNA-protein complex, mRNP, in regulating gene expression is central to cellular growth, differentiation, and diseases. The composition of an mRNP is dynamically altered throughout its lifetime, with several RNA-binding proteins and RNA-protein complexes known to associate with and regulate different steps of mRNP metabolism. While the function of individual

proteins and factors have been studied extensively, how mRNAs and proteins are involved in assembling the 3D mRNP assembly is still unclear. As described above, the 3D organisation of an mRNP is critical at different stages of a lifetime, from preventing R-loop formation to facilitating mRNP splicing, export, translation and degradation. Therefore, understanding the principles that govern mRNP organisation in cells and the factors that alter it is key to understanding how cells regulate gene expression.

The aim of my PhD thesis was to develop an RNA-centric view of mRNP organisation in cells and characterise the role of cellular processes in determining and altering it. For this, I have developed an approach combining super-resolution microscopy with single-molecule visualisation of different regions of specific mRNAs, an approach that provides a single-RNA perspective of mRNP conformations in a cellular context. I present two studies in my thesis that have helped characterise mRNP conformations in the nucleus and cytoplasm. In the first study, I aimed to get an overview of mRNP organisation in the different cellular compartments, with a specific focus on determining the role of translation in mRNP compaction and organisation. In the second study, I expanded upon the tools I had developed and used single-molecule microscopy in the context of mRNP processing to determine the assembly and organisation of pre-mRNPs in the nucleus.

2. Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway

2.1 Aims of Article 1

For my first project, I worked on determining the spatial organisation of mRNPs in human cell lines. The metabolism of mRNPs in cells is regulated through proteins binding to mRNAs at different stages in its lifetime and through different regions of mRNAs interacting through proteins bound to them. One such interaction is between the 5' and 3' end of an mRNA during translation when the ends of the mRNA are brought together through interactions between the cap-binding protein eIF4E, adapter protein eIF4G and poly(A) binding protein PABPC1. However, how mRNAs are organised to bring together regions, separated by hundreds to thousands of nucleotides in space, is poorly understood. Previously, electron microscopy and cryo-electron tomography studies have visualised certain mRNPs and polysomes in cells, but due to the lack of specificity for mRNAs and the inability to visualise the RNA within the mRNP particles, the exact nature of the organisation of specific mRNPs and factors that contribute to it is not understood. To determine mRNA organisation in cells, I developed a microscopy-based approach by combining smFISH with super-resolution microscopy to visualise different regions of specific mRNAs and used this approach to determine mRNA organisation in cells. The results of this study were published in *Molecular Cell* in November 2018 where I was the first author (Adivarahan, Srivathsan, et al. "Spatial organisation of single mRNPs at different stages of the gene expression pathway." Molecular Cell 72.4 (2018): 727-738.).

2.2 Article 1

Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway

Srivathsan Adivarahan[1], Nathan Livingston[2], Beth Nicholson[3], Samir Rahman[1], Bin Wu[2], Olivia S. Rissland[3,4] and Daniel Zenklusen[1]*

Affiliations:

(1) Département de Biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, Canada

(2) The Department of Biophysics and Biophysical Chemistry, the Solomon Snyder Department of Neuroscience, Center for Cell Dynamics, Johns Hopkins School of Medicine, Baltimore, USA

(3) Molecular Medicine Program, The Hospital for Sick Children, Toronto, Canada

(4) Current address: Department of Biochemistry and Molecular Genetics, RNA Bioscience Initiative, University of Colorado School of Medicine, Aurora, USA

*Corresponding author: daniel.r.zenklusen@umontreal.ca

2.2.1 Summary

mRNAs form ribonucleoprotein complexes (mRNPs) by association with proteins that are crucial for mRNA metabolism. While the mRNP proteome has been well characterised, little is known about the mRNP organisation. Using a single-molecule approach, we show that mRNA conformation changes depending on its cellular localisation and translational state. Compared to nuclear mRNPs and lncRNPs, association with ribosomes decompacts individual mRNAs, while pharmacologically dissociating ribosomes or sequestering them into stress granules leads to increased compaction. Moreover, translating mRNAs rarely show co-localised 5' and 3' ends, indicating that mRNAs are either not translated in a closed-loop configuration, or that mRNA circularisation is transient, suggesting that a stable closed-loop conformation is not a universal state for all translating mRNAs.

2.2.2 Introduction

RNAs are single-stranded nucleic acid polymers. Intramolecular base pairing and binding of RNA-binding proteins (RBPs), many of which contain homo and hetero-dimerisation domains, assemble mRNAs into RNPs (Singh et al., 2015). Assembly of mRNPs is initiated co-transcriptionally, and mRNP composition is thought to constantly change during the different processing and maturation steps and upon translocation to the cytoplasm when mRNAs meet with ribosomes for translation. Proteomic approaches have identified many RBPs assembling to mRNA at these different stages, and recent cross-linking approaches have identified binding sites for many of these proteins, leading to comprehensive maps of mRNP composition (Hentze et al., 2018; Marchese et al., 2016). Similarly, recent transcriptome-wide chemical mapping approaches have identified single and double-stranded regions within mRNAs revealing extensive internal secondary structures (Strobel et al., 2018). More broadly, mRNA organisation is crucial for many aspects of mRNA metabolism, especially steps where different regions within (pre-) mRNAs are thought to communicate, such as splicing, translation regulation or miRNA-mediated regulation (Fabian and Sonenberg, 2012; Imataka et al., 1998; TarunSachs, AB, 1996). Despite the importance of mRNA organisation, little is known about how mRNPs are organised as three-dimensional assemblies.

Much of our understanding of mRNP organisation comes from in vivo and in vitro electron microscopy studies. Electron tomography studies of the 35kB-long Balbiani Ring (BR) mRNA in C. tentans salivary glands revealed a dense nuclear particle with a diameter of about 50 nm where 5’ and 3’ are in close proximity (Skoglund et al., 1986). A different architecture was observed for nuclear mRNAs purified from yeast and analysed by electron microscopy (EM) that revealed particles with a homogenous width but variable length (5 nm wide, 20–30 nm long), suggesting a linear assembly with the ends separated (Batisse et al., 2009). The organisation of cytoplasmic mRNAs, on the other hand, has been primarily studied by visualising ribosomes as a proxy for visualising mRNA. Polysomes containing various numbers of ribosomes and in different conformations have been observed in vivo, as well as in vitro. Polysomes are found either in spiral arrangements, forming double rows of ribosomes, arranged in circles as well as in less defined, open conformations. However, how the mRNA is organised within these polysomes is not visible in these experiments (Afonina et al., 2015; Brandt et al., 2010; Lu et al., 2016; Ramani et al., 2015; Rech, 1995). Considering a ribosome footprint of about 30nt and an average ribosome density of

about one ribosome per 200-900nt, large regions of the mRNA must be exposed between individual ribosomes as well as in their 5' and 3' UTRs(Hendrickson et al., 2009; Steitz, 1969; C. Wang et al., 2016; Yan et al., 2016). However, the organisation of this higher-order structure is not known.

The best-studied example for the role of intramolecular mRNA organisation in gene regulation is the communication between 5' and 3' ends during translation (Gallie, 1991). The current model of initiation is that mRNAs are organised in a circular conformation, mediated by a series of interactions between the 5' cap, the cap-binding protein eIF4E, the adaptor protein eIF4G, poly(A) binding protein (PABPC1), and the poly(A) tail. These interactions have been proposed to bring together the ends of the transcript to stimulate translation (Jackson et al., 2010). This closed-loop model is supported by many studies showing physical interactions between eIF4E, eIF4G and PABPC1, in vitro experiments reconstituting mRNA circularisation using purified components, as well as the observation of circular polysomes in cells (Christensen et al., 1987; Imataka et al., 1998; TarunSachs, AB, 1996; Wells et al., 1998). However, many polysomes in cells observed by EM show configurations that do not suggest a closed-loop, and so it is unclear whether closed-loops represent transient states, polysomes with mRNAs with connected ends but where ribosomes are positioned distant from the 5' and 3' ends, or different classes of transcripts for which translation of only some might occur in a closed-loop configuration. Furthermore, although there are examples of factors that repress gene expression by connecting the 5' and 3' ends, how the ends are physically brought together to establish these complexes is not known (Jonas and Izaurralde, 2015).

Thus, the fundamental issue of how mRNAs are organised as mRNPs *in vivo* remains unresolved. In this study, we investigate mRNA organisation within cells by combining Structured Illumination Microscopy (SIM) with single-molecule resolution fluorescent in situ hybridisation (smFISH) to investigate the spatial relationship of various regions within mRNAs in different cellular compartments. We observe that mRNAs exist in different levels of compaction depending on their cellular localisation and translation state and show that translation, at least for a subset of mRNAs studied here, results in the separation of 5' and 3' ends, suggesting that these RNAs are not translated in a stable closed-loop.

2.2.3 Results

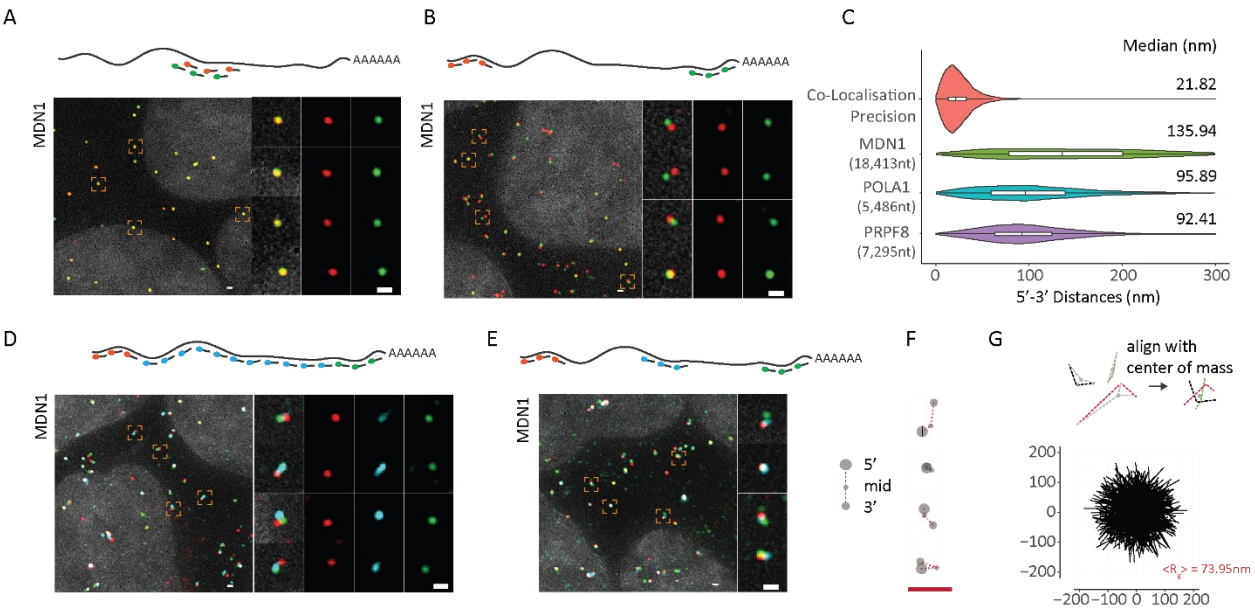
2.2.3.1 Visualising different regions within mRNAs using smFISH and SIM

To determine whether combining smFISH and SIM allows to spatially resolve different regions within single mRNAs, we first measured co-localisation precision by hybridising a mix of 44 20nt-long DNA probes, alternatingly labelled with cy3 and cy5, and spanning a 1.2 kb region within the 18,413 nt-long MDN1 mRNA in paraformaldehyde-fixed HEK293 cells (Figure 2-1A). Probes spreading the 1.2kB region were used to ensure enough single smFISH probes, each with similar annealing temperature, bind the mRNA and together result in sufficient signal for robust detection and localisation of individual mRNAs. Images were acquired spanning the entire cell volume, and 3D datasets were reduced to 2D by maximum intensity projection. We then determined the centre of each signal by 2D Gaussian fitting and measured the distance between signals from both channels. 2D Gaussian fitting calculates the centroid of the signal emitted from individual fluorescent particles spread over multiple pixels on the detector and allows sub-diffraction localisation precision (R. Thompson et al., 2002; Zenklusen et al., 2008). Measuring the distances between co-localising spots showed a co-localisation precision of 21 nm, indicating that we can resolve discrete regions within mRNAs when separated by more than 20 nm (Figure 2-1C).

We then positioned labelled probes to the 5' and 3' ends of the *MDN1* mRNA to determine RNA extension in cells (Supplementary Figure 2-1, Table S2-2 and Table S2-3), which, in a hypothetical scenario with 0.59 nm spacing between nucleotides for a rigid ssRNA, would measure about 10.8 µm in length when fully extended (Liphardt et al., 2001). However, as a flexible polymer, mRNA is unlikely to exist in such a conformation, which will depend on different parameters, including the stiffness of the polymer chain, as well as by thermodynamics and intra-molecular interactions, which will reduce the overall extension of the mRNA (Borodavka et al., 2016; Chen et al., 2012; Gopal et al., 2012; Liphardt et al., 2001). Analysing cytoplasmic *MDN1* mRNAs, we observed few overlapping 5' and 3' signals; instead, a majority of 5' signals had a 3' signal within close proximity (Figure 2-1B), with distances up to 300 nm between the two signals (Figure 2-1 C). A similar distribution was observed when measured in 3D, and distances were indistinguishable when the EtOH step in the hybridisation protocol was omitted (Supplementary Figure 2-2 A, B). To determine if 5' and 3' signals were part of the same mRNA molecule, we used a third set of

FISH probes tiling the entire length of the mRNA between the 5' and 3' regions in 500 nt intervals. The tiling signal overlapped with either one of the two regions and connected the 5' and 3' within the 300 nm radius, confirming that 5' and 3' end signals belonged to the same molecule and, moreover, pointing towards an elongated conformation of cytoplasmic *MDN1* mRNPs (Figure 2-1D). To better understand the spatial relationship between different regions within these mRNAs, we replaced the tiling probes with a probe set hybridising to the middle region of the *MDN1* mRNA (Figure 2-1E, Supplementary Figure 2-1). Using these probes, we observed cytoplasmic mRNAs where the three different regions could be spatially resolved (Figure 2-1E and F). To measure the average volume of these cytoplasmic mRNAs, we aligned individual mRNAs using their centre of mass and calculated the radius of gyration (<Rg>), a measure of the global size of the mRNP, and found a mean radius of gyration of 73.95 nm (Figure 2-1G). These dimensions are comparable to the size of polysomes imaged by electron microscopy and super-resolution microscopy, in which polysomes containing 6-10 ribosomes, as suggested for the ribosome occupancy for *MDN1*, typically have a diameter ranging from around 90-150 nm (Brandt et al., 2010; Christensen et al., 1987; Floor et al., 2016; Viero et al., 2015).

To determine whether such open conformations are particular to the long *MDN1* mRNA or are a more common feature of cytoplasmic mRNAs, we measured the compaction of two shorter mRNAs encoding for the splicing factor PRPF8 (7,295 nt) and the DNA polymerase alpha catalytic subunit POLA1 (5,486 nt) and found similar open conformations (Supplementary Figure 2-2C). End-to-end distances showed narrower distributions compared to *MDN1* mRNAs, indicating that maximum expansion in cells as a function of mRNA length (Figure 2-1C). Together, these data show that cytoplasmic mRNAs predominantly exist in an open conformation where 5' and 3' are rarely found in close proximity.



Adivarahan et al. Figure 1

Figure 2-1: Visualising single mRNA reveals open conformations of cytoplasmic mRNAs.

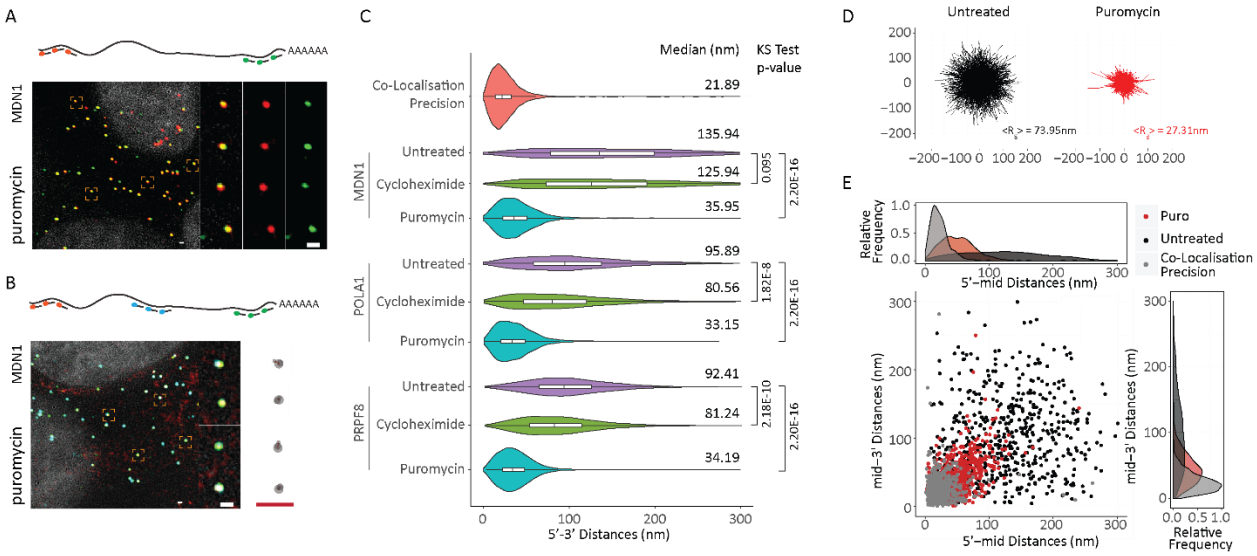
A) *smFISH images using alternating probes labelled in cy3 (red) and cy5 (green) to the middle region of MDN1 mRNA (Probe Set#1, Table S2-3) in paraformaldehyde-fixed HEK 293 cells. Nuclei are visualised by DAPI staining (grey). Magnified images of individual RNAs marked by dashed squares are shown on the right. Schematic position of probes shown on top.* ***(B)*** *smFISH using probes targeting the 5' (red) and 3' (green) ends of MDN1 (Probe Set#2, Table S2-3).* ***(C)*** *Violin plots showing distance distribution of co-localization precision of co-localising spots from A, and 5'-3' distances for MDN1, POLA1, PRPF8 mRNAs determined by Gaussian fitting. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances are shown on the right.* ***(D, E)*** *smFISH using 5' (red), 3' (green), and tiling or middle probes (cyan), respectively (Probe Sets#3,4, Table S2-3).* ***(F)*** *Cartoon depicting different mRNA conformations from E.* ***(G)*** *Projections of superimposed conformations from E with their centres of mass in the registry, n=563. Mean Radius of gyration (<R_g>). Scale bars, 500 nm*

2.2.3.2 Open mRNP conformation corresponds to translating mRNA

Translating mRNPs are thought to exist in a closed-loop conformation where 5' and 3' ends of the mRNA are brought together through interactions between the cap-binding eIF4F complex and the polyA binding protein PABPC1 (Imataka et al., 1998; TarunSachs, AB, 1996; Wells et al., 1998). Surprisingly, we rarely observed 5'-3' conformations consistent with such a closed-loop configuration. One possibility could be that most mRNAs with separated 5' and 3' ends are not in the process of being translated and that only the fraction with co-localising ends represents the pool of translating mRNAs. If that were the case, interfering with translation should further reduce the fraction of mRNAs with co-localising 5' and 3' ends. To test this hypothesis, we treated cells, prior to fixation, with drugs that affect translation via different mechanisms: cycloheximide, which inhibits elongation by binding to the E-site of the 60S ribosomal unit and stabilises polysomes, and puromycin, which causes premature chain termination and disassembles polysomes (Bhat et al., 2015). Treatment with cycloheximide only modestly affected the distribution of 5'-3' *MDN1, PRPF8* and *POLA1* end distances when compared to untreated cells, with slightly lower end-to-end distances, suggesting that elongating ribosomes contribute to the openness of the mRNA (Figure 2-2C). In contrast, the disassembly of polysomes following a short treatment with puromycin (10 min) resulted in an unexpected phenotype where the 5'-3' ends of most transcripts were co-localising (Figure 2-2A). For *MDN1,* distance measurements showed a narrow distribution with a median of 36 nm. A one-hr treatment with the translation inhibitor homoharringtonine, which stalls ribosomes at the initiation site, yielded similar results (Supplementary Figure 2-3A). Similarly, *POLA1* and *PRPF8* ends also showed a high degree of co-localisation with similar median 5'-3' end distances (Figure 2-2C and Supplementary Figure 2-3B).

These observations could be due to either a change in mRNP conformation resulting in increased levels of 5'-3' interaction or the result of general compaction of the mRNP because of the loss of bound ribosomes. To distinguish between these possibilities, we repeated the experiment, this time using probes that hybridise to the middle region of *MDN1* mRNA or tile along its entire length and found that puromycin treatment resulted in general compaction of the mRNPs (Figure 2-2B, Supplementary Figure 2-3C). Overlaying mRNA conformations revealed a less extended form of these mRNPs compared to untreated cells (Figure 2-2D and E). These observations suggest that

most of these cytoplasmic mRNAs are translating, that mRNAs within translating mRNPs are not arranged in a closed-loop conformation, and that disassembly of polysomes results in highly compact mRNAs.



Adivarahan et al. Figure 2

Figure 2-2: Open mRNP conformation corresponds to translating mRNA.

*(**A, B**) 5' and 3' (Probe Set#2, Table S2-3) or three colour MDN1 smFISH (Probe Set#4, Table S2-3) in HEK 293 cells treated with puromycin (10 min, 100 µg/ml). **(C)** Violin plots showing 5'-3' distances for MDN1, POLA1, PRPF8 mRNAs in untreated and cells treated with cycloheximide and puromycin. White box plot inside the violin plot shows the first quartile, median and third quartile. Median distances and p-values calculated using the Kolmogorov-Smirnov test are shown on the right. **(D)** Projections of superimposed conformations from three colour MDN1 smFISH (Probe Set#4, Table S2-3) in untreated and puromycin treated cells with their centres of mass in the registry, n=563. Mean Radius of gyration (<R_g>). **(E)** Scatter plot showing 5'mid and mid-3' distances for individual RNAs. Frequency distribution is shown on top and on the right. Scale bars, 500 nm.*

2.2.3.3 Inhibiting eIF4G1-PABPC1 interaction does not alter open mRNP conformations

Nonetheless, we noted that a small fraction of *MDN1* mRNAs in untreated cells had ends in close proximity. If we consider 50 nm as an upper limit for a closed-loop configuration, 12.5% of *MDN1* mRNAs show ends that are closer than 50 nm as judged by 2D analysis. Because 2D projection analysis overestimates proximity due to the projection of the z dimension, we re-analysed the data in 3D to refine our estimate of mRNAs potentially in the closed-loop confirmation and found that only 4.4% are found closer than 50 nm, further suggesting that the close proximity of the ends is a rare event (Supplementary Figure 2-2A). To determine whether this small fraction indeed represents closed-loop conformations mediated by PABPC1–eIF4G1 interactions, we used CRISPR/Cas9 gene editing to construct two different cell lines mutating key residues in PABPC1 or eIF4G1 needed for the interaction as well as matched wild-type controls (Figure 2-3A). Both mutant cell lines showed reduced interactions, but these mutations had minimal impact on cell survival and overall translation, although there was a slight increase in the monosome: polysome ratio in the eIF4G1 mutant cell line (Figure 2-3B-D). Although other paralogues of PABPC1 and eIF4G1 are present in the human genes expressed at lower levels in HEK293 cells, these proteins are not sufficient to compensate for a knock-out of eIF4G1 and are expressed at lower levels (Hart et al., 2015). When we analysed the conformation of cytoplasmic *MDN1* mRNAs in these mutant cell lines, we observed a 5'-3' distance distribution similar to those in WT cells (Figure 2-3E), although we observe a modest increase in end-to-end distances for the mutant cell lines. Notably, the fraction of *MDN1* mRNAs with 5'–3' distances below 50 nm remained unchanged, indicating that the small 5'-3' colocalising fraction is not dependent on the PABPC1–eIF4G1 interaction. Treatment with puromycin resulted in increased proximity of 5' and 3' ends, indicating that the open conformations that we observe represent translating mRNAs (Supplementary Figure 2-4 A, B). Although we cannot formally exclude compensation by other paralogues, our data strongly suggest that the colocalisation fraction may instead be non-translating mRNAs or mRNPs where the ends are close to each other independent of the PABPC1–eIF4G1 interaction, possibly due to the flexibility of the RNA polymer.

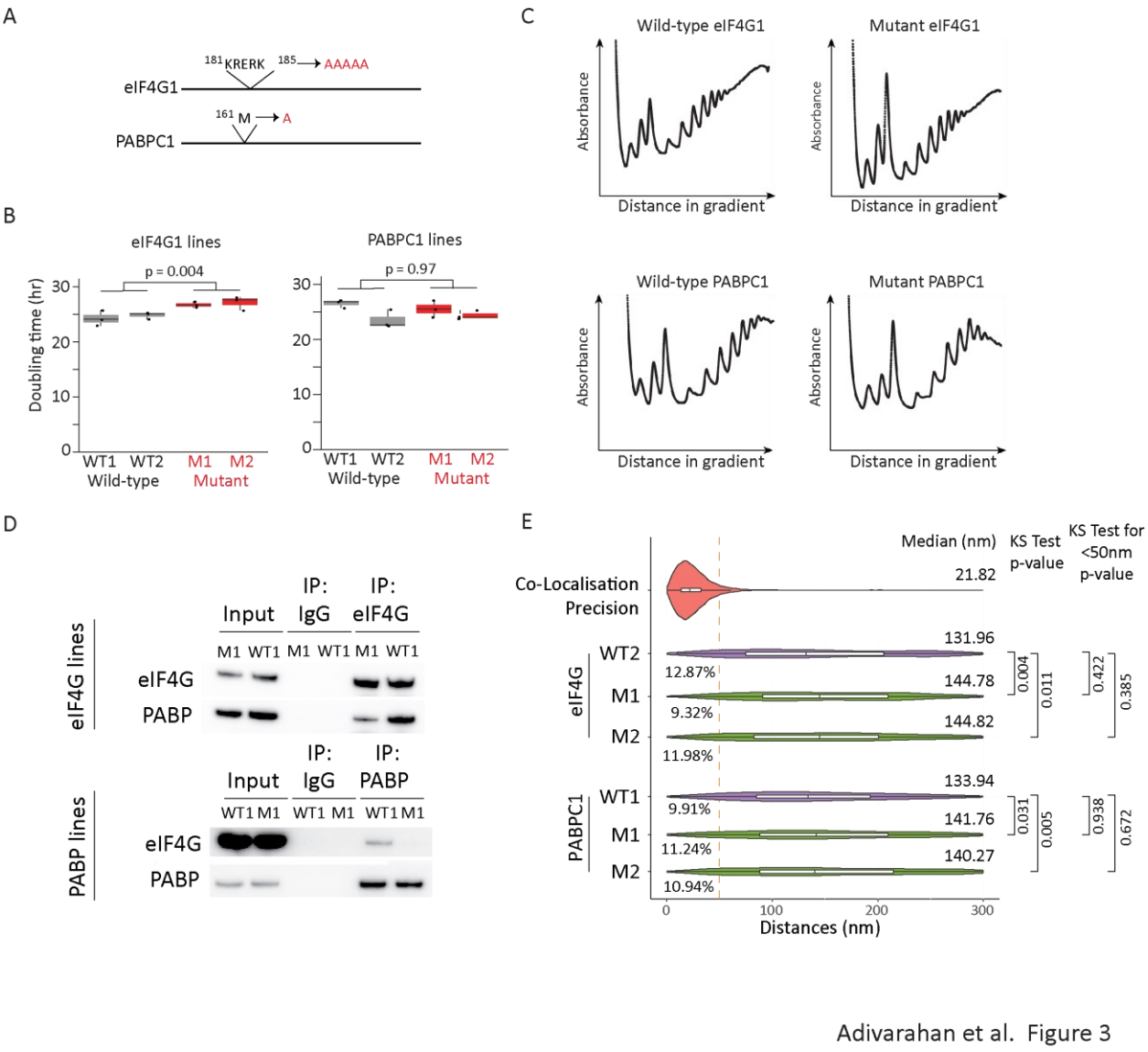


Figure 2-3: Inhibiting eIF4G1-PABC1 interactions does not alter 5'-3' distances.

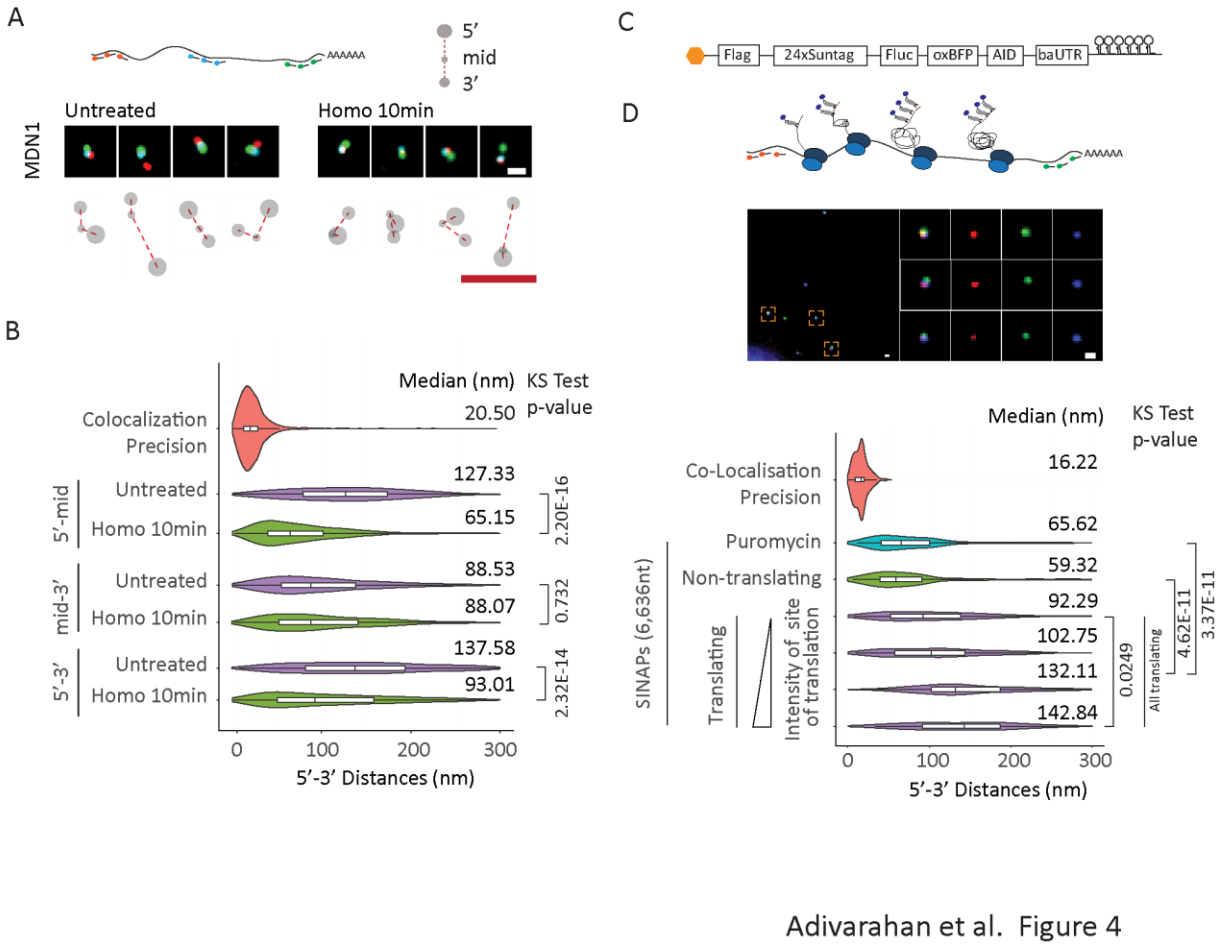
(A) *Sites of amino acid substitutions in eIF4G1 and PABPC1 cell lines.* ***(B)*** *Doubling time for eIF4G1 and PABPC1 CRISPR-edited lines. Shown are the doubling times calculated for three independent biological replicates for two independent wild-type and mutant eIF4G1 and PABPC1 lines.* ***(C)*** *Polysome profiles for wild-type eIF4G1, wild-type PABPC1, mutant eIF4G1, and mutant PABPC1 lines.* ***(D)*** *Immuno-precipitation of eIF4G1 and PABPC1 from wild-type and mutant cell lines using anti- eIF4G1 and PABPC1 anti-bodies.* ***(E)*** *Violin plots showing distance distribution of co-localization precision from 5'-3' distances for MDN1 in wild type and mutant cell lines (Probe Set#2, Table S2-3). White box plot inside the violin plot shows first quartile, median and*

third quartile. Median distances and p-values calculated using Kolmogorov-Smirnov test are shown on the right. WT1/WT2/M1/M2 represent different clonal cell lines.

2.2.3.4 Ribosome occupancy determines mRNP compaction

To further investigate the role of ribosome occupancy on mRNP compaction, we performed a ribosome run-off experiment using the translation inhibitor homoharringtonine. Treatment with homoharringtonine inhibits new initiation but allows elongating ribosomes to continue translating until reaching the stop codon, allowing us to determine local compaction within the *MDN1* mRNA upon short drug treatment. Translation elongation in human cells is thought to occur at about 5aa per second, therefore for the 16,791 nt *MDN1* open-reading frames, after a 10 min treatment, the first half will be devoid of ribosomes, whereas the second half will still contain ribosomes(Bin Wu et al., 2016). Consistent with the requirement of ribosome occupancy for RNA decompaction, the 5'-to-mid region of *MDN1* mRNA became compacted after the 10-min homoharringtonine treatment, whereas the mid-to-3' region remained in an open conformation (Figure 2-4A, B, Supplementary Figure 2-5).

To further investigate the relationship between translation and 5'–3' proximity, we employed a reporter system developed for Single-Molecule Imaging of Nascent Peptides (SINAPs), where nascent proteins are rapidly bound at the ribosome exit channel by a fluorescently labelled single-chain antibody (scFv-sfGFP) and fluorescence intensity, therefore, is proportional to the number of ribosomes on a specific mRNA (Figure 2-4C) (Bin Wu et al., 2016; Pichon et al., 2016; C. Wang et al., 2016; Yan et al., 2016). The SINAPs reporter was transfected into U2OS cells stably expressing the scFv-sfGFP fusion, cells were fixed after 24 hrs, and ribosome occupancy and mRNA conformation were simultaneously measured by smFISH and immunofluorescence targeting scFv-sfGFP using an anti-GFP antibody (Figure 2-4D). Consistent with our previous analysis, translating mRNAs had more open conformations relative to non-translating mRNAs, as judged by both nascent peptide signal and puromycin treatment. Significantly, the RNA 5' -3' distance increased with the relative intensity of nascent peptides. Taken together, our data indicate that ribosome occupancy decompacts mRNA and separates the ends.



Adivarahan et al. Figure 4

Figure 2-4: Ribosome occupancy determines mRNP compaction.

(A) *smFISH using 5' (red), 3' (green), and middle probes (cyan) respectively (Probe Set#4, Table S2-3) for untreated and homoharringtonine (100µg/ml, 10min) treated cells and cartoon depicting different mRNA conformations. At a translation speed of 5 amino acids per second, all translating ribosomes will have reached at least the mid region of MDN1.* ***(B)*** *Violin plots showing 5'-mid, mid-3' and 5'-3'distance distribution of cytoplasmic MDN1 mRNAs in untreated and homoharringtonine treated cells.* ***(C)*** *Cartoon depicting the SINAPs construct* ***(D)*** *Images showing 5' and 3' smFISH and anti-GFP immunofluorescence (Probe Set#15, Table S2-3) (top), and violin plots depicting 5'-3' distances for puromycin treated, non-translating and translating mRNAs. Translating mRNAs were clustered in 4 groups (k-means) according to intensity of anti-GFP signal (bottom). White box plot inside the violin plot shows first quartile, median and third quartile. Median distances and p-values calculated using Kolmogorov-Smirnov test are shown on the right. Scale bars, 500 nm.*

2.2.3.5 Compaction state of lncRNAs and mRNA sequestered to stress-granules

If translation is the main cause for an open mRNP conformation, we hypothesised that non-translating RNAs, such as cytoplasmic long non-coding RNAs (lncRNAs), might show a similar level of compaction than non-translating mRNAs, and, moreover, their compaction should be unaffected by translation inhibitors. To test this model, we measured end-to-end distances for two lncRNAs, TUG1 (7,469nt) and OIP5-AS1 (8,829nt), previously found to be present in the nucleus and the cytoplasm (Cabili et al., 2015). Both lncRNAs contain short putative ORFs that could lead to their association with ribosomes, however, their translation will be limited to the very 5'-end of the transcript (van Heesch et al., 2014). As shown in Figure 2-5A, 5' and 3' labelled cytoplasmic TUG1 and OIP5-AS1 lncRNAs displayed a more compact conformation compared to the similarly-sized *PRPF8* mRNA. In addition, 5'-3' distances of OIP5-AS1 lncRNA was unaffected by puromycin, further suggesting that decompaction of cytoplasmic mRNAs requires the formation of polysomes (Figure 2-5B). Interestingly, we observe a small but significant change of end-to-end distance for TUG1 upon puromycin treatment (Figure 2-5B). Unlike OIP5-AS1, TUG1 has been shown to associate with higher polysome fractions despite its very short putative ORFs, which could explain this observation (Floor et al., 2016).

We next hypothesised that if eviction of ribosomes from translating mRNAs by puromycin results in strong compaction of mRNA, then mRNAs that are translationally repressed in response to external stimuli or environmental triggers should also acquire a compact conformation. Treatment with sodium arsenite inhibits translation through phosphorylation of eIF2α and results in disassembly of polysomes and sequestration of mRNAs in stress granules (Buchan and Parker, 2009; Panas et al., 2016). We induced stress granule assembly in U2OS cells upon treatment with arsenite for 1 hr and found that this treatment relocated cytoplasmic *MDN1* mRNAs to stress granules (Figure 2-5C). Furthermore, mRNAs show a highly compact conformation, observed by measuring end-to-end distances using 5'- 3' probes and tiling probes spanning the entire transcript up to 3' regions labelled with differently labelled probes (Figure 2-5C). End-to-end measurements for *MDN1* mRNAs in stress granules showed a level of compaction similar to that seen in puromycin-treated cells (Figure 2-5D), and similar compaction was also observed for *POLA1* and *PRPF8* mRNAs under the same conditions (Supplementary Figure 2-6A). Interestingly, not all *POLA1* and *PRPF8* mRNAs accumulated in stress granules, but those mRNAs that remained

outside showed the same level of compaction as those within stress granules, suggesting that translation inhibition occurs independently of mRNA sequestration to stress granules, as previously suggested (Mollet et al., 2008; Panas et al., 2016; Souquere et al., 2009; Khong et al., 2017). Moreover, a fraction of TUG1 and OIP5-AS1 was also found localised to stress granules, and this localisation did not alter their compaction (Supplementary Figure 2-6B).

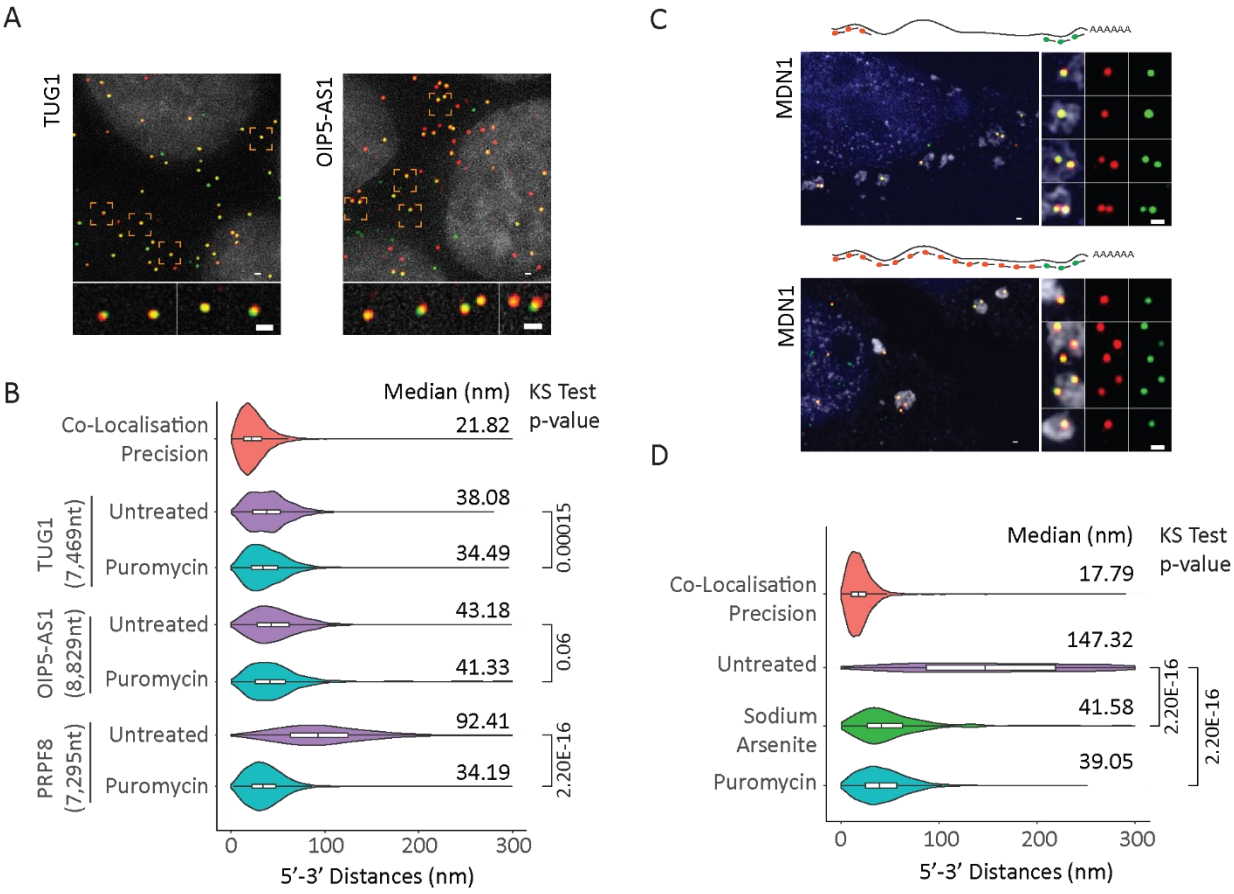


Figure 2-5: lncRNAs in the cytoplasm and mRNAs sequestered to stress granules show compact conformations.

(A) *smFISH visualising 5' and 3' ends of TUG1 and OIP5-AS1 lncRNAs (Probe Sets#7,8, Table S2-3). Nuclei are visualised by DAPI staining (grey).* ***(B)*** *Violin plots showing 5'-3' distance distribution of cytoplasmic TUG1 and OIP5-AS1 RNAs in untreated and puromycin treated cells*

compared to PRPF8 mRNAs. ***(C)*** *5' - 3' (Probe Set#9, Table S2-3) or 3' and tiling (Probe Set#10, Table S2-3) MDN1 smFISH in U2OS cells treated with arsenite (1 hr, 2 mM). Stress granules are visualised using an oligo dT probe (white). Nuclei are visualised by DAPI staining (blue).* ***(D)*** *Violin plots comparing MDN1 mRNA 5'-3' distance distribution for untreated, arsenite and puromycin treated U2OS cells. For arsenite treated cells, only mRNAs in stress granules were considered. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances and p-values calculated using Kolmogorov-Smirnov test are shown on the right. Scale bars, 500 nm.*

2.2.3.6 Organisation of nuclear mRNAs

We finally asked whether the compacted state of mRNAs found within stress granules or after puromycin treatment reflects a default state for non-translating cellular mRNPs. In the nucleus, nascent mRNAs are co-transcriptionally spliced and assembled into mRNPs resulting in the binding of a large set of RBPs, including the exon-junction complex and SR proteins (Le Hir et al., 2000; Müller-McNicoll and Neugebauer, 2013; Singh et al., 2012). The ribosome evicts many RBPs bound to the open reading frame during translation in the cytoplasm. mRNAs that have been translated and then go into a translationally silent state might therefore be bound by fewer proteins than cytoplasmic mRNAs prior to their first round of translation or nuclear mRNPs before their export to the cytoplasm.

To determine whether a default compaction state exists for non-translating mRNPs, we investigated the organisation of nuclear *MDN1* mRNAs. Compared to cytoplasmic *MDN1* mRNAs upon puromycin treatment, nuclear *MDN1* mRNAs were found in an extended conformation, although it was more compacted than translating cytoplasmic mRNAs (Figure 2-6A-E). Moreover, 5' to mid and mid to 3' distances were shorter than the 5' to 3' distance and the distances larger than cytoplasmic mRNAs upon puromycin treatment (Supplementary Figure 2-7A). Unlike for cytoplasmic mRNAs, open mRNP conformations of nuclear MDN1 were still observed upon puromycin (10 min) or homoharringtonine (1 h) treatment, although we measure a small overall reduction in 5'-3' distances (Supplementary Figure 2-7B, C). This might, in part, be due to the difficulty to accurately segment nuclear-cytoplasmic borders so that our analysis of nuclear mRNAs includes a small fraction of cytoplasmic mRNAs. Together, these observations suggest

that assembly of nuclear mRNPs results in more extended mRNP compared to translationally inhibited mRNPs.

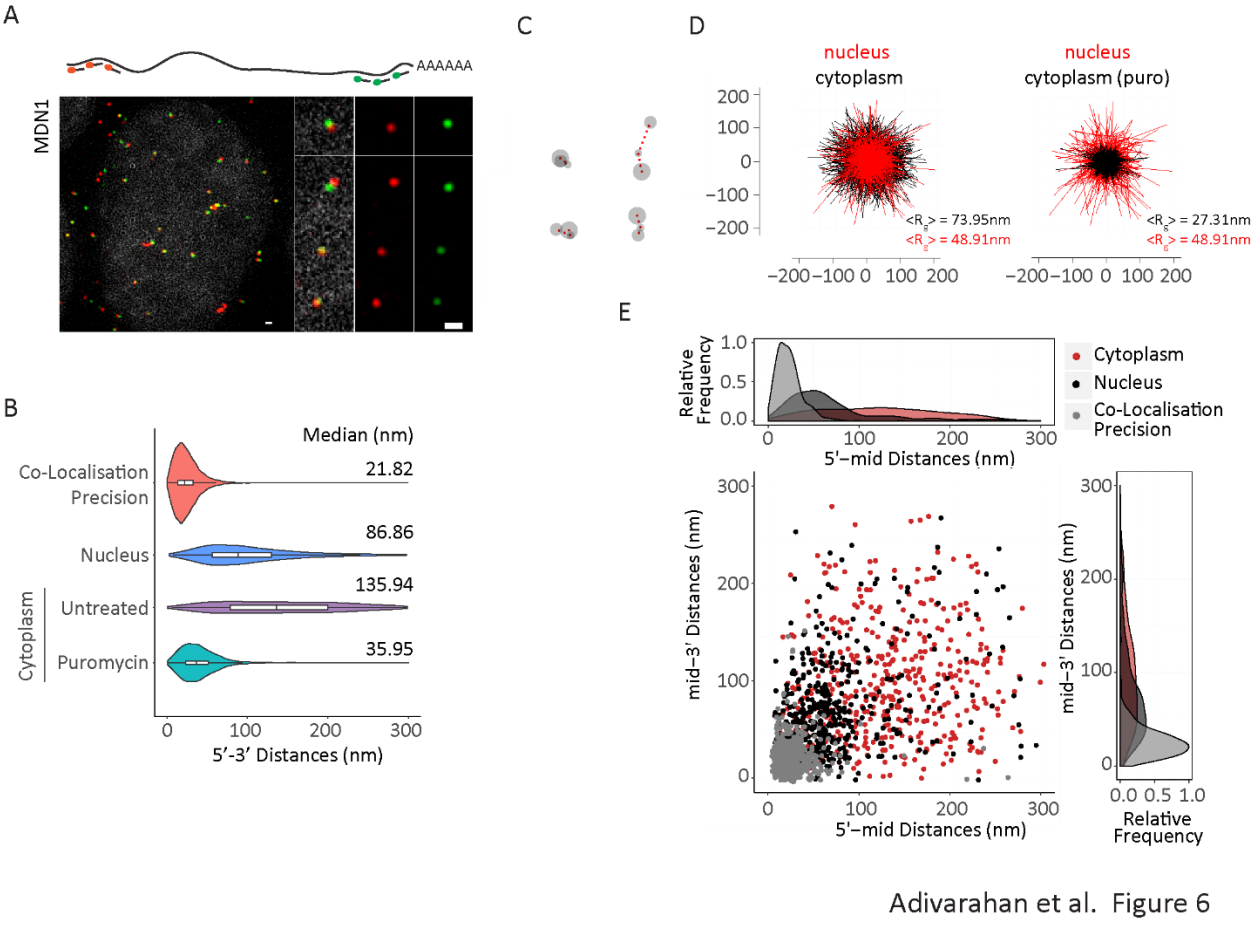


Figure 2-6: Organisation of nuclear MDN1 mRNAs.

(A) *5'- 3' MDN1 smFISH (Probe Set#2, Table S2-3) of nuclear mRNAs. The nucleus was stained with DAPI (gray).* ***(B)*** *Violin plots comparing MDN1 mRNA 5'-3' distance distribution of nuclear and cytoplasmic mRNAs. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances are shown on the right.* ***(C)*** *Representative conformations of nuclear MDN1 mRNAs measured by 5, middle and 3' labeling as in 1E.* ***(D)*** *Projections of superimposed conformations from C with their centers of mass in registry, compared to untreated or puromycin treated cytoplasmic MDN1 mRNAs, n=452. Mean Radius of gyration (<R_g>).* ***(E)*** *Scatter plot comparing 5'-mid and mid-3' distances for individual nuclear and cytoplasmic MDN1 mRNAs. Frequency distribution are shown on top and on the right. Scale bars, 500 nm.*

2.2.4 Discussion

Although the proteome of mRNPs has been studied extensively, the understanding of how mRNA and proteins organise into mRNPs is still poorly understood. Here, using a single molecule super-resolution microscopy approach to describe features of mRNP organisation in cells, our data shows that mRNA in cells are found at different levels of compaction depending on their subcellular localisation and translation state, with actively translating mRNAs and mRNAs sequestered to stress granules representing two extremes of open and compacted mRNAs states *in vivo*. Furthermore, we show that decompaction during translation results in the separation of the 5' and 3' ends of mRNAs, indicating that at least for the mRNAs investigated here, translation does not occur in a stable closed-loop conformation.

2.2.4.1 Nuclear mRNPs show a linear organisation

EM studies visualising the 35kb-long nuclear BR mRNPs show mRNPs assembled as compact particles with a croissant shape where 5' and 3' ends are in close proximity (Mehlin et al., 1995). The formation of this particle occurs sequentially and co-transcriptionally, starting with the formation of a rod-like structure with about a 12 nm diameter that further compacts into stalk and finally results in a croissant-shaped mRNP with a ~50 nm diameter and ~15 nm thickness. Considering a hypothetically fully extended, linear mRNA with a spacing between nucleotides of 0.59 nm, and the 50 nm diameter of the BR mRNPs as the maximal extension of the BR mRNP, BR mRNPs are compacted about ~413-fold. We also observe a high degree of compaction of about 111-fold for nuclear MDN1 mRNPs, considering a diameter of 97 nm (double the radius of gyration), suggesting mRNPs are generally highly compact in the nucleus.

However, in contrast to BR mRNAs, we do not observe the 5' and 3' ends in close proximity, but rather 5'-3' end further apart than the 5' to the middle region and middle to the 3', suggesting a more linear conformation of the nuclear mRNP (Figure 2-6). This structure could be the result of the sequential assembly of RNPs, such as the EJC to nascent mRNAs and the further compaction through binding to other proteins containing homo- and heterodimerisation domains, such as SR proteins, as suggested in (Singh et al., 2012). Indeed, a recently developed RNA ImmunoPrecipitation and Proximity Ligation in Tandem (RIPPLiT) approach investigating the proximity of different regions within mRNAs identified only local intramolecular contacts but

failed to observed long-range intramolecular mRNA interactions (Metkar et al., 2018). These observations suggest that mammalian nuclear mRNPs may be organised as rod-like structures, similar to the nuclear mRNPs previously purified from yeast (Batisse et al., 2009). Interestingly, we also observed a small fraction in nuclear *MDN1* mRNAs with a more open conformation. One possibility could be that these mRNAs are not fully spliced, although we view this explanation as unlikely because analysis of nuclear mRNA sequencing datasets from HEK293 cells does not suggest inefficiently spliced introns for *MDN1* (not shown) (Neve et al., 2016). Alternatively, if mRNPs assemble linearly, mediated by the binding of EJC and other RBPs, inefficient assembly of these complexes might result in more open mRNPs.

2.2.4.2 Variable levels of RNP compaction in cells

The compaction state of nuclear mRNA represents an intermediate state relative to the compacted and extended states observed for cytoplasmic mRNAs. Only a few examples of large RNP structures have been described that allow a direct comparison of the different levels of compaction found for cytoplasmic mRNAs. For instance, the 80S eukaryotic ribosome is a highly compact RNP with a diameter of about 30 nm and containing 7,216nt, resulting in an RNA compaction of ~142-fold. Nonetheless, this is less compact than *MDN1* mRNAs upon puromycin treatment, where we observe compaction of ~199-fold. Interestingly, the compaction of *MDN1* mRNA upon puromycin treatment is similar to packaged viruses. For instance, the ~7,500 nt RNA genome of the Hepatitis A virus is packed into a capsid with an inner diameter of about 22 nm, leading to a ~200-fold compaction of its genome (X. Wang et al., 2017), and the Zika genome (~11,000nt, 30 nm capsid inner diameter) gets similarly compacted (Sirohi et al., 2016). Interestingly, viral RNAs, when transcribed in vitro, were shown to acquire a condensed conformation, as measured using cryo-EM or SAXS, but the volume occupied by these RNAs *in vitro* is larger than when the RNA gets packaged into the viral capsid, consistent with the idea that compaction into the capsid is an active packaging mechanism (Gopal et al., 2012). Finally, a recent study showed that different *in vitro* transcribed mRNAs and lncRNA get compacted *in vitro* to a level similar, or sometimes even greater than rRNA (Borodavka et al., 2016). Together, these results suggest that different levels of RNA compaction *in vivo* are likely mediated by a combination of RNA sequence as well as associated proteins, and it will be interesting to determine whether the high level of mRNA compaction observed for mRNAs upon ribosome eviction or sequestration into stress granules, is

an active process that requires specific proteins, or whether it rather reflects the collapse of the RNA polymer onto itself due to the absence of ribosomes and other RBPs.

2.2.4.3 Closed-loop translation and regulation of gene expression

Our end-to-end measurements revealed that translating mRNAs rarely show co-localising 5' and 3' ends, and the sun-tag reporter data further suggest that separation of the ends increases as a function of ribosome occupancy. Similar results were observed in another study using a similar approach but different mRNAs, suggesting that open conformations of translating mRNAs are a widespread phenomenon (Khong and Parker, 2018). Thus, our results are seemingly at odds with the current view that translating mRNAs exist in a stable closed-loop conformation. One possibility is that the eIF4G–PABP interaction may be transient and only occurs during specific steps of the translation cycle. Recent studies have shown that translation of many transcripts occurs in a bursting pattern, and the variable ribosome occupancy during 'on' and 'off' times of translation bursts is likely to cause altered mRNA compaction (Bin Wu et al., 2016; Pichon et al., 2016; Yan et al., 2016). Translation bursting could therefore induce structural reorganisation of mRNAs that facilitate 5'-3' proximity during 'off' times, allowing transient eIF4G–PABP interactions to occur. Interestingly, *in vitro* transcribed mRNAs were shown to obtain conformations where the 5' and 3' end are close in space, which is also suggested using computational predictions (Lai et al., 2018; Leija-Martínez et al., 2014; Yoffe et al., 2011). It will be interesting to investigate whether this occurs for mRNAs *in vivo*, maybe as a result of translation bursting or as a result of translation inhibition in response to an external stimulus, and whether this will facilitate transient, eIF4G–PABP dependent, closed-loop configurations.

Closed-loop interactions could also occur during the pioneer round of translation. However, arguing against such a model is that pre-translation, EJC-containing mRNAs have a rod-like organisation where the 5' and 3' are not in proximity, making it unlikely that this mRNP would be able to acquire a closed-loop conformation without further reorganisation (Metkar et al., 2018). Nevertheless, mRNP reorganisation at the cytoplasmic side of the nuclear pore has been shown in yeast, and recent studies suggest two populations of EJC containing mRNAs, with a cytoplasmic EJC-mRNP fraction that contains far fewer proteins and therefore possibly a different architecture (Mabin et al., 2018; Tran et al., 2007).

An alternate possibility to bring ends together could be a long poly(A) tail. As it is not possible to design probes for the tail that do not hybridise to all polyadenylated RNAs, our probes targeting to the 3' of the RNA only hybridise up to the start of the poly(A) tail. However, it is unlikely that the tail is long enough to bridge a gap of up to 300 nm, even if fully extended, as recent TAIL-seq studies revealed that polyA tails in mammals are on average only 50-100nt long (Chang et al., 2014; Subtelny et al., 2014).

Alternatively, it may be that only a subset of mRNAs is translated in a closed-loop conformation. EM and cryo-ET have shown polysomes in various conformation, and only some of these conformations are compatible with a possible closed-loop conformation of the mRNAs (Brandt et al., 2010; Christensen et al., 1987; Christensen and Bourne, 1999). Interestingly, recent studies demonstrate that not all mRNAs are bound to the same extent by the closed-loop factors, supporting the idea that closed-loop formation might preferentially occur for some mRNA and/or during distinct phases of polysome assembly (Archer et al., 2015; Costello et al., 2015; Rissland et al., 2017; M. K. Thompson et al., 2016). A closed-loop configuration could also be more difficult to achieve for longer mRNAs where the ends could be separated by larger distances. Furthermore, it has been suggested that the formation of a closed-loop might be more complex than the interaction of eIF4G and PABPC1, that this interaction might not be a prerequisite for translation at all times, or that it can be mediated by additional factors. Indeed, non-polyadenylated mRNAs can associate with polysomes and produce proteins, and *S. cerevisiae* strains with impaired closed-loop components are viable, and we show there that mammalian cell lines limited phenotypes upon reduced eIF4G–PABPC1 interactions (Figure 2-3) (Park et al., 2011; Proweller and Butler, 1997; Wilusz et al., 2012). Together with our observations showing that ribosome occupancy results in a decompaction of the mRNA and separation of the ends, all these observations argue against a model where a stable closed-loop conformation can be considered as a universal state of translating mRNAs.

Finally, some of the strongest functional evidence for 5'–3' proximity comes from the numerous examples of regulatory elements in the 3'UTR that modulate processes at the 5' end, such as de-capping or translation repression or initiation (Fabian and Sonenberg, 2012; Rissland, 2016). Signal transmission from the 3' to the 5' likely requires the mRNP to be flexible to allow both ends to meet, and it is unclear whether this flexibility is possible when mRNAs are in polysomes, as we show that ribosome occupancy leads to the separation of the ends. In general, we have little

understanding of the biophysical properties of mRNPs *in vivo*. Obtaining a better mechanistic understanding of different aspects of mRNP metabolism involving intramolecular communication will therefore require a better understanding of the biophysical properties of RNPs in cells and will likely require new tools that allow us to study mRNP organisation *in* vivo, with single-molecule resolution and in real-time.

2.2.5 Acknowledgement

We thank members of the Zenklusen and Rissland laboratories, Marlene Oeffinger, Nicole Francis, Steve Michnick, Jeff Kieft and Julie Claycomb, for critical discussion and comments on the manuscript, and Melissa Moore, Job Dekker and Roy Parker for sharing data prior to publication. This work has been supported CIHR (Project Grant-366682), FRQ-S (Chercheur-boursier Junior 2), CFI (DZ), and the NIH (**R35GM128680) to OSR.**

2.2.6 Author contributions

SA, NL, SR, OSR, BW and DZ conceived the study; SA performed and analyzed the smFISH experiments, NL the SINAPs reporter FISH-IF experiments, and BN and OSR created and characterized CRISPR/Cas9 cell lines. OSR, BW and DZ supervised the work, and DZ wrote the paper with input from all authors.

2.2.7 References

Afonina, Z.A., Myasnikov, A.G., Shirokov, V.A., Klaholz, B.P., Spirin, A.S., 2015. Conformation transitions of eukaryotic polyribosomes during multi-round translation. Nucleic Acids Res 43, 618–628.

Archer, S.K., Shirokikh, N.E., Hallwirth, C.V., Beilharz, T.H., Preiss, T., 2015. "Probing the closed-loop model of mRNA translation in living cells." rnabiology 12, 248–254.

Batisse, J., Batisse, C., Budd, A., Böttcher, B., Hurt, E., 2009. Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. J Biol Chem 284, 34911–34917.

Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J., Topisirovic, I., 2015. Targeting the translation machinery in cancer. Nature Reviews Drug Discovery 14, 261–278.

Bin Wu, Eliscovich, C., Yoon, Y.J., Singer, R.H., 2016. Translation dynamics of single mRNAs in live cells and neurons. Science aaf1084.

Borodavka, A., Singaram, S.W., Stockley, P.G., Gelbart, W.M., Ben-Shaul, A., Tuma, R., 2016. Sizes of Long RNA Molecules Are Determined by the Branching Patterns of Their Secondary Structures. Biophysical Journal 111, 2077–2085.

Brandt, F., Carlson, L.-A., Hartl, F.U., Baumeister, W., Grünewald, K., 2010. The three-dimensional organization of polyribosomes in intact human cells. Mol Cell 39, 560–569.

Buchan, J.R., Parker, R., 2009. Eukaryotic stress granules: the ins and outs of translation. Mol Cell 36, 932–941.

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

Chang, H., Lim, J., Ha, M., Kim, V.N., 2014. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3′ End Modifications. Molecular Cell 53, 1044–1052.

Chen, H., Meisburger, S.P., Pabit, S.A., Sutton, J.L., Webb, W.W., Pollack, L., 2012. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. Proceedings of the National Academy of Sciences 109, 799–804.

Christensen, A.K., Bourne, C.M., 1999. Shape of large bound polysomes in cultured fibroblasts and thyroid epithelial cells. The Anatomical Record 255, 116–129.

Christensen, A.K., Kahn, L.E., Bourne, C.M., 1987. Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am. J. Anat. 178, 1–10.

Costello, J., Castelli, L.M., Rowe, W., Kershaw, C.J., Talavera, D., Mohammad-Qureshi, S.S., Sims, P.F.G., Grant, C.M., Pavitt, G.D., Hubbard, S.J., Ashe, M.P., 2015. Global mRNA selection mechanisms for translation initiation. Genome Biology 16, 10.

Fabian, M.R., Sonenberg, N., 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol 19, 586–593.

Floor, S.N., Doudna, J.A., Green, R., 2016. Tunable protein synthesis by transcript isoforms in human cells. eLife 5, e10921.

Gallie, D.R., 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev 5, 2108–2116.

Gopal, A., Zhou, Z.H., Knobler, C.M., Gelbart, W.M., 2012. Visualizing large RNA molecules in solution. RNA 18, 284–299.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F.P., Rissland, O.S., Durocher, D., Angers, S., Moffat, J., 2015. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163, 1515–1526.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., Brown, P.O., 2009. Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. PLoS Biol 7, e1000238.

Hentze, M.W., Castello, A., Schwarzl, T., Preiss, T., 2018. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol 19, 327–341.

Imataka, H., Gradi, A., Sonenberg, N., 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. The EMBO Journal 17, 7480–7489.

Jackson, R.J., Hellen, C., Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nature reviews Molecular cell ….

Jonas, S., Izaurralde, E., 2015. Towards a molecular understanding of microRNA-mediated gene silencing. Nature Reviews Genetics 16, 421–433.

Khong A, Matheny T, Jain S, Mitchell SF, Wheeler JR, Parker R. The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. Mol Cell 2017 Nov 16;68(4):808-820.e5

Khong, A., Parker, R., 2018. mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. bioRxiv 366690.

Lai, W.-J.C., Kayedkhordeh, M., Cornell, E.V., Farah, E., Bellaousov, S., Rietmeijer, R., Mathews, D.H., Ermolenko, D.N., 2018. The formation of intramolecular secondary structure brings mRNA ends in close proximity. bioRxiv 289496.

Le Hir, H., Izaurralde, E., Maquat, L.E., Moore, M.J., 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. The EMBO Journal 19, 6860–6869.

Leija-Martínez, N., Casas-Flores, S., Cadena-Nava, R.D., Roca, J.A., Mendez-Cabañas, J.A., Gomez, E., Ruiz-Garcia, J., 2014. The separation between the 5“-3” ends in long RNA molecules is short and nearly constant. Nucleic Acids Res 42, 13963–13968.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., de Turris, V., Lopez-Jones, M., Singer, R.H., 2011. A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods

Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I., Bustamante, C., 2001. Reversible Unfolding of Single RNA Molecules by Mechanical Force. Science 292, 733–737.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R., Chang, H.Y., 2016. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell 165, 1267–1279.

Mabin, J.W., Woodward, L.A., Patton, R., Yi, Z., Jia, M., Wysocki, V., Bundschuh, R., Singh, G., 2018. The exon junction complex undergoes a compositional switch that alters mRNP structure and nonsense-mediated mRNA decay activity. bioRxiv 355495. doi:10.1101/355495

Marchese, D., de Groot, N.S., Gotor, N.L., Livi, C.M., Tartaglia, G.G., 2016. Advances in the characterization of RNA-binding proteins. WIREs RNA 7, 793–810. doi:10.1002/wrna.1378

Mehlin, H., Daneholt, B., Skoglund, U., 1995. Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J Cell Biol 129, 1205–1216. doi:10.1083/jcb.129.5.1205

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., Moore, M.J., 2018. Higher-Order Organization Principles of Pre-translational mRNPs. bioRxiv 278747.

Mollet, S., Cougot, N., Wilczynska, A., Dautry, F., Kress, M., Bertrand, E., Weil, D., 2008. Translationally Repressed mRNA Transiently Cycles through Stress Granules during Stress. Mol Biol Cell 19, 4469–4479.

Müller-McNicoll, M., Neugebauer, K.M., 2013. How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nature Reviews Genetics 14, 275–287.

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., Zimmer, C., 2013. FISH-quant: automatic counting of transcripts in 3D FISH images. Nat Methods 10, 277–278.

Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M., Furger, A., 2016. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome Res 26, 24–35.

Panas, M.D., Ivanov, P., Anderson, P., 2016. Mechanistic insights into mammalian stress granule dynamics. J Cell Biol 215, 313–323.

Park, E.H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., Hinnebusch, A.G., 2011. Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. The EMBO Journal 30, 302–316.

Pichon, X., Bastide, A., Safieddine, A., Chouaib, R., Samacoits, A., Basyuk, E., Peter, M., Mueller, F., Bertrand, E., 2016. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. J Cell Biol 126, jcb.201605024.

Proweller, A., Butler, J.S., 1997. Ribosome concentration contributes to discrimination against poly(A)- mRNA during translation initiation in Saccharomyces cerevisiae. J Biol Chem 272, 6004–6010.

Rahman, S., Zorca, C.E., Traboulsi, T., Noutahi, E., Krause, M.R., Mader, S., Zenklusen, D., 2017. Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. Nucleic Acids Res 45, 3017–3030.

Rakheja, D., Chen, K.S., Liu, Y., Shukla, A.A., Schmid, V., Chang, T.-C., Khokhar, S., Wickiser, J.E., Karandikar, N.J., Malter, J.S., Mendell, J.T., Amatruda, J.F., 2014. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. Nat Comms 2, 4802.

Ramani, V., Qiu, R., Shendure, J., 2015. High-throughput determination of RNA structure by proximity ligation. Nat Biotechnol 33, 980–984.

Rech, J., 1995. An Ultrastructural Characterization of in Vitro-Assembled hnRNP C Protein-RNA Complexes. J. Struct. Biol. 114, 84–92.

Rissland, O.S., 2016. The organization and regulation of mRNA–protein complexes. WIREs RNA.

Rissland, O.S., Subtelny, A.O., Wang, M., Lugowski, A., Nicholson, B., Laver, J.D., Sidhu, S.S., Smibert, C.A., Lipshitz, H.D., Bartel, D.P., 2017. The influence of microRNAs and poly(A) tail length on endogenous mRNA–protein complexes. Genome Biol 18, 211.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., Moore, M.J., 2012. The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. Cell. doi:10.1016/j.cell.2012.10.007

Singh, G., Pratt, G., Yeo, G.W., Moore, M.J., 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annual Review of Biochemistry 84, 325–354.

Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T.C., Rossmann, M.G., Kuhn, R.J., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science 352, 467–470.

Skoglund, U., Andersson, K., Strandberg, B., Daneholt, B., 1986. Three-dimensional structure of a specific pre-messenger RNP particle established by electron microscope tomography. Nature 319, 560–564.

Souquere, S., Mollet, S., Kress, M., Dautry, F., Pierron, G., Weil, D., 2009. Unravelling the ultrastructure of stress granules and associated P-bodies in human cells. J Cell Sci 122, 3619–3626.

Steitz, J.A., 1969. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature 224, 957–964.

Strobel, E.J., Yu, A.M., Lucks, J.B., 2018. High-throughput determination of RNA structures. Nature Reviews Genetics 319, 1.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., Bartel, D.P., 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71.

Tarun, S.Z., Sachs, AB, 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. The EMBO Journal 15, 7168–7177. doi:10.1002/wrna.1430

Thompson, M.K., Rojas-Duran, M.F., Gangaramani, P., Gilbert, W.V., Hinnebusch, A.G., 2016. The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. eLife 5, e11154.

Thompson, R., Larson, D., Webb, W., 2002. Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82, 2775–2783.

Tran, E.J., Zhou, Y., Corbett, A.H., Wente, S.R., 2007. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. Mol Cell 28, 850–859.

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., MacInnes, A.W., Cuppen, E., Simonis, M., 2014. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol 15, R6.

Viero, G., Lunelli, L., Passerini, A., Bianchini, P., Gilbert, R.J., Bernabò, P., Tebaldi, T., Diaspro, A., Pederzolli, C., Quattrone, A., 2015. Three distinct ribosome assemblies modulated by translation are the building blocks of polysomes. J Cell Biol 208, 581–596.

Wang, C., Han, B., Zhou, R., Zhuang, X., 2016. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell 165, 990–1001.

Wang, X., Zhu, L., Dang, M., Hu, Z., Gao, Q., Yuan, S., Sun, Y., Zhang, B., Ren, J., Kotecha, A., Walter, T.S., Wang, J., Fry, E.E., Stuart, D.I., Rao, Z., 2017. Potent neutralization of hepatitis A virus reveals a receptor mimic mechanism and the receptor recognition site. Proceedings of the National Academy of Sciences of the United States of America 114, 770–775.

Wells, S., Hillner, P., Vale, R., Sachs, A., 1998. Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell 2, 135–140.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.-D., Joshua-Tor, L., Sharp, P.A., 2012. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. Genes Dev 26, 2392–2407.

Yan, X., Hoek, T.A., Vale, R.D., Tanenbaum, M.E., 2016. Dynamics of Translation of Single mRNA Molecules In Vivo. Cell 165, 976–989.

Yoffe, A.M., Prinsen, P., Gelbart, W.M., Ben-Shaul, A., 2011. The ends of a large RNA molecule are necessarily close. Nucleic Acids Res 39, 292–299.

Zenklusen, D., Larson, D.R., Singer, R.H., 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 15, 1263–1271.

2.2.8 Materials and Methods

Data and Software Availability

The uncompressed imaging files can be found using this link: https://data.mendeley.com/datasets/rjwfnvykd5/draft?a=28b31b35-4523-4cec-8dae-f39b57b44010

Reagents used, stock concentrations, working concentrations and treatment conditions

Puromycin dihydrochloride (Sigma P8833) – stock at 5 mg/ml in water, Cycloheximide (Sigma C7698-1G) – stock 5 mg/ml in ethanol, Sodium Arsenite (Sigma 35000-1L-R) – stock 50 mM in water, Homoharringtonine (Sigma SML1091-10MG) – stock 10mg/ml in DMSO. The drugs were diluted in warm media to get final working concentrations, and cells were treated prior to fixation as follows: Puromycin (100 µg/ml for 10 min), Cycloheximide (100 µg/ml for 10 min), Homoharringtonine – 100µg/ml for 10 min or 1hr and Sodium Arsenite (2 mM for 1 hr).

Cell culture and drug treatment

HEK293 (American Type Culture Collection CRL-1573) and U2OS osteosarcoma (American Type Culture Collection HTB-96) cell lines were maintained at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) (Wisent, 319-005-CL) supplemented with 10% fetal bovine serum (FBS) (Wisent, 080-150) and passaged every 2-3 days with Trypsin (Wisent 325-043-EL). Cells were plated on poly-L-Lysine (Sigma, P8920) coated coverslips the day before treatment and fixation. On the day of the experiment, media was replaced with fresh warm media containing drugs in indicated concentrations and placed back in the incubator. After treatment, the cells were briefly washed with 1xPBS, fixed with 4% paraformaldehyde in 1xPBS (pH 7.4) for 10 minutes at room temperature, washed three times with 1xPBS and stored overnight in 70% ethanol at -20ºC for permeabilisation. Alternatively, the cells were permeabilised using 0.1% TritonX-100 + 0.5%BSA in 1x PBS for 15 min, after which they were washed 2 times with 1X PBS for 5 min each immediately before using the samples for smFISH (Figure S2B).

Plasmid Preparation: The phage-ubc-flag-24xSunTag-Fluc-oxBFP-AID-baUTR-24xMS2 plasmid was prepared as described in (Bin Wu et al., 2016).

Generation and screening of eIF4G1 and PABPC1 mutant cell lines

Mutant cell lines were generated using CRISPR-Cas9. To produce sgRNAs targeting either eIF4G1 or PABPC1, annealed DNA oligos (Table S2-1) were ligated into the BbsI site of plasmid pX330 (Ran 2013). Homology repair constructs containing the intended mutations and upstream and downstream homology arms (~1 kb in total) were ligated into the plasmid Lox-Stop-Lox-TOPO-Δstop (Rakheja 2014), in which homology arms are cloned surrounding a puromycin resistance cassette flanked by loxP sites (Table S2-1).

HEK293 cells (5 x 105 cells in one well of a 6-well plate) were transfected with 250 ng of the pX330-sgRNA construct and 1 μg of the repair construct using Lipofectamine 2000 according to the manufacturer instructions, and then incubated in EMEM supplemented with 10% FBS in a humidified incubator at 37°C with 5% CO2. Two days following transfection, cells were trypsinised and 10% of the cells were moved into a 15-cm dish. After 24 h, puromycin was added to a final concentration of 3 μg/ml, and the media was changed daily for the next 3 days. The following day, single cells were seeded into each well of a 96-well plate on a MoFlo Astrios cell sorter (Beckman Coulter) at the Flow and Mass Cytometry Facility at SickKids Hospital, Toronto. Following expansion of single colonies, cells were harvested and screened by PCR using primers that anneal to the genome outside of the homology arm region (Table S2-1). To excise the puromycin resistance cassette from positive clones, the cells were transfected with 1μg of pgk-Cre (Rakheja et al., 2014) and incubated for 3 days before single-cell seeding, expansion, and screening for loss of the puromycin resistance gene by PCR as described above. The PCR products were analysed by Sanger sequencing to ensure that the intended mutations were present.

Cell viability assays

Cell viability was measured using PrestoBlue Cell Viability Reagent (Invitrogen) according to the manufacturer's instructions. Briefly, cells were seeded in triplicate in 96-well plates at 1000 cells per well in 90 μL of EMEM supplemented with 10% FBS, and then incubated at 37°C with 5% CO_2. At 24 h, 48 h, and 72 h after seeding, 10 μL of PrestoBlue reagent was added to each well. After a further 6.5-h incubation at 37°C with 5% CO_2, the fluorescence of each well was read on a SpectraMax M2 microplate reader (Molecular Devices).

Polysome profiling

To generate polysome profiles, cycloheximide was added to cells in a 10-cm dish to a final concentration of 100 µg/ml, and the cells were incubated for 10 min at 37°C. The cells were then placed on ice and washed twice with ice-cold PBS containing 100 µg/ml cycloheximide. Cells were lysed by shearing four times through a 26-gauge needle in 500 µL of lysis buffer (10 mM Tris-HCl pH 7.4, 5 mM MgCl2, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 500 U/ml Rnasin (Promega), EDTA-free protease inhibitor cocktail (Sigma), 100 µg/ml cycloheximide). Following centrifugation at 1300 × g for 10 min at 4°C, the supernatant was collected, flash-frozen in liquid nitrogen, and stored at -80°C until further processing.

Lysates were separated by loading 300 µL onto a 10-50% (w/v) sucrose gradient prepared with a Gradient Master (BioComp Instruments) and centrifuged for 2 h at 36,000 rpm in a SW41Ti rotor (Beckman Coulter) at 4°C. Gradients were fractionated on a Piston Gradient Fractionator (BioComp) coupled to an EM-1 Econo UV detector (Bio-Rad). UV profile data were recorded using Gradient Profiler software v 2.07 (BioComp).

smRNA FISH

Custom DNA probe sets were designed using Stellaris® Probe Designer, synthesised by Biosearch Technologies containing 3' amine-reactive group and labelled with far-red dye Cy5 (GEPA25001), red dyes Cy3 (GEPA23001) from Sigma or Dylight 550 (Thermo Scientific 62263) or green dye Dy488 (Thermo Scientific 46403) as described in (Rahman et al., 2017). For the mRNAs and the lncRNAs, the isoforms used to design the probes are mentioned in Figure S1. For the mRNAs, these isoforms were verified as the predominantly expressed transcripts in HEK293 using RNA-seq datasets from human protein atlas. For the lncRNAs, the probes were designed such that they hybridise to the longer isoforms. Probe sequences are shown in Table S2-2. Probe combinations used areshown in Table S2-3 and the probe combinations used for the experiment is mentioned in the figure legends. smFISH was done as described in (Rahman et al., 2017). Prior to hybridisation, cells were rehydrated in 1xPBS, then washed with 10% formamide/2xSSC for 10 minutes at room temperature. The cells were hybridised with 10-20 ng of each probe mix plus 40 µg of ssDNA/tRNA resuspended in the hybridisation solution (10% dextran sulfate/10% formamide/2xSSC/2 mM VRC/0.1 mg/ml BSA) for 3 hrs in the dark at 37ºC. Post hybridisation washes (2x 30 min) were carried out at 37ºC with 10% formamide/2xSSC. Samples were then

rinsed with 1xPBS and mounted with ProLong Gold antifade reagent with DAPI (P36935, Invitrogen).

Image Acquisition and pixel shift correction

Images were acquired with a 63x NA 1.46 oil objective on a Zeiss Elyra PS.1 system equipped with an Andor EMCCD iXon3 DU-885 CSO VP461 camera (1004x1002 pixels), the following filter sets: DAPI: BP420-480 + LP750 (Zeiss SR cube 07), Cy2: BP495-590+LP750 (Zeiss SR cube 13), Cy3: LP570 (Zeiss SR cube 14), Cy5: LP655 (Zeiss SR cube 10) and the following lasers: 50 mW 405 nm HR diode, 100 mW 488 nm HR diode, 100 mW 561 nm HR DPSS, 150 mW 642 nm HR diode. Each image was acquired using 3 rotations and a grid size of 42 µm for all channels. The microscope was located in a temperature-controlled room and samples were kept in the room for at least an hr before imaging to minimize thermal fluctuations. To correct for pixel shifts between channels, 0.1 µm TetraSpec beads (Invitrogen T-7279) were imaged in all channels, and the channel shift values and chromatic aberration were calculated and corrected using the built-in channel alignment tool in ZEN 2012 SP5, which uses an affine image alignment algorithm and later applied to the images. This correction was calculated for each day of imaging.

Combined smFISH and Immunofluorescence for simultaneous detection of mRNA conformation and nascent translation

Human U2OS osteosarcoma cell line (American Type Culture Collection HTB-96) expressing stdMCP-Halotag , phR-scFV-GCN4-sfGFP-GB1-NLS-dWPRE, and pBabe-TIR1-9myc was prepared as described in (Bin Wu et al., 2016). Single-molecule FISH immunofluorescence was performed as described in (Bin Wu et al., 2016). In brief, cells were plated on 18mm diameter, 0.13mm thick collagen-coated coverslips (Fisher) in a 12-well dish. Cells were then transfected with 250 ng of the phage-ubc-flag-24xSunTag-Fluc-oxBFP-AID-baUTR-24xMS2 construct using X-tremeGENE 9 transfection reagent (XTG9-RO ROCHE). Six hrs after transfection, IAA (Sigma-Aldrich) was added to a final concentration of 250 µM. 20 hrs after transfection, fresh IAA was added to a final concentration of 250 µM. 24 hrs after transfection, cells were fixed for 10 minutes in PBS + 5 mM $MgCl_2$ (PBSM), permeabilised for 15 minutes in PBSM + 0.1% Triton-X and 0.5 % BSA, and incubated with 100 nM MS2v5-Cy5 and 50 nM SunTagV4-Qusar 570 smFISH probe sets (Table S2-2) and primary antibody against GFP (GFP-1010, Aves labs, Inc.)

and incubated for three hrs at 37°C. After washing, cells were incubated with Alexa Fluor 488 labeled secondary antibody (ThermoFischer) and mounted in ProLong Diamond antifade reagent with DAPI (Life Technologies). Images were acquired on a custom inverted wide-field Nikon Eclipse Ti-E microscope equipped with three Andor iXon DU897 EMCCD cameras (512x512 pixels) , Apochromatic TIRF 100X Oil Immersion Objective Lens/1.49 NA (Nikon MRD01991), encoded Stage with 150 micron Piezo Z (ASI), and LU-n4 four laser unit with solid-state 405 nm, 488 nm, 561 nm, and 640 nm lasers (Nikon), a TRF89901-EM ET-405/488/561/640 nm Laser Quad Band Filter Set for TIRF applications (Chroma), and Nikon H-TIRF system. Images were acquired using in-unit intermediate 1.5x magnification changer for a final magnification of 150x and independent epi-illumination from the 488, 561, and 640 nm lasers. Image pixel size: XY, 106.7 nm; Z-step, 200 nm. A total of 29 cells without drug treatment (total of individual 396 mRNAs) and 40 cells (97 individual mRNAs) upon puromycin treatment were analysed.

Immunoprecipitations and Western blotting

Cells were washed with 1X PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4, 1.47 mM KH2PO4, pH 7.4) and then lysed with 1 ml ice-cold lysis buffer A (100 mM KCl, 0.1 mM EDTA, 20 mM Hepes, pH 7.6, 0.4% NP-40, 10% glycerol, with freshly added 1 mM DTT and complete mini EDTA-free protease inhibitors [Roche; one tablet per 25 ml lysis buffer]) per 2.5 million cells. 50 µl was saved as the input sample. Cells were incubated with antibody (diluted according to manufacturer's instructions) for 2 hrs, rotating at 4°C. α-PABPC1 antibody was purchased from Abcam (ab21060), and α-eIF4G1 from MBL International. EZ view protein G Sepharose (Sigma) was washed twice with lysis buffer and added to lysate with 40 µl slurry used per ml of lysate. The beads and lysate were incubated with the lysate for an additional hr, rotating at 4°C. The beads were washed 3X with cold lysis buffer. After the first wash, the beads were transferred to a new tube. The beads were then resuspended in protein loading dye (Life Technologies) with freshly added reducing agent, according to manufacturer's instructions, and boiled for 3 min. 2% lysate and 10% immunoprecipitants were loaded onto an SDS-PAGE gel and probed for PABPC1 and eIF4G1. α-PABPC1 and α-eIF4G1 were used at 1:1000, and α-rabbit IgG HRP (at 1:10,000) was used as the secondary antibody.

RNA spot detection, spot assignment and distance measurements

For image analysis, 3D datasets were reduced to 2D data using maximum projections in FiJi. Spot detection was done by 2D Gaussian fitting as described in (Thompson et al., 2002; Zenklusen et al., 2008). For 3D analysis, the spots were detected using AIRLOCALIZE as described in (Lionnet et al., 2011). To separate cytoplasmic and nuclear mRNPs, masks were created in FiJi by manual segmentation using DAPI stained nuclei as a reference, while ensuring that regions with overlapping spots within the same channel were not included. Assignment of the 5', 3' and/or the mid spots to either the cytoplasmic or the nuclear masks was done using MATLAB (MathWorks). To measure distances between different regions of mRNPs, spots from different channels were first grouped to assign neighbouring spots corresponding a single RNA. This was achieved by using spots from one channel as a reference and finding spots from the other channels within a defined radius using the coordinates from 2D Gaussian fitting or 3D Gaussian fitting using a custom MATLAB script. 300 nm for 2D analysis and 400 nm for 3D analysis were chosen as radii to limit assigning signals from neighbouring RNAs. These values were chosen as we observed very few RNAs with distances larger than these thresholds. Moreover, a threshold was required to ensure that there was no wrongful assignment of the signals. Groups with more than one spot from each channel, which could correspond to overlapping mRNPs or mRNPs close together in space, were discarded. For 2 colour imaging, the 5' signal was taken as reference and for 3 colour imaging, the middle was taken as reference. Switching references yielded comparable results (not shown). 2D or 3D distances between different regions of the mRNPs were then calculated for each signal within a group.

Combined smFISH and Immunofluorescence Data Analysis

All image analysis was performed using existing or custom build packages in MatLab (MathWorks). Gaussian fitting of smFISH and immunofluorescence spot intensities was performed using FISH-quant (Mueller et al., 2013). Briefly, cytoplasmic FISH spots were fit to a 3D Gaussian to determine the mRNA and translation site coordinates in each colour. Both 5'-end, 3'-end, and translation site intensities were detected independently by this method. Image registration was performed by imaging 100 nm TetraSpeck Microspheres (ThermoFisher) and calibrating the field correction based on an affine transformation in a custom-built MatLab package. The transformation matrix was first verified for reproducibility on other microsphere samples and then applied to mRNA samples (data not shown). Only 2D distances were considered

for this analysis. To determine the end-to-end mRNA distance, we first assigned the Quasar 570 channel (SunTag Probes) to FITC channel (Alexa 488 labelled translation site) by setting a colocalisation threshold of 300 nm after image correction. We then assigned the Quasar 570 to Cy5 (MS2 Probes), again with a colocalisation threshold of 300 nm. We first grouped mRNA with both Cy3 and Cy5 colocalisation and then determined if there was also a colocalised translation site signal. We then binned two-colour mRNA based on the presence (translating) or absence (non-translating) of the translation site signal. We then determined the end-to-end distance and, in the case of the translating mRNAs, the associated translation site intensity.

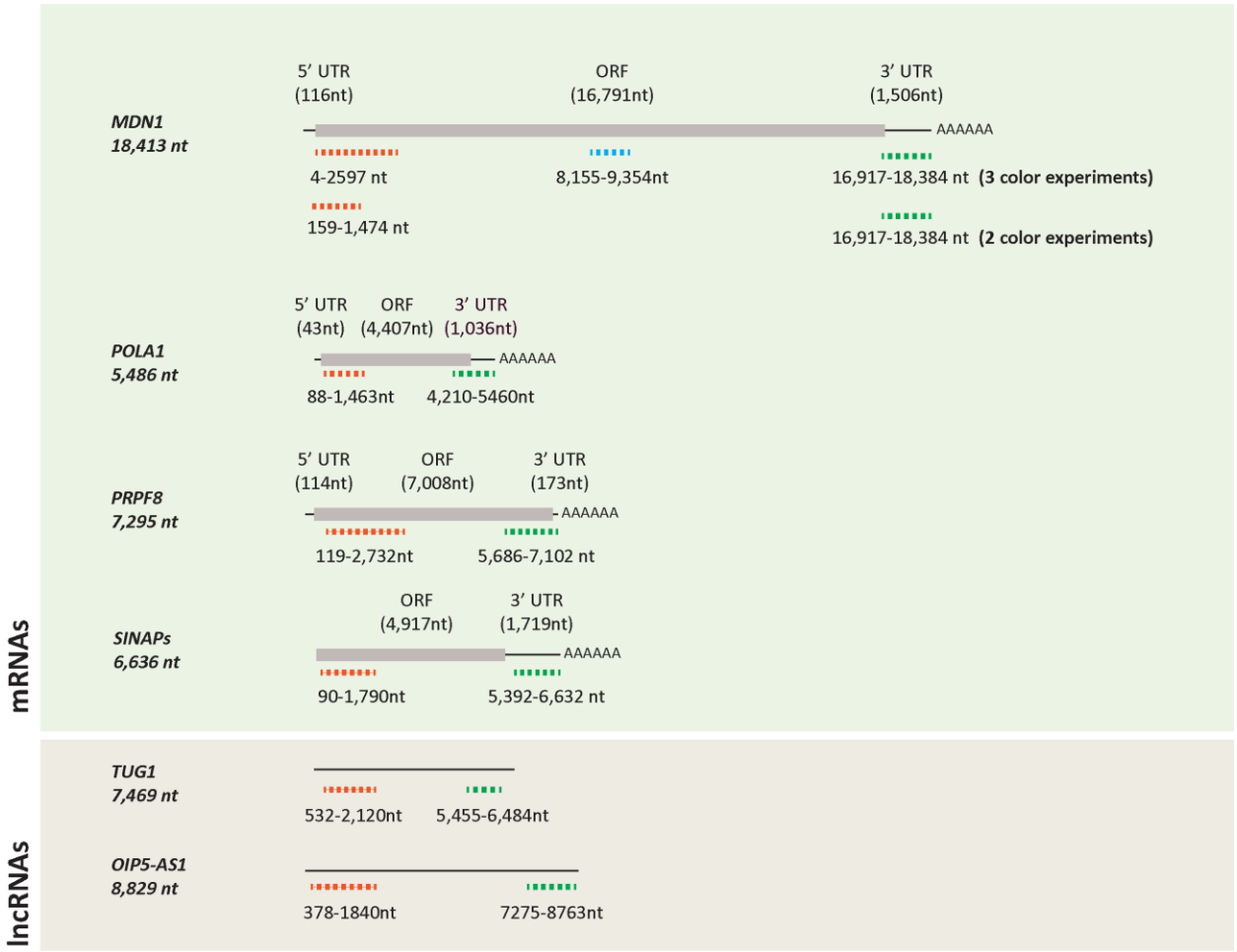
Data Plotting

All measurements were made for at least 2 independent biological replicates and the data plotted are representative from one of the replicates. For each measurement, at least 5 different fields were imaged, with each image containing a minimum of 10 cells to make a total of at least 50 cells. For the smFISH plots, a minimum of 593 RNAs was considered for cytoplasmic plots and a minimum of 430 RNAs were considered for the nuclear plots for data from HEK293 cells and a minimum of 308 RNAs were considered for data from U2OS cells, unless mentioned otherwise. For the FISH-IF plots, a total of 323 data points for translating, 97 for puromycin and 73 for non-translating were considered. The translating mRNAs were clustered using the k-means algorithm in R according to the intensity of the site of translation. After clustering, the four groups contained 64, 115, 104 and 40 RNAs from lower to higher intensity. The p-values were calculated using the Kolmogorov-Smirnov test in R for the data points plotted. The centre of mass plots in Figure 2-1G, 2D, 6D were made using R. The centre of mass was calculated as the mean of the coordinates of the three regions. The different conformations were then aligned using their centre of masses. For the 3-colour scatter plot in Figure 2-2E, 2-6E, Supplementary Figure 2-5 and Supplementary Figure 2-7B, to get a pair of co-localization precision values, two values were chosen randomly from our data. These values were taken as the X and Y coordinates for the scatter plot. The values that served as the X and Y coordinates were used to get density plots in the same figure. The mean Radius of gyration ($<R_g>$) was calculated using:

$$< R_g >= \sqrt{\frac{1}{3}\sum_{k=1}^{3}(r_k\text{-}r_{mean})^2}$$

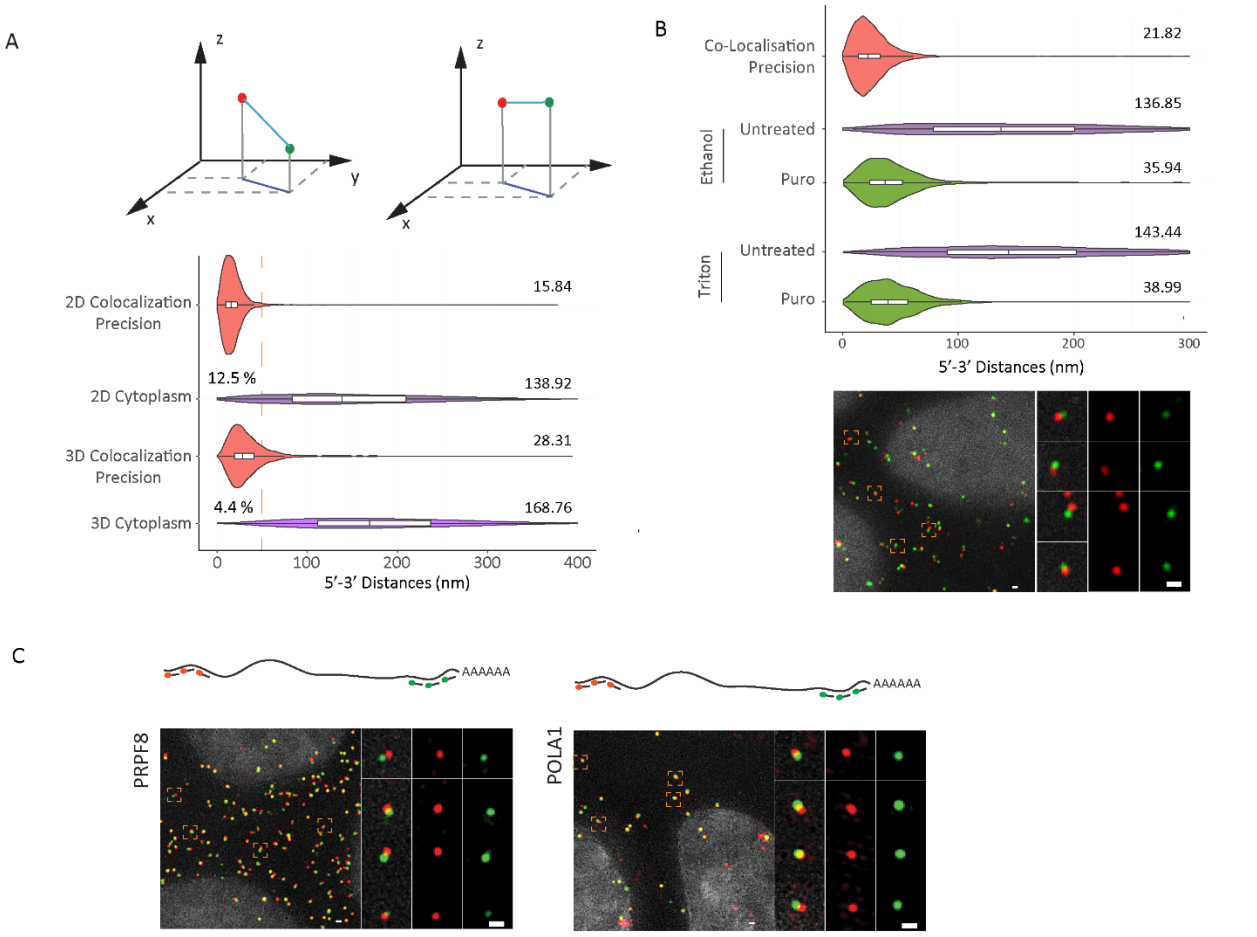
where k represents one of the three regions of the mRNP and r_k the position of the corresponding position in space as determined by 2D Gaussian fitting.

2.2.9 Supplementary Figures



Supplementary Figure 2-1: Positions of smFISH probes used in this study.

Cartoons illustrating the positions of the probes used for the different genes used. See Table S2-2 for probe sequences. The transcripts sequences were obtained from ensembl – MDN1 (ENST00000369393), POLA1 (ENST00000379068), PRPF8 (ENST00000304992), TUG1 (ENST00000644773) and OIP5-AS1 (ENST00000500949).

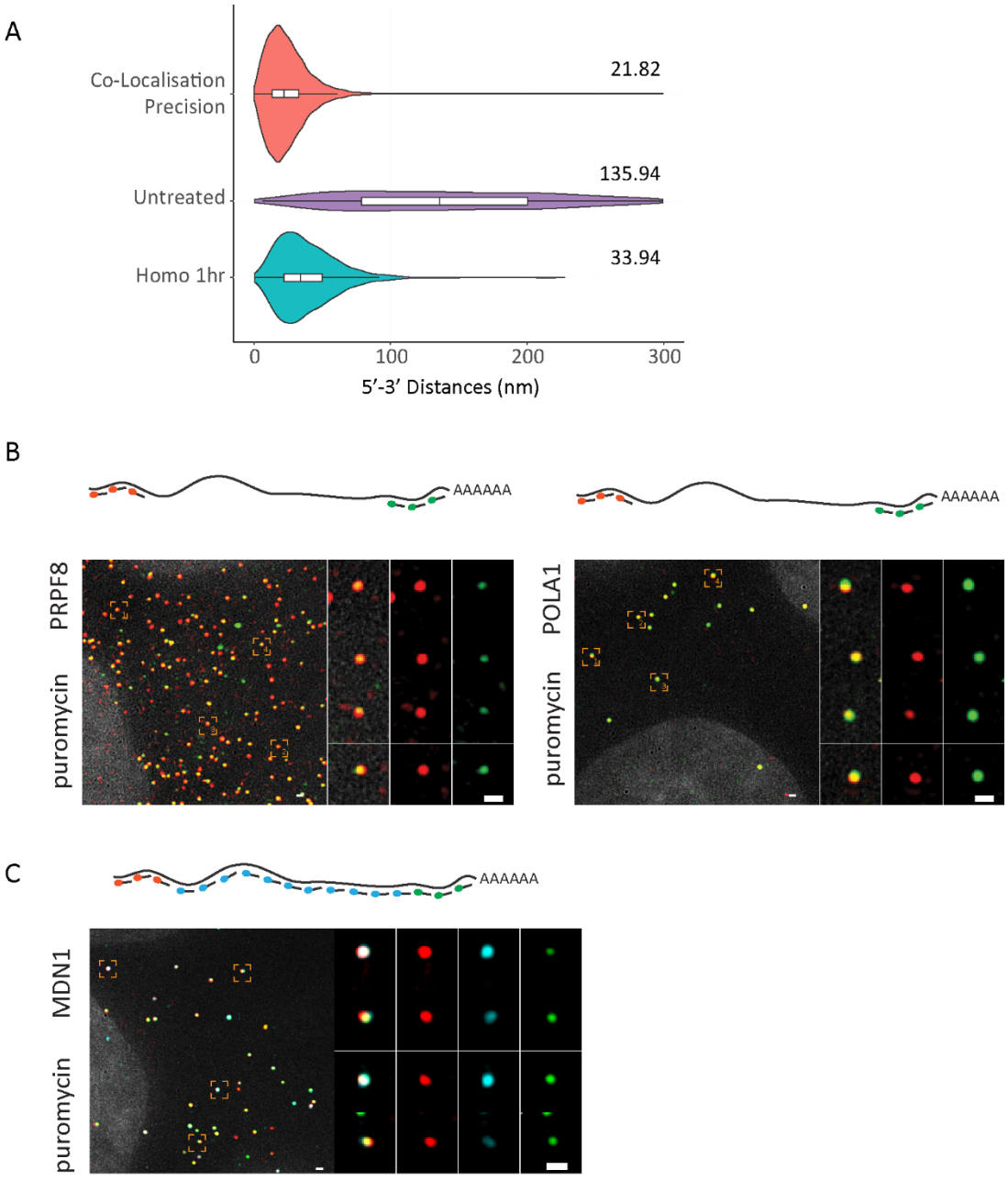


Adivarahan et al. Figure S2

Supplementary Figure 2-2: Visualising mRNP conformations for MDN1, POLA1 and PRPF8 and measurement of 5'-3' distances for MDN1 in 2D, 3D and in cells permeabilized using either TritonX or Ethanol

(A) *Cartoon illustrating how 2D projection alters 5'-3' distances measured (above) and violin plots showing distance distribution of co-localization precision and 5'-3' distances for MDN1 mRNAs calculated in 2D and 3D. Dotted line delineates the percentage of MDN1 mRNA with 5'-3' distances less than 50 nm.* ***(B)*** *Violin plots showing distance distribution of co-localization precision (Probe Set#1, Table S2-3) and 5'-3' distances for MDN1 mRNAs (Probe Set#2, Table S2-3) in cells permeabilised with either Ethanol or TritonX-100 and smFISH images using hybridising to the 5' and 3' ends of MDN1 in cells permeabilised with TritonX-100,* ***(C)****. smFISH images using probes hybridising to the 5' and 3' ends of PRPF8 (Probe Set#5, Table S2-3) and POLA1(Probe Set#6, Table S2-3) mRNAs in paraformaldehyde fixed HEK293 cells. Scale bars,*

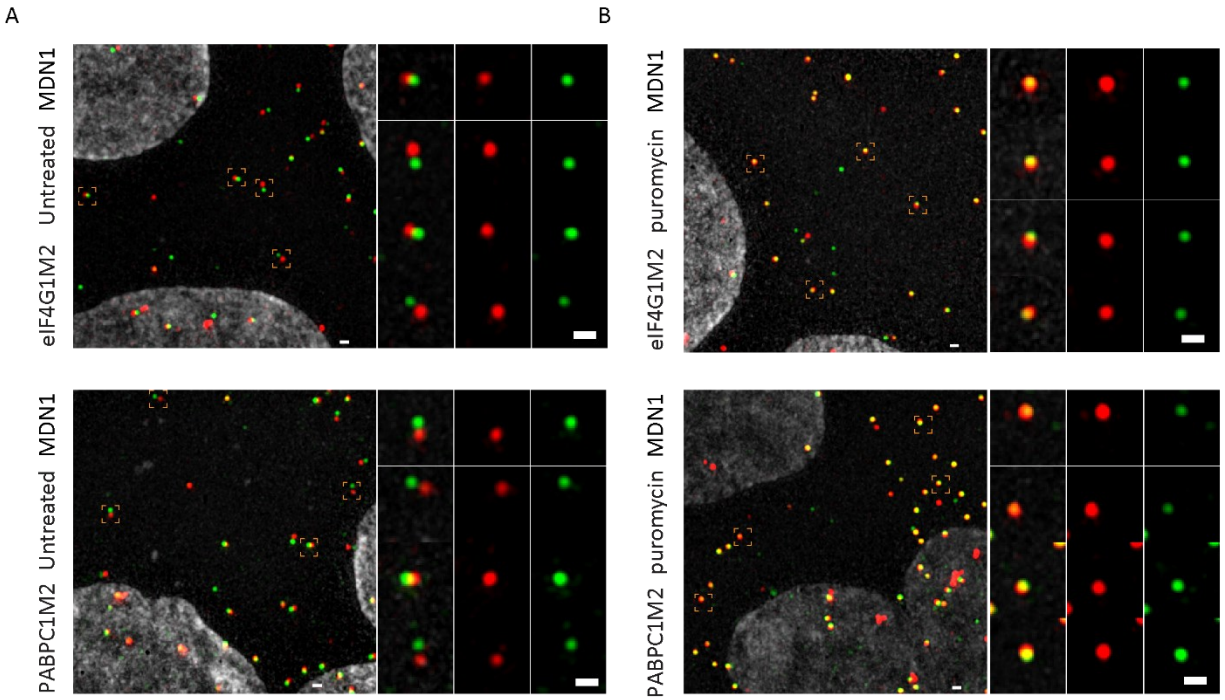
500 nm. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances are shown on the right.



Adivarahan et al. Figure S3

Supplementary Figure 2-3: 5'-3' distance measurements for MDN1 mRNA upon treatment with homoharringtonine and visualising mRNP conformation of single POLA1, PRPF8 and MDN1 mRNAs when treated with Puromycin.

(A) *Violin plots showing distance distribution of co-localisation precision (Probe Set#1, Table S2-3) and 5'-3' distances for MDN1 mRNAs (Probe Set#2, Table S2-3) determined by Gaussian fitting from untreated and homoharringtonine (1hr) treated cells* ***(B)*** *smFISH images using probes hybridising to the 5' and 3' ends of PRPF8 (Probe Set#5, Table S2-3) and POLA1 (Probe Set#6, Table S2-3) mRNAs in paraformaldehyde fixed HEK293 cells treated with puromycin (10 min, 100 µg/ml)* ***(C)*** *smFISH using 5' (red), 3' (green), and tiling (cyan) for MDN1 mRNA (Probe Set#3, Table S2-3) in paraformaldehyde fixed HEK293 cells treated with puromycin (10 min, 100 µg/ml). Nuclei are visualised by DAPI staining (grey). Magnified images of individual RNAs marked by dashed squares are shown on the right. Schematic position of probes shown on top. Scale bars, 500 nm. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances are shown on the right.*

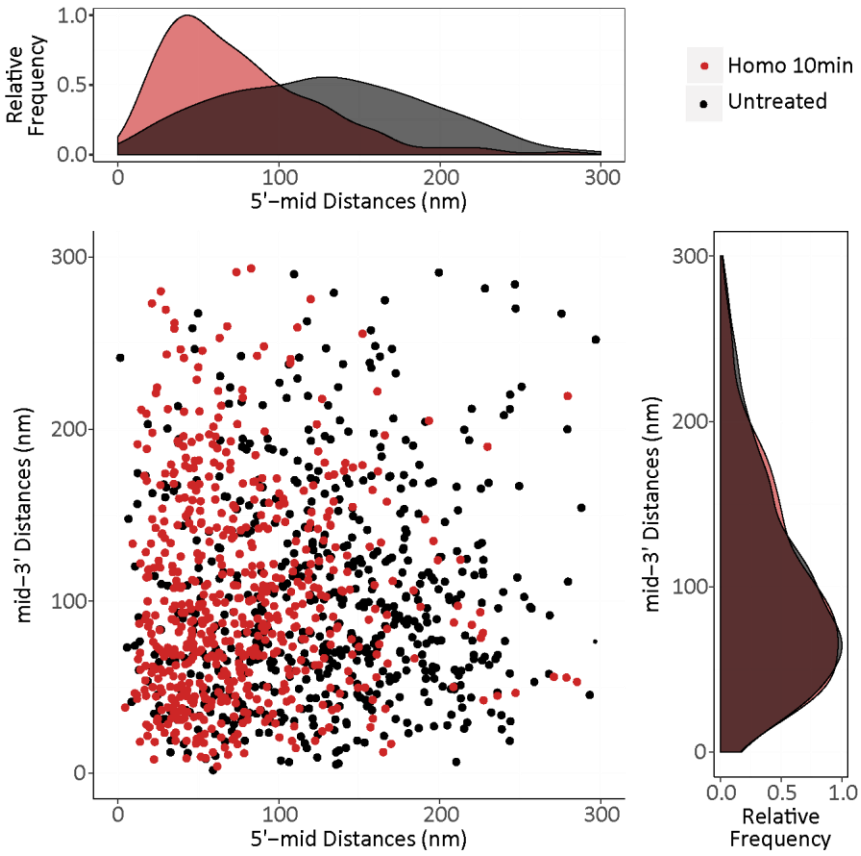


Adivarahan et al. Figure S4

Supplementary Figure 2-4: Visualising mRNP conformations in mutant cell lines of PABPC1 and eIF4G1.

(A) *smFISH images using probes hybridising to the 5' and 3' ends (Probe Set#2, Table S2-3) of MDN1 in mutant cell lines of eIF4G1 and PABPC1 in paraformaldehyde-fixed HEK 293 cells.* ***(B)***

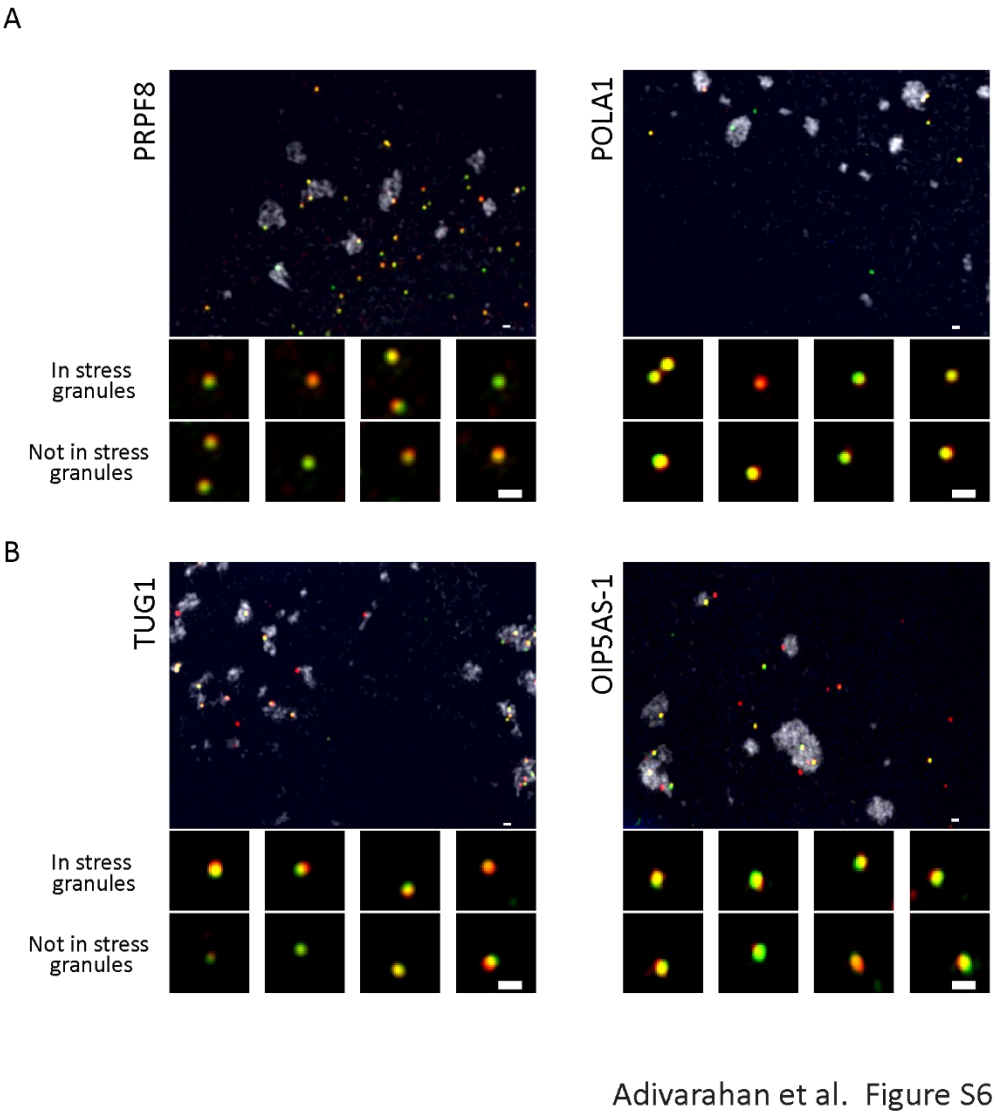
smFISH images using probes hybridising to the 5' and 3' ends of MDN1 in mutant cell lines of eIF4G1 and PABPC1 in HEK 293 cells treated with puromycin (10 min, 100 µg/ml). Scale bars, 500 nm.



Adivarahan et al. Figure S5

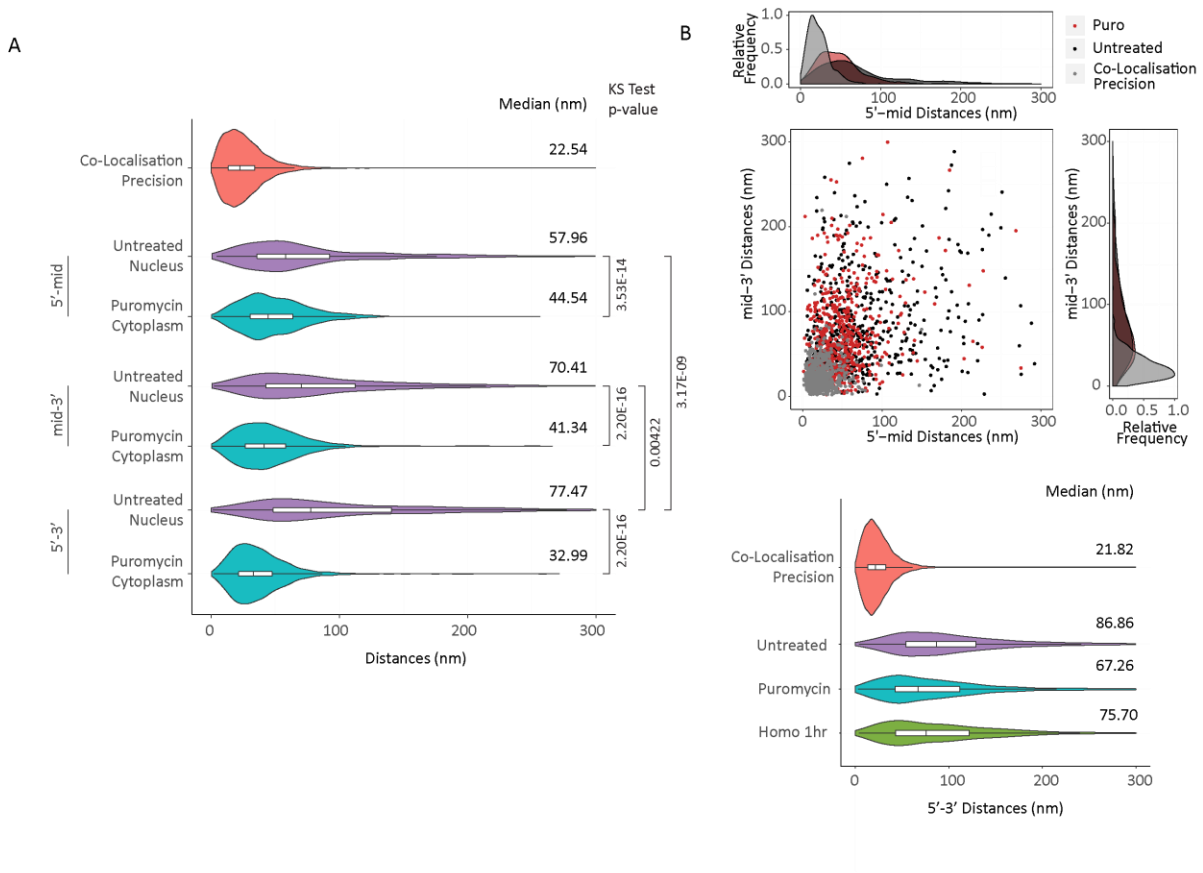
Supplementary Figure 2-5: Compaction of the 5' end is altered upon a pulsed homoharringtonine treatment for 10min.

Scatter plot showing 5'- mid and mid-3' distances for individual cytoplasmic MDN1 mRNAs from untreated cells (black) and cells treated with homoharringtonine (100µg/ml, 10min) (red). Frequency distribution are shown on top and on the right.



Supplementary Figure 2-6: mRNA and lncRNA compaction and accumulation in stress granules.

smFISH visualising 5' and 3' ends of PRPF8 (Probe Set#11, Table S2-3) and POLA1 (Probe Set#12, Table S2-3) mRNAs (A) or TUG1 (Probe Set#13, Table S2-3) and OIP5-AS1(Probe Set#14, Table S2-3) lncRNAs (B) in U2OS cells treated with arsenite (1-hr, 2 mM). Only a selected cytoplasmic region of cells is shown. Stress granules are visualised using an oligo dT probe (grey). Magnified images of individual RNAs localised inside or outside of stress granules are shown on the bottom of the images. For POLA1 and OIP5-AS1, not all magnified single RNAs shown in the bottom are from the corresponding image above. Scale bars, 500 nm.



Adivarahan et al. Figure S7

Supplementary Figure 2-7: Compaction of nuclear MDN1 mRNA upon puromycin or homoharringtonine treatment.

(A) *Violin plots showing distance distribution of co-localization precision (Probe Set#1, Table S2-3) and 5'-3' distances for MDN1 mRNAs (Probe Set#2, Table S2-3) determined by Gaussian fitting from nuclear mRNAs in untreated cells and cytoplasmic mRNAs in cells treated with puromycin (100µg/ml, 10 min) (Probe Set#4, Table S2-3).* ***(B)*** *Scatter plot showing 5'- mid and mid-3' distances for individual nuclear MDN1 mRNAs from untreated cells (black) and cells treated with puromycin (100µg/ml, 10min) (red). Co-localization precision is shown in grey. Frequency distribution are shown on top and on the right.* ***(C)*** *Violin plots showing distance distribution of co-localization precision (Probe Set#1, Table S2-3) and 5'-3' distances for MDN1 (Probe Set#2, Table S2-3) mRNAs determined by Gaussian fitting from untreated, puromycin (100µg/ml, 10min) or homoharringtonine (100µg/ml, 1 hr) treated HEK293 cells. White box plot inside the violin plot shows first quartile, median and third quartile. Median distances and p-values calculated using Kolmogorov-Smirnov test are shown on the right.*

2.2.10 Supplementary Tables

Table S2-1: Primers used for making CRISPR/Cas9 cells lines

Name	Sequence (5'-3')	Purpose
PABPC1 M161A gRNA 1 Fw	CACCGAATCTGTTAGCCATCTAACC	Guide RNA targeting PABPC1; annealed and ligated into pX330
PABPC1 M161A gRNA 1 Rev	AAACGGTTAGATGGCTAACAGATTC	
PABPC1 region Fw NotI	GCGCACTAGCGGCCGCGAGGAAGCGTTCAACTGTGA	To amplify 5' arm region of PABPC1
PABPC1 Not I 5' arm Rev guide 1 XhoI	GCGCACTAGCGGCCGCCTCGAGACCTGGATATTTGTGAAATAAAG	
PABPC1 BamHI 3' arm Fw guide 1	CGCGGATCCTAGATGGCTAACAGATTGTCTCTC	To amplify 3' arm region of PABPC1
PABPC1 region Rev BamHI	CGCGGATCCTTGGTCAGGCTGGTCTCAAA	
PABPC1 mutation Fw	GAAAGAGCTATTGAAAAAATGAATGGAGCGCTCCTAAATGATCGCAAAGTATTTGTTGG	Internal primers for stitch PCR to make PABPC1 mutation
PABPC1 mutation Rev	CCAACAAATACTTTGCGATCATTTAGGAGCGCTCCATTCATTTTTTCAATAGCTCTTTC	
PABPC1 region Fw	GGCGAGAGATTGCGTCAAGAA	Screening for puromycin cassette insertion/excision
PABPC1 region Rev	CCCTGGTAACAGGCATTTGTGAG	
PABP seq Fw1	GCAATATGGAATTCTTTTATATG	To sequence PCR products from PABPC1 mutant cell line screening
PABP seq Fw2	GGAACTGTGCAGTAATGGATATC	
PABP seq Rev1	CAATCTTGTCGCCCAGACTGG	
eIF4G1 mutant gRNA 1 Fw	CACCGTGCTGCTGGGACATTGTGC	Guide RNA targeting eIF4G1; annealed and ligated into pX330

eIF4G1 mutant gRNA 1 Rev	AAACGCACAATGTCCCAGCAGCAC	
eIF4G1 5' HR Arm Fw NotI	GCGCACTAGCGGCCGC GAGACAGGAACTAGACTCAAG	To amplify 5' arm region of eIF4G1
eIF4G1 5' HR Arm Rev NotI XhoI	GCGCACTAGCGGCCGC CTCGAG CAATGTCCCAGCAGCACCTGACC	
eIF4G1 3' HR Arm Fw BamHI	CGC GGATCC TGCCGGAAAGAGCAGTGACTTG	To amplify 3' arm region of eIF4G1
eIF4G1 3' HR Arm Rev BamHI	CGC GGATCC GGCACCCTATTCTGGGCACC	
eIF4G1 mutation Fw	TGCAGCCGCTGCCGCGGGAGCAATCTGGGGTGGCTGGTTC	Internal primers for stitch PCR to make eIF4G1 mutation
eIF4G1 mutation Rev	CGCCAGTGGGAAACTGCTGCACCCCTTGGGCTGGATAGTAGG	
eIF4G1 region Fw	GTAGTCGCACAGTCTTGGCTC	Screening for puromycin cassette insertion/excision
eIF4G1 region Rev	GAGTCCAGGGCAGAACAGAC	
4G seq Rev1	CACCCCTCGTAGGCAGGCACTC	To sequence PCR products from eIF4G mutant cell line screening
4G seq Fw 1	CAGAGTATGTGTGTACATGTTG	
4G seq Rev2	CTTCCTCGCTAGGCACTTCAG	
4G seq Rev 3	CCAGCAGTCCCCAAGTCAGTGG	

Table S2-2: List of smFISH probes used

MDN1 5'	tcgttcttggctgcgattaa	POLA1 3'	tgtacagggacttgtcagaa
MDN1 5'	taaggtactcaggacacact	POLA1 3'	taccggtaaaagcacagctg
MDN1 5'	cacagtacagtccttatcca	POLA1 3'	gtgcacactccgcatcaaaa
MDN1 5'	agcaaatccaaaaggagagg	POLA1 3'	tctcatgatcggtagtaagt
MDN1 5'	ttgaaagactggggatgtgt	POLA1 3'	ctgtagtcctgcagaacttt
MDN1 5'	gcatctgaactctctaggaa	POLA1 3'	ctctgctgtgttcttgagtt
MDN1 5'	gtgctcttcattcatacagg	POLA1 3'	tagccacttcgggacaagaa
MDN1 5'	cctcaacctgaaatggatca	POLA1 3'	tttgctcagattcacttcgg
MDN1 5'	aaaaccaaggccttctccaa	POLA1 3'	ttaggatttcacggcacaac
MDN1 5'	aaagggagacttctggattg	POLA1 3'	cttggttactcctgggattc
MDN1 5'	tcagacgaaacaagatgtcc	POLA1 3'	ggaagctgggatttttcaac
MDN1 5'	accagcacataagacctaag	POLA1 3'	caagggagaaacagatgctg
MDN1 5'	gaagacttttgcagacagac	POLA1 3'	acacaaacatgagacacagt
MDN1 5'	acagcattctgagaagcaac	POLA1 3'	gactcaacatttttgcagcc
MDN1 5'	tcctattggtccttccaaca	POLA1 3'	ctcagaaaccgggtcttcag
MDN1 5'	cctgtcactgcagctaaata	POLA1 3'	gctactctcaatccaagtag
MDN1 5'	aagctggactttgagaagct	POLA1 3'	gcctggggtcacttacattc
MDN1 5'	acatctgtgcagcgatacat	POLA1 3'	cataggctaaaggccctgag
MDN1 5'	atatcctccagaaggatcca	POLA1 3'	ttcagtcaggctctgagaag
MDN1 5'	aagagctctccattctccaa	POLA1 3'	tgaaaaagcaaaacgtcagc
MDN1 5'	aaatccaggtgccactttca	POLA1 3'	ttagaccgggtttaattggc
MDN1 5'	caactacgcccgcaggaaag	POLA1 3'	actcctggatggctggagaa
MDN1 5'	agcaagaagtgctccatgac	POLA1 3'	agacaagactgaaaaggaca
MDN1 5'	caagaacctgcccaactcac	POLA1 3'	agtgcaaggcttctaaatct
MDN1 5'	cgatcttgaggtgtccacac	POLA1 3'	gagcaattcaacaaacaagc
MDN1 5'	aatggcttcggcattccttt	POLA1 3'	cagtgtgtgtctgttggact
MDN1 5'	gttcatgcagatcatggttg	POLA1 3'	atgtgagtgtaaaacacctg
MDN1 5'	gtttgctcatcgacacacat	POLA1 3'	gcactttctatttaaggggc
MDN1 5'	gacatcaggatggttaccaa	POLA1 3'	ccctacacatgttaatggat

MDN1 5'	aatatctcagggcaaacggg	POLA1 3'	tgaatacaacacagtgatcc
MDN1 5'	tacgtccatagcgtactgga	POLA1 3'	ggccaattaagcatccctct
MDN1 5'	gcagaaacttgaaggctgct	POLA1 3'	agaaaaaaatagcaagcgcc
MDN1 5'	cggaacacagactgctcctg	POLA1 3'	agacggtcctattgtgaaga
MDN1 5'	aaggtgtcatggcttctgag	POLA1 3'	aaatacattttgctgctgcc
MDN1 5'	acaattggctgtataccagc	POLA1 3'	agagaggaaagactgccata
MDN1 5'	acaccacaaacagctgtcac	POLA1 3'	cagcataattgtacaagggg
MDN1 5'	tcttactgtcagtctgatct	POLA1 3'	agaagaaggcacaacatact
MDN1 5'	gaagaactcctattaccacc	POLA1 5'	aactccctgaatctgacaga
MDN1 5'	ctgccacacaaactctccag	POLA1 5'	ttgatttttttctcgccgg
MDN1 5'	cagaaaccacgtctaagggg	POLA1 5'	tttctagggcttcttggcgc
MDN1 5’	caatttcctccacagctcaa	POLA1 5'	ccagctttagccttttcag
MDN1 middle	aagcagcaagattgaccaca	POLA1 5'	taaacacctgtgaagtcctc
MDN1 middle	ggactagtgcatcacaaagt	POLA1 5'	cctgaaccagcttcgaatac
MDN1 middle	actggcatcagacaccattc	POLA1 5'	caatccagtcatcatcctgg
MDN1 middle	accgcagagaacctaagatc	POLA1 5'	ggcatcatcttcaaggtcat
MDN1 middle	ggcatctacttttactgtgt	POLA1 5'	tcttgtctttattgcgtgct
MDN1 middle	tgccaatggagggcaagaag	POLA1 5'	cactgcgagcttctttacat
MDN1 middle	ctggtggaccagatgtttta	POLA1 5'	tctgcagttttctttccagc
MDN1 middle	aattcatcagaagtcgggggg	POLA1 5'	accatccttggacaagtcta
MDN1 middle	ctgagacagtctgaacttct	POLA1 5'	cctgtagaatgtcacctagc
MDN1 middle	tccccagacaattctgaata	POLA1 5'	tcagtatcattacaggtggt
MDN1 middle	cttctttataccagcgaagc	POLA1 5'	tgaagctccaatggatcttt
MDN1 middle	aaacggtcgtcccaggaact	POLA1 5'	tgtgcacagagaaaggattc
MDN1 middle	ccaccagcttgtccttaaaa	POLA1 5'	gggaagcaatttttcctgaa
MDN1 middle	ggaccttcagttgtgaaaaa	POLA1 5'	agttaatggaggctcctttc
MDN1 middle	ctgatggcaaggactttgtt	POLA1 5'	cagcacgtttaagaggaaca
MDN1 middle	agacggttaatgtcttcctg	POLA1 5'	tctcgacctgtacatcatcg
MDN1 middle	ccactgagaagcaaccactt	POLA1 5'	tgactcctgctcttcttctg
MDN1 middle	cttgcaggagacttttcttt	POLA1 5'	ggctcatcaaagtcaccatc
MDN1 middle	atttgctctgaggatcagtc	POLA1 5'	agccataggctccaggtcca
MDN1 middle	ttcatctaggctgacatctt	POLA1 5'	tctctttgtcccaagccttg
MDN1 middle	ctgagcatgcacaaaattct	POLA1 5'	ttgtttcacttcctctgctg

MDN1 middle	cctttggctttcagttctaa	POLA1 5'	agtaggacacggtccctttc
MDN1 middle	ctccagaaaaccaagtgaga	POLA1 5'	catccgggagaaaacttcct
MDN1 middle	aggaagcttcatcatgcttt	POLA1 5'	tgatcaatgtcccaacaaga
MDN1 middle	gtcaagtctggatgggacag	POLA1 5'	tgcactgagaaactgctatc
MDN1 middle	tcctggtgaggtggattacg	POLA1 5'	gactggaatccacttgaact
MDN1 middle	attgcaggccacaactgaac	POLA1 5'	gccccttttaccaatgggag
MDN1 middle	tttgtaccgccaaagcatag	POLA1 5'	agtggaatacttgttcctca
MDN1 middle	gtgccataaaatcagctgtc	POLA1 5'	ggttggttgtactgatcctc
MDN1 middle	ctgcatcttctgagacaagc	POLA1 5'	ctcggctgattcaatccaaa
MDN1 middle	tttatctggggttgttgatt	POLA1 5'	atgacacaacagctcacatg
MDN1 middle	agatgaggtgacttatctcc	POLA1 5'	agcgttcgctcgatattttt
MDN1 middle	acaggtgtgtgataaagaca	POLA1 5'	gtttctttccccgtatttag
MDN1 middle	atctcgaagctcttgaggtg	POLA1 5'	gcatagttcttttccactgg
MDN1 middle	tgatgatgcagcaaggacca	POLA1 5'	ctggaacatcaggtatctca
MDN1 middle	aatttcttcaggagacacct	POLA1 5'	aatcttgaggaagctgtggc
MDN1 middle	ataactcggaccacaaagat	POLA1 5'	agatgtgttggtcccaaata
MDN1 middle	agtactgctccagaaagaca	OIP5-AS1 5’	acgggaaagtaactgggtaa
MDN1 middle	agtactctggatttgtggtc	OIP5-AS1 5’	gaagtgatagccacatttca
MDN1 middle	ggcaaagggttccacattag	OIP5-AS1 5’	ccaaatcatggaggtaatgt
MDN1 middle	tccaaaacagacttgggtgc	OIP5-AS1 5’	aagaaaagtggccttttcca
MDN1 middle	tgggtctattgagattgccg	OIP5-AS1 5’	ctcctctgtgttgaatttga
MDN1 middle	tcaaagcagcacttagagaa	OIP5-AS1 5’	agcaggataactggaatcct
MDN1 middle	tctccagctgctggttaaga	OIP5-AS1 5’	cttctttcttcctggagatg
MDN1 3'	aagtctcactttggactctt	OIP5-AS1 5’	agcaaaagacccatgtgagt

MDN1 3'	aatgtgaccttctgaccaca	OIP5-AS1 5’	agccttctgcttgcaaaatg
MDN1 3'	aaaagggagcacctgggtaa	OIP5-AS1 5’	gcacctgtttcaaatgaaac
MDN1 3'	agcattctgtaggctgtaag	OIP5-AS1 5’	attcacaaataccaccacct
MDN1 3'	tggtataaaaacctcagccc	OIP5-AS1 5’	cttctttagtcttcttgctt
MDN1 3'	actcttctctagttacgagt	OIP5-AS1 5’	ccctgaagtcattcatactt
MDN1 3'	cttccaaggcagggaagaag	OIP5-AS1 5’	gtggctgaaactggaagaca
MDN1 3'	aagaaaacaggcagctgggc	OIP5-AS1 5’	ctcaactgtgcttatcatgg
MDN1 3'	acaaaggactgtcagagtcc	OIP5-AS1 5’	atttggaatttactgctgca
MDN1 3'	aaaagggcagctccctttag	OIP5-AS1 5’	ttttagtatctttcacgtca
MDN1 3'	gcaaggcagagcttagaaca	OIP5-AS1 5’	caggatgagccaggatttaa
MDN1 3'	tttgggcacacactatgggc	OIP5-AS1 5’	gttcctattcagtaaatcta
MDN1 3'	ctgtcttggccacttgacag	OIP5-AS1 5’	ttttgtttaaaattgggcct
MDN1 3'	cctcatactctccagaaacg	OIP5-AS1 5’	aactcctctatgtccaaagt
MDN1 3'	tctaagagaaggtagttcct	OIP5-AS1 5’	ggaaaattctctcatcctcc
MDN1 3'	ctataatgtccagttgcttt	OIP5-AS1 5’	catgagggatttttgcttct
MDN1 3'	tttttatagatgacctgggc	OIP5-AS1 5’	caccataaagtcagatggca

MDN1 3'	tttttacacagcccaaggat	OIP5-AS1 5’	gggttgcaggaagagttaat
MDN1 3'	gaggatactgaaaagccact	OIP5-AS1 5’	caaacatccaagtatccacc
MDN1 3'	cattgcatagtctcccgaag	OIP5-AS1 5’	ccttttcagcctagaaatca
MDN1 3'	ataaaggggcaatcaccttc	OIP5-AS1 5’	ctggggaaagtacctgagtc
MDN1 3'	tacaacaacagggaccatgg	OIP5-AS1 5’	tgatgagaaagttcagtccc
MDN1 3'	agtgtgaggaatcactcttc	OIP5-AS1 5’	tgggaacattgcttctgagg
MDN1 3'	tggctcagtccagcttgaaa	OIP5-AS1 5’	cacgatgacccaaccacaag
MDN1 3'	cgtttctttcccagaatgag	OIP5-AS1 5’	ttcttatttgaggtttcttt
MDN1 3'	atagatggagctgctgagtt	OIP5-AS1 5’	ttctacgacagtctgttctt
MDN1 3'	atcagtttctttcgactgga	OIP5-AS1 5’	gctgatttggaagcaaagac
MDN1 3'	ggccaagtaaaaactgccta	OIP5-AS1 5’	acaatacatacaatggtcct
MDN1 3'	caagtattcagcactgcttt	OIP5-AS1 3’	accacagatctgtcagtatt
MDN1 3'	agtagaacagagcacacagt	OIP5-AS1 3’	gccaaacccatcaaggataa
MDN1 3'	atcatgacatactgcctaca	OIP5-AS1 3’	gctttagtcaagaaattgca
MDN1 3'	tcgcagacttcacagtgtaa	OIP5-AS1 3’	aagtttctcctgtttaacct
MDN1 3'	atctgtgtctttgatgacca	OIP5-AS1 3’	gcaaggattttcttctagtg

MDN1 3'	tacatgctttgggacacttg	OIP5-AS1 3’	gcaaggattttcttctattg
MDN1 3'	aagatcagtcccctagcata	OIP5-AS1 3’	gtcaggaattttctcaagga
MDN1 3'	ctgactgactgatccagcag	OIP5-AS1 3’	tgtcacaatcacttgtactt
MDN1 3'	gcacagcatcaactagtaac	OIP5-AS1 3’	tgaatgccatttttacgtca
MDN1 3'	gaagtaggaggggatcatgt	OIP5-AS1 3’	tttaggcttgtttcagacat
MDN1 3'	cctttgtagtaagagcaaca	OIP5-AS1 3’	accttgaatgtgctttgtga
MDN1 3'	cagcctaccatggacataaa	OIP5-AS1 3’	tcccaaagttatttgtagca
MDN1 3'	ctgcaaagccagcatattat	OIP5-AS1 3’	gttaccattccactttattc
MDN1 3'	gcctccttataaggctacac	OIP5-AS1 3’	ttcaattctgaaccactgga
MDN1 tiling	acacacactgcaagctgctc	OIP5-AS1 3’	accaaacatcacagtcacaa
MDN1 tiling	ctggagcagccaagttaatc	OIP5-AS1 3’	ggatatgacatctcacttca
MDN1 tiling	agagtgagcgctgaatgttg	OIP5-AS1 3’	tgctatcaagtaattgggga
MDN1 tiling	ccctgtcaagctgtgttaag	OIP5-AS1 3’	gttcttaatctttactggct
MDN1 tiling	ttcacggttatcatccaaca	OIP5-AS1 3’	gccttagtggaataatcgga
MDN1 tiling	tcctgtgtttctgttactag	OIP5-AS1 3’	tgatgtctggtttcccaaaa
MDN1 tiling	agttacactccaatgtggat	OIP5-AS1 3’	gatttaggacttcatggaga

MDN1 tiling	ctcagtccacacgatatggc	OIP5-AS1 3’	agatacctattctgacttta
MDN1 tiling	ccatatttctgtaaaccggt	OIP5-AS1 3’	acctgaatactgtcagtaac
MDN1 tiling	taaatcttcacggcttgtgc	OIP5-AS1 3’	tgacatatcaccaagggaga
MDN1 tiling	attattcctgtgtaggacag	OIP5-AS1 3’	tccctgggagaaaatatttt
MDN1 tiling	tgcactgagtgcatagtctg	OIP5-AS1 3’	gtattaccagaatacctttc
MDN1 tiling	cccaaaaatcttcgtctttt	OIP5-AS1 3’	tgccaagactgttactgttt
MDN1 tiling	cttgtctaaagggattctga	OIP5-AS1 3’	gaaatgggtcggttttgtga
MDN1 tiling	aatgactggtgcaggagcaa	OIP5-AS1 3’	attcttacaaggcagtaggt
MDN1 tiling	tcatctgcacacacttcatg	OIP5-AS1 3’	gtaaacacagtgagcaactc
MDN1 tiling	caactgcgtaagcttcacac	OIP5-AS1 3’	tggagaatatggaggacagc
MDN1 tiling	tcaaatgtgccatggctatg	OIP5-AS1 3’	acgttatacaccaatggtgc
MDN1 tiling	gcaggtactggacaattagt	OIP5-AS1 3’	aagcaagctgccttttgtaa
MDN1 tiling	aggctctgtctaaactcagc	OIP5-AS1 3’	accccaaagatgcatagatt
MDN1 tiling	agccatttaactctgagcat	OIP5-AS1 3’	ccaacttttttaacagttcc
MDN1 tiling	gcggtatattcttggctaaa	TUG1 5’	cactaaggcggcataaggag
MDN1 tiling	ggactagtgcatcacaaagt	TUG1 5’	accccacacacattgatagg
MDN1 tiling	actggcatcagacaccattc	TUG1 5’	ttgtgaagggtcccaaatga
MDN1 tiling	gttctaaacactgagcatgc	TUG1 5’	ggactcaaacagggcttcaa
MDN1 tiling	aaccaagtgagagtcctttg	TUG1 5’	agaaacagctcacatatccc

MDN1 tiling	ggcaaagggttccacattag	TUG1 5’	aaccccgaatatccattatg
MDN1 tiling	tccaaaacagacttgggtgc	TUG1 5’	aatagaagccaagcagggga
MDN1 tiling	gaggtgagcaactggcagag	TUG1 5’	ccccatggtgaaagaaaagg
MDN1 tiling	ttcaccaacaaagtggtgca	TUG1 5’	attgtaccatatgcatcagc
MDN1 tiling	tttagaaggctctgggctac	TUG1 5’	tcctaattgtagctgcttta
MDN1 tiling	cggaactggtggtgtgactg	TUG1 5’	ggatgctacagaacacatcc
MDN1 tiling	acatgtctgaagagctgcag	TUG1 5’	gttgacgggccaaaggataa
MDN1 tiling	ttctcttgggctatgcgttc	TUG1 5’	accagtacaagcagcagata
MDN1 tiling	tagaagtcatagggcccatc	TUG1 5’	tatggtcaatgagagtcaga
MDN1 tiling	tgcttctggaacattgggat	TUG1 5’	atggatgacaaccatggtcg
MDN1 tiling	gcaacatcaaggtcatctgt	TUG1 5’	cctagacatgtaaagtagga
MDN1 tiling	aagtcgcacatggaactctc	TUG1 5’	ctattcaccaccaaccacac
MDN1 tiling	ctgacaccactctggaattg	TUG1 5’	ctggcagcaccatgtaaaaa
MDN1 tiling	tttggcaagcgacgcagaag	TUG1 5’	cgtgcaccattaattagctg
MDN1 tiling	gttgctagtgctgcgttaag	TUG1 5’	tgagcccgcttgctaaaagt
MDN1 tiling	aatgtgctgagaaccctctg	TUG1 5’	aattccatgccaggttcagt
MDN1 tiling	cagctgacttccaggtattg	TUG1 5’	aaaaaccccaaacatcttca
MDN1 tiling	caaaggtgatcctgtttccg	TUG1 5’	ctactgattttgaggttccg
MDN1 tiling	cataatcttcttgtgggctt	TUG1 5’	ttagctttgaatcacttcca
MDN1 tiling	ttagatgtcctgattgcagt	TUG1 5’	agaacacaaggagggccaag
MDN1 tiling	cccttctgaaatgtatcttc	TUG1 5’	tcctagtattacacttgcaa
MDN1 tiling	tcgcccttaatatcagattt	TUG1 5’	gcaaaaactggcatcttgga
MDN1 tiling	ccacctcattttcatcatag	TUG1 5’	tttgttggataatggcctga
MDN1 tiling	ttcctgattgccatggtaag	TUG1 5’	tgagcaccactccacaaaac
MDN1 tiling	ggccaactgggcaatgaaat	TUG1 5’	atgtgtttgtgttcacttgc
MDN1 tiling	tttttcctggtgtgcttctg	TUG1 5’	agcataactggctaacatct
MDN1 tiling	gtaactctgccacatctcag	TUG1 5’	tctctgccctttaggaaaag
MDN1 tiling	gtgaaagaggcgctgttaag	TUG1 5’	gtcatttctctggaagtgag
MDN1 tiling	gagcagctgcaaacatgttg	TUG1 5’	accaacacttttttttctcc
MDN1 tiling	aactgatgttctgcgagagc	TUG1 5’	gaggtacatccggatttaat
MDN1 5’ additional	atgactgtttagcggtcgat	TUG1 5’	tcaccacagtcttaagtctt
MDN1 5’ additional	ccaggtgaattttggtccaa	TUG1 5’	ttctgctgaggaaagcatct
MDN1 5’ additional	cagttctctcttatccaggt	TUG1 3’	gagtgacagggtcagcagag

MDN1 5’ additional	gtttctctccagtaagttgg	TUG1 3’	gtttaggagtctatgctaca
MDN1 5’ additional	gaactatcactccaagagtg	TUG1 3’	cctcaatcagacttgaggtt
MDN1 5’ additional	aacttcttcaggtgcctgtt	TUG1 3’	actttcagctcaggagaagg
MDN1 5’ additional	ctcaagggttggtcttttgt	TUG1 3’	tttcttgctcatctgccaaa
MDN1 5’ additional	gggcaatcctattacaccaa	TUG1 3’	agtcaggtacatcagctttc
MDN1 5’ additional	ctcagaaagcattgctgtga	TUG1 3’	ggatacagttttctgatctt
MDN1 5’ additional	gcttccaataacttctgcca	TUG1 3’	ggttaatccatagggcttat
MDN1 5’ additional	tttgctgacacacactgcaa	TUG1 3’	gtttttactctgggtactca
MDN1 5’ additional	atggtagaggttttgccagt	TUG1 3’	ccagcagatcaataaggaag
MDN1 5’ additional	cctgtaatgtgagccaagta	TUG1 3’	gccagaatatgatctggaag
MDN1 5’ additional	attgacaaccctcaaacggt	TUG1 3’	agaaaggccagccataccaa
MDN1 5’ additional	gcaagtctgcagtatcactt	TUG1 3’	aatagacaagcagggtacct
MDN1 5’ additional	atggtccaccggtttataac	TUG1 3’	aaaggcaggcaagagctgag
MDN1 5’ additional	gtaagggtagccaaataagc	TUG1 3’	gtctaattgcagcaacatgt
MDN1 5’ additional	aagagttcctcaaatgcctc	TUG1 3’	aattgactgtagtcctcacg
MDN1 5’ additional	gccccaagaacgtaaagttt	TUG1 3’	aaatccttgttgtattgggc
MDN1 5’ additional	ctgtctgtaacaggtctgaa	TUG1 3’	aggtgtgggttgttactatt
MDN1 5’ additional	ttagtctcaggagatcatgc	TUG1 3’	gaacattgagtcctagtggg
MDN1 5’ additional	agcagacttgtgtacatgct	TUG1 3’	aagttgctggtcaaggagta
MDN1 5’ additional	cactgtcttttccatccttg	TUG1 3’	acatacaggaatagaggcct
MDN1 5’ additional	ccaaatgcttcccatttctc	TUG1 3’	aggttccaggtgctgaaaat
MDN1 5’ additional	ttgggcatggttgagtctaa	TUG1 3’	aactgctgtattccttccag
PRPF8 5’	ctcgataaggaaacactccg	TUG1 3’	agagaaatggacgcggcttt
PRPF8 5’	caattgctgccattttcgag	TUG1 3’	actcatttctgcactactga
PRPF8 5’	acttccgcttttctgcatag	TUG1 3’	tctgtgtgtactggtgaatc
PRPF8 5’	tggtctcgaatgatcttcct	TUG1 3’	tgctaggttgaactggtaca
PRPF8 5’	aacttcctgttggtcatgtc	TUG1 3’	acagtggaaacttttcttct
PRPF8 5’	catgttctccaggagtttga	TUG1 3’	ctgtgaggcaatttgagtca
PRPF8 5’	attgacgaaggaaatggctc	TUG1 3’	acaactagccttctacatca
PRPF8 5’	gtagacaggttcaatgaccc	TUG1 3’	aactgtcctgctgatctgaa
PRPF8 5’	ttttctcggcgcatcataat	TUG1 3’	agtggtcatgagtctgagag
PRPF8 5’	ggaaaacgcatcctcttgaa	TUG1 3’	ccacagtttcaacacaagca
PRPF8 5’	cggctcctcatcatcaaaag	TUG1 3’	gtccaatagcatatgttgga

PRPF8 5’	ggatgttgtcagcatagtcc	TUG1 3’	cctcaagaagtctgtaatcc
PRPF8 5’	ggtcatagaaccagtccaac	TUG1 3’	atgagagataagtttgtcct
PRPF8 5’	tggtaagtggagccatttac	TUG1 3’	gtttcttccttggttataaa
PRPF8 5’	cggtagagagtcgacatcat	TUG1 3’	ttgaatggtaacagctggca
PRPF8 5’	agtctgtcaggagctgatta	TUG1 3’	agcttaatctctgcttaaga
PRPF8 5’	gacgtaaagaaggccttcaa	PRPF8 3’	gtgacaatgatctgcttggg
PRPF8 5’	caggaatggccatattgagt	PRPF8 3’	gtgggtccagcatgcccttc
PRPF8 5’	gtctcgaacaagaggttcaa	PRPF8 3’	gggaagtccagtaagtgcac
PRPF8 5’	ctgccggatgataatcttgt	PRPF8 3’	ctccgatcctttgatgacaa
PRPF8 5’	tcttgtactcagtgcggata	PRPF8 3’	cttgagacacgcctggaaag
PRPF8 5’	aagtagaaagctggcaagtc	PRPF8 3’	tgagatccccgaatttttcc
PRPF8 5’	gggagctcaaattcctcatc	PRPF8 3’	ctggggctcagtggctttaa
PRPF8 5’	ccattggctgtattgtctgt	PRPF8 3’	catagaggttgaagagaacc
PRPF8 5’	cttaagcagcttctggtagg	PRPF8 3’	tgaaatagtcttgagccagt
PRPF8 5’	cttcttttgagccttagggg	PRPF8 3’	cgcagaatcaggatgagacg
PRPF8 5’	cagctttgtggactgaaaga	PRPF8 3’	atcgttgttcacatgtaggg
PRPF8 5’	ttttcacaggcttgaggttg	PRPF8 3’	gcttcaggatcacttttgcc
PRPF8 5’	gtggaaagcattcccaaaac	PRPF8 3’	cgaccttgatccattcttcg
PRPF8 5’	cagcttagtcaaacgcagaa	PRPF8 3’	caagatcagatccttgagct
PRPF8 5’	atactgcacgtgactatcca	PRPF8 3’	tgtttttcttgccgtagtca
PRPF8 5’	gcccaacatgggcaaatata	PRPF8 3’	gtcagtgatgccacgttcac
PRPF8 5’	tatcgatacatgcccgtcaa	PRPF8 3’	ggatgatgtctcgaatttct
PRPF8 5’	ctgtcgcatcagcttgtatt	PRPF8 3’	ggtgccgagatctccatacc
PRPF8 5’	atgaaaaagagccagactcg	PRPF8 3’	gatctgctgccgctgctgtg
PRPF8 5’	gccatcgctctaataaaggg	PRPF8 3’	ttggtctgcttctcgatctc
PRPF8 5’	ttgttactgtctttgccacc	PRPF8 3’	gttgccgtcagctgcgattg
PRPF8 5’	ggtcaaaatgtgactccact	PRPF8 3’	cttgttgacagtgcgagtct
PRPF8 5’	atatcatgcatcacagctgc	PRPF8 3’	gaggtgatgatctcatcgcc
PRPF8 5’	catattctctatgggcgtcg	PRPF8 3’	tctgggtctcatagttgctg
PRPF8 5’	aacagtcttgtccacagtgg	PRPF8 3’	tccactcagtcttggatgag
PRPF8 5’	tcctgttctgccttcagata	PRPF8 3’	tgattggtccttaggtgcag
PRPF8 5’	ggtatatactgccactgctt	PRPF8 3’	tgatgtcgtcagatgaaaca
PRPF8 5’	tttccaaccaatgcactgtg	PRPF8 3’	atgtaggtgtagccagtctc

PRPF8 5’	aagatgagcaacttggtgtc	PRPF8 3’	tcttaagcacattcttggga
PRPF8 5’	ctataagcttccttgagccg	PRPF8 3’	cgaaggtcagatatgcagat
PRPF8 5’	ctggttcaaccgagacttca	PRPF8 3’	ccccatataggtatcctgca
PRPF8 5’	ttatcgtaggcctgctcgat	PRPF8 3’	cggatctccttcacctgggg
dT	ttttttttttttttttttttttttttttttttttttttttttt	PRPF8 3’	actgcggcaccatcacaatg
MDN1 tiling additional	tccagcaagaagtgctccat	PRPF8 3’	caacctaagggttccatctc
MDN1 tiling additional	cgttcttggctgcgattaac	PRPF8 3’	gttgtcagccatgatcttgg
MDN1 tiling additional	tgaaggctgcttccattagg	PRPF8 3’	ataatggtcttctcgccatc
MDN1 tiling additional	aaggtgtcatggcttctgag	PRPF8 3’	cttgtaggccgtcagtgtac
MDN1 tiling additional	cagcattctgagaagcaacc	PRPF8 3’	cccattcgtagccactgggg
MDN1 tiling additional	cctattggtccttccaacaa	PRPF8 3’	ttgcccttgtctgtgttctg
MDN1 tiling additional	aagcaggtgatcaaccactg	PRPF8 3’	ctctcatagtgtgaaggcag
MDN1 tiling additional	ccaacagaactatcactcca	PRPF8 3’	gtccgacagcagcatctgca
MDN1 tiling additional	tgagaacagaggacggccgt	PRPF8 3’	ccatgaagaagccaaggaaa
MDN1 tiling additional	acacacactgcaagctgctc	PRPF8 3’	gaagttgtagttccacgagg
MDN1 5’ additional	tcgttcttggctgcgattaa	PRPF8 3’	agctcatatttcatgttggg
MDN1 5’ additional	taaggtactcaggacacact	PRPF8 3’	ctctttggggttcgccagct
MDN1 5’ additional	cacagtacagtccttatcca	PRPF8 3’	gcctgtgcacctcgtggtag
MDN1 5’ additional	agcaaatccaaaaggagagg	PRPF8 3’	ctgcaggagagcaaagttga
MDN1 5’ additional	ttgaaagactggggatgtgt	PRPF8 3’	gatccgcagagtaaacctcc
MDN1 5’ additional	gcatctgaactctctaggaa	SunTagV4	ccacttcgttctcaagatga
MDN1 5’ additional	gtgctcttcattcatacagg	SunTagV4	ccctttttcagtctagctac
MDN1 5’ additional	cctcaacctgaaatggatca	SunTagV4	aatttttgctcagcaactcc

MDN1 5’ additional	aaaaccaaggccttctccaa	SunTagV4	ttctttagtcgtgctacttc
MDN1 5’ additional	aaagggagacttctggattg	SunTagV4	tttcgagagtaactcctcac
MDN1 5’ additional	tcagacgaaacaagatgtcc	SunTagV4	ccacttcgttttcgagatga
MDN1 5’ additional	accagcacataagacctaag	SunTagV4	acttccctttttaagcgtg
MDN1 5’ additional	gaagacttttgcagacagac	SunTagV4	tcttggatagtagctcttca
MDN1 5’ additional	acagcattctgagaagcaac	SunTagV4	acctcgttctcaagatgata
MDN1 5’ additional	tcctattggtccttccaaca	SunTagV4	cggaacccttcttcaaacgc
MDN1 5’ additional	cctgtcactgcagctaaata	SunTagV4	agttcttcgagagcagttcc
MDN1 5’ additional	aagctggactttgagaagct	SunTagV4	gatccctttttaatcgagc
MDN1 5’ additional	acatctgtgcagcgatacat	SunTagV4	tgaaagtagttcctcaccac
MDN1 5’ additional	atatcctccagaaggatcca	SunTagV4	cttcgttttcgaggtggtaa
MDN1 5’ additional	aagagctctccattctccaa	SunTagV4	ccctgaacctttctttaatc
MDN1 5’ additional	aaatccaggtgccactttca	SunTagV4	tactcagtaattcttcaccc
MDN1 5’ additional	caatttcctccacagctcaa	SunTagV4	tttcgatagcaactcttcgc
MS2V5	tgattgtgaagtgtcgggtg	SunTagV4	tttttgagcctagcaacttc
MS2V5	tccacccttgtgtattgtac	SunTagV4	ttttcgagagcaactcctcg
MS2V5	tgtaatgtgtctggagggtg	SunTagV4	acctcattttccaagtggta
MS2V5	gcttctgtttgattggattt	SunTagV4	tttgctcaataactcctcgc
MS2V5	gatggtgattccttgttgta	SunTagV4	cgcgacttcgttctctaaat
MS2V5	gtatattgcacagggaatcc	SunTagV4	ttcgataagagttcttcgcc
MS2V5	gatattcgggaggcgtgatc	SunTagV4	ctcattttcgaggtggtagt
MS2V5	acgcactgaattcgaaagcc	SunTagV4	agtggtagttcttgctcaag
MS2V5	attcgactctgattggctgc	SunTagV4	ttcaatctcgcgacctcatt
MS2V5	ctcttcgcgaaagtcgactt	SunTagV4	attcttgctgagcaattcct
MS2V5	taagaatggcgcgaaggctg	SunTagV4	cgacttcgttctccaaatga
MS2V5	gtaggggagagtgtggtttg	SunTagV4	cgacttcattttccaagtgg
MS2V5	caggaacgctgatgctgttc	SunTagV4	ttgctcaataactcttcgcc
MS2V5	ttttcttgagttgggtactg	SunTagV4	ttcgttctccaagtggtaat
MS2V5	tgatgctgcatggggacata	SunTagV4	agttcttcgataagagctcc
MS2V5	ttggggatgtattcttgggg	SunTagV4	gcgacttcattctctaagtg
MS2V5	ttggtgctcggatgtgattt	SunTagV4	ttcttgctcaagagctcttc
MS2V5	aagaaacaacactccgagcc	SunTagV4	cacctcattttccaagtggt
MS2V5	atggagggtttgtccagttg	SunTagV4	ttagatagtaactcttcccc

MS2V5	tttgtcttgttggtgagagt	SunTagV4	cctcgttctcgagatgataa
MS2V5	ctgatgctgcttcgagaaga	SunTagV4	gatagttcttcgacaggagt
MS2V5	gtatgctcgagtgtttcgaa	SunTagV4	ccttttaagtcttgcaacc
MS2V5	gatcgtccacccaagaaata	SunTagV4	ttactgagtagttcctcacc
MS2V5	aattcgtgagagcatgggtg	SunTagV4	ttcgttttccaggtggtaat
MS2V5	tcgtattggacgtggaacga	SunTagV4	tcctgatcctttcttcaaac
MS2V5	tcgtgatcccgaaaggtaag	SunTagV4	cttttgagagcagttcttcg
MS2V5	atcgtgcatgcttgaatgtc	SunTagV4	gcaacctcattttccaaatg
MS2V5	gttgagacttgtggagcatg	SunTagV4	tgccacttccctttttaaa
MS2V5	tgaacccatttggtagtttc	SunTagV4	tttcgacagaagttcctcac
MS2V5	tttgaggtaggagtgggttc	SunTagV4	gctacttcattctcgagatg
MS2V5	ttgccagttttgtgggaaga	SunTagV4	gagccagaaccctttttaag
MS2V5	tttggtatgttggaatgggc	MDN1 middle Even	ggactagtgcatcacaaagt
MS2V5	gatgctgtaccagtaattgt	MDN1 middle Even	accgcagagaacctaagatc
MS2V5	tagtagtgagagatgtgggc	MDN1 middle Even	tgccaatggagggcaagaag
MS2V5	tgctgaacggtttggttttt	MDN1 middle Even	aattcatcagaagtcggggg
MS2V5	ttgatttttccgtgtgtacc	MDN1 middle Even	tccccagacaattctgaata
MS2V5	gtctttcgtatttgtaaacc	MDN1 middle Even	aaacggtcgtcccaggaact
MS2V5	ttgcgctggacgaaagcgtg	MDN1 middle Even	ggaccttcagttgtgaaaaa

MS2V5	ccgtcggatgtttttcgtaa	MDN1 middle Even	agacggttaatgtcttcctg
MS2V5	ggttgtaagtttgtgggttg	MDN1 middle Even	cttgcaggagacttttcttt
MS2V5	ctgaggtgtttgatgtacgg	MDN1 middle Even	ttcatctaggctgacatctt
MDN1 middle Odd	aagcagcaagattgaccaca	MDN1 middle Even	cctttggctttcagttctaa
MDN1 middle Odd	actggcatcagacaccattc	MDN1 middle Even	aggaagcttcatcatgcttt
MDN1 middle Odd	ggcatctacttttactgtgt	MDN1 middle Even	tcctggtgaggtggattacg
MDN1 middle Odd	ctggtggaccagatgtttta	MDN1 middle Even	tttgtaccgccaaagcatag
MDN1 middle Odd	ctgagacagtctgaacttct	MDN1 middle Even	ctgcatcttctgagacaagc
MDN1 middle Odd	cttctttataccagcgaagc	MDN1 middle Even	agatgaggtgacttatctcc
MDN1 middle Odd	ccaccagcttgtccttaaaa	MDN1 middle Even	atctcgaagctcttgaggtg
MDN1 middle Odd	ctgatggcaaggactttgtt	MDN1 middle Even	aatttcttcaggagacacct

MDN1 middle Odd	ccactgagaagcaaccactt	MDN1 middle Even	agtactgctccagaaagaca
MDN1 middle Odd	atttgctctgaggatcagtc	MDN1 middle Even	ggcaaagggttccacattag
MDN1 middle Odd	ctgagcatgcacaaaattct	MDN1 middle Even	tgggtctattgagattgccg
MDN1 middle Odd	ctccagaaaaccaagtgaga	MDN1 middle Even	tctccagctgctggttaaga
MDN1 middle Odd	gtcaagtctggatgggacag		
MDN1 middle Odd	attgcaggccacaactgaac		
MDN1 middle Odd	gtgccataaaatcagctgtc		
MDN1 middle Odd	tttatctggggttgttgatt		
MDN1 middle Odd	acaggtgtgtgataaagaca		
MDN1 middle Odd	tgatgatgcagcaaggacca		
MDN1 middle Odd	ataactcggaccacaaagat		
MDN1 middle Odd	agtactctggatttgtggtc		
MDN1 middle Odd	tccaaaacagacttgggtgc		
MDN1 middle Odd	tcaaagcagcacttagagaa		

Table S2-3: Probe and antibody combinations used

Probes Set#	Experiment	Combination of probes/antibodies and dyes used (from Table S2)
1	MDN1 middle alternating probes	MDN1 middle odd – Cy5
		MDN1 middle even – Cy3
2	MDN1 5'-3'	MDN1 5' -Cy5
		MDN1 3'- Dy550
3	MDN1 5'-tiling-3'	MDN1 5'+ MDN1 5' additional – Dy488
		MDN1 tiling – Cy5
		MDN 3' – Dy550
4	MDN1 5'-middle-3'	MDN1 5'+ MDN1 5' additional – Dy488
		MDN1 middle – Cy5
		MDN 3' – Dy550
5	PRPF8 5'-3'	PRPF8 5' -Cy5
		PRPF8 3'- Cy3
6	POLA1 5'-3'	POLA1 5' -Cy3
		POLA1 3'- Cy5
7	TUG1 5'-3'	TUG1 5' -Cy5
		TUG1 3'- Cy3
8	OIP5-AS1 5'-3'	OIP5-AS1 5' -Cy5
		OIP5-AS1 3'- Cy3
9	MDN1 5'-3'-dT	MDN1 5' -Cy5
		MDN1 3'- Dy550
		dT – Cy2
10	MDN1 tiling-3'-dT	MDN1 tiling + MDN1 tiling additional– Cy5
		MDN1 3'– Dy550
		dT- Cy2
11	PRPF8 5'-3'-dT	PRPF8 5' -Cy5
		PRPF8 3'- Cy3

		dT – Cy2
12	POLA1 5’-3’-dT	POLA1 5’ -Cy3
		POLA1 3’- Cy5
		dT – Cy2
13	TUG1 5’-3’-dT	TUG1 5’ -Cy5
		TUG1 3’- Cy3
		dT – Cy2
14	OIP5-AS1 5’-3’-dT	OIP5-AS1 5’ -Cy5
		OIP5-AS1 3’- Cy3
		dT – Cy2
15	SINAPs 5’-3’	SuntagV4 5’ – Quasar 570
		MS2v5 3’ – Cy5
		Chicken anti-GFP Antibody
		Goat-Anti Chicken Alexa 488

3. Single-molecule imaging suggests compact and spliceosome dependent organisation of long introns

3.1 Aims of Article 2

For my second project, I determined the assembly, packaging and organisation of long introns in cells. For introns to be spliced, their 5' and 3' ends need to be brought together for the subsequent assembly of the spliceosome complex. Intronic sequences contribute towards ~95% of the transcribed genome, and individual introns can range from 30nt to >2 million nt in length. Most introns - long or short, however, are spliced relatively efficiently either co- or post-transcriptionally. How cells organise and package introns co-transcriptionally to facilitate their efficient splicing, subsequently leading to the formation of a mature and export-competent mRNA, is unknown. Recent biochemical and structural studies have suggested a model where the 5' end of the intron could be tethered to the polymerase through direct interaction between the U1 snRNP and RNA Pol II to facilitate the interaction with the 3' splice site, although the validity of this model in cells is yet to be tested. To determine the organisation and compaction of introns, I visualised different regions of two introns within the sparsely transcribed POLA1 gene using the method that I had developed in Article1. The results of this study will be submitted first to bioRxiv in August 2021, and then to Molecular Cell as a short article, where I will be the first author.

3.2 Article 2

Single-molecule imaging suggests compact and spliceosome dependent organisation of long introns

Srivathsan Adivarahan, A.M.S.Kalhara Abeykoon and Daniel Zenklusen*

Affiliations:

Département de Biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, Canada

*Corresponding author: daniel.r.zenklusen@umontreal.ca

3.2.1 Summary

Removal of introns is a critical step in processing RNA polymerase II transcripts, requiring the co-transcriptional assembly of the nascent RNA and the formation of the spliceosome across the 5' and 3' ends of individual introns. However, introns in higher eukaryotes vary vastly in their length, from a few tens to hundreds of thousands of nucleotides, and it remains unclear how such long regions of RNAs are packaged and organised to facilitate the communication between the ends of introns. Here, we use a single-molecule RNA in situ hybridisation approach to investigate the spatial organisation of long introns in cells. We show the co- and post-transcriptional intron organisation using two model long introns located within the long POLA1 gene. We find that the 5' splice site of co-transcriptional introns is maintained proximal to the furthest transcribed region, consistent with the U1 snRNP-Pol II tethering model. Additionally, fully transcribed introns are assembled into compact particles with their ends in proximity, an organisation that requires the assembly of U2 snRNP on the intron. Together, our study reveals details of intron organisation and supports a model that co-transcriptional intron organisation facilitates the splicing of long introns.

3.2.2 Introduction

In eukaryotes, mRNAs are initially transcribed as precursor mRNAs (pre-mRNAs) that need to be assembled and processed before being exported to the cytoplasm for translation. One of the critical steps for mRNA processing involves the removal of introns, which is carried out by the spliceosome, a megadalton complex comprising 5 UsnRNPs and numerous non snRNP proteins that assemble at the 5' splice site (ss), branchpoint site, and the 3' ss of the intron (Will and Lührmann 2011). Splice site recognition and spliceosome assembly need to be robust as errors in splicing can result in mRNAs either quickly degraded or translated into truncated and potentially toxic proteins. Consistent with this, defects in splicing are associated with a wide range of diseases, including cancer and neurodegenerative diseases (Scotti and Swanson 2015). While spliceosome assembly and splicing catalysis have been extensively studied, very little is known about how an intron is co-transcriptionally packaged and organised and how this facilitates the assembly of the spliceosome across the ends of introns, separated by a few tens to hundreds of thousands of nucleotides (Wilkinson, Charenton, and Nagai 2019).

Intron assembly occurs co-transcriptionally, likely through packaging by hnRNPs and other RNA-binding proteins (Singh et al. 2015). This role is best described for hnRNP C, one of the core components of hnRNP particles. hnRNP C has been shown to form tetramers, and the tetrameric form of hnRNP C can package RNAs *in vitro* into compact particles of uniform size and shape, and CLIP-seq data from different cell lines suggest the possible existence of such structures *in vivo* (Huang et al. 1994; König et al. 2010; Van Nostrand et al. 2016). In addition, RNA as a single-stranded polymer can form extensive secondary and tertiary structures. A recent transcriptome-wide chemical mapping approach has identified extensive folding within intronic regions of the pre-mRNA, suggesting that RNA structure could play an important role in intron compaction (Sun et al. 2019). However, the impact of RNA folding and RNA-binding proteins on intron compaction is largely unknown, as to date, the architecture of single introns has not been visualised in cells, and how co-transcriptional assembly of introns and its subsequent organisation facilitates its splicing is yet to be determined.

In cells, splicing can occur co-transcriptionally and post-transcriptionally, with the majority of the introns known to be spliced when the RNA is still tethered to the polymerase, with co-transcriptional splicing known to increase the efficiency and accuracy of splicing (Coulon et al.

2014; Vargas et al. 2011; Bentley 2014). However, how the spliceosome functionally pairs the ends of introns in metazoans, especially when separated by several thousands of nucleotides, and how the transcriptional machinery facilitates this process has been largely unclear. Interactions between RNA Pol II and U1 snRNP led to the model that the U1 snRNP bound to the 5' ss could be tethered to the elongating polymerase keeping the 5' splice site in proximity when the polymerase reaches the 3' end (Nojima et al. 2018; Harlen et al. 2016; David et al. 2011; Robert et al. 2002; Hollander et al. 2016). A cryo-EM structure of the U1 snRNP-Pol II complex was recently obtained, showing a direct interaction between U1-70K of the U1 snRNP and RPB2 and RPB12 subunits of RNA Pol II, supporting the U1 snRNP-Pol II tethering model (Zhang et al. 2021). However, evidence for such interactions or the effect of tethering on intron organisation and splicing has not been observed in cells yet.

Here, we use single-molecule resolution fluorescent RNA in situ hybridisation (smFISH) in combination with SIM super-resolution microscopy to determine the organisation of introns, an approach we have previously used to study mRNA organisation in cells (Adivarahan et al. 2018). We find that the introns are organised co-transcriptionally into compact particles, with the organisation of the intron changing during transcription. While partially transcribed introns have the 5' end closer to the furthest transcribed region, introns containing the 3' end are assembled with the ends in proximity. This conformation, however, was found to depend on the assembly of the spliceosome, disruption of which alters intron organisation, opening up the intron. Together, our results provide the first overview of intron organisation and compaction in cells and evidence for co-transcriptional tethering of U1 snRNP to the elongating RNA Pol II.

3.2.3 Results

3.2.3.1 Introns are organised as compact particles with the ends in proximity

To study the spatial organisation and compaction of introns, we chose the POLA1 gene, encoding a subunit of DNA polymerase, as a model gene. POLA1 is a 303 kb long gene with 36 introns and is transcribed at low frequency, resulting in few nascent transcripts being present at the site of transcription at any given time (Figure 3-1 A-B). Within the POLA1 gene, we chose intron 36, 65,255 nt in length, and hybridised probes to the 5' and middle of this intron and the 5' end of the POLA1 exons in paraformaldehyde-fixed HEK293T cells. Figure 3-1 B shows 5' exon signals in

the nucleus and cytoplasm, with nuclear signals clustering in one or two sites representing transcription sites. Intron signals can be observed at and near the site of transcription. The nature of the signal at the site of transcription suggests that we detect one or more nascent introns. However, the intronic signals in the nucleoplasm near the site of transcription represent single RNAs and can be seen to overlap with the signal from the 5' exon of the gene, suggesting that these spots represent pre-mRNAs that have not been spliced co-transcriptionally or intron signal only, likely representing intron lariats. Upon quantifying the overlapping intron and exon signals, we found that 185 out of 209 individual spatially distinct intron signals colocalised with an exon signal indicating that most single introns we observe are unspliced and are still part of pre-mRNAs. Moreover, we rarely observed colocalisation of the 5' exon signal with just the middle region of the intron and lacking the 5' intron signal, suggesting that if recursive splicing does occur for long introns, it is not very common for the intron 36 of POLA1 (for more details on recursive splicing read section 4.3.4). Overall, we detect single introns at sites of transcription, as well as non-spliced nucleoplasmic pre-mRNAs and nucleoplasmic intron lariats. Henceforth, we used this approach to study intron organisation.

Next, we wanted to determine how this 65,255 nt long intron is spatially organised. Therefore, we first hybridised cells with a combination of probes tiling along the length of the intron (2 probes ~ every 1,000 nt) and probes targeting the middle and 3' regions of the intron (Figure 3-1C). Previously, we had observed a diffused and extended signal when using tiling probes against translating MDN1 mRNAs in the cytoplasm, suggestive of an elongated molecule with volume greater than the diffraction limit (Adivarahan et al 2018). However, the tiling signal for probes hybridised against the intron 36 of POLA1, although stronger than what we observed for probes against specific regions of introns, had a nearly circular shape near the diffraction limit for most introns, indicating a highly compact and possibly globular molecule (Figure 3-1 C). Interestingly, we observed that the middle and 3' regions were spatially separated for most single introns within this compact particle.

To better understand the organisation of this intron in cells, we replaced the tiling probes with probes targeting the 5' region of the intron, allowing us to quantify the relative position of the three regions. We measured the distances between different regions by first localising the centre of the signal using the 3D Gaussian fitting, converting the 3D coordinates to 2D coordinates, and measuring the distances between colocalising signals from different channels. As a measure of our

experimental and localisation error, we targeted probes to a ~1,200 nt region of 18,413 nt-long MDN1 mRNA alternately labelled with Cy3 and Cy5 and measured the distances between signals from the two channels to obtain what we term as "Colocalization precision", typically ranging between 15-25 nm. Figure 3-1D shows introns imaged with the probes against the 5', middle and 3' regions. Interestingly, though the introns had compact conformations, the introns were almost always organised with the ends in proximity compared to the middle region. Measuring distances between different regions result in end-to-end distances with a median of ~57 nm compared to distances of ~100 nm between the ends and the middle region of the intron (Figure 3-1E).

To further dissect the size and organisation of this intron, we divided the intron into four sections and designed smFISH probes against the boundaries of these sections (5', 1/4th, mid, 3/4th and 3', see Supplementary Figure 3-1), which were then used in various combinations, allowing us to calculate the distances between these regions (Supplementary Figure 3-2 and Table S 3-2). In summary, we found that the distances were the longest between the ends of the intron and the middle region and the shortest between the two ends of the intron. All other regions were separated by distances between the end-to-end and end-to-middle distances, suggesting that intron 36 molecules are possibly organised in a looped conformation (Figure 3-1E and Supplementary Figure 3-2). Furthermore, the distance measurements suggest the intron particles have a diameter of ~100 nm. When stretched out, a 65, 255nt long RNA would have a contour length of 38.5 µm considering an internucleotide distance of 0.59 nm, indicating that the introns we observed have a high level of compaction. Together, we show that most POLA1 intron 36 molecules are still part of pre-mRNAs that can be found at the site of transcription. However, most POLA1 intron 36 molecules are found near transcription sites, likely representing pre-mRNAs that are spliced post-transcriptionally and are organised into compact particles with looped conformations having their ends in proximity.

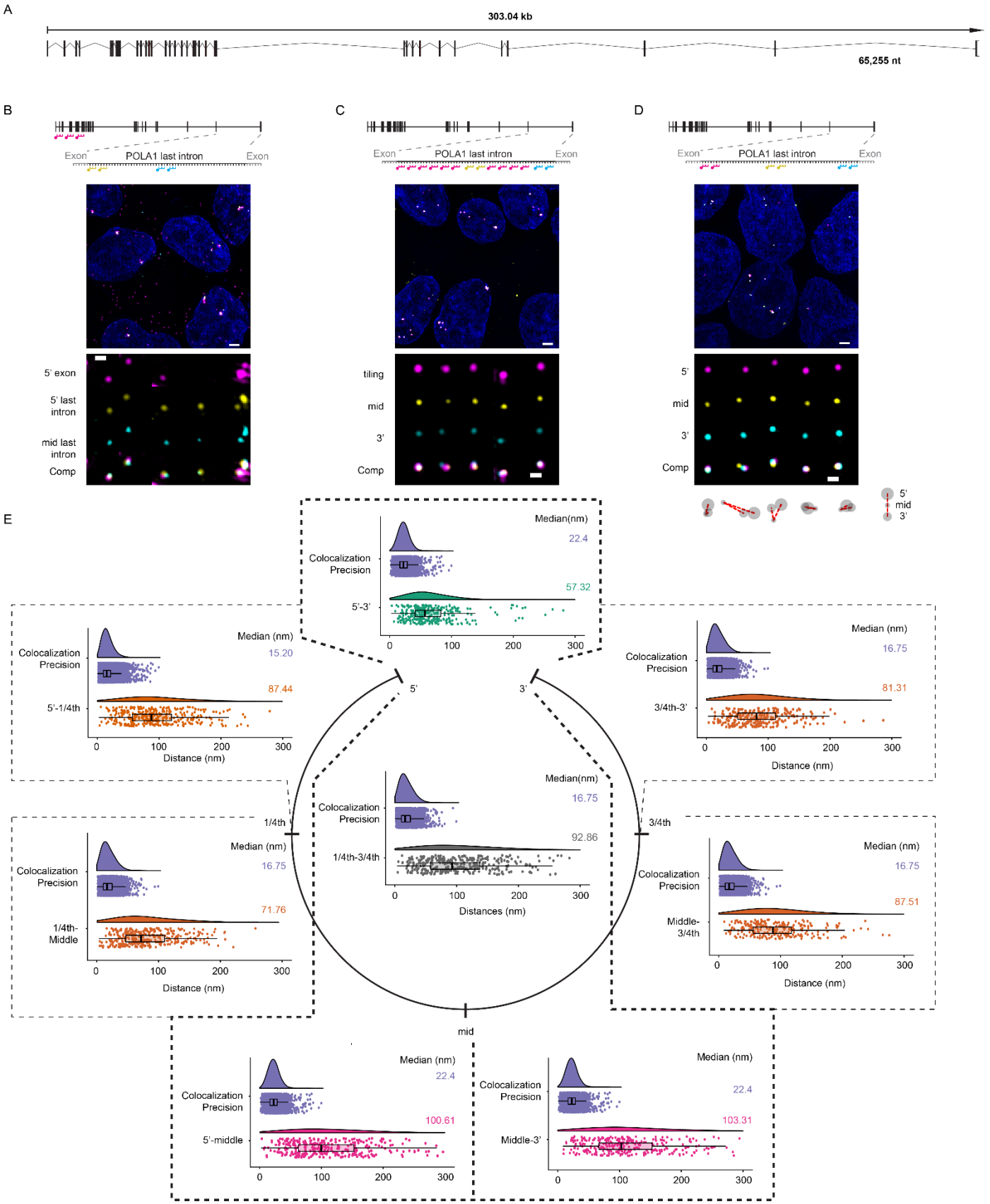

A
303.04 kb
65,255 nt
B
Exon
POLA1 last intron
Exon
5' exon
5' last intron
mid last intron
Comp
C
Exon
POLA1 last intron
Exon
tiling
mid
3'
Comp
D
Exon
POLA1 last intron
Exon
5'
mid
3'
Comp
5'
mid
3'
E
Colocalization Precision
5'-3'
Median(nm)
22.4
57.32
0
100
200
300
5'
3'
Colocalization Precision
5'-1/4th
Median (nm)
15.20
87.44
Distance (nm)
Colocalization Precision
3/4th-3'
Median (nm)
16.75
81.31
Distance (nm)
1/4th
3/4th
Colocalization Precision
1/4th-3/4th
Median(nm)
16.75
92.86
Distances (nm)
Colocalization Precision
1/4th-Middle
Median (nm)
16.75
71.76
Distance (nm)
Colocalization Precision
Middle-3/4th
Median (nm)
16.75
87.51
Distance (nm)
mid
Colocalization Precision
5'-middle
Median (nm)
22.4
100.61
Colocalization Precision
Middle-3'
Median (nm)
22.4
103.31

Figure 3-1: Organisation of intron 36 of the POLA1 gene

(A) *Exon-intron structure of human POLA1 gene for the transcript ENST00000379059, with the length of intron 36 shown at the bottom, smFISH images using probes hybridising to* ***(B)*** *5' exons and 5' and middle region of the intron 36 (Probe Set#3, Table S 3-2) in paraformaldehyde-fixed HEK293T cells* ***(C)*** *the middle (yellow), 3' (cyan), and tiling along the intron (magenta) (Probes Set#7, Table S 3-2)* ***(D)*** *5'(magenta), middle (yellow), 3' (cyan) regions (Probes Set#6, Table S 3-2) Nuclei are visualised by DAPI staining (blue). Magnified images of individual RNAs are shown at the bottom, and cartoons depicting different RNA conformations are shown below the images (D). Schematic position of probes shown on top.* ***(E)*** *Raincloud plots for distances between different regions. Individual plots show distance distribution of co-localization precision, distances for POLA1 introns as violin plots. The box plot shows the first quartile, median and third quartile and the distances corresponding to single RNAs are shown as spots overlayed on top of the box plots. Median distances are shown on the right. Scale bars, 2 μm in larger images, and 500 nm in zoomed-in images*

3.2.3.2 Intron organisation is altered during transcription

Next, we wanted to test whether other introns show similar organisation as observed for intron 36. We chose the 42,330nt long intron 35 of the POLA1 gene and hybridised probes against the 5' end of the intron and the 5' end of POLA1 exons. Compared to intron 36, which is most frequently found near and at the site of transcription, we mostly observe intron 35 smFISH signals at transcription sites (Figure 3-2A and Figure 3-2B), suggesting more efficient splicing of this intron and a fast turnover of the lariat. Combined with the low transcription frequency of the POLA1 gene in HEK293T cells, we found that this allowed us to identify single introns at the transcription site based on the intensity and size of the smFISH signals (Figure 3-2B).

To get a detailed understanding of this intron's organisation, we divided it into three regions and designed probes against the boundaries of these regions (Supplementary Figure 3-1). Three-colour imaging of the intron using probes hybridising against the 5', 3' ends with the third region reveals an organisation of intron 35 to be similar to what was observed for intron 36, with the median end-to-end distances ~40 nm and the distances between the ends and the third region larger than the end-to-end distances (Figure 3-2C-F).

As most intron 35 molecules were observed at the site of transcription, it also allowed us to visualise the organisation of partially transcribed introns. If the cotranscriptional assembly of introns is involved in ensuring the proximity of their ends, as described by the U1-snRNP-Pol II tethering model, we would expect intron organisation to be altered during transcription. To test this model, we quantified distances for introns that contained signals for the 5' and the 2/3rd region of the POLA1 intron 35 but were missing the 3' signal, indicative of transcription not yet reaching the 3' end. We hypothesised that if the 5' ss stays attached to the elongating polymerase complex, partially transcribed introns containing signals against just the 5', and 2/3rd regions should have polymerases closer to the 2/3rd region, resulting in shorter distances than observed for the same regions within introns containing the 3' signal. Visualising single introns showed an increased overlap of the 5' and 2/3rd signals when the 3' signal was absent, which was reflected in shorter end-to-end distances (Figure 3-2G-H). Figure 3-2 H shows that 5'-2/3rd distances for partially transcribed introns lie between the end-to-end and 5'-2/3rd distances for fully assembled introns. Together, our observations show that POLA1 intron 35 has an organisation similar to intron 36, with a looped conformation, which is altered during transcription, in a manner that is suggestive of the tethering of the polymerase to U1 snRNP bound to the 5' ss.

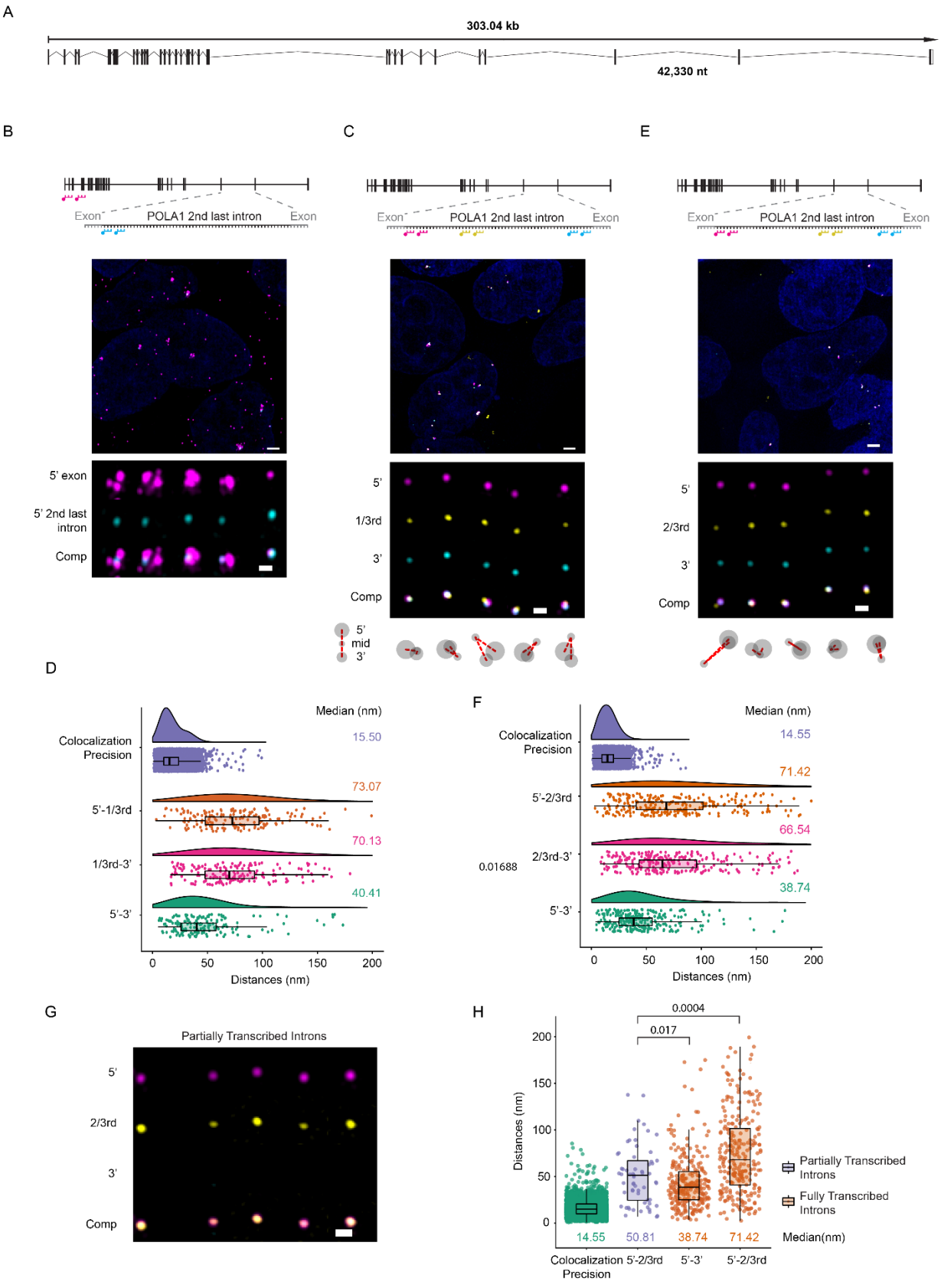

A
303.04 kb
42,330 nt
B
C
E
Exon
POLA1 2nd last intron
Exon
5' exon
5' 2nd last intron
Comp
5'
1/3rd
3'
Comp
5'
mid
3'
5'
2/3rd
3'
Comp
D
Median (nm)
Colocalization Precision
15.50
5'-1/3rd
73.07
1/3rd-3'
70.13
5'-3'
40.41
0
50
100
150
200
Distances (nm)
F
Median (nm)
Colocalization Precision
14.55
5'-2/3rd
71.42
2/3rd-3'
66.54
0.01688
5'-3'
38.74
0
50
100
150
200
Distances (nm)
G
Partially Transcribed Introns
5'
2/3rd
3'
Comp
H
0.0004
0.017
200
150
100
50
0
Distances (nm)
14.55
50.81
38.74
71.42
Median(nm)
Colocalization Precision
5'-2/3rd
5'-3'
5'-2/3rd
Partially Transcribed Introns
Fully Transcribed Introns

Figure 3-2: Co-transcriptional assembly of intron 35 of the POLA1 gene

(A) *Exon-intron structure of human POLA1 gene for the transcript ENST00000379059, with the length of intron 35 shown at the bottom* ***(B)*** *smFISH images using probes hybridising to 5' exons and 5' end of the intron 36 (Probe Set#4, Table S 3-2) in paraformaldehyde-fixed HEK293T cells.* ***(C, E)*** *smFISH using probes hybridising to 5'(magenta), 3' (cyan) regions and either 1/3rd or 2/3rdregions (Probes Set#13 and #14 respectively, Table S 3-2) and cartoons depicting different RNA conformations. Nuclei are visualised by DAPI staining (blue). Magnified images of individual RNAs are shown at the bottom. Schematic position of probes shown on top.* ***(D, F)*** *Raincloud plots for distances between different regions. Individual plots show distance distribution of co-localization precision, distances for POLA1 intron 35 shown as violin plots. The box plot shows the first quartile, median and third quartile and the distances corresponding to single RNAs are shown as spots overlayed on top of the box plots. Median distances are shown on the right,* ***(G)*** *Magnified smFISH images of partially transcribed individual RNAs using smFISH probes hybridising to the 5'(magenta), 2/3rd (yellow) and 3' regions (cyan) (Probes Set#14, Table S 3-2).* ***(H)*** *Box plot with distances between different regions for partially and wholly transcribed introns 35 from (E) and (G). The box plot shows the first quartile, median and third quartile and the individual RNAs shown as spots overlayed on top of the box plots. P-values calculated using Kolmogorov–Smirnov test are shown above. Scale bars, 2 µm in larger images, and 500 nm in zoomed-in images in (B), (C), (E) and (G).*

3.2.3.3 Spliceosome assembly on pre-mRNAs determines the final organisation of introns

The intronic organisation with the 5' and 3' ends in proximity is expected for lariats, but not necessarily for pre-mRNAs, except if these would already contain partially assembled spliceosome complexes or if the co-transcriptional assembly of introns was, by itself, sufficient for maintaining this proximity. To distinguish between these possibilities, we inhibited splicing with Pladienolide B (PB), a small molecule that binds to SF3B1 and interferes with U2 snRNP assembly at the branch point. Consistent with splicing inhibition and transcript release without splicing, we find more spots for both introns of *POLA1* in the nucleoplasm upon treatment with 100nM PB for 4h (Supplementary Figure 3-3). Treatment with PB also resulted in increased premature termination

as there was an increase in nuclear mRNAs containing just the 5' exon signal without the 3' exon signal outside the site of transcription as described earlier (Sousa-Luís et al. 2021) (Supplementary Figure 3-6 A). To determine if splicing inhibition resulted in altered intronic conformations, we targeted probes to the 5' and 3' ends of both introns in combination with probes targeting a third region of each intron (middle in case of intron 36 and 1/3rd and 2/3rd in case of intron 35). When all three signals were visualised for these introns, we noticed an increased separation of signals corresponding to the 5' and 3' ends of introns in cells treated with PB (Figure 3-3 A-C). This increased separation was reflected in our distance measurements where we see the median end-to-end distances increases from ~57 nm to ~144 nm for intron 36 and from ~40 nm to > 120 nm for intron 35, suggesting that establishment of the proximity between the 5' and 3' ends of introns is dependent on the assembly of U2 snRNP (Figure 3-3 D-F). Interestingly, we also observe a small increase in distances between the ends and the third region for both introns suggesting an increased decompaction of the intron, possibly due to the unfolding of an otherwise compact intron (Figure 3-3 D-F). Furthermore, analysing distances of individual introns shows that treatment with PB results in introns with end-to-end distances larger than distances between ends and the third region, suggesting an overall change in the organisation of introns upon inhibition of U2 assembly (Figure 3-3G and Supplementary Figure 3-4).

To monitor if the changes in intronic conformations affected the overall conformations of pre-mRNPs, we hybridised probes to different regions of three mRNAs - POLA1, MDN1and *AHNAK*. While *POLA1* and *MDN1* contain multiple introns that have previously been observed to be affected by the PB treatment, *AHNAK* is an mRNA with five exons and four introns, with the last exon having a length of 18,173nt (Kim Guisbert, Mossiah, and Guisbert 2020). Our probes to *AHNAK* were designed to hybridise to different regions within this long exon, and any changes in intronic conformations due to splicing inhibition by PB is likely not to affect the organisation of the long exon within the *AHNAK* mRNA (Supplementary Figure 3-1). We imaged these mRNAs in untreated cells and cells treated with PB and found an increased separation between different regions for *MDN1* and *POLA1* mRNAs but not AHNAK mRNAs (Supplementary Figure 3-5A and Supplementary Figure 3-6 A, D). This was reflected in our distance measurement, where the end-to-end distances for POLA and distances between different regions for *MDN1* increased upon PB treatment, while they were largely unaltered for AHNAK mRNAs (Supplementary Figure 3-5C and Supplementary Figure 3-6 B, C, F). For three colour measurements of *AHNAK* and *MDN1*,

we found that these changes in distances also resulted in changes in the size of the pre-mRNP as determined by the radius of gyration (Supplementary Figure 3-5B and Supplementary Figure 3-6E), indicating that intron assembly during transcription determines the final organisation of pre-mRNPs. Together, our observations suggest that intron organisation is dependent on the binding of U2 snRNP and possibly other spliceosomal components, inhibition of which separates the ends of the intron and alters the pre-mRNP conformation.

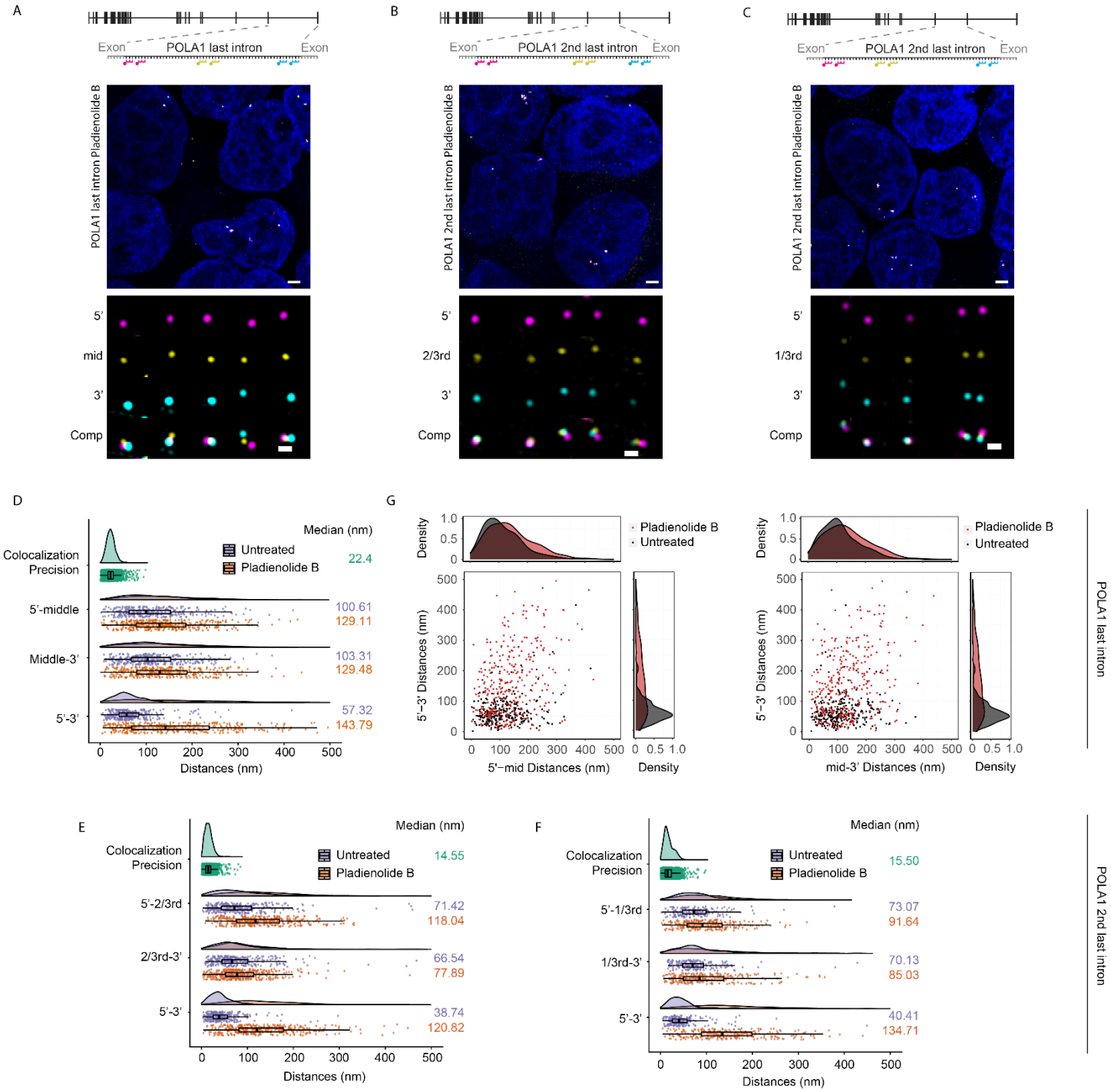


Figure 3-3: Intron Organisation is altered upon inhibition of U2 assembly

(A-C) *smFISH images in HEK293T cells treated with 100nM Pladeienolide B for 4h. Nuclei are visualised by DAPI staining (blue). Magnified images of individual RNAs are shown at the bottom. Schematic position of probes shown on top,* ***(A)*** *Probes hybridising to 5'(magenta), middle*

(yellow), 3' (cyan) regions of the intron 36 (Probes Set#6, Table S 3-2), ***(B, C)*** *Probes hybridising to 5'(magenta), 3' (cyan) regions and either 2/3rd or 1/3rdregions (Probes Set#14 and #13 respectively, Table S 3-2)* ***(D-F)*** *Raincloud plots for distances between different regions for untreated and treated cells from (A-C). Individual plots show distance distribution of co-localization precision and distances for POLA1 introns as violin plots. The box plot shows the first quartile, median and third quartile and the individual RNAs shown as spots. Median distances are shown on the right.* ***(G)*** *Scatter plot showing 5'mid and mid-3' distances for individual introns in untreated (black) and PB treated (red) cells. Frequency distribution is shown on top and on the right. Scale bars, 2 µm in larger images, and 500 nm in zoomed-in images*

3.2.4 Discussion

Although spliceosome assembly and the catalytic reaction has been extensively studied, how introns are co-transcriptionally compacted and organised to facilitate the communication between the 5' and 3' ends are still poorly understood. Here, using single-molecule localisation microscopy, we find that introns are co-transcriptionally packaged into compact particles, organised with their ends in proximity. This proximity is possibly achieved through U1 snRNP -Pol II tethering as we observe that the 5' end is proximal to the farthest transcribed region of intron 35 within the POLA1 gene. Finally, we show that the assembly of introns is defined by recognition of its boundaries, as inhibition of U2 assembly using Pladienolide B separates the ends of the intron.

3.2.4.1 Introns are co-transcriptionally packaged into compact particles

Previously, we had shown that MDN1 mRNAs are linearly organised in the nucleoplasm, and our observations for intron 35 and 36 suggest that they are linearly packaged during transcription (Adivarahan et al 2018). Our results show that these introns, though normally organised with the ends in proximity, undergo a conformational change upon inhibition of U2 assembly, indicative of the opening of the loop. Though we cannot rule out a change in intron packaging upon treatment with Pladienolide B, the observation that the opened intronic loops have end-to-end distances larger than distances between 5'-middle and middle-3' suggests that these introns are possibly linearly organised.

Furthermore, our distance measurements between different regions of intron 35 and 36 suggest their assembly into compact particles. Intron 36 has a contour length of 38.5 μm, assuming an inter-nucleotide spacing of 0.59 nm and a diameter of 100 nm. This results in compaction of ~400 fold, similar to what is observed in BR mRNPs and higher than what was observed for MDN1 mRNPs in our previous study (Adivarahan et al., 2018; Mehlin et al., 1995). These high compaction levels are possibly achieved through a combination of RNA folding and associated RBPs, likely hnRNP proteins, and it will be interesting to test the effects of intron compaction upon depletion of factors from the hnRNP family of proteins, including hnRNP C (Huang et al. 1994; König et al. 2010; Van Nostrand et al. 2016).

3.2.4.2 5'-3' ends are possibly bridged through U1snRNP-Pol II tethering

We observe that the organisation of intron 35 is altered during transcription. The distances between 5' and 2/3rd regions are lower for introns lacking the 3' signal than those where the intron has been fully transcribed. These results agree with the model proposing the tethering of the 5'ss assembled U1 snRNP to RNA Pol II and are the first indications for such tethering in cells. Whether this results in the immediate assembly of the spliceosome once the polymerase reaches the 3' end is unclear. We find that inhibition of U2 assembled separates the ends of intron 35. Furthermore, our end-to-end measurements for POLA1 intron 35 yield a median distance of ~40 nm, similar in size to a fully assembled spliceosome (~25-30 nm). Taken together, these observations could indicate the possible assembly of spliceosomes within these introns.

Though we were able to detect co-transcriptional assembly of intron 35, we could not do the same for intron 36 of the POLA1 gene. The slower kinetics of splicing of this intron means that most single spatially distinct introns were not at but near the site of transcription. However, intron 36 shows a similar conformation to intron 35, with most introns having their ends in proximity, though with a slightly larger end-to-end distance, and has comparable changes in conformation in response to PB treatment. These similarities could indicate that tethering could cause the proximity of the ends within this intron. However, whether U1-snRNP Pol II tethering is the cause for end-to-end proximity for intron 36 and is a general mechanism to facilitate the splicing of all introns needs to be determined. Furthermore, our observations cannot help determine whether tethering is achieved through interactions observed in the recently obtained U1snRNP- Pol II structure and future experiments with other long introns and in cells with mutations to residues involved in interactions

between U1-70K and RNA Pol II need to be done to establish the generality of U1snRNP-Pol II tethering (Zhang et al. 2021).

Alternate models have also been proposed to explain the splicing of long introns, among which recursive splicing (RS) has gained recent prominence (more details can be found in Section 4.3.4). While RS has been characterised in *Drosophila*, its role in splicing long introns in humans remains unclear, despite recent transcriptome-wide studies (Sibley et al., 2016; Wan et al., 2021). We do not observe extensive recursive splicing for intron 35 and intron 36 of POLA1. Additionally, the presence of fully transcribed introns in a conformation capable of undergoing splicing suggests that recursive splicing alone could be insufficient to explain the splicing of all long introns.

3.2.5 Materials and Methods

Reagents used

Splicing inhibitor Pladienolide B was bought from Cayman Chemicals(#16538) – stock at 100 µM in DMSO. The drug was diluted in warm media to get its final working concentration of 100nM, and cells were treated for 4 hrs before fixation.

Cell culture

HEK293T (American Type Culture Collection CRL-3216) cells were maintained at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) (Wisent, 319-015-CL) supplemented with 10% fetal bovine serum (FBS) (Wisent, 080-150) and passaged every 2-3 days with Trypsin (Wisent 325-043-EL). Prior to the day of treatment and fixation, cells were plated on poly-L-Lysine (Sigma, P8920 - final concentration of 0.01% w/v) coated coverslips. On the day of treatment, the media was replaced with media containing the drug and placed back in the incubator. After incubation, the cells were briefly rinsed with 1x PBS, after which they were fixed using 4% PFA, 1X PBS for 10min at room temperature. After that, the cells were washed twice with 1X PBS for 5 min each and permeabilised with ice-cold 70% ethanol. The coverslips were stored at -20°C for at least 12 hrs before being used for smFISH.

smRNA FISH probe design and labelling

The sequences for the introns and mRNAs were obtained from Ensembl, and the designed probes are listed in Table S 3-1, and their distribution is shown in Supplementary Figure 3-1. Probe combinations used for experiments and distance measurements are listed in Table S 3-2 and mentioned in the figure legends. smFISH was done as previously described in (Adivarahan et al. 2021). Custom DNA probe sets targeting different regions of RNAs of interest were designed using Stellaris® Probe Designer and either synthesised by Biosearch Technologies containing 3' amine-reactive group or ordered as DNA oligos from Biobasic and ThermoFisher and modified in house by the addition of amine-modified ddUTP as previously described in Gaspar et al. (Gaspar, Wippich, and Ephrussi 2017). The amine-modified probes were labelled with far-red dye Cy5 (GEPA25001), orange dyes Cy3 (GEPA23001) from Sigma or Dylight 550 (Thermo Scientific 62263) or green dyes Dylight488 (Thermo Scientific 46403) or Atto488 (Thermo Scientific A20000) as previously described in (Adivarahan et al. 2021).

smFISH

Prior to hybridisation, cells were rehydrated in 1xPBS, then rinsed with 10% formamide/2xSSC for 10 mins at room temperature. The cells were hybridised with 10-20 ng (1.2-2.5 pmol) of each probe mix along with 40 µg of ssDNA/tRNA. The probes were resuspended in the hybridisation solution (10% dextran sulfate/10% formamide/2xSSC/2 mM VRC/0.1 mg/ml BSA) and incubated for 3 hrs in the dark at 37ºC. Post hybridisation, the coverslips were washed 2x with 10% formamide/2xSSC solution for 30 min at 37°C. The second wash was carried out in the presence of 0.5µg/ml. Samples were then rinsed with 1xPBS and mounted using ProLong Gold antifade reagent with DAPI (P36935, Invitrogen).

Image Acquisition and pixel shift correction

Images were acquired with a 63x NA 1.46 oil objective on a Zeiss Elyra PS.1 system equipped with an Andor EMCCD iXon3 DU-885 CSO VP461 camera (1004x1002 pixels), and the following filter sets: BP420-480 + LP750 (Zeiss SR cube 07), BP495-590+LP750 (Zeiss SR cube 13), LP570 (Zeiss SR cube 14), LP655 (Zeiss SR cube 10) and the following lasers: 50 mW 405 nm HR diode, 100 mW 488 nm HR diode, 100 mW 561 nm HR DPSS, 200 mW 639 nm HR diode. Each image was acquired using three rotations and a grid size of 42 µm for all channels. The channels were registered using coverslips containing 0.1 µm TetraSpec beads (Invitrogen T-7279).

Images for the beads were acquired in all channels, and the correction between channels was calculated and corrected using the built-in channel alignment tool in ZEN 2012 SP5 using the affine transformation. This correction was calculated for each day of imaging.

RNA spot detection, spot assignment and distance measurements

3D images were processed using ImageJ, and the spots in different channels were separated using custom ImageJ scripts. 3D spot detection was carried out using FISH-Quant (Mueller et al. 2013). Only the X and Y coordinates were used from here on for further analysis. Masks were created in FiJi by manual segmentation as described in Adivarahan et al. 2021 for either separating nuclear and cytoplasmic RNAs or assigning spots corresponding to a single RNA. Assignment of the 5', 3' and/or the mid spots to either the cytoplasmic or the nuclear masks was done using MATLAB (MathWorks). To measure distances between different regions of mRNPs, spots from different channels were first grouped to assign neighbouring spots corresponding to a single RNA. This was achieved by using spots from one channel as a reference and finding spots from the other channels within a defined radius using the coordinates from the Gaussian fitted spots as described in Adivarahan et al 2021. Alternatively, spots from different channels corresponding to a single mRNA were assigned manually within the mask. 2D distances between different regions of the RNA were then calculated for signals within a group.

Data Plotting

All measurements were made for at least two independent biological replicates and the data plotted are representative from one of the replicates. For each measurement, at least ten different fields were imaged, with each image containing a minimum of 15 cells. A minimum of 200 RNAs was analysed for each dataset unless stated otherwise. For partially transcribed intron 35, a total of 61 introns were analysed from 36 different fields, primarily due to the lack of abundance of this intron. The centre of mass plots in Supplementary Figure 3-5B and Supplementary Figure 3-6E were made in R. The centre of mass was calculated as the mean of the coordinates of the three regions. The different conformations were then aligned using their centre of masses. The mean radius of gyration ($<R_g>$) was calculated using:

$$< R_g >= \sqrt{\frac{1}{3}\sum_{k=1}^{3}(r_k - r_{mean})^2}$$

where k represents one of the three regions of the mRNP and r_k the position of the corresponding position in space as determined by 3D Gaussian fitting, but using the X and Y coordinates. All plots were made in R.

3.2.6 Acknowledgement

We thank members of the Zenklusen laboratory for critical discussion and comments on the project. This work has been supported CIHR (Project Grant-366682), FRQ-S (Chercheur-boursier senior -DZ), FRQ-S (Formation de doctorat- SA) and CFI (DZ)

3.2.7 Author contributions

SA and DZ conceived the study; SA developed the image analysis toolbox, performed and analyzed the smFISH experiments with the assistance of AMSKA. DZ supervised the work.

3.2.8 References

Adivarahan, Srivathsan, Nathan Livingston, Beth Nicholson, Samir Rahman, Bin Wu, Olivia S. Rissland, and Daniel Zenklusen. 2018. “Spatial Organization of Single MRNPs at Different Stages of the Gene Expression Pathway.” Molecular Cell 72 (4): 727-738.e5.

Adivarahan, Srivathsan, and Daniel Zenklusen. 2021. “Probing the Conformational State of MRNPs Using SmFISH and SIM.” In RNA Remodeling Proteins: Methods and Protocols, edited by Marc Boudvillain, 267–86. New York, NY: Springer US.

Bentley, David L. 2014. “Coupling MRNA Processing with Transcription in Time and Space.” Nature Reviews. Genetics 15 (3): 163–75.

Coulon, Antoine, Matthew L. Ferguson, Valeria de Turris, Murali Palangat, Carson C. Chow, and Daniel R. Larson. 2014. “Kinetic Competition during the Transcription Cycle Results in Stochastic RNA Processing.” ELife 3 (October). https://doi.org/10.7554/eLife.03939.

David, Charles J., Alex R. Boyne, Scott R. Millhouse, and James L. Manley. 2011. “The RNA Polymerase II C-Terminal Domain Promotes Splicing Activation through Recruitment of a U2AF65-Prp19 Complex.” Genes & Development 25 (9): 972–83.

Gaspar, Imre, Frank Wippich, and Anne Ephrussi. 2017. “Enzymatic Production of Single-Molecule FISH and RNA Capture Probes.” RNA 23 (10): 1582–91.

Harlen, Kevin M., Kristine L. Trotta, Erin E. Smith, Mohammad M. Mosaheb, Stephen M. Fuchs, and L. Stirling Churchman. 2016. “Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue.” Cell Reports 15 (10): 2147–58.

Hollander, Dror, Shiran Naftelberg, Galit Lev-Maor, Alberto R. Kornblihtt, and Gil Ast. 2016. “How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing?” Trends in Genetics: TIG 32 (10): 596–606.

Huang, M., J. E. Rech, S. J. Northington, P. F. Flicker, A. Mayeda, A. R. Krainer, and W. M. LeStourgeon. 1994. “The C-Protein Tetramer Binds 230 to 240 Nucleotides of Pre-MRNA and Nucleates the Assembly of 40S Heterogeneous Nuclear Ribonucleoprotein Particles.” Molecular and Cellular Biology 14 (1): 518–33.

Kim Guisbert, Karen S., Isiah Mossiah, and Eric Guisbert. 2020. “Titration of SF3B1 Activity Reveals Distinct Effects on the Transcriptome and Cell Physiology.” International Journal of Molecular Sciences 21 (24). https://doi.org/10.3390/ijms21249641.

König, Julian, Kathi Zarnack, Gregor Rot, Tomaz Curk, Melis Kayikci, Blaz Zupan, Daniel J. Turner, Nicholas M. Luscombe, and Jernej Ule. 2010. “ICLIP Reveals the Function of HnRNP Particles in Splicing at Individual Nucleotide Resolution.” Nature Structural & Molecular Biology 17 (7): 909–15.

Mehlin, H., Daneholt, B., and Skoglund, U. (1995). Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J. Cell Biol. 129, 1205–1216.

Nojima, Takayuki, Kenny Rebelo, Tomás Gomes, Ana Rita Grosso, Nicholas J. Proudfoot, and Maria Carmo-Fonseca. 2018. “RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-Transcriptional Splicing.” Molecular Cell 72 (2): 369-379.e4.

Robert, François, Marco Blanchette, Olivier Maes, Benoit Chabot, and Benoit Coulombe. 2002. “A Human RNA Polymerase II-Containing Complex Associated with Factors Necessary for Spliceosome Assembly*.” The Journal of Biological Chemistry 277 (11): 9302–6.

Scotti, Marina M., and Maurice S. Swanson. 2015. “RNA Mis-Splicing in Disease.” Nature Reviews. Genetics, November. https://doi.org/10.1038/nrg.2015.3.

Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. 17, 407–421.

Singh, Guramrit, Gabriel Pratt, Gene W. Yeo, and Melissa J. Moore. 2015. “The Clothes Make the MRNA: Past and Present Trends in MRNP Fashion.” Annual Review of Biochemistry 84 (1): 325–54.

Sousa-Luís, Rui, Gwendal Dujardin, Inna Zukher, Hiroshi Kimura, Carika Weldon, Maria Carmo-Fonseca, Nick J. Proudfoot, and Takayuki Nojima. 2021. “POINT Technology Illuminates the Processing of Polymerase-Associated Intact Nascent Transcripts.” Molecular Cell 81 (9): 1935-1950.e6.

Sun, Lei, Furqan M. Fazal, Pan Li, James P. Broughton, Byron Lee, Lei Tang, Wenze Huang, Eric T. Kool, Howard Y. Chang, and Qiangfeng Cliff Zhang. 2019. “RNA Structure Maps across Mammalian Cellular Compartments.” Nature Structural & Molecular Biology 26 (4): 322–30.

Van Nostrand, Eric L., Gabriel A. Pratt, Alexander A. Shishkin, Chelsea Gelboin-Burkhart, Mark Y. Fang, Balaji Sundararaman, Steven M. Blue, et al. 2016. “Robust Transcriptome-Wide Discovery of RNA-Binding Protein Binding Sites with Enhanced CLIP (ECLIP).” Nature Methods 13 (6): 508–14.

Vargas, Diana Y., Khyati Shah, Mona Batish, Michael Levandoski, Sourav Sinha, Salvatore a. E. Marras, Paul Schedl, and Sanjay Tyagi. 2011. “Single-Molecule Imaging of Transcriptionally Coupled and Uncoupled Splicing.” Cell 147 (5): 1054–65.

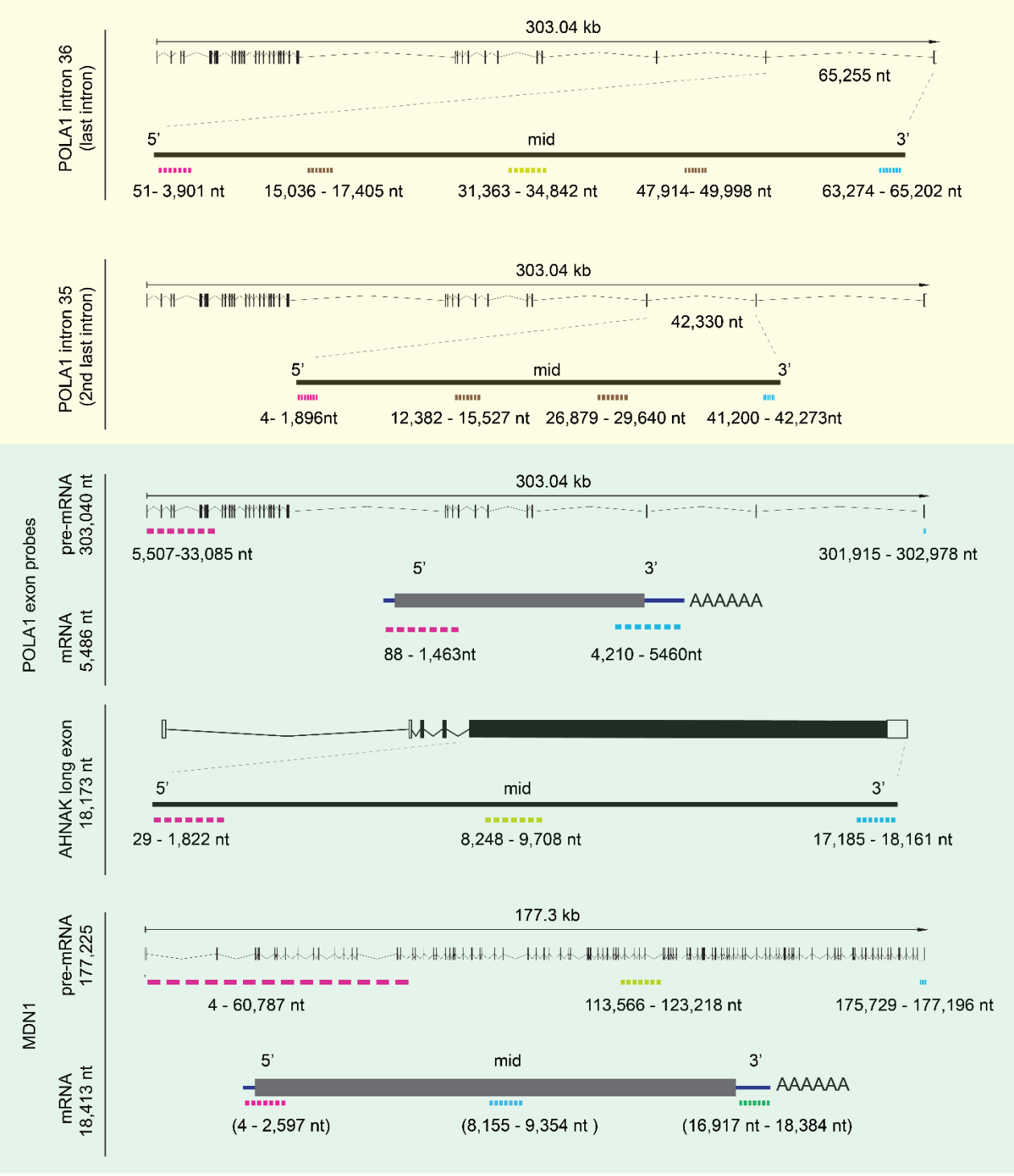
Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell.

Wilkinson, Max E., Clément Charenton, and Kiyoshi Nagai. 2019. “RNA Splicing by the Spliceosome.” Annual Review of Biochemistry, December. https://doi.org/10.1146/annurev-biochem-091719-064225.

Will, Cindy L., and Reinhard Lührmann. 2011. “Spliceosome Structure and Function.” Cold Spring Harbor Perspectives in Biology 3 (7). https://doi.org/10.1101/cshperspect.a003707.

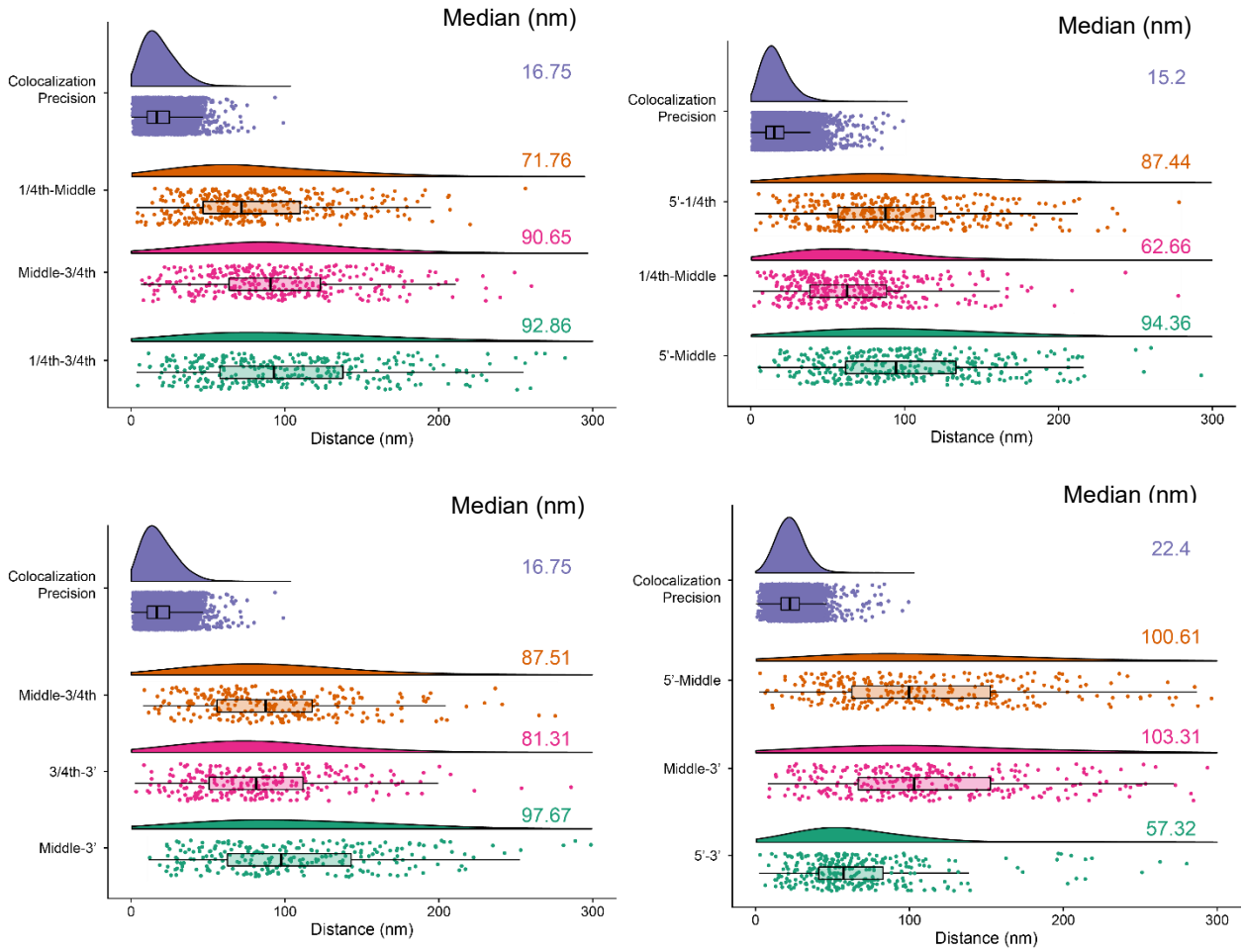
Zhang, Suyang, Shintaro Aibara, Seychelle M. Vos, Dmitry E. Agafonov, Reinhard Lührmann, and Patrick Cramer. 2021. “Structure of a Transcribing RNA Polymerase II–U1 SnRNP Complex.” Science 371 (6526): 305–9.

3.2.9 Supplementary Figures



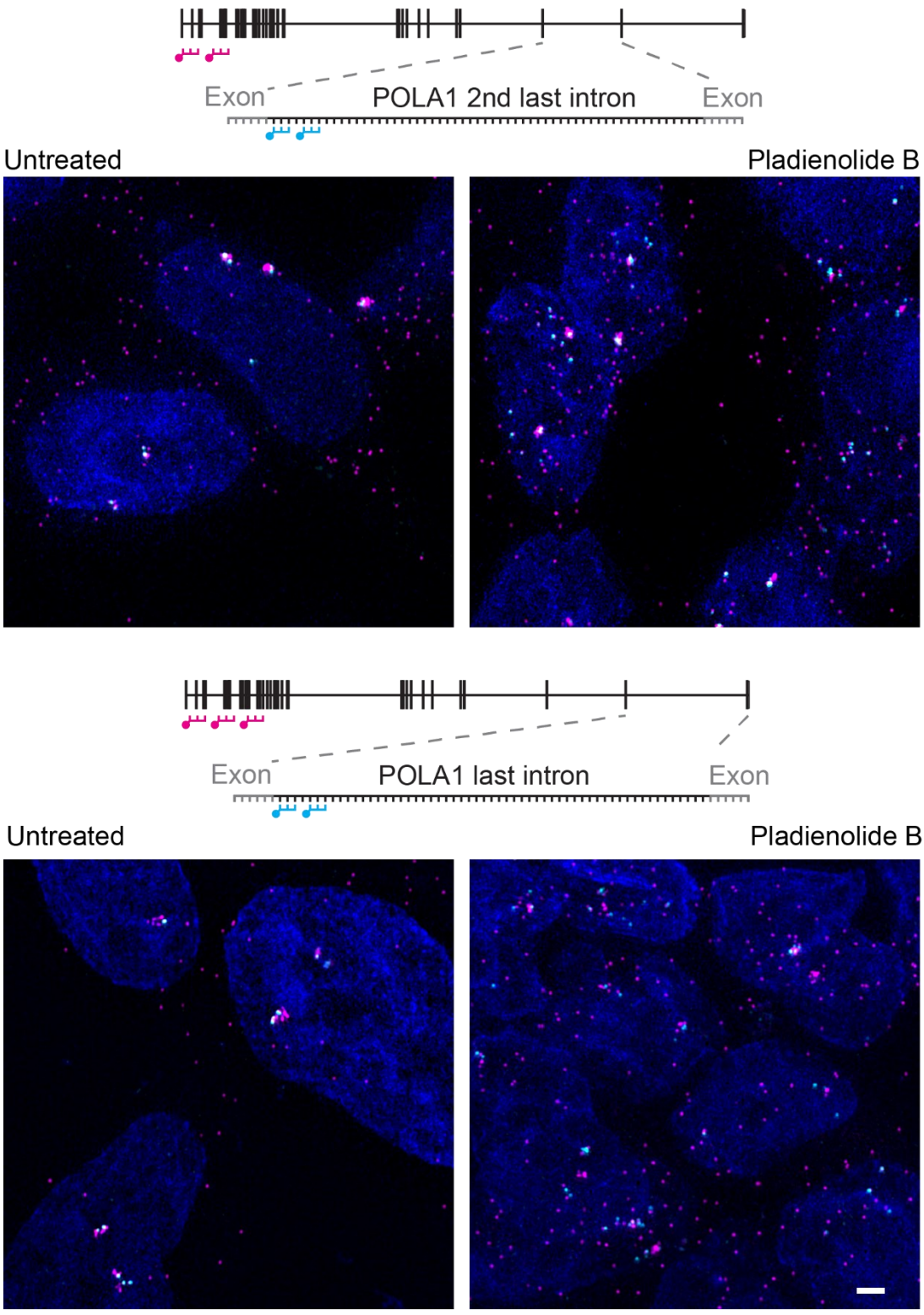
Supplementary Figure 3-1: Positions of smFISH probes used in this study

Cartoons illustrating the positions of the probes used for the different genes used. See Table S 3-1 for probe sequences. The intron and transcript sequences were obtained from ensembl.



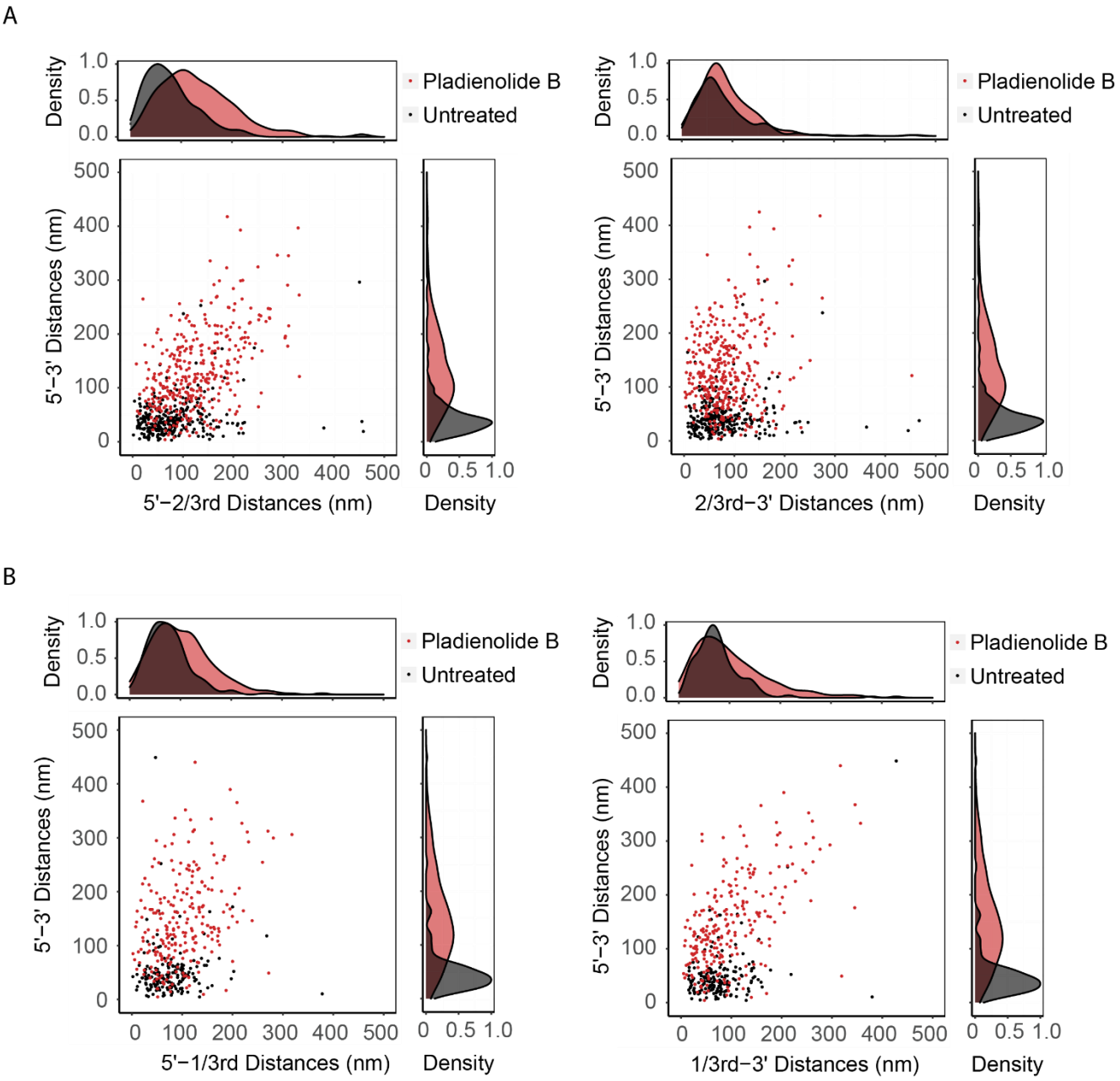
Supplementary Figure 3-2: Intron 36 smFISH distance distributions

Raincloud plots for distances between different regions used in Figure 3-1 E. Individual plots represent individual smFISH experiments using the probe combinations as follows: Top-left: Probe Set#9, Table S 3-2, Top-right: Probe Set#8, Table S 3-2, Bottom-left: Probe Set#10, Table S 3-2 and Bottom-right: Probe Set#2, Table S 3-2. Individual plots show distance distribution of co-localization precision and distances for POLA1 introns as violin plots. The box plot shows the first quartile, median and third quartile and the distances corresponding to single RNAs are shown as spots overlayed on top of the box plots. Median distances are shown on the right



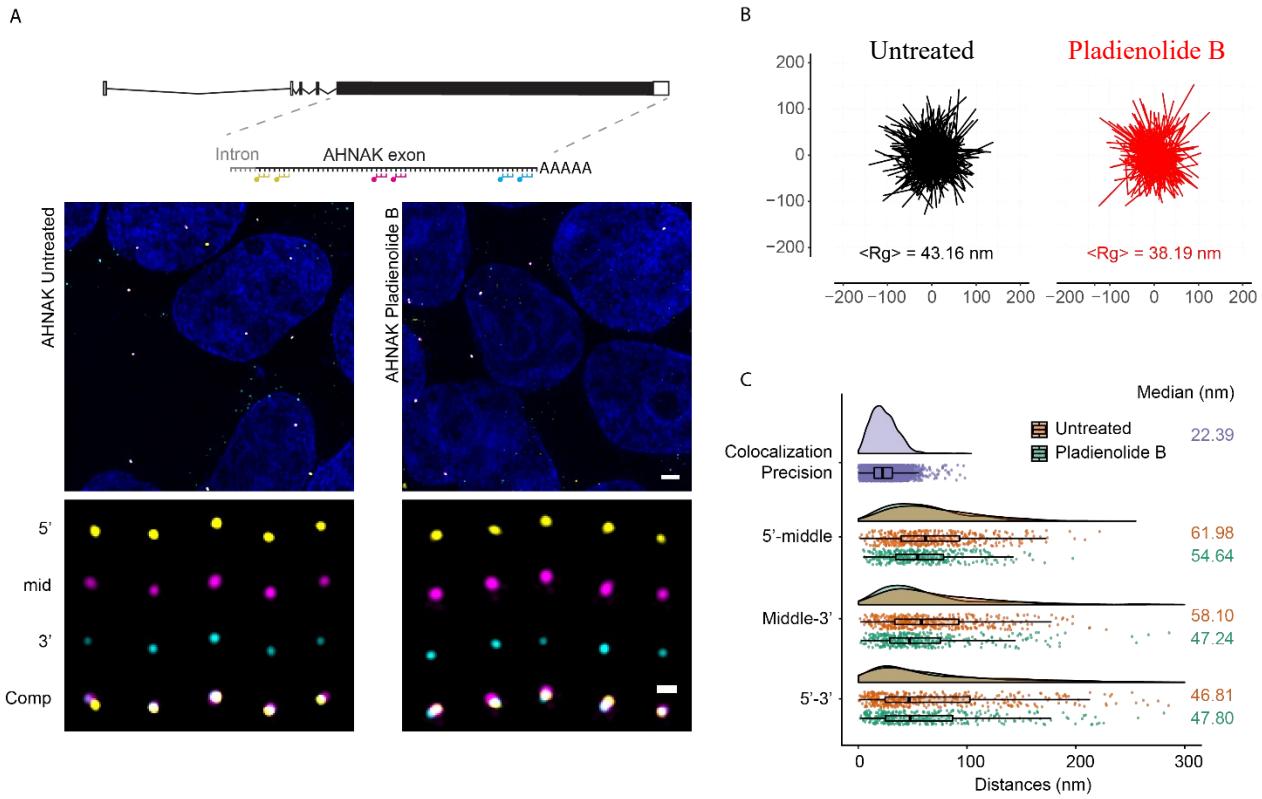
Supplementary Figure 3-3: Splicing inhibition upon Pladienolide B treatment

smFISH images in mock HEK293T cells and cells treated with 100nM Pladeienolide B for 4h with probes hybridising to the 5' exon and 5' end of the intron 35 (Probe Set#4, Table S 3-2) (top) or 5' end of intron 36 (bottom) (Probe Set#16, Table S 3-2). Nuclei are visualised by DAPI staining (blue). Schematic position of probes shown on top. Scale bar is 2 µm



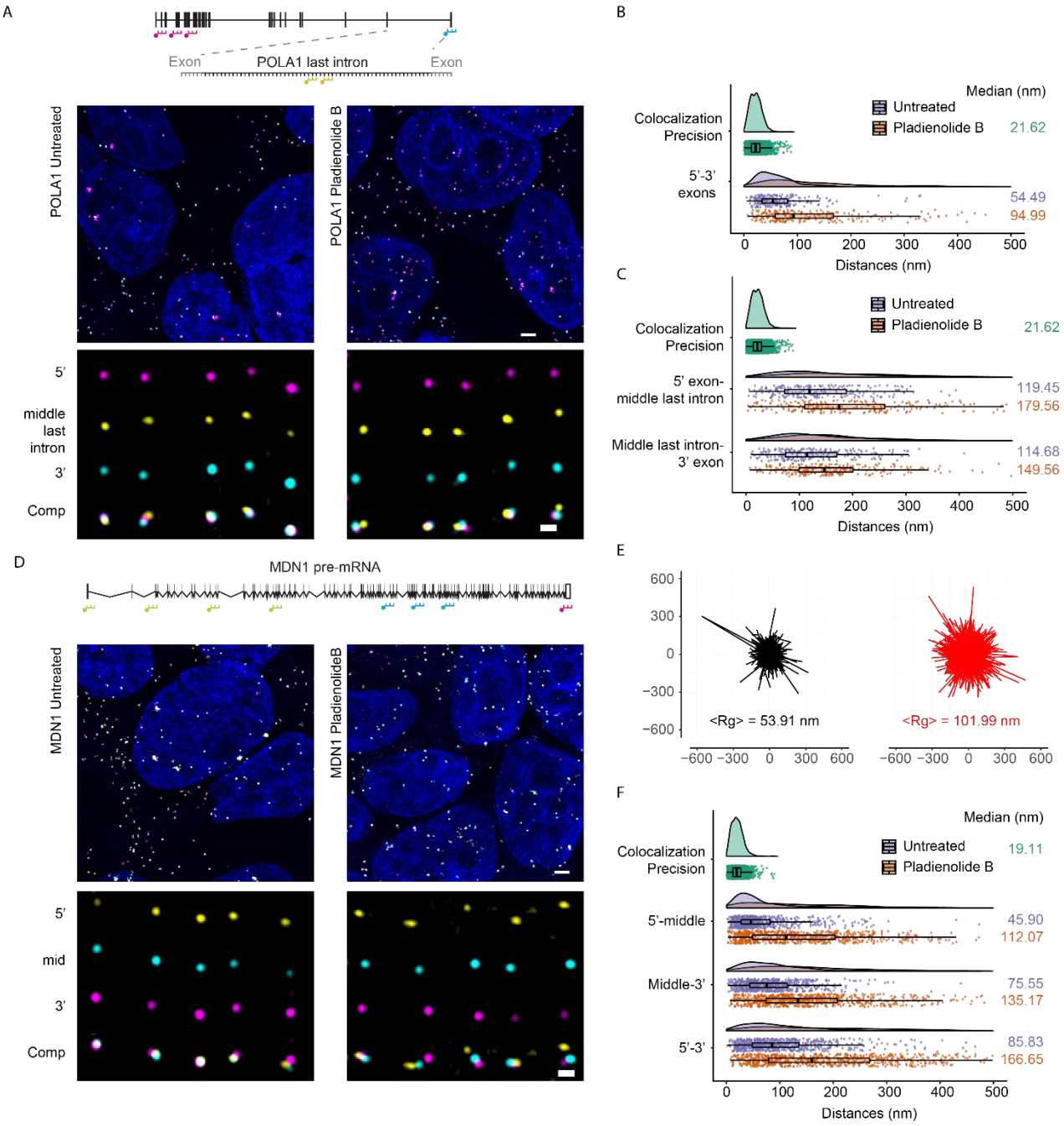
Supplementary Figure 3-4: Pladienolide B treatment alter intron organisation

Scatter plot showing distances for individual introns for intron 35 for mock (black) and Pladienolide B (red) treated HEK293T cells. POLA1 intron 35 imaged using probes hybridising to 5'(magenta), 3' (cyan) regions and either ***(A)*** *1/3rd or* ***(B)*** *2/3rdregions (Probes Set#13 and #14 respectively, Table S 3-2). Frequency distribution is shown on top and on the right.*



Supplementary Figure 3-5: Organisation of long AHNAK exon

(A) *smFISH images in HEK293T cells for probes hybridising to 5'(yellow), middle (magenta), 3' (cyan) regions of AHNAK mRNAs (Probes Set#15, Table S 3-2).in mock and treated with 100nM Pladeienolide B for 4h. Nuclei are visualised by DAPI staining (blue). Magnified images of individual RNAs are shown at the bottom. Schematic position of probes shown on top,* ***(B)*** *Projections of superimposed conformations from A with their centres of mass. The radius of gyration is shown at the bottom* ***(C)*** *Raincloud plots for distances between different regions used in A. Individual plots show distance distribution of co-localization precision and distances for AHNAK in mock and Pladienolide B treated cells shown as violin plots. The box plot shows the first quartile, median and third quartile and the individual RNAs shown as spots overlayed on top of the box plots. Median distances are shown on the right. Scale bars, 2 µm in larger images, and 500 nm in zoomed-in images*



Supplementary Figure 3-6: mRNA organisation upon Pladienolide B treatment

(A, D) *smFISH images in HEK293T cells in mock and treated with 100nM Pladeienolide B for4h. Nuclei are visualised by DAPI staining (blue). Magnified images of individual RNAs are shown at the bottom. Schematic position of probes shown on top, Probes hybridising to* ***(A)*** *5' exons (magenta), 3'exons (cyan) and middle region of intron 36 (yellow) of POLA1 gene (Probes Set#5, Table S 3-2). and (D) to 5'(yellow), middle (magenta), 3' (cyan) exonic regions of MDN1 mRNAs*

(Probes Set#2, Table S 3-2). ***(B,C,F))*** *Raincloud plots for distances between different regions used in (A),(D). Individual plots show distance distribution of co-localization precision and distances between different regions as violin plots. The box plot shows the first quartile, median and third quartile and the individual RNAs shown as spots overlayed on top of the box plots. Median distances are shown on the right.* ***(E)*** *Projections of superimposed conformations from D with their centres of mass. The radius of gyration is shown below. Scale bars, 2 μm in larger images, and 500 nm in zoomed-in images*

3.2.10 Supplementary Tables

Table S 3-1: List of smFISH probes used

MDN1 exons 5’	tcgttcttggctgcgattaa	POLA1 intron 36 tiling	gtgaaaaagcagttagtggc
MDN1 exons 5’	taaggtactcaggacacact	POLA1 intron 36 tiling	agggtgacctgaagaatagt
MDN1 exons 5’	cacagtacagtccttatcca	POLA1 intron 36 tiling	cccatagcactttgtgagaa
MDN1 exons 5’	agcaaatccaaaaggagagg	POLA1 intron 36 tiling	cagcaacagtttttcgatgg
MDN1 exons 5’	ttgaaagactggggatgtgt	POLA1 intron 36 tiling	gtgtgctttacctgttatta
MDN1 exons 5’	gcatctgaactctctaggaa	POLA1 intron 36 tiling	aaacttgggtagtccttctg
MDN1 exons 5’	gtgctcttcattcatacagg	POLA1 intron 36 tiling	gccaggttgtgatatgaatt
MDN1 exons 5’	cctcaacctgaaatggatca	POLA1 intron 36 tiling	ctttctcgttcttgtagtga
MDN1 exons 5’	aaaaccaaggccttctccaa	POLA1 intron 36 tiling	atgctgccatcagacaatac
MDN1 exons 5’	aaagggagacttctggattg	POLA1 intron 36 tiling	ccactggaatagccttttaa
MDN1 exons 5’	tcagacgaaacaagatgtcc	POLA1 intron 36 tiling	aaccatggaaagccttccac
MDN1 exons 5’	accagcacataagacctaag	POLA1 intron 36 tiling	agtatacctgtatctaaccc
MDN1 exons 5’	gaagacttttgcagacagac	POLA1 intron 36 tiling	agcactaattttggggagga
MDN1 exons 5’	acagcattctgagaagcaac	POLA1 intron 36 tiling	tgatgggctactaagtctgt
MDN1 exons 5’	tcctattggtccttccaaca	POLA1 intron 36 tiling	ggcctttatacagggagaat
MDN1 exons 5’	cctgtcactgcagctaaata	POLA1 intron 36 tiling	atttctttcaaagcaccacc
MDN1 exons 5’	aagctggactttgagaagct	POLA1 intron 36 tiling	gatgtatcctgtaggatgga
MDN1 exons 5’	acatctgtgcagcgatacat	POLA1 intron 36 tiling	acatggtagggtaattcctg
MDN1 exons 5’	atatcctccagaaggatcca	POLA1 intron 36 tiling	tagtccatgtctggaatagc
MDN1 exons 5’	aagagctctccattctccaa	POLA1 intron 36 tiling	cattttgaggacactgcttc
MDN1 exons 5’	aaatccaggtgccactttca	POLA1 intron 36 tiling	gcttacatcatttgagccaa
MDN1 exons 5’	caactacgcccgcaggaaag	POLA1 intron 36 tiling	atgcttccttttatctaggaa
MDN1 exons 5’	agcaagaagtgctccatgac	POLA1 intron 36 tiling	gctacattagcgtatacagc
MDN1 exons 5’	caagaacctgcccaactcac	POLA1 intron 36 tiling	gaaaccacaccagccaaagg
MDN1 exons 5’	cgatcttgaggtgtccacac	POLA1 intron 36 tiling	tgggttgaggttggctaacc
MDN1 exons 5’	aatggcttcggcattccttt	POLA1 intron 36 tiling	atctggagcagagagctaaa
MDN1 exons 5’	gttcatgcagatcatggttg	POLA1 intron 36 tiling	aggctagccaaccaaataac
MDN1 exons 5’	gtttgctcatcgacacacat	POLA1 intron 36 tiling	cagccttttttattcagttga
MDN1 exons 5’	gacatcaggatggttaccaa	POLA1 intron 36 tiling	cacccaaaggaatcagcagt
MDN1 exons 5’	aatatctcagggcaaacggg	POLA1 intron 36 tiling	acatgcctcaagatattttt
MDN1 exons 5’	tacgtccatagcgtactgga	POLA1 intron 36 tiling	ggaggggggctggttaaaaaa
MDN1 exons 5’	gcagaaacttgaaggctgct	POLA1 intron 36 tiling	tggaaggcaatctaagcaga
MDN1 exons 5’	cggaacacagactgctcctg	POLA1 intron 36 tiling	gctacctgtgtatgattcaa
MDN1 exons 5’	aaggtgtcatggcttctgag	POLA1 intron 36 tiling	aatgtactggtgaagtgggg
MDN1 exons 5’	acaattggctgtataccagc	POLA1 intron 36 tiling	ataaacctgtcactgcagtg
MDN1 exons 5’	acaccacaaacagctgtcac	POLA1 intron 36 tiling	gacatcacatagcaaggctt
MDN1 exons 5’	tcttactgtcagtctgatct	POLA1 intron 36 tiling	tagctgagccatgtcataac
MDN1 exons 5’	gaagaactcctattaccacc	POLA1 intron 36 tiling	aatacacagagtgcttccag

MDN1 exons 5’	ctgccacacaaactctccag	POLA1 intron 36 tiling	gacctccgtacataatttct
MDN1 exons 5’	cagaaaccacgtctaagggg	POLA1 intron 36 tiling	ggtacagtcacaatcatgca
MDN1 exons 5’	caatttcctccacagctcaa	POLA1 intron 36 tiling	tatgcacccattacagatga
MDN1 exons 5’	atgactgtttagcggtcgat	POLA1 intron 36 tiling	aattgtgttcctgagtcact
MDN1 exons 5’	ccaggtgaattttggtccaa	POLA1 intron 36 tiling	acaatggggactgcctaatg
MDN1 exons 5’	cagttctctcttatccaggt	POLA1 intron 36 tiling	taaactatgccctgcatcta
MDN1 exons 5’	gtttctctccagtaagttgg	POLA1 intron 36 tiling	gccaaaattctattctagcg
MDN1 exons 5’	gaactatcactccaagagtg	POLA1 intron 36 tiling	tatacagctgaggctagagg
MDN1 exons 5’	aacttcttcaggtgcctgtt	POLA1 intron 36 tiling	atttttactcttaacctccc
MDN1 exons 5’	ctcaagggttggtcttttgt	POLA1 intron 36 tiling	gccagtctttttctaaactg
MDN1 exons 5’	gggcaatcctattacaccaa	POLA1 intron 36 tiling	aaatgactgctggcatctca
MDN1 exons 5’	ctcagaaagcattgctgtga	POLA1 intron 36 tiling	ctttatcccaaaagacttca
MDN1 exons 5’	gcttccaataacttctgcca	POLA1 intron 36 tiling	gaaggcatggtgtgttaagc
MDN1 exons 5’	tttgctgacacacactgcaa	POLA1 intron 36 tiling	ctagtgaatgtcctcactga
MDN1 exons 5’	atggtagaggttttgccagt	POLA1 intron 36 tiling	attgtgctggttaggataca
MDN1 exons 5’	cctgtaatgtgagccaagta	POLA1 intron 36 tiling	agagacttgccattaagtgg
MDN1 exons 5’	attgacaaccctcaaacggt	POLA1 intron 36 tiling	acagcagttgactgttgaga
MDN1 exons 5’	gcaagtctgcagtatcactt	POLA1 intron 36 tiling	attcaacctgtcaaagctgc
MDN1 exons 5’	atggtccaccggtttataac	POLA1 intron 36 tiling	gaacgtatcacccactattg
MDN1 exons 5’	gtaagggtagccaaataagc	POLA1 intron 36 tiling	atgatggttaatgggaaggt
MDN1 exons 5’	aagagttcctcaaatgcctc	POLA1 intron 36 tiling	ctgggtattatagttacctc
MDN1 exons 5’	gccccaagaacgtaaagttt	POLA1 intron 36 tiling	ttctcaccttgctttttca
MDN1 exons 5’	ctgtctgtaacaggtctgaa	POLA1 intron 36 tiling	tttcttatccctgacttcag
MDN1 exons 5’	ttagtctcaggagatcatgc	POLA1 intron 36 tiling	cgtttcttgtgtttttgacc
MDN1 exons 5’	agcagacttgtgtacatgct	POLA1 intron 36 tiling	ctgcgtgtgaagctcaaatt
MDN1 exons 5’	cactgtcttttccatccttg	POLA1 intron 36 tiling	agccacactattatatgttt
MDN1 exons 5’	ccaaatgcttcccatttctc	POLA1 intron 36 tiling	aattcacccttacactctc
MDN1 exons 5’	ttgggcatggttgagtctaa	POLA1 intron 36 tiling	gatctgccaacataccttac
MDN1 exons middle	aagcagcaagattgaccaca	POLA1 intron 36 tiling	aacatgatcacggactccac
MDN1 exons middle	ggactagtgcatcacaaagt	POLA1 intron 36 tiling	aaatactaggtctcctctct
MDN1 exons middle	actggcatcagacaccattc	POLA1 intron 36 tiling	aaaggctcatttttcccatg
MDN1 exons middle	accgcagagaacctaagatc	POLA1 intron 36 tiling	gtgtctagaaaaccattgct
MDN1 exons middle	ggcatctacttttactgtgt	POLA1 intron 36 tiling	caaagaccaaggccattttc
MDN1 exons middle	tgccaatggagggcaagaag	POLA1 intron 36 tiling	ggtagctgaatttgctgaga
MDN1 exons middle	ctggtggaccagatgtttta	POLA1 intron 36 tiling	accttagctctagaactctt
MDN1 exons middle	aattcatcagaagtcggggg	POLA1 intron 36 tiling	aaggcatgctgtgccaagac
MDN1 exons middle	ctgagacagtctgaacttct	POLA1 intron 36 tiling	ggcttattctcttgatgagg
MDN1 exons middle	tccccagacaattctgaata	POLA1 intron 36 tiling	ccataaatcattagccttgt
MDN1 exons middle	cttctttataccagcgaagc	POLA1 intron 36 tiling	ttggggacatcatctttagt
MDN1 exons middle	aaacggtcgtcccaggaact	POLA1 intron 36 tiling	aaatgcatagtcgtcacctg
MDN1 exons middle	ccaccagcttgtccttaaaa	POLA1 intron 36 tiling	ggtaagagtcggatctgtta
MDN1 exons middle	ggaccttcagttgtgaaaaa	POLA1 intron 36 tiling	aaacctgacagaggtgggtg

MDN1 exons middle	ctgatggcaaggactttgtt	POLA1 intron 36 tiling	accaataacgtagcatagct
MDN1 exons middle	agacggttaatgtcttcctg	POLA1 intron 36 tiling	ggcataatgcccactttaaa
MDN1 exons middle	ccactgagaagcaaccactt	POLA1 intron 36 tiling	agattggttttccattgtgc
MDN1 exons middle	cttgcaggagacttttcttt	POLA1 intron 36 tiling	aaattgcgtttgggcttctc
MDN1 exons middle	atttgctctgaggatcagtc	POLA1 intron 36 tiling	accatatctttcactggcaa
MDN1 exons middle	ttcatctaggctgacatctt	POLA1 intron 36 tiling	aaactgaaactcgccatcct
MDN1 exons middle	ctgagcatgcacaaaattct	POLA1 intron 36 tiling	tgcacatctcagattcttct
MDN1 exons middle	cctttggctttcagttctaa	POLA1 intron 36 tiling	ccaaatgggggttattaga
MDN1 exons middle	ctccagaaaaccaagtgaga	POLA1 intron 36 tiling	tactgactccaatggctgac
MDN1 exons middle	aggaagcttcatcatgcttt	POLA1 intron 36 tiling	cctcttttcaactcttctta
MDN1 exons middle	gtcaagtctggatgggacag	POLA1 intron 36 tiling	ggttgatcactaacatagcc
MDN1 exons middle	tcctggtgaggtggattacg	POLA1 intron 36 tiling	caaggcactgtcaaacaggt
MDN1 exons middle	attgcaggccacaactgaac	POLA1 intron 36 tiling	atctgtggataacctgctac
MDN1 exons middle	tttgtaccgccaaagcatag	POLA1 intron 36 tiling	cagtttctgagaccatgtta
MDN1 exons middle	gtgccataaaatcagctgtc	POLA1 intron 36 tiling	cacaggctgacggattagag
MDN1 exons middle	ctgcatcttctgagacaagc	POLA1 intron 36 tiling	acatattcaaccagtgttca
MDN1 exons middle	tttatctggggttgttgatt	POLA1 intron 36 tiling	atcctgggaagtttatttgt
MDN1 exons middle	agatgaggtgacttatctcc	POLA1 intron 36 tiling	atttaactctgttacctgca
MDN1 exons middle	acaggtgtgtgataaagaca	POLA1 intron 36 tiling	ctgtgtttgaaattgctggt
MDN1 exons middle	atctcgaagctcttgaggtg	POLA1 intron 36 tiling	aatgctggaagacaagctta
MDN1 exons middle	tgatgatgcagcaaggacca	POLA1 intron 36 tiling	aacgaaagaggggactcggg
MDN1 exons middle	aatttcttcaggagacacct	POLA1 intron 36 tiling	ttagccaatgcctagaggac
MDN1 exons middle	ataactcggaccacaaagat	POLA1 intron 36 tiling	cagtgatcaatgtgtctatc
MDN1 exons middle	agtactgctccagaaagaca	POLA1 intron 36 tiling	agggtagcctagatactaat
MDN1 exons middle	agtactctggatttgtggtc	POLA1 intron 36 tiling	cccctatgaaaacgatggaa
MDN1 exons middle	ggcaaagggttccacattag	POLA1 intron 36 tiling	ctgtggaacttaatggccat
MDN1 exons middle	tccaaaacagacttgggtgc	POLA1 intron 36 tiling	agcaggtggtagcattatag
MDN1 exons middle	tgggtctattgagattgccg	POLA1 intron 36 tiling	acccacagttcaaagaacta
MDN1 exons middle	tcaaagcagcacttagagaa	POLA1 intron 36 tiling	cacagattctactattcctt
MDN1 exons middle	tctccagctgctggttaaga	POLA1 intron 36 tiling	gtgcctgcattagatcataa
MDN1 exons 3’	aagtctcactttggactctt	POLA1 intron 36 tiling	tgacaccctgacatctgatc
MDN1 exons 3’	aatgtgaccttctgaccaca	POLA1 intron 36 tiling	aattgctgcagtccattcac
MDN1 exons 3’	aaaagggagcacctgggtaa	POLA1 intron 36 tiling	acaaggttaactaggttcgt
MDN1 exons 3’	agcattctgtaggctgtaag	POLA1 intron 35 1/3rd	tcagtgtttgtgattatact
MDN1 exons 3’	tggtataaaaacctcagccc	POLA1 intron 35 1/3rd	ttcacagtgtcactttcaca
MDN1 exons 3’	actcttctctagttacgagt	POLA1 intron 35 1/3rd	atccccaaatgccttgaaaa
MDN1 exons 3’	cttccaaggcagggaagaag	POLA1 intron 35 1/3rd	gtcctgttgaatgctgaact
MDN1 exons 3’	aagaaaacaggcagctgggc	POLA1 intron 35 1/3rd	aaatgcatctgcctacatct
MDN1 exons 3’	acaaaggactgtcagagtcc	POLA1 intron 35 1/3rd	aaggatagcacagagggcct
MDN1 exons 3’	aaaagggcagctccctttag	POLA1 intron 35 1/3rd	ggaaggtaaatgcaggaaca
MDN1 exons 3’	gcaaggcagagcttagaaca	POLA1 intron 35 1/3rd	ttgcactccattgcatagaa
MDN1 exons 3’	tttgggcacacactatgggc	POLA1 intron 35 1/3rd	ttcaggactgccaataaaca

MDN1 exons 3’	ctgtcttggccacttgacag	POLA1 intron 35 1/3rd	ccattcagtacctctgcaaa
MDN1 exons 3’	cctcatactctccagaaacg	POLA1 intron 35 1/3rd	ggaggatttggaaatctcag
MDN1 exons 3’	tctaagagaaggtagttcct	POLA1 intron 35 1/3rd	aaagagtgctgccctaaaac
MDN1 exons 3’	ctataatgtccagttgcttt	POLA1 intron 35 1/3rd	atttgggactcagaaatcct
MDN1 exons 3’	tttttatagatgacctgggc	POLA1 intron 35 1/3rd	tcaaggttatgtctgaagga
MDN1 exons 3’	tttttacacagcccaaggat	POLA1 intron 35 1/3rd	cattccaaggacacacagcc
MDN1 exons 3’	gaggatactgaaaagccact	POLA1 intron 35 1/3rd	agcataaaatgatctctgct
MDN1 exons 3’	cattgcatagtctcccgaag	POLA1 intron 35 1/3rd	cgtcacaggggaagaaggat
MDN1 exons 3’	ataaaggggcaatcaccttc	POLA1 intron 35 1/3rd	aaggaatctctcctccattt
MDN1 exons 3’	tacaacaacagggaccatgg	POLA1 intron 35 1/3rd	attgagaatcagtagcctgt
MDN1 exons 3’	agtgtgaggaatcactcttc	POLA1 intron 35 1/3rd	tccataggcaagatcttagt
MDN1 exons 3’	tggctcagtccagcttgaaa	POLA1 intron 35 1/3rd	agtaaggatactgctatgct
MDN1 exons 3’	cgtttctttcccagaatgag	POLA1 intron 35 1/3rd	atttttttcccctaaagtct
MDN1 exons 3’	atagatggagctgctgagtt	POLA1 intron 35 1/3rd	ggctagactgtgagctgaaa
MDN1 exons 3’	atcagtttctttcgactgga	POLA1 intron 35 1/3rd	ggtttctctggttctgtaat
MDN1 exons 3’	ggccaagtaaaaactgccta	POLA1 intron 35 1/3rd	ctgtacatgcatgtacaggg
MDN1 exons 3’	caagtattcagcactgcttt	POLA1 intron 35 1/3rd	gtgcttctcatgactttcag
MDN1 exons 3’	agtagaacagagcacacagt	POLA1 intron 35 1/3rd	cacacccatctgagacaatg
MDN1 exons 3’	atcatgacatactgcctaca	POLA1 intron 35 1/3rd	cttcaacacacctgcatatg
MDN1 exons 3’	tcgcagacttcacagtgtaa	POLA1 intron 35 1/3rd	ttgttaattacttcccacca
MDN1 exons 3’	atctgtgtctttgatgacca	POLA1 intron 35 1/3rd	gcagaagcacagagaaggg a
MDN1 exons 3’	tacatgctttgggacacttg	POLA1 intron 35 1/3rd	gagtagaaggcaccaaatcc
MDN1 exons 3’	aagatcagtcccctagcata	POLA1 intron 35 1/3rd	actggagcatcttctaaata
MDN1 exons 3’	ctgactgactgatccagcag	POLA1 intron 35 1/3rd	cctcctaaggctgttttcat
MDN1 exons 3’	gcacagcatcaactagtaac	POLA1 intron 35 1/3rd	ctgtttctaagtagatgtca
MDN1 exons 3’	gaagtaggaggggatcatgt	POLA1 intron 35 1/3rd	gctctaaagatccaacttca
MDN1 exons 3’	cctttgtagtaagagcaaca	POLA1 intron 35 1/3rd	actggagtctaacagtcaca
MDN1 exons 3’	cagcctaccatggacataaa	POLA1 intron 35 1/3rd	gaagggcatagtctgatacc
MDN1 exons 3’	ctgcaaagccagcatattat	POLA1 intron 35 1/3rd	cctttaatctataaccatcc
MDN1 exons 3’	gcctccttataaggctacac	POLA1 intron 35 1/3rd	atttcaagatttgccctaga
MDN1 middle Odd	aagcagcaagattgaccaca	POLA1 intron 35 1/3rd	aacagagtatggagcacacc
MDN1 middle Odd	actggcatcagacaccattc	POLA1 intron 35 1/3rd	ggaaggacaaggcataagg a
MDN1 middle Odd	ggcatctacttttactgtgt	POLA1 intron 35 1/3rd	ggaagaggagagggaagac t
MDN1 middle Odd	ctggtggaccagatgtttta	POLA1 intron 35 1/3rd	ggtgctggttaatcatcaag
MDN1 middle Odd	ctgagacagtctgaacttct	POLA1 intron 35 1/3rd	caataaagcagggagacagt
MDN1 middle Odd	cttctttataccagcgaagc	POLA1 intron 35 1/3rd	tacgagtgtctttgtacatt
MDN1 middle Odd	ccaccagcttgtccttaaaa	POLA1 intron 35 1/3rd	atgttccaatcagtctttct
MDN1 middle Odd	ctgatggcaaggactttgtt	POLA1 intron 35 1/3rd	caggatgctggagatagaca
MDN1 middle Odd	ccactgagaagcaaccactt	POLA1 intron 35 1/3rd	aacgtttttttgtttcccacg
MDN1 middle Odd	atttgctctgaggatcagtc	POLA1 intron 35 1/3rd	aaaggcagaacccattgttt

MDN1 middle Odd	ctgagcatgcacaaaattct	POLA1 intron 35 1/3rd	cttagttccccatattcaag
MDN1 middle Odd	ctccagaaaaccaagtgaga	POLA1 intron 35 1/3rd	gttgttttgcttcaaaggtc
MDN1 middle Odd	gtcaagtctggatgggacag	POLA1 intron 35 1/3rd	ctcttcagcactgatgtatt
MDN1 middle Odd	attgcaggccacaactgaac	POLA1 intron 35 1/3rd	ctttactaactactttaccc
MDN1 middle Odd	gtgccataaaatcagctgtc	POLA1 intron 35 1/3rd	ccttactaactactttatcc
MDN1 middle Odd	tttatctggggttgttgatt	POLA1 intron 35 2/3rd	ccagtaattttaaaccagga
MDN1 middle Odd	acaggtgtgtgataaagaca	POLA1 intron 35 2/3rd	caccttgctcagatatactg
MDN1 middle Odd	tgatgatgcagcaaggacca	POLA1 intron 35 2/3rd	ccaatatgactcctctgtaa
MDN1 middle Odd	ataactcggaccacaaagat	POLA1 intron 35 2/3rd	atctgcaggcctgtaaggag
MDN1 middle Odd	agtactctggatttgtggtc	POLA1 intron 35 2/3rd	aagccactggaggcagtatg
MDN1 middle Odd	tccaaaacagacttgggtgc	POLA1 intron 35 2/3rd	aataggtgggtggggcttaa
MDN1 middle Odd	tcaaagcagcacttagagaa	POLA1 intron 35 2/3rd	aatactgttccccttcagaa
MDN1 middle Even	ggactagtgcatcacaaagt	POLA1 intron 35 2/3rd	gtatgtgaatttgctaagcc
MDN1 middle Even	accgcagagaacctaagatc	POLA1 intron 35 2/3rd	agcaaattggtagatcatca
MDN1 middle Even	tgccaatggagggcaagaag	POLA1 intron 35 2/3rd	ttatgctcttcaacacaggc
MDN1 middle Even	aattcatcagaagtcggggg	POLA1 intron 35 2/3rd	ccacagaatgcagcttcaaa
MDN1 middle Even	tccccagacaattctgaata	POLA1 intron 35 2/3rd	ggatatctttggatagagcc
MDN1 middle Even	aaacggtcgtcccaggaact	POLA1 intron 35 2/3rd	gaacccctttttcttttagat
MDN1 middle Even	ggaccttcagttgtgaaaaa	POLA1 intron 35 2/3rd	gttcacaatattttcttggt
MDN1 middle Even	agacggttaatgtcttcctg	POLA1 intron 35 2/3rd	tttcatttgtccaggtgagt
MDN1 middle Even	cttgcaggagacttttcttt	POLA1 intron 35 2/3rd	ctcaagttctcattatctgt
MDN1 middle Even	ttcatctaggctgacatctt	POLA1 intron 35 2/3rd	ggaaatggggatgactggtt
MDN1 middle Even	cctttggctttcagttctaa	POLA1 intron 35 2/3rd	ttttgctgttagcaacactt
MDN1 middle Even	aggaagcttcatcatgcttt	POLA1 intron 35 2/3rd	agataaagtaatgcaccccc
MDN1 middle Even	tcctggtgaggtggattacg	POLA1 intron 35 2/3rd	ctcagcagtttttccaaagg
MDN1 middle Even	tttgtaccgccaaagcatag	POLA1 intron 35 2/3rd	gataatccatgtcctcattc
MDN1 middle Even	ctgcatcttctgagacaagc	POLA1 intron 35 2/3rd	gtggcactcaattactctga
MDN1 middle Even	agatgaggtgacttatctcc	POLA1 intron 35 2/3rd	ctggcacattagtcagtaca
MDN1 middle Even	atctcgaagctcttgaggtg	POLA1 intron 35 2/3rd	gcccatttatctcatttgaa
MDN1 middle Even	aatttcttcaggagacacct	POLA1 intron 35 2/3rd	gcaggtctgtttctctgaaa
MDN1 middle Even	agtactgctccagaaagaca	POLA1 intron 35 2/3rd	acagaagtgcctcaatctga
MDN1 middle Even	ggcaaagggttccacattag	POLA1 intron 35 2/3rd	caatcctcttggctctaatt
MDN1 middle Even	tgggtctattgagattgccg	POLA1 intron 35 2/3rd	gcaggctgtgttaacatgaa
MDN1 middle Even	tctccagctgctggttaaga	POLA1 intron 35 2/3rd	ccaatatcttatctgattcc
POLA1 exons 3’	tgtacagggacttgtcagaa	POLA1 intron 35 2/3rd	aattcttcaattcccaagcc
POLA1 exons 3’	taccggtaaaagcacagctg	POLA1 intron 35 2/3rd	aggttcaaggatcaacaaca
POLA1 exons 3’	gtgcacactccgcatcaaaa	POLA1 intron 35 2/3rd	cctctatctgggatttgaac
POLA1 exons 3’	tctcatgatcggtagtaagt	POLA1 intron 35 2/3rd	catatttagatgccattcca
POLA1 exons 3’	ctgtagtcctgcagaacttt	POLA1 intron 35 2/3rd	taataggcttcattctagcc
POLA1 exons 3’	ctctgctgtgttcttgagtt	POLA1 intron 35 2/3rd	tacatcccattgatgttctc
POLA1 exons 3’	tagccacttcgggacaagaa	POLA1 intron 35 2/3rd	aagtatgggaagtcctttga
POLA1 exons 3’	tttgctcagattcacttcgg	POLA1 intron 35 2/3rd	tatacatctccccagtctag

POLA1 exons 3’	ttaggatttcacggcacaac	POLA1 intron 35 2/3rd	cactcaattcctgtaacttt
POLA1 exons 3’	cttggttactcctgggattc	POLA1 intron 35 2/3rd	gtgttctctctcaactactt
POLA1 exons 3’	ggaagctgggatttttcaac	POLA1 intron 35 2/3rd	gctaatggattttccttgga
POLA1 exons 3’	caagggagaaacagatgctg	POLA1 intron 35 2/3rd	tgctttctacatgcgataga
POLA1 exons 3’	acacaaacatgagacacagt	POLA1 intron 35 2/3rd	tcccactgaaagcttaccta
POLA1 exons 3’	gactcaacatttttgcagcc	POLA1 intron 35 2/3rd	gcttgtatacaagatctggt
POLA1 exons 3’	ctcagaaaccgggtcttcag	POLA1 intron 35 2/3rd	gcactgcaaaaacagagcca
POLA1 exons 3’	gctactctcaatccaagtag	POLA1 intron 35 2/3rd	attcactaattccttgtcta
POLA1 exons 3’	gcctggggtcacttacattc	POLA1 intron 35 2/3rd	gtgattttcctccatgatat
POLA1 exons 3’	cataggctaaaggccctgag	POLA1 intron 35 2/3rd	tactgaatcacggtagaggt
POLA1 exons 3’	ttcagtcaggctctgagaag	POLA1 intron 35 2/3rd	ggtttggactttctttagga
POLA1 exons 3’	tgaaaaagcaaaacgtcagc	POLA1 intron 35 2/3rd	gagatcatttagtggagtcc
POLA1 exons 3’	ttagaccgggtttaattggc	POLA1 intron 35 2/3rd	ccagccatcataatgagaga
POLA1 exons 3’	actcctggatggctggagaa	POLA1 intron 35 2/3rd	aggacacagactctctatcc
POLA1 exons 3’	agacaagactgaaaaggaca	POLA1 intron 35 2/3rd	ctaaacatggctgctaggct
POLA1 exons 3’	agtgcaaggcttctaaatct	POLA1 intron 35 2/3rd	ccagctcagtcaatgaaata
POLA1 exons 3’	gagcaattcaacaaacaagc	POLA1 intron 36 1/4th	tgcagattattcgtgtttca
POLA1 exons 3’	cagtgtgtgtctgttggact	POLA1 intron 36 1/4th	ctccctttctatatcaaagt
POLA1 exons 3’	atgtgagtgtaaaacacctg	POLA1 intron 36 1/4th	attgttatctagatgccacc
POLA1 exons 3’	gcactttctatttaaggggc	POLA1 intron 36 1/4th	gtgtttgacttcattttcct
POLA1 exons 3’	ccctacacatgttaatggat	POLA1 intron 36 1/4th	gcttcctttatctaggaaat
POLA1 exons 3’	tgaatacaacacagtgatcc	POLA1 intron 36 1/4th	ttagcgtatacagcaccata
POLA1 exons 3’	ggccaattaagcatccctct	POLA1 intron 36 1/4th	ttcactttatctttgatgct
POLA1 exons 3’	agaaaaaaatagcaagcgcc	POLA1 intron 36 1/4th	tacttttccctagcctattt
POLA1 exons 3’	agacggtcctattgtgaaga	POLA1 intron 36 1/4th	ccagaaaacagttccttctt
POLA1 exons 3’	aaatacattttgctgctgcc	POLA1 intron 36 1/4th	tttcatcttcagtttgctca
POLA1 exons 3’	agagaggaaagactgccata	POLA1 intron 36 1/4th	gccaattttgttgaaatgca
POLA1 exons 3’	cagcataattgtacaagggg	POLA1 intron 36 1/4th	gggggaaaataatgtgcttt
POLA1 exons 3’	agaagaaggcacaacatact	POLA1 intron 36 1/4th	tattccttcattcatgacgc
POLA1 exons 5’	aactccctgaatctgacaga	POLA1 intron 36 1/4th	tccagtctccttgaaagatc
POLA1 exons 5’	ttgatttttttctcgccgg	POLA1 intron 36 1/4th	cttaatatctccccagttac
POLA1 exons 5’	tttctagggcttcttggcgc	POLA1 intron 36 1/4th	aagtaccacgtttgggttag
POLA1 exons 5’	ccagctttagccttttcag	POLA1 intron 36 1/4th	ggtctcagaaatccaaacca
POLA1 exons 5’	taaacacctgtgaagtcctc	POLA1 intron 36 1/4th	agaagaggacaggaggtgg g
POLA1 exons 5’	cctgaaccagcttcgaatac	POLA1 intron 36 1/4th	cagtcagctgagtgtcaagg
POLA1 exons 5’	caatccagtcatcatcctgg	POLA1 intron 36 1/4th	taaagttgcggaaggctgct
POLA1 exons 5’	ggcatcatcttcaaggtcat	POLA1 intron 36 1/4th	cgcttggaggacgtgcaaaa
POLA1 exons 5’	tcttgtctttattgcgtgct	POLA1 intron 36 1/4th	aaaccctctactactaccat
POLA1 exons 5’	cactgcgagcttctttacat	POLA1 intron 36 1/4th	acaggacacaggaatatcct
POLA1 exons 5’	tctgcagttttctttccagc	POLA1 intron 36 1/4th	aagagatgagcaccccaaga
POLA1 exons 5’	accatccttggacaagtcta	POLA1 intron 36 1/4th	gacaccccaggtaagatgag
POLA1 exons 5’	cctgtagaatgtcacctagc	POLA1 intron 36 1/4th	cagtcaagctatttctcctg

POLA1 exons 5’	tcagtatcattacaggtggt	POLA1 intron 36 1/4th	actattttataactccagct
POLA1 exons 5’	tgaagctccaatggatcttt	POLA1 intron 36 1/4th	tccttctaggcaaaaagggg
POLA1 exons 5’	tgtgcacagagaaaggattc	POLA1 intron 36 1/4th	tgggtaccaacaaggtgcag
POLA1 exons 5’	gggaagcaatttttcctgaa	POLA1 intron 36 1/4th	ggatggagtcagaaagtgct
POLA1 exons 5’	agttaatggaggctcctttc	POLA1 intron 36 1/4th	atgacacagttgaaggacat
POLA1 exons 5’	cagcacgtttaagaggaaca	POLA1 intron 36 1/4th	ggtatgggctaaatgtttga
POLA1 exons 5’	tctcgacctgtacatcatcg	POLA1 intron 36 1/4th	aagtccatgtggaatgagga
POLA1 exons 5’	tgactcctgctcttcttctg	POLA1 intron 36 1/4th	gtagacagattaaggtggga
POLA1 exons 5’	ggctcatcaaagtcaccatc	POLA1 intron 36 1/4th	cagcccaactctattagtaa
POLA1 exons 5’	agccataggctccaggtcca	POLA1 intron 36 1/4th	atcttctggagatacatgat
POLA1 exons 5’	tctctttgtcccaagccttg	POLA1 intron 36 1/4th	atttctaggctggaatttat
POLA1 exons 5’	ttgtttcacttcctctgctg	POLA1 intron 36 1/4th	gactgagcatcagtttgatt
POLA1 exons 5’	agtaggacacggtccctttc	POLA1 intron 36 1/4th	ctttcaaacttaatcccact
POLA1 exons 5’	catccgggagaaaacttcct	POLA1 intron 36 1/4th	gttgagacagtcctcatttg
POLA1 exons 5’	tgatcaatgtcccaacaaga	POLA1 intron 36 1/4th	aattctggagggcaaggagt
POLA1 exons 5’	tgcactgagaaactgctatc	POLA1 intron 36 1/4th	gggaaaacaaccagctgcta
POLA1 exons 5’	gactggaatccacttgaact	POLA1 intron 36 1/4th	catctggagcagagagctaa
POLA1 exons 5’	gccccttttaccaatgggag	POLA1 intron 36 1/4th	taaaggctagccaaccaaat
POLA1 exons 5’	agtggaatacttgttcctca	POLA1 intron 36 1/4th	cctccagacatttcaagata
POLA1 exons 5’	ggttggttgtactgatcctc	POLA1 intron 36 1/4th	ttcacgagagctttctggag
POLA1 exons 5’	ctcggctgattcaatccaaa	POLA1 intron 36 1/4th	tctgaaagtctcttagcggt
POLA1 exons 5’	atgacacaacagctcacatg	POLA1 intron 36 1/4th	gctgctcaactgaaagtcac
POLA1 exons 5’	agcgttcgctcgatattttt	POLA1 intron 36 1/4th	attgccattaatacgtgtgc
POLA1 exons 5’	gtttctttccccgtatttag	POLA1 intron 36 1/4th	ccagcctattagtagtattg
POLA1 exons 5’	gcatagttcttttccactgg	POLA1 intron 36 1/4th	cttgtgggctatttttctcc
POLA1 exons 5’	ctggaacatcaggtatctca	POLA1 intron 36 1/4th	tcaagacaggtatctccaga
POLA1 exons 5’	aatcttgaggaagctgtggc	POLA1 intron 36 1/4th	agagtatatcttgcaggtta
POLA1 exons 5’	agatgtgttggtcccaaata	POLA1 intron 36 1/4th	tatcgtgcaatgattactct
POLA1 intron 36 3’	tatcccatagccagtcaata	POLA1 intron 36 1/4th	ttttctaactcgctcagtct
POLA1 intron 36 3’	tgctcttcctcctaaaagac	POLA1 intron 36 3/4th	ttgaactgtgggtatcatct
POLA1 intron 36 3’	agaatgacaaccagcgtcac	POLA1 intron 36 3/4th	taaatcattagccttgtgca
POLA1 intron 36 3’	ctatttacgctgccacacag	POLA1 intron 36 3/4th	catctttagttattcatgca
POLA1 intron 36 3’	accaggttggttagaaaggg	POLA1 intron 36 3/4th	atctctaacaaagttgggga
POLA1 intron 36 3’	ctctgctgatgggatattga	POLA1 intron 36 3/4th	gacaaccgatccattaagga
POLA1 intron 36 3’	atagctcctgatgtgtttac	POLA1 intron 36 3/4th	gagctgtaagttggtgaagt
POLA1 intron 36 3’	aagagctgcggtaatacgga	POLA1 intron 36 3/4th	aggcattttggcaagtttta
POLA1 intron 36 3’	catctgatgactggggagag	POLA1 intron 36 3/4th	aatttttttctcccttcagtc
POLA1 intron 36 3’	ccgtgcatgtgaatgacaca	POLA1 intron 36 3/4th	cttaatactagttctctgct
POLA1 intron 36 3’	aaacaacaatggccccaacc	POLA1 intron 36 3/4th	gcggctaatttacattttct
POLA1 intron 36 3’	ctaatgatctgagtctggga	POLA1 intron 36 3/4th	cacttgccttctctaatact
POLA1 intron 36 3’	tacattcgcatttgctctcc	POLA1 intron 36 3/4th	acctgtggtaaaaatggctt
POLA1 intron 36 3’	gtctacaaaacacccggaaa	POLA1 intron 36 3/4th	atgactcagtcagcaacaga

POLA1 intron 36 3’	accggtcaggaaggaaacac	POLA1 intron 36 3/4th	attttaggtttcagtgacct
POLA1 intron 36 3’	gacatcagctgatcctgagg	POLA1 intron 36 3/4th	gttacgcgatgcttacaaca
POLA1 intron 36 3’	acatagaaatttgcaggcct	POLA1 intron 36 3/4th	tgtcttaagtggatcttgga
POLA1 intron 36 3’	ctgccactacatcactactc	POLA1 intron 36 3/4th	ccatctcagtgatgcagaaa
POLA1 intron 36 3’	tgaggagaggctcacttctg	POLA1 intron 36 3/4th	cttcagcacattattgatgg
POLA1 intron 36 3’	taaggtgtctggagaggctg	POLA1 intron 36 3/4th	ttccctgtatagagagctaa
POLA1 intron 36 3’	atgacgctgtgttcaggaac	POLA1 intron 36 3/4th	acttctgtacattttgctgg
POLA1 intron 36 3’	gaagagagtgatcacgtggg	POLA1 intron 36 3/4th	tgaccagagattccagtgaa
POLA1 intron 36 3’	atgctcaggtctgatctttt	POLA1 intron 36 3/4th	ttagcaacacatgggctgac
POLA1 intron 36 3’	taccagccagagaaaagacc	POLA1 intron 36 3/4th	ccatgggtcaaatcaaacca
POLA1 intron 36 3’	tgctcctttaacccattcat	POLA1 intron 36 3/4th	gtgtaattgcctggattcag
POLA1 intron 36 3’	actgtgtgaactcagggtat	POLA1 intron 36 3/4th	gtggacagtagaaccttaa
POLA1 intron 36 3’	ctgaaagaggcgagaggtga	POLA1 intron 36 3/4th	gacatttctcaagatgagga
POLA1 intron 36 3’	gcctagtaagcattttctag	POLA1 intron 36 3/4th	ttcagatgcaggatatgtgc
POLA1 intron 36 3’	agtctcaaagggagctatgg	POLA1 intron 36 3/4th	ccatttttgtttcttgtgtt
POLA1 intron 36 3’	ctggacgtggcaggaaacac	POLA1 intron 36 3/4th	gtgaaagcactcttcaggca
POLA1 intron 36 3’	cacatgcatgcaactgcttg	POLA1 intron 36 3/4th	cagctgtcccttgattaatg
POLA1 intron 36 3’	agacaaggtgagctctactc	POLA1 intron 36 3/4th	cagacttaattcagcaggcc
POLA1 intron 36 3’	actgatgtgaactgcaggtg	POLA1 intron 36 3/4th	cactcgcagtctgactctta
POLA1 intron 36 3’	ctagctgactcagtgcactg	POLA1 intron 36 3/4th	tcacctaaacacagaggcag
POLA1 intron 36 3’	ttcagtattccaccagagaa	POLA1 intron 36 3/4th	ttacacttgctaaagtggcc
POLA1 intron 36 5’	cagtaggtctggcttttata	POLA1 intron 36 3/4th	tctcacctgagtctatacat
POLA1 intron 36 5’	ctgcttctgagacagcaatg	POLA1 intron 36 3/4th	cccaaactcaaggggcaaaa
POLA1 intron 36 5’	gccttgctttaatagaggac	POLA1 intron 36 3/4th	agatctggaggtcatttctc
POLA1 intron 36 5’	agagtccttttgtactgtt	POLA1 intron 36 3/4th	tcctccacagtgaatgatat
POLA1 intron 36 5’	gttgcctatgcttactttat	POLA1 intron 36 3/4th	atccctaaggggaaattacc
POLA1 intron 36 5’	agcgaacagcagagacacac	POLA1 intron 36 3/4th	tctccctattaatgcataca
POLA1 intron 36 5’	gcctatatactctcttagga	POLA1 intron 36 3/4th	atttgtgactctctctct
POLA1 intron 36 5’	tgcaacagctgatcatacgg	POLA1 intron 36 3/4th	aggcatagacacttaccata
POLA1 intron 36 5’	ccgcaagtcttgtataagga	POLA1 intron 36 3/4th	ttttcctgacgagttctaca
POLA1 intron 36 5’	cttatgcggggaaatctagc	POLA1 intron 36 3/4th	aaatgcatagtcgtcacctg
POLA1 intron 36 5’	gcctatcttgtggatggaaa	POLA1 intron 36 3/4th	ggtaagagtcggatctgtta
POLA1 intron 36 5’	tctccaaggacagcaaaacc	POLA1 intron 36 3/4th	atgttgttaactggtcgtct
POLA1 intron 36 5’	atgctgtggagatagctcaa	POLA1 intron 36 3/4th	tcaaaaaccatcactgcctt
POLA1 intron 36 5’	atatgggctagacattgacc	POLA1 intron 36 3/4th	ttcagagagaaactcgaggt
POLA1 intron 36 5’	actgtctttaggaaaggtcc	POLA1 intron 36 3/4th	ctaggggtttctacagtgtg
POLA1 intron 36 5’	gccgtatagggaagaggaat	POLA1 intron 36 3/4th	aattctgtgcctgacagagc
POLA1 intron 36 5’	aaagactgattctgcctctc	POLA1 intron 36 3/4th	attaatttttcttccctggg
POLA1 intron 36 5’	agttgccagacctttgaaa	POLA1 intron 36 3/4th	aagagaacacagccttctcc
POLA1 intron 36 5’	aaatctccagtgactttcca	POLA1 intron 36 3/4th	gctcttgctgctaaatatca

POLA1 intron 36 5’	cagtgtcctctaagaggatg	POLA1 intron 36 3/4th	aagaaggaaacatccctctc
POLA1 intron 36 5’	gtgggtgtaataacgtgcac	POLA1 intron 36 3/4th	gtgtttattgtaactactgc
POLA1 intron 36 5’	gatcctgtcaatacttttcc	POLA1 intron 36 3/4th	gagtcctctgagatttggaa
POLA1 intron 36 5’	tcactcattcccaaactgta	POLA1 intron 35 3’	acagccaagtcatttgtaca
POLA1 intron 36 5’	gcctggagcatgtaaatttc	POLA1 intron 35 3’	ctgccttggaatgacatgta
POLA1 intron 36 5’	ataccctaataccacttagc	POLA1 intron 35 3’	aacatgttaccagtttctcc
POLA1 intron 36 5’	agttttggtttatatcccac	POLA1 intron 35 3’	agccctcagaacttttcaaa
POLA1 intron 36 5’	atgtgttaggaaacctctgc	POLA1 intron 35 3’	gtcctaccaaggtgaaacaa
POLA1 intron 36 5’	aggagagggcttatcagatg	POLA1 intron 35 3’	gctttggtgttttctttgag
POLA1 intron 36 5’	cctatgaactatgaccttgc	POLA1 intron 35 3’	ttgcgagctggggaacaatg
POLA1 intron 36 5’	gttgctaggtactgacagta	POLA1 intron 35 3’	aggggttagacacttgggaa
POLA1 intron 36 5’	ggggttaatccacattatgg	POLA1 intron 35 3’	aaactttctttccatcttcc
POLA1 intron 36 5’	tatgccatctgtgtacattc	POLA1 intron 35 3’	ggggagtcacattcaaatgc
POLA1 intron 36 5’	cctatgttcctgtaaaaccg	POLA1 intron 35 3’	cctgaggttatttagagtca
POLA1 intron 36 5’	aggaatgtcactcaacacct	POLA1 intron 35 3’	gtaggagacgtgtctgaact
POLA1 intron 36 5’	taatttcaccctattggtgg	POLA1 intron 35 3’	cttgccaagttccaattcaa
POLA1 intron 36 middle	ctgtggaacttaatggccat	POLA1 intron 35 3’	cacccatgtcaaaacactcg
POLA1 intron 36 middle	agcaggtggtagcattatag	POLA1 intron 35 3’	tatgatctctgagaagtacc
POLA1 intron 36 middle	ctattcattccacacctatt	POLA1 intron 35 3’	tggtaatcctaagtgggcaa
POLA1 intron 36 middle	tcctatgaatctgcagcata	POLA1 intron 35 3’	aacatcttgactatttgctc
POLA1 intron 36 middle	gtgagatttcatggggttac	POLA1 intron 35 3’	gattggctcacagactcaca
POLA1 intron 36 middle	ttgggaccattcttatatag	POLA1 intron 35 3’	ttgtggtgacaatccattgc
POLA1 intron 36 middle	acccacagttcaaagaacta	POLA1 intron 35 3’	atattgctataccatgagcc
POLA1 intron 36 middle	cacagattctactattcctt	POLA1 intron 35 3’	tgctatgtgcttcctaactg
POLA1 intron 36 middle	ctttagaactgaatgcctcc	POLA1 intron 35 3’	atttgcactgttaattggga
POLA1 intron 36 middle	atcagacatgccattgagtt	POLA1 intron 35 3’	actgcaaattcctatgcact
POLA1 intron 36 middle	gctctttcttcatcacgatg	POLA1 intron 35 3’	cctgctgtcaaaaacctgat
POLA1 intron 36 middle	tgagcaagagttggggcaag	POLA1 intron 35 3’	attactacttgcgaagctgc
POLA1 intron 36 middle	gatcacttacaatcctgtgc	POLA1 intron 35 3’	ataaatcctgccacttcgat
POLA1 intron 36 middle	ggctgtaacccctgaaagag	POLA1 intron 35 3’	tagcatatacacctttcata

POLA1 intron 36 middle	ccaaacctatgggttctcta	POLA1 intron 35 3’	ctaatggtttgtccctctta
POLA1 intron 36 middle	aggaacccaggttttctgac	POLA1 intron 35 3’	gggtaagaagggcagaatca
POLA1 intron 36 middle	ctcacagcacagtaatcttc	POLA1 intron 35 3’	cactctcagcagagtgtatc
POLA1 intron 36 middle	gtccctattttaagactgga	POLA1 intron 35 3’	tgcctctgcatttaagaagt
POLA1 intron 36 middle	actctttaaaggacaggcat	POLA1 intron 35 3’	tgtgtatcaacccttttcag
POLA1 intron 36 middle	gtggcatttactggatttca	POLA1 intron 35 3’	ctgtgtgtacatatttttct
POLA1 intron 36 middle	tgggattttgctgacatagt	POLA1 intron 35 3’	cctgactagttgccaaattt
POLA1 intron 36 middle	gcagggagttattagtttgg	POLA1 intron 35 3’	acgatggtagctatttgcag
POLA1 intron 36 middle	cttctacgctgccaaaacag	POLA1 intron 35 3’	ttcattgtgatggcaaccat
POLA1 intron 36 middle	ctccttcaaatttcacaact	POLA1 intron 35 5’	ctttagcattatgagactgt
POLA1 intron 36 middle	cttaagagtaacggcagcac	POLA1 intron 35 5’	ctctaccctcttctacatgg
POLA1 intron 36 middle	acttgactcaattagcacct	POLA1 intron 35 5’	tctacttctcaaagcaagca
POLA1 intron 36 middle	gtgcctgcattagatcataa	POLA1 intron 35 5’	cttcatagggcaaatactca
POLA1 intron 36 middle	tgacaccctgacatctgatc	POLA1 intron 35 5’	tatgcatatggtagtagggc
POLA1 intron 36 middle	aatgcctttgggtagagctt	POLA1 intron 35 5’	gaaagcatgtatctttctga
POLA1 intron 36 middle	ctagctagaaggtgtgtctg	POLA1 intron 35 5’	gtattgctcatgacatgtga
POLA1 intron 36 middle	gtctgtgaattggtggatct	POLA1 intron 35 5’	gaattagtcagtttcccttt
POLA1 intron 36 middle	agcatgaagtggcatatgac	POLA1 intron 35 5’	gtaattatgtagtggcaacc
POLA1 intron 36 middle	gcgcaaggcagtaagctaaa	POLA1 intron 35 5’	aatggaatggccatgtcttc
POLA1 intron 36 middle	ttcaatacttcttgtgtgct	POLA1 intron 35 5’	cccaaagacgatagcagttt
POLA1 intron 36 middle	tttatttggtgtcagctacc	POLA1 intron 35 5’	ctggctcttacaaatgggat
POLA1 intron 36 middle	cttgaaattacagattccct	POLA1 intron 35 5’	gggtacagaattgggaggat
POLA1 intron 36 middle	tgcctttcaaaatacgcagt	POLA1 intron 35 5’	accccacaaaataaaccatt

POLA1 intron 36 middle	aggtgaatgtgactggtgtc	POLA1 intron 35 5’	caattaatccagcagagggg
POLA1 intron 36 middle	tctgtgaggaactactcaga	POLA1 intron 35 5’	ctactcatctatgcagctac
POLA1 intron 36 middle	aaccatatgtttttccccat	POLA1 intron 35 5’	tacacctcaggtgtgtatat
POLA1 intron 36 middle	gcctatggaattatgagact	POLA1 intron 35 5’	ctttctagttgggagtcatt
POLA1 intron 36 middle	ggctgaggacagctacaaat	POLA1 intron 35 5’	agaagtgtttccagatgcca
POLA1 intron 36 middle	aggtttcttaccatcttcat	POLA1 intron 35 5’	gggaaggcatcttaattacc
POLA1 intron 36 middle	ggtatagatagcttaccagc	POLA1 intron 35 5’	gatgctcaaaggggtcaaca
POLA1 intron 36 middle	aattgctgcagtccattcac	POLA1 intron 35 5’	ggaaaatgctgtttgggtcg
POLA1 intron 36 middle	acaaggttaactaggttcgt	POLA1 intron 35 5’	gaacctgcaaacttctcact
POLA1 intron 36 middle	agtgatctgatataggggga	POLA1 intron 35 5’	ccttcttctcactaaaaacc
POLA1 intron 36 middle	ccctctactgggaattttaa	POLA1 intron 35 5’	aaacaacaggttcccaagtc
POLA1 intron 36 middle	gtttgacatccagtcagatt		
POLA1 intron 36 middle	acagtgtacaacggatacca		
POLA1 intron 36 middle	ctagcacatactgattctgc		
POLA1 intron 36 middle	tctggtaggcttttagtttg		
POLA1 intron 36 middle	ctgatccttcacttaccaag		
POLA1 intron 36 middle	aagaagccagatttgggctg		
POLA1 intron 36 middle	taactctatctggagaccac		
POLA1 intron 36 tiling	cctctctttgcttccaaaat		

Table S 3-2: Probe combinations used

Probes Set#	Experiment	Combination of probes/antibodies and dyes used (from Table S2)
1	MDN1 middle alternating probes	MDN1 middle odd – Cy5
		MDN1 middle even – Cy3
2	MDN1 5'-middle-3'	MDN1 5'+ MDN1 5' additional – Dy488
		MDN1 middle – Cy5
		MDN 3' – Dy550
3	POLA1 intron 36-exon	POLA1 5' exon Cy3
		POLA1 intron 36 5' Cy5
		POLA1 intron 36 middle Atto488
4	POLA1 intron 35-exon	POLA1 5' exon Cy3
		POLA1 intron 35 5' Cy5
5	POLA1 exons 5'-3' intron 36 middle	POLA1 5' exons -Cy3
		POLA1 3' exons- Cy5
		POLA1 intron 36 middle Atto488
6	POLA1 intron 36 5'-middle-3'	POLA1 intron 36 5' Cy5
		POLA1 intron 36 middle Atto488
		POLA1 intron 36 3' Cy3
7	POLA1 intron 36 tiling middle 3'	POLA1 intron 36 tiling Cy5
		POLA1 intron 36 middle Atto488
		POLA1 intron 36 3' Cy3
8	POLA1 intron 36 5' 1/4th middle	POLA1 intron 36 5' Cy5
		POLA1 intron 36 1/4th Cy3
		POLA1 intron 36 middle Atto488
9	POLA1 intron 36 1/4th middle-3/4th	POLA1 intron 36 1/4th Cy3
		POLA1 intron 36 middle Atto488
		POLA1 intron 36 3/4th Cy5
10	POLA1 intron 36 middle-3/4th-3'	POLA1 intron 36 middle Atto488
		POLA1 intron 36 3/4th Cy5

		POLA1 intron 36 3’ Cy3
11	POLA1 intron 36 5’-1/4th-3’	POLA1 intron 36 5’ Cy5
		POLA1 intron 36 1/4th Atto488
		POLA1 intron 36 3’ Cy3
12	POLA1 intron 36 5’-3/4th-3’	POLA1 intron 36 5’ Cy5
		POLA1 intron 36 3/4th Atto488
		POLA1 intron 36 3’ Cy3
13	POLA1 intron 35 5’-1/3rd-3’	POLA1 intron 35 5’ Cy5
		POLA1 intron 35 1/3rd Atto488
		POLA1 intron 35 3’ Cy550
14	POLA1 intron 35 5’-2/3rd-3’	POLA1 intron 35 5’ Cy5
		POLA1 intron 35 2/3rd Atto488
		POLA1 intron 35 3’ Cy550
15	AHNAK 5’-middle-3’	AHNAK 5’ – Dy488
		AHNAK middle – Cy5
		AHNAK 3’ – Cy3
16	POLA1 intron 36-exon	POLA1 5’ exon Cy3
		POLA1 intron 36 5’ Cy5

4. Discussion

4.1 General goals of the thesis

Despite their central role in modulating mRNA metabolism, mRNP organisation and factors that determine mRNP organisation in cells are poorly understood. The goal of my PhD thesis was to use single-molecule resolution microscopy approaches and biochemical methods to develop an RNA-centric view on mRNP organisation in cells. To achieve this, I developed an smFISH-based approach that allows for separating different regions of single mRNAs with high spatial precision. Using this, I visualised mRNP conformations in different cellular compartments, including the nucleus, cytoplasm, and stress granules. I will elaborate on the findings from two studies where I have used this approach – the first one mainly focusing on the organisation of cytoplasmic and nuclear mRNAs and the second on pre-mRNAs and introns, discussing them in the broader context of the current literature.

4.2 Article 1: Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway

4.2.4 Objectives and summary of results

The role of mRNP packaging and organisation in regulating mRNA metabolism is manifold. For instance, communication between the 5' and 3' ends of mRNAs is required to regulate their translational status and decay. This communication has been theorised to result from a closed-loop structure in the cytoplasm, keeping the ends in proximity. The most popular model for enabling this closed-loop structure is through interactions between eIF4E-eIF4G-PABPC1. However, to date, the presence of such structures has never been visualised in cells, and the organisation of specific cytoplasmic mRNPs at different steps in their lifetime has remained unclear. Similarly, nuclear mRNP organisation is required for regulating several functions, including preventing R-loop formation, RNA splicing and export. However, BR mRNPs remain the only nuclear mRNPs in higher eukaryotes whose structure has been determined. Overall, very little is understood about mRNPs conformations in cells and how they are altered in response to changes in cellular processes.

In this project, I visualised the spatial organisation of single mRNP molecules within the nucleus and cytoplasm of two human cell lines, HEK293 and U2OS. To determine the organisation of

mRNPs, I developed an imaging approach by combining single-molecule localisation microscopy, smFISH and structured illumination microscopy (SIM), a method with a high spatial resolution that can separate regions farther than ~20 nm apart. I used three different mRNAs with varying lengths to determine mRNP organisation in the cytoplasm and nucleus. Surprisingly, I found that cytoplasmic mRNPs have elongated conformations with their ends separated in space. This separation was found to depend on the length of the mRNA and its ribosome occupancy, with higher end-to-end distances observed for longer and highly translated mRNAs. Inhibition of translation by removal of ribosomes, either pharmacologically or with oxidative stress, resulted in mRNAs having a compact conformation. Finally, CRISPR cell lines where eIF4G-PABPC1 interaction was admonished showed very few phenotypes in translation or growth, and mRNA conformations within these cells were similar to what was observed in wild-type cells. On the other hand, nuclear MDN1 mRNAs have compaction levels between translating and non-translating cytoplasmic mRNA and an organisation that reflects a linear packaging of mRNPs. In the following sections, I will be discussing different observations for nuclear and cytoplasmic mRNPs from this study in the context of our current understanding of mRNA metabolism and RNA biophysical properties.

4.2.5 Nuclear mRNP organisation

4.2.5.1 Many nuclear mRNPs are linearly organised into rod-like conformations

EM studies with Balbiani ring mRNPs have provided a detailed insight into their assembly. This assembly begins co-transcriptionally and goes through multiple stages of compaction, resulting in a molecule with a linear organisation but a croissant-like shape folded onto itself (Figure 1-6) (Björk and Wieslander, 2015; Skoglund et al., 1983). A different organisation was observed for nuclear mRNPs purified from *S. cerevisiae*, which were found to have a rod-like shape when visualised using electron microscopy. These rod-like structures were found to have a constant width but varying length, with larger mRNPs found to contain longer mRNAs, suggesting their assembly possibly results from linear packaging of mRNPs (Batisse et al., 2009). Finally, the data from more than 450 mRNPs purified using the exon junction complexes showed they follow the rule of a linear organisation of nuclear mRNPs. Computational polymer analysis of the data

showed that this organisation would result in the formation of a flexible rod-like structure (Metkar et al., 2018). Our data for MDN1 mRNAs in the nucleus supports the existing linear nuclear mRNP organisation model and represents the first mRNPs whose conformations have been visualised in human cells. By targeting smFISH probes against three different regions, the 5', the middle and the 3' end, we were able to determine the relative distances between different regions of the mRNA. We found that the 5' and 3' ends were farther apart than the 5' - middle and middle-3' regions, suggesting a linear organisation of the mRNP, perhaps similar to the rod-like shapes of nuclear mRNPs purified from yeast and predicted for EJC containing mRNPs, including MDN1 (Supplementary Figure 2-7). Other nuclear mRNPs (POLA1, PRPF8, PITRM1, GART, DYNC1I2) also show end-to-end distances larger than those observed for cytoplasmic mRNAs in conditions of translation inhibition (Figure A-1). However, as we do not have probes targeting a third region within these mRNAs, we cannot ascertain whether these observations result from a linear mRNP organisation similar to MDN1.

Advantages of a linear mRNP organisation in the nucleus

A linear rod-like organisation provides multiple biological, biophysical and mechanical advantages to mRNPs. First, a linear organisation is likely to result from its co-transcriptional assembly, which is essential in preventing R-loop formation in cells. Second, a linear architecture provides the framework for a cell to organise mRNPs regardless of the length of an mRNA. This could have had a critical role during evolution, as the length of genes and transcripts have gradually increased with the increased complexity of organisms. Third, through their linear packaging, mRNP could be compacted to prevent regions of RNAs from being exposed, which could prevent attack from endonucleases or cause physical breakage of the RNA, ensuring that the functional integrity of the RNA is preserved during its diffusion in the nucleoplasm and export through the NPC. Fourth, a rod-like architecture resulting from linear packaging could facilitate the diffusion of the mRNP through the nucleoplasm. Recent studies have observed that diffusion of cylindrical rod-like particles is greater in complex polymeric environments, even when compared to spherical particles with a diameter the same as the minor axis of the cylindrical rod (Wang et al., 2018). The nucleoplasmic environment provides multiple avenues for weak interactions with the mRNPs, as is observed for mRNPs when diffusing through dense regions of the chromatin (Vargas et al., 2005). Finally, organising mRNPs into rods of uniform diameter could assist in their efficient

passage through the nuclear pore complex. The transit of globular particles through the pore is inversely correlated with their size (Hinde et al., 2017; Paci et al., 2020). Interestingly, for similarly sized objects, rod and worm-shaped nanoparticles were found to traverse more effectively through the NPC (Hinde et al., 2017). Consistent with this observation, BR mRNPs are unfolded at the nuclear baskets before their entry through the central channel of the NPC.

4.2.5.2 Expanding our understanding of nuclear mRNP organisation

Our observations of MDN1 mRNAs and the results from Metkar et al. present the first *in vivo* evidence for nuclear mRNP organisation in humans (Metkar et al., 2018). However, it is unclear how we could extrapolate our findings to determine the general principles of nuclear mRNP packaging and organisation. The human genome has more than 20,000 genes, with most expressing multiple isoforms. As a result, there is considerable heterogeneity between genes in their intron-exon structure and mRNAs in their sequence and secondary structures, impacting the final composition of different mRNPs. Furthermore, cellular processes like transcription, mRNP processing, and export could each influence the organisation of mRNPs.

Gene structure

MDN1 mRNA is 18,413 nt in length and contains 102 exons and 101 introns, with a relatively even distribution of introns and exons throughout the gene. Most exons are less than 700 nt in length, except for the last exon, which is nearly 1,600 nt. The presence of multiple exons within this long gene provides several avenues for the canonical deposition of the exon junction complex within the body of the mRNA and the formation of local SR protein-EJC multimeric complexes. Such multimerisation has been observed for MDN1 and ~450 other mRNPs within the RIPPLiT dataset, suggesting the role of EJC in organising the mRNP into a linear rod-like shape (Metkar et al., 2018). However, not all genes have an intron-exon structure similar to MDN1. For example, Balbiani ring mRNP, which to date remains the only nuclear mRNP with a well-defined 3D organisation, is a linearly compacted molecule that finally folds into a croissant-like shape with the ends in proximity (Björk and Wieslander, 2015).

Several factors could contribute to the observed differences between MDN1 and BR mRNPs. The Balbiani Ring mRNPs have a unique exon-intron structure, with a long exon 4, limiting the number of canonical depositions of EJCs within the mRNP (Figure 1-5). Consistent with this, the amount

of EJC proteins found on BR mRNPs has been observed to correlate with the number of exon-exon junctions (Björk et al., 2015). The lack of EJCs from the majority of the mRNP limits its capacity to form local EJC-SR protein-mediated heteromeric complexes required for the assembly of mRNPs into linear rods (Metkar et al., 2018; Singh et al., 2012). We have since visualised other mRNPs in the nucleus, among which AHNAK shows a vastly different conformation compared to MDN1. Similar to BR2 and BR6 genes, AHNAK has a gene structure with one long exon at the 3' end (18,173 nt), which constitutes most of the nucleotide content of the mRNA (Supplementary Figure 3-1). Despite the similarity in lengths of MDN1 and AHNAK mRNAs, the median end-to-end distances observed for nuclear AHNAK and MDN1 mRNPs are ~47 nm and ~86 nm, respectively, suggesting that AHNAK, like BR mRNPs, could fold onto itself deviating from the rod-like conformation (Figure A-2 and Figure 4-1). Whether this is limited to AHNAK or is a general characteristic of mRNAs incapable of forming EJC mediated multimers is still unclear. However, more than 80% of exons in humans are less than 200 nt in length, and with genes containing an average of 8.8 exons, the local EJC mediated local multimerisation could have an important role in organising most mRNAs into linear rods (Sakharkar et al., 2004).

Curiously, purified mRNPs from *S. cerevisiae* have also been observed to have a similar rod-like shape. Budding yeasts mRNPs do not contain EJCs and lack the diversity in SR and SR-like proteins, they are shorter on average than their human counterparts, and most are devoid of introns, indicating that co-transcriptional assembly of mRNPs could by itself be sufficient for mRNPs to assemble into rod-like structures, and perhaps EJC mediated mRNP organisation might not be required for all mRNAs (Batisse et al., 2009). Whether this is true for human mRNAs either containing very few exons or single exon genes is yet to be determined. Alternatively, it is possible that EJC-mediated mRNP organisation could be important for longer mRNAs. While shorter mRNPs could maintain their shape due to their rigidity, it could become harder for mRNPs containing longer mRNAs to do the same.

In addition to intron-exon structure, MDN1 and BR mRNPs also differ in their length. The BR mRNPs are ~2x in length compared to MDN1 mRNAs, and there is some evidence suggesting that longer mRNAs are more compacted than shorter mRNPs (Mor et al., 2010). This has also been observed for BR mRNPs that undergo multiple stages of compaction during their assembly (Björk and Wieslander, 2015). However, almost all mRNAs in humans are shorter than MDN1 and BR

mRNAs, and whether the shorter length of mRNAs results in alternated mRNP organisation needs to be determined.

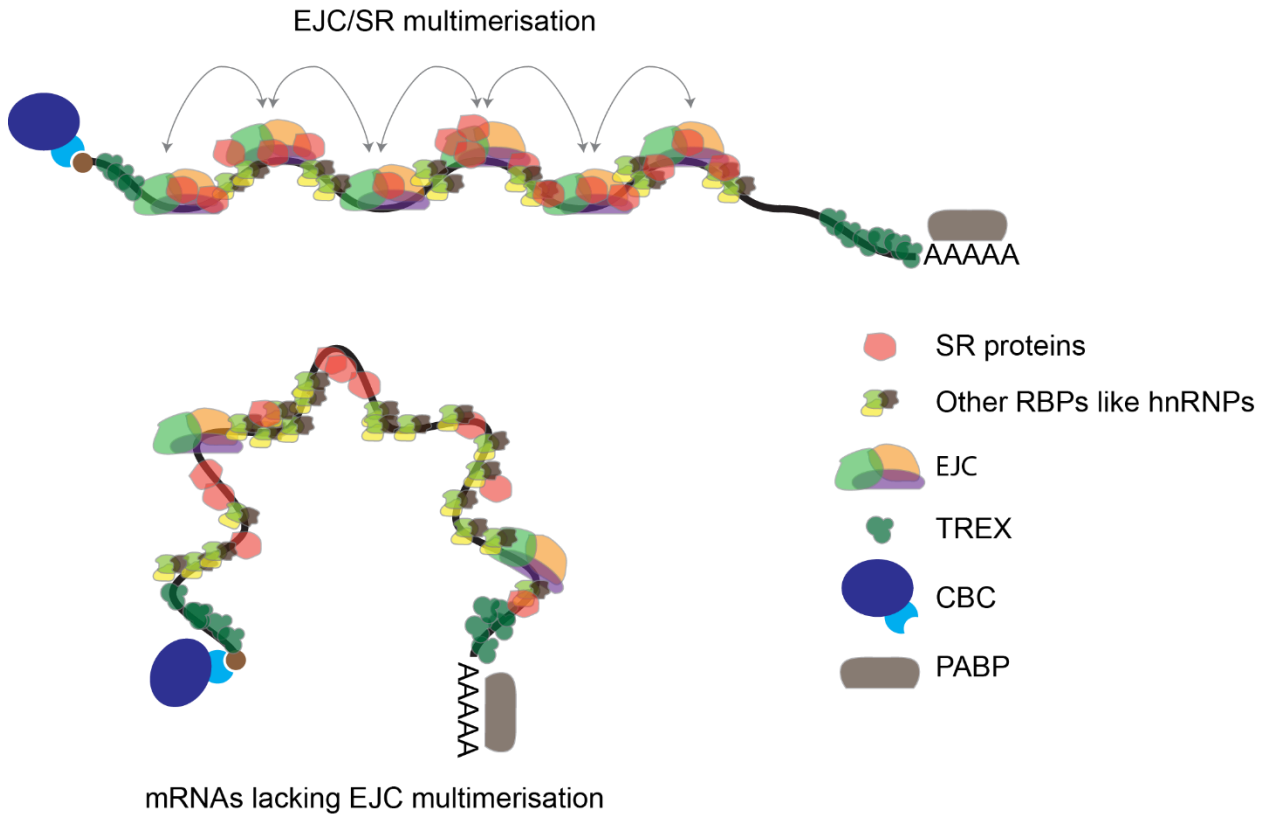


Figure 4-1: EJC multimerisation can result in mRNP organisation into a linear-rod

This figure illustrates how local EJC-SR protein multimerisation can influence its organisation. mRNAs with multiple exons can have several EJCs deposited on them, allowing for such local interactions.

Composition of nuclear mRNPs

The final structure of the mRNP depends on its co-transcriptional packaging and subsequent compaction. While EJCs play a prominent role in this process, their splicing dependent recruitment to mRNAs could suggest that they could act as 'late' factors, possibly after initial packaging of the nascent transcript as it emerges from the polymerase and once the spliceosome has been assembled and perhaps introns have been spliced. The hnRNP proteins, SR proteins and the TREX complex, among others, have been hypothesised to be involved in the early assembly of mRNPs. hnRNPs

were initially identified as part of the nascent hnRNP particles, making them prime candidates for packaging the nascent transcripts. On the other hand, SR proteins and the TREX complex are recruited through interactions with the RNA pol II CTD or the CBC, suggesting their co-transcriptional assembly on the nascent transcript. In addition to these proteins and protein complexes, several different RBPs are known to be recruited co-transcriptionally to nascent mRNAs, assisting in the processing and exporting of mRNAs (Singh et al., 2015). Despite our knowledge of the proteins capable of binding RNAs, the temporal pattern of the recruitment of these proteins to mRNPs, and a general understanding of the composition of specific nuclear mRNPs at different stages of its assembly is still lacking.

Though the multimerisation of EJC assemblies might influence the linear organisation of MDN1, additional factors could determine its local and global compaction and our observations of nuclear mRNP compaction of AHNAK point in this direction. We observed that AHNAK mRNAs also form compact structures in the nucleus. The probes targeting AHNAK are positioned within the body of the long exon and should reflect the compaction of this exon in an EJC-independent manner. However, we cannot rule out the non-canonical binding of EJCs within the exon body, as has been observed using CLIP-Seq for EJC factors, and whether the compaction observed for AHNAK is due to the function of RBPs recruited co-transcriptionally to the mRNA or through EJCs bound within the body of the exon is yet to be determined (Saulière et al., 2012). An overall comprehensive outlook of the RBP composition of individual mRNAs can help understand if there is differential recruitment of RNA-binding proteins to mRNAs and how this could influence mRNA compaction and overall organisation.

4.2.5.3 Is there heterogeneity in nuclear mRNPs of the same transcript?

Studies with Balbiani ring mRNPs have observed a surprisingly reproducible manner of organisation for all four BR genes - BR1, BR2.1, BR2.2, and BR6, finally assembling a particle with nearly identical structures. However, whether such homogeneity in mRNP assembly is a universal process applicable to all genes is unclear. BR3 mRNPs visualised with EM were found to exist with many different morphologies at the site of transcription, and our data for MDN1 mRNPs shows heterogeneity in nuclear conformations. Using three colour imaging of MDN1, targeting the 5', middle and 3' regions, we found that though most mRNAs are compact, a small fraction of nuclear MDN1 mRNPs has more open conformations (Supplementary Figure 2-7B).

Currently, we do not know what causes some MDN1 mRNPs to have such open conformations. One possible explanation is that this fraction represents mRNAs that have not yet been completely spliced, though RNA-seq datasets from nuclear fractions of HEK293 cells do not suggest the presence of inefficiently spliced introns within MDN1 (Neve et al., 2016). Alternatively, these mRNPs could be assembled differently either through alterations in the RBP composition or secondary structure formation, which eventually affects the overall compaction resulting in more open mRNPs. For, e.g., transcriptome-wide analysis has observed that loading of EJCs could vary between exon-exon junctions, even within the same transcript (Saulière et al., 2010, 2012; Singh et al., 2012). Similar observations have also been made for SR proteins (Björk et al., 2009). Whatever the cause for the heterogeneity, it is unclear if mRNPs with open conformations are processed and exported to the cytoplasm for protein synthesis.

4.2.5.4 mRNP transit through the nuclear pore complex

Though the kinetics of nuclear export have been explored extensively, BR mRNPs remain the only mRNPs whose conformations have been observed while transiting through the pore. Due to their croissant-like structure, BR mRNPs need to be unfolded at the nuclear basket before they enter into the central channel of the nuclear pore complex, where they are exported as linear rods. Whether such unfolding is necessary for the transit of all mRNPs in the nucleus is still unknown. Despite our best efforts, we were not able to capture MDN1 mRNAs transiting through the pore. The low expression levels of MDN1 in HEK293 cells, coupled with the fast kinetics of export, which has been observed to occur within a few hundred milliseconds, provide a possible explanation for our lack of success (Grünwald and Singer, 2010; Siebrasse et al., 2012). However, a few interesting observations can be made from the existing literature, which opens up avenues for speculation about mRNP conformations at the basket and during export.

Single-molecule imaging of MS2 labelled β-actin mRNPs has revealed two possible rate-limiting steps during export - at the nuclear basket and the cytoplasmic side of the pore (Grünwald and Singer, 2010). While the increased residency at the cytoplasmic side of the NPC was also recently observed during the export of the pre-60S ribosomal RNA, the rate-limiting step at the nuclear basket was not, suggesting that the delay at the NPC basket could be specific to β-actin mRNPs or perhaps mRNPs in general. This increased residency of β-actin mRNPs could assist in quality control steps or represent the time it takes to unfold the mRNP before its transit through the central

channel. However, not all mRNPs might need to undergo unfolding at the basket. A study using fluorescently labelled RNA-binding protein hrp36 visualised the export of all hrp36 bound mRNAs (Siebrasse et al., 2012). They found that export times for mRNAs could span over multiple orders of magnitude, with only a subset of mRNPs, believed to be the large BR mRNPs, found stalled at the nuclear basket. A linear organisation of mRNPs, resulting in a rod-like conformation, as observed for MDN1 mRNP in our study and >450 other mRNPs identified in the RIPPLiT dataset, could negate the need for remodelling at the nuclear basket and result in the fast export times observed in this study. Future studies capturing mRNPs at the pore either using smFISH or through cryoEM, similar to the observations for pre-ribosomes, could allow for the determination of mRNP organisation during export (Delavoie et al., 2019).

4.2.6 Cytoplasmic mRNP organisation

Once mRNPs reach the cytoplasm, they are translated into polypeptides and eventually degraded. The metabolism of mRNPs in the cytoplasm is tightly regulated by several factors binding to the mRNA's 5' and 3' ends. Moreover, there is a known functional interlink between these factors binding at either end of the mRNP. As a result, mRNAs are thought to exist in structures with their 5' and 3' ends in proximity, allowing for protein factors to interact with one another, regulating the behaviour and function of mRNAs. However, previous visualisations of polysome conformations in cells have yielded vastly varying results, suggesting a lack of consensus on the structures taken by polysomes in cells. Our single-molecule analysis of translating cytoplasmic mRNPs contradicts the popular closed-loop model. By targeting probes to the 5' and 3' ends of multiple mRNAs, we and others have found that the ends are too far apart than would be expected for closed-loop structures (Figure 2-1 and Figure A-3) (Khong and Parker, 2018; Koch et al., 2020).

Furthermore, our results also suggest a possible additional consequence of ribosome occupancy. We find a strong correlation between increased ribosome numbers on the mRNA and the separation between the ends of the mRNA, with highly translated mRNAs having their ends further apart. Consistent with this, non-translating mRNAs, translationally inhibited mRNAs, and mRNAs within stress granules were found to be highly compact.

4.2.6.1 Ribosomes regulate local mRNP compaction

Ribosomes have a strong helicase activity suggesting their role in decompacting mRNPs by melting the local secondary structures within the body of the mRNA, and our observations confirm this role of ribosomes in mRNA decompaction. MDN1, POLA1, and PRPF8 mRNPs have extended conformations during translation compared to when their translation is inhibited. Furthermore, we found that the ribosome mediated decompaction correlates with the ribosome occupancy on the mRNA, as witnessed from our SUNtag experiment. Similarly, the mRNA region with active translation is less compact than regions with fewer ribosomes, as witnessed from our ribosome runoff experiments (Figure 2-4). Curiously, elongating ribosomes were found to have a larger effect on decompaction than stalled ribosomes, as observed by the slight but consistent difference in the end-to-end distances for MDN1, POLA1 and PRPF8 mRNAs, possibly due to stabilisation of the secondary structures between different ribosomes (Figure 2-2).

Despite the ribosome-mediated decompaction, translating mRNPs are still much more compact compared to their contour length. An MDN1 mRNA, considering an inter-nucleotide separation of 0.59 nm, has a contour length of 10.8 μm. However, our measurements suggest that, *in vivo,* MDN1 mRNAs rarely have end-to-distances >300 nm. This observation could be due to the absence of ribosomes from the majority of the mRNA. Translatome-wide studies have found that ribosomes are present every 156 and 183 nucleotides on average (Arava et al., 2003; Hendrickson et al., 2009). Single-molecule imaging studies have suggested an ever sparser ribosome occupancy on mRNAs, with an average inter-ribosome distance found to exist between 200 and 2000 nt (Morisaki et al., 2016; Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). Polysome profiling of HEK293 cells found MDN1 present in multiple polysome fractions (2-8) peaking at six ribosomes, suggesting an average ribosome every ~2000-8000nt (Floor and Doudna, 2016). Due to the difference in kinetics between RNA folding and translation elongation, and considering that ribosomes have a footprint of 30nt, this could mean that large regions of the mRNA are not occupied by the ribosomes and could subsequently fold into secondary structures and that ribosomes would have to unwind the mRNA continuously as they elongate (Figure 4-2).

Additionally, translational bursting has been observed using single-molecule imaging of translation, suggesting that ribosomes translating on the mRNA could be unevenly distributed on

the mRNA. Such a distribution could leave large regions of mRNAs unoccupied by ribosomes, resulting in their folding into stable secondary structures (Figure 4-2). (Wu et al., 2016; Yan et al., 2016). Whether such structures are unfolded by ribosomes alone or require the assistance of other RNA helicases is unclear. Nevertheless, it is possible that such a folding could slow down translation elongation. However, while translational bursts might result in the lead ribosome slowing down, the same might not always be true for the trailing ribosomes. If the trailing ribosomes are close enough, they could avoid the need for melting stable secondary structures, resulting in faster elongation times. While such differential translational rates could result in increased speed of translation by the trailing ribosomes, they could also likely lead to higher ribosome collisions. Supporting this hypothesis is the observation that an increase in RNA structure results in higher mRNA stability (Mauger et al., 2019). If increased RNA folding slows down all ribosomes, mRNA decay as a result of ribosome collisions could be avoided in highly structured mRNAs. However, whether the spacing between ribosomes allows for forming such stable structures needs to be determined.

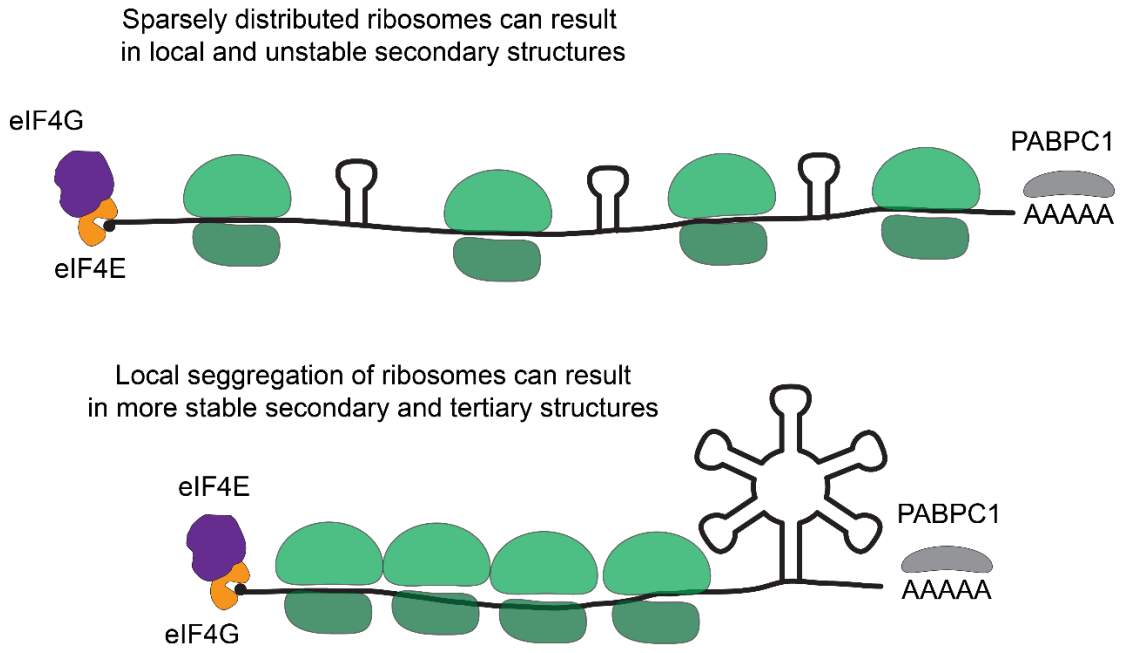


Figure 4-2: The effect of ribosome occupancy and ribosome distribution on local secondary structures

This figure illustrates how ribosome distribution can affect the folding of the mRNA. While an even distribution of ribosomes could result in local structures, an uneven distribution could result in large regions of RNAs that can fold into more stable structures, possibly influencing mRNA translation.

4.2.6.2 Non-translating mRNPs and stress granule dynamics

Using three colour imaging of MDN1, we found that mRNA forms compact structures upon translation inhibition by puromycin (Supplementary Figure 2-7). Similar compaction was also observed for MDN1 mRNAs assembled in stress granules using a combination of probes targeting the 5' and 3' ends and probes tiling along the length of MDN1 (Figure 2-5). Though the mechanism for compaction of RNAs *in vivo* has not been determined, RNA folding and RBP binding are possibly essential in this process.

Earlier *in silico* analysis had predicted that RNA folding could result in multiple conformations, each with similar free energies (Yoffe et al., 2008). Such folding-driven compaction has been observed *in vitro* for different RNAs, including mRNAs, using cryo-EM and SAXS (Borodavka et al., 2016; Gopal et al., 2012). Consistent with the prediction, these mRNAs were found to form a diverse array of structures mediated through stochastic events in RNA folding. Without the helicase activity of the ribosomes, RNAs can fold, forming secondary and tertiary structures *in vivo*. However, a few things need to be considered for non-translational mRNAs. First, *in vitro* studies of RNA folding allow for the formation of long-range interactions. While puromycin treatment, which results in almost simultaneous removal of all ribosomes, could allow for such long-range interactions, translational repression through ribosome runoff might not. The difference in kinetics of translation (~3-60nt/sec) and RNA folding (microseconds range) would likely result in the preference for local secondary structures (Gralla and Crothers, 1973; Morisaki et al., 2016; Pichon et al., 2016; Pörschke, 1974; Takamoto et al., 2004; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). Whether these structures are altered to assemble RNA structures with the least free energy once all the ribosomes have run off is unclear. Several cytoplasmic proteins like DEAD-box and DEAH box helicases could assist such a remodelling by dissolving and annealing

local secondary structures. Second, cytoplasmic proteins, like YB-1, could assist in this folding. YB-1 is one of the most abundant cytoplasmic proteins and is predominantly associated with non-translational mRNPs *in vivo.* Furthermore, YB-1 has been shown to multimerise on RNAs *in vitro,* helping in their compaction (Skabkin et al., 2004). Interestingly, RNA granules containing translationally repressed mRNAs have been found to associate with YB-1 *in vivo* (Mateu-Regué et al., 2019). Whether the cellular concentration of YB-1 allows for the formation of such structures when translation is inhibited for all mRNAs is an interesting question. In fact, an overview of the proteome of non-translational mRNPs could help understand how these RNPs fold *in vivo* and whether such folding has a functional relevance. Third, while the 5' UTR and ORF might be devoid of proteins due to the action of the elongating ribosome, the same is not true for the 3' UTR and the poly(A) tail. How the 3' UTR influences RNA folding and its resultant organisation could be crucial, primarily if the communication between the 5' and 3' ends is mediated during translational silencing (see below).

The similarity in mRNP compaction for non-translational mRNPs and mRNPs in stress granules raises interesting questions. It is known that mRNAs need to be translationally silenced before their recruitment to SGs (Khong and Parker, 2018). Furthermore, SG formation has been hypothesised to require RNA-protein and intermolecular RNA-RNA interactions (Van Treeck and Parker, 2018). We observe that translationally silent mRNAs have a compact organisation, likely due to intramolecular RNA-RNA interactions and perhaps RNA-protein interactions. Whether recruitment of translationally silent mRNAs to SGs requires alterations to local structures, enabling them to form intermolecular RNA-RNA structures is unknown. A more comprehensive view of single RNA structures using approaches like PARIS and COMRADES across different conditions, including the introduction of different types of stresses, can help determine how RNA-RNA interactions are altered in these conditions and how that could influence the dynamics of assembly and disassembly of SGs (Lu et al., 2016; Ziv et al., 2018).

4.2.6.3 Implication for the closed-loop model and 5'-3' communication

The classic textbook model for mRNPs in the cytoplasm shows structures with the 5' and 3' ends bridged through a series of interactions between the cap-binding protein eIF4E, the adapter protein eIF4G and the poly(A) binding protein PABPC1. The evidence for such a model comes from several biochemical studies showing specific interactions between the RNA elements and these

proteins, disruption of which leads to functional effects on mRNA translation (Imataka et al., 1998; Kahvejian et al., 2005; Tarun and Sachs, 1996; Wakiyama et al., 2000). However, this closed-loop model is just one explanation of the communication observed between the 5' and 3' ends of the mRNA, and there is ample evidence to suggest that such communication is essential in regulating the fate and function of cytoplasmic mRNPs (Rissland, 2017). This communication has traditionally been interpreted as resulting from spatial proximity between the ends of the mRNA. The spatial separation between the ends of MDN1, POLA1 and PRPF8 mRNAs in our study and similar observations made for AHNAK and DYNC1H1 in Khong et al. do not support a model where the ends of these mRNAs are always in proximity (Khong and Parker, 2018). Additionally, CRISPR/Cas9 cell lines with mutations that inhibit the interaction between eIF4G and PABPC1 are not only viable but rarely show any defect in translation or growth, suggesting that eIF4G-PABPC1 interaction is not the only means to achieve 5'-3' communication (Figure 2-3). Therefore, alternate models are required to fit our data in the context of existing biochemical studies to determine the mechanism through which this communication is achieved in cells.

5'-3' communication could be dynamic

Unlike the static pictures depicted in most textbook models, most cellular processes are dynamic and thus require a spatiotemporal view to understand their kinetics. Similarly, the spatial proximity between the 5' and 3' ends of mRNAs might not be a constant and could be achieved either dynamically for brief periods or at specific periods during the lifetime of the mRNA. Our data of mRNP conformations were captured in paraformaldehyde-fixed cells. As a result, we only observe static pictures of mRNP conformations achieved at specific periods during their lifetime, and whether mRNPs with extended conformations end up bridging their ends together through diffusion is unclear.

Very little is understood about the biophysical properties of polysomes or mRNPs in cells, making it harder to determine whether their inherent flexibility could be sufficient for the ends to meet dynamically. Measuring the flexibility of an mRNP requires developing assays to visualise the 5' and 3' ends simultaneously over time. We attempted to determine these dynamic changes in 5'-3' distances by using a reporter mRNA with MS2 stem-loops on the 3' end and translatable PP7 stem loops at the 5' end of a ~6,000 ntmRNA (Halstead et al., 2015). The translational status of this reporter mRNA was measured using the SUNtag system placed downstream of the PP7 stem-

loops, which allowed us to understand the spatiotemporal separation of the ends of the mRNA while knowing its ribosome occupancy. However, the reporter system failed to work in our hands as the PP7 stem loops within the ORF inhibited translation of the mRNA. An alternate unobtrusive method for tagging the 5' end of mRNP is thus required for such visualisations. With the advent of single protein tracking using Halotag, translation initiation factors could provide this alternative. However, whether the 5' and 3' ends could be visualised simultaneously using this system still needs to be determined. Such a system will further require the mRNA to be tethered to the cellular membrane as changes in distances could also result from the rotation of the mRNP.

5'-3' communication could also be achieved transiently at stages of an mRNPs lifetime when the ends are in proximity. Our three colour experiments with MDN1 suggest that the mRNP becomes compact when translation is inhibited. These compact molecules were organised with the shortest distances between the 5' and 3' ends (Supplementary Figure 2-7). Similar observations have also been made for AHNAK mRNAs in Khong et al. (Khong and Parker, 2018). While translation inhibition for our study was attained artificially, similar compact conformations could be achieved when mRNPs go through periods when they are not translated (Morisaki et al., 2016; Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016).

Computational prediction and *in vitro* experiments have shown that mRNAs tend to fold into structures with the ends in proximity, regardless of their length or sequence (Lai et al., 2018; Leija-Martínez et al., 2014; Yoffe et al., 2008). Similar folding is also observed for viral RNAs, many of which require communication between the 5' and 3' for their biogenesis (Leija-Martínez et al., 2014). As discussed earlier, RNA folding likely plays a significant role in the compaction of non-translational mRNPs, and these compacted non-translational mRNPs could have conformations that allow for biological interactions between proteins bound to their ends. Such analysis would require a resolution greater than what we get from our methodology, capable of distinguishing regions separated by a few nanometers in space. Recently smFRET, which can separate regions separated by <10 nm, has been used *in vivo* to study protein folding, and such an approach could be explored to study mRNP organisation in cells under different conditions (Okamoto et al., 2020).

5'-3' communication through closed-loop factors might occur for a subset of mRNAs

Additionally, closed-loop conformations mediated through eIF4E-eIF4G-PABPC1 could be specific for certain classes of mRNPs. For example, EM and cryo-ET studies found polysomes

with various conformations, some of which might resemble mRNPs translating in a closed-loop (Brandt et al., 2010; Christensen and Bourne, 1999; Christensen et al., 1987). For example, many ER-bound polysomes were observed to have circular or hairpin-shaped conformations, perhaps due to closed-loop interactions. These ER bound polysomes were imaged in cell types known to express mRNAs coding for specific secretory proteins, suggesting that most polysomes could represent a single transcript. However, the most significant evidence for mRNA specificity comes from various RNA-protein interaction studies in *S. cerevisiae*. In contrast to PABP, which was found to bind to most budding yeast mRNAs, closed-loop factors associating at the 5' end of the mRNAs – eIF4E, eIF4G1 and eIF4G2 were enriched explicitly for specific mRNAs (Costello et al., 2015; Thompson et al., 2016). Interestingly, these mRNAs have two main characteristics – they are short in length, typically less than 1,000 nt, and are highly translated (Figure 4-3) (Arava et al., 2003; Costello et al., 2015). However, whether this observed correlation between mRNA length and translational efficiency results from mRNAs forming a closed-loop structure or from increased stability and/or translation due to the association of the eIF4F complex at the cap is unclear.

The correlation between mRNA length and translational efficiency has been observed across different eukaryotic systems, from budding yeasts to plants, and this correlation could have significant importance in cellular function (Guo et al., 2015; Thompson and Gilbert, 2017). mRNAs with short open reading frames generally code for highly expressed proteins with housekeeping functions like ribosomal proteins, mitochondrial components and histones, the functions of which are tightly coupled to cellular growth. Hence, a mechanism to ensure high translation of these mRNAs under optimal conditions and repress it under conditions of stress or nutrient unavailability provides the cells with the ability to regulate the usage of their resources. Consistent with this, inhibition of eIF4G-PABP interactions in budding yeast has been associated with decreased translation and growth (Park et al., 2011). However, we do not observe a change in cellular growth in HEK293 CRISPR/Cas9 cell lines with impaired eIF4G-PABPC1 interactions. Thus, communication between 5'-3' ends of mRNAs in higher eukaryotes may involve additional factors that could compensate for the absence of eIF4G-PABPC1 interactions to enhance the translation of certain classes of mRNAs. However, the presence or identity of these factors and their resultant role in 5' -3' communication still needs to be determined.

Additionally, though short mRNAs show high relative density in metazoans, the correlation between mRNA length and the association of closed-loop factors has not been established yet. However, metazoans have been shown to regulate the translation of certain classes of mRNAs containing the 5' terminal oligopyrimidine (TOP) motif under changes in nutrient availability and conditions of stress (Thoreen et al., 2012). This alteration in translational status has been shown to occur due to changes in the binding of the eIF4F complex, suggesting a possible differential enrichment of specific mRNAs with the closed-loop factors similar to the observations in budding yeast (Costello et al., 2015; Jia et al., 2021). Whether this explains the conserved translational privilege of short mRNAs or if other mechanisms have evolved to ensure higher translation of short mRNAs in metazoans, perhaps through closed-loop interactions, is still to be determined.

In our study, we have used three mRNAs – MDN1 (18,413nt), POLA1 (5,486nt) and PRPF8 (7,295nt), which do not fit the characteristics for mRNAs with an increased association of closed-loop factors as found in budding yeast or high translation rates as found in most eukaryotes. However, we could still make some observations from our study that could help understand whether short and highly translated mRNPs could form a closed-loop structure. First, we observe that the longer MDN1 mRNAs have a higher median end-to-end distance than the shorter PRPF8 and POLA1 mRNAs, suggesting a possible length-dependence for the maximal separation between the ends. By extrapolating these findings, shorter mRNAs could have a higher chance of having their ends meet, perhaps by random diffusion (Guo et al., 2015). Second, in addition to the length, we also found that the translational status of an mRNA could influence its conformation, with highly translated mRNAs showing larger separations between their ends. While the decreased mRNA length provides a shorter maximal extension of the mRNA, short mRNAs are generally highly translated, and how this affects the spatial separation between 5' and 3' ends are not yet known. In addition to MDN1, PRPF8 and POLA1, we have visualised the 5'-3' separation of three shorter and highly translated mRNAs – GART (3,341 nt), PITRM1 (3,450 nt) and DYNC1I2 (2,589 nt), previously found to be present in higher fractions of polysome gradient (Floor and Doudna, 2016). All three mRNAs showed end-to-end separation that decreases upon translation inhibition with puromycin, suggesting a translation dependent separation of the ends (Figure A-1 and Figure A-3). It is possible that the mRNAs we have tested do not belong to the class of mRNAs, which are preferentially associated with closed-loop factors, and potentially that the factors determining the end-to-end separation for the mRNAs we have visualised might not apply

to mRNAs preferentially associating with closed-loop factors. Potential candidates belonging to mRNAs coding for ribosomal proteins, mitochondrial and translation initiation factors need to be tested to verify this hypothesis.

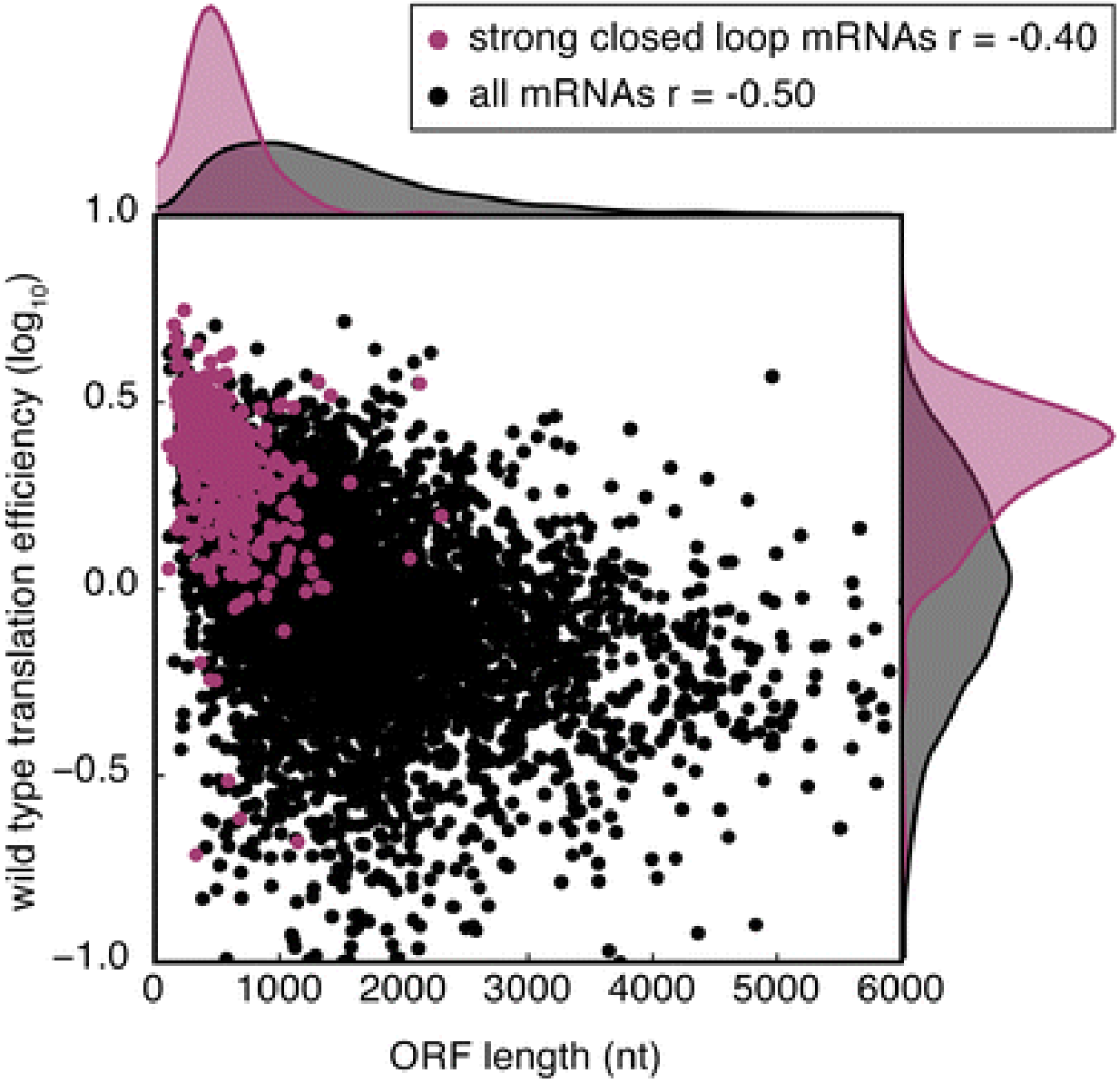


Figure 4-3: Translation status and length of mRNAs associating with closed-loop factors in *S. cerevisiae*

The figure illustrates how the length of mRNAs correlates with translation efficiency in budding yeast. The mRNAs with increased association with closed-loop factors are shown in pink, while

the rest are shown in black. The Spearman correlation coefficient for each group is indicated at the top. Figure adapted from (Thompson and Gilbert, 2017)

4.2.6.4 The compaction of 3' UTRs and its effect on mRNP metabolism

Unlike the 5' UTR and the open reading frame, the 3' UTR of an mRNA is not continuously remodelled by the elongating ribosome. Instead, local regions of mRNAs have been shown to form clearly defined and stable secondary and tertiary folds, some of which have functional roles in cells (Badis et al., 2004; Brümmer et al., 2017; Jambor et al., 2014). Furthermore, several RBPs have also been found to be enriched within the 3' UTR, including proteins from the hnRNP family - hnRNP U, L, E, FUS, and PTB (Geuens et al., 2016; Singh et al., 2015). The 3' UTR is influential in mRNP metabolism with several regulatory elements like miRNA-binding sites and RNA localisation elements known to exist within this region of the mRNA. Therefore, understanding the compaction of 3' UTRs and accessibility of regulatory elements within this region of the mRNA is requisite to understanding the function of these elements in mRNP metabolism.

Due to the short sizes of 3' UTR, we could not design probes targeting multiple regions to get a spatial understanding of their spatial organisation. However, an alternate approach can be used to get an estimate of the compaction of 3'UTRs. lncRNAs generally have a short ORF, a cap and a poly(A) tail. Hence, they could perhaps be perceived as mRNAs with long 3' UTRs. We visualised the end-to-end separation for two lncRNAs – TUG1 (7,469 nt) and OIP5AS1 (8,829 nt) and found that these distances are similar to our observations for mRNAs under translation inhibition, suggesting that 3' UTRs could potentially have high levels of compaction (Figure 2-5). It is possible that such compaction could influence the accessibility to regulatory elements, like miRNA-binding sites and RNA localisation elements, determining the fate of the mRNA (Brümmer et al., 2017; Yang et al., 2020).

4.3 Article 2: Single-molecule imaging suggests compact and spliceosome dependent organisation of long introns

4.3.1 Objectives and summary of results

In the second project, I studied the co-transcriptional assembly and processing of pre-RNPs. On average, human pre-mRNAs are much longer than the final processed transcript, with introns making up to 95% of the entire RNA sequence (Venter et al., 2001). Therefore, pre-mRNAs, including introns, need to be packaged and compacted to prevent them from binding to DNA and forming R-loops. Additionally, introns also need to be excised, which requires the spliceosome to assemble at their 5' and 3' ends, sometimes separated by tens to hundreds of thousands of nucleotides. How the transcription and splicing machinery work together to co-transcriptionally compact and organise an intron that helps bridge the 5' and 3' ends of long introns and assemble the spliceosome remains unclear.

In this project, I used the tools that I developed in my previous work to determine the compaction and spatial organisation of long introns and test how it is altered upon inhibition of spliceosome assembly. I chose two model introns for this study, intron 35 (42,330 nt) and 36 (65,255nt), within the long POLA1 gene, each with differing splicing kinetics. While intron 35 is spliced efficiently and is predominantly found at the site of transcription, the same is not true for intron 36 containing pre-mRNAs that are found at and close to the site of transcription. Regardless, our observations suggest that both introns are assembled into compact assemblies with their 5' and 3' ends in proximity, with the rest of the intron found to loop out, suggesting that while the textbook pictures of introns are accurate in showing the organisation, they might need revision in their depiction of introns as long unstructured particles.

Interestingly, our data also indicate a possible mechanism to ensure the proximity of the 5' and 3' ends of long introns. When visualising intron 35 using probes targeting three regions (5', 2/3rd and 3'), we found that the distances between the 5' end and 2/3rd regions were considerably shorter for introns lacking the 3' intron signal. These observations could result from a possible tethering of the U1 snRNP bound to the 5' ss to the elongating RNA polymerase II. Finally, we found that intronic organisation depends on the assembly of U2 snRNP at the 3' end of the intron. Inhibition of U2 assembly using Pladienolide B altered the end-to-end distances for introns 35 and 36,

suggesting that the nucleoplasmic pre-mRNPs we observe have introns with partially assembled spliceosomes. This alteration in intron organisation was found to affect the final assembly of the pre-mRNP, highlighting the importance of correct spliceosome assembly and pre-mRNP organisation in influencing mRNA metabolism. In the following sections, I will discuss these different observations in the context of existing models of splicing and known RNPs that have been visualised in cells or *in vitro*.

4.3.2 Introns are compact assemblies *in vivo*

Our previous observations suggested that nuclear mRNPs are assembled into compact particles, possibly due to their linear co-transcriptional packaging, and single-molecule imaging of introns in this study shows that similar conclusions can be made for introns. By targeting probes to multiple regions of introns 35 and 36 (4 and 5 regions, respectively) and determining the spatial separation between these regions, we estimated these introns' overall size within pre-mRNPs (Figure 3-1 and Figure 3-2). Furthermore, the looped intronic structure in normal conditions, which opens up when splicing is inhibited, suggests that these introns are possibly packaged and compacted co-transcriptionally in a linear manner, similar to our observations for MDN1 and cryoEM structures of BR mRNPs.

Our distance measurements between different regions of intron 36 suggest that it is organised into a compact particle with a diameter of ~100 nm. Considering a hypothetical fully extended, linear mRNA with a spacing between nucleotides of 0.59 nm and defining compaction as the ratio of the fully extended distance to the diameter of the assembled RNP, this would suggest compaction of ~400 fold. Such a high degree of compaction is similar to what is observed for RNAs packaged within the BR mRNPs (~413 fold). In comparison, 80S ribosomes, considering a diameter of 30 nm and containing 7,216 nt of RNA, are compacted by ~142 fold, while hepatitis C (~7,500 nt genomes and 22 nm inner capsid diameter) and Zika viruses (~11,000 nt and 30 nm inner capsid diameter) have a ~200 fold compaction (Sirohi et al., 2016; Wang et al., 2017). High compaction numbers could have two possible explanations 1) A high density of RNA within the RNP and 2) a more globular architecture. Cryo-EM studies have shown that BR mRNPs are compacted multifold to assemble the final mRNP that is released into the nucleoplasm. Furthermore, BR mRNPs have a croissant-like structure with the ends in proximity. The high RNA density and a more globular architecture result in high compaction levels. On the contrary, we found that MDN1

nuclear mRNPs have a compaction of ~111 fold. These comparatively lower compaction levels result possibly from their linear rod-shaped organisation, though whether there is a difference in density of RNA packaging between MDN1 and BR mRNPs is unclear. Introns are organised as a loop, suggesting an architecture that is more akin to BR mRNPs and their comparable compaction levels, despite their vastly different sizes, would attest to that.

POLA1 introns show a U2 snRNP dependent looped intronic organisation suggesting a linear assembly of the intron within this structure. While the exact factors involved in the compaction of these introns are unknown, there is plenty of room for speculation based on available data. A recent transcriptome-wide RNA structure analysis study found possible evidence for intron compaction through secondary structure formation. Using icSHAPE to determine folding of chromatin associated pre-mRNAs, it was found that RNA folding was much more prevalent within the intronic regions when compared to exons *in vivo,* though it could not be determined whether such folding results from short or long-range interactions (Sun et al., 2019). While many of the secondary structures could correspond to regions with known functional roles, like structures near the 5' and 3' splice sites, it could also suggest the general importance of RNA folding in intron compaction (Lewis et al., 2017; Xu et al., 2021). However, this dataset lacked reads for POLA1 introns and the prevalence of RNA structures within these introns and their role in compacting them is unclear.

In addition to RNA folding, several RBPs are known to bind to intronic regions preferentially. Prominent among these are the members of the hnRNP family, with hnRNP C hypothesised to play a crucial role in nascent RNA packaging. hnRNP C tetramer mediated compaction of RNAs has been observed *in vitro,* and recent CLIP-Seq datasets have suggested the possibility of such compaction *in vivo* (König et al., 2010). eCLIP datasets for hnRNP C show peaks separated by a regular interval of ~165 nt and ~300nt in several regions of intron 35 and 36 of the POLA1 gene, indicating that these introns could be assembled, at least in part, by hnRNP C multimerisation. However, large regions of the introns also lack hnRNP C peaks, and whether these regions are folded and compacted by other hnRNPs or through secondary structure formation is unclear (Van Nostrand et al., 2020). Additional variability in hnRNPC-mediated compaction could be caused by the difference in the stoichiometry of hRNPC binding or the position of hnRNP C tetramers within the body of different introns.

Though we do not know how many hnRNP C tetramers assemble onto these introns, we can estimate how hnRNPC-mediated packaging could compact introns, assuming an even binding of hnRNP C throughout the length of the introns. *In vitro* studies have suggested that three hnRNP C tetramers compact 700 nt of RNAs into a 19S isosceles triangular complex, and this assembly is repeated over longer stretches of RNA. The isosceles sides of the 19S complex measure 23 nm, the shorter side is 18 nm in length, and this complex's thickness is ~9 nm (Figure 1-1). Thus each 19S complex occupies a volume of ~1700 nm^3. Assuming the entire intron to be occupied by hnRNP C proteins, we should expect ~93 and ~60 19S triangular complexes to assemble on intron 36 and 35, respectively. Further, if the resultant intron is compacted and organised as a sphere with no spacing between the assembled 19S tetramers, these spheres will have a diameter of ~67 and ~58 nm, respectively. Our observations for POLA1 intron 36 yield sizes similar to those occupied by these theoretical particles, and it is to be determined if much of the compaction we observe *in vivo* for this intron results from such hnRNP C-mediated assembly.

4.3.3 Spliceosome assembly and co-transcriptional intron organisation

The pairing of the splice sites happens at the earliest stages of spliceosome assembly; however, the mechanism through which U1 and U2 snRNPs bound to the opposite ends of introns are brought together has remained elusive, especially when many introns in higher eukaryotes are tens of thousands of nucleotides long. However, despite the variability in length of introns, very little difference has been observed in splicing kinetics, suggesting that additional mechanisms exist to ensure the communication between the 5' and 3' splice sites and splicing of long introns (Coulon et al., 2014; Hollander et al., 2016; Martin et al., 2013; Singh and Padgett, 2009). One such model suggests a direct tethering of U1 snRNP to the elongating polymerase. This model gained popularity due to studies showing interactions between the U1 snRNP and RNA Pol II (Harlen et al., 2016; Nojima et al., 2018). More recently, a cryo-EM structure was reported for the U1 snRNP-Pol II complex, showing a direct interaction between the positively charged α-helices of the U1-70k and the negatively charged pocket formed by the RNA Pol II domains RBP2 and RBP12. However, no definitive evidence for such tethering had been observed *in vivo*. Our observations for intron 35 within the POLA1 gene provide the first indication for such a co-transcriptional tethering and the growing loop model (Zhang et al., 2021).

The growing loop model suggests that the direct interaction between the elongating polymerase and the U1 snRNP bound to the 5' splice site would result in the intron looping out from the nascent pre-mRNA, facilitating the scanning and recognition of the downstream branching point and the 3' splice site. Our data for introns 35 and 36 of the POLA1 gene show that these introns exist as looped structures within pre-mRNPs, with fully transcribed introns having their ends in proximity. Though we do not observe the steps that lead to the formation of this looped structure, the proximity of 5' and 2/3rd regions within the partially transcribed intron 35 suggests that tethering of RNA Pol II to U1 snRNP bound to the 5' ss could facilitate this (Figure 3-2). However, whether tethering is observed for all other introns, including intron 36 of POLA1, and is achieved using interactions observed in the cryo-EM structure needs to be determined, and mutagenesis within the residues of U1-70k, RBP2 and RBP12 can help address this question.

Our observations of a U1-tethering model raise several questions about the spliceosome assembly once the polymerase reaches the 3' end of the intron. For instance, it is unclear how and when U2 snRNP associates with the 3' splice site and whether U2 assembly is coupled to the release of U1 snRNP from the elongating polymerase. Several studies have identified the association of RNA Pol II with factors of the U2 snRNP. For example, U2AF65 has been known to interact directly with RNA Pol II CTD and indirectly through interaction with the transcription elongation factor TCERG1 (David et al., 2011). Similarly, TCERG1 is also known to interact with SF1, which is the first protein deposited at the branch point site (Hollander et al., 2016). Such interactions could indicate the role of RNA Pol II in the co-transcriptional recruitment and assembly of U2 as soon as the 3' splice site emerges. Whether such recruitment immediately results in the formation of the A complex or if U1 release is a requirement for the interaction between U1 and U2 is not known.

The release of U1 snRNP from the RNA Pol II is harder to address. Zhang et al. suggest a model where U1 release could be coupled to the formation of the pre-B complex that results in the eventual displacement of U1 from the spliceosome resulting in the B complex. This model could be supported by biochemical studies showing RNA Pol II accumulation at exons and exon-intron junctions, suggesting possible sites of Pol II pausing (Brodsky et al., 2005; Jonkers et al., 2014). Furthermore, RNA Pol II is observed to have increased elongation rates in regions of the chromosome with decreased exon abundancy, and measurements of Pol II elongation rates also suggest that it could slow down at the 3' end of introns (Jonkers et al., 2014; Veloso et al., 2014). These pausing or polymerase slowdown effects at the 3' end of the intron or the downstream exon

could provide enough time for the recruitment of the tri-snRNP complex and the formation of the pre-B complex, eventually resulting in the release of U1 snRNP from the pre-mRNA. Our data for POLA1 intron 35 and 36 conformations could support this model. We find that the proximity between the ends of the introns is only maintained in normal conditions, and disruption of U2 assembly using Pladienolide B increases the distances between the ends of the introns. These observations indicate that despite the difference in splicing kinetics, intron 35 and intron 36 have some components of the spliceosomes assembled, ensuring end-to-end proximity. However, as we cannot determine the stage of spliceosome assembly within individual introns, nor whether U1 snRNP is still associated with these introns, we cannot determine whether the end-to-end proximity we observe results from the formation of the pre-B complex.

In addition to the difference in kinetics of splicing between the two introns, we also observe a consistent but slight difference in end-to-end distances for the two introns. The median end-to-end distances for intron 35 and 36 are ~40 nm and ~57 nm, respectively (Figure 3-1 and Figure 3-2). One possible explanation for this observation is that the difference results from the smFISH probe distribution on these introns. The probes for POLA1 intron 36 are spread over ~3,900 nt and ~1980 nt (5' and 3' respectively), and for intron 35, they are distributed over 1,900 and 1,100 nt (5' and 3' respectively). Regardless of the differences, both introns show a similar phenotype with increased end-to-end separation upon inhibition of splicing using Pladienolide B, suggesting that the separation for both introns depends on U2 assembly. However, it is still possible that the assembly of U2 at the 3' splice site alone might not be sufficient for the formation of A complex in all introns, and the difference in end-to-end distances might reflect such introns. If the A complex or U2 snRNP assembly is coupled to the binding of U1 snRNP on the downstream intron, it is possible that intron 36, which is the last intron within the POLA1 gene, could have a different conformation in comparison to an intron earlier within the body of the gene. Alternatively, the differences could be a result of altered processing of last introns in general. The splicing of the last intron is coupled with the 3' end processing of the pre-mRNA, with physical and functional interactions known to occur between splicing factors and 3' end processing factors (Kaida, 2016). Such interactions could result in altered conformations for the last introns, and more comprehensive visualisation of end-to-end distances for different introns can help address this question.

The increased separation between the ends of introns upon Pladienolide B also raises questions about whether such introns would end up bridging the gap between the 5' and 3' splice sites, assemble the spliceosome and get spliced eventually. Or in general, whether co-transcriptional intron organisation and spliceosome assembly within long introns is essential for their eventual excision. For instance, though most introns have 5'-3' ends in proximity, a small fraction has more open conformations with their 5' and 3' ends far apart under normal conditions. Whether these introns are spliced post-transcriptionally, or pre-mRNAs containing these introns can be degraded is unclear as smFISH only allows for a static view of these introns during their lifetime. A more dynamic measure of intron organisation is thus required to test its effect on intron metabolism. Live-cell imaging of pre-mRNA using MS2 and PP7 stem-loops placed either within different regions of long introns or distributed between exons and introns can help yield a more comprehensive intron organisation and its effect on splicing. One such study found a vast variation in splicing kinetics ranging from minutes to > 1 hr, even for pre-mRNA synthesised in the same locus (Wan et al., 2021). While these could represent introns that have been unable to assemble the spliceosome, it is also possible that the slow kinetics result from the introns with increased spatial separation between the ends.

4.3.4 An alternate model for splicing of long introns – recursive splicing

Our single-molecule analysis with POLA1 introns suggests that their organisation depends on the spliceosome assembly, defining the boundaries of the intron. When the boundaries are correctly determined, we observe the introns to be organised in a looped conformation and disruption of the boundary recognition using splicing inhibitor Pladienolide B ends up resulting in the separation of the 5' and 3' ends of the intron. These results suggest that despite the background of cryptic and alternative splice sites, the spliceosome assembly is robust, and only canonical splice sites and branch points are recognised within these introns. However, whether this holds for other long introns is unclear.

Cryptic splice sites within long introns have been proposed to play a role in recursive splicing. This splicing model suggests the progressive removal of chunks of the intron as the polymerase travels along the gene and requires the presence of a pair of 5' and 3' cryptic splice sites juxtaposed next to each other within the body of the intron (known as recursive splice site or RS site). First discovered to occur within three long introns in *Drosophila,* recursive splicing has since been

extensively characterised, with ~130 introns known to undergo recursive splicing in fruitflies (Sibley et al., 2016). However, the prevalence of recursive splicing within mammalians introns is unclear. Early studies had identified a handful of RS sites in humans (Sibley et al., 2016). However, the identity of RS sites has been expanded, with recent transcriptome-wide CLIP-Seq data showing the pervasive binding of the U2AF heterodimer along the body of many introns. These observations indicated potential spliceosome assembly at different regions within the intron, which could result in recursive splicing (Wan et al., 2021). This data was supported with long read-sequencing and lariat sequencing methods that showed that splicing of introns could occur recursively within several introns. However, the usage of these RS splice sites was still found to be highly stochastic, with single-molecule imaging of one intron (first intron within the RAB7A gene ~62,000 nt in length) only estimated to be removed in more than one piece ~57 % of the time, despite it containing 22 RS sites within the intron body.

Our smFISH images of POLA1 introns 35 and 36 suggest that recursive splicing is not a very common occurrence for these introns. Using probes against the 5' and 3' regions, we detect multiple introns containing either just the 5' signal, indicating partially transcribed introns or colocalising 5' and 3' signals, resulting from fully transcribed introns. However, individual introns with just the 3' signal, resulting from recursive splicing, are extremely rare. It is possible that the stochastic binding of the spliceosome within the body of the intron, as observed in Wan et al., could contribute to some molecules of intron 35 and 36 being recursively spliced (Wan et al., 2021). Nevertheless, the high frequency of introns containing the 5' and 3' ends observed in our study and those observed in Wan et al. suggest that recursive splicing alone might not be sufficient to ensure the efficient splicing of long introns.

4.3.5 Determining the model for spliceosome deposition – intron vs exon definition

Two models have been suggested to determine the deposition of the spliceosome, termed intron definition and exon definition. The intron definition model suggests that the splicing machinery recognises the boundaries of the intronic unit and places the spliceosome across the intron. Hence, according to its canonical meaning, intron definition is restricted to shorter introns as increasing the length of the intron separates the ends, making it harder for the splicing machinery to recognise the boundaries. This mechanism is proposed to be predominant in lower eukaryotes like *S.*

cerevisiae and *Drosophila* (De Conti et al., 2013). On the other hand, exon definition occurs when the basal splicing machinery is deposited on either end of an exon, hence, constraining the length of the exon. Exon definition is believed to be the widespread mechanism for spliceosome deposition in vertebrate species containing many long introns. According to the exon definition model, the splicing enhancers within an exon, typically SR proteins, concurrently interact with U1 snRNP and U2AF proteins to deposit them on the two ends of the exon (De Conti et al., 2013; Hollander et al., 2016). Spliceosome deposition on either end of the exon is hypothesised to form a cross-exon spliceosome, later converted to a cross-intron complex linking the U1 and U2 snRNPs bound to the upstream intron (Hollander et al., 2016; Li et al., 2019b). The evidence for exon definition comes from findings that show altered splicing of upstream intron when the 5' ss of downstream introns is mutated (Kuo et al., 1991).

While the intron and exon definition models help define spliceosome assembly and explain how an exon can be retained or excluded from the final mRNP, they do not provide a mechanism for the efficient assembly of the spliceosome across long introns. The U1-Pol II tethering model could act as a means for bridging the ends, and the factors determining spliceosome assembly across the intron and the downstream exon could determine whether its intron or exon definition. For example, if the deposition of U2 snRNP on the 3' splice site is independent of the recognition of the 5' splice site on the downstream intron, this will represent an intron definition model. However, as observed for several exons, the 5' ss on the downstream intron can influence the splicing of the upstream intron (De Conti et al., 2013). Therefore, if U2 assembly and spliceosome formation on the upstream intron depend on U1 deposition or 5' ss recognition on the downstream intron, this would agree with the exon definition. An i*n vitro* study using human nuclear extracts found that U1 snRNPs bound to the 5' ss of upstream and downstream introns could have a distinct but additive effect on the recruitment of U2 snRNP and splicing, suggesting the role of the downstream 5' ss in U2 recruitments (Braun et al., 2018). However, the exact steps in the assembly of U2 snRNP at the 3' splice site of the intron, once the polymerase reaches and crosses the downstream exon, are unknown and further biochemical and structural analyses characterising this process are required to get a better picture of how long introns are spliced (Figure 4-4).

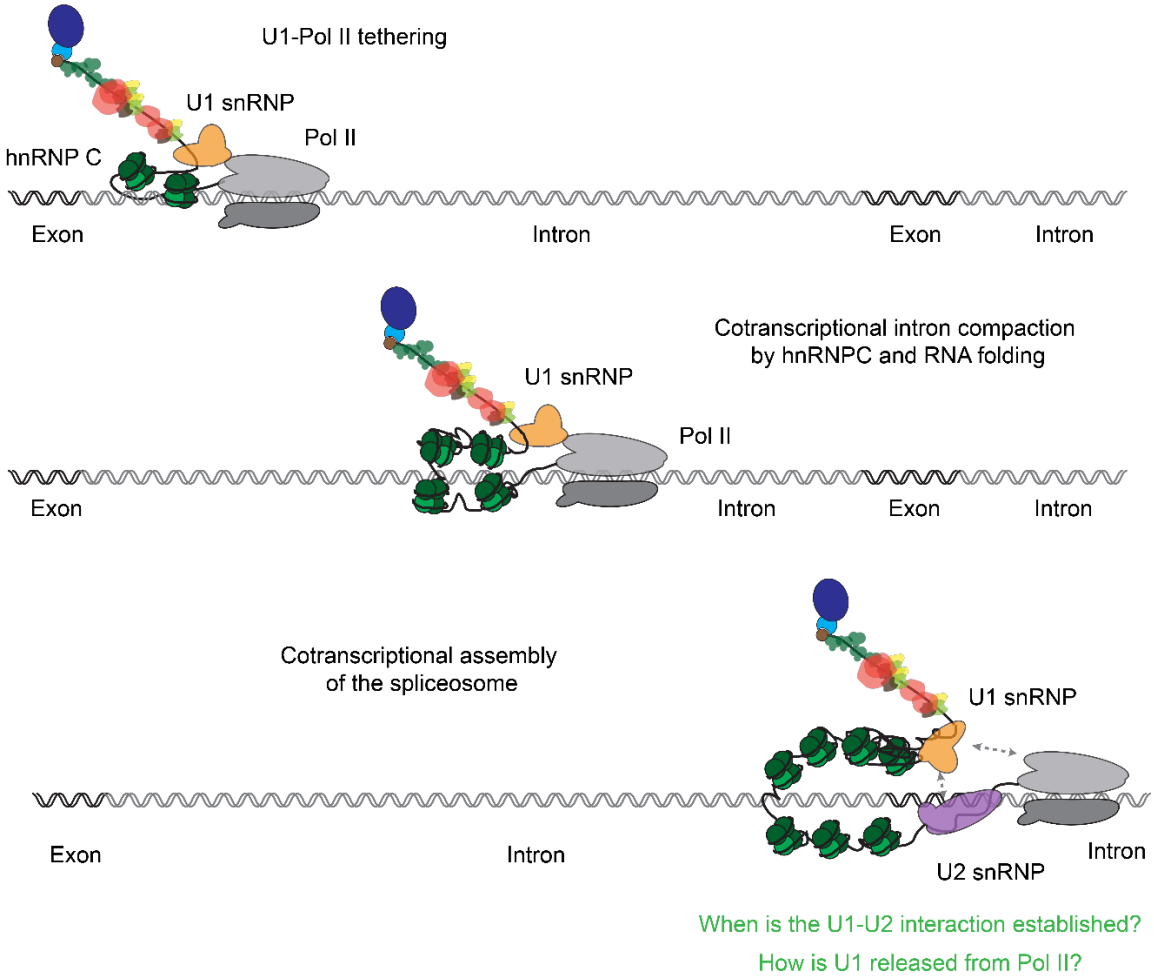


Figure 4-4: Cotranscriptional assembly and compaction of intron and the possible establishment of the spliceosome

This model illustrates how introns could be co-transcriptionally packaged and organised. As the polymerase transcribes the 5' end of the intron, the deposited U1 snRNP can interact with the polymerase, as observed in (Zhang et al., 2021)*. The transcribed intron can be packaged through RNA folding and assembly by hnRNPC and possibly other hnRNPs. The mechanism of assembly of the spliceosome and the release of the tethered U1 snRNP once the polymerase reaches the 3' end is not yet known.*

4.3.6 Effect of intron looping on (pre-)mRNP organisation

Our observations of a looped intronic organisation could have implications beyond the splicing of that particular intron. As discussed earlier, final nuclear mRNPs in humans have been observed to

be organised in a linear rod-like shape by forming local multimers through interactions between EJC and SR proteins and from our observations for MDN1 mRNAs (Adivarahan et al., 2018; Metkar et al., 2018; Singh et al., 2012). Co-transcriptional assembly of introns that results in a looped conformation could assist in this process, ensuring an efficient EJC multimerisation through interactions with SR proteins within the body of the exon, irrespective of whether an active spliceosome has been assembled. Deposition of EJC, which is thought to occur through recruitment by CWC22 in an active spliceosome, can further stabilise these interactions (Boehm and Gehring, 2016). The conformations of POLA1 pre-mRNPs points towards such a role. We find that upon treatment with Pladienolide B, the end-to-end distances for POLA1 increase. However, they do not increase proportionally with the end-to-end distances for either intron 35 or 36, suggesting alternate interactions within the pre-mRNP exist that keeps the ends in proximity (Figure 3-3 and Supplementary Figure 3-6). Therefore, intron looping could be essential for the proper assembly and organisation of mRNAs in the nucleus, especially for ones containing inefficiently spliced introns, though the validity of such a hypothesis needs to be tested.

4.4 Concluding remarks

The focus of my thesis was to elucidate the organisation of mRNPs at different stages of their life cycle. To achieve this, I have developed a microscopy-based approach combining single-molecule RNA fluorescence in situ hybridisation (smFISH) and structured illumination microscopy (SIM) and used it to reveal novel aspects of mRNP organisation in cells.

In my first project, I demonstrated that mRNP organisation and its compaction levels could vary between different compartments and is determined by cellular processes. In the nucleus, we found that mRNPs are organised into compact particles, and together with another study, we found that these particles could have linear rod-like organisation. When mRNPs reach the cytoplasm, translation alters their conformations, opening up mRNPs and separating the ends, an organisation that is contradictory to the textbook models of translation. Furthermore, the end-to-end separation of mRNAs was observed to scale with the ribosome occupancy on the mRNA and inhibiting translation through ribosome removal compacts mRNPs. This data opens future investigations for mechanistic and structural studies to determine how the communication between different regions of mRNPs are mediated and how cellular processes regulate it.

In my second project, I expanded my work on mRNP organisation and determined how co-transcriptional assembly of pre-mRNPs influence the splicing of introns. The findings of this study are the first view of intron organisation in cells and, together with a recent cryo-EM structure, provide a possible mechanism for splicing long introns. Our observations suggest that long introns are organised co-transcriptionally into compact particles. While nascent introns have an organisation with the 5' end closest to the furthest transcribed region, introns containing the 3' end are organised with their ends in proximity. These observations are possibly a result of the 5' ss tethered to the elongating RNA Pol II, facilitating the interaction between the ends of the intron and the formation of the intron loop. Furthermore, intron organisation depends on the assembly of the U2 snRNP, inhibition of which separates the ends of the intron.

In summary, both of my PhD projects share a common theme where we show (pre-)mRNP compaction and organisation is highly variable and is changed throughout its lifetime depending on the localisation of the RNA and cellular processes like translation and splicing. Further, I have established an RNA-centric methodology that can be combined with future approaches to get a comprehensive view of mRNP organisation in cells.

4.5 The 3D organisation of mRNPs and perspectives for future research

Though my PhD work has provided significant insight into mRNP organisation, several questions remain unanswered, many of which require improvements in methodology, combining with other known techniques or developing entirely new techniques.

Several modifications can be made to our experiment setup to overcome some of its drawbacks. First, our approach is low throughput, which includes both the number of mRNAs and the regions within single mRNAs that can be visualised. This can be overcome by using the recently developed multiplexed FISH approaches that have allowed for the imaging of hundreds to thousands of mRNAs in cells (Chen et al., 2015). However, whether imaging of multiple mRNAs still allows for the reliable assignment of spots that correspond to a single mRNA while still maintaining the high spatial resolution observed using our setup will have to be determined. Second, we largely restricted our analysis to two dimensions as we found that the experimental variability between images and the resultant error introduced by analysing in 3D was near the dimensions of the molecules visualised in our studies. However, the use of fiducials could help overcome this

limitation, and our initial attempts have shown promising signs. Third, we are restrained in terms of mRNA length and spatial resolution. Our ~20 nm spatial separation results from RNAs imaged using two different channels, and we still need an RNA coverage of ~1,000 nt to have enough probes for a reliable smFISH signal. As a result, we have restricted our analysis to longer mRNAs mRNA and regions separated by more than 1,000 nt. Amplification-based approaches like branched DNA FISH, FISH-STICs or clampFISH, can be used instead of the traditional smFISH to reduce the RNA coverage, helping us analyse shorter mRNAs (Adivarahan and Zenklusen, 2019). Additionally, higher-resolution approaches like STED and dSTORM could further help separate probes binding to regions that are otherwise non-separable due to the diffraction limit of light. Finally, getting a detailed understanding of the structures adopted by mRNPs requires the use of approaches that have a resolution beyond what is capable with light microscopy. Electron microscopy can be combined with *in* situ hybridisation (EM-ish) as an alternative to smFISH to get a 3D organisation of mRNPs (Trzaskoma et al., 2020). Additionally, the advent of cryoEM allows for near atomic-level resolution and has been used extensively to determine structures of ribosomes and other RNA-protein complexes. However, whether the inherent variability in mRNP structures allows for accurate prediction of structures using this approach will need to be determined.

The study of mRNP organisation is intertwined with the knowledge of mRNP composition. While improvements have been made in identifying RBPs associating with mRNAs, we still have a minimal idea of the composition of individual mRNPs and how it changes dynamically over time. RNA-centric approaches involving mass spectrometry have been applied for specific mRNAs, and this has provided a new perspective of a single mRNA proteome (Jazurek et al., 2016). These have been expanded to include proximity biotinylation based labelling and identification of RNA-binding proteins (Han et al., 2020). As this biotinylation has a fixed radius within which proteins can be identified, positioning the bait within certain regions of the mRNA using an aptamer tag or using Cas13-based RNA tethering could provide a spatial proteomic map of the mRNA.

Furthermore, as the RBP composition is altered throughout an mRNP's lifetime, it requires a dynamic view to understand the residency times, kinetics and the stoichiometry of RBPs on the mRNA. Single-particle tracking using Halo-tagged proteins has recently been used to identify the residency of several transcription factors and other DNA binding proteins, and such an approach could be adopted to study the dynamics of RNA-binding proteins (Hansen et al., 2017; Presman

et al., 2017). More recently, a cross-linking based approach was used to determine the kinetics of the RBP DAZL on individual RNA-binding sites in cells, and its use can be expanded for similar detection for other RBPs (Sharma et al., 2021). Similarly, the stoichiometry of RBPs can be measured using fluorescence correlation spectroscopy or single-step photobleaching assays, and some attempts have been made in this direction (Wu et al., 2015).

The main question left unanswered for cytoplasmic mRNPs is whether the conformations we observe for the subset of mRNAs are applicable for all mRNAs and what it means for the closed-loop model. In addition to determining the organisation of the likely closed-loop candidates like ribosomal protein-coding mRNAs and mitochondrial factors, reporter mRNAs can be used, where the 5' and 3' UTRs could be modified to include elements like the TOP motif or miRNA-binding sites to determine if specific elements within mRNAs are responsible in determining the mRNP architecture. Such a system should be complemented with assays to identify specific targets of closed-loop factors in higher eukaryotes, and together they can help uncover the mystery of the closed-loop model. Furthermore, as discussed before, a better understanding of the biophysical properties of mRNPs is required to determine whether communication between different regions can be achieved through mere diffusion between the ends of the mRNAs. Such a model could be important for shorter mRNAs if the communication between ends is established dynamically. Finally, we do not know the composition or organisation of non-translational mRNPs. How these mRNPs are compacted and if their organisation in these compact structures is vital in determining their stability and/or influencing translation initiation is an interesting question worth pursuing.

The factors determining nuclear (pre-)mRNP organisation are even less clear. We do not currently understand whether all mRNPs have a linear organisation, nor do we know if different RBPs have distinct roles in the packaging and compaction of nuclear mRNPs. For example, TREX could be involved in the compaction of the 5' end of the mRNA and hnRNP C the intron. Additionally, if EJC is required for mRNP compaction or whether EJC-mediated multimerisation is only responsible for mRNP organisation is unknown. The effect of these proteins on mRNP organisation needs to be studied through depletions either using siRNAs or labelling the proteins endogenously using the auxin-induced degron (AID) system (Li et al., 2019a).

Our understanding of intron organisation is minimal. While POLA1 long introns have a consistent looped organisation, whether this is the universal mechanism for intron organisation in cells or if

other introns are organised differently needs to be ascertained. Genome-wide studies have found that nearly 90-95% of all genes undergo some form of alternative splicing, some of which is affected by cellular factors like the elongation rate of RNA Pol II (Wang et al., 2008). Additionally, introns can vary in their lengths and be back spliced. Similarly, introns can also be retained within pre-mRNAs, with some known to be spliced in response to specific stimuli (Mauger et al., 2016). How the organisation of such introns differs from POLA1 long introns is unknown, and further studies need to be conducted with such introns in mind to get a better overview of intron organisation in cells. Altogether, such studies will be essential in determining mRNP composition and organisation and how it affects various cellular processes.

Annex

a. Additional Figures

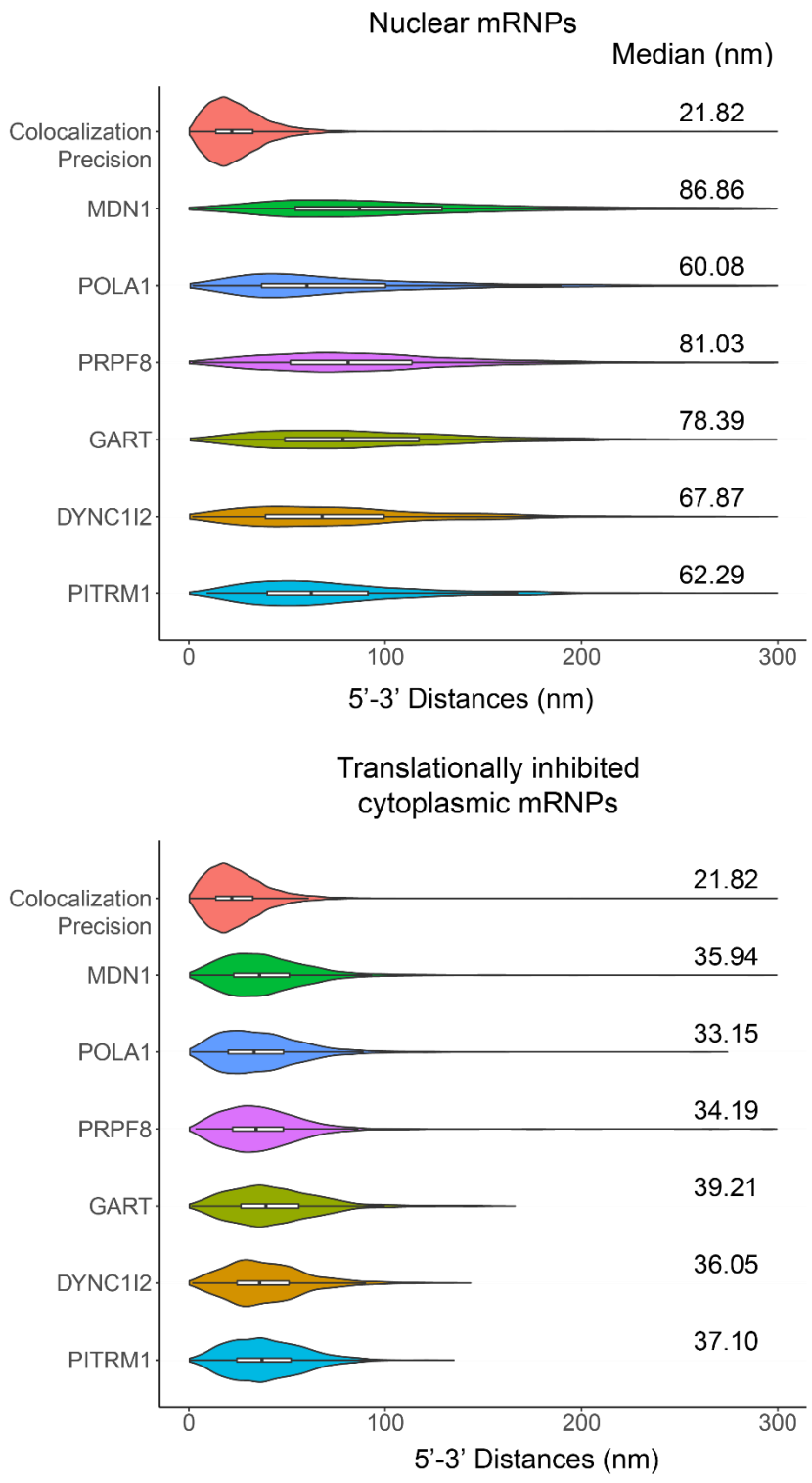


Figure A-1: Comparing 5'-3' distances for nuclear and cytoplasmic non-translating mRNPs

*Violin plots showing distance distribution of co-localization precision and 5'-3' distances for MDN1, POLA1, PRPF8, GART, DYNC1I2 and PITRM1 mRNAs determined by Gaussian fitting for nuclear mRNAs in untreated cells (top) and cytoplasmic mRNAs in cells treated with puromycin (100*μg*/ml, 10 min; bottom).*

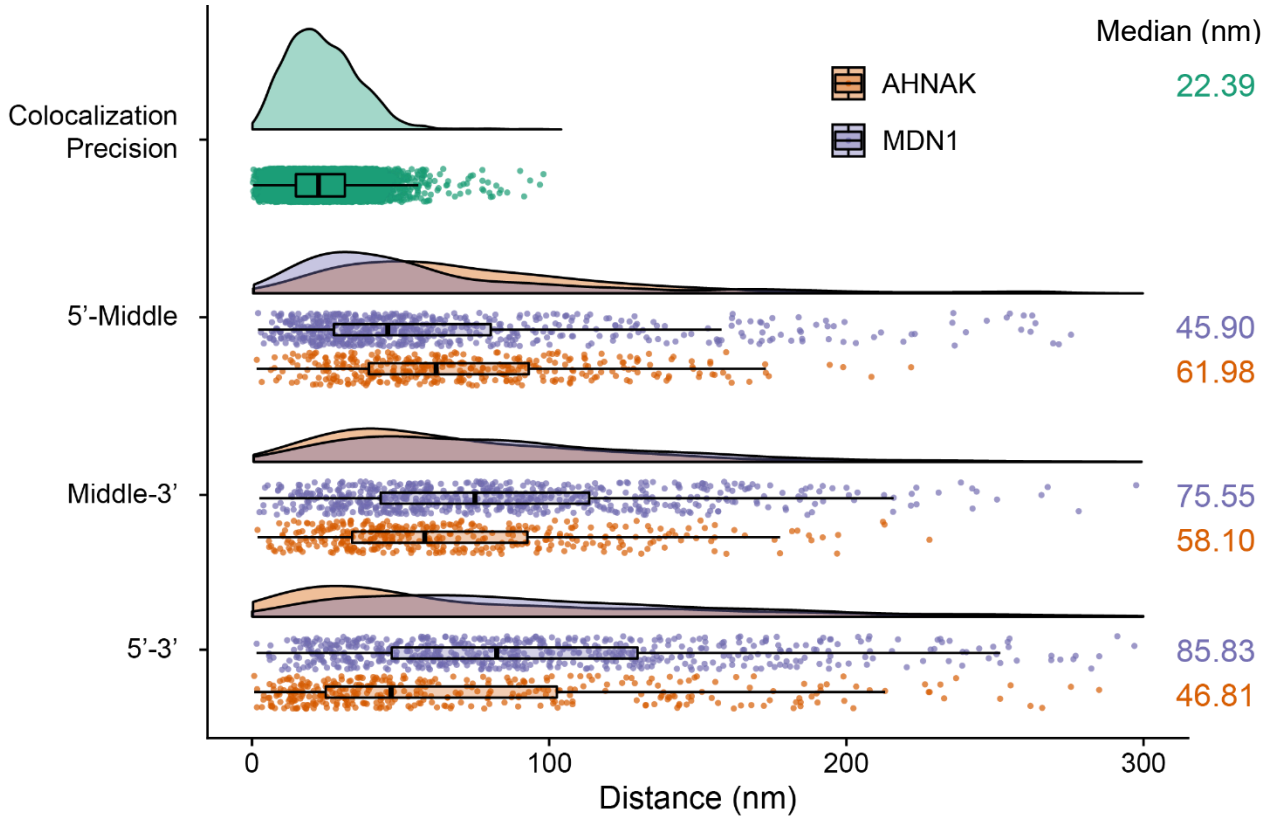


Figure A-2: Distances for MDN1 and AHNAK nuclear mRNAs

Raincloud plots for distances between different regions for AHNAK and MDN1 nuclear mRNAs. Plot shows distance distribution of co-localization precision, distances between different regions of AHNAK and MDN1 mRNAs. The box plot shows the first quartile, median and third quartile and the individual RNAs shown as spots. Median distances are shown on the right. Distances in nm.

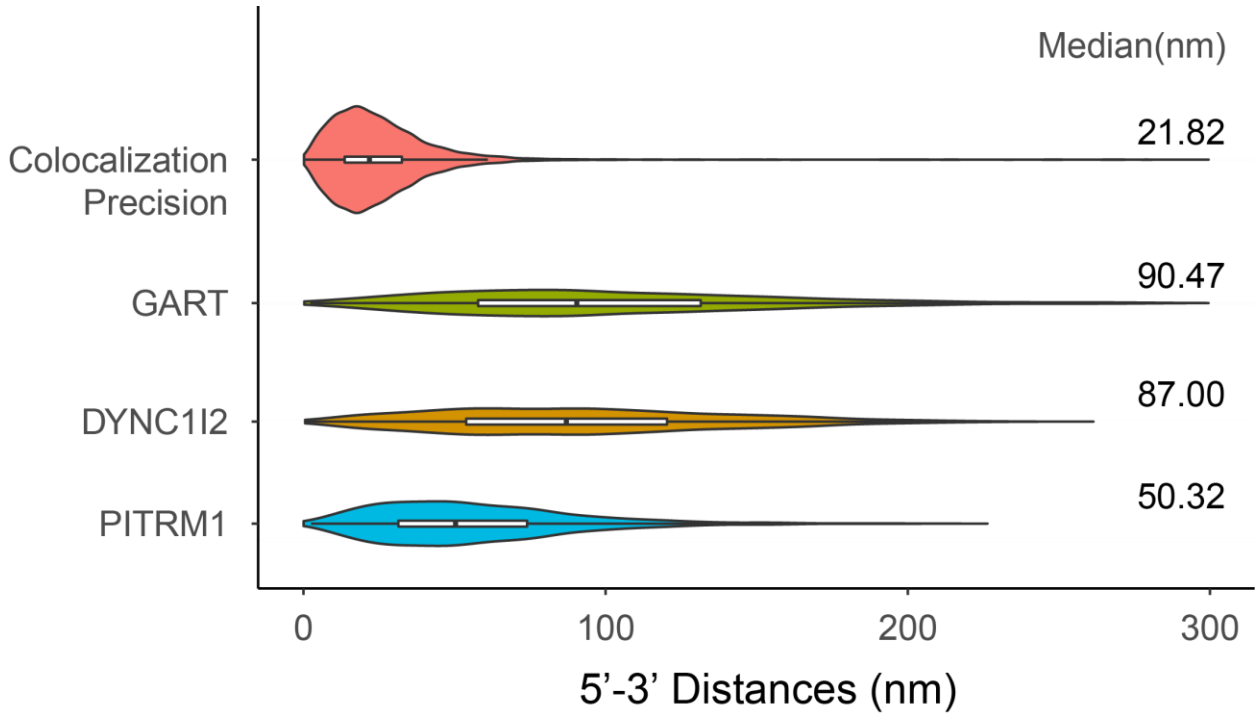


Figure A-3: Cytoplasmic mRNP conformations – shorter mRNAs

Violin plots showing distance distribution of co-localization precision and 5'-3' distances for GART, DYNC1I2 and PITRM1 mRNAs determined by Gaussian fitting for cytoplasmic mRNAs in untreated cells.

b. Lessons from (pre-) mRNA imaging

Srivathsan Adivarahan and Daniel Zenklusen*

Affiliations:
Département de Biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, Canada

*Corresponding author: daniel.r.zenklusen@umontreal.ca

Published in The Biology of mRNA: Structure and Function (2019)

Summary

Cells are complex assemblies of molecules organized in organelles and membrane less compartments, each playing important roles in ensuring cellular homeostasis. The different steps of the gene expression pathway take place within these various cellular compartments and studying gene regulation and RNA metabolism requires incorporating the spatial, as well as temporal separation and progression of these processes. Microscopy has been a valuable tool to study RNA metabolism, as it allows the study of biomolecules in the context of intact individual cells, embryos or tissues, preserving cellular context often lost in experimental approaches that require the collection and lysis of cells in large numbers to obtain sufficient material for different types of assays. Indeed, from the first detection of RNAs and ribosomes in cells to today's ability to study the behaviour of single RNA molecules in living cells, or the expression profile and localization of hundreds of mRNA simultaneously in cells, constant effort in developing tools for microscopy has extensively contributed to our understanding of gene regulation. In this chapter, we will describe the role various microscopy approaches have played in shaping our current understanding of mRNA metabolism and outline how continuous development of new approaches might help finding answers to outstanding questions or help to look at old dogmas with a new perspective.

Keywords: mRNA, mRNPs, electron microscopy, *in situ* hybridization, smFISH, polysomes, RNA imaging, single molecule microscopy, gene expression,

1. Tools for RNA visualization at different scales

The dynamic regulation of gene expression is critical for cells and organisms to develop and maintain homeostasis. mRNAs play a central role in this process, acting as messenger molecules that connect the information stored in the genome and the machineries translating them into proteins. However, despite their often-short-lived role as templates for proteins synthesis, controlling mRNA metabolism is among the most complex cellular processes composed of numerous steps, many subject to regulation and quality control and involving hundreds of proteins. Therefore, a longstanding and critical effort has been made towards studying gene regulation and mRNA metabolism, as mis-regulation in any step can lead to wide range of diseases (Cooper et al., 2009). Microscopy approaches have been critical tools in this effort, as they allow to study these processes in the context of the native environment of the cell.

Many imaging-based approaches have been developed to observe mRNA/mRNPs in cells, either in a fixed cell or a living cell context, and each of these approaches come with their own strengths and limitations. While fixation prior to any kind of labelling for RNA detection comes with the benefit of allowing complex labelling protocols and long exposures during image acquisition that is often required for robust detection and multiplexing, the dynamics of interactions within the cell is lost and can only be captured using live cell approaches. Moreover, though light microscopy techniques are most commonly used to study mRNAs, they are limited in their resolution, although the development of single molecule approaches and super-resolution microscopy have made it possible to overcome the limit set by the physical properties of light. Electron microscopy, on the other hand, is superior in terms of resolution and has been used in combination of different staining protocols or combined with immunolabeling to detect RNPs in cells or *in vitro*, however, it is limited in terms of labelling efficiencies for specific RNAs and multiplexing. In the paragraphs below, we will discuss the most commonly used techniques for visualizing mRNPs both in fixed and live cells, at low and high resolution, before discussing in more detail how RNA imaging has contributed to the current understanding of mRNA metabolism.

1.1. RNA detection in fixed cells and tissues

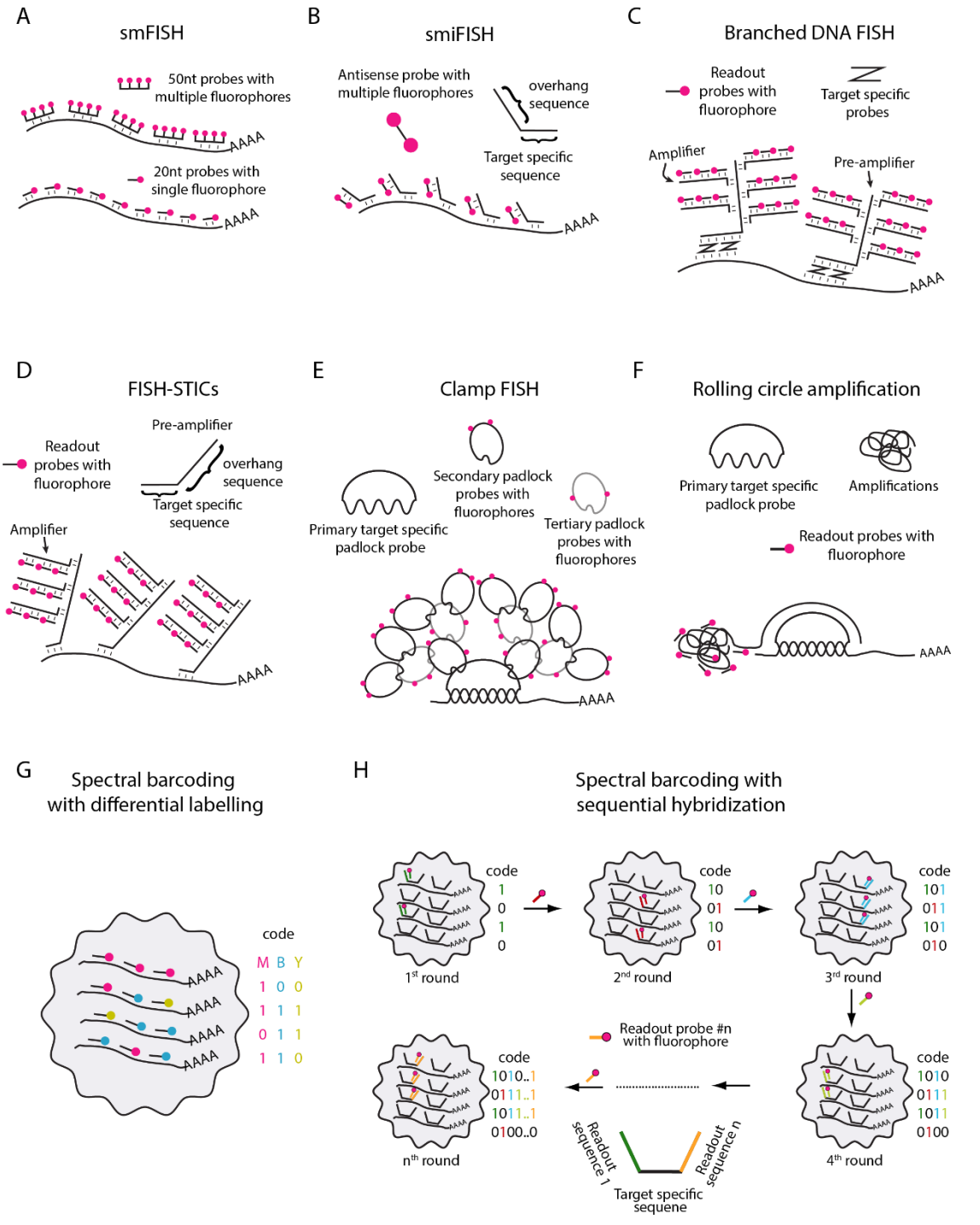
Most studies involving mRNA detection using microscopy have been performed in fixed cells and tissues. The main reason to work in a fixed cell environment is largely technical, allowing access

to a wider range of tools and methodologies that are easier to implement and requiring less sophisticated microscopy setups to image, making them a preferable choice over live-cell imaging. Moreover, due to the crudeness of sample preparation and the destructive nature of high energy electron beams, EM studies have to be restricted to fixed cells. Two approaches are used generally to visualize mRNPs – either through direct targeting of the mRNA, or indirect targeting of associated RNA-binding proteins within the mRNA-protein complex. When using RNA as the target for mRNA imaging, the most common tools use anti-sense probes, most often DNA probes of various lengths, that hybridize specifically to an mRNA of interest. These probes can be coupled with labels that can be recognized using either electron microscopy or tomography (EM-in situ hybridization), fluorescent *in situ* hybridization (FISH) or, in the early days of RNA detection, radioactivity. The use of fluorescent dyes instead of heavy metals in EM or radioactive materials is advantageous as it allows for multiplexing using probes labelled with spectrally differentiable fluorescent dyes. Early RNA studies were often limited to the detection of either highly abundant mRNAs, or mRNAs that show high local concentrations within specific cellular compartments, such as localized RNAs, largely due to the limited sensitivity and low signal to noise ratio of RNA FISH when using single and often long (>1kB) fluorescent probes. The development of methods that allowed for detection of single mRNA molecules in cells, independent of their abundance, represented a milestone in RNA imaging and opened the door for more quantitative approaches to mRNA imaging in cells. However, adoption of the technique as the standard tool for cellular mRNA imaging was a slow process. The development of single molecule resolution RNA FISH (smFISH) by the Singer laboratory in 1998 was the first of many crucial steps towards this process (Femino et al., 1998). In a seminal paper by Femino et al., multiple DNA oligonucleotides probes ~50nt in length were targeted to hybridize with the beta and gamma actin mRNAs allowing for the simultaneous detection of single mRNAs molecules of multiple transcripts within the same cell (Figure 1A top). However, the limited availability of sensitive cameras and high end imaging equipment, combined with the need for custom synthesis of densely labelled probes that were both expensive and harder to generate for laboratories that did not had access to a DNA synthesizer, limited the adoption of the technique. Over the last decade however, various modifications to the initial approach have been made that have made single mRNA detection much more accessible. The approach that is currently most widely adopted uses 35-50 DNA oligo-nucleotides, each 20nt in length and coupled to a single fluorescent dye. The probes are hybridized in paraformaldehyde

fixes cells in low formamide concentrations, resulting in robust single molecule detection (Figure 1A bottom) (Raj et al., 2008). The high signal to noise ratio observed for single mRNAs has seen its wide adoption for mRNAs imaging in many organisms, cells and tissues. A more cost-efficient adaptation of this approach, termed single molecule inexpensive FISH (smiFISH), has also been developed that uses target specific probes containing a gene specific sequence as well as an overhang that can hybridize with a common set of fluorescently labelled anti-sense probes (Figure 1B) (Tsanov et al., 2016). Additionally, alternative approaches have been successful implemented, using either branched probes (Sinnamon and Czaplinski, 2014; Wang et al., 2012), rolling circle amplification (Larsson et al., 2010) or click chemistry to padlock probes to the target mRNA or probes hybridized to the target mRNA (Rouhanifard et al., 2019) to increase signal amplification (Figure 1C-F) . Furthermore, to overcome the low detection capability of RNA FISH due to the limited availability of spectrally differentiable fluorophores, spectral barcoding has often been used, either by using probes labelled with specific combinations of dyes for specific RNAs, or through sequential rounds of hybridization with a subset of probes followed by rounds of imaging and stripping of hybridized probes, and has allowed for the detection of tens to hundreds of RNAs in the same cell (Figure 1G-H) (Chen et al., 2015; Codeluppi et al., 2018; Eng et al., 2017, 2019; Jakt et al., 2013; Levsky et al., 2002; Lubeck and Cai, 2012; Lubeck et al., 2014; Wang et al., 2018). In addition to development of new techniques for improved mRNA detection, smFISH has also been combined with super-resolution microscopy approaches like structured illumination microscopy (SIM) to determine intramolecular organization of RNAs (Adivarahan et al., 2018) and see below.

mRNAs in cells are part of mRNPs and mRNAs can also be visualized indirectly by visualizing protein bound to mRNAs, either using antibodies to specific RNA-binding proteins (RBPs) or using fluorescent protein fusions. Similar to FISH probes, antibodies can be conjugated either with fluorescent dyes or heavy metals to be imaged using either fluorescent or electron microscopy respectively. However, there are important differences to direct RNA detection, as most RBPs bind to many different mRNAs, and in addition, exist in cells in RNA-bound, as well as in free fractions. Imaging RBPs therefore reveals a different kind of information than the RNA centric information obtained from hybridization approaches that target specific transcripts. Nevertheless, combining RBP imaging and FISH is a powerful tool to study regulatory mechanisms acting on mRNAs.

For electron microscopy and tomography studies, mRNPs can be labelled using heavy metal salt solutions such as uranyl acetate and lead citrate. These salts can react with cellular structures including RNA and RNA-binding protein to increase their contrast when imaging with electron microscopy (Bozzola and Russell, 1999). This methodology can either be used alone or combined with EM-ISH or antibody-based targeting of RNA-binding proteins to further increase the labelling of mRNPs or identification of specific proteins as part of the mRNP complex. More recently, the advent of cryo-electron microscopy/tomography has made it possible to determine structures of different RNA-protein complexes *in-vitro* without the need for crystallization (Kühlbrandt, 2014), however its usage in imaging mRNPs in cells might be limited as mRNPs are heterogeneous both in protein and mRNA composition. Moreover, the crowded environment of the cell combined with the low contrast while imaging, has limited the usage of cryo-EM to specific regions of the cell, where it is possible to spatially separate mRNPs (Mahamid et al., 2016).



Adivarahan and Zenklusen Figure 1

***Figure 1: Methods used to visualize mRNA in fixed cells**. **(A)** Single molecule RNA in situ hybridization (smFISH) uses either multiple 50nt ssDNA oligos labeled with multiple dyes, or 20nt ssDNA oligos labelled at a single position. **(B)** smiFISH uses a 20-35nt target specific sequence plus a 28nt overhang which can hybridize to an antisense FLAP probe coupled to fluorescent dyes. **(C)** Branched DNA FISH requires hybridization of two gene specific probes to allow the hybridization of a pre-amplifier and subsequent amplifier probes, that are then detected with dye labelled readout probes. **(D)** FISH-STICS is similar to Branched DNA FISH, with the pre-amplifier sequence present as an overhang to the gene specific probe. **(E)** and **(F)** Padlock based systems for detection rely on single-stranded target specific probes with ligatable ends. ClampFISH uses multiple round of hybridization with padlock probes with each round amplifying the signal. Rolling circle amplification uses one padlock probe from which the signal can be amplified. **(G)** and **(H)** Spectral barcoding approaches to detect multiple mRNA targets either through differential labelling (G), or multiple round of hybridization and stripping to generate unique barcodes for a specific mRNA. See text for more details.*

1.2. RNA detection in living cells

Cellular processes are dynamic and the various steps within the gene expression pathway involve spatial progression through different cellular structures and compartments. Investigating such dynamic processes is limited when using approaches that rely on cell fixation and various imaging techniques have been developed to visualize mRNPs in living cells (Figure 2). Early approaches relied on hybridizing single fluorescently labelled probes to target mRNAs. These probes contained either a single fluorescent dye, a FRET dye pair or were designed such as to contain a dye and quencher on the same oligonucleotide sequence where the dye is quenched when the probe is not hybridized to its target (Figure 2A, B) (Bao et al., 2009; Molenaar et al., 2001; Santangelo et al., 2004; Tyagi and Kramer, 1996). However, the usage of these probes for visualizing mRNAs in live cells was challenging and often limited due to their low signal to noise ratio, fast degradation and difficulty to permeate through the cell membrane, requiring the use of delivery methods such as microinjection, electroporation, cell membrane permeabilization or packaging in cell-penetrating peptides (discussed in (Bao et al., 2009)).

To overcome many of these drawbacks, aptamer-based RNA visualization approaches have been developed that allow detection of RNAs using either endogenously expressed fluorescent proteins

(aptamer-protein combination) or through membrane permeable fluorescent dyes (aptamer-dye combination) (Figure 2C, D). The most commonly used aptamer-protein combinations are derived from bacteriophage capsid proteins that bind with high affinity and specificity to short stem-loop RNA structures. Because they are derived from bacteriophages, these proteins do not have endogenous targets in eukaryotic systems. Coat-protein/RNA stem loop combination of the MS2 and PP7 bacteriophages are most frequently used, but other combinations, such as lambda N, or U1A have also been used (Brodsky and Silver, 2000; Daigle and Ellenberg, 2007; Urbanek et al., 2014). Insertion of a specific aptamer sequence to an RNA of interest, and co-expression of a coat-protein fused to a fluorescent protein results in a fluorescently labelled RNA. However, insertion of a single stem-loop does not allow for detection of single RNAs, and aptamer sequences need to be multimerized to amplify the signal. To obtain robust single molecule sensitivity, typically 12-24 stem-loops are inserted to an mRNA of interest, often within the 3' untranslated region. Over the years, many modifications have been made to the system to fine tune signal to noise ratio and to adopt the system for the study of specific processes, either through modifications to the RNA aptamer sequences or through dimerization of the proteins (Tutucci et al., 2017; Wu et al., 2012, 2015a). Aptamer labelled RNAs are either ectopically expressed, integrated into genomic loci or the aptamer sequence can be integrated to endogenously expressed mRNAs. Common in lower eukaryotes such as *S cerevisiae* for a long time, genomic integration only recently got adapted in higher eukaryotes using different genome editing approaches such as TALEN (Ochiai et al., 2014) or CRISPR/Cas9 (Spille et al., 2019).

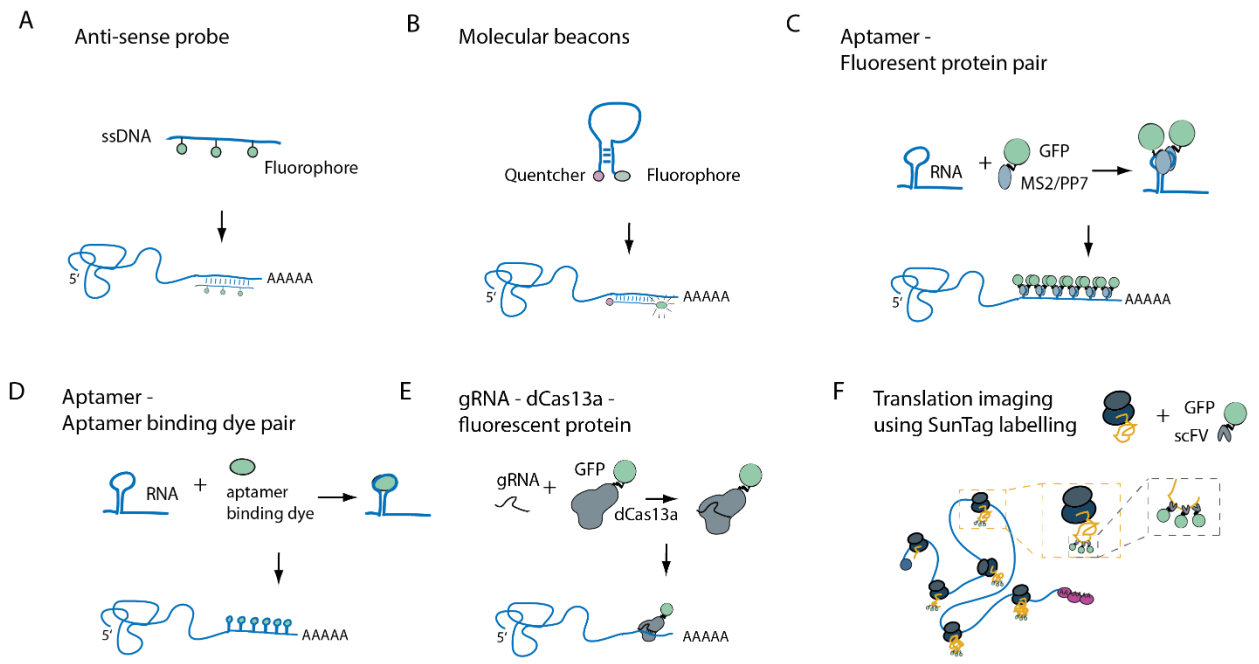
One limitation of the MS2/PP7 systems is that it requires the expression of the aptamer binding proteins fused to a fluorescent protein. To circumvent this problem, dye binding aptamers have been developed such as Mango and Spinach (Dolgosheina et al., 2014; Paige et al., 2011). Aptamer-dye combinations provide a distinct advantage in terms of theoretically stronger signal because of the use of organic dyes that are generally brighter and more photostable than fluorescent proteins, and, for dyes that change their fluorescent properties upon binding to the aptamer, can further reduce background. However, nevertheless of having great potential, aptamer-dye pairs have not yet shown to result in robust single molecule detection, possibly due to issues in RNA folding, cell permeability and or/dye binding properties. Aptamer based imaging systems are discussed more in detail in (Dolgosheina and Unrau, 2016; Urbanek et al., 2014). Moreover, the limitations of aptamer-based methods still apply to aptamer-dye combinations requiring genetic

manipulations to insert aptamer sequences to the RNA of interest and is therefore laborious for studying endogenously expressed RNAs.

More recently, a CRISPR-Cas9 based method was developed that uses an RNA-targeting Cas9 to recognize RNAs of interest (Nelles et al., 2016), as well as Cas13a, which directly binds to RNA (Abudayyeh et al., 2017) (Figure 2E). Though successfully applied for detecting the population of highly abundant endogenously expressed mRNAs, signal-to-noise ratio sufficient for single molecule detection has not yet been reported.

In addition to new methodologies for mRNP imaging in cells, tools for image analysis have simultaneously been developed with the aim to facilitate detection, localization and tracking of single RNA molecules. Single particle tracking algorithms initially developed for tracking receptor diffusion on cell surfaces were later utilized to track of single molecules in cells with a very high spatial accuracy (~10-20 nm) (Cherry et al., 1998). These algorithms were further developed by Thompson et al to enable sub-diffraction resolution localization of single particles for a wide range of circumstances (Thompson et al., 2002). To overcome the resolution limit determined by the wavelength of the light, the signal emitted from individual spatially distinct particles was fit to a 2D-Gaussian with the centroid of the Gaussian being able to determine the localization to a very high precision. This and similar approaches have since been widely adapted to create tools for localization and counting of single mRNPs in fixed cells and localization, as well as counting and tracking of single mRNPs in living cells (Jaqaman et al., 2008; Lionnet et al., 2011; Mueller et al., 2013; Tinevez et al., 2017).

Together, these imaging techniques have been used to study various aspects of mRNP metabolism, starting from transcription to degradation, as well as have been used to study biophysical properties of mRNPs. Below, we will discuss how imaging approaches have contributed to the current understanding of these processes.



Adivarahan and Zenklusen Figure 2

Figure 2: Tools for life cell RNA imaging. Cartoons illustrating different methods to visualize mRNA in living cells. ***(A)*** *Anti-sense probes. Singe stranded DNA or RNA probes labeled with fluorescent dyes can be inserted into cells using different transfection or injection strategies where they hybridize to specific mRNAs.* ***(B)*** *Molecular beacons change their fluorescent properties when binding to a target mRNA, thereby reducing background.* ***(C)*** *Aptamer – fluorescent protein pair. RNA stem-loops bound with high affinity and specificity by RNA-binding proteins such as the capsid proteins from the bacteriophages MS2 and PP7, which when fused to a fluorescent protein result in a fluorescent labelled RNA.* ***(D)*** *Aptamer – aptamer binding dye pair. Molecules designed to bind to RNA aptamers such as spinach or Mango can result in fluorescent RNAs.* ***(E)*** *Cas13a RNA imaging. Cas13a binds RNA specifically, mediated by a guide RNA (gRNA). Co-expressing a gRNA and a catalytic dead mutant termed dCas13a fused to a fluorescent protein results in a fluorescent labelled mRNA. Multiple gRNAs to an mRNA are used to enhance the signal.* ***(F)*** *Translation imaging using the SunTag peptide labeling system. An anti-body fused to GFP (scFv-GFP) that recognizes a short multimerized peptide sequence at the N-terminus of a nascent protein allows imaging of translating mRNAs. See text for more details.*

2. Visualizing nuclear (pre-) mRNPs

The life of an mRNA starts with its synthesis by RNA polymerase II, when mRNAs are produced as precursors that require extensive processing and maturation before being released from chromatin to find their way to the nuclear periphery to be exported to the cytoplasm for translation. Imaging has been an important tool to study all aspects of nuclear RNA metabolism, including mRNA synthesis and processing, and has revealed important aspects of the kinetics and dynamics of these various processes (Figure 3).

2.1. RNA imaging to study transcription

Transcription regulation is a complex process initiated by recruitment of the pre-initiation complex at the promoter region, a process itself influenced by a multitude of factors including transcription factors, chromatin remodelling and interaction with regulatory elements such as enhancers (Hager et al., 2009). Extensively studied for a long time using different experimental systems and approaches, including *in vitro* assays to determine binding affinities of TF and the role of general and specific factors in modulating the transcription reaction, the emergence of RNA imaging to study transcription quickly revealed the limitations of some of these approaches to recapitulate many aspects of transcription regulation in the context of a living cell (Coulon et al., 2013). One factor that made application of in particular *in vitro* studies to living systems difficult, is that many regulatory elements that have to assemble at the promoters are present only in finite numbers in cells. This implies that transcription in cells can best be described as a stochastic rather than a purely deterministic process, an effect of which would result in high variability of RNA numbers expressed in different cells, even between clonal cells grown under identical conditions. Indeed, the stochastic nature of transcription has since been described in many different organisms, mainly by using two experimental approaches. One approach uses the ability to determine cellular mRNA levels as a measure for transcription output, similar to measuring mRNA levels using RNAseq or qRT-PCR, but at the single cell level, typically using some variant of smFISH. In addition to total RNA, smFISH also allows determining the presence and numbers of nascent transcripts at individual loci, revealing transcriptional activity. These two measurements can be combined with modelling approaches to describe transcription behaviour in single cells, and this has been applied in many studies (Bartman et al., 2019; Halpern et al., 2015; Paré et al., 2009; Raj and van Oudenaarden, 2008; Raj et al., 2006; Senecal et al., 2014; Zenklusen et al., 2008). Alternatively,

transcription can be monitored in real time by inserting aptamers repeats and measuring the intensity of fluorescence signals of nascent mRNAs. As each initiation and termination event leads to fluctuation in transcription site intensity, these measurements can reveal transcription dynamics, including initiation frequencies (Chubb et al., 2006; Darzacq et al., 2007; Golding et al., 2005; Larson et al., 2011; Muramoto et al., 2012; Yunger et al., 2010). Such measurements have revealed important features on how genes are transcribed that could not be obtained using classical approaches. One such observation was that most genes are transcribed in a discontinuous manner, where periods of active transcription are interspaced with periods where there is no new initiation by RNA polymerase II (Chubb et al., 2006; Golding et al., 2005; Muramoto et al., 2012). Thereafter, various studies have showed that both the duration of 'on' and 'off' periods, as well as the initiation frequency during the 'on' time, often described as a transcription burst, are extensively regulated. Factors such as histone modifications, promoter architecture, binding of transcription factors, formation of enhancer-promoter loops, cell volume and position of genes in the genome were all found to regulate transcription bursting and is something that continues to be extensively studied (Bartman et al., 2016; Chen and Larson, 2016; Lenstra et al., 2016; Nicolas et al., 2018; Padovan-Merhar et al., 2015; Raj and van Oudenaarden, 2008; Raj et al., 2006; Senecal et al., 2014; Suter et al., 2011).

Assays used to study transcription initiation have also been used to measure the speed of an RNA polymerase along the template, either by modelling smFISH data or by correlating signal fluctuations from time traces of aptamer labelled RNAs. These measurements revealed a high amount of variability in the elongation speed of RNA polymerase II, ranging from ~25 nt/sec in E coli (Golding and Cox, 2004; Golding et al., 2005) to 14-1000 nt/sec in eukaryotes, (Ben-Ari et al., 2010; Brody et al., 2011; Darzacq et al., 2007; Femino et al., 1998; Hocine et al., 2013; Larson et al., 2011; Maiuri et al., 2011; Wada et al., 2009; Yunger et al., 2010). Moreover, transcription elongation rates were found to vary from cell to cell, with some of the variations sensitive to the cell cycle (Hocine et al., 2013; Larson et al., 2011). One cause of the variability in elongation rates was attributed to RNA polymerase pausing, previously suggested by ChIP studies that showed non-uniform distribution of RNA polymerase II across genes, with intermittent spikes (Churchman and Weissman, 2011; Jonkers et al., 2014; Zeitlinger et al., 2007). This was confirmed by FRAP studies on the transcription site of MS2 labelled mRNA, revealing that RNA polymerases can stochastically pause during the elongation step, or at the 3' terminus post the polyadenylation site

(Boireau et al., 2007; Darzacq et al., 2007). In addition to stochastic pausing events, ChIP results also indicated the RNA polymerases could pause throughout the body of the gene with particularly enrichment near the promoter (called promoter-proximal pausing), before nucleosome dyads and at intron-exon junctions, suggesting a wider role for RNA polymerase in regulation of gene expression (Churchman and Weissman, 2011; Kwak et al., 2013; Lenstra et al., 2016). However, single molecule FRAP observations on reporter mRNAs did not observe pausing in the body of intron containing genes (Brody et al., 2011), indicating that pausing might not be a universal for all intron-exon junctions.

mRNA detection in single cells has allowed investigating the relationship between mRNA numbers and cell size, showing that in higher eukaryotes, mRNA concentration scales with cell size for most genes (Kempe et al., 2015; Padovan-Merhar et al., 2015). Although single cell sequencing is also able to measure mRNA numbers in single cells, it still lacks the sensitivity to detect low abundant transcripts, and, in addition, does not provide spatial information, characteristics all preserved with smFISH.

2.2. pre-mRNA maturation

mRNAs are first transcribed as precursors and the process of mRNA maturation involves multiple processing steps, including modification of the 5', excision of introns, 3' end cleavage and polyadenylation as well as modification of various bases. Many of these processes occur co-transcriptionally, are coupled to each other and are essential to ensure that mRNPs are properly assembled to allow their export and subsequent translation. Already early into the discovery of RNA processing steps, imaging provided important clues into this complex process. Hybridization approaches combined with electron microscopy were critical for the discovery of introns, when experiments in the Roberts and Sharp laboratories showed DNA segments looped from DNA-RNA R-loop regions, when RNA was hybridized to viral genomic DNA fragments (Berget et al., 1977; Chow et al., 1977). Similarly, electron micrographs of chromatin spreads from *Drosophila melanogaster* showed RNP assemblies at the intron-exon junctions which were later identified as spliceosomes, indicating that splicing might be a co-transcriptional process (Osheim et al., 1985). Co-transcriptional spliceosome assembly was later shown by many studies and approaches, including using variants of chromatin immunoprecipitation that allowed to crosslink snRNPs and splicing factors to chromatin in transcription and/or splicing dependent manner (Alpert et al., 2017;

Görnemann et al., 2005; Kotovic et al., 2003) as well as using an *in vitro* TIRF microscopy system with labelled RNAs and spliceosome components (Hoskins et al., 2011). Dynamics of spliceosome association at sites of transcription was further studied using single molecule microscopy approaches either using antibodies against U snRNP proteins or FISH probes against U snRNAs (Brody et al., 2011; Schmidt et al., 2011; Wetterberg et al., 2001) Interestingly, it was found that recruitment of the U1 snRNP to active transcription sites could occur independent of the presence of introns in the pre-mRNA, indicating an RNA independent recruitment possible mediated by RNA pol II (Brody et al., 2011). The same study found that mRNAs with higher number of introns had more spliceosome components recruited to the transcription site, suggesting that multiple spliceosomes could potentially assemble onto the same mRNA. However, the number of spliceosomes acting on the pre-mRNA is likely to vary depending on the strength of the 5' and 3' splice sites, the presence of RNA secondary structures and splicing of adjacent introns. Moreover, it is not clear which proteins are recruited as preassembled complexes and which join as individual proteins. Co-transcriptional recruitment to sites of transcription or loading onto pre-mRNPs was also observed for various splicing regulators, including several SR proteins either using immunofluorescence or immuno-EM (Björk et al., 2006, 2009; Brody et al., 2011; Misteli et al., 1998; Wetterberg et al., 1996)

Co-transcriptional assembly of splicing factors resulting in the co-transcriptional splicing was first observed on chromatin miller spreads from *Drosophila* embryos. These electron micrographs showed nascent pre-mRNA with multiple stages of intron excision with loops of introns 5' and 3' in the process of getting excised (Beyer and Osheim, 1988). Since then, co-transcriptional splicing has been reported for several intron containing mRNAs using either dual colour RNA labelling where specific intron and exon sequences were labelled with MS2, lambda N or PP7 aptamer sequences allowing to monitor splicing in live cells, or by *in situ* hybridization in fixed cells (Brody et al., 2011; Coulon et al., 2014; Martin et al., 2013; Schmidt et al., 2011; Vargas et al., 2011). These experiments also showed that not all introns are spliced at the site of transcription, as a small fraction of intron containing RNAs either was observed to be retained at the site of transcription, close to the site of transcription or within in the nucleoplasm, with some indication that splicing of mRNAs is enhanced post-transcriptionally (Boireau et al., 2007; Brody et al., 2011; Coulon et al., 2014; Vargas et al., 2011; Waks et al., 2011). Co- or post-transcriptional splicing might be defined my many factors, including the position of an intron within a pre-mRNA, the strength of

splice sites and possibly other regulatory processes that might facilitate faster or slower splicing of specific introns. Splicing of the first intron was shown to stimulate transcription elongation as well as splicing of the following introns and recent studies also suggest that splicing of neighbouring introns can influence the choice between different splice sites (Blazquez et al., 2018; Boehm et al., 2018). Moreover, intron retention has recently been suggested as a mechanism to regulate mRNA and protein expression, and imaging approaches will likely play an important role on dissecting the mechanisms of this regulatory process (Bahar Halpern et al., 2015; Wegener and Müller-McNicoll, 2018).

When reaching the 3' of a gene, RNA polymerase has to terminate transcription and the mRNAs are cleaved and polyadenylated before being released into the nucleoplasm. Studies using various experimental approaches have shown that termination, cleavage and release of the mRNA are possibly linked to other transcriptional process including elongation and splicing (Bentley, 2014; Kyburz et al., 2006; Niwa and Berget, 1991). Single molecule live cell imaging was able to visualize polymerases stalling at the 3' end of the transcript and to determine the relationship between splicing and release of mRNAs from the transcription site (Coulon et al., 2014; Martins et al., 2011). Using MS2 labelled mRNAs, it was found that inhibition of splicing using spliceostatin A did not result in increase in release time of the mRNAs, on the contrary, it was found that beta-globin mRNAs were released faster after treatment with the drug, suggesting that RNA polymerases could pause at the 3' end of the gene as a quality control mechanism. Similar experiments were used to measure the post-transcriptional dwell times of transcripts at sites of transcription. These measurements showed varying release times for different genes and across the cell cycle, ranging from 60 s to 8 min (Boireau et al., 2007; Coulon et al., 2014; Darzacq et al., 2007; Larson et al., 2011). The mechanisms modulating mRNPs release are poorly understood and a few models have to proposed to explain the observations (Lenstra et al., 2016).

2.3. Movement of mRNPs within the nucleoplasm

It is thought that nascent mRNA do not exist in cells as long extended polymers, but that pre-RNAs are co-transcriptionally folded and packaged into pre-mRNPs, a process mediated in part by RNA-binding proteins, many of which contain homo or hetero dimerizing domains (Singh et al., 2015). Very little is known about how (pre-) mRNP formation is achieved, and much of our knowledge comes from electron microscopy experiments visualizing the long BR mRNPs expressed from

polytene chromosomes of the dipteran *Chironomus tenants*. Due to the large size of mRNAs expressed from the *BR1*, *BR2.1*, *BR2.2* and *BR6* loci (between 35-40kB in length), the resulting mRNPs are sufficiently electron dense to be visualized using election microscopy, and have been an extremely valuable model system to study different aspects of mRNP metabolism (Björk and Wieslander, 2015). These studies showed that mRNA packaging begins sequentially with the formation of a short 19-20 nm thick fibre which is later packaged into a globular particle of ~50 nm diameter, before being released into the nucleoplasm (Skoglund et al., 1986). Complementing these studies, measuring diffusion characteristics of BR mRNPs using single molecule tracking of mRNPs labelled using complementary oligonucleotides suggest similar sized particles for these BR mRNPs (Siebrasse et al., 2008). A somewhat different kind of organization was observed for nuclear 18kB long MDN1 mRNPs in human tissue culture cells using smFISH and super-resolution microscopy. These mRNPs were found to have a more linear architecture and were similar to purified nuclear mRNPs from yeast, which showed elongated, rod-like structures with variable length but a constant width when visualized by electron microscopy (Batisse et al., 2009; Adivarahan et al., 2018). Such a linear organization is also consistent with data from a recent developed RNA-RNA proximity ligation approach (Metkar et al., 2018). Due to the limited number of studies investigating the organization of nuclear mRNPs it is still difficult to assess whether there exists a universal mechanism that mediates organization for mRNPs in the nucleus.

Once released from the site of transcription, mRNPs need to reach the nuclear pore to be exported to the cytoplasm. Though very early studies suggested that there might be directed movement of mRNPs from the site of transcription to the nuclear pore, as stated by the 'gene gating hypothesis' proposed by (Blobel, 1985) , various studies using either EM or fluorescent microscopy approaches since then have shown that mRNPs move within the nucleoplasm and reach the nuclear pore through diffusion. First indications for this non-directed movement came again from visualizing nuclear BR mRNPs which were observed to have a random distribution within the nucleoplasm, suggesting that these mRNPs do not have a defined path from the site of transcription towards nuclear pores, but possibly diffuse throughout the nucleoplasm in a random manner (Singh et al., 1999). However, the first direct measurement of diffusion kinetics of nuclear mRNAs used fluorescently labelled oligo dT probes that when allowed to penetrate cells hybridized to nuclear poly(A) RNA and permitted monitoring diffusion of all poly(A) RNAs in cells. These studies showed that poly(A) RNA move freely within in the non-chromosomal space of the nucleus with

properties characteristic of diffusion (Politz et al., 1999). Thereafter, various other studies have found mRNPs to have a wide distribution of diffusion coefficients (Calapez et al., 2002; Molenaar et al., 2004; Politz et al., 1998, 1999). Using single particle imaging approaches such as anti-sense oligonucleotides targeted to specific mRNA or using the MS2 tagging system, it was then revealed that nuclear mRNP diffusion was, although random in its movement, restricted to the extranucleolar space (Mor et al., 2010; Shav-Tal et al., 2004; Siebrasse et al., 2008; Vargas et al., 2005). Moreover, diffusion was slowed while passing through high density chromatin, suggesting possible interactions with chromatin, and resolution of this stalling required the presence of ATP (Miralles et al., 2000; Shav-Tal et al., 2004; Vargas et al., 2005).

At least for some mRNPs, the path taken from the site of transcription towards the nuclear pore might be more complex than simple diffusion through the interchromatin space to reach the nuclear periphery. mRNAs containing inverted Alu repeats elements in their 3' UTRs have been shown to localize to paraspeckles, from where they can be released upon further processing or binding of specific RNA-binding proteins that promote their export. A first example for such localization was the CTN-RNA, an alternatively processed transcript expressed from the mCAT2 locus that contains alternative 5' and 3' UTRs but is otherwise identical to the protein-coding mCAT mRNA. The longer CTN-RNA 3'UTR contains Alu-like SINE repeats that are A-to-I edited, resulting in the RNA localizing to paraspeckles. Upon stress, the transcript is processed to the mCAT2 mRNA and transported to the cytoplasm (Prasanth et al., 2005). Alternatively, paraspeckle localization of different Alu repeat containing mRNAs was shown to be mediated by the binding of the Staufen 2 protein (Elbarbary et al., 2013), however the mechanism that results in this localization is not known. Paraspeckles are often located adjacent to nuclear speckles, nuclear domains located in the interchromatin regions and enriched in splicing factors, poly(A) RNAs and noncoding RNAs, however, the abundance of paraspeckles is lower than that of nuclear speckles (Galganski et al., 2017; Staněk and Fox, 2017). The dynamics of mRNP localization to the paraspeckles remains unclear. It is possible that mRNPs might be transcribed and spliced in or close to nuclear speckles and are subsequently handed over to paraspeckles. Alternatively, it remains possible that mRNAs once released from the transcription site diffuse though the interchromatin space to reach either the nuclear speckles or paraspeckles. In addition to Alu containing mRNAs, recent studies have shown that many other mRNAs are retained within the nucleus and that the process of mRNA retention is regulated. Combining fractionation with RNA sequencing and smFISH, Bahar Halpern

et al showed that in different mouse metabolic tissues such as beta cells, liver, and gut, many mRNAs showed varying retention within the nucleus depending on exposure to different metabolic conditions (Bahar Halpern et al., 2015). Similarly, a study aimed towards determining expression variability of over 900 different mRNAs in HeLa cell suggested that mRNAs can be nuclear retained and that their slow export buffers expression noise in the cytoplasm (Battich et al., 2015). Furthermore, many transcripts can contain retained introns that leads to their nuclear retention (Wegener and Müller-McNicoll, 2018). However, the mechanistic details on how nuclear retention is achieved and whether these mRNAs are retained in specific subnuclear compartments is not yet known.

2.4. Export through the nuclear pore complex

Diffusion takes mRNPs to the nuclear periphery where they interact with the nuclear pore complexes to be exported. The time for reaching the nuclear periphery widely varies across organisms and largely depends on the size of the nucleus, taking only a few seconds in lower eukaryotes such as *S cerevisiae,* but possibly up to minutes in human cell nuclei (Grünwald and Singer, 2010; Mor et al., 2010; Oeffinger and Zenklusen, 2012; Saroufim et al., 2015; Shav-Tal et al., 2004; Siebrasse et al., 2012; Smith et al., 2015). Live cell single molecule fluorescence microscopy has shown that when mRNPs reach the periphery, they often first scan the nuclear periphery possibly making contact with multiple nuclear pore complexes before stably docking on to the nuclear pore for export (Grünwald and Singer, 2010; Mor et al., 2010; Saroufim et al., 2015; Siebrasse et al., 2012). At the NPC, mRNPs first interact with the nuclear basket, a structure attached to the central framework of the nuclear pore complex that protrudes toward the nuclear interior (Buchwalter et al., 2018). Docking to the NPC has been shown to be a rate limiting step for the export of mRNPs. Different single molecule studies found prolonged residency times of mRNPs at NPC, with some of these studies being able to map prolonged residency at the basket (Grünwald and Singer, 2010; Mor et al., 2010; Saroufim et al., 2015; Siebrasse et al., 2012). Such a rate limiting step might be a result of mRNPs being rearranged at the basket which could result in the release of specific proteins, or the remodelling of mRNPs which might be required to get to access to the NPC and facilitate translocation through the central channel. Indeed EM studies of the BR mRNPs showed that the large BR mRNPs are unfolded at the distal ring of the nuclear basket before entering the basket with the 5’ of the mRNA first (Mehlin et al., 1992). However, it

is not known whether such remodelling is required for all mRNPs, as most mRNPs are at least one order magnitude smaller than BR mRNPs. Moreover, as mentioned above, recent single molecule super-resolution microscopy studies suggest that mRNPs in mammalian cells, as well as mRNPs purified from yeast show a linear organization which could negate the need for such a reorganization (Batisse et al., 2009; Adivarahan et al., 2018).

Once accessing the central framework of the NPC, translocation is a very fast process (Grünwald and Singer, 2010; Siebrasse et al., 2012). Using a super-registration approach to follow the translocation process of MS2 labelled beta-actin mRNAs through the NPC, Grünwald and Singer showed that translocation only takes about ~20ms. Moreover, mRNPs can move in either direction within the central channel, suggesting the directionality is not encoded by the central channel, but by events at either side of the NPC (Grünwald and Singer, 2010). Consistent with such a model, residency times at the cytoplasmic side of the pore are similar to the residency times at the basket, about 80ms. Moreover, mRNP rearrangements, in part mediated by RNA helicases, are thought to be required at the cytoplasmic side of the NPC to facilitate their release to the cytoplasm (Alcázar-Román et al., 2006; Smith et al., 2015; Weirich et al., 2006).

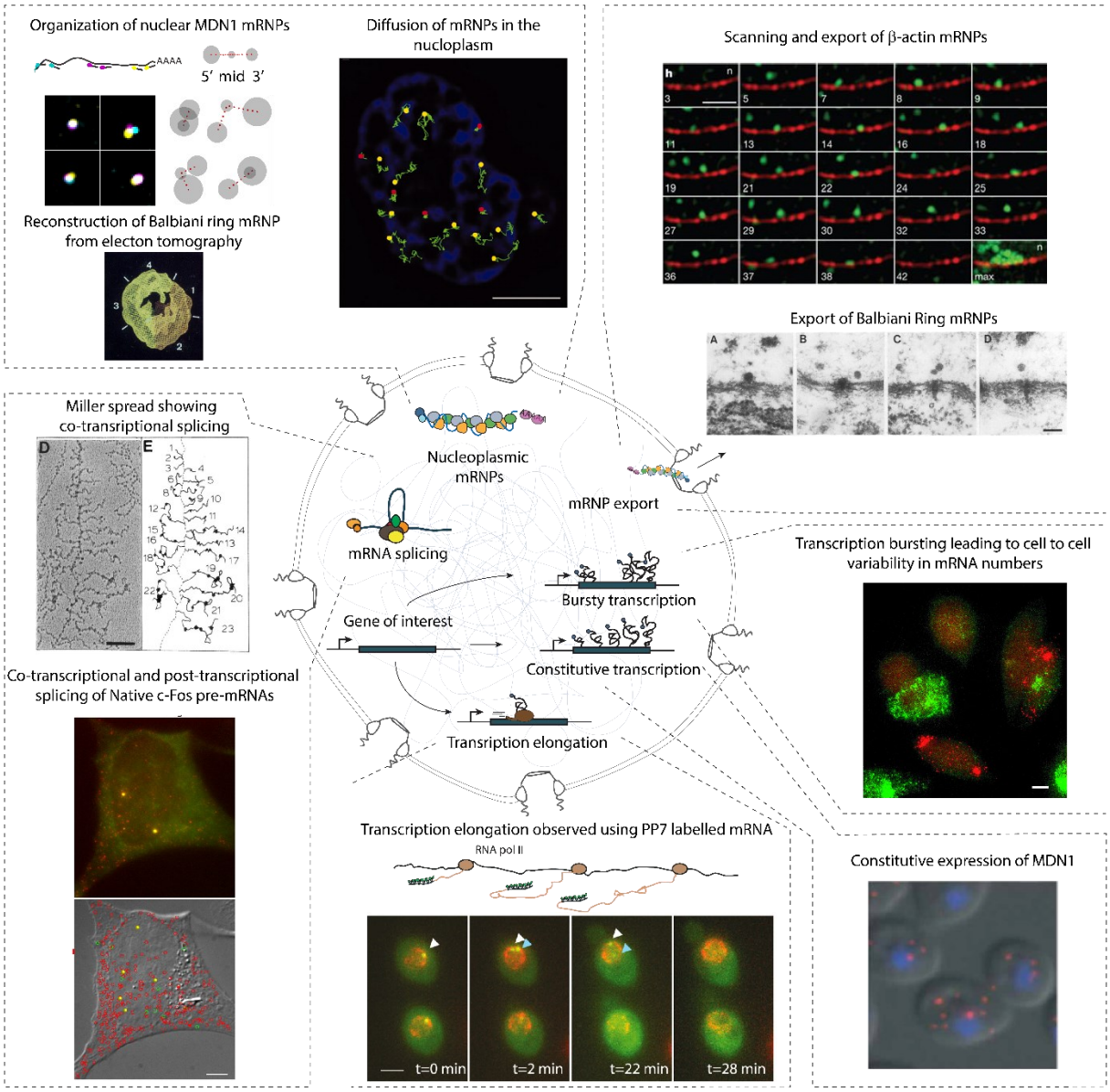


Figure 3: Visualizing nuclear mRNA metabolism. *Clockwise starting at the top left.* ***Organization and dynamics of nuclear mRNPs.*** *(Top) Conformations of nuclear MDN1 mRNAs in HEK293 cells visualized using smFISH from Adivarahan et al., 2018. (Bottom) Reconstruction of BR mRNPs observed with electron tomography. Modified with permission from Mehlin et al., 1992. (Right) Restricted diffusion of mRNPs thorough interchromatin space visualized using molecular beacons. Blue shows DAPI signal. Modified from Vargas et al., 2005. Copyright (2005) National Academy of Sciences, U.S.A.* ***mRNPs export.*** *Scanning and export of beta-actin mRNA*

visualized using MS2 RNA tagging. Figure shows time series of single□□-actin mRNP reaching the nuclear periphery before export. Bottom right panel shows overlay of time-series. Nuclear pores are labeled in red. Modified with permission from Grünwald and Singer, 2010. Export of BR mRNPs observed using electron microscopy. mRNPs were found to dock to the nuclear basket and getting remodeled before exiting through the central channel. Modified with permission from Mehlin et al., 1992. ***Transcription bursting.*** *smFISH illustrating transcription bursting of a inducible reporter (green) and the large subunit of RNA polymerase II (red). Modified with permission from Raj et al., 2006.* ***Constitutive transcription.*** *Image showing constitutive expression of MDN1 mRNA in Saccharomyces cerevisiae using smFISH. Modified from Zenklusen et al., 2008.* ***Measuring transcription elongation****. Images showing expression of PP7 labelled reporter mRNA (green) in live cells. Nuclear pores are shown in red. Panels represents different time points with arrows pointing sites of active transcription. Modified from Larson et al., 2011.* ***mRNA processing.*** *(Top) Electron micrograph of a Miller spread of chromatin isolated from Drosophila embryos and drawing showing tracing of the micrograph. The numbers represent different mRNA templates transcribed. Looping of introns can be observed along with the co-transcriptional splicing of mRNPs identified by the deposition of the spliceosome complexes. Modified with permission from Beyer and Osheim, 1988. (Bottom) Co- and post-transcriptional splicing of c-Fox pre-mRNAs visualized by smFISH. Exons shown in red, introns in green. Bottom shows overlay of the localized signals on a brightfield image with yellow spots representing unspliced pre-mRNAs not colocalizing with transcription sites. Modified with permission from Vargas et al., 2011. See text for more details.*

3. Cytoplasmic mRNPs

The main function for cytoplasmic mRNAs is to associate with ribosomes for translation. However, following their translocation though the nuclear pore complex, many mRNAs are not immediately translated but first transported to various cytoplasmic compartments before associating with ribosomes and initiating protein synthesis. Mechanisms for mRNA localization are diverse, including diffusion followed by local retention to motor driven movement, a process best described in neurons (Buxbaum et al., 2014). Moreover, mRNAs can switch between a translationally active and repressed states upon certain stimuli such as stress, and this is, at least in part, concurrent with their accumulation in membrane-less organelles such a stress granules (SG)

(Guzikowski et al., 2019). Similarly, degradation has been linked to membrane-less organelles called processing bodies (P-bodies) that contain high concentration of proteins involved in RNA degradation and are distinct from SGs. Imaging has been pivotal in the identification and characterization of all these processes, in particular RNA localization and local translation, with many methods now applied to study mRNA metabolism using microscopy were first developed to study mRNA localization, including smFISH and the MS2 aptamer system (Bertrand et al., 1998; Femino et al., 1998). Moreover, recent developments now allow monitoring translation at the single mRNA level in real time, as well as to study localization and mRNA turnover more directly in their relation to translation regulation and dynamic association with phase separated compartments, such as p-bodies and stress granules.

3.1. Discovery of mRNA localization and local mRNA translation

Much like for the study of nuclear mRNA metabolism, electron microscopy and transcript specific fluorescent RNA imaging have both played important roles towards todays understanding of many aspects of cytoplasmic RNA metabolism. When scientist became aware of the extensive cytoplasmic compartmentalization, the question arose as to how the proteins are targeted to these subcellular structures and organelles. One proposed mechanism was that protein targeting would occur through post-translational transport, with mRNAs translated anywhere in the cytoplasm and proteins finding their final location by diffusion or through some active transport mechanism. However, electron microscopy visualizing polysomes, clusters of ribosomes translating a single mRNA, provided a first indication that translation of specific proteins might occur in a more regulated and localized manner. Polysomes were observed to localize at different cellular structures such as the endoplasmic reticulum (Christensen et al., 1987; Lin and Chang, 1975) and mitochondria (Kellems et al., 1975), as well as within dendritic spines (Steward and Levy, 1982). Later, *in situ* hybridization approaches showed the localization of specific cytosolic protein coding mRNAs, such as the actin coding mRNAs in *Styela plicata* embryos, have a distinct localization pattern (Jeffery et al., 1983). Localization of few specific examples of mRNAs has since been observed in many organisms such as yeast (Long et al., 1997), *Xenopus* (Melton, 1987; Melton and Yisraeli, 1988), *Drosophila* (Akam, 1983) and mammalian cells (Lawrence and Singer, 1986). For a long time thought to be a process restricted to only few specific transcripts, a high-throughput

in situ hybridization study that surveyed localization patterns revealed that 71% of the 3370 transcripts examined preferentially localized to distinct subcellular compartments in *Drosophila* embryos, suggesting that RNA localization, and possibly localized translation is the rule rather the exception (Lécuyer et al., 2007). Similarly, recent studies have identified subcellular localization of a large number of mRNAs in specialized cells like neurons, as well as single celled eukaryotes like yeast, further establishing the role of mRNA localization in regulation of gene expression (Cajigas et al., 2012; Gonsalvez et al., 2005; Jung et al., 2014). mRNA localization and localized translation offers distinct advantages compared to protein targeting trough diffusion, in particular in larger cells such a neurons, but also in dividing cells (Buxbaum et al., 2014). Localized translation can quickly increase the local concentration of proteins, circumventing time and energy that would otherwise be required to transport each individual protein molecule. Moreover, a single mRNA that can undergo many rounds of translation, and therefore allow a fast response to stimuli at the site of localization, either by increasing or decreasing translation. Additionally, localization of mRNA might help restrict synthesis and hence localization of proteins to subcellular compartments which could be essential in case the proteins are either toxic to the cell or have alternate functions based on their localization.

RNA imaging has also been extensively used as a readout when determining the mechanisms that mediate mRNA localization and localized translation across different transcripts and organisms. mRNA localization commonly depends on the presence of cis-acting localization elements, often called ‘zip-codes’, and frequently located within the 3’ UTR of an mRNA. *In situ* hybridization is most often used as a functional readout for localization, such as during the characterization of one of the first RNA localization elements located within 3’ UTR of the chicken β-actin mRNA that mediates the localization of the mRNA to the leading lamellae of chicken embryo fibroblasts . The short sequence was shown to be sufficient to mediate localization of the reporter mRNA when isolated from its host RNA context and placed into a reporter RNA, an assay often used to define localization sequences. In addition, mutations to the localization element deterred, but did not abolish the localization of mRNAs, suggesting that multiple elements within the sequence could play a role (Kislauskis et al., 1994). The same study determined that this ‘zip-code’ was conserved across species both in sequence and function, as replacing the 3’-UTR with one from the human β-actin gene did not alter localization pattern for β-galactosidase mRNAs in chicken cells. Similar localization elements have been found for mRNAs in other organisms targeting mRNAs to various

cellular compartments, including the ASH1 mRNA in yeast, (Bertrand et al., 1998; Long et al., 1997; Takizawa et al., 1997), Vg1 in Xenopus (Mowry and Melton, 1992), oskar (Kim-Ha et al., 1993), nanos (Gavis and Lehmann, 1992; Gavis et al., 1996), biocoid (Macdonald and Struhl, 1988) in *Drosophila* and MBP mRNA in neurons (Ainger et al., 1997). In addition to the use of FISH, the RNA aptamer systems has been extensively applied to the study of RNA localization in live cells. The MS2 system was first developed to study the localization of the *ASH1* mRNA to the bud tip of the daughter cell in dividing cells in *S cerevisiae* and allowed to demonstrate a motor driven localization of the mRNA to the daughter cell. This process was found to depend on a protein complex with the myosin protein She3p as a core component, that allowed the mRNA to move along the actin cables (Bertrand et al., 1998). Since then, the MS2 and other aptamer systems have become indispensable tools for studying localization dynamics, with many examples showing motor driven movements, such as in neurons, or diffusion based localization and retention, as observed in fly oocytes (Becalska and Gavis, 2009; Buxbaum et al., 2014; Lee; et al., 2016; Wu et al., 2016).

The cis-RNA localization elements work in conjunction with trans-acting factors, i.e. RNA-binding proteins that bind to these zip-code sequences and are required for transportation of mRNAs to their cellular destination. Many RNA-binding proteins have been implicated in transport of mRNAs, some of which function through interaction with other protein partners that can link them to motor proteins such as myosin, kinesin or dynein (Buxbaum et al., 2014). Examples for trans-acting factors include the Staufen protein, required for the localization of oskar and bicoid mRNAs in *Drosophila*, and the Imp1/ZBP1 and ZBP2 that are important for localization of β-actin mRNA in mammals and Vera for Vg1 localization in *Xenopus* (Deshler et al., 1997; Farina et al., 2003; Hüttelmaier et al., 2005; Johnston et al., 1991; Martin and Ephrussi, 2009). However, the use of imaging in characterizing the role of these proteins in the localization process has been challenging. In comparison to RNA imaging that allows visualizing individual mRNAs, visualizing of single proteins, and in particular their association with mRNAs is still challenging. RBPs are typically labelled using fluorescent proteins or by immunolabeling, but signals from such staining is difficult to attribute to specific mRNPs. Therefore, the readout from RBP imaging is much less direct and can represent both unbound and bound fraction, with the bound fraction possibly representing RBPs associated with multiple mRNA targets. Nevertheless, RBP imaging has been an important tool for studying mRNA localization, as RBPs co-localizing

with mRNAs or in transport granules in neurons were shown to have similar localization dynamics to mRNAs, and can therefore be used to study RNA localization mechanisms. To bridge the gap between the single molecule sensitivity of mRNA imaging and protein imaging, approaches have been developed to measure interactions of RBPs with localized mRNAs, such as fluorescence fluctuation spectroscopy (FFS) which was used by Wu et al to characterize the interaction and stoichiometry of ZBP1 association with β-actin mRNA in living cells (Wu et al., 2015b). However, further technological development is needed in order to use RBPs as targets to monitor mRNP localization dynamics (see outlook).

In addition to the sequence specific 'zip-code' binding proteins, localization of certain mRNAs has been found to depend on proteins deposited during splicing, including proteins that are part of the exon junction complex, however the mechanism behind the role of these proteins in mRNA localization is less well understood (Martin and Ephrussi, 2009). Furthermore, recent studies in *Drosophila* suggest yet another mechanism regulating localization of mRNAs through modulation of local stability of mRNAs within the cell as is observed for *Hsp83* mRNAs (Bashirullah et al., 2001; Martin and Ephrussi, 2009). The different mechanisms regulating mRNA localization are only being unravelled, with the localization elements determining localization of a vast majority of mRNAs yet to be identified. Moreover, it has been shown mRNA localization is linked with translation, with at least some mRNAs getting transported in a translationally silenced form with translation of these mRNAs only initiated upon reception of specific signals at the site of localization (Buxbaum et al., 2014; Halstead et al., 2015; Hüttelmaier et al., 2005; Yoon et al., 2016). Combining recently developed translation imaging assays with single molecule RNA imaging and advancements in proteins imaging will be essential to dissect these processes in even more details (see also below).

3.2. Monitoring translation

Although some mRNAs are translationally repressed after reaching the cytoplasm, many mRNAs are thought to quickly associate with ribosome and start translation. Once more, imaging of BR mRNPs was the first indication for the ability of ribosomes to quickly initiate translation, with ribosomes shown to assemble on BR mRNPs even before the entire mRNP was fully exported to the cytoplasm (Mehlin et al., 1992). However, because of their large size, translocation for BR mRNPs might be slower than for most mRNPs and translation might not initiate as rapidly for

most cellular mRNAs. Nevertheless, using an elegant live cell imaging approach, Halstead et al showed that translation initiation for certain reporter mRNAs can be quite fast. They developed a new translation imaging approach, termed 'Translating RNA Imaging by Coat protein Knock-off (TRICK)', that allows distinguishing between untranslated mRNAs and mRNAs that have undergone at least one round of translation. The TRICK system consists of an mRNA reporter with two aptamer sequences within the body of the mRNA (MS2 and PP7). While the MS2 sequence is placed in the 3' untranslated regions, the PP7 sequence, generally incorporated in the 3' UTR of mRNAs, was instead placed within the open reading frame of the mRNA and the entire sequence was translated along with an upstream ORF. This required modification of the PP7 repeat sequence to separate the individual stem-loops, so that translating ribosomes could efficiently displace the PCP-GFP proteins during the first round of translation. As the PCP-GFP contains a nuclear localization signal, ribosome displaced PCP-GFP would be transported back to the nucleus depleting their abundance in the cytoplasm to allow rebinding, whereas the MS2 signal is maintained. Therefore, any cytoplasmic mRNA will lose one label but maintain the second label after it has been translated at least once. Using this system, they showed that 94% of cytoplasmic mRNA from their TRICK reporter had undergone translation of least once, suggesting that translation occurs most likely within minutes after export to the cytoplasm, if not faster (Halstead et al., 2015). This system was also used to study the role of Oskar proteins in osk mRNA localization and localized translation and will be a useful tool for studying translation regulation in the future. However, one limitation of this assay is that it does not allow to directly test whether mRNAs are actually associated with ribosomes.

One way of attempting to distinguish between translating and non-translating mRNAs started with the reasoning that polysomal mRNAs, part of much larger assemblies in comparison to non-translating mRNAs should show altered diffusion characteristics. Tracking labelled ribosomal proteins together were with MS2/PP7 labelled β-actin mRNAs in living cells found that association with ribosomes significantly slowed down mRNA diffusion in a manner that scaled with the ribosome occupancy (Katz et al., 2016; Wu et al., 2015b) and see below. Furthermore, it was found that mRNAs in focal adhesions showed slowed and confined diffusion suggesting that β-actin mRNAs localized in the regions were heavily translating (Katz et al., 2016)

Though polysomes can be tracked by labelling ribosomes, association of mRNAs with ribosomes does not imply translation in all cases. Moreover, diffusion of mRNAs might become restricted for other reasons than their association with ribosomes. An early attempt to quantify translation in single cells used a reporter mRNA expressing a protein that contained a tetra-cysteine motif in its N-terminus that can be bound by the biarsenial dyes FlAsH and ReAsH. Using pulse-chase labelling in living cells allowed to visualize newly synthesized proteins and to spatially correlate them with the sites of β-actin mRNA localization (Rodriguez et al., 2006). Though it was possible to visualize sites of localized translation, it did not allow monitoring of translation at a single molecule level. This became possible with the development of protein tagging systems that used multiple epitopes within the N-terminus of the reporter protein, which upon expression could amplify the signal of nascent peptides, a concept that mimics the signal amplification achieved by aptamer based systems for imaging mRNAs. Two such tags have been used to image translation at the single molecule level, the SunTag and 'spaghetti monster' (SM) tag (Tanenbaum et al., 2014; Viswanathan et al., 2015). The SunTag system uses endogenously expressed single chain antibody fragments (scFV) against a short epitope of the yeast Gcn4p that when fused to GFP can specifically bind to proteins containing, in general, multiples of this epitope (Figure 2F). The SM-tag contains multimerized epitopes recognized by either anti-myc or anti-Flag Fabs, introduced into cells by injection or through bead loading. Although the signal intensity emitted by an individual nascent protein is not different than for a mature protein, the signal at translating mRNAs is amplified as translation in polysomes results in multiple nascent peptides at a single mRNA, all of which containing epitope sequences. This results in signal intensities at translating mRNAs that are integer multiples compared to the signal of a single proteins and therefore allow to determine ribosome occupancy on individual mRNAs, similar to determining polymerase density and dynamics at a transcription site by determining nascent mRNA signal intensities and fluctuations (Morisaki et al., 2016; Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). Moreover, to further facilitate the detection of translation sites, some studies have inserted degradation tags to their reporter proteins, resulting in low background except from nascent peptides (Wu et al., 2016).

The ability to monitor translation in real time and at the single mRNA level revealed important features of translation regulation, with some of them being analogous to observations first made when imaging transcription. Monitoring signal intensities of nascent peptides at individual

translating mRNAs revealed that translation, similar to transcription, occurs in bursts, with mRNAs alternating between active and inactive states of translation (Morisaki et al., 2016; Pichon et al., 2016; Wu et al., 2016; Yan et al., 2016). Moreover, similar to RNA polymerase, ribosome stalling was observed at a fraction of mRNAs, even for codon optimized transcripts and introduction of previously suggested stalling sequences further increased this fraction (Yan et al., 2016). mRNAs in polysomes were also found to have slower diffusion coefficients in comparison to ones that are not translating, as previously observed (Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). Monitoring fluctuations in ribosome occupancy also allowed to calculate initiation and elongation rates at individual mRNAs. This showed that translation rates vary significantly between mRNAs with different 5' untranslated regions, but also between identical mRNAs, with initiation rates ranging from every 13 seconds to every 45 seconds being observed. Elongation rates also varied significantly between 3-18 amino acids per second. Using these values also allowed to determine the average spacing between ribosomes, showing number that vary significantly across the different studies due to the different in reporter mRNAs and data analysis, suggesting spacing between 160 - 910 nucleotides between individual ribosomes (Pichon et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016)..

Expanding the toolbox for translation imaging with the development of additional epitope-scFV combinations or SM-tags has further widened the scope of their usage for imaging of translation dynamics (Boersma et al., 2018; Zhao et al., 2018). Using two different epitope tags translated in different reading frames (SunTag and MoonTag), Boersma and co-workers revealed heterogeneity in start site selection that varied for different genes as well as for a specific mRNAs during different stages its life cycle (Boersma et al., 2018). Furthermore, although no evidence for frame-shifting was observed in human mRNAs, viral RNAs seem to exhibit frame-shifting allowing for synthesis of multiple proteins from the same RNA template (Boersma et al., 2018; Lyon et al., 2018).

3.3. Spatial organization of translating mRNAs

Translation is regulated by a set of proteins that help recruit the 43S pre-initiation complex to the 5' end of the mRNA. The cap binding protein eIF4E, the scaffold protein eIF4G and the DEAD box helicase eIF4A, together forming the eIF4F complex, have been shown to have a critical role in regulating translation initiation. Moreover, eIF4G interacts with the poly(A) binding protein PABC1, and this interaction was shown to stimulate translation of mRNAs *in vitro* (Imataka et al.,

1998; Tarun and Sachs, 1996; Tarun et al., 1997; Wakiyama et al., 2000). The interactions between all these components can be reconstituted *in vitro* using *in vitro* transcribed mRNAs and purified proteins, resulting in a closed-loop configuration of the mRNA mediated by PABC1 and eIF4F that can be visualized using electron microscopy (Wells et al., 1998). Together with the biochemical evidence, these observations have resulted in a model that suggests that translating mRNAs are present in cells in a closed loop configuration. This model is also supported, at least in part, by early electron microscopy studies in cells that showed polysome conformations resembling a closed loop state for ER associated polysomes (Christensen and Bourne, 1999; Christensen et al., 1987). However, in addition to circular-like conformations, various EM studies have found polysomes in many different configurations, including spiral, G-spiral and hairpin shapes, questioning whether all translating mRNAs exist in such a closed-loop configuration (Christensen and Bourne, 1999; Christensen et al., 1987). Furthermore, a recent cryo-electron tomography study in human glioblastoma cells found the majority of polysomes had a helical organization indicating an open conformation with the ends separate (Brandt et al., 2010). Interestingly, polysome conformations observed in this study were very similar to conformations that had been observed previously in bacteria, suggesting that there could be an evolutionary pressure on the organization of polysomes (Brandt et al., 2009). However, one limitation of using polysome imaging to understand mRNA organization is that the RNA is not visible in these images. Ribosome densities determined using different assays suggest a spacing between ribosomes in the order of several hundreds of nucleotides, making it is difficult to ascertain where most of the mRNA is located within polysomes. Furthermore, mRNAs with long 3' UTRs will have long regions that will not be occupied by ribosomes. In an attempt to obtain a more mRNA centric view of mRNA organization during translation, a recent study combined smFISH and super-resolution microscopy to determine the spatial relationship of different regions within mRNAs in human cell lines (Adivarahan et al., 2018). A similar attempt made by another study using smFISH with widefield microscopy (Khong and Parker, 2018). These studies did not observe closed-loop conformations for the mRNAs studied, but rather suggested that translation results in a decompaction of mRNAs and separation of the ends. However, it is still unclear whether the interaction between eIF4F and PABC1 could reflect a transient state during translation initiation, during when mRNAs become compact, something that could happen frequently during the lifetime of a cytoplasmic mRNA as a result of translation bursting. Alternatively, it is also possible that

closed-loop translation happens for specific classes of mRNAs. New assays allowing to study the dynamics of mRNA conformation and compaction in living cells will be required to test these models (Vicens et al., 2018).

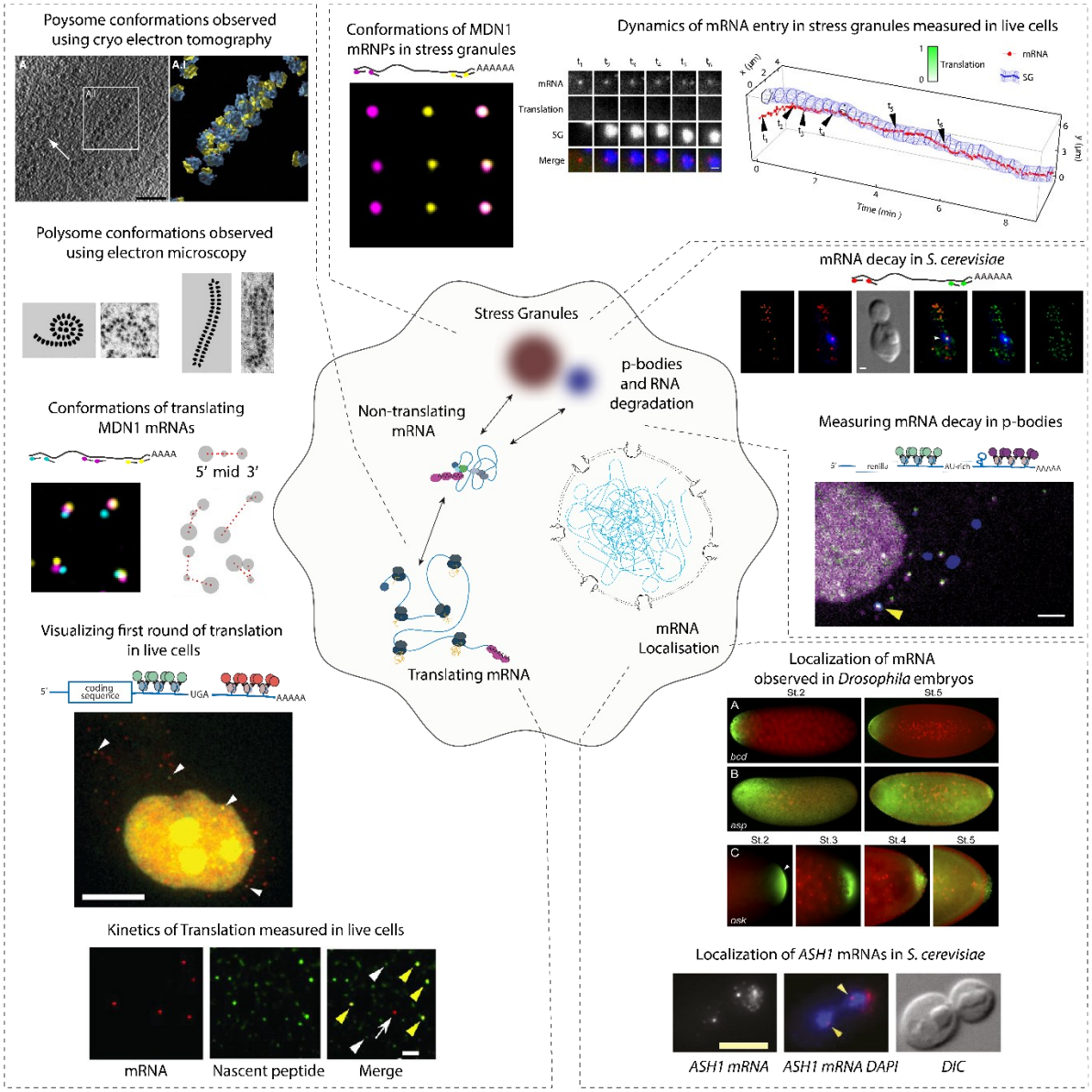


Figure 4: Visualizing different steps of cytoplasmic mRNA metabolism. Organization of mRNAs during translation and translation kinetics. *(Top to bottom) Polysome conformations in human glioblastoma cells visualized by cryo-electron tomography. Isosurface model shown on the right representing a helical polysome conformation. Modified with permission from Brandt et al., 2010.*

Sample imagines of spiral and hairpin configurations of endoplasmic reticulum localized polysomes in cultured fibroblasts visualized by electron microscopy. Modified with permission from Christensen and Bourne, 1999. Open conformation of cytoplasmic MDN1 mRNAs in HEK293 cells visualized using smFISH and their representations. Modified from Adivarahan et al., 2018. TRICK assay to visualize the first round of translation. Red signals correspond to MS2 signal, green signal to PP7 signal. Overlapping red and green signals, visualized as yellow, represent mRNAs that yet to undergo first round of translation. Modified with permission from Halstead et al., 2015. Measuring translation kinetics using the SunTag labeling system. The red signals correspond to the MS2 tagged mRNAs, green signals nascent peptides. Modified with permission from Wu et al., 2016. ***Stress granules (SG)*** *(from left to right) - Conformations of MDN1 mRNAs in stress granules as visualized using smFISH and super resolution microscopy. Cartoon showing regions of probe hybridization. Modified from Adivarahan et al., 2018. Dynamics of mRNA entry into SGs observed for MS2 tagged mRNAs in living cells. The plot on the right shows the path of an mRNA from initial entry to stable association with SG. Modified with permission from Moon et al., 2019.* ***RNA degradation*** *(top to bottom) mRNA decay kinetics in S. cerevisiae measured using smFISH. The red signal corresponds to probes hybridizing to the 5', green probes to the 3' end. Modified with permission from Trcek et al., 2011. Measuring mRNA decay using the TREAT reporter. The green signal corresponds to the PP7 signals, magenta the MS2 signals and blue represents Dcp1a, a marker for P-bodies. Modified with permission from Horvathova et al., 2017.* ***RNA localization*** *(top to bottom) Visualizing mRNA localization in Drosophila at different stages of development for transcripts (from top to bottom) bcd, asp and osk mRNAs (mRNAs in green, nucleus in red). Modified with permission from Lécuyer et al., 2007. Localization of ASH1 mRNA observed in S. cerevisiae using FISH. Nuclei were stained with DAPI. Arrowheads indicate transcription sites. From Powrie et al., 2001.*

3.4. mRNA conformation of translationally inhibited RNAs/stress granules

Translation is a highly energy consuming process and producing new proteins is the main requirement for cells to grow and divide. Upon cellular stress, cells need to conserve energy to ensure their survival, a consequence of which is that translation of most mRNAs is inhibited with only selected mRNAs being translated efficiently. Moreover, various stresses induce the formation of phase-separated compartments termed stress-granules (SG) made up of various mRNA-binding

proteins including components from the preinitiation complex as well as poly(A) mRNAs (Decker and Parker, 2012). Initially suggested as static structures of mRNA storage sequestering translationally inactive mRNAs, recent studies revealed that association of mRNA with SGs can be dynamic, with inhibition of translation alone not sufficient to target an mRNA to these structures. Guided by data from stress granule purifications and RNA sequencing experiments and monitoring the localization for a subset of mRNAs using smFISH, it was found that not all translationally inhibited mRNAs accumulate in stress granules, but that large mRNAs are preferentially found to localize to SG compared than shorter mRNAs (Khong et al., 2017). *In situ* hybridization and live-cell tracking using MS2 tagged mRNAs in combination with SunTag/SM-tag labelling showed that formation of stress granules in cells precedes the recruitment of mRNAs, with stable association of mRNAs with SG requiring run off of all ribosomes from the translating mRNA (Khong and Parker, 2018; Moon et al., 2019; Wilbertz et al., 2019). Interestingly, mRNAs in SGs were found to have a very compact conformation when compared to nuclear mRNAs or translating cytoplasmic mRNAs, and similar to the compact conformations obtained by mRNAs in cells when translation was inhibited pharmacologically (Khong and Parker, 2018; Adivarahan et al., 2018). Together these results suggest that compaction of mRNAs might be a pre-requisite for their recruitment to SGs. mRNA recruitment to SGs was also found to be influenced by cis-elements like the presence of a TOP motif, a sequence motif found within the 5' UTR of many highly translated mRNAs, as it was shown that TOP motif containing mRNAs more frequently showed a stable association of an mRNA with these membrane-less compartments (Halstead et al., 2015). In addition, single protein tracking of the SG proteins G3BP1 and IMP1 showed dynamic biphasic partition of these proteins within SGs, suggesting that SGs contain relatively immobile nanocores, however it is unclear whether static association of mRNAs within SGs requires localization to these regions (Niewidok et al., 2018).

Consistent with the suggested role of SGs as storage compartment during stress, mRNAs localizing to SGs were shown to be capable of resuming translation once the stress was dissolved, with their translation kinetics indistinguishable from non-SG localized cytoplasmic mRNAs (Wilbertz et al., 2019). However, much of the rules that define why some translationally inactive mRNAs accumulate in SGs upon stress whereas others do not still need to be determined.

3.5. Towards death of an mRNA

In the cytoplasm, the processes modulating translation are believed to be in direct competition with the mRNA decay pathway. For example the eIF4F complex that binds to the 5' cap and is responsible for translation initiation, is thought to compete with the decapping complex for access to the cap (Decker and Parker, 2012; Schwartz and Parker, 1999, 2000). This competition between mRNA translation and decay factors ultimately determines mRNA stability. The balance between the two processes can further be regulated through regulatory elements like miRNA-binding sites or AU-rich elements (Duchaine and Fabian, 2019; Grudzien-Nogalska and Kiledjian, 2017). *In situ* hybridization has also proven to be a useful tool for studying mRNA degradation in cells as targeting FISH probes to different regions of the mRNAs allows for the detection of degradation intermediates. Using such an approach, it was shown that decay of mRNAs in yeast could be regulated by the promoter sequence, regulating the *SWI5* and *CLB2* mRNAs decay in a cell-cycle dependent manner, modulated by the co-transcriptional assembly of Dbf2p (Trcek et al., 2011). Using a similar *in situ* hybridization approach, a study in Trypanosome was able to establish that Xrn1 mediated decay was the predominant mode of mRNA degradation (Kramer, 2016).

Many of the factors implicated in RNA degradation were found to accumulate in membrane-less organelles called processing bodies (P-bodies) (Decker and Parker, 2012). P-bodies exist in most cells under normal growth conditions, however their size and number are dependent on the pool of non-translating mRNAs and are greatly enhanced upon exposure to stress (Decker and Parker, 2012; TEIXEIRA et al., 2005). The presence of the degradation factors including the decapping factors Dcp1 and Dcp2, as well as many regulatory proteins of the RNA degradation pathway such as Ccr4 and GW182, led to the hypothesis that P-bodies act as degradation factories, where mRNAs are brought at the end of their lives to be degraded. To directly test this model, Horvathova and colleagues developed a reporter that allowed the visualization of degradation intermediates in living cells termed '*3(three)'-RNA end accumulation during turnover' or TREAT* (Horvathova et al., 2017). The reporter was designed such that it contained pseudo-knots (PKs) in between PP7 and MS2 stem loops (Figure 4). The PKs sequences, derived from insect-borne flaviviruses, are resistant to degradation by the 5-'3' exonuclease Xrn1, believed to be the dominant pathway for mRNA degradation in mammals, thereby protecting the downstream MS2 loops and resulting in stable 3' degradation intermediates. Using this system, it was therefore possible to distinguish between full length and Xrn1 degraded

mRNAs through monitoring the presence of both PP7 and MS2 signals. This study found that full length mRNAs, but no degradation intermediates localized to P-bodies, suggesting that degradation does not occur within P-bodies. Consistent with such a model, they also found that individual degradation events occurred within the cytosol rather than in P-bodies.

P-bodies are often found adjacent to SGs and early models suggested that mRNAs might transit from SGs to P-bodies. However, using MS2 tagged mRNAs, Wilbertz et al observed that although specific reporter mRNAs localize to both SGs and P-bodies during stress, exchange between the two compartments, when they are adjacent to each other, was an extremely rare event, though mRNAs could contact multiple SGs and P-bodies before stable association with either (Moon et al., 2019; Wilbertz et al., 2019). These new observations suggest that the two membrane-less compartments function independently in their roles of compartmentalizing specific activities during stress. However, whether these observations are true for all mRNAs and whether the proximity of SG and P-bodies play an active role in the regulation of RNA stability under different conditions for specific mRNAs still needs to be determined.

4. Outlook

mRNA imaging has contributed extensively to the current understanding of different aspects of RNA metabolism and has in recent years become an increasingly important tool to study quantitative aspects, as well as the dynamics, of the different processes along the gene expression pathway. The continuous development and refinement of RNA imaging approaches will further facilitate studying these processes in even more detail, allowing finding answers to new questions or look at old questions with a new set of eyes. Single-molecule imaging is likely to continue providing an ideal platform to study different processes that regulate mRNA metabolism, having the advantage over traditional methods in being able to provide high resolution spatial and temporal information.

One of the main limitations for cellular RNA imaging today is the low-throughput nature of the many of these approaches. While it is relatively straightforward to image hundreds or even thousands of cells using automated image acquisition and image analysis, most approaches still only allow to study one or few mRNAs at the time. However, recent developments in *in situ* hybridization approaches in fixed cells using sequential hybridization and barcoding have enabled for imaging of thousands of RNAs within the same cell (MERFISH, SeqFISH, seqFISH+) (Chen

et al., 2015; Eng et al., 2019; Lubeck et al., 2014; Shah et al., 2018). These approaches have the potential to become complementary, or even more powerful, than (single cell -) RNA sequencing methodologies, and might open doors for a microscopy centric transcriptome analysis that is not limited by expression levels, and the same time is able to provide high-resolution spatial information. Moreover, combining these with expansion or clearing protocols might further increase resolution and facilitate the use of such approaches in tissues and animals.

The spatial organization of mRNPs is one of the last unexplored topics in mRNA research that is likely to profit from future advances in imaging methods. Recent approaches combining smFISH and super-resolution microscopy have already revealed important new insights into mRNP organization in cells and showed that even longstanding dogmas, such as the closed-loop model for translation, have to be revisited (Pierron and Weil, 2018; Adivarahan et al., 2018; Vicens et al., 2018). Further adaptations of super-resolution approaches, including STED and dSTORM/PALM and yet to be developed methods will allow to delve deeper into the structural organization of RNA-protein complexes and its role in mRNA metabolism. Similarly, recent improvements in cryo-EM revolution, that have led to high resolution structures of many large protein/RNA-protein complexes, can provide an interesting avenue towards exploring the topic of mRNP organization. However, purifying specific mRNPs from heterogenous mRNP populations in sufficient quantities and homogeneity for cryo-EM analysis will likely be challenging. In addition, correlative imaging in-cells by combining cryo-EM with fluorescence microscopy will allow to combine the strength of specific labelling of fluorescent approaches with the resolution of election microscopy and could be a powerful approach to study mRNPs in cells.

Lastly, further expanding the tools to image mRNPs in living cells will be essential for moving towards the ability to follow mRNP metabolism through its different stages, and will require tools that allow imaging single mRNAs, as well as its associated proteins, through time and space. Current methods only allow this for very short time periods, and/or in limited sub regions of the cells. Advances in labelling, illumination and image acquisition will be required to move towards this goal. Further improvement of aptamer-based RNA visualization approaches that make use of bright, photostable and membrane permeable dyes such as the like Janelia Fluor dyes are already helping to overcome some of these drawbacks (Grimm et al., 2015, 2016, 2017). Similarly, the use of proteins labelling systems such as Halo-tags, CLIP-tags and SNAP tags that allow single protein imaging will allow to better understand the dynamics and stoichiometry of RBPs on mRNA and

how this regulates RNA metabolism (Grimm et al., 2015, 2016, 2017; Keppler et al., 2002; Los et al., 2008). However, achieving all this will also require further improvements in the fluorophores for live cell imaging in terms of photon emission, photostability, cell permeability and labelling efficiencies to their respective tags as well as combining them with less-phototoxic imaging methodologies like light-sheet microscopy, or the development of entirely new tools to follow biomolecules in cells.

Acknowledgement

We thank members of the Zenklusen laboratory for discussion and comments on the manuscript. This work has been supported Canadian Institutes of Health Research (Project Grant-366682), Le Fonds de recherche du Québec – Santé (Chercheur-boursier Junior 2) and Natural Sciences and Engineering Research Council of Canada for DZ.

References

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B., Kellner, M.J., Regev, A., et al. (2017). RNA targeting with CRISPR–Cas13. Nature 550, 280.

Adivarahan, Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol Cell 72, 727-738.e5.

Ainger, K., Avossa, D., Diana, A.S., Barry, C., Barbarese, E., and Carson, J.H. (1997). Transport and Localization Elements in Myelin Basic Protein mRNA. J Cell Biology 138, 1077–1087.

Akam, M.E. (1983). The location of Ultrabithorax transcripts in Drosophila tissue sections. Embo J 2, 2075–2084.

Alcázar-Román, A.R., Tran, E.J., Guo, S., and Wente, S.R. (2006). Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. Nat Cell Biol 8, 711–716.

Alpert, T., Herzel, L., and Neugebauer, K.M. (2017). Perfect timing: splicing and transcription rates in living cells. Wiley Interdiscip Rev Rna 8.

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. Cell Reports 13, 2653–2662.

Bao, G., Rhee, W., and Tsourkas, A. (2009). Fluorescent Probes for Live-Cell RNA Detection. Annu Rev Biomed Eng 11, 25–47.

Bartman, C.R., Hsu, S.C., Hsiung, C., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol Cell 62, 237–247.

Bartman, C.R., Hamagami, N., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A., and Raj, A. (2019). Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol Cell 73, 519-532.e4.

Bashirullah, A., Cooperstock, R.L., and Lipshitz, H.D. (2001). Spatial and temporal control of RNA stability. Proc National Acad Sci 98, 7025–7028.

Batisse, J., Batisse, C., Budd, A., Böttcher, B., and Hurt, E. (2009). Purification of Nuclear Poly(A)-binding Protein Nab2 Reveals Association with the Yeast Transcriptome and a Messenger Ribonucleoprotein Core Structure. J Biol Chem 284, 34911–34917.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. Cell 163, 1596–1610.

Becalska, A.N., and Gavis, E.R. (2009). Lighting up mRNA localization in Drosophila oogenesis. Development 136, 2493–2503.

Ben-Ari, Y., Brody, Y., Kinor, N., Mor, A., Tsukamoto, T., Spector, D.L., Singer, R.H., and Shav-Tal, Y. (2010). The life of an mRNA in space and time. J Cell Sci 123, 1761–1774.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15, nrg3662.

Berget, S., Moore, C., and Sharp, P. (1977). Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. Proc National Acad Sci 74, 3171–3175.

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

Beyer, A., and Osheim, Y. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Gene Dev 2, 754–765.

Björk, P., and Wieslander, L. (2015). The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA–Protein Complexes. Annu Rev Biochem 84, 65–92.

Björk, P., Wetterberg-Strandh, I., Baurén, G., and Wieslander, L. (2006). Chironomus tentans-Repressor Splicing Factor Represses SR Protein Function Locally on Pre-mRNA Exons and Is Displaced at Correct Splice Sites. Mol Biol Cell 17, 32–42.

Björk, P., Jin, S., Zhao, J., Singh, O., Persson, J.-O., Hellman, U., and Wieslander, L. (2009). Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions. J Cell Biology 184, 555–568.

Blazquez, L., Emmett, W., Faraway, R., Pineda, J., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., et al. (2018). Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. Mol Cell 72, 496-509.e9.

Blobel, G. (1985). Gene gating: a hypothesis. Proc National Acad Sci 82, 8527–8529.

Boehm, V., Britto-Borges, T., Steckelberg, A.-L., Singh, K.K., Gerbracht, J.V., Gueney, E., Blazquez, L., Altmüller, J., Dieterich, C., and Gehring, N.H. (2018). Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. Mol Cell 72, 482-495.e7.

Boersma, S., Khuperkar, D., Verhagen, B.M.P., Sonneveld, S., Grimm, J.B., Lavis, L.D., and Tanenbaum, M.E. (2018). Multi-color single molecule imaging uncovers extensive heterogeneity in mRNA decoding. bioRxiv, p.477661.

Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Bäcker, V., Kornblihtt, A., Marcello, A., and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biology 179, 291–304.

Bozzola, J.J., and Russell, L. (1999). Electron microscopy : principles and techniques for biologists.

Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, U.F., and Baumeister, W. (2009). The Native 3D Organization of Bacterial Polysomes. Cell 136, 261–271.

Brandt, F., Carlson, L.-A., Hartl, U.F., Baumeister, W., and Grünewald, K. (2010). The Three-Dimensional Organization of Polyribosomes in Intact Human Cells. Mol Cell 39, 560–569.

Brodsky, A., and Silver, P. (2000). Pre-mRNA processing factors are required for nuclear export. Rna New York N Y 6, 1737–1749.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.-M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing. Plos Biol 9, e1000573.

Buchwalter, A., Kaneshiro, J.M., and Hetzer, M.W. (2018). Coaching from the sidelines: the nuclear periphery in genome regulation. Nat Rev Genet 20, 1.

Buxbaum, A.R., Haimovich, G., and Singer, R.H. (2014). In the right place at the right time: visualizing and understanding mRNA localization. Nat Rev Mol Cell Bio 16, nrm3918.

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

Calapez, A., Pereira, H.M., Calado, A., Braga, J., Rino, J., Carvalho, C., Tavanez, J., Wahle, E., Rosa, A.C., and Carmo-Fonseca, M. (2002). The intranuclear mobility of messenger RNA binding proteins is ATP dependent and temperature sensitive. J Cell Biology 159, 795–805.

Chen, H., and Larson, D.R. (2016). What have single-molecule studies taught us about gene expression? Gene Dev 30, 1796–1810.

Chen, K., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090.

Cherry, R.J., Smith, P.R., Morrison, I., and Fernandez, N. (1998). Mobility of cell surface receptors: a re-evaluation. Febs Lett 430, 88–91.

Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell 11, 819–836.

Christensen, K.A., and Bourne, C.M. (1999). Shape of large bound polysomes in cultured fibroblasts and thyroid epithelial cells. Anatomical Rec 255, 116–129.

Christensen, K.A., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat 178, 1–10.

Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional Pulsing of a Developmental Gene. Curr Biol 16, 1018–1025.

Churchman, S.L., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368.

Codeluppi, S., Borm, L.E., Zeisel, A., Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat Methods 15, 932–935.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and Disease. Cell 136, 777–793.

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Rev Genet 14, 572–584.

Coulon, A., Ferguson, M.L., de Turris, V., Palangat, M., Chow, C.C., and Larson, D.R. (2014). Kinetic competition during the transcription cycle results in stochastic RNA processing. Elife 3, e03939.

Daigle, N., and Ellenberg, J. (2007). λN-GFP: an RNA reporter system for live-cell imaging. Nat Methods 4, nmeth1065.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., enoy, S., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 14, 796–806.

Decker, C.J., and Parker, R. (2012). P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation. Csh Perspect Biol 4, a012286.

Deshler, J.O., Highett, M.I., and Schnapp, B.J. (1997). Localization of Xenopus Vg1 mRNA by Vera Protein and the Endoplasmic Reticulum. Science 276, 1128–1131.

Dolgosheina, E.V., and Unrau, P.J. (2016). Fluorophore-binding RNA aptamers and their applications. Wiley Interdiscip Rev Rna 7, 843–851.

Dolgosheina, E.V., Jeng, S.C., Panchapakesan, S.S., Cojocaru, R., Chen, P.S., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014). RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. Acs Chem Biol 9, 2412–2420.

Duchaine, T.F., and Fabian, M.R. (2019). Mechanistic Insights into MicroRNA-Mediated Gene Silencing. Csh Perspect Biol 11, a032771.

Elbarbary, R.A., Li, W., Tian, B., and Maquat, L.E. (2013). STAU1 binding 3′ UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. Gene Dev 27, 1495–1510.

Eng, C.-H., Shah, S., Thomassie, J., and Cai, L. (2017). Profiling the transcriptome with RNA SPOTs. Nat Methods 14, 1153–1155.

Eng, C.-H., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature 1–5.

Farina, K.L., Hüttelmaier, S., Musunuru, K., Darnell, R., and Singer, R.H. (2003). Two ZBP1 KH domains facilitate β-actin mRNA localization, granule formation, and cytoskeletal attachment. J Cell Biology 160, 77–87.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of Single RNA Transcripts in Situ. Science 280, 585–590.

Galganski, L., Urbanek, M.O., and Krzyzosiak, W.J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. Nucleic Acids Res gkx759-.

Gavis, E.R., and Lehmann, R. (1992). Localization of nanos RNA controls embryonic polarity. Cell 71, 301–313.

Gavis, E.R., Curtis, D., and Lehmann, R. (1996). Identification ofcis-Acting Sequences That ControlnanosRNA Localization. Dev Biol 176, 36–50.

Golding, I., and Cox, E.C. (2004). RNA dynamics in live Escherichia coli cells. P Natl Acad Sci Usa 101, 11310–11315.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-Time Kinetics of Gene Activity in Individual Bacteria. Cell 123, 1025–1036.

Gonsalvez, G.B., Urbinati, C.R., and Long, R.M. (2005). RNA localization in yeast: moving towards a mechanism. Biol Cell 97, 75–86.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Mol Cell 19, 53–63.

Grimm, J.B., English, B.P., Chen, J., Slaughter, J.P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J.J., Normanno, D., Singer, R.H., et al. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. Nat Methods 12, nmeth.3256.

Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P., Brown, T.A., Lippincott-Schwartz, J., Liu, Z., Lionnet, T., et al. (2016). Bright photoactivatable fluorophores for single-molecule imaging. Nat Methods 13, nmeth.4034.

Grimm, J.B., Muthusamy, A.K., Liang, Y., Brown, T.A., Lemon, W.C., Patel, R., Lu, R., Macklin, J.J., Keller, P.J., Ji, N., et al. (2017). A general method to fine-tune fluorophores for live-cell and in vivo imaging. Nat Methods 14, nmeth.4403.

Grudzien-Nogalska, E., and Kiledjian, M. (2017). New insights into decapping enzymes and selective mRNA decay. Wiley Interdiscip Rev Rna 8.

Grünwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature 467, 604.

Guzikowski, A.R., Chen, Y.S., and Zid, B.M. (2019). Stress-induced mRNP granules: Form and function of processing bodies and stress granules. Wiley Interdiscip Rev Rna e1524.

Hager, G.L., McNally, J.G., and Misteli, T. (2009). Transcription Dynamics. Mol Cell 35, 741–753.

Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty Gene Expression in the Intact Mammalian Liver. Mol Cell 58, 147–156.

Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). An RNA biosensor for imaging the first round of translation from single cells to living animals. Science 347, 1367–1671.

Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A., and Singer, R.H. (2013). Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nat Methods 10, 119.

Horvathova, I., Voigt, F., Kotrys, A.V., Zhan, Y., Artus-Revel, C.G., Eglinger, J., Stadler, M.B., Giorgetti, L., and Chao, J.A. (2017). The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. Mol Cell 68, 615-625.e9.

Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J., and Moore, M.J. (2011). Ordered and Dynamic Assembly of Single Spliceosomes. Science 331, 1289–1295.

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

Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. Embo J 17, 7480–7489.

Jakt, L., Moriwaki, S., and Nishikawa, S. (2013). A continuum of transcriptional identities visualized by combinatorial fluorescent in situ hybridization. Development 140, 216–225.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. Nat Methods 5, nmeth.1237.

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

Johnston, D., Beuchle, D., and Nüsslein-Volhard, C. (1991). staufen, a gene required to localize maternal RNAs in the Drosophila egg. Cell 66, 51–63.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife 3, e02407.

Jung, H., Gkogkas, C.G., Sonenberg, N., and Holt, C.E. (2014). Remote Control of Gene Function by Local Translation. Cell 157, 26–40.

Katz, Z.B., English, B.P., Lionnet, T., Yoon, Y.J., Monnier, N., Ovryn, B., Bathe, M., and Singer, R.H. (2016). Mapping translation “hot-spots” in live cells by tracking single molecules of mRNA and ribosomes. Elife 5, e10415.

Kellems, R.E., Allison, V.F., and Butow, R.A. (1975). Cytoplasmic type 80S ribosomes associated with yeast mitochondria. IV. Attachment of ribosomes to the outer membrane of isolated mitochondria. The Journal of Cell Biology 65.

Kempe, H., Schwabe, A., Crémazy, F., Verschure, P.J., and Bruggeman, F.J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol Biol Cell 26, 797–804.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2002). A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat Biotechnol 21, 86–89.

Khong, A., and Parker, R. (2018). mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. J Cell Biol 217, jcb.201806183.

Khong, A., Matheny, T., Jain, S., Mitchell, S.F., Wheeler, J.R., and Parker, R. (2017). The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. Mol Cell 68, 808-820.e5.

Kim-Ha, J., Webster, P., Smith, J., and Macdonald, P. (1993). Multiple RNA regulatory elements mediate distinct steps in localization of oskar mRNA. Dev Camb Engl 119, 169–178.

Kislauskis, E., Zhu, X., and Singer, R. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J Cell Biology 127, 441–451.

Kotovic, K.M., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional Recruitment of the U1 snRNP to Intron-Containing Genes in Yeast. Mol Cell Biol 23, 5768–5779.

Kramer, S. (2016). Simultaneous detection of mRNA transcription and decay intermediates by dual colour single mRNA FISH on subcellular resolution. Nucleic Acids Res 45, gkw1245.

Kühlbrandt, W. (2014). Cryo-EM enters a new era. Elife 3, e03678.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. Science 339, 950–953.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3′ End Processing and Splicing. Mol Cell 23, 195–205.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. Science 332, 475–478.

Larsson, C., Grundberg, I., Söderberg, O., and Nilsson, M. (2010). In situ detection and genotyping of individual mRNA molecules. Nat Methods 7, 395.

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

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. Cell 131, 174–187.

Lee;, B., Bae, S.-W., Shim, J., Sung, and Park, H. (2016). Imaging Single-mRNA Localization and Translation in Live Neurons. Mol Cells.

Lenstra, T.L., Rodriguez, J., Chen, H., and Larson, D.R. (2016). Transcription Dynamics in Living Cells. Annu Rev Biophys 45, 1–23.

Levsky, J.M., Shenoy, S.M., Pezo, R.C., and Singer, R.H. (2002). Single-Cell Gene Expression Profiling. Science 297, 836–840.

Lin, C., and Chang, J. (1975). Electron microscopy of albumin synthesis. Science 190, 465–467.

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

Long, R.M., Singer, R.H., Meng, X., Gonzalez, I., Nasmyth, K., and Jansen, R.-P. (1997). Mating Type Switching in Yeast Controlled by Asymmetric Localization of ASH1 mRNA. Science 277, 383–387.

Los, G.V., Encell, L.P., ugall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R., Urh, M., et al. (2008). HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. Acs Chem Biol 3, 373–382.

Lubeck, E., and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat Methods 9, 743.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat Methods 11, nmeth.2892.

Lyon, K.R., Aguilera, L.U., Morisaki, T., Munsky, B., and Stasevich, T.J. (2018). Live-cell single RNA imaging reveals bursts of translational frameshifting. bioRxiv, p.478040.

Macdonald, P.M., and Struhl, G. (1988). Cis- acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. Nature 336, 336595a0.

Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L., Förster, F., Hyman, A.A., Plitzko, J.M., and Baumeister, W. (2016). Visualizing the molecular sociology at the HeLa cell nuclear periphery. Science 351, 969–972.

Maiuri, P., Knezevich, A., Marco, A., Mazza, D., Kula, A., McNally, J.G., and Marcello, A. (2011). Fast transcription rates of RNA polymerase II in human cells. Embo Rep 12, 1280–1285.

Martin, K.C., and Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial Dimension. Cell 136, 719–730.

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-Cell Visualization of Pre-mRNA Splicing with Single-Molecule Sensitivity. Cell Reports 4, 1144–1155.

Martins, S., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J., de Almeida, S., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3′ end of human genes. Nat Struct Mol Biol 18, 1115.

Mehlin, H., Daneholt, B., and Skoglund, U. (1992). Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. Cell 69, 605–613.

Melton, D. (1987). Translocation of a localized maternal mRNA to the vegetal pole of Xenopus oocytes. Nature 328, 328080a0.

Melton, D., and Yisraeli, J.K. (1988). The maternal mRNA Vg1 is correctly localized following injection into Xenopus oocytes. Nature 336, 592.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. Mol Cell 72.

Miralles, F., Öfverstedt, L.-G., Sabri, N., Aissouni, Y., Hellman, U., Skoglund, U., and Visa, N. (2000). Electron Tomography Reveals Posttranscriptional Binding of Pre-Mrnps to Specific Fibers in the Nucleoplasm. J Cell Biology 148, 271–282.

Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and Spector, D.L. (1998). Serine Phosphorylation of SR Proteins Is Required for Their Recruitment to Sites of Transcription In Vivo. J Cell Biology 143, 297–307.

Molenaar, C., Marras, S., Slats, J., Truffert, J.-C., Lemaître, M., Raap, A., Dirks, R., and Tanke, H. (2001). Linear 2′ O-Methyl RNA probes for the visualization of RNA in living cells. Nucleic Acids Res 29, e89–e89.

Molenaar, C., Abdulle, A., Gena, A., Tanke, H.J., and Dirks, R.W. (2004). Poly(A)+ RNAs roam the cell nucleus and pass through speckle domains in transcriptionally active and inactive cells. J Cell Biology 165, 191–202.

Moon, S.L., Morisaki, T., Khong, A., Lyon, K., Parker, R., and Stasevich, T.J. (2019). Multicolour single-molecule tracking of mRNA interactions with RNP granules. Nat Cell Biol 21, 162–168.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12, 543.

Morisaki, T., Lyon, K., DeLuca, K.F., DeLuca, J.G., English, B.P., Zhang, Z., Lavis, L.D., Grimm, J.B., Viswanathan, S., Looger, L.L., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. Science 352, 1425–1429.

Mowry, K., and Melton, D. (1992). Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in Xenopus oocytes. Science 255, 991–994.

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. Nat Methods 10, 277.

Muramoto, T., Cannon, D., Gierliński, M., Corrigan, A., Barton, G.J., and Chubb, J.R. (2012). Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. Proc National Acad Sci 109, 7350–7355.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell 165, 488–496.

Nicolas, D., Zoller, B., Suter, D.M., and Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. Proc National Acad Sci 115, 201722330.

Niewidok, B., Igaev, M., da Graca, A., Strassner, A., Lenzen, C., Richter, C.P., Piehler, J., Kurre, R., and Brandt, R. (2018). Single-molecule imaging reveals dynamic biphasic partition of RNA-binding proteins in stress granules. J Cell Biol 217, jcb.201709007.

Niwa, M., and Berget, S. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Gene Dev 5, 2086–2095.

Ochiai, H., Sugawara, T., Sakuma, T., and Yamamoto, T. (2014). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci Rep-Uk 4, 7125.

Oeffinger, M., and Zenklusen, D. (2012). To the pore and through the pore: A story of mRNA export kinetics. Biochimica Et Biophysica Acta Bba - Gene Regul Mech 1819, 494–506.

Osheim, Y.N., Miller, J., and Beyer, A.L. (1985). RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell 43, 143–151.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, S.L., Singh, A., and Raj, A. (2015). Single Mammalian Cells Compensate for

Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. Mol Cell 58, 339–352.

Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011). RNA Mimics of Green Fluorescent Protein. Science 333, 642–646.

Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009). Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting. Curr Biol 19, 2037–2042.

Pichon, X., Bastide, A., Safieddine, A., Chouaib, R., Samacoits, A., Basyuk, E., Peter, M., Mueller, F., and Bertrand, E. (2016). Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. J Cell Biology 214, 769–781.

Pierron, G., and Weil, D. (2018). Re-viewing the 3D Organization of mRNPs. Mol Cell 72, 603–605.

Politz, J.C., Browne, E.S., Wolf, D.E., and Pederson, T. (1998). Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. Proc National Acad Sci 95, 6043–6048.

Politz, J.C., Tuft, R.A., Pederson, T., and Singer, R.H. (1999). Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. Curr Biol 9, 285–291.

Powrie, E.A., Zenklusen, D., and Singer, R.H. (2011). A nucleoporin, Nup60p, affects the nuclear and cytoplasmic localization of ASH1 mRNA in S. cerevisiae. Rna 17, 134–144.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, F.C., Zhang, M.Q., and Spector, D.L. (2005). Regulating Gene Expression through RNA Nuclear Retention. Cell 123, 249–263.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell 135, 216–226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. Plos Biol 4, e309.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5, nmeth.1253.

Rodriguez, A.J., Shenoy, S.M., Singer, R.H., and Condeelis, J. (2006). Visualization of mRNA translation in living cells. J Cell Biology 175, 67–76.

Rouhanifard, S.H., Mellis, I.A., Dunagin, M., Bayatpour, S., Jiang, C.L., Dardani, I., Symmons, O., Emert, B., Torre, E., Cote, A., et al. (2019). ClampFISH detects individual nucleic acid molecules using click chemistry–based amplification. Nat Biotechnol 37, 102.

Santangelo, P.J., Nix, B., Tsourkas, A., and Bao, G. (2004). Dual FRET molecular beacons for mRNA detection in living cells. Nucleic Acids Res 32, e57–e57.

Saroufim, M.-A., Bensidoun, P., Raymond, P., Rahman, S., Krause, M.R., Oeffinger, M., and Zenklusen, D. (2015). The nuclear basket mediates perinuclear mRNA scanning in budding yeast. J Cell Biology 211, 1131–1140.

Schmidt, U., Basyuk, E., Robert, M.-C., Yoshida, M., Villemin, J.-P., Auboeuf, D., Aitken, S., and Bertrand, E. (2011). Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. J Cell Biology 193, 819–829.

Schwartz, D.C., and Parker, R. (1999). Mutations in Translation Initiation Factors Lead to Increased Rates of Deadenylation and Decapping of mRNAs inSaccharomyces cerevisiae. Mol Cell Biol 19, 5247–5256.

Schwartz, D.C., and Parker, R. (2000). mRNA Decapping in Yeast Requires Dissociation of the Cap Binding Protein, Eukaryotic Translation Initiation Factor 4E. Mol Cell Biol 20, 7933–7942.

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., and Darzacq, X. (2014). Transcription Factors Modulate c-Fos Transcriptional Bursts. Cell Reports 8, 75–83.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H., Koulena, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell 174, 363-376.e16.

Shav-Tal, Y., Darzacq, X., Shenoy, S.M., Fusco, D., Janicki, S.M., Spector, D.L., and Singer, R.H. (2004). Dynamics of Single mRNPs in Nuclei of Living Cells. Science 304, 1797–1800.

Siebrasse, J., Veith, R., Dobay, A., Leonhardt, H., Daneholt, B., and Kubitscheck, U. (2008). Discontinuous movement of mRNP particles in nucleoplasmic regions devoid of chromatin. Proc National Acad Sci 105, 20291–20296.

Siebrasse, J., Kaminski, T., and Kubitscheck, U. (2012). Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. Proc National Acad Sci 109, 9426–9431.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem 84, 1–30.

Singh, O., Björkroth, B., Masich, S., Wieslander, L., and Daneholt, B. (1999). The Intranuclear Movement of Balbiani Ring Premessenger Ribonucleoprotein Particles. Exp Cell Res 251, 135–146.

Sinnamon, J.R., and Czaplinski, K. (2014). RNA detection in situ with FISH-STICs. Rna 20, 260–266.

Skoglund, U., Andersson, K., Strandberg, B., and Daneholt, B. (1986). Three-dimensional structure of a specific pre-messenger RNP particle established by electron microscope tomography. Nature 319, 319560a0.

Smith, C., Lari, A., Derrer, C., Ouwehand, A., Rossouw, A., Huisman, M., Dange, T., Hopman, M., Joseph, A., Zenklusen, D., et al. (2015). In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. J Cell Biology 211, 1121–1130.

Spille, J.-H., Hecht, M., Grube, V., Cho, W., Lee, C., and Cissé, I.I. (2019). A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells. Methods 153, 35–45.

Staněk, D., and Fox, A. (2017). Nuclear bodies: news insights into structure and function. Curr Opin Cell Biol 46, 94–101.

Steward, O., and Levy, W. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. J Neurosci 2, 284–291.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science 332, 472–474.

Takizawa, P.A., Sil, A., Swedlow, J.R., Herskowitz, I., and Vale, R.D. (1997). Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. Nature 389, 90–93.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell 159, 635–646.

Tarun, S., and Sachs, A. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. Embo J 15, 7168–7177.

Tarun, S.Z., Wells, S.E., Deardorff, J.A., and Sachs, A.B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc National Acad Sci 94, 9046–9051.

TEIXEIRA, D., SHETH, U., VALENCIA-SANCHEZ, M.A., BRENGUES, M., and PARKER, R. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. Rna 11, 371–382.

Thompson, R.E., Larson, D.R., and Webb, W.W. (2002). Precise Nanometer Localization Analysis for Individual Fluorescent Probes. Biophys J 82, 2775–2783.

Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: An open and extensible platform for single-particle tracking. Methods 115, 80–90.

Trcek, T., Larson, D.R., Moldón, A., Query, C.C., and Singer, R.H. (2011). Single-Molecule mRNA Decay Measurements Reveal Promoter- Regulated mRNA Stability in Yeast. Cell 147, 1484–1497.

Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.-M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., et al. (2016). smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res 44, e165–e165.

Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2017). An improved MS2 system for accurate reporting of the mRNA life cycle. Nat Methods 15, 81.

Tyagi, S., and Kramer, F. (1996). Molecular Beacons: Probes that Fluoresce upon Hybridization. Nat Biotechnol 14, 303.

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

Vargas, D.Y., Raj, A., Marras, S.A., Kramer, F., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. P Natl Acad Sci Usa 102, 17008–17013.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S., Schedl, P., and Tyagi, S. (2011). Single-Molecule Imaging of Transcriptionally Coupled and Uncoupled Splicing. Cell 147, 1054–1065.

Vicens, Q., Kieft, J.S., and Rissland, O.S. (2018). Revisiting the Closed-Loop Model and the Nature of mRNA 5′–3′ Communication. Mol Cell 72, 805–812.

Viswanathan, S., Williams, M.E., Bloss, E.B., Stasevich, T.J., eer, C., Nern, A., Pfeiffer, B.D., Hooks, B.M., Li, W.-P., English, B.P., et al. (2015). High-performance probes for light and electron microscopy. Nat Methods 12, 568–576.

Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A., et al. (2009). A wave of nascent transcription on activated human genes. Proc National Acad Sci 106, 18357–18361.

Wakiyama, M., Imataka, H., and Sonenberg, N. (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. Curr Biol 10, 1147–1150.

Waks, Z., Klein, A.M., and Silver, P.A. (2011). Cell-to-cell variability of alternative RNA splicing. Mol Syst Biol 7, 506–506.

Wang, C., Han, B., Zhou, R., and Zhuang, X. (2016). Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell 165, 990–1001.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., and Luo, Y. (2012). RNAscope A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. J Mol Diagnostics 14, 22–29.

Wang, G., Moffitt, J.R., and Zhuang, X. (2018). Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. Sci Rep-Uk 8, 4847.

Wegener, M., and Müller-McNicoll, M. (2018). Nuclear retention of mRNAs – quality control, gene regulation and human disease. Semin Cell Dev Biol 79, 131–142.

Weirich, C.S., Erzberger, J.P., Flick, J.S., Berger, J.M., Thorner, J., and Weis, K. (2006). Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. Nat Cell Biol 8, ncb1424.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by Eukaryotic Translation Initiation Factors. Mol Cell 2, 135–140.

Wetterberg, I., Baurén, G., and Wieslander, L. (1996). The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. Rna New York N Y 2, 641–651.

Wetterberg, I., Zhao, J., Masich, S., Wieslander, L., and Skoglund, U. (2001). In situ transcription and splicing in the Balbiani ring 3 gene. Embo J 20, 2564–2574.

Wilbertz, J.H., Voigt, F., Horvathova, I., Roth, G., Zhan, Y., and Chao, J.A. (2019). Single-Molecule Imaging of mRNA Localization and Regulation during the Integrated Stress Response. Mol Cell 73, 946-958.e7.

Wu, B., Chao, J.A., and Singer, R.H. (2012). Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells. Biophys J 102, 2936–2944.

Wu, B., Miskolci, V., Sato, H., Tutucci, E., Kenworthy, C.A., Donnelly, S.K., Yoon, Y.J., Cox, D., Singer, R.H., and Hodgson, L. (2015a). Synonymous modification results in high-fidelity gene expression of repetitive protein and nucleotide sequences. Gene Dev 29, 876–886.

Wu, B., Buxbaum, A.R., Katz, Z.B., Yoon, Y.J., and Singer, R.H. (2015b). Quantifying Protein-mRNA Interactions in Single Live Cells. Cell 162, 211–220.

Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. Science 352, 1430–1435.

Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016). Dynamics of Translation of Single mRNA Molecules In Vivo. Cell 165, 976–989.

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

Yunger, S., Rosenfeld, L., Garini, Y., and Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. Nat Methods 7, 631.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 39, ng.2007.26.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biology 15, nsmb.1514.

Zhao, N., Kamijo, K., Fox, P., Oda, H., Morisaki, T., Sato, Y., Kimura, H., and Stasevich, T.J. (2018). A genetically encoded probe for imaging HA-tagged protein translation, localization, and dynamics in living cells and animals. bioRxiv, p.474668.

c. Probing the conformational state of mRNPs using smFISH and SIM

Srivathsan Adivarahan and Daniel Zenklusen

Département de Biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, Canada

Published in RNA Remodeling Proteins, Humana, New York, NY, 2021.

Abstract

mRNAs and lncRNAs assemble with RNA-binding proteins (RBPs) to form ribonucleoprotein complexes (RNPs). The assembly of RNPs initiates co-transcriptionally and their composition and organization is thought to change during the different steps of an RNP live-cycle. Modulation of RNP structural organization have been implicated in the regulation of different aspects of RNA metabolism, including establishing interactions between the 5' and 3' ends in regulating mRNA translation and turnover. In this chapter, we describe a single molecule microscopy approach that combines fluorescent RNA *in situ* hybridization (smFISH) and Structured Illumination Microscopy (SIM) and allows to measure different aspects of RNP organization in cells, including distances between different regions within individual mRNAs, as well as the overall compaction state of RNAs in different subcellular compartments and environmental conditions. Moreover, we describe a detailed workflow required for image registration and analysis that allows determining distances at sub-diffraction resolution.

Key words: smFISH, mRNA, mRNPs, lncRNAs, RNA organization, RNA structure, single molecule resolution microscopy, Structured Illumination Microscopy

Introduction

The regulation of gene expression involves many steps and requires the interplay of RNA and RNA-binding proteins (RBPs). RNP organization and composition changes during the RNA lifecycle as RBPs associate with mRNAs at different steps of the gene regulation pathway [1, 2]. RNAs itself can form extensive local secondary structures as well as long distance intramolecular interactions through Watson-Crick base pairing and many different non-Watson-Crick interactions, that, in turn can provide additional domains for the binding of RBPs [2, 4]. RBPs, aside from playing a crucial role in regulating many catalytic steps during RNA metabolism such as splicing, 3' processing or degradation, act as important structural components of the RNP contributing to their organization as 3D assemblies. RBP binding can influence RNP organization either by stabilizing or destabilizing local secondary structures or by allowing additional short and long distances intramolecular interactions with other RBPs through their homo and hetero-dimerization domains[5].

Propagation though the different steps of RNA metabolism requires RNP reorganization at different scales. Among the best studied long-distance rearrangement is establishing the communication between the 5' and 3' ends within mRNAs. During translation, the 5' and 3' ends of the mRNA are thought to be brought together through interactions between the cytoplasmic cap-binding protein eIF4E, the adapter protein eIF4G and the poly(A) binding protein PABC1. In addition to stimulating translation, these interactions could potentially play a role in regulating RNA decay by bringing the poly(A) tail and cap in close proximity[6]. In contrast to the detailed view of the factors participating in 5'-3' communications that were largely obtained through *in vitro* and *in vivo* biochemical and genetic approaches, few current methods exist that allow to directly interrogate such RNP (re-) organization in cells.

RNP organization has been studied using different approaches, with X-ray crystallography, NMR and cryo-electron microscopy playing a crucial role in constructing high resolution structures of RNPs at different scales, from tRNAs to the ribosome. However, aside from ribosomes that have a well-defined 3D structure, neither high resolution structures nor principles that define the 3D organization of large RNPs been investigated[3, 6]. In particular, much less is known on the overall structural organization of mRNAs and lncRNAs, thought to be more difficult to study using these structural approaches due to the presumption that these molecules are either less well defined as 3D assemblies and/or get frequently rearranged during and while transiting between different steps of the gene regulation pathway[1]. Despite these difficulties, numerous *in vivo* and *in vitro*

approaches have been developed to interrogate RNP organization, some focused on determining local secondary structures using various chemical probing approaches, others able to map local as well as long range contacts mediated by protein-RNA, protein-protein or RNA-RNA interactions[3, 7].

Microscopy approaches provides another powerful tool to study RNP organization in cells[7, 13]. Here we describe a single molecule imaging approach that combines single molecule resolution *in situ* hybridization (smFISH) approach with Structured Illumination Microscopy (SIM) to study RNA organization in fixed cells. The method makes use of the ability to localize the signals emitted from single molecules with sub-diffraction resolution using 2D or 3D Gaussian fitting, allowing for a localization precision of around 20 nm using the setup described here. RNAs are detected using 20nt long DNA probes, each labeled with a single fluorescent dye and hybridized to paraformaldehyde fixed tissue culture cells. Probes hybridizing to different regions can be labeled with different dyes allowing for measurement of intramolecular distances between different regions within an RNA. Moreover, using probes targeting multiple regions within an mRNA allows determining the overall compaction state of RNAs[17]. However, accurately measuring intramolecular distances requires careful image registration and a stable microscopy setup. We describe different aspects of the imaging and image analysis workflow required to achieve such measurements (Figure 1).

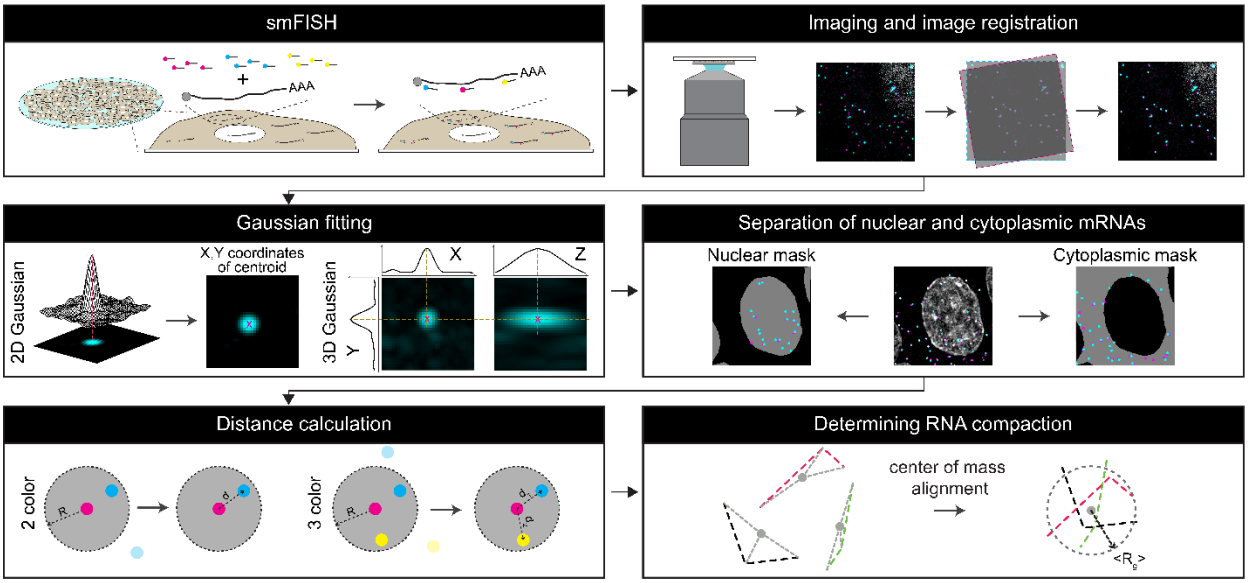


Figure 1

***Figure 1**: Schematic illustrating the imaging and image analysis workflow used to measure conformational states and compaction of single mRNPs using smFISH and Structured Illumination Microscopy.*

Materials

Prepare all reagents and solutions using ultrapure water and analytical grade reagents. Store all reagents at room temperature unless stated otherwise.

Chemicals

1. Ethanol
2. 20mg/ml Bovine Serum Albumin (BSA; NEB: Cat# B9000S)
3. Formamide (Sigma: Cat#F9037-100ML)
4. Puromycin dihydrochloride (Sigma: Cat#P8833)
5. Cyclohexamide (Sigma: Cat#C7698-1G)
6. Sodium arsenite (Sigma : Cat#35000-1L-R)
7. Homoharringtonine (Sigma: Cat#SML1091-10MG)
8. 50% Dextran sulfate solution (Millipore: Cat# S4030)
9. 32% paraformaldehyde in H_2O, methanol free and RNase free (Electron Microscopy Sciences: Cat#15714)
10. 200mM ribonucleoside vanadyl complexes (RVC; NEB: Cat# S1402S)
11. Triton X-100
12. Sodium chloride (NaCl)
13. Potassium chloride (KCl)
14. Sodium phosphate ($Na_2HPO_4.7H_2O$)
15. Potassium dihydrogen phosphate (KH_2PO_4)
16. Sodium citrate dihydrate (HOC(COONa)(CH2COONa)2 · 2H2O)
17. 4′,6-Diamidino-2-phenylindole dihydrochloride (DAPI)
18. Poly-L-Lysine (0.1 % (w/v) in H_2O; Sigma: Cat#P8920)
19. Hydrochloric acid (HCl)
20. Sodium Hydroxide (NaOH)
21. Sodium bicarbonate ($NaHCO_3$)

22. E. coli tRNA (Roche: Cat# 10109541001)
23. Salmon sperm DNA Solution (ssDNA) – 10mg/ml (Thermofisher: Cat# 15632011)

Buffers and Solutions

1. 10X PBS – Dissolve 80g NaCl, 2g KCl, 26.8g $Na_2HPO_4.7H_2O$ and 2.4g KH_2PO_4 in 800ml water. Adjust pH to 6.8 before making up the volume to 1L. pH when diluted to 1X will be 7.4.
2. 20X SSC (3M NaCl and 0.3M Sodium Citrate) - Dissolve 175.3 g of NaCl and 88.2 g of sodium citrate dihydrate in 800 ml of H2O. Adjust the pH to 7.0 with HCl and make up the volume to 1L.
3. Hybridization solution (10% Formamide, 2X SSC, 10% Dextran Sulfate) – For 10ml hybridization solution, add 1ml of 100% Formamide, 1ml of 20X SSC and 2ml of 50% dextran sulfate. Make up the volume to 1mL with deionized water. Aliquot and store at -20°C.
4. Washing solution (10% Formamide, 2X SSC) – For 10ml, add 1ml of 100% Formamide and 1ml of 20X SSC. Make up the volume to 10ml with deionized water. Prepare fresh every time.
5. Labeling solution (0.1M sodium bicarbonate, pH = 8.3) – To 84.01 mg of sodium bicarbonate, add 9 ml of deionized water. Adjust pH with NaOH/HCl if necessary, to 8.3. Make up the volume to 10ml. Aliquot and store at -20°C.
6. 70% Ethanol - diluted in deionized water
7. 1X PBS – 10X PBS diluted in deionized water.
8. 5% BSA stock solution
9. Permeabilization solution (0.1% Triton X-100, 0.5%BSA and 1x PBS) – for 50ml – add 50µl of Triton X- 100, 5ml of 5% BSA stock solution and 5ml of 10X PBS to a 50ml tube and make up the volume to 50ml with deionized water.

General laboratory equipment

1. Fine-tip forceps
2. Aluminium foil
3. Parafilm
4. 37°C incubator

5. Fume hood
6. Heating blocks for tubes or a thermocycler
7. Sterile 0.2 µm syringe filters
8. Syringes
9. Speed vacuum centrifuge

Cell culture

1. Culture medium: DMEM (Wisent: Cat#319-005-CL) + 10% Fetal Bovine Serum (Wisent: Cat#080-150)
2. Cultured adherent cell lines grown and passaged regularly (two-three times a week) using Trypsin (Wisent: Cat#325-043-EL)
3. 12 well plates
4. Circular 18mm diameter #1.5 thickness cover glass (Electron Microscopy Sciences Cat#72222-01)
5. 0.01% Poly-L-Lysine in deionized water. Prepare fresh.
6. Fixation Solution (4% PFA and 1X PBS) – For 10ml, add 1.25ml of 32% PFA, 1ml of 10X PBS and make up the volume to 10ml using deionized water.

Sample preparation

1. smFISH probes: 30-48 20mer DNA oligonucleotides with a 3' amine modification designed using Stellaris RNA FISH probes designer and ordered from Biosearch technologies.
2. Succinimidyl (NHS) Ester form of dye of interest. For far red channel – Cy5 (GE Healthcare Cat# GEPA25001)/Alexa647 (Thermo Scientific Cat# A37573), red/orange channel- Cy3 (GE Healthcare Cat# GEPA23001), Dylight 550 (Thermo Scientific Cat#62263) and green channel – Dylight 488 (Thermo Scientific Cat#46403), Atto488 (Sigma Cat# 41698-1MG-F).
3. Qiagen Nucleotide removal kit (Cat# 28304)
4. 100 nm tetraspeck beads (Thermofisher Cat# T7279)
5. DAPI stock solution: 5 mg/ml in deionized H_2O. Store at −20°C.
6. DAPI working solution: 2.5 µg/ml in 1x PBS, freshly prepared for each experiment and stored in dark.

7. Clean glass microscopy slides
8. Antifade mountant (e.g., Prolong Gold, Prolong Diamond, Prolong Glass).
9. Transparent nail polish
10. tRNA solution – 10 mg/ml E. coli tRNA. Store at -20°C.
11. ssDNA/tRNA mix – 5mg/ml ssDNA and 5mg/ml tRNA – mix 500 µl of ssDNA solution with 500 µl of tRNA solution. Store at -20°C.

Microscope setup

Zeiss Elyra PS1 imaging system used for image acquisition in the experiments described below is equipped with 50mW 405 nm, 100mW 488 nm, 100mW 561 nm and 150mW 642 nm lasers and 4 emission filters – i) BP420-480+LP750 ii) BP495-590+LP750 iii) LP570 and iv) LP655, a 63X 1.4 NA Plan Apo DIC II oil objective with a working distance of 0.1mm, EMCCD camera (Andor iXon3 DU-885K with a chip size of 1004x1002 pixels and pixel size of 8 µm), a piezo stage for high precision focus in Z and step size of 25 nm and a high performance computer workstation for imaging, image processing and image analysis. The microscope is stored in a temperature-controlled room to minimize temperature fluctuations.

Image analysis and data representation

1. ZEN black
2. FIJI
3. Localize
4. AIR Localize
5. MATLAB
6. R Studio for plotting

Methods

Probe designing and labeling – Depending on the length of an mRNA of interest, probes can be designed against two or more regions within the mRNA. Probes targeting the different regions have to be labeled with dyes that can be spectrally separated using a specific microscope setup (Note 1).

1. To design probes against multiple regions, choose a ~1000 nt region within the region as template and use the Stellaris probe designer tool to design 30-48 probes against the target. Once designed, probes can either be ordered conjugated with fluorescent dyes (steps 2-7 are not required), or containing a single amine modification at the 3' end that allows post-synthesis labeling with different fluorophores (Note 2). An alternative protocol for probe labeling has been described by Gaspar et al. [8, 14]
2. For post synthesis labeling, the probes corresponding to one target region are pooled in equimolar amounts prior to labeling. Pooled probes can be stored at -20°C for many months. We have efficiently labeled probes stored for several years.
3. The pooled probes can be used for labeling using NHS ester dyes.
4. NHS ester dyes can be purchased as value packs (available as Mono 5-pack for Cyanine dyes, 5x50µg packs for Dylight dyes), in which case one vial can be used for labeling 20µg of probes. Alternatively, dyes can be aliquoted to equivalent concentrations as described in [7] (Note 3)
5. Use one vial of NHS ester dye pack or equivalent aliquoted dye to label 2x 10µg of probes. Aliquot 10µg of probes and dry the probe mix using a speed vacuum evaporator. Resuspend one vial of dye in 30 µL of labeling solution and add 15µl of the resuspended dye to the probes. Incubate o/n in the dark.
6. Remove unincorporated dye using a nucleotide removing spin column such as Qiagen nucleotide removal kit. Alternatively, probes can also be purified using a size exclusion column as described in [7]
7. Measure the concentration and determine the labeling efficiency to ensure a high labeling efficiency for the probes. If the labeling is low (<80%), relabel the purified labelled probes using another vial of dye. Labeling efficiency is calculated by taking the ratio of the picomolar concentration of dye and the oligo in the sample.

$$Labelling\ Efficiency\ =\ (A_{dye} *\ 20 * \varepsilon_{base})/(A_{base} * \varepsilon_{dye})$$

where A_{dye} is the absorption of the dye measure at peak **λ**max , A_{base} is the absorption at 260 nm, ε_{base} is the average molar extinction coefficient of ss DNA base (8,919 M-1cm-1) and ε_{base} is the molar extinction coefficient of the dye (in M-1cm-1)

8. Store the probes in the dark at -20°C (Note 4)

Preparing coverslips, seeding cells, treatment and fixation – Adherent cells are plated on 18mm circular coverslips in 12 well dishes before treatment with drugs and fixation.

1. Coverslips are thoroughly washed with deionized water and stored in 70% Ethanol before use. #1.5 thickness coverslips compatible with a high NA oil objective should be used.
2. To prepare the coverslips for adherent cells, first take out individual coverslips and place them vertically resting on the side walls of a 10cm dish in a laminar hood. After the ethanol has evaporated, immerse the coverslips in freshly prepared 0.01% Poly-L-Lysine solution (filtered using a 0.2 μm syringe filter) in a 10cm dish for 15-30min. Use 15ml of 0.01% per 10 cm dish.
3. Remove the Poly-L-Lysine solution and wash the coverslips twice with autoclaved deionized water. After second wash, remove the coverslips and rest them vertically against the wall of a 10cm dish till they dry. After drying, each coverslip is placed in a well of a 12 well plate. Some cell types might require coverslips treated with a different matrix.
4. Cells should be added such that they reach 60-80% confluency on the day of fixation. To get an optimal coverslip for imaging, cells should be added such that they are uniformly distributed throughout the coverslips and not exist in clusters. Addition of cells that results in multiple layers or a lot of cell debris should be avoided. 1-2ml of media is added per well and cells are allowed to adhere overnight.
5. On the day of fixation, first check the coverslips under the microscope to ensure that they are ready for fixation. If they need to be treated with drugs, dilute the stock concentration of drug in culture media and replace the media within the well with media containing the appropriate concentration of the drug. For mock, add the same volume of solvent used for preparing the stock concentration of the drug should be used.
6. After drug treatment, remove media and rinse with 1ml of warm 1X PBS.
7. Remove the PBS and add 1ml of fixation solution (4% PFA, 1X PBS) and incubate for 10min at RT.
8. Wash with 1x PBS for 5min. Repeat the wash step.

9. Remove the remaining PBS and add 2ml of ice-cold 70% ethanol for permeabilization and storage. Seal the 12 well plate with parafilm to ensure that the ethanol doesn't dry. Store at -20°C. Coverslips can be stored for up to 6 months.
10. Alternately, cells can be permeabilized for 15 min using the permeabilization solution, washed 2 times with 1X PBS for 5 min. Store the coverslips at 4°C in 1X PBS. Coverslips can be stored for 2-3 days.

Hybridization and sample preparation – Fluorescent *in situ* hybridization can be performed using either two or three sets of probes labelled with spectrally distinguishable probes targeting different regions of a single transcript. If using a thermocycler, mix the hybridization solution in a PCR tube, else a 1.5ml reaction tube can be used. 20µl of hybridization solution is used for an 18mm diameter coverslip. If using a different size coverslip, ensure that the volume is sufficient to cover the entire area of the coverslip.

1. Prepare a fresh stock of washing solution (10% Formamide 2X SSC).
2. Use the fixed coverslips and transfer them to a single well of a 12 well plate containing 1ml of 1X PBS for rehydration. If using the coverslips stored in ethanol, ensure you remove excess ethanol. Do not let the coverslips dry.
3. Keep the coverslips in 1X PBS for 5 min.
4. Remove the PBS and replace it with 1ml of washing solution and leave for 5 min (Note 5).
5. At the same time, prepare the hybridization mix by adding 10-20 ng of each labelled probe to 4µl of ssDNA/tRNA mix and dry the mix using a speed vacuum centrifuge. This would be sufficient for one hybridization. (Note 6)
6. Once the probes have dried, add 20µl of hybridization solution (10% Formamide, 2X SSC and 10% Dextran Sulfate) and keep it at 95°C for 3-5 min for denaturation.
7. Spin down the probe solution and add 0.2µl of 200mM RVC and 0.2µl of 20mg/ml BSA (final concentration of RVC is 2mM and BSA is 0.2mg/ml BSA). (Note 7)
8. Vortex and spin down the probe solution.
9. Prepare hybridization chamber by stretching a film on parafilm on a square glass plate.
10. Add the hybridization solution as a single drop on top of the parafilm. Ensure that there are no bubbles.

11. Gently place the coverslip on the top of the droplet such that the cells are in contact with the hybridization solution. Ensure that the hybridization solution is spread uniformly across between the coverslip and the parafilm. Do not press down on the coverslip.
12. Seal the hybridization chamber using parafilm. Check to ensure that the top layer of parafilm doesn't press down on the coverslips. A small piece of paper or cardboard can be used for this purpose. A small wet piece of paper can also be added in between the two layers of parafilm to ensure that chamber remains humid.
13. Cover the chamber with aluminium foil and keep it at 37°C in the dark for 3 hrs.
14. After the hybridization, transfer the coverslip back to a 12 well plate. Add 1 ml of washing solution and cover the plate with aluminium foil. Keep at 37°C in the dark for 30 min.
15. Repeat Step 14.
16. Rinse with 1X PBS.
17. Add 1ml of 1X PBS containing DapI (DapI working solution) and incubate for 5 min at RT in the dark.
18. Rinse 2 times with 1X PBS
19. Take a clean slide and add a single drop of antifade mountant solution. Ensure that there are no air bubbles.
20. Remove the coverslip from the 12 well plate and remove excess PBS using Kimwipes.
21. Mount the coverslips such that the cells are on the side of the glass slide with the antifade mountant. Keep it at RT for at least 24 hrs before imaging to ensure that the antifade mountant hardens and attains the desired refractive index. For long term storage, keep the slides at 4°C or -20°C.
22. After the antifade mountant has hardened, seal the edges of the coverslips using nail polish.

Acquisition of 3D SIM images – The microscope should be switched on at least three hrs before acquiring images. This is essential to ensure that all the lasers have reached a state where the output is relatively stable during image acquisition. Furthermore, switching on the microscope before time allows for the components to reach their optimal temperature and can reduce thermal drift during imaging. If the coverslips were stored at 4°C or -20°C, keep them at RT for 1 hr. To reduce mechanical drift while imaging, ensure to wait for some time after placing each slide into the slide holder of the stage. For z stacks, generally around 25-30 stacks separated by 0.180 µm are acquired (Note 8. For all lasers, a grid size of 42 µm is used (Note 9). To reduce imaging time, images

corresponding to different phases are acquired for each stack and rotation and this is repeated for different channels. The exposure times are generally between 50 and 80 ms and the lasers adjusted to achieve maximum signal to noise ratio with minimal bleaching of the signal over the course of acquisition (Note 10). The images are processed using the inbuilt structured illumination processing tool in Zen Black 2012 (Note 11). Similarly, registration is performed using the inbuilt channel alignment tool.

1. To determine and correct for channel misalignment and chromatic aberration, use 100 nm Tetraspeck beads on a coverslip prepared such that the beads are interspaced uniformly through the coverslip. To achieve this, typically dilute 1-1.5µl of stock beads solution in 20 µl of tap water, spread them uniformly on a 18mm diameter coverslip and allow them to dry in the dark. Once dried, mount the coverslip using Prolong antifade mountant and keep in the dark at RT for at least 24 hrs before imaging. Seal the coverslips with nail polish (Note 12).
2. Acquire images of Tetraspeck beads for the channels that require correction and process the images using the default settings and the 3D-SIM option. The 3D-SIM uses the adjacent z stacks in order to process images that yield a higher resolution in the z direction (Note 13, 14).
3. Perform registration using the affine option. The affine option estimates parameters necessary to correct misalignment between channels due to translation, rotation, scale and shear related misalignments. Save the parameters.
4. Use probes targeting the same region of an mRNA and labelled such that alternating probes have either Cy5 or Cy3 to determine the precision of localization of the FISH signal achieved after registration (Note 15). Process the images and apply registration on the images from the data saved previously.
5. Repeat the process for all samples. Typically, image at least 50 cells (or 5-10 fields) for each sample (Note 16).

Image analysis and data quantification – The acquired 3D images can be processed either directly or converted to 2D images before analyzing them. Single mRNAs in cells are localized to high precision using Gaussian fitting algorithms developed for single particle localization in 2D and 3D. For 2D analysis, we use LocalizeApp, a progam that uses 2D Gaussian fitting as described

in [16, 21]. For 3D analysis, we use AIRLocalize [8, 15] (Note 17). The nuclear and cytoplasmic mRNAs are separated using manual masks created using ImageJ (Note 18). To separate nuclear and cytoplasmic mRNAs, identify spots from different channels that correspond to the same mRNA and calculate the distance between these spots, we use custom MATLAB scripts. For data visualization and statistical analysis, we use R (Note 19).

mRNP conformation analysis in 2D

1. Open 3D images in ImageJ and convert them to 2D using maximum intensity projection.
2. Split the images by channels, and save all individual channels containing RNA FISH signal within a field of view in the same folder (Note 20).
3. To separate and select for cytoplasmic RNAs, open the maximum intensity projected image containing all channels in imagej and use the free hand selection tool to manually select for non-overlapping and spatially separate RNA spots that do not overlap with the DapI signal. Select RNAs that have signal in all channels through the free hand selection tool in ImageJ.
4. Add the selected regions to the ROI. Repeat for all cytoplasmic mRNAs in the field.
5. Once all mRNAs have been selected, use the make masks plugin. Save the mask in the same folder as the other images.
6. Repeat Steps 3-5 for nuclear mRNAs (Note 21).
7. Use 'LocalizeApp.sav' to determine the position of RNA spots in 2D. Adjust the parameters such that only true signals are detected. Compare with the original image to determine the quality of localization. The coordinates of the localized RNAs are saved '.loc' files
8. Repeat for all channels
9. Open MATLAB and navigate to the parent folder containing subfolders, each containing the nuclear and cytoplasmic masks as well as the file containing the coordinates from the localized RNA spots in all channels.
10. To analyze two color images, run 'RNA2.m'. For three color images, run 'RNA3.m'. To analyze data to check for colocalization precision, use the script 'RNAloc.m'
11. The script will analyze the data and save the distance information in two files - 'Nuclear Distances.csv' and 'Cytoplasmic Distances.csv' within these subfolders.

12. In addition, ‘RNA3.m’ also saves the coordinates of the spots corresponding to an individual mRNA in two different files - ‘Cytoplasmic Spots.csv’ and ‘Nuclear Spots.csv’. These coordinates can be used to determine the radius of gyration (see below).

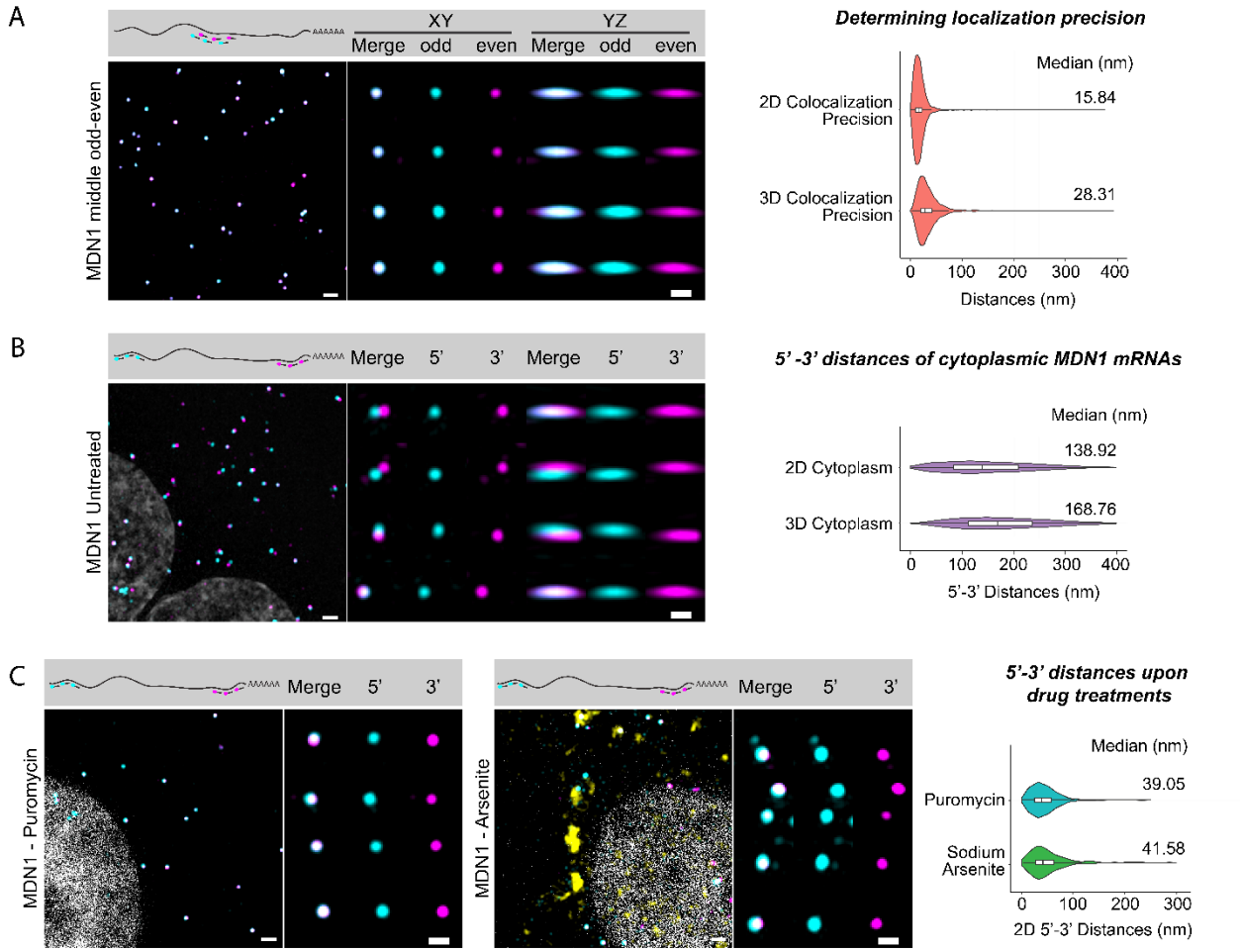


Figure 2

Figure 2: Measuring localization precision and 5’-3’ distances of single mRNAs using two color imaging. ***A)*** *Determining localization precision. 3D SIM images of HEK293T cells hybridized with smFISH probes alternatingly labeled with Cy3 and Cy3 targeting a region in the middle the MDN1 mRNA (left). Zoom in images of single mRNAs and violin plots representing 2D and 3D distance distribution are shown on the right. Box plots represent the median, first and third quartiles.* ***B)*** *Determining 5’ – 3’ distances of single mRNAs. 3D SIM Images of HEK293T cells hybridized with smFISH probes targeting the 5’ and 3’ regions of the MDN1 mRNA as well as zoom in and violin plots representing 5’-3’ distances as in A).* ***C)*** *Measuring RNA conformations*

upon translation inhibition and Arsenite stress. smFISH images and quantifications of 5'-3' ends for MDN1 mRNA as shown in B for U2OS cells treated with 100µg/ml Puromycin for 10min (left), 2mM Sodium Arsenite for 1hr (middle). The yellow signal in cells treated with Arsentie represents stress granules visualized using dT probes hybridizing to mRNAs polyA tails. Scale bar= 1 µm, scale bar in zoomed-in images is 500 nm. Median values are shown on the right of violin plots. Probe combination and positions are shown on the top of the images.

mRNP conformation analysis in 3D

1. Open 3D images in ImageJ and split the images according to the channels and save them as done for 2D analysis (Note 20).
2. In Imagej, make nuclear and cytoplasmic masks as shown previously for 2D image analysis (see above).
3. Open MATLAB and run 'AIRLOCALIZE.m'. Open each individual image with RNA spots and determine the position RNAs using AIRLOCALIZE. The localized spots are saved in '.loc3' files.
4. Repeat for all images.
5. Navigate to the parent folder within MATLAB that contains subfolders, similar to when analyzing 2D images.
6. Run RNA2_3D.m. This will calculate distances between RNAs within the masks and save the results in two files - 'Cytoplasmic Distances.csv' and 'Nuclear Distances.csv'.

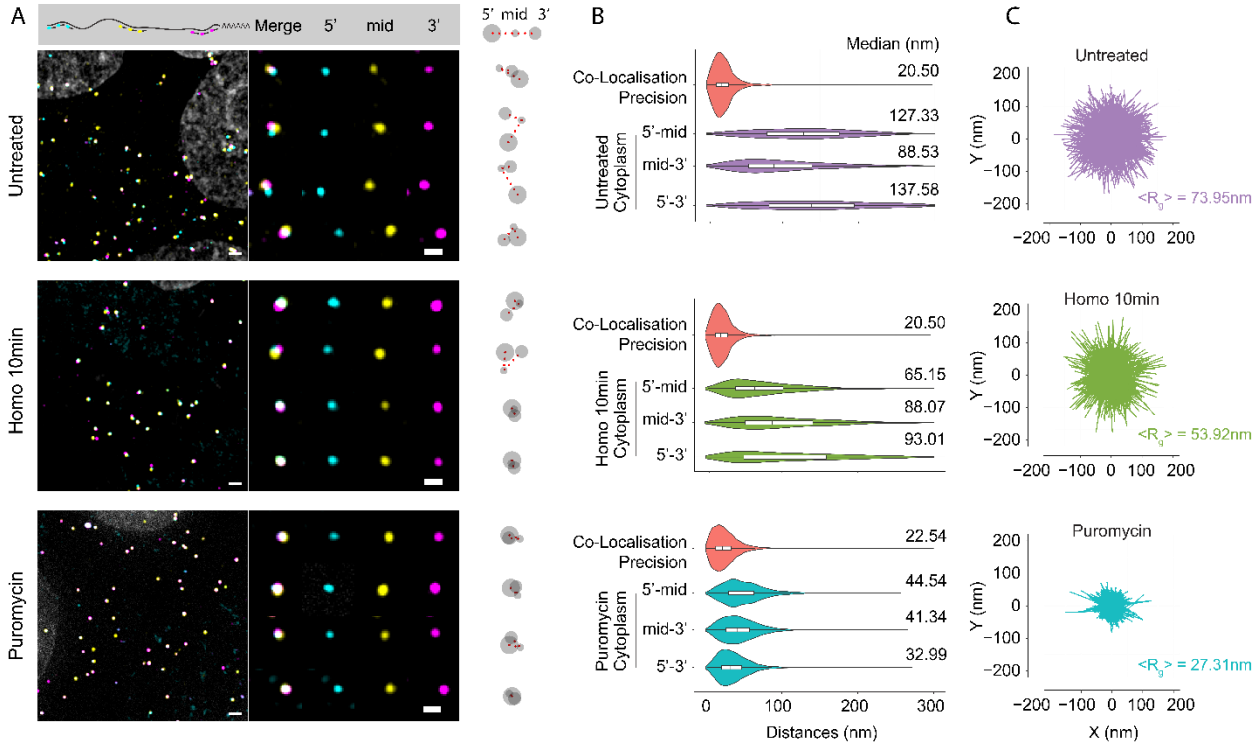


Figure 3

Figure 3: Determining mRNP conformations and compaction using three color imaging. *HEK293T cells were hybridized with MDN1 probes targeting the 5', middle and 3' regions.* ***A)*** *Maximum intensity projection of a field of view (left), zoom showing signal from single mRNAs (middle) and cartoon representing the corresponding mRNA conformations (right) for Untreated (top), 10min with 100µg/ml Homoharringtonine (middle) and 10min with 100µg/ml Puromycin (bottom). Scale bar= 1 µm, scale bar in zoomed in RNAs = 500 nm.* ***B)*** *Violin plot representing 2D distance distribution between 5'-middle, middle-3' and 5'-3' regions of cytoplasmic mRNA in Untreated (top), Homoharringtonine (middle) and Puromycin (bottom). Box plot representing the median, first and third quartiles. Median values are shown on the right of the plot.* ***C)*** *Overlay of all 2D projected mRNA conformations aligned using their center of mass for Untreated (top), Homoharringtonine (middle) and Puromycin (bottom). Calculated mean radius of gyration (<R-$_g$>) is shown on the bottom right. The axes represent the coordinates (in nm).*

Estimating p**arameters of mRNP compaction and performing statistical analysis -** Compaction levels of mRNAs can be determined when relative position of 3 or more regions of an mRNA are known. We use our 3 color imaging dataset to determine mRNP compaction using the factor - 'mean radius of gyration'. The mean radius of gyration(<Rg>) is calculated as follows:

$$< R_g > = \sqrt{1/3 \sum_{k=1}^{3} (r_k - r_{mean})^2}$$

where k represents one of the three regions of the mRNP and rk the position of the corresponding region in space determined by gaussian fitting.

Use the coordinates of different regions of mRNPs saved in the files 'Cytoplasmic Spots.csv' and 'Nuclear Spots.csv' (see above) and the above formula to calculate the mean radius of gyration.

The statistical significance of changes in distance distributions between regions of an mRNA is measured using the Kolmogorov-Smirnov test using the function 'ks.test' in R.

Notes

1. It is important to take into account the expression levels and isoforms for the investigated transcripts. If the expression level of the mRNA of interest is very high, it becomes challenging to spatially separate individual mRNAs and therefore difficult to determine their individual conformations. Furthermore, if mRNAs have multiple isoforms, it is important to consider which of the isoforms are expressed in a particular cell line and to ensure that probes are specific to the isoforms that are studied.

2. The number of probes is an important consideration to get a good RNA signal, which in turn is essential to achieve a high localization precision. The number of probes that can be designed to hybridize to a specific region depends on the length and composition of the target sequence. We use a probe design software from LGC Biosearch Technologies called *Stellaris Probe designer,* which searchers for probes of 20nt in length and a GC content of ~35-55%. Moreover, it excludes repetitive sequences within genomes of different organisms to ensure specificity. Other oligo design tools can also be used, but we have predominantly used this software and obtained good results. *Stellaris Probe designer* recommends the use of a minimum of 25 probes for a single target sequence, though in our hands we have obtained good RNA signal with as little as 15 probes using our imaging setup for probes labeled in the red and far red channels. Length and spacing between probes can also been alter if needed. Due to higher autofluorescence in the green channel, we

found that we needed at least 35 probes in the green channel to achieve acceptable signal-to-noise and localization precisions.

3. Aliquoting the dyes must be done in a moisture free environment and using anhydrous DMSO. This is to preserve the reactivity of the dye for longer periods. Aliquoted dyes in DMSO can be stored at -80°C.

4. We haven't found a considerable loss in FISH signal using aliquots of labelled FISH probes thawed multiple times over a short period of time. However, it is recommended that if the probes are to be used over longer periods, they be aliquoted and stored at -20°C.

5. It is important to be careful while pipetting as some cell lines are prone to detaching from the coverslips when pipetted directly on the face of the coverslip.

6. Ensure that all steps involved in handling of labelled probes or dyes are done with minimal light exposure.

7. Addition of BSA is optional. We have not observed differences in signal when omitting BSA in the buffer.

8. The Z stack interval was optimized to get the minimal interval that resulted in highest localization precision for our imaging setup. Moreover, the number for the z stacks was chosen to cover a considerable volume of the cell while keeping acquisition times low (see also Note 10)

9. By default, the grid sizes allocated to the lasers are different. For the 405 nm, it is 23 µm, for the 488 nm laser it is 28 nm, for the 561 and 642 nm lasers it is 34 µm. However, in our hands we observed that the best signal and localization was achieved for the 42 µm grid for all lasers.

10. It is important to ensure that the acquisition time is kept to a minimum. We have experienced that longer acquisitions due to either increased number of z stacks or increased

exposure time resulted in a decrease in co-localization precision. This is mostly likely due to mechanical or thermal drift that the sample experiences during acquisition.

11. A high NA objective with a very small point spread function (PSF) is preferable for imaging to get the best results. The PSF of the objective should be measured regularly using 100 nm TetraSpeck beads and this PSF can be used for reconstruction of the 3D SIM image. Regular measurement of the PSF also helps avoid artifacts during reconstruction due to possible defects in the objective.

12. Registration with beads should be performed every day of imaging. Though, the microscope system is relatively stable for short periods of time, we have found that the parameters for alignment of channels can change over time.

13. Image processing using the 3D SIM option is essential when analyzing the images in 3D. However, images can be processed in using the 2D-SIM which does not yield a high z resolution but provides similar localization results in 2D. We have found that using the experimentally measured PSF did not yield in a higher localization precision when compared to using the option of a theoretical PSF when processing images. However, this could vary depending on the quality of the objective.

14. TetraSpeck beads are not very bright when illuminated with a 405 nm laser and as this channel is essentially used for nuclear marker, it is skipped when performing registration.

15. We chose two different regions of MDN1 (each around 1.2-1.5 kb in length) to test for precision of localization (or 'colocalization precision'). We did not find a difference in colocalization precision between the two regions tested. However, this might not apply to longer sequences which correspond to regions of mRNAs that are no longer diffraction limited

16. Although the number of mRNAs can vary based on transcript or cell line of choice, for many of the mRNA studied, we found that 50 cells gave us at least 500 RNAs. We also found that quantifying data from more than 500 mRNAs did not significantly alter our

observations when measuring mRNA intramolecular distances. However, for low abundant RNAs or the cell type, this number might sometimes be difficult to achieve and required imaging many more cells.

17. LocalizeApp is freely available on Dan Larson's lab website (https://ccr.cancer.gov/Laboratory-of-Receptor-Biology-and-Gene-Expression/daniel-r-larson).

18. AIRLocalize is available in Timothee Lione's lab webiste (http://www.timotheelionnet.net/software/). In addition to AIRLocalize and LocalizeApp, different other software packages allow 2D and 3D quantifications, such as FISH-Quant which allows both 2D and 3D quantification) [19]

19. The plugin was developed in Robert Singer's laboratory and can be downloaded from [https://github.com/zenklusenlab/ImageJ-plugins]. The plugin creates a mask using all the ROI in the ROI Manager in ImageJ.

20. It is assumed that the user has a minimal understanding of MATLAB. All the MATLAB scripts required for data quantification can be found here [https://github.com/zenklusenlab/MolCell_DistanceCalc].

21. The scripts rely on a predefined folder structure and nomenclature of different input files. The scripts are designed to run and analyze data within subfolders and save the output files within the same subfolders. Each subfolder, for instance, contains all the information relevant for data quantification of one field - the localization ('.loc' or '.loc3' files, the nuclear and cytoplasmic masks - Also see Note 21). For 2 color images, the image files corresponding to RNA FISH signal should be saved as Cy5.tif and Cy3.tif (and the corresponding localization files as Cy5.loc and Cy3.loc). For 3 color images, the image files should be saved as '5p.tif', '3p.tif' and 'mid.tif', where 5p, 3p and mid correspond to RNA signal for the 5' end, 3' end and middle region respectively.

22. The mask files should be saved as 'Cymask.tif' and 'Nucmask.tif' (for cytoplasmic and nuclear masks respectively) and should be saved in the same subfolder as other data from the same field of view.

References

1. Björk P, Wieslander L (2015) The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA–Protein Complexes. Annu Rev Biochem 84:65–92. doi: 10.1146/annurev-biochem-060614-034150

2. Brandt F, Carlson L-A, Hartl UF, Baumeister W, Grünewald K (2010) The Three-Dimensional Organization of Polyribosomes in Intact Human Cells. Mol Cell 39:560–9. doi: 10.1016/j.molcel.2010.08.003

3. Christensen KA, Bourne CM (1999) Shape of large bound polysomes in cultured fibroblasts and thyroid epithelial cells. Anatomical Rec 255:116–129. doi: 10.1002/(sici)1097-0185(19990601)255:2<116::aid-ar2>3.0.co;2-o

4. Christensen KA, Kahn LE, Bourne CM (1987) Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat 178:1–10. doi: 10.1002/aja.1001780102

5. Dimitrova-Paternoga L, Jagtap P, Chen P-C, Hennig J (2019) Integrative Structural Biology of Protein-RNA Complexes. Structure 28:6–28. doi: 10.1016/j.str.2019.11.017

6. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM (2019) The roles of structural dynamics in the cellular functions of RNAs. Nature Reviews Molecular Cell Biology 20. doi: 10.1038/s41580-019-0136-0

7. Gáspár I, Wippich F, Ephrussi A (2018) Terminal Deoxynucleotidyl Transferase Mediated Production of Labeled Probes for Single-molecule FISH or RNA Capture. Bio-protocol 8. doi: 10.21769/bioprotoc.2750

8. Hentze MW, Castello A, Schwarzl T, Preiss T (2018) A brave new world of RNA-binding proteins. Nat Rev Mol Cell Bio 19:327. doi: 10.1038/nrm.2017.130

9. Jones S (2016) Protein–RNA interactions: structural biology and computational modeling techniques. Biophysical Rev 8:359–367. doi: 10.1007/s12551-016-0223-9

10. Khong A, Parker R (2018) mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. J Cell Biol 217:jcb.201806183. doi: 10.1083/jcb.201806183

11. Koch A, Aguilera L, Morisaki T, Munsky B, Stasevich TJ (2020) Quantifying the spatiotemporal dynamics of IRES versus Cap translation with single-molecule resolution in living cells. Biorxiv 2020.01.09.900829. doi: 10.1101/2020.01.09.900829

12. Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y, Wells AL, Chao JA, Park H, de Turris V, Lopez-Jones M, Singer RH (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods 8:165. doi: 10.1038/nmeth.1551

13. Lu Z, Chang HY (2016) Decoding the RNA structurome. Curr Opin Struc Biol 36:142–8. doi: 10.1016/j.sbi.2016.01.007

14. Mitchell D, Assmann SM, Bevilacqua PC (2019) Probing RNA structure in vivo. Curr Opin Struc Biol 59:151–158. doi: 10.1016/j.sbi.2019.07.008

15. Mitchell SF, Parker R (2014) Principles and Properties of Eukaryotic mRNPs. Mol Cell 54:547–558. doi: 10.1016/j.molcel.2014.04.033

16. Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, Shav-Tal Y (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12:543. doi: 10.1038/ncb2056

17. Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, Basyuk E, Bertrand E, Darzacq X, Zimmer C (2013) FISH-quant: automatic counting of transcripts in 3D FISH images. Nat Methods 10:277. doi: 10.1038/nmeth.2406

18. Pelletier J, Sonenberg N (2019) The Organizing Principles of Eukaryotic Ribosome Recruitment. Annu Rev Biochem 88:307–335. doi: 10.1146/annurev-biochem-013118-111042

19. Singh G, Pratt G, Yeo GW, Moore MJ (2015) The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem 84:1–30. doi: 10.1146/annurev-biochem-080111-092106

20. Thompson RE, Larson DR, Webb WW (2002) Precise Nanometer Localization Analysis for Individual Fluorescent Probes. Biophys J 82:2775–2783. doi: 10.1016/s0006-3495(02)75618-x

21.Adivarahan S, Livingston N, Nicholson B, Rahman S, Wu B, Rissland OS, Zenklusen D (2018) Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol Cell 72:727-738.e5. doi: 10.1016/j.molcel.2018.10.010

22. Adivarahan S, Zenklusen D (2019) Lessons from (pre-)mRNA Imaging. Adv Exp Med Biol 1203:247–284. doi: 10.1007/978-3-030-31434-7_9

23. Vicens Q, Kieft JS, Rissland OS (2018) Revisiting the Closed-Loop Model and the Nature of mRNA 5′–3′ Communication. Mol Cell 72:805–812. doi: 10.1016/j.molcel.2018.10.047

24. West JA, Mito M, Kurosaka S, Takumi T, Tanegashima C, Chujo T, Yanaka K, Kingston RE, Hirose T, Bond C, Fox A, Nakagawa S (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. J Cell Biol 214:817–830. doi: 10.1083/jcb.201601071

25. Wheeler EC, Nostrand EL, Yeo GW (2017) Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley Interdiscip Rev Rna 9:e1436. doi: 10.1002/wrna.1436

26. Yamazaki T, Souquere S, Chujo T, Kobelke S, Chong Y, Fox AH, Bond CS, Nakagawa S, Pierron G, Hirose T (2018) Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Mol Cell 70:1038-1053.e7. doi: 10.1016/j.molcel.2018.05.019

27. Zenklusen D, Larson DR, Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biology 15:nsmb.1514. doi: 10.1038/nsmb.1514

28. Zhao J, Qian X, Yeung P, Zhang Q, Kwok C (2019) Mapping In Vivo RNA Structures and Interactions. Trends Biochem Sci 44:555–556. doi: 10.1016/j.tibs.2019.01.012

Bibliography

Adivarahan, S., and Zenklusen, D. (2019). Lessons from (pre-)mRNA Imaging. Adv. Exp. Med. Biol. *1203*, 247–284.

Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol. Cell *72*, 727-738.e5.

Afonina, Z.A., Myasnikov, A.G., Khabibullina, N.F., Belorusova, A.Y., Menetret, J.-F., Vasiliev, V.D., Klaholz, B.P., Shirokov, V.A., and Spirin, A.S. (2013). Topology of mRNA chain in isolated eukaryotic double-row polyribosomes. Biochemistry *78*, 445–454.

Afonina, Z.A., Myasnikov, A.G., Shirokov, V.A., Klaholz, B.P., and Spirin, A.S. (2014). Formation of circular polyribosomes on eukaryotic mRNA without cap-structure and poly(A)-tail: a cryo electron tomography study. Nucleic Acids Res. *42*, 9461–9469.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. PNAS *100*, 3889–3894.

Badis, G., Saveanu, C., Fromont-Racine, M., and Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. Mol. Cell *15*, 5–15.

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

Batisse, J., Batisse, C., Budd, A., Böttcher, B., and Hurt, E. (2009). Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. J. Biol. Chem. *284*, 34911–34917.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754–765.

Beyer, A.L., Christensen, M.E., Walker, B.W., and LeStourgeon, W.M. (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. Cell *11*, 127–138.

Björk, P., and Wieslander, L. (2015). The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA–Protein Complexes. Annu. Rev. Biochem. *84*, 65–92.

Björk, P., and Wieslander, L. (2017). Integration of mRNP formation and export. Cell. Mol. Life Sci.

Björk, P., Jin, S., Zhao, J., Singh, O.P., Persson, J.-O., Hellman, U., and Wieslander, L. (2009). Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions. J. Cell Biol. *184*, 555–568.

Björk, P., Persson, J.-O., and Wieslander, L. (2015). Intranuclear binding in space and time of exon junction complex and NXF1 to premRNPs/mRNPs in vivo. J. Cell Biol. *211*, 63–75.

Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., et al. (2018). Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. Mol. Cell *72*, 496-509.e9.

Boehm, V., and Gehring, N.H. (2016). Exon Junction Complexes: Supervising the Gene Expression Assembly Line. Trends Genet. *32*, 724–735.

Boo, S.H., and Kim, Y.K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. Exp. Mol. Med. *52*, 400–408.

Borodavka, A., Tuma, R., and Stockley, P.G. (2013). A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. RNA Biol. *10*, 481–489.

Borodavka, A., Singaram, S.W., Stockley, P.G., Gelbart, W.M., Ben-Shaul, A., and Tuma, R. (2016). Sizes of Long RNA Molecules Are Determined by the Branching Patterns of Their Secondary Structures. Biophys. J. *111*, 2077–2085.

Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. Nat. Rev. Mol. Cell Biol. *17*, 426–438.

Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. Cell *136*, 261–271.

Brandt, F., Carlson, L.-A., Hartl, F.U., Baumeister, W., and Grünewald, K. (2010). The three-dimensional organization of polyribosomes in intact human cells. Mol. Cell *39*, 560–569.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol. Cell *44*, 120–133.

Braun, J.E., Friedman, L.J., Gelles, J., and Moore, M.J. (2018). Synergistic assembly of human pre-spliceosomes across introns and exons. Elife *7*.

Brodsky, A.S., Meyer, C.A., Swinburne, I.A., Hall, G., Keenan, B.J., Liu, X.S., Fox, E.A., and Silver, P.A. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. Genome Biol. *6*, R64.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.-M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. *9*, e1000573.

Brümmer, A., Yang, Y., Chan, T.W., and Xiao, X. (2017). Structure-mediated modulation of mRNA abundance by A-to-I editing. Nat. Commun. *8*, 1255.

Bruun, G.H., Doktor, T.K., Borch-Jensen, J., Masuda, A., Krainer, A.R., Ohno, K., and Andresen, B.S. (2016). Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. BMC Biol. *14*, 54.

Buratti, E., and Baralle, F.E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol. Cell. Biol. *24*, 10505–10514.

Caponigro, G., and Parker, R. (1995). Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. Genes Dev. *9*, 2421–2432.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285–298.

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

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

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat. Struct. Mol. Biol. *18*, 1218–1226.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. *10*, 741–754.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389–1400.

Christensen, A.K., and Bourne, C.M. (1999). Shape of large bound polysomes in cultured fibroblasts and thyroid epithelial cells. Anat. Rec. *255*, 116–129.

Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am. J. Anat. *178*, 1–10.

Clote, P., Ponty, Y., and Steyaert, J.-M. (2012). Expected distance between terminal nucleotides of RNA secondary structures. J. Math. Biol. *65*, 581–599.

Costello, J., Castelli, L.M., Rowe, W., Kershaw, C.J., Talavera, D., Mohammad-Qureshi, S.S., Sims, P.F.G., Grant, C.M., Pavitt, G.D., Hubbard, S.J., et al. (2015). Global mRNA selection mechanisms for translation initiation. Genome Biol. *16*, 10.

Coulon, A., Ferguson, M.L., de Turris, V., Palangat, M., Chow, C.C., and Larson, D.R. (2014). Kinetic competition during the transcription cycle results in stochastic RNA processing. Elife *3*.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol. Cell *73*, 398–411.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. Genes Dev. *20*, 1100–1109.

David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev. *25*, 972–983.

De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip. Rev. RNA *4*, 49–60.

Delavoie, F., Soldan, V., Rinaldi, D., Dauxois, J.-Y., and Gleizes, P.-E. (2019). The path of pre-ribosomes through the nuclear pore complex revealed by electron tomography. Nat. Commun. *10*, 497.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A., et al. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. Mol. Cell *70*, 854-867.e9.

Domínguez-Sánchez, M.S., Barroso, S., Gómez-González, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet. *7*, e1002386.

Eliseeva, I.A., Kim, E.R., Guryanov, S.G., Ovchinnikov, L.P., and Lyabin, D.N. (2011). Y-box-binding protein 1 (YB-1) and its functions. Biochemistry *76*, 1402–1433.

Evdokimova, V.M., Wei, C.L., Sitikov, A.S., Simonenko, P.N., Lazarev, O.A., Vasilenko, K.S., Ustinov, V.A., Hershey, J.W., and Ovchinnikov, L.P. (1995). The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. J. Biol. Chem. *270*, 3186–3192.

Fang, L.T. (2011). The end-to-end distance of RNA as a randomly self-paired polymer. J. Theor. Biol. *280*, 101–107.

Floor, S.N., and Doudna, J.A. (2016). Tunable protein synthesis by transcript isoforms in human cells. Elife *5*, e10921.

Gall, J.G. (1956). On the submicroscopic structure of chromosomes. Brookhaven Symp. Biol. 17–32.

Gamberi, C., Izaurralde, E., Beisel, C., and Mattaj, I.W. (1997). Interaction between the human nuclear cap-binding protein complex and hnRNP F. Mol. Cell. Biol. *17*, 2587–2597.

Gehring, N.H., Wahle, E., and Fischer, U. (2017). Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. Trends Biochem. Sci. *42*, 369–382.

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

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum. Genet. *135*, 851–867.

Gopal, A., Zhou, Z.H., Knobler, C.M., and Gelbart, W.M. (2012). Visualizing large RNA molecules in solution. RNA *18*, 284–299.

Gralla, J., and Crothers, D.M. (1973). Free energy of imperfect nucleic acid helices. II. Small hairpin loops. J. Mol. Biol. *73*, 497–511.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E.G., and Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell *115*, 739–750.

Grünwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature *467*, 604–607.

Guo, J., Lian, X., Zhong, J., Wang, T., and Zhang, G. (2015). Length-dependent translation initiation benefits the functional proteome of human cells. Mol. Biosyst. *11*, 370–378.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. J. Vis. Exp.

Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science *347*, 1367–1671.

Han, H.S.W., and Reidys, C.M. (2012). The 5'-3' distance of RNA secondary structures. J. Comput. Biol. *19*, 867–878.

Han, S., Zhao, B.S., Myers, S.A., Carr, S.A., He, C., and Ting, A.Y. (2020). RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. Proc. Natl. Acad. Sci. U. S. A. *117*, 22068–22079.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. Biochem. J *473*, 2911–2935.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol. *7*, e1000238.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nat. Rev. Mol. Cell Biol. *19*, 327–341.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol.

Hinde, E., Thammasiraphop, K., Duong, H.T.T., Yeow, J., Karagoz, B., Boyer, C., Gooding, J.J., and Gaus, K. (2017). Pair correlation microscopy reveals the role of nanoparticle shape in intracellular transport and site of drug release. Nat. Nanotechnol. *12*, 81–89.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. *83*, 779–812.

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A.R., and Ast, G. (2016). How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet. *32*, 596–606.

Howard, J.M., and Sanford, J.R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. Wiley Interdiscip. Rev. RNA *6*, 93–110.

Huang, M., Rech, J.E., Northington, S.J., Flicker, P.F., Mayeda, A., Krainer, A.R., and LeStourgeon, W.M. (1994). The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. Mol. Cell. Biol. *14*, 518–533.

Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO J. *17*, 7480–7489.

Jambor, H., Mueller, S., Bullock, S.L., and Ephrussi, A. (2014). A stem–loop structure directs oskar mRNA to microtubule minus ends. RNA *20*, 429–439.

Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K., and Krzyzosiak, W.J. (2016). Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. Nucleic Acids Res. *44*, 9050–9070.

Jean-Philippe, J., Paz, S., and Caputi, M. (2013). hnRNP A1: the Swiss army knife of gene expression. Int. J. Mol. Sci. *14*, 18999–19024.

Jia, J.-J., Lahr, R.M., Solgaard, M.T., Moraes, B.J., Pointet, R., Yang, A.-D., Celucci, G., Graber, T.E., Hoang, H.-D., Niklaus, M.R., et al. (2021). mTORC1 promotes TOP mRNA translation through site-specific phosphorylation of LARP1. Nucleic Acids Res.

Jolma, A., Zhang, J., Mondragón, E., Morgunova, E., Kivioja, T., Laverty, K.U., Yin, Y., Zhu, F., Bourenkov, G., Morris, Q., et al. (2020). Binding specificities of human RNA binding proteins towards structured and linear RNA sequences.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.-N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev. *19*, 104–113.

Kaida, D. (2016). The reciprocal regulation between splicing and 3′-end processing. Wiley Interdiscip. Rev. RNA *7*, 499–511.

Katz, Z.B., English, B.P., Lionnet, T., Yoon, Y.J., Monnier, N., Ovryn, B., Bathe, M., and Singer, R.H. (2016). Mapping translation "hot-spots" in live cells by tracking single molecules of mRNA and ribosomes. Elife *5*.

Khong, A., and Parker, R. (2018). mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. J. Cell Biol. *217*, 4124–4140.

Kierzek, E., and Kierzek, R. (2003). The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. Nucleic Acids Res. *31*, 4472–4480.

Koch, A., Aguilera, L., Morisaki, T., Munsky, B., and Stasevich, T.J. (2020). Quantifying the dynamics of IRES and cap translation with single-molecule resolution in live cells. Nat. Struct. Mol. Biol. *27*, 1095–1104.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. *17*, 909–915.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kretov, D.A., Clément, M.-J., Lambert, G., Durand, D., Lyabin, D.N., Bollot, G., Bauvais, C., Samsonova, A., Budkina, K., Maroun, R.C., et al. (2019). YB-1, an abundant core mRNA-binding protein, has the capacity to form an RNA nucleoprotein filament: a structural analysis. Nucleic Acids Res. *47*, 3127–3141.

Kudla, G., Wan, Y., and Helwak, A. (2020). RNA Conformation Capture by Proximity Ligation. Annu. Rev. Genomics Hum. Genet. *21*, 81–100.

Kuo, H.C., Nasim, F.H., and Grabowski, P.J. (1991). Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science *251*, 1045–1050.

Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2013). Determination of in vivo RNA structure in low-abundance transcripts. Nat. Commun. *4*, 2971.

Lai, D., Proctor, J.R., and Meyer, I.M. (2013). On the importance of cotranscriptional RNA structure formation. RNA *19*, 1461–1473.

Lai, W.-J.C., Kayedkhordeh, M., Cornell, E.V., Farah, E., Bellaousov, S., Rietmeijer, R., Salsi, E., Mathews, D.H., and Ermolenko, D.N. (2018). mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. Nat. Commun. *9*, 4328.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol. Cell *54*, 887–900.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. *19*, 6860–6869.

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

Leija-Martínez, N., Casas-Flores, S., Cadena-Nava, R.D., Roca, J.A., Mendez-Cabañas, J.A., Gomez, E., and Ruiz-Garcia, J. (2014). The separation between the 5'-3' ends in long RNA molecules is short and nearly constant. Nucleic Acids Res. *42*, 13963–13968.

Leppek, K., Das, R., and Barna, M. (2017). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol.

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

Li, S., Prasanna, X., Salo, V.T., Vattulainen, I., and Ikonen, E. (2019a). An efficient auxin-inducible degron system with low basal degradation in human cells. Nat. Methods *16*, 866–869.

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. Nat. Chem. Biol. *12*, 311–316.

Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R.C., Espinosa, S., Shi, S., Cui, Y., Kappel, K., Das, R., et al. (2019b). A unified mechanism for intron and exon definition and back-splicing. Nature *573*, 375–380.

Lin, D.H., and Hoelz, A. (2019). The Structure of the Nuclear Pore Complex (An Update). Annu. Rev. Biochem. *88*, 725–783.

Ling, J., Morley, S.J., Pain, V.M., Marzluff, W.F., and Gallie, D.R. (2002). The histone 3'-terminal stem-loop-binding protein enhances translation through a functional and physical interaction with eukaryotic initiation factor 4G (eIF4G) and eIF3. Mol. Cell. Biol. *22*, 7853–7867.

Liu, S.-R., Hu, C.-G., and Zhang, J.-Z. (2016). Regulatory effects of cotranscriptional RNA structure formation and transitions. Wiley Interdiscip. Rev. RNA *7*, 562–574.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem. J *417*, 15–27.

Lönnroth, A., Alexciev, K., Mehlin, H., Wurtz, T., Skoglund, U., and Daneholt, B. (1992). Demonstration of a 7-nm RNP fiber as the basic structural element in a premessenger RNP particle. Exp. Cell Res. *199*, 292–296.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279.

Lyabin, D.N., Eliseeva, I.A., and Ovchinnikov, L.P. (2014). YB-1 protein: functions and regulation. Wiley Interdiscip. Rev. RNA *5*, 95–110.

Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R., and Singh, G. (2018). The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. Cell Rep.

Mahen, E.M., Watson, P.Y., Cottrell, J.W., and Fedor, M.J. (2010). mRNA secondary structures fold sequentially but exchange rapidly in vivo. PLoS Biol. *8*, e1000307.

Maiuri, P., Knezevich, A., De Marco, A., Mazza, D., Kula, A., McNally, J.G., and Marcello, A. (2011). Fast transcription rates of RNA polymerase II in human cells. EMBO Rep. *12*, 1280–1285.

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep. *4*, 1144–1155.

Mateu-Regué, À., Christiansen, J., Bagger, F.O., Winther, O., Hellriegel, C., and Nielsen, F.C. (2019). Single mRNP Analysis Reveals that Small Cytoplasmic mRNP Granules Represent mRNA Singletons. Cell Rep. *29*, 736-748.e4.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical

insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Mol. Cell *54*, 751–765.

Matsumoto, K., and Bay, B.-H. (2005). Significance of the Y-box proteins in human cancers. J. Mol. Genet. Med. *1*, 11–17.

Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J., et al. (2019). mRNA structure regulates protein expression through changes in functional half-life. Proc. Natl. Acad. Sci. U. S. A. *116*, 24075–24083.

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. Neuron *92*, 1266–1278.

McAfee, J.G., Soltaninassab, S.R., Lindsay, M.E., and LeStourgeon, W.M. (1996). Proteins C1 and C2 of Heterogeneous Nuclear Ribonucleoprotein Complexes Bind RNA in a Highly Cooperative Fashion: Support for Their Contiguous Deposition on Pre-mRNA during Transcription. Biochemistry *35*, 1212–1222.

Mehlin, H., Daneholt, B., and Skoglund, U. (1992). Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. Cell *69*, 605–613.

Mehlin, H., Daneholt, B., and Skoglund, U. (1995). Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J. Cell Biol. *129*, 1205–1216.

Meinel, D.M., and Sträßer, K. (2015). Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in S. cerevisiae. Bioessays *37*, 666–677.

Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Söding, J., Holstege, F.C.P., and Sträßer, K. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. PLoS Genet. *9*, e1003914.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. Mol. Cell *72*, 715-726.e3.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635–1646.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat. Cell Biol. *12*, 543–552.

Morisaki, T., and Stasevich, T.J. (2018). Quantifying Single mRNA Translation Kinetics in Living Cells. Cold Spring Harb. Perspect. Biol. *10*.

Morisaki, T., Lyon, K., DeLuca, K.F., DeLuca, J.G., English, B.P., Zhang, Z., Lavis, L.D., Grimm, J.B., Viswanathan, S., Looger, L.L., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. Science *352*, 1425–1429.

Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev. *8*, 855–866.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat. Rev. Genet. *14*, 275–287.

Nachtergaele, S., and He, C. (2018). Chemical Modifications in the Life of an mRNA Transcript. Annu. Rev. Genet. *52*, 349–372.

Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M., and Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome Res. *26*, 24–35.

Nishimura, T., Padamsi, Z., Fakim, H., Milette, S., Dunham, W.H., Gingras, A.-C., and Fabian, M.R. (2015). The eIF4E-Binding Protein 4E-T Is a Component of the mRNA Decay Machinery that Bridges the 5′ and 3′ Termini of Target mRNAs. Cell Rep.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol. Cell *72*, 369-379.e4.

Oeffinger, M., and Zenklusen, D. (2012). To the pore and through the pore: a story of mRNA export kinetics. Biochim. Biophys. Acta *1819*, 494–506.

Okamoto, K., Hibino, K., and Sako, Y. (2020). In-cell single-molecule FRET measurements reveal three conformational state changes in RAF protein. Biochim. Biophys. Acta Gen. Subj. *1864*, 129358.

Okunola, H.L., and Krainer, A.R. (2009). Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol. Cell. Biol. *29*, 5620–5631.

Ozgur, S., Basquin, J., Kamenska, A., Filipowicz, W., Standart, N., and Conti, E. (2015). Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. Cell Rep. *13*, 703–711.

Paci, G., Zheng, T., Caria, J., Zilman, A., and Lemke, E.A. (2020). Molecular determinants of large cargo transport into the nucleus. Elife *9*.

Park, E.-H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. EMBO J. *30*, 302–316.

de la Parra, C., Ernlund, A., Alard, A., Ruggles, K., Ueberheide, B., and Schneider, R.J. (2018). A widespread alternate form of cap-dependent mRNA translation initiation. Nat. Commun. *9*, 3068.

Pichon, X., Bastide, A., Safieddine, A., Chouaib, R., Samacoits, A., Basyuk, E., Peter, M., Mueller, F., and Bertrand, E. (2016). Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. J. Cell Biol. *214*, 769–781.

Piñol-Roma, S., Choi, Y.D., Matunis, M.J., and Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. *2*, 215–227.

Pörschke, D. (1974). Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. Biophys. Chem. *1*, 381–386.

Portman, D.S., O'Connor, J.P., and Dreyfuss, G. (1997). YRA1, an essential Saccharomyces cerevisiae gene, encodes a novel nuclear protein with RNA annealing activity. RNA *3*, 527–537.

Presman, D.M., Ball, D.A., Paakinaho, V., Grimm, J.B., Lavis, L.D., Karpova, T.S., and Hager, G.L. (2017). Quantifying transcription factor binding dynamics at the single-molecule level in live cells. Methods *123*, 76–88.

Pühringer, T., Hohmann, U., Fin, L., Pacheco-Fiallos, B., Schellhaas, U., Brennecke, J., and Plaschka, C. (2020). Structure of the human core transcription-export complex reveals a hub for multivalent interactions. Elife *9*.

Puton, T., Kozlowski, L., Tuszynska, I., Rother, K., and Bujnicki, J.M. (2012). Computational methods for prediction of protein-RNA interactions. J. Struct. Biol. *179*, 261–268.

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

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. RNA Biol. *4*, 118–130.

Rech, J.E., Huang, M.H., LeStourgeon, W.M., and Flicker, P.F. (1995). An ultrastructural characterization of in vitro-assembled hnRNP C protein-RNA complexes. J. Struct. Biol. *114*, 84–92.

Rissland, O.S. (2017). The organization and regulation of mRNA-protein complexes. Wiley Interdiscip. Rev. RNA *8*.

Robert, F., Blanchette, M., Maes, O., Chabot, B., and Coulombe, B. (2002). A Human RNA Polymerase II-containing Complex Associated with Factors Necessary for Spliceosome Assembly*. J. Biol. Chem. *277*, 9302–9306.

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. *137*, 2107–2115.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Rouya, C., Siddiqui, N., Morita, M., Duchaine, T.F., Fabian, M.R., and Sonenberg, N. (2014). Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA *20*, 1398–1409.

Russell, R., Millett, I.S., Tate, M.W., Kwok, L.W., Nakatani, B., Gruner, S.M., Mochrie, S.G.J., Pande, V., Doniach, S., Herschlag, D., et al. (2002). Rapid compaction during RNA folding. Proc. Natl. Acad. Sci. U. S. A. *99*, 4266–4271.

Safaee, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J., and Gehring, K. (2012). Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. Mol. Cell *48*, 375–386.

Sakharkar, M.K., Chow, V.T.K., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol. *4*, 387–393.

Saldi, T., Riemondy, K., Erickson, B., and Bentley, D.L. (2021). Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. Mol. Cell *0*.

Samarina, O.P., Krichevskaya, A.A., and Georgiev, G.P. (1966). Nuclear Ribonucleoprotein Particles Containing Messenger Ribonucleic Acid. Nature *210*, 1319–1322.

Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R.M., and Tartaglia, G.G. (2019). RNA structure drives interaction with proteins. Nat. Commun. *10*, 3246.

Sapra, A.K., Ankö, M.-L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.-M., and Neugebauer, K.M. (2009). SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. Mol. Cell *34*, 179–190.

Saulière, J., Haque, N., Harms, S., Barbosa, I., Blanchette, M., and Le Hir, H. (2010). The exon junction complex differentially marks spliced junctions. Nat. Struct. Mol. Biol. *17*, 1269–1271.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat. Struct. Mol. Biol. *19*, 1124–1131.

Sharma, D., Zagore, L.L., Brister, M.M., Ye, X., Crespo-Hernández, C.E., Licatalosi, D.D., and Jankowsky, E. (2021). The kinetic landscape of an RNA-binding protein in cells. Nature.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. Mol. Cell *62*, 618–626.

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. *18*, 655–670.

Shi, M., Zhang, H., Wu, X., He, Z., Wang, L., Yin, S., Tian, B., Li, G., and Cheng, H. (2017). ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. Nucleic Acids Res. *45*, 9640–9653.

Si, J., Cui, J., Cheng, J., and Wu, R. (2015). Computational Prediction of RNA-Binding Proteins and Binding Sites. Int. J. Mol. Sci. *16*, 26303–26317.

Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. *17*, 407–421.

Siebrasse, J.P., Kaminski, T., and Kubitscheck, U. (2012). Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. Proc. Natl. Acad. Sci. U. S. A. *109*, 9426–9431.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat. Struct. Mol. Biol. *16*, 1128–1133.

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

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem. *84*, 325–354.

Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T.C., Rossmann, M.G., and Kuhn, R.J. (2016). The 3.8 Å resolution cryo-EM structure of Zika virus. Science *352*, 467–470.

Skabkin, M.A., Kiselyova, O.I., Chernov, K.G., Sorokin, A.V., Dubrovin, E.V., Yaminsky, I.V., Vasiliev, V.D., and Ovchinnikov, L.P. (2004). Structural organization of mRNA complexes with major core mRNP protein YB-1. Nucleic Acids Res. *32*, 5621–5635.

Skoglund, U., and Daneholt, B. (1986). Electron microscope tomography. Trends Biochem. Sci. *11*, 499–503.

Skoglund, U., Andersson, K., Björkroth, B., Lamb, M.M., and Daneholt, B. (1983). Visualization of the formation and transport of a specific hnRNP particle. Cell *34*, 847–855.

Solomon, O., Di Segni, A., Cesarkas, K., Porath, H.T., Marcu-Malina, V., Mizrahi, O., Stern-Ginossar, N., Kol, N., Farage-Barhom, S., Glick-Saar, E., et al. (2017). RNA editing by ADAR1 leads to context-dependent transcriptome-wide changes in RNA secondary structure. Nat. Commun. *8*, 1440.

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A.S., Nguyen, V.T., Hébert, C., Schmidt, U., Bertrand, E., Darzacq, X., and Bensaude, O. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. J. Cell Sci. *123*, 2085–2093.

Spitale, R.C., Flynn, R. a., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E. a., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G., et al. (2014). Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature *516*, 272–275.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., Ambrogio, A.D., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas ofmRNA secondary structures recognized by Staufen 1. Nature *519*, 491–494.

Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y., and Zhang, Q.C. (2019). RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol. *26*, 322–330.

Takamoto, K., Das, R., He, Q., Doniach, S., Brenowitz, M., Herschlag, D., and Chance, M.R. (2004). Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. J. Mol. Biol. *343*, 1195–1206.

Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. Cell *120*, 49–58.

Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C., and Burge, C.B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. Mol. Cell *64*, 294–306.

Tarun, S.Z., Jr, and Sachs, A.B. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. EMBO J. *15*, 7168–7177.

Tarun, S.Z., Jr, Wells, S.E., Deardorff, J.A., and Sachs, A.B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9046–9051.

Thompson, M.K., and Gilbert, W.V. (2017). mRNA length-sensing in eukaryotic translation: reconsidering the "closed loop" and its implications for translational control. Curr. Genet. *63*, 613–620.

Thompson, M.K., Rojas-Duran, M.F., Gangaramani, P., and Gilbert, W.V. (2016). The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. Elife *5*.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109–113.

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

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2018). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell.

Trzaskoma, P., Ruszczycki, B., Lee, B., Pels, K.K., Krawczyk, K., Bokota, G., Szczepankiewicz, A.A., Aaron, J., Walczak, A., Śliwińska, M.A., et al. (2020). Ultrastructural visualization of 3D chromatin folding using volume electron microscopy and DNA in situ hybridization. Nat. Commun. *11*, 2120.

Tuszynska, I., Matelska, D., Magnus, M., Chojnowski, G., Kasprzak, J.M., Kozlowski, L.P., Dunin-Horkawicz, S., and Bujnicki, J.M. (2014). Computational modeling of protein-RNA complex structures. Methods *65*, 310–319.

Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., et al. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. Nat. Commun. *10*, 990.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Van Nostrand, E.L., Gelboin-Burkhart, C., Wang, R., Pratt, G.A., Blue, S.M., and Yeo, G.W. (2017). CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. Methods *118–119*, 50–59.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711–719.

Van Treeck, B., and Parker, R. (2018). Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. Cell *174*, 791–802.

Vargas, D.Y., Raj, A., Marras, S. a. E., Kramer, F.R., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. Proc. Natl. Acad. Sci. U. S. A. *102*, 17008–17013.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Viero, G., Lunelli, L., Passerini, A., Bianchini, P., Gilbert, R.J., Bernabò, P., Tebaldi, T., Diaspro, A., Pederzolli, C., and Quattrone, A. (2015). Three distinct ribosome assemblies modulated by translation are the building blocks of polysomes. J. Cell Biol. *208*, 581–596.

Wakiyama, M., Imataka, H., and Sonenberg, N. (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. Curr. Biol. *10*, 1147–1150.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell.

Wang, C., Han, B., Zhou, R., and Zhuang, X. (2016). Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell *165*, 990–1001.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, J., Yang, Y., Yu, M., Hu, G., Gan, Y., Gao, H., and Shi, X. (2018). Diffusion of rod-like nanoparticles in non-adhesive and adhesive porous polymeric gels. J. Mech. Phys. Solids *112*, 431–457.

Wang, X., Zhu, L., Dang, M., Hu, Z., Gao, Q., Yuan, S., Sun, Y., Zhang, B., Ren, J., Kotecha, A., et al. (2017). Potent neutralization of hepatitis A virus reveals a receptor mimic mechanism and the receptor recognition site. Proc. Natl. Acad. Sci. U. S. A. *114*, 770–775.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. *35*, 169–178.

Wei, C.C., Balasta, M.L., Ren, J., and Goss, D.J. (1998). Wheat germ poly(A) binding protein enhances the binding affinity of eukaryotic initiation factor 4F and (iso)4F for cap analogues. Biochemistry *37*, 1910–1916.

Weighardt, F., Biamonti, G., and Riva, S. (1996). The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. Bioessays *18*, 747–756.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol. Cell *2*, 135–140.

Wende, W., Friedhoff, P., and Sträßer, K. (2019). Mechanism and Regulation of Co-transcriptional mRNP Assembly and Nuclear mRNA Export. In The Biology of MRNA: Structure and Function, M. Oeffinger, and D. Zenklusen, eds. (Cham: Springer International Publishing), pp. 1–31.

Wetterberg, I., Baurén, G., and Wieslander, L. (1996). The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. RNA *2*, 641–651.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip. Rev. RNA *8*.

Wu, B., Buxbaum, A.R., Katz, Z.B., Yoon, Y.J., and Singer, R.H. (2015). Quantifying Protein-mRNA Interactions in Single Live Cells. Cell *162*, 211–220.

Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. Science *352*, 1430–1435.

Wurtz, T., Lönnroth, A., and Daneholt, B. (1990). Higher order structure of Balbiani ring premessenger RNP particles depends on certain RNase A sensitive sites. J. Mol. Biol. *215*, 93–101.

Xu, B., Meng, Y., and Jin, Y. (2021). RNA structures in alternative splicing and back-splicing. Wiley Interdiscip. Rev. RNA *12*, e1626.

Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016). Dynamics of Translation of Single mRNA Molecules In Vivo. Cell *165*, 976–989.

Yang, M., Woolfenden, H.C., Zhang, Y., Fang, X., Liu, Q., Vigh, M.L., Cheema, J., Yang, X., Norris, M., Yu, S., et al. (2020). Intact RNA structurome reveals mRNA structure-mediated regulation of miRNA cleavage in vivo. Nucleic Acids Res. *48*, 8767–8781.

Yang, X., Yang, Y., Sun, B.-F., Chen, Y.-S., Xu, J.-W., Lai, W.-Y., Li, A., Wang, X., Bhattarai, D.P., Xiao, W., et al. (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. Cell Res. *27*, 606–625.

Yang, Y., Wang, L., Han, X., Yang, W.-L., Zhang, M., Ma, H.-L., Sun, B.-F., Li, A., Xia, J., Chen, J., et al. (2019). RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. Mol. Cell *75*, 1188-1202.e11.

Yoffe, A.M., Prinsen, P., Gopal, A., Knobler, C.M., Gelbart, W.M., and Ben-Shaul, A. (2008). Predicting the sizes of large RNA molecules. Proc. Natl. Acad. Sci. U. S. A. *105*, 16153–16158.

Yoffe, A.M., Prinsen, P., Gelbart, W.M., and Ben-Shaul, A. (2011). The ends of a large RNA molecule are necessarily close. Nucleic Acids Res. *39*, 292–299.

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N.M., and Ule, J. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell *152*, 453–466.

Zhang, S., Aibara, S., Vos, S.M., Agafonov, D.E., Lührmann, R., and Cramer, P. (2021). Structure of a transcribing RNA polymerase II–U1 snRNP complex. Science *371*, 305–309.

Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. *18*, 31–42.

Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol. Cell *8*, 1351–1361.

Ziv, O., Gabryelska, M.M., Lun, A.T.L., Gebert, L.F.R., Sheu-Gruttadauria, J., Meredith, L.W., Liu, Z.-Y., Kwok, C.K., Qin, C.-F., MacRae, I.J., et al. (2018). COMRADES determines in vivo RNA structures and interactions. Nat. Methods *15*, 785–788.