TYING THE CNOT: CNOT1 REGULATES POLY(A) TAIL LENGTH AT THE END AND THE BEGINNING

KATHRYN WILLIAMS, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

FEBRUARY 2021

Declaration

Except where acknowledged in the text, I declare that this thesis is my own work and is based on research that was undertaken by me in the School of Pharmacy, Faculty of Science, The University of Nottingham.

Acknowledgements

I thought this would be really easy compared to the part where I did a PhD, but it turns out a there are a hell of a lot of people who helped...

Firstly I would like to thank my supervisor, Dr Cornelia de Moor. It has been such a privilege to work with someone who is so fascinated by the world and who has such brilliant insight, but has not lost a drop of their humanity. In someone else's hands I might have quit, but Lia brought incredible balance and pushed me when I needed it, but always resolved my panic when things went wrong.

I would of course not have been able to undertake any of this research without the University of Nottingham or funding from the BBSRC. Several elements of the work presented here involved collaboration with others, in particular DeepSeq (University of Nottingham) and the Casanueva group (Babraham Institute), and this is refenced in text. Most of my immunofluorescence experiments did not make the final cut of this thesis, however, I would like to thank Hilary Collins for spending many hours with me in the Confocal room, training me on the microscope, taking a LOT of images, and generally having a great chat in the dark. I would also like to thank Dr Jonathan Wattis, my second supervisor, for undertaking the mammoth task of proofreading and commenting on an entire thesis which was outside of his main subject area at rather short notice...

My PhD would not have been the same without the wider Gene Regulation and RNA Biology lab. It's been a wonderful place to learn and make mistakes. In particular I'd like to thank:

- Barbara Rampersad and Trudi Gee who provided technical support for much of my PhD.
- Raj Gandhi, Jialiang Lin, Richa Singhania and Graeme Thorn for training me in various regards in the early days.
- Chris Roberts for the miscellaneous lessons I was only half paying attention to from the other side of the bench.
- Steven Lawrence for always having spare western blotting buffer.
- Poppy Winlow….just for being Poppy. (And for bringing us Attenborough.)

- Hannah Tomlin for *[!DANGER-DANGER!]* turning me upside down and inside out.
- Nicole-*Nicolino*-Zordan. I don't know what GRRB did to deserve her, but she makes the world a better place. She is so kind. (And I know that she would *never* put an eagle in the microwave, even if there were apple trees chasing her and it turned out she hadn't destroyed her thesis's last horcrux.)
- Stephen Hall, our new technical support who has been a ray of [slightly bizarre] sunshine since we returned after lockdown. I am sure it helped spur me on for the final push.

Circus Hub transformed my time in Nottingham and I only wish it had been around longer. Aerial has filled the gap the sea left and given me something else to love and to work hard for. Circus Hub would of course not be the same without its amazing community, which from a small survey, seems to be pretty unique even within the circus world. It has been so fun, and I will miss training there dearly. In particular, I would like to thank Kay Monrose for being so welcoming and encouraging and giving me so many opportunities, and for creating such a brilliant space with Ria. Under this canopy I would also like to give a 'shout out' to Emma Ward for hijacking the skill share chat so many times to ask me science questions. Her enthusiasm always made me smile, and getting to explain things to her often reminded me just how cool biology is.

Dana 'DAY-NAH' Baltmane has been one of the best things Circus Hub brought into my life. There are not that many people that you still like after they drop you on your head and punch you in the nose, but Dana is one of the few. I could always depend on her to make the evenings and weekends into fun breaks from work, and to go on weird adventures with (when we were allowed). In connection with Dana, I'd also like to thank Joe White and Kiki, other members of the wider Dana complex.

Aimée Parsons was to all intents and purposes, my lab big sister. Her generosity is astounding. I will be forever grateful for the many times she had me round for cat cuddles and dinner, and the many hours she spent listening when things were difficult. (I am also grateful for her Netflix account.)

Dylan Chegwin won't read this, but he helped me immensely at a time when I really wasn't ok, and without him I would not have finished my PhD.

It also seems right to thank Jere Tidey who encouraged me to ‘smell the roses’ and without whom I might not have chosen this path.

Harriet Lander requires something much more profound than I can come up with. She is one of the only people who ‘gets it’ and she has been a huge support throughout in more ways than I can count or remember.

For most of my PhD I did not know my neighbours, but I knew their cat, Leia. It is a little absurd to thank a cat with written words, but she always understood, better than most people.

Finally, I’d like to thank my parents for giving me a childhood so close to nature, for encouraging me to be amazed at the world around me, and for providing a supportive environment growing up. I have been so blessed to have them as a safety net. I am most thankful to them though for my sister, who has been there for the ups and downs, in life and in my PhD.

Abstract

Poly(A) tails affect multiple aspects of gene regulation: they help identify mRNAs for nuclear export, enhance translation efficiency, and are essential to regulating mRNA degradation rate. A poly(A) tail of 200-250 residues is thought to be uniformly added to newly synthesised mRNA and later gradually removed in the cytoplasm, allowing degradation of the mRNA itself.

Previous work in the lab showed that poly(A) tail addition is not uniform; soon after serum induction, transiently expressed mRNAs are exported with long poly(A) tails, but towards the end of the transcription pulse the tails of new transcripts are much shorter. In contrast, housekeeping mRNAs consistently receive only 30-70 adenosines both before and throughout the serum response and do not appear to be gradually deadenylated.

Given these controversial findings, the work presented here began by assessing the suitability of the PCR-based PAT assay for detecting differences in poly(A) tail length. This was achieved by comparing poly(A) length measurements obtained using the PAT assay with those using RNase H northern blots, which detected RNA directly. PAT assays of chromatin-associated, nucleoplasmic and cytoplasmic fractions then revealed that in NIH 3T3 cells, the poly(A) tail lengths of most mRNAs tested were determined before release from the chromatin. For the remainder of mRNAs, poly(A) tail lengths were determined in the nucleoplasm. Genome-wide analysis of poly(A) tails using adapted RNA-Seq (PQ-Seq) showed that in NIH 3T3 cells, nuclear regulation of polyadenylation was widespread and was not limited to the mRNAs previously selected for PAT. Short poly(A) tails were associated with reduced stability of transiently expressed transcripts, and it logically follows that nascent poly(A) regulation resulting in production of short-tailed transcripts at the end of the response may enhance the precision with which gene expression is controlled. Specifically, production of unstable transcripts at the end of the serum response would sharpen the peak in mature mRNA levels, limiting the time during which translation can occur.

Knockdown of the mRNA encoding the CCR4-NOT deadenylase subunit, CNOT1, increased chromatin and/or nucleoplasmic poly(A) tail size for all mRNAs tested, indicating it was involved in nuclear poly(A) tail regulation. Furthermore, the magnitudes of these changes were gene-specific. Preliminary data suggested that CCR4-NOT-dependent initial poly(A) regulation may also occur in human cells. As well as increasing initial poly(A) length (and therefore presumably enhancing transcript stability), Cnot1 knockdown caused decreases in pre-mRNA levels of all

mRNAs tested. Together, these data suggested that the CCR4-NOT complex may mediate mRNA homeostasis in mammalian cells.

Several lines of enquiry were followed to explore the mechanism through which CNOT1 both limited initial poly(A) length and seemingly promoted mRNA production. Although the exact mechanism remained elusive, differential effects of RNAi-mediated depletion versus pharmacological inhibition of the complex's CAF1 subunit suggested that the documented involvement of CCR4-NOT in both deadenylation and transcription elongation may have been significant.

The above findings complement work from other groups showing changes to CCR4-NOT subunit levels in different physiological conditions (e.g. nutrient deprivation, B cell activation). Specifically, levels of CCR4-NOT may be adjusted to simultaneously affect both mRNA production (through promoting transcription elongation) and nuclear determination of cytoplasmic mRNA stability (through promoting nuclear deadenylation). In this way, the mammalian CCR4-NOT complex may mediate high and low mRNA turnover states according to the state of the cell.

List of figures

List of tables

List of abbreviations

4E-T	eIF4E-Transporter
4SU	4-thiouridine
α-^{32}P-dCTP	deoxycitidine triphosphate, labelled with ^{32}P on the alpha phosphate group
AcD	Actinomycin D
Actb	Actin beta
APA	Alternative polyadenylation
APS	Ammonium persulfate
ARE	AU-rich element
Arl6ip5	ADP ribosylation factor like GTPase 6 interacting protein 5
β-ME	Beta mercaptoethanol
BPM	Bins per million mapped reads
BPS	Branch point sequence
BSA	Bovine serum albumin
BTG2	BTG anti-proliferation factor 2
CAF1	CCR4-associated factor 1
CBC	Cap binding complex
CBP	Cap-binding protein
CCR4-NOT	Carbon Catabolite Repression-Negative On TATA-less
CDK	Cyclin dependent kinase
cDNA	Complementary deoxyribonucleic acid
CFIm	Mammalian cleavage factor 1
CFIm25	Mammalian cleavage factor 1, 25 kDa subunit
CFIm68	Mammalian cleavage factor 1, 68 kDa subunit
CFIIm	Mammalian cleavage factor 2
CHTOP	Chromatin target of PRMT1
MYB	MYB proto-oncogene
CNOT1	CCR4-NOT transcription complex subunit 1
CPA	Cleavage and polyadenylation
CPE	Cytoplasmic polyadenylation element
CPEB	CPE-binding
CPSF	Cleavage and polyadenylation specificity factor
CRM1	Exportin 1
cryo-EM	Cryogenic electron microscopy
CstF	Cleavage stimulation factor
Ct	Cycle threshold
CTD	C-terminal domain
daf-21	90 kDa heat shock protein
DCP1	Decapping protein 1
DCP2	Decapping protein 2
DDX6	DEAD-box helicase 6
DEPC	Diethyl pyrocarbonate
Dhh1	DExD/H-box ATP-dependent RNA helicase DHH1
DIS3	DIS3 homolog, exosome endoribonuclease and 3'-5' exoribonuclease

Dld	Dihydrolipoamide dehydrogenase
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNase	Deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
DRB	5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSE	Downstream sequence element
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
Egr1	Early growth response 1
Egr2	Early growth response 2
EGTA	Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
eIF	Eukaryotic translation initiation factor
EJC	Exon junction complex
eNOS	Nitric oxide synthase 3
ERK	Extracellular signal-related kinase
exo-	Without exonuclease activity
EXOSC10	Exosome component 10
FAM46A	Family with sequence similarity 46 member A
FAM46C	Family with sequence similarity 46 member C
FBS	Foetal bovine serum
FDR	False discovery rate
FIP1	Factor interacting with PAPOLA and CPSF1
Fos	Fos proto-oncogene, AP-1 transcription factor
Fosb	FosB proto-oncogene, AP-1 transcription factor subunit
Gal4 DBD	Gal4 (galactose-responsive transcription factor GAL4) gene DNA binding domain
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
Gcn5	Histone acetyltransferase GCN5
GLD2	Terminal nucleotidyltransferase 2
GO	Gene ontology
GW182	GW protein family, containing multiple GW repeats in their N-terminal half
H3K27me3	Trimethylation at lysine 27 of the histone H3 protein
HELZ	Helicase with zinc finger
hnRNP-Q1	Heterogenous ribunucleoprotein Q, isoform 1
HO1	Heme oxygenase 1
HRP	Horseradish peroxidase
Hsp	Heat shock protein
HuR (ELAV1)	Human antigen R (ELAV like RNA binding protein 1)
IAV	Influenza A virus
IgM	Immunoglobulin M
IGV	Integrative Genomics Viewer
IRES	Internal ribosome entry site
Itga1	Integrin subunit alpha 1

Jhd2	Jumonji C domain-containing histone demethylase 2
JNK	c-Jun N-terminal kinase
KSRP	KH-type splicing regulatory protein
LARP1	La-related protein 1
LARP4	La-related protein 4
lncRNA	Long non-coding RNA
LPS	Lipopolysaccharide
LSM1-7	SM-like proteins 1-7
LUZP4	Leucine zipper protein 4
m^6A	N 6-methyladenosine
Malat1	Metastasis associated lung adenocarcinoma transcript 1
miRNA	micro RNA
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein
mTORC1	Mechanistic target of rapamycin complex 1
MYC	MYC proto-oncogene, bHLH transcription factor
Nab2	Nuclear polyadenosine RNA-binding 2
NCS	Newborn calf serum
NLS	Nuclear localization signal
NMD	Nonsense-mediated decay
Not3	Ccr4-Not core subunit NOT3 (yeast)
NPC	Nuclear pore complex
NPM1	Nucleophosmin 1
Nrg1	Negative regulator of glucose-repressed genes
NS1	Nonstructural protein 1
ntl-1	Not-like 1
NUDT21	Nudix (nucleoside diphosphate linked moiety X)-type motif 21
NXF1	Nuclear RNA export factor 1
NXT1	Nuclear transport factor 2 like export factor 1
oligo	Oligonucleotide
OPMD	Oculopharyngeal muscular dystrophy
P bodies	mRNA processing bodies
Pab1	Polyadenylate-binding protein
PABP	Poly(A) binding protein
PABPC	Cytoplasmic poly(A) binding protein
PABPN1	Nuclear poly(A) binding protein
PAN	Poly(A) nuclease
PAP	Poly(A) polymerase
PAPD	PAP associated domain containing
PAPOLA	Poly(A) polymerase alpha
PAPOLB	Poly(A) polymerase beta
PAPOLG	Poly(A) polymerase gamma
PARN	poly(A)-specific ribonuclease
PAS	Poly(A) signal
PAT	Poly(A) tail test
PAT1	DNA topoisomerase 2-associated proein PAT1

PBS	Phosphate buffered saline
PCF11	PCF11 cleavage and polyadenylation factor subunit
P:C:I	Phenol:chloroform:isoamyl
PCPA	Premature cleavage and polyadeylation
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PHF3	PHD finger protein 3
PI3K	Phosphoinositide 3-kinase
PIC	Pre-initiation complex
PLE	Poly(A) limiting element
PMSF	Phenylmethylsulfonyl fluoride
PNUTS	Phosphatase 1 nuclear targeting subunit
Pol I / II / III	RNA Polymerase I / II / III
PP1	Protein phosphatase 1
Ppia	Peptidylprolyl isomerase A
PROMPT	Promoter upstream transcripts
PVDF	Polyvinylidene fluoride
RBD	RNA binding domain
rf	relative front
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNase	Ribonuclease
RPAIN	RPA interacting protein
RPB1	DNA-directed RNA polymerase II subunit RPB1
Rpb	DNA-directed RNA polymerase II subunit
RPL10A	60S Ribosomal protein L10A
Rpl28	60S Ribosomal protein L28
RRM	RNA recognition motif
rRNA	Ribosomal RNA
Rrp6	Exosome complex exonuclease RRP6
Rsp5	NEDD4 family E3 ubiquitin-protein ligase
RT	Reverse transcription
SAGA	Spt-Ada-Gcn5-acetyltransferase (19 subunit complex)
SDS	Sodium docecyl sulfate
Ser 2/5/7 (p)	Serine 2/5/7 (phosphorylated)
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA/snRNP	small nuclear RNA / small nuclear ribonucleoprotein
SOX	Shutoff and exonuclease protein
Spt5	Suppressor of Ty 5 homolog
Sqstm1	Sequestosome 1
SRSF	Serine and arginine rich splicing factor
SS	Splice site
SS III	SuperScript III
SSC	Saline-sodium citrate
ssRNA	single-stranded RNA

STRE	Stress response element
SV40	Simian Virus 40
TAB182	Tankyrase 1-binding protein of 182 kDa
TBE	Tris-borate-EDTA buffer
TBST	Tris-buffered saline and Tween 20
TDP-43	TAR DNA binding protein 43
TE	Translation efficiency
TEC	Transcription elongation complex
TEMED	N,N,N',N'-Tetramethylethylenediamine
TENT	Terminal nucleotidyltransferase
TES	Transcription end site
TFIID	Transcription factor II D
TFIIH	Transcription factor II H
TFIIS	Transcription factor II S
THOC	THO complex
Thr4	Threonine 4
TOB	Transducer of ERBB2
TOP	Terminal oligopyrimidine tract
TORC1	Target of rapamycin complex 1
TPM	Transcripts per million mapped reads
TREX	Transcription and export [complex]
tRNA	transfer RNA
TSS	Transcription start site
TTP	Tristetraprolin
TUT1	Terminal uridylyl transferase 1
Tyr1	Tyrosine 1
U2AF	U2 auxiliary factor
U2AF65	U2 auxiliary factor 65 kDa subunit
UAP56	U2AF65-Associated protein 56
UIF	UAP56-interacting factor
USE	Upstream sequence element
UTR	Untranslated region
WDR33	WD repeat domain 33
WNK1	WNK Lysine Deficient Protein Kinase 1
WT	Wild type
XRN1	Exoribonuclease 1
XRN2	Exoribonuclease 2
YTHDF	YTH N6-methyladenosine RNA binding protein
ZC3H14	Zinc finger CCCH-type containing 14
ZCCHC14	Zinc finger CCHC-type containing 14
ZFC3H1	Zinc finger C3H1-type containing
ZFP36 (L1/L2)	Zinc finger protein 36 (like 1 / like 2)

1 Introduction

Correct gene expression relies on the orchestrated synthesis and degradation of mRNA, and on its successful translation. The eukaryotic messenger RNA poly(A) tail, present at the 3' end of the molecule, has roles in both mRNA stability and translation efficiency as well as permitting export of mRNA from the nucleus (1, 2). At the other end of the mRNA is a modified nucleotide, the 7-methylguanosine cap, which also serves a protective function, identifies the molecule as 'self' and interacts with nuclear export adapters (3–5). Changes in polyadenylation are crucial to several dynamic cell- and development–specific processes (6–14), and dysregulation of poly(A) site choice or deadenylation is well documented in different disease states (15–18). The highly variable phenotypes achieved by adjusting such a widespread modification indicate extensive regulation of a system which begs investigation.

The 3' poly(A) tail and 5' cap bind to cytoplasmic poly(A) binding proteins (PABPC) and the eIF4F complex respectively, which can interact via the eIF4G subunit and may take on a closed loop conformation (19, 20). This closed loop structure has been proposed to protect the mRNA from exonuclease activity and enhance translation efficiency (20–22). The poly(A) tail is gradually shortened by cytoplasmic deadenylases, dislodging PABPC and leaving the mRNA vulnerable to exonucleic degradation (23–28).

One such deadenylase, CCR4-NOT, boasts two catalytic subunits (exhibiting distinct kinetics) alongside several additional functional modules which mediate both message specific targeting of the complex, and its involvement in a number of other processes (28–51). The best-studied of these additional roles is in transcription, however various other putative functions seem to be emerging (34, 36–39, 41–44, 46). This coupling of deadenylation with other processes may indicate a central role for CCR4-NOT in mediating cross talk between different elements which contribute to correct gene expression.

Regulated length of the poly(A) tail, which accompanies the mRNA from the beginning of its lifetime and provides a temporary roadblock to degradation, is thus a prime candidate for connecting mRNA stability with transcription (52–58). Poly(A) tails are added by poly(A) polymerases and removed by deadenylases, both

machineries which can be regulated by environmental or internal stimuli to tune gene expression according to the cellular program (58–70). In this way the poly(A) tail can be considered a signal, the precise encoding and decoding of which remains to be elucidated.

1.1 Overview of messenger RNA synthesis and turnover

Messenger RNAs are synthesised in the nuclei of cells by RNA polymerase II (RNAP II/Pol II) which is first recruited into the pre-initiation complex (PIC). The PIC includes a set of general transcription factors (TFIIB, TFIID, TFIIE, TFIIF, TFIIH) (71, 72) as well as the multiprotein complex, Mediator, which recruits Pol II (73). Changes in Pol II phosphorylation state and consequent recruitment of elongation factors enables escape from the promoter, immediately followed by promoter-proximal pausing, which is likely required to facilitate capping of the nascent transcript (74). The elongation complex (EC) then proceeds along the template DNA strand. Upon reaching the poly(A) signal (PAS) at the 3' end of the gene, the nascent mRNA is cleaved and Pol II continues along the DNA until transcription is terminated in a linked but distinct process (75–77). An overview of this process is outlined in figure 1.1.

Before these newly synthesised mRNAs undergo translation they must be properly processed and exported from the nucleus. The main processing steps in mRNA biogenesis include 5' capping, splicing of introns and 3' cleavage and polyadenylation, all of which must be completed correctly to produce a stable and translatable mature mRNA. Splicing and 3' processing are triggered by the recognition of sequence elements. Both will be detailed further in due course, but it is appropriate to introduce the poly(A) signal (PAS) here as the extensive crosstalk between processes makes it almost impossible to discuss other elements of mRNA metabolism without describing it first. The PAS is a well conserved hexamer, the canonical form of which is AAUAAA. It sits 10-30 nt upstream of the cleavage site and is recognised by components of the cleavage and polyadenylation (CPA) complex (75, 78–80).

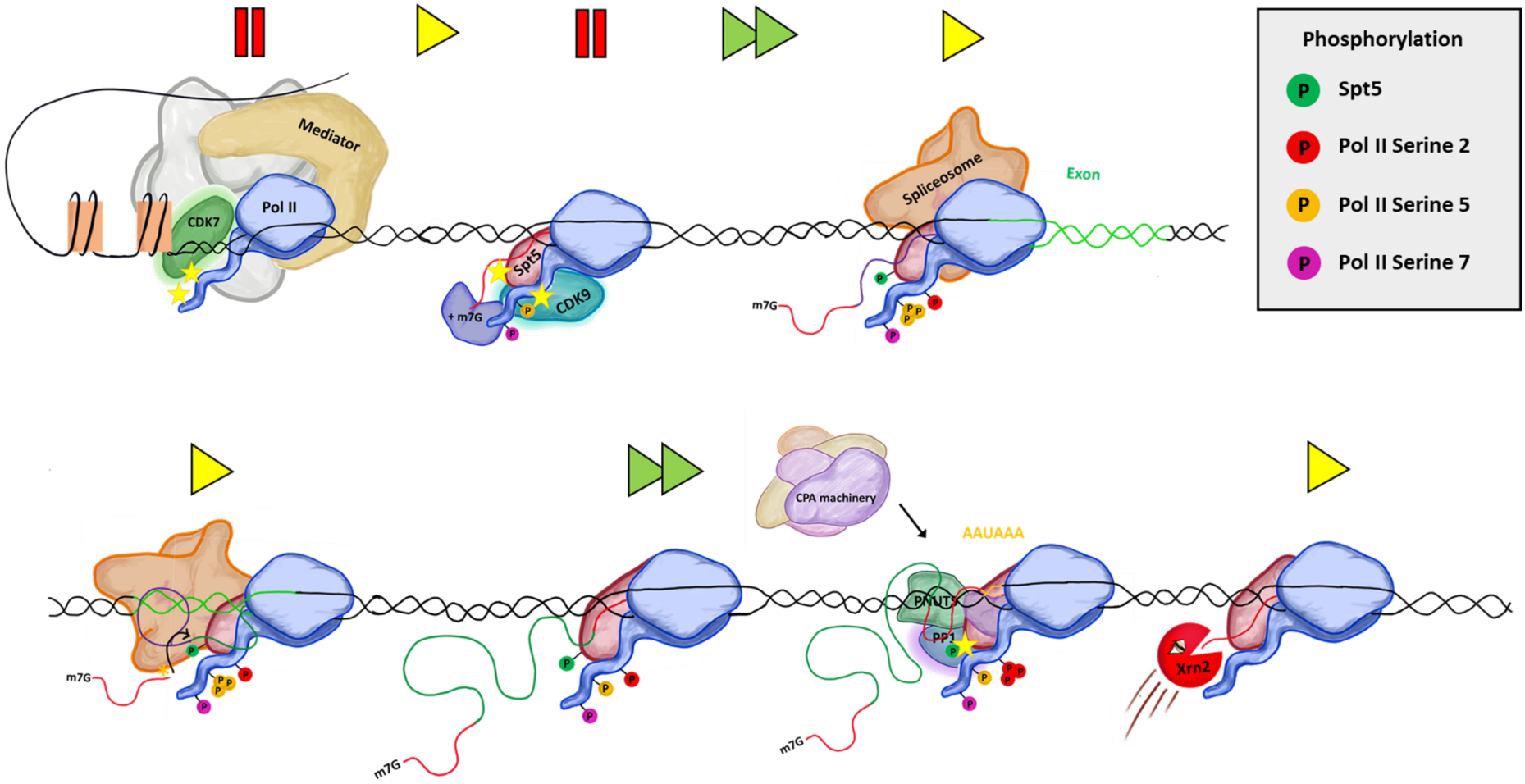


Figure 1.1 Overview of mRNA biogenesis and its coordination with RNA Pol II CTD phosphorylation. Symbols above indicate elongation rate.

Errors in pre-mRNA processing can lead to nuclear retention and degradation by either XRN2 (a 5' exonuclease), or the nuclear RNA exosome (a complex containing 3' exonuclease activity) (81–87). Protein-coding transcripts which are not identified by nuclear surveillance are exported to the cytoplasm where they undergo a further stage of quality control during the first round of translation, failure of which leads to decay of the mRNA, and thereby its removal from the pool of translatable transcripts (88–91). It has been suggested that the balance between nuclear and cytoplasmic mRNA quality control could differ between yeast and higher eukaryotes, and cytoplasmic quality control should not be inferred as a back-up system (92). In any case, correct processing of nascent transcripts is essential to efficient and tightly controlled gene expression.

The following section gives an overview of the various processes contributing to mRNA biogenesis, headed by an introduction the RNA polymerase II C terminal domain (Pol II CTD) which seems to play a central role in their coordination, and appended by a brief summary of the pathways available for mRNA decay.

1.1.1 RNA polymerase II C terminal domain

Eukaryotic cells contain three DNA-dependent RNA polymerases. While all protein-coding RNAs are transcribed by RNA polymerase II, Pol I is responsible for transcribing rRNA and Pol III for tRNA and 5S rRNA (93). Many non-coding RNAs are also transcribed by Pol II, but Pol I and III also contribute (94). Pol II is made up of 12 subunits, the largest of which is RBP1 (95). In addition to providing the polymerase's catalytic activity, RBP1 possesses a notable C-terminal domain (CTD) which has been and continues to be the subject of much curiosity. This interest stems from the changing modifications of the CTD over the course of transcription which help coordinate mRNA processing with transcriptional progress (96–99). This is outlined in figure 1.1.

The CTD is characterised by tandem repeats of the consensus sequence YSPTSPS. Although the sequence is consistent across eukaryotes, the number of these heptad repeats varies between organisms with 52 in mammals and 26 in yeast (73). Mammalian Pol II CTDs contain a mixture of consensus and non-consensus sequences. While the N terminal 26 repeats are similar to that of the yeast CTD and deviate only minimally from the consensus, the distal C terminal half is highly

enriched for non-consensus sequences (100). The 'CTD code' arises from its ability to be phosphorylated on five of the seven residues in each repeat, facilitating interactions of Pol II with factors required at each stage of the mRNA synthesis journey: from transcription initiation, through elongation to cleavage, and onto every major RNA processing step (73, 74, 96–99, 101–106). While this function may be mediated through direct interactions with other machinery (107), the CTD is intrinsically disordered and as such can participate in phase-separated nuclear condensates. These membraneless structures are formed through liquid-liquid phase separation and often include low complexity regions of proteins (108). Incorporation of the Pol II CTD into condensates may allow for co-localisation with the transcription initiation or RNA processing machinery to enhance efficiency (99, 109, 110).

Prior to incorporation into the transcription pre-initiation complex (PIC), the Pol II CTD is hypophosphorylated; indeed in yeast experiments, premature phosphorylation could inhibit its recruitment to the promoter (111). Upon incorporation, the CTD is phosphorylated on Ser5 and Ser7 by CDK7 which is a component of the multiprotein general transcription factor, TFIIH (112, 113). Inhibition of CDK7 caused a reduction in promoter-proximal pausing - an important step for coordination of elongation with capping of the nascent transcript in metazoans - for a majority of genes (74, 114).

In order for the polymerase to begin transcription elongation, CDK9 (and CDK12 in metazoans) phosphorylates the CTD on Ser2, a mark which it carries throughout elongation (73, 115, 116). During elongation, Ser5 is gradually dephosphorylated by specific phosphatases, while Ser7 phosphorylation dynamics are less clear (73, 117, 118). By the time Pol II approaches the 3' end, Ser5p phosphorylation is low, Ser7 may have been dephosphorylated and re-phosphorylated and Ser2p levels remain elevated, facilitating interaction with 3' processing and termination machinery (119).

Although Ser5 phosphorylation is considered a promotor-proximal feature, more recent work utilising mNET-Seq, in which Pol II-protected RNA fragments are sequenced, also observed Ser5p enrichment at 5' splice sites in HeLa cells, possibly indicating differences in phosphorylation pattern between yeast and mammals

(120). The authors suggest Pol II may pause here to allow the spliceosome to perform the first catalytic step, however this is in contrast to direct RNA measurements by Nanopore that indicate the majority of splicing in human cells occurs when the polymerase is at least 4 kb downstream of the 3' splice site (121). It is perhaps the case that the limited co-transcriptional splicing that does take place in human cells is aided by pausing of Ser5p Pol II, or that Ser5p may pause long enough to facilitate recruitment of the spliceosome, but that splicing is not completed until Pol II is much further downstream. It is of course also possible that the process differs between cell types, or that the experimental techniques gave rise to artefacts which were not accounted for. In particular, across many studies of this type it is assumed that the antibodies used to precipitate particular phosphorylation states of the CTD are accurate, however the binding specificity likely varies between antibodies and may also be influenced by the surrounding context of the modification.

Though the limelight has largely fallen on the serine residues, both Tyr1 and Thr4 are also available for phosphorylation. Like Ser2, Thr4 is phosphorylated by CDK9 (118). This phosphorylation has a documented role in recruitment of histone 3' processing machinery and in snoRNA transcription, as well as post-transcriptional mRNA splicing in yeast, though more T4p dependent processes may be revealed by further investigation (122–124). While substitution of Thr4 with valine is lethal in higher eukaryotes, yeast viability is not affected, though slower growth is observed (122, 123, 125). This difference may relate to yeast histone mRNAs being polyadenylated, unlike their higher eukaryotic counterparts where distinct 3' processing pathways are employed (126). Tyr1 phosphorylation seems to be important for RPB1 stability and promoter directionality and in yeast can also prevent premature transcription termination (127–129).

The Pol II CTD thereby acts as a platform through which correctly timed recruitment of processing factors to the nascent mRNA is orchestrated, and is instrumental in determining RNA production and fate.

1.1.2 Transcription initiation and elongation

Transcription is preceded by formation of the pre-initiation complex (PIC). This involves binding of the multisubunit complex, Mediator, to transcriptional activators

at upstream enhancer sequences and the concomitant assembly of general transcription factors on the core promoter. Mediator can then associate with the general transcription factors, uniting the PIC components and causing the chromatin to loop around (130). As part of this process, Mediator delivers unphosphorylated Pol II to the promoter, where its CTD is phosphorylated on Ser5 and Ser7 by the CDK7/CyclinH subunit of the general transcription factor TFIIH. Binding of Mediator to this subunit stimulates CDK7 activity, generating the Ser5,7p form of Pol II which can no longer be bound by Mediator and is therefore free to escape the promoter (73, 105).

In metazoans, 20-50 nt after clearing the promoter, Pol II undergoes promoter-proximal pausing, an important step for the protection of the emerging mRNA which is capped both to avoid degradation and to enhance pre-mRNA processing (3–5, 74, 98, 102, 131–133). Here, the Ser5p CTD, along with the elongation factor Spt5, are recognised and bound by RNA guanylyltransferase, an enzyme integral to 5' cap formation (74, 98).

Following release from the promoter-proximal pause into the elongation phase, Pol II can still encounter sites which induce backtracking, pausing or deceleration, for example at DNA lesions, splice sites, polyadenylation signals, or other regulatory sequences (58, 134–142). Pauses may in some cases lead to premature transcription termination (134, 136), and backtracking caused by DNA lesions can either be rescued, or result in ubiquitination of Pol II and presumably its degradation by the proteasome (138, 143–146). Rescue is mainly thought to involve the elongation factor, TFIIS, however the CCR4-NOT complex has also been implicated in yeast, suggesting another facet to its role in transcription (39, 143, 144, 147, 148). Elongation rate has been linked to mRNA stability in mammalian cells by affecting m^6A deposition, high levels of which may elicit recruitment of CCR4-NOT via the m^6A readers YTHDF1-3 and consequently, deadenylation (58). This points to a possible role for the CCR4-NOT complex in regulating both entry and exit of mRNAs from the system.

1.1.3 Capping

The 5' cap on Pol II-transcribed RNA is important for protecting the mRNA and marking it as 'self', as well as for promoting pre-mRNA processing and then export

(3–5, 133). The cap is connected to the first transcribed nucleotide by a 5'-5' triphosphate bridge (4). Function is initially mediated through interaction of the cap with the cap binding complex (CBC) a heterodimer of CBP80 and CBP20, which in general seems to be exchanged for the multiprotein translation initiation factor, eIF4F, in the cytoplasm after the first round of translation (149). In the case of mRNA, the methylation states of the first and second transcribed nucleotides also contribute to stability by affecting how readily the molecule is decapped (4).

Capping of nascent 5' ends of mammalian RNAs occurs immediately after transcription of the end, and is complete by the time 50 nt of nascent RNA have been transcribed (150, 151). The cap is added in stages (4). RNGTT (RNA guanylyltransferase and 5' phosphatase), possessing both triphosphatase and guanylyltransferase activities, first removes the terminal phosphate, creating a 5' terminal disphosphate to which guanosine is then added through hydrolysis of GTP to GMP + pp. Methylation of the resulting G(5')ppp(5')X is carried out by RNMT (RNA guanine-7 methyltransferase) which transfers a methyl group from the methyl donor, s-adenosyl methionine to the N-7 position of the terminal guanosine. In vertebrates, the miniprotein RAM (RNMT-activating miniprotein) stabilises RNMT and enhances binding of SAM to its active site.

1.1.4 Splicing

Introns are removed from the pre-mRNA by splicing in order to produce a continuous transcript from which to translate the correct protein product. Selective inclusion of some introns to generate different mRNA isoforms – known as alternative splicing - allows the production of multiple protein variants from the same gene.

Splicing relies on activity of the spliceosome which assembles on the RNA in a stepwise fashion. The location of assembly is defined by the 5' (donor, /GU) and 3' (acceptor, AG/) splice sites (SS), a branch point sequence (BPS, YUNAY) ~18-40 nt upstream of the 3' SS and in metazoans, a 12-17 nt polypyrimidine tract between the BPS and 3' SS (152). Donor and acceptor sites usually sit within wider consensus sequences, mutations in which can lead to a variety of abnormal transcripts, such as those missing all or part of an exon, or aberrantly including portions of introns through the use of normally unencountered cryptic splice sites (153). Splice site

usage can also vary in a more regulated manner to produce multiple isoforms of a gene with varying exon inclusion, a phenomenon termed alternative splicing (154).

Splicing is initiated through recognition of the donor site by U1 snRNP and the binding of SF1 (splicing factor 1) and U2AF65 to the BPS and polypyrimidine tract respectively (155). A series of steps follow, including recruitment of the U4, U5, U6 snRNP trimer and two ATP-dependent reactions (152). In brief, an A residue protruding from the branch point:U2 snRNA duplex carries out a nucleophilic attack on the 5' SS, forming a covalent attachment with the 5' end of the intron; this results in the lasso-shaped intron lariat. The now free 3' end of the 5' exon then attacks the 3' SS, connecting the two exons via a phosphodiester bond and releasing the intron lariat which is thought to usually be rapidly degraded (152, 156, 157).

Whether the 3' and 5' splice sites used reside on the same intron depends on organism complexity. The comparatively short introns native to yeast and lower complexity metazoa are removed by spliceosomes forming on splice sites of the same intron. Longer introns however, which contribute significantly to the increased overall gene lengths observed in humans (compared to yeast), are thought to be too long for intron definition and instead require interaction of the 3' and 5' splice sites at either end of the same exon (121, 152, 158–160). Consequently, in mammalian cells the 5' and 3' terminal introns require interaction with the capping and 3' processing machinery respectively in order to define the 3' and 5' splice sites respectively, thus linking the three major pre-mRNA processing steps (132, 133, 161–164).

1.1.5 3' processing and termination

The 3' ends of mRNAs -transcribed by highly Ser2 phosphorylated Pol II - are usually denoted by the poly(A) signal (PAS), which is located ~10-30 nt upstream of the cleavage site (119, 165, 166). The PAS and cleavage sites are nestled between several upstream and downstream sequence elements which are recognised, along with the PAS, by well-characterised multimeric complexes (78, 167–176). The components of these complexes vary slightly between yeast and higher eukaryotes. Of particular importance in mammals is the CPSF complex which contains two subunits able to recognise the PAS (CPSF4 and WDR33), the enzyme responsible for nascent mRNA cleavage (CPSF3), and another subunit (FIP1) which can both

recognise upstream sequences and recruit poly(A) polymerase (PAP) which catalyses addition of the poly(A) tail (55, 78, 175, 177–180). A more in-depth appraisal of cleavage and polyadenylation can be found in section 1.2.

1.1.5.1 Transcription termination

Tightly linked to cleavage of the nascent mRNA is termination of the transcription elongation complex (TEC) itself. Following cleavage, the TEC (containing Ser2 phosphorylated Pol II) continues transcribing along the DNA template until the unprotected RNA is degraded by XRN2 and the complex disengages (181). The order of these events has been described by two opposing putative models: allosteric and torpedo, both of which depend - either directly or indirectly - on the presence of a poly(A) signal. In the allosteric model, transcription termination signals cause a conformational change in the TEC which leads to dissociation (182). In the torpedo model, cleavage of the nascent mRNA exposes the 5' end of the remaining RNA, providing a substrate for the 5'-3' exonuclease XRN2 (Rat1 in yeast) which degrades the remaining uncapped RNA and dislodges the TEC (76).

Evidence for both models exists in the literature in yeast as well as higher eukaryotes. In favour of the allosteric model, Rat1 (yeast XRN2 homologue) was insufficient to terminate Pol II *in vitro*, and in Xenopus oocyte nuclei, terminating complexes were observed by electron microscopy (EM) on uncleaved transcription products of a plasmid vector (183, 184). In the latter study, terminating complexes were also observed on cleaved transcripts if a strong PAS was present, suggesting that the torpedo model may be relevant for efficiently cleaved transcripts (184). This is consistent with subsequent work by the same group showing that the majority – but not all – of over 100 Pol II transcribed Drosophila genes were terminated prior to cleavage (185).

Supporting the torpedo model, or at least a variant thereof, work in HeLa cells found that XRN2 activity was sufficient to terminate transcription of the human β-globin gene; however, the nuclease was found to be loaded onto the RNA downstream of the cleavage site, following autocatalysis of the RNA at its co-transcriptional cleavage (CoTC) site (186). A key principal of the torpedo model is the ability of XRN2 to catch up with elongating Pol II in order to dislodge it. In addition to checking the effects of human XRN2 nuclease inactivation, Fong et al therefore also

examined the impact of Pol II elongation rate on distance of termination site from the PAS (187). In the case of the dominant negative XRN2 mutant, termination was delayed substantially and the use of a slow or fast Pol II mutant resulted in more proximal or distal termination sites respectively. While both findings are consistent with the torpedo model, one could posit that allosteric inhibition may also be more efficient on an already slowly elongating polymerase. This pleasantly intuitive relationship between Pol II elongation rate and termination efficiency has been suggested elsewhere, this time induced by changes to CTD phosphorylation (114). Further exemplifying this relationship is the finding that CDK9, which phosphorylates the Pol II CTD to enable release of paused Pol II into the elongation phase, also phosphorylates XRN2 to enhance its enzymatic activity and recruitment to chromatin (188).

Further supporting a dependence of termination on cleavage are the observations of readthrough transcripts following CPSF3 depletion (and rescue with a catalytically inactive form) or infection with Influenza A virus (IAV), whose NS1 protein binds CPSF4 to prevent recognition of the PAS (77, 189). Of course, IAV can interfere will cellular processes at many levels, and given its notoriety for cap-snatching at the 5' end, additional effects on Pol II elongation do not seem implausible (190). Indeed, an IAV strain which could not bind CPSF4 also caused a moderate termination defect, and expression of the CPSF4-binding NS1 protein alone only seemed able to impair termination of a subset of genes; taken together – and assuming that CPSF4 is essential to cleavage - these findings suggest that cleavage-dependent termination (i.e. the torpedo model) is likely important for some genes, but may not universally applicable to polyadenylated Pol II transcripts (191). These conflicting findings may of course also indicate a difference in dominant mechanism between species, tissues or cellular states. Despite their differences, these studies agree on the necessity of the poly(A) signal for 3' end transcription termination.

Models unifying the two theories have recently been presented, based on experiments in human cell lines, in which encountering the poly(A) signal triggers dephosphorylation of the elongation factor Spt5 by PNUTS/PP1, reducing its stimulation of Pol II. This reduced stimulation of Pol II is thought to slow down the TEC sufficiently for XRN2 to catch up with and dislodge it (77, 142). The exact

mechanism though which encountering the PAS triggers Spt5 dephosphorylation has not been elucidated, but expression of IAV NS1 prevented the PAS-triggered slowdown of Pol II observed in untreated cells, confirming requirement of CPSF4 in triggering Spt5 dephosphorylation (142). Since PNUTS/PP1 can associate with the 3' processing machinery, it is proposed that its phosphatase activity is influenced by conformational changes occurring upon binding of CPSF to the PAS (142, 192).

Another potentially important contributor to mRNA transcription termination is the cleavage factor II (CFIIm, mammals)/cleavage factor CFIA (yeast) component, PCF11, which preferentially binds Ser2 phosphorylated Pol II (193, 194). The remit of PCF11 seems to be in control of the TEC-RNA interaction, with studies in yeast and mammals returning slightly different models (106, 195). *In vitro*, the yeast Pcf11 CID (CTD-interacting domain) is sufficient to dismantle the TEC by dissociating from the template DNA whilst bound to both the nascent mRNA and the Pol II CTD (106). Indeed, work by the same group in Drosophila found that depletion of dPcf11 caused increased readthrough transcription and that it again acts to terminate transcription through some undetermined mechanism involving formation of a bridge between the CTD and the RNA (196).

In vertebrates, phosphorylation of PCF11 by the serine-threonine kinase, WNK1 is required for its dissociation from Pol II and reduction in this phosphorylation caused retention of MYC mRNA on the chromatin (195). In this way, PCF11 is integral in linking transcription termination with export. More recent work indicates that dependence on PCF11 for transcription termination is variable across the transcriptome and seems to be more important for closely spaced genes (197).

1.1.6 Export

Export of mRNAs to the cytoplasm requires transit through the nuclear pore complex (NPC), which forms a channel across the nuclear envelope enabling passage of macromolecules. The NPC is composed of 3 main regions: the nuclear basket, cytoplasmic fibrils, and a central channel lined with nucleoporins (198). Since RNAs are unable to directly interact with the NPC, transit through it requires packaging into RNPs with adapter proteins, a requirement which in theory comprises an additional layer of regulation and quality control. Three main pathways exist for transit of RNAs out of the nucleus: NXF1/NXT1, CRM1 (Exportin) and Exportin-5. In

higher eukaryotes, miRNAs, rRNA and snRNA are preferentially exported by Exportin-5 and CRM1, whereas NXF1/NXT1 is thought to be the major pathway for export of mRNA (199). A subset of mRNPs are too large to fit through the channel and instead exit the nucleus via budding of the nuclear envelope (200).

An elegant solution to efficiently transporting newly made mRNA to the pore while filtering out those which are improperly processed is the dependence of export factor recruitment on conserved indicators of competent mRNAs such as the cap, poly(A) tail, and splice junctions (2, 5, 201–203). These indicators are interpreted through the various protein complexes which bind them, most prominently the cap binding complex (CBC), and exon junction complexes (EJCs), which are deposited after splicing 20-24 nt upstream of exon-exon junctions (5, 201, 204).

1.1.6.1 Nxf1/Nxt1

The prevailing model for NXF1/NXT1 – mediated export of mRNA is that the RNA is bound co-transcriptionally by components of the TREX (transcription and export) complex. TREX then interacts with NXF1 – which usually exists in a closed conformation with its RBD hidden to avoid non-selective RNA export – and elicits a change in conformation that allows binding of NXF1 to the RNA (199, 205). Once RNA is bound, NXF1 interacts with nucleoporins to mediate transport though the NPC (206).

Mammalian TREX includes the hexameric THO subcomplex along with additional proteins including the helicase UAP56 (DDX39B) and adapters and co-adapters which mediate interaction with NXF1, a role for which Aly/REF (yeast Yra1) has been best studied (207–212). These can lead to different variations of TREX (207). Two adapter proteins, UIF and LUZP4 have been identified as able to at least partially compensate for Aly/REF depletion and to interact with both NXF1 and the essential TREX component, UAP56 whose ATPase activity seems important for the complex's assembly (203, 211–215). While knockdown of UIF did not affect bulk poly(A)+ localisation, combined knockdown with REF caused substantially greater nuclear accumulation than REF knockdown alone (211). Interestingly, although binding of NXF1 by REF and UIF was mutually exclusive, the two proteins could interact in an RNA – dependent manner, suggesting that multiple TREX complexes can be loaded onto a single transcript. LUZP4 is normally only expressed in the testes but can be

upregulated in cancers, perhaps to enhance the export of a subset of transcripts or to compensate for reduced Aly/REF expression (212).

The mammalian THO subcomplex comprises six subunits: THOC 1-3 and THOC5-7, with THOC2 proposed to act as a scaffold (207). Although theoretically essential, depletion of THO complex subunits does not uniformly affect all transcripts and cell types in higher eukaryotes. Depletion of THOC1 for example causes apoptosis in cancer cells, but is permissible for growth of normal fibroblasts (216). Similarly, THOC5 seems to be required during embryogenesis and differentiation, but its depletion has little effect in differentiated cells (217). These inconsistent effects may be due to the particular transcripts which are reliant on THO for their export; indeed, while THO was required for export of heat shock mRNAs, it was not required - unlike ALY and TAP (NXF1) – for bulk mRNA export (218, 219). Interestingly, the reliance of Hsp70 mRNA on THOC5 for export in a mouse embryo fibroblast cell line is temperature dependent, with export succeeding in the absence of THOC5 at 37 °C but not 42 °C (220). The same group also identified a requirement for THOC5 in the expression and export of differentiation- and migration-related mRNAs (221). In keeping with a role for THO in dynamically changing systems, missense variants in THOC2 have been implicated in neurocognitive and growth disorders and intellectual disability (222, 223).

The THO subcomplex is recruited early in mRNA synthesis to mediate packaging of the nascent mRNA into an mRNP and thereby avoid generation of R-loops (through wayward interaction with the unwound DNA), which threaten genome integrity (224, 225). Whereas THO recruitment is directly linked to transcription in yeast, in human cells it relies on splicing; this is perhaps due to varying genomic complexity and architecture requiring different systems to sort their transcripts (226, 227). One key link between TREX assembly and splicing is common involvement of UAP56 which is required for early spliceosome formation as well as being an integral TREX component (209, 213, 215, 227–230). Direct interaction of ALY with the CBC has also been well documented, though for mRNAs which are spliced, this appears to be a transient step prior to EJC binding (2, 5). Inhibition of polyadenylation does not affect distribution of ALY along the mRNA and TREX recruitment is therefore considered to be co-transcriptional; however, ALY can interact with PABPN1 and has

been proposed to compete with ZFC3H1 for PABPN1 binding, resulting in rescue of the mRNA from nuclear degradation and successful export (2, 231). Another poly(A) binding protein, ZC3H14 - which despite its ubiquitous expression is particularly important for higher order brain function – interacts with THO, coupling processing and export for specific mRNAs (232). As well as recognising features of processed mRNA, some TREX complexes associate specifically with m^6A-modified transcripts to mediate their export (233, 234).

Interestingly, export factors can themselves affect mRNA processing. Aly/REF for example can influence splicing outcomes on poorly spliced introns (5). The same study found that deposition of the co-adapter, CHTOP is enriched towards the 3' ends of mRNAs and promotes use of distal poly(A) signals, resulting in mRNAs with longer 3'UTRs. THOC5 can also affect choice of 3'UTR isoform by interacting with the CFIm68 subunit of mammalian cleavage factor I, with THOC5 depletion specifically reducing the level of mRNAs polyadenylated at the distal PAS (235).

While a stepwise pathway involving TREX and then NXF1 seems rational, recent work suggests that although there is considerable overlap, not all TREX substrates may depend on NXF1 for their export (236). In particular, transcripts with few exons which are long or have a high A/U content showed greater dependence on NXF1. Depletion of TREX components disproportionately reduced export of spliced and G/C rich mRNAs, and both single and multi-exon transcripts were affected (236). Furthermore, in addition to recruitment by TREX in the canonical pathway, NXF1 can interact with certain SR proteins, a family initially characterised as splicing factors (237, 238). SRSF3 and SRFS7 can bind to NXF1 as well as the last exon of mRNA, and have opposite effects on 3'UTR length, thereby coupling alternative polyadenylation with export (237). The constitutive transport element (CTE) of type D retroviral RNA also directly interacts with NFX1 to circumvent splicing and allow export of its unspliced mRNA, avoiding other host machinery (239).

1.1.6.2 Crm1 (Exportin)

While NXF1 is important for bulk mRNA export, another NPC-interacting protein, CRM1 may provide an alternative pathway for mRNAs which are engaged with specific proteins through sequence and structural elements in their 3' UTRs (240–244). 3'UTR AREs for example can by bound by HuR (ELAV1), which as well as

stabilising transcripts, can mediate export of the bound mRNA through interactions with CRM1-interacting proteins (245–247). CRM1-dependent export of these mRNAs is thought to depend on stress conditions, with HuR and CRM1 only coimmunoprecipitating after heat shock (248). AREs are not the only element to promote export via CRM1; the eIF4E sensitivity element (4E-SE) is a structural element in the 3'UTR of Cyclin D and other cell cycle progression mRNAs which also promotes export via CRM1 (243, 244). This reliance on CRM1 is mediated by eIF4E and depends on its binding of both the 4E-SE and the cap. Notably, those mRNAs whose nuclear export is promoted by eIF4E are distinct from the mRNAs whose translation is eIF4E-dependent, indicating differing nuclear and cytoplasmic roles and targets (244).

1.1.7 Degradation

1.1.7.1 Cytoplasmic mRNA degradation

To maintain dynamic and agile cellular control, mRNA cannot persist indefinitely and must therefore be degraded. The average lifetime of an mRNA varies from minutes to days (249, 250), and stability depends on a variety of factors, including the cap and poly(A) tail, sequence elements and bound proteins, secondary structure, base modifications, and presence of coding defects which may be detected during translation and lead to premature decay (3, 19, 32, 45, 58, 90, 91, 251–271).

Translation-coupled mRNA decay can occur via three pathways, depending on the nature of the defect. Presence of a premature stop codon results in nonsense-mediated decay (NMD), while absence of a stop codon or stalling of the ribosome (eg. at highly structured regions) cause degradation by non-stop decay or no-go decay respectively (90, 91, 266–268). Should the mRNA survive translation, it persists until it undergoes canonical degradation, which can be targeted or non-specific (25, 26, 45, 254–256, 258, 269). Degradation may not always be immediate, with some transcripts being stored in the cytoplasm whilst maintaining translational silence (8, 272–274).

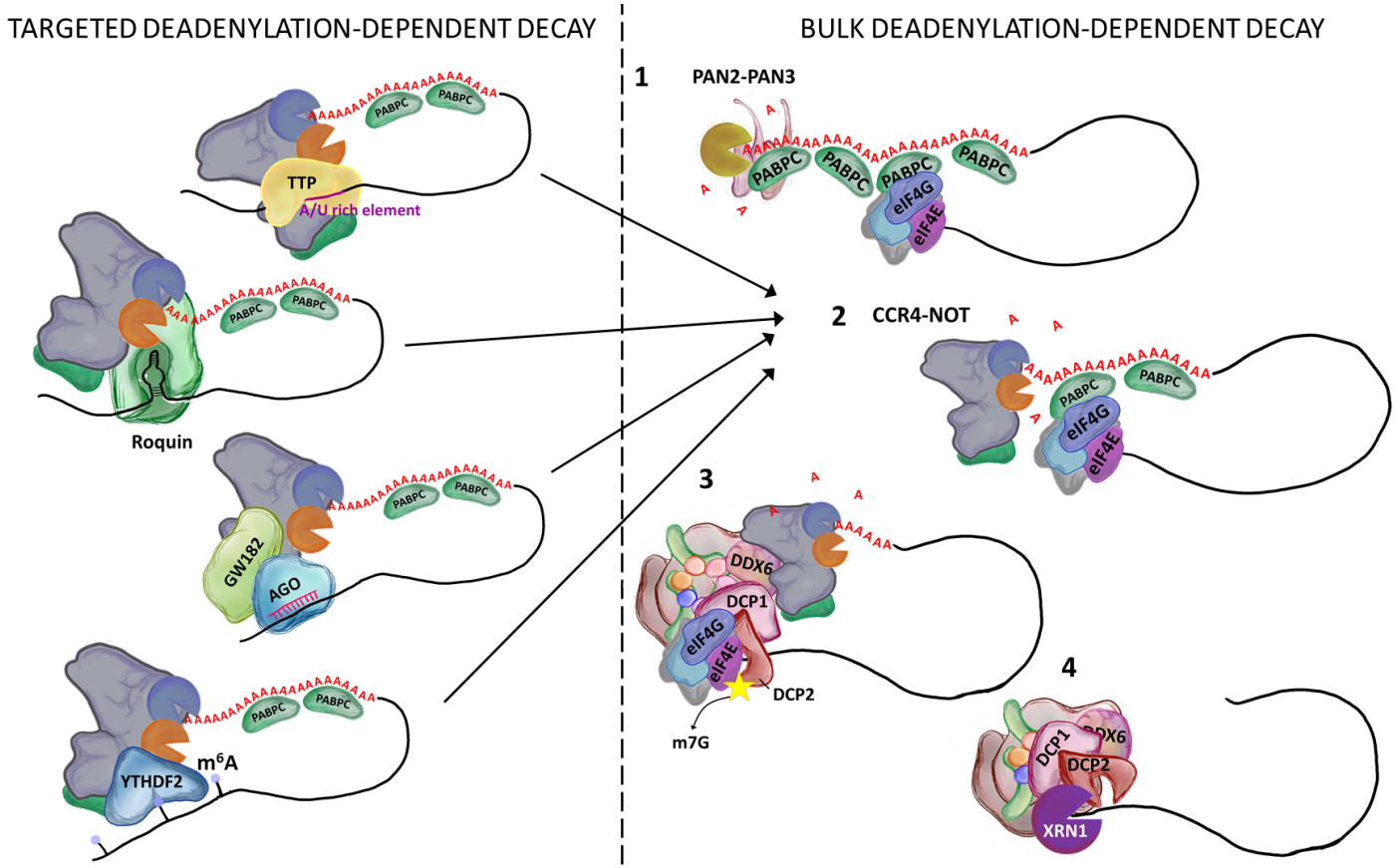


Figure 1.2 Overview of deadenylation-dependent decay. **Left:** examples of specific recruitment of CCR4-NOT. **Right:** Biphasic model of deadenylation-dependent decay. **1)** PABPC-stimulated PAN2-PAN3 trims long poly(A) tails. Adjacent PABPC binding may promote recognition by PAN2-PAN3 by arranging the poly(A) tail into repeated arches. **2)** CCR4-NOT trims shorter tails or tails with lower PABPC density. **3)** Decapping machinery can be recruited once PABPC-eIF4G interaction is disrupted. **4)** XRN1 associates with decapping machinery and elicits 5'-3' degradation starting with the newly exposed 5' monophosphate.

Based on synergistic translation activation by the cap and poly(A) tail, and later on interactions between PABPC and the eIF4F complex, mature mRNAs are thought to exist in a closed loop conformation which enhances their stability and translatability (19–21). More recent work has challenged the prevalence of closed loop formation *in vivo* (22, 275). In canonical degradation, the PABPC-coated poly(A) tail is degraded by cytoplasmic deadenylases in a process suggested to involve two sequential phases (i.e. PAN2-PAN3 then CCR4-NOT) (26, 50). Once the tail is too short to bind a single PABPC (< ~27 nt), the interaction between the 5' and 3' ends is interrupted, leaving the 3' end vulnerable to decay, and the 5' end to decapping by DCP1-DCP2 (26, 28, 50, 276–279). A schematic of deadenylation-dependent decay is provided in figure 1.2.

Although the 3' end theoretically becomes available to the cytoplasmic RNA exosome, the 5'-3' exonuclease XRN1 is thought to contribute more to cytoplasmic mRNA degradation, with 3' decay by the exosome being implicated for a few individual mRNAs (90, 253, 280, 281). This bias is initially surprising, since by this model the 3' end is exposed before the 5' end. It seems however, that the deadenylation machinery may promote recruitment of the capping machinery and its activators such that there is a minimal lag in cap removal once the 3' end is exposed (256, 282–286). Degradation by XRN1 is in turn coupled with cap removal via interactions with components of the decapping complex (282, 287). 5' → 3' degradation is likely dominant because the alternative option - degradation from the 3' end of the mRNA - could still allow translation from the 5' end and lead to the consequent generation of incomplete proteins.

Messenger RNAs can be specifically targeted for deadenylation-dependent degradation in the cytoplasm by several means; in particular through recognition of sequence elements by certain proteins or miRNAs (32, 45, 254–256, 258, 288, 289). Of particular note is the recognition of AU-rich elements (AREs) in 3' UTRs by Tristetraprolin (TTP) which then recruits CCR4-NOT to promote transcript degradation (51, 290–292). This is best characterised in inflammatory mRNAs, however targeting of AREs in the mRNAs of other tightly controlled genes is also well-documented (288, 293–296). TTP can shuttle between the nucleus and cytoplasm (297). Given the importance of mRNA stability in determining the timing

of the inflammatory - and perhaps other - response(s) (298, 299), it is interesting to consider the possibility that TTP – and even CCR4-NOT – could be recruited in the nucleus and behave as a timer in the cytoplasm to safeguard against sustained gene expression.

A relationship between translation efficiency and mRNA stability (outside the detection of aberrant transcripts) has been reported on several occasions (28, 44, 300–303). More recently, the mechanisms underlying these links have begun to be revealed, predominantly through studies in yeast (28, 44, 301). For an overview of various steps in translation, see review by Andreev et al (304). Transcripts with low codon optimality exhibit accelerated degradation which is Dhh1-dependent (Dhh1 being the yeast orthologue of DDX6) (301). Since DDX6 can bind both the ribosome and the deadenylation and decapping machinery (256, 283, 305), it was suggested that tRNAs compete with DDX6 for ribosome binding and that less frequent displacement of DDX6 by tRNAs leaves enough time for the degradation machinery to associate with the translationally-engaged transcript (306). Such a model is supported by the findings that Ccr4-Not can associate with polysomes and that association of DDX6 with CCR4-NOT enhances activity of the human CNOT7 (CAF1) nuclease subunit (305, 307). Furthermore, the same subunit was recently suggested to preferentially deadenylate poorly translated mRNAs (28). Recently, the model for this relationship was updated in yeast to involve direct association of Ccr4-Not with the ribosome when the A site lacks a tRNA (44). This is not due to direct competition between Ccr4-Not and tRNA but rather between Ccr4-Not and eIF5A (which rescues translation in the case of slow peptidyl transfer). While Not5 and eIF5A can both bind to the ribosome E site if it does not contain a tRNA, eIF5A is only able to interact with the E site if the A site is occupied (since this elicits a permissive conformational change in the ribosome). In the case of codon non-optimal transcripts, there is a higher likelihood of simultaneously vacant A and E sites, allowing Ccr4-Not (but not eIF5A) binding, and promoting transcript degradation (44). Importantly, many of the documented direct links between yeast Ccr4-Not and the ribosome rely on portions of Ccr4-Not which are less well conserved between yeast and mammals (44, 307, 308).

Cytoplasmic mRNA stability can also vary according to additional factors such as base modifications and structural features, but the wider significance of these is still being evaluated (47, 58, 260–265).

Deadenylation is covered in more depth in section 1.4 and the relationship between the poly(A) tail and mRNA stability in section 1.3.5.

1.1.7.2 Nuclear RNA turnover

RNA degradation is not a phenomenon isolated to the cytoplasm; indeed, nuclear mRNA decay is central to transcription termination and more recently has been shown to tune gene expression and counteract pervasive transcription (85, 92, 181, 186, 188, 189, 309–315). Decay of nuclear RNA can be performed by the 5' – 3' exonuclease, XRN2 – known for its role in transcription termination, or by the nuclear RNA exosome which in mammals includes the 3' – 5' nucleases EXOSC10 and DIS3, the latter of which also possesses endonuclease activity (92, 181, 186, 310, 316, 317). While some nuclear decay is directed towards noncoding transcripts, widespread degradation of transcripts from protein-coding promoters also occurs, largely mediated by the nuclear exosome (85, 92, 309–311, 318, 319). Exosome targets originating from protein-coding promoters include premature cleavage and polyadenylation (PCPA) products, PROMPTs (promoter upstream transcripts), whereas XRN2 was shown to contribute to transcriptional repression of H3K27me3-marked genes (92, 311, 320).

Factors determining susceptibility to nuclear decay, and the mechanisms that underlie targeting have only been partially uncovered (86, 231, 313, 319, 321–323). At present, nuclear residence time seems to be important, with apparent competition between factors which promote export and those which target transcripts for degradation (86, 231, 323).

1.2 Cleavage and Polyadenylation

With the exception of replication dependent histone mRNAs, all nascent metazoan mRNAs receive a non-templated 3' poly(A) tail which is usually removed prior to degradation (23, 269, 324–328). In certain cell types/developmental phases, poly(A) tails may be extended in the cytoplasm in order to switch the attached mRNA from dormant to translationally active (9, 329–331).

1.2.1 Cleavage and nuclear polyadenylation

Nuclear poly(A) tail synthesis is initiated in concert with 3' cleavage and utilises some of the same machinery. Sequences in the 3'UTR are recognised in mammalian cells by cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF) and cleavage factor 1 (CFIm) and joined by other factors including cleavage factor 2 (CFIIm) and Symplekin to form the cleavage machinery (78, 80, 168, 173–175, 177, 332–335). Poly(A) polymerase (PAP) is recruited by the FIP1 subunit of CPSF (175) and catalyses poly(A) tail synthesis following cleavage (55, 178–180). An overview of the process is depicted in figure 1.3. Cleavage of replication-dependent histone mRNAs - which are not polyadenylated and instead are stabilised by a 3' UTR stem loop - also involves some CPSF subunits but these are active in a different complex (327, 328, 336).

The poly(A) signal (AAUAAA), located 10-30nt upstream of the cleavage site, is recognised by WDR33 and CPSF4 in the CPSF complex (78–80). Single base deviation from the canonical poly(A) signal (PAS) - commonly AUUAAA - has been widely recorded, and AAUAAA may be present in <60 % human protein-coding genes (166, 337, 338). Analysis of the CPSF-PAS interaction by cryo-EM showed that the hexamer bends to form a Hoogsteen base pair between U3 and A6 and that CPSF1 pre-arranges WDR33 and CPSF4 for binding of the PAS (80). Around 30 nucleotides downstream of the cleavage site is the G/U or U-rich downstream sequence element (DSE) which is present in the majority of transcripts and is bound by the CstF complex (167, 169, 170, 173, 174). Located upstream of the poly(A) signal is the auxiliary upstream sequence element (USE) (334). The USE – often 'UGUA' or a similarly U-rich motif - is bound by the NUDT21 subunit of Cleavage Factor I (CFIm) and its presence is thought to enhance cleavage and polyadenylation efficiency, particularly of intronless mRNAs (334, 339–341).

Cleavage and polyadenylation are linked to transcription through interactions with the Pol II body and CTD (96, 101, 342). While CPSF, or at least some of its subunits, seem to be recruited at the promoter and travel with Pol II along the gene, complete assembly of the cleavage and polyadenylation (CPA) machinery does not occur until the poly(A) signal has been transcribed (101, 103, 104, 342). Arrival at the PAS

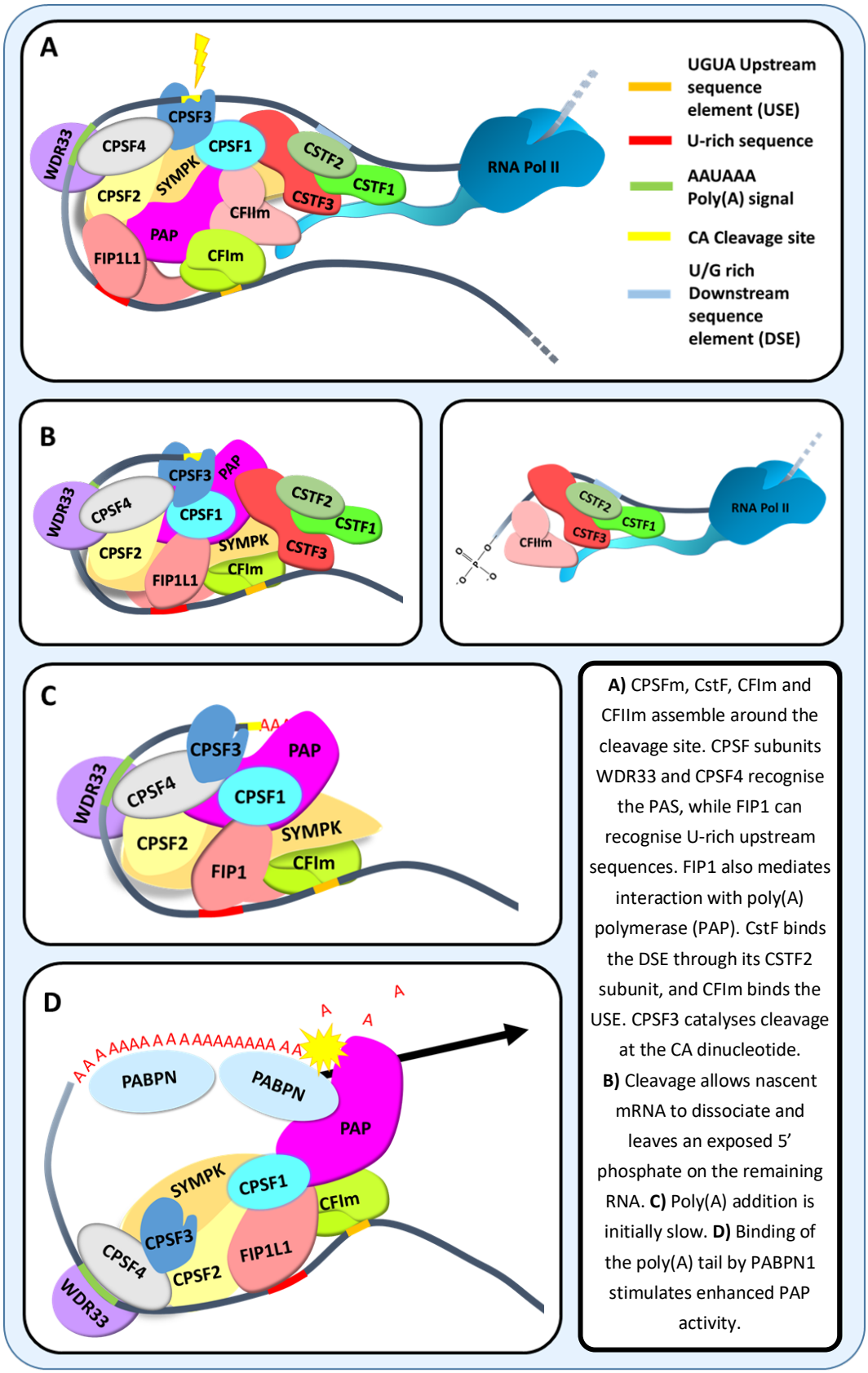


Figure 1.3 Overview of cleavage and polyadenylation.

coincides with increased phosphorylation of Ser2 in the Pol II CTD, a process which seems interdependent with CPA recruitment (103, 104, 343, 344). Blockage of PAS recognition using the Influenza A NS1 protein (which binds CPSF4) prevented 3' end Ser2p enrichment; meanwhile, pausing of Pol II itself seems sufficient to elicit a Ser2p peak (104, 345, 346). In addition, recruitment of other CPA factors relies on, or is enhanced by, Ser2p (103, 104). Taken together, these indicate an integrated, stepwise process in which recognition of the PAS by Pol II-bound CPSF4 causes deceleration of Pol II and increased Ser2 phosphorylation of the CTD, which then facilitates assembly of the remaining CPA machinery. The PNUTS-PP1 phosphatase, which has been found in CPA complexes, was recently shown to reduce Pol II speed by dephosphorylating the transcription elongation factor Spt5 after reaching the PAS (77, 142, 192). It is not yet clear exactly how transcription of the PAS triggers Pol II deceleration, but dephosphorylation of Spt5 coincides with increase phosphorylation of Ser2, and both events depend on inhibition of CDK9 activity (344).

Following assembly of the CPA complex, cleavage is carried out by the CPSF3 subunit of CPSF (177). While PAP does not seem to be tightly associated with the cleavage machinery (192), its presence seems to be required for the mammalian cleavage reaction (347–349). PAP catalyses addition of the poly(A) tail in a reaction which is initially slow, but is then greatly enhanced by binding of PABPN1 once the tail is ~10-12 nt (55, 56, 178, 179, 350, 351).

A process composed of so many factors, yet so central to competent RNA production, is unsurprisingly targeted by pathogens and can be affected in disease (345, 352–355). As previously mentioned, CPSF4 is targeted by the influenza A NS1 protein to block PAS recognition and nascent mRNA cleavage, resulting in readthrough transcripts which cannot fulfil their original function but whose exact fates have not been well characterised (345, 353). CPSF4 is upregulated in lung adenocarcinoma cell lines and patient tumours, and its knockdown in lung cancer cells results in reduction in phosphorylation of PI3K, AKT, ERK1/2 and JNK (355). These changes in phosphorylation state indicate either an interplay between 3' processing and signal transduction – possibly as part of some wider cellular control scheme – or additional roles for CPSF4 itself. Interestingly, CPSF4 was not

detected in normal human tissues, suggesting either that is was present in very small quantities, or that it is not required in more modest growth conditions or can be compensated, perhaps by CPSF4L. CPSF4 is known to be a core component of the 3' processing machinery (78, 80) and its apparent absence from normal human tissues highlights the potential effect that choice of experimental system can have upon conclusion. Both CPSF4 and the other PAS recognition subunit, WDR33 seem to be regulated in their localisation; induction of the inflammatory response with LPS caused a shift of the two subunits from mostly cytoplasmic to both nuclear and cytoplasmic in RAW 264.7 macrophages (354).

1.2.1.1 Poly(A) polymerase

Three canonical poly(A) polymerases (PAPs) have been identified: PAPα (PAPOLA), PAPβ (PAPOLB) and PAPγ (PAPOLG) (178, 349, 356–359). PAP expression varies between tissues; according to the Bgee database (www.bgee.org), PAPα is highly expressed throughout the body in humans, and PAPγ similarly so, though with a considerably more variation in expression level. PAPβ is only highly expressed in the testes and is expressed to a lower level in several other tissues (356, 360). Consequently, much investigation of canonical PAPs has focused on PAPα and the specific paralog is often not disclosed in publications.

PAP can be recruited to the mRNA by CPSF1 and FIP1 (175, 361) and also, in the absence of the canonical PAS, by CFIm (via FIP1) (334, 362). Canonical PAP is stimulated by interaction with PABPN1 once the tail is long enough (~10-12 nt) to be bound bind it (55, 56, 179, 350, 351). This interaction may also function to maintain association with CPSF (56). The PAP-PABPN1 interaction causes PAP to switch from distributive activity (detaching and reattaching) to processive (continuous) activity (52, 56, 179, 363). It is thought that once a certain number of PABPNs are bound, the interaction with CPSF cannot be maintained, explaining a sudden decline in activity above 200-250 adenosines (56, 363, 364).

There is some limited evidence of regulated PAP activity. Hyperphosphorylation of PAPα occurs during M phase - coincident with translational repression of cell cycle transcripts - and was shown to inhibit its activity in vitro (59, 62, 63, 365). This is unlikely to be the only factor contributing to impaired translation since bulk poly(A) lengths showed little difference between phases (62); of course, the necessarily

short half-lives of cell cycle transcripts would make them more susceptible to a temporary reduction in PAP activity, but deadenylation could also contribute. So far, CDK1 and ERK have been shown to phosphorylate PAP, enabling control in response to both intra- and extracellular signals (59, 63, 64). PAPγ does not seem to be phosphorylated despite possessing multiple suitable residues, perhaps due to selection of experimental conditions towards PAPα regulation (349).

Early work recorded a difference in kinetics between chromatin-associated and soluble PAP, with the soluble enzyme exhibiting greater efficiency (366). Though this is a common feature of enzymes, it is interesting to consider its biological relevance here, particularly since the two forms were seemingly stimulated by different ions (Mg^{2+} for chromatin-associated and Mn^{2+} for soluble). Both forms of the polymerase were isolated from HeLa nuclear extracts and could be inhibited by the same antibody, suggesting both activities belonged to the same enzyme; this of course depends on the antibody binding site meaning it is theoretically possible that two different canonical PAP paralogs were detected (366). While the presence of PAP in the nucleoplasm could be due to nuclear import after translation, another possibility is that the soluble form is responsible for the hyperadenylation of non-coding nuclear RNAs which is thought in some cases to precede their degradation (319, 321, 322, 348, 367). The poly(A) tails produced by nucleoplasmic PAP were longer than those from its chromatin-associated counterpart (400-800 vs 200-250) (348, 366). The idea that chromatin association inherently imparts some limit on tail length through slower enzymatic activity is one possibility, however this difference could also have arisen from differences in the enzyme's modifications or surrounding complex. As noted above, although both PAPs could be bound by the same antibody, they may not have been the same enzyme.

While canonical PAPs are thought to enact the bulk of nuclear pre-mRNA polyadenylation, they are not the only mammalian enzymes capable of 3' non-templated nucleotide addition. Another group exists – terminal nucleotidyltransferases (TENTs) – which can catalyse the 3' tailing of both messenger and non-coding RNAs with adenosine as well as other ribonucleotides (368–370). TUT4 and TUT7 for example, add oligo(U) tails to various RNA species which in the context of mRNA, seem to be markers for degradation (368, 369, 371,

372). Of the 11 TENTs encoded by the mammalian genome, only one - TUT1 - has a documented role in nuclear mRNA polyadenylation (65, 370, 373, 374).

TUT1 (TENT1/Star-PAP) is mainly considered a terminal uridyl transferase before a poly(A) polymerase and its primary role is in maturation of the spliceosome component, U6 snRNA (375, 376). Although it has a preference for uridine, TUT1 can polyadenylate a subset of (pre-)mRNAs in the nucleus, with the best studied target being the oxidative stress response gene, HO1 (65, 373, 374). The substrates of canonical PAP and TUT1 appear to have little overlap, and the choice of polymerase is thought to be dictated by the presence and quality of the DSE; suboptimal DSEs are not bound by CstF64 and therefore do not recruit canonical PAP well (65, 377). In addition to being regulated differently to PAPα, TUT1 is thought to be more heavily involved in the cleavage reaction than its canonical counterpart, being required for CPSF3 and CPSF1 recruitment to the mRNA (65).

FAM46A (TENT5A) is an evolutionary conserved protein essential to several dynamic processes including bone formation and embryonic development (67, 378, 379). It has putative signalling function and was thought to be mainly cytosolic, however it too has recently been recognised to have poly(A) polymerase activity (67). In its unphosphorylated state, FAM46A localises to the nucleus where it is enriched mainly in unwound chromatin regions and in the nuclear matrix and lamina, consistent with a transcription-coupled role (67). This preliminary work also showed regulation of its protein levels over the cell cycle and found indications that its poly(A) polymerase activity was important for erythroid differentiation (67). Further study is required to clarify the extent and biological significance of its nuclear role.

1.2.2 Alternative polyadenylation (APA)

Around a third of all mouse, and half of all human protein coding genes contain multiple poly(A) signals which result in mRNAs of varying 3'UTR length (166). Though this does not affect amino acid sequence of the encoded protein, longer 3'UTRs confer increased opportunity for regulation of transcript stability (45, 288, 380), and may also facilitate interactions with factors during translation which affect protein localisation and interaction partners (381, 382). Intra-gene 3'UTR diversity is thought to contribute to organism complexity, likely tailoring control of gene expression in different tissues with those genes expressed in greater numbers of

tissues also exhibiting more 3' UTR isoforms (383). In addition, APA is employed in many dynamic cellular processes such as differentiation, autophagy, induction of pluripotency and immune cell activation, and short 3'UTR oncogene isoforms are frequently favoured in cancer cell lines (380, 381, 384–386).

These poly(A) signals are of varying 'strengths' depending on the combination of surrounding auxiliary elements and fidelity to consensus sequences, and the choice of PAS seems to be at least partially dependent on availability of CPA subunits (381, 387, 388). The effect of different CPA components on poly(A) site choice can largely be categorised into promoting either proximal or distal site usage, with FIP1 (CPSF subunit) and PCF11 (CFIIm subunit) promoting proximal PAS usage and CFIm25/68, PABPN1 and PABPC promoting distal site use (388–390). Shorter 3'UTRs as a result of PABPN1 knockdown abrogated miRNA regulation of the Cyclin D1 mRNA in U2OS cells, and this 3'UTR shortening seemed to be mediated through a lack of repressive binding of proximal PAS sites by PABPN1 (389).

An early study found that CstF had a higher affinity for the distal PAS of the IgM heavy chain mRNA, and use of this PAS produced the membrane-bound form of the protein (381). The 64 kDa subunit of CstF was found to be the limiting factor in its assembly and during B cell differentiation CstF-64 is specifically repressed to reduce CstF formation; this thereby promotes use of the proximal PAS which results in secreted IgM (381, 391). This may not be indicative of a general preference of CstF for distal poly(A) signals; instead it is possible this site had better quality flanking sequences since CstF-64 recognises the GU-rich downstream element (173). Indeed, CstF-64 was specifically upregulated following LPS stimulation of RAW macrophages, leading to an increase in proximal PAS selection in a reporter mini-gene (392).

Not only can the abundance of 3' processing factors affect poly(A) site choice, but so it seems, can their post translational modifications. CFIm regulates poly(A) site choice depending on presence of the USE and interacts with the CPSF subunit, FIP1 through arginine-serine repeat domains in CFIm68 and CFIm59 (334, 362). Hyperphosphorylation of CFIm68/59 disrupts this interaction, presumably negating the pro-cleavage effect of USE presence, though this does not seem to have been tested (362). Upon repression of TORC1 activity in Drosophila, CPSF6 (CFIm68) is phosphorylated, promoting its nuclear translocation and consequent 3' UTR

lengthening of a certain transcripts involved in autophagy (386). APA can be further regulated by factors outside the canonical 3' processing machinery. The U1 snRNP can prevent premature cleavage within the gene body and also promote use of distal poly(A) sites by blocking cleavage ~1 kb downstream of itself in a process known as telescripting (393, 394). This may be achieved by U1 binding to the CPA machinery and preventing its interaction with the CFIm68 subunit (395).

1.2.3 Cytoplasmic polyadenylation

The canonical role for cytoplasmic polyadenylation is in translational activation of dormant mRNAs during oogenesis and early development by extending short poly(A) tails to a length sufficient for translation (9, 13, 396–400). This is thought to be the main setting in metazoans in which poly(A) tail length correlates with translation efficiency, whereas the current consensus is that this correlation does not generally exist in somatic cells (11, 13, 62, 325, 326, 401). This absence of a correlation may be confounded by an abundance of highly expressed mRNAs possessing only medium length tails which could mask any trends in less stable transcripts (11, 326). Among this group of mRNAs are the ribosomal protein transcripts which interestingly also experience only minimal changes in poly(A) length during egg activation (13). Outside of development, cytoplasmic polyadenylation has been described in neuronal cells and the liver, as well as in mitotically dividing HeLa cells, with functions in long-term memory, circadian control and entry to M-phase respectively (66, 330, 331, 402). These dynamic processes perhaps benefit from cytoplasmic polyadenylation as a means to respond rapidly to signals without needing to wait for transcription and export; this is particularly relevant in the case of neurons where the site of translation may be far from the nucleus (403). Similar rapid-response systems have been observed, such as nuclear storage of polyadenylated pre-mRNAs which are spliced upon neuronal activation (404).

Cytoplasmic polyadenylation relies on the presence of both the canonical PAS hexamer and a cytoplasmic polyadenylation element which are bound by CPSF and one of four CPE-binding proteins (CPEBs 1-4), the latter of which interact via Symplekin (329, 331, 396, 397, 405, 406). Polyadenylation is itself carried out by either canonical PAP or the specific cytoplasmic PAP, GLD2 (PAPD4) (9, 399, 400).

Additional cytoplasmic-acting non-canonical PAPs have been identified. PAPD5 (TENT4A) and PAPD7 (TENT4B) confer additional stability to target mRNAs by incorporating occasional guanosine residues to generate a mixed tail (407). These G residues cause the major deadenylases to stall, perhaps through disruption of the stacked formation that can be observed in adenosine-only tails which is thought to be important for recognition by these complexes (407, 408). The increased stability conferred by a mixed tail is exploited by hepatitis B and human cytomegalovirus; these incorporate sequences in their transcripts which form pentaloop structures that are recognised by ZCCHC14, leading to recruitment of TENT4 (PAPD5/7) (409). FAM46C (TENT5C) was recently identified to possess PAP activity, affecting the stability of a subset of mRNAs which pass through the ER/Golgi apparatus and promote cell death (410).

1.2.4 Control of poly(A) tail length

1.2.4.1 Inconsistencies in poly(A) length studies

It was originally thought that a poly(A) tail of 150-250 nucleotides was uniformly added to all higher eukaryotic transcripts, save those coding for histones (56, 179, 269, 324, 327, 363, 411). This figure was initially obtained by radiolabelling experiments in HeLa cells, ranging from 12 minute to 48 hour labelling periods (269, 324, 411). These measurements were corroborated by *in vitro* experiments showing that PAP could perform rapid processive polyadenylation up to around 250 nt through co-stimulation by PABPN1 and CPSF; above this length the co-stimulation ceased (55, 56, 179, 363, 412). More recent endeavours to study poly(A) tail length on a global scale have found that the majority of poly(A) tails present in cells are significantly shorter than earlier measurements (median ~60 nt) and that considerable heterogeneity exists, however only total RNA (i.e. a mixture of nascent and partially deadenylated transcripts) has been studied, and information about the poly(A) lengths of newly made mRNAs (i.e. initial poly(A) tail length) is lacking (11, 325, 326). Thus, partial deadenylation presents an obvious argument to explain differences in observed poly(A) tail lengths. Additional study of nascent poly(A) length control has been very limited (250, 413).

Although deadenylation could feasibly contribute substantially to the differences observed, some inconsistencies remain. In particular, a 48 hour labelling experiment

– which should have approximately captured steady state - still showed a substantial population of cytoplasmic (or at least polysome-associated) RNAs to possess ~200 nt tails (324). This was accompanied by a shorter and broader peak below 80 nt, more akin to the results obtained by global deep-sequencing experiments. Though it would be tempting to speculate that the longer-tailed population comprises some different RNA species, this seems unlikely since the only peaks in the nuclear fraction are a similarly long peak and another shorter and broader peak which has a much lower modal length that its cytoplasmic counterpart. This either suggests that more modern experiments have failed to capture the long tails originally detected, or that the original measurements were flawed in some way. For example, size determination may have been inaccurate, or the long labelling period may have affected cell behaviour. Crucially, the position of the longer-tailed peak is similar in the nuclear and polysomal fractions, suggesting that cytoplasmic deadenylation may not lead to radically different poly(A) distributions between fractions (324). This raises the possibility that recent measures may be more reflective of nuclear poly(A) dynamics than is often assumed.

One factor compounding the idea that all mRNAs are made with long tails is the frequent use of reporter constructs incorporating exogenous PAS /3'UTR constructs eg SV40, which are assumed to represent endogenous mRNAs (179, 337). Although these transcripts may normally undergo polyadenylation by the host machinery, it may be incorrect to assume that the process fastidiously reflects treatment of cellular transcripts. Illustrative of this: when incubated with HeLa nuclear extract, substrates containing the PAS of CFIm68 had considerably shorter poly(A) tails that were more heterogeneously distributed compared to those containing SV40 late or bovine growth hormone poly(A) signals (415, figure 1B).

1.2.4.2 Initial poly(A) length regulation

It is widely accepted that the metazoan nuclear poly(A) binding protein PABPN1 stimulates polyadenylation and limits mRNA poly(A) tail length to ~250 nucleotides, however it has also been recorded to promote hyperadenylation and nuclear decay of viral and noncoding RNA (56, 321, 363, 364). On the other hand, nuclear accumulation of PABPC following expression of a viral nuclease which promotes cytoplasmic mRNA degradation is associated with hyperadenylation, however there

are conflicting results as to the direction of this relationship (271, 415–417). While it is possible that nuclear accumulation of PABPC is the driving force which promotes mRNA hyperadenylation and nuclear retention, one group found evidence that PABPC is reliant on association with mRNA for export and therefore that mRNA hyperadenylation causes nuclear accumulation of PABPC (417). The yeast homologue to PABPC, Pab1 instead seems to limit poly(A) length in a reconstituted reaction (418).

In yeast, the predominantly nuclear poly(A) binding protein, Nab2 stimulates polyadenylation and provides protection against hyperadenylation, but does not behave with any specificity (419). Interestingly, the nuclear Pab1:Nab2 ratio may be important for determining poly(A) length since the exosome subunit, Rrp6 can displace Nab2 but not Pab1 and elicit degradation of the newly exposed stretch of tail (420). The mammalian orthologue to Nab2, ZC3H14, also binds the poly(A) tract with high affinity and prevents hyperadenylation of bulk mRNA in N2A cells (421).

Another general factor involved in poly(A) length control is nucleophosmin which associates with the majority of poly(A)+ mRNAs upstream of the PAS to limit tail length, its depletion leading to hyperadenylation (414, 422). Reports exist both of NPM1 association requiring active polyadenylation (pre-cleaved and pre-polyadenylated transcript failed to associate with NPM1 in vitro) and of a requirement for correct termination of PAP activity (premature termination of PAP activity by cordycepin abrogated NPM1 recruitment) (414, 422). Taken together, these findings suggest that NPM1 may be recruited during and contribute to normal termination of polyadenylation.

As introduced earlier, PAP activity itself can also be attenuated via hyperphosphorylation, documented so far to occur during M phase in the somatic cell cycle (59, 62, 63).

1.2.4.3 mRNA-specific poly(A) regulation

While the control of bulk mRNA poly(A) tail length has been relatively well documented, the few studies indicating that poly(A) tails may be regulated in somatic cells in a more targeted manner have been considered interesting exceptions (423–428). Early experiments found mouse globin mRNAs possess ~50 nt

tails at steady state (18 hour incubation with radioisotopes), a result consistent with widespread reports of prevalent medium length tails, particularly on highly expressed mRNAs (325, 326, 423). Even shorter tails (12-30 nt) were discovered on a subset of mRNAs enriched for translation, mitochondrial and growth functions at steady state in human liver tissue (428). The mRNAs in this subset contained both the nuclear PAS and the cytoplasmic polyadenylation element (CPE), suggesting that production of stable mRNAs with short tails is not a phenomenon limited to early development. Although these are not themselves evidence for message-specific control of initial poly(A) length, they indicate that the story is likely not as simple as a blanket 200-250 nt tail acting as timer across all mRNAs as previously laid out.

Studies of poly(A) regulation of nascent mRNA have mainly focused on the poly(A) limiting element (PLE), a 23 nt sequence present in the terminal exon of affected mRNAs which limits poly(A) tails to 17 nt (425, 427). The PLE was discovered in Xenopus albumin mRNA and its tail length limiting effects are evident on the pre-mRNA, a result recapitulated using transferrin mRNA and pre-mRNA (424, 426). This suggests either impaired tail synthesis or early targeted rapid deadenylation is at play. Plasmids encoding PLE-containing transcripts also yielded mRNAs with short poly(A) tails in mammalian cells, indicating that the mechanism is conserved across species (426). The exact mechanism by which presence of the PLE in the terminal exon restricts poly(A) tail length is not clear, but it seems to be bound by the auxiliary splicing factor, U2AF which has been implicated in crosstalk between polyadenylation and splicing (162, 427, 429).

Outside of work on the PLE, the main example of regulated nuclear poly(A) length is on the eNOS mRNA. In this case, increases in laminar flow stress on endothelial cells cause a switch from <25 nt to long poly(A) tails, simultaneously with increased transcription (413). As with cytoplasmic polyadenylation this led to an increase in translational efficiency, but crucially the changes to poly(A) tails on eNOS transcripts were nuclear in origin and occurred in somatic cells.

Recently detected widespread deviation from the purported 200 nt length has been suggested to occur via targeted cytoplasmic pruning by deadenylases (326). This is not a new idea; yeast Saccharomyces cerevisiae 3’ end processing extracts lacking poly(A) nuclease (PAN) produced transcripts with 200 nt poly(A) tails instead of the

wild type 60-80 nt, indicating an additional role for deadenylation in 3' processing (430). While targeted trimming by deadenylases unifies otherwise contradictory early and recent poly(A) length measurements, it does not sit comfortably as a universal rule. Trimming represents a straightforward system for tailoring the tails of unstable mRNAs which may already be bound in their 3'UTRs by deadenylase-interacting factors, however, it seems a wasteful solution for constitutively highly expressed mRNAs which are destined to possess much shorter poly(A) tails for the majority of their lifetimes. However, it cannot be ruled out that a 250 nt tail fulfils some brief but essential function, perhaps in escaping immediate degradation.

Previous work showed differences in poly(A) tail length between reporters containing poly(A) signals of different origins (415, figure 1B). Although this implies that 3' end sequences have the capacity to influence poly(A) tail length at the point of synthesis, investigation of such a relationship has been minimal (431–433). More recently, a weak positive correlation between 3'UTR length and poly(A) length was observed when investigated with the poly(A) deep sequencing technique, FLAM-Seq. FLAM-Seq uses the PacBio platform to measure the tail lengths of full length cDNAs (and can thus distinguish between 3'UTR isoforms) (433). Though this correlation is weak, it may indicate existence of subsets of mRNAs with stronger and/or different relationships. Any potential specificity need not be limited to poly(A) signals and their surrounding elements but could depend on other sequences in the 3'UTR; nuclear recruitment of Tristetraprolin (TTP) to AU-rich elements for example, inhibits polyadenylation of the containing mRNA (431). This is purportedly though simultaneous binding of both PABPN1 and PAP by TTP. This finding is complicated by the known role of TTP in recruiting the CCR4-NOT deadenylase complex for cytoplasmic deadenylation, however, in vitro binding assays do lend some credibility to a model of direct PAP inhibition (255, 431).

1.3 Roles of polyadenylation

The eukaryotic mRNA poly(A) tail has established roles in stability, export and translation efficiency (1, 2, 10, 11, 13, 19, 23, 26, 62, 269–271, 274, 325, 398, 434). Current understanding is that the relationship between each of these properties and poly(A) length is however, not universal and instead varies depending on the mRNA, subcellular location, and cell state (8, 13, 62, 250, 272, 274, 326, 417, 435).

1.3.1 Poly(A) binding proteins

The presence and lengths of poly(A) tails are detected through binding of poly(A) binding proteins (PABPs), which can mediate interactions with other cellular machinery, such as that involved in nuclear export, deadenylation and translation (2, 20, 24, 28, 50, 321, 436–441). Importantly, the canonical PABPs have been documented to bind the poly(A) tail in multiple copies, causing signature surges and pauses in deadenylation, affecting recognition by deadenylases, and behaving as a measuring stick during poly(A) synthesis (28, 50, 364, 412, 442). At least one report suggests that multiple PABP binding is not homogenous, and that PABPN1 and PABPC1 may bind the poly(A) tail simultaneously, with both proteins able to immunoprecipitate a number of pre-mRNA transcripts, but not histone mRNAs (443). An early study also suggested that PABP II (PABPN1) and PABP I (PABPC1) have opposing effects on PAP activity, though a role for PABPC in limiting poly(A) length is incongruous with reports that increased nuclear localisation of PABPC is associated with nuclear retention of hyperadenylated transcripts (350, 415, 417). One study found that depletion of PABPN1 in human cell lines did not affect β-actin poly(A) length, but resulted in compensatory upregulation and nuclear localisation of PABPC isoforms, suggesting that the two proteins have some functional redundancy (444).

1.3.1.1 PABPN1

As its name suggests, nuclear poly(A) binding protein (PABPN1) is a predominantly nuclear poly(A) binding protein that is conserved across metazoans and has a known homologue in fission (but not budding) yeast (54, 445). It is thought to transiently locate to the cytoplasm while attached to newly exported mRNAs, before being dislodged by the ribosome during the pioneer round of translation (54, 363, 420, 446–448). Examination by cryoimmunoelectron microscopy found little evidence of cytoplasmic PABPN1, but did find that it was associated with RNA Pol II (54).

The most well-known function of PABPN1 is its stimulation of poly(A) polymerase (PAP) during nuclear polyadenylation. In the absence of PABPN1, PAP activity is minimal, meaning that the first ~10-11 nucleotides – the number required for PABPN1 to bind – is very slow before accelerating, resulting in two observed phases of tail addition (55, 56, 179, 350, 351, 363, 412). *In vitro* experiments using pre-

cleaved RNA substrates found that PAP activity was stimulated significantly upon binding of the first PABPN1, up to a maximum tail length of 200-250 nt, above which PAP activity returned to being minimal (56, 412). Maximum PAP stimulation was also reliant on CPSF (56, 412). The ability of PABPN1 to stimulate PAP up to a maximum length gave rise to the idea that PABPN1 could act as a molecular ruler, with multiple copies binding until PAP is mechanically unable to maintain association with CPSF (364, 412). Consistent with a role for PABPN1 in stimulating PAP, depletion of PABPN1 caused shorter poly(A) tails in mouse myoblasts (436).

In addition to promoting polyadenylation, PABPN1 has been reported both to mediate the export of mature mRNAs through interaction with Aly/REF, and to target hyperadenylated RNAs to the nuclear exosome for degradation (2, 231, 322, 367). The necessity of the PABPN1 interaction for mRNA export is unclear, as Aly/REF also has documented interactions with both the cap binding complex (CBC) and exon junction complexes (EJCs) (2, 5); however, it may be that the Aly/REF – PABPN1 interaction is important for redirecting mature mRNAs from a nuclear degradation pathway (231). More rigorous binding assays between PABPN1 and export components are required in order to isolate any direct effects on export from those attributable to the established role of PABPN1 in promoting poly(A) tail synthesis.

Despite its apparent ubiquity and - by current understanding - universal role in poly(A) tail formation and length control, additional alanine residues in the N-terminal region of PABPN1 specifically affect a specific subset of muscles, causing oculopharyngeal muscular dystrophy (OPMD) (449). Recent work showed that Pabpn1 mRNA is unstable and exists at particularly low steady state mRNA and protein levels in the muscles affected in OPMD (450). The instability of the Pabpn1 mRNA apparently derives from an AU-rich element (ARE) within its 3'UTR which is bound by the ARE-binding protein HuR (451). Although HuR binding is usually considered to stabilise transcripts, this may not be by preventing deadenylation; indeed, HuR itself can also recruit deadenylases (452–454).

PABPN1 also exhibits autoregulation, achieved through binding to an A-rich stretch in its 3'UTR (312). Binding to this sequence inhibits splicing of the 3' terminal intron, promoting nuclear retention and degradation by the nuclear RNA exosome. In

conditions of low PABPN1 protein, the 3' terminal intron is efficiently spliced and the mRNA successfully exported and translated.

1.3.1.2 PABPC

Unlike PABPN1, many isoforms of PABPC exist, with PABPC1 the best studied, likely due to its more widespread expression (455). Additional cytoplasmic poly(A) binding proteins have been described, including PABPCs 3-5, PABPC1L and PABPC4L but these rarely feature in the literature (455–457). PABPC1 and PABPC4 contain 4 RNA recognition motifs (RRMs) which each bind 8 nt of RNA (28, 458). When multiple Pab1s (PABPC1s) are bound, RRM4 can interact with RRMs 1 and 2 of the adjacent PABP, forming repeated arch structures which are recognised by the PAN2-PAN3 deadenylase (442).

The conventional roles for PABPC1 centre around its interaction with the eIF4G subunit of eIF4F, the complex which replaces CBC in binding the 5' cap after the pioneer round of translation (19–21, 438, 447). This has been formalised as the 'closed loop' model, in which the PABPC-eIF4G interaction serves the dual purpose of protecting both ends of the mRNA from degradation by blocking exonuclease access, and of enhancing translation efficiency by reducing the need for repeated ribosome recruitment and assembly (19–21). Of note is the correlation between poly(A) tail length and translation efficiency of the Pabpc1 mRNA in the adult heart (434). In response to stress and hypertrophic stimuli, the poly(A) tail of Pabpc1 mRNA is lengthened, increasing the amount of PABPC1 available for loop formation and thereby enhancing global protein synthesis within the cell.

Contrary to its aforementioned role in protecting mRNA, PABPC1 has also long been recognised to stimulate activity of some deadenylases (24, 28, 50, 459). PABPC1 was initially thought only to stimulate the PAN2-PAN3 deadenylase - which has limited activity in its absence - and was thought to inhibit both nuclease subunits of the other major mammalian deadenylase, CCR4-NOT (24, 460, 461). More recent *in vitro* investigations of CCR4-NOT kinetics have found that PABPC1 can directly interact with and stimulate CCR4, and that CAF1 can also be stimulated indirectly via an interaction with BTG/Tob family proteins (28, 50, 462, 463). There have also been suggestions that, rather than protecting mRNAs from deadenylation, PABPC1 may

be more critical for preventing premature terminal uridylation, an emerging signal for mRNA decay (50, 369).

Exactly what delineates the protective and de-protective roles of PABPC1 in the cytoplasm is unclear. It seems likely that, as with PABPN1, at least some element of control is through choice of PABPC1 binding partner, and that binding of the poly(A) tail by PABPC1 may not itself commit the mRNA to a particular fate. Binding density of PABPC1 is not consistent across mRNAs (464, 465) and one could speculate that this is important for determining interactions. Alternatively, post translational modifications of PABPC1 could influence its function (466).

In addition to its cytoplasmic roles in mRNA [de-]stabilisation, PABPC shuttles to the nucleus (271, 415–417, 467–470). This is particularly evident in viral infection and under other stress conditions (271, 415–417, 468–470). Nuclear PABPC accumulation is accompanied by accumulation of hyperadenylated transcripts in the nucleus and a reduction in gene expression (415, 470, 471). This co-accumulation led to the hypothesis that PABPC induces hyperadenylation (415). The dependency of viral-induced nuclear accumulation of hyperadenylated transcripts on PABPC was confirmed by combined knockdown of PABPC1 and 4. PABPC1/4 knockdown prevented nuclear enrichment of poly(A)+ RNA which normally occurs in response to expression of the viral SOX endonuclease (415). Given the many levels of PABPC involvement and long duration of RNAi-mediated depletion, it is hard to conclude a direct effect on nuclear hyperadenylation. The sufficiency of PABPC1 translocation in causing mRNA hyperadenylation and nuclear retention was however confirmed by expressing a nuclear localization signal (NLS) – tagged PABPC1. Forced nuclear localisation of PABPC1 recapitulated viral protein-induced observations of nuclear retention (measured by *in situ* hybridization) and hyperadenylation (northern blotting of a reporter mRNA) (415). The effect on hyperadenylation was reliant on the first two (of four) RNA recognition motifs (RRMs) of PABPC1, suggesting that the effect on hyperadenylation may via direct interaction with the RNA, though these RRMs also interact with eIF4G (415, 472). A more recent study involving expression of the same viral endonuclease found that siRNA knockdown of PABPC1/4 also prevented endonuclease-induced depletion of Pol II from promoters (470).

Additional work by the same group showed that PABPC1 is routinely shuttled to the nucleus and is mainly exported via association with mRNAs, leading to - or at least contributing to - its nuclear accumulation when mRNA export is impaired (416). In this model, PABPC1 is transported to the nucleus by default via importin α, through an interaction involving regions of its RRMs. Messenger RNA competes for binding of these RRMs such that when cytoplasmic mRNA/poly(A) levels are high, PABPC1 is not transported to the nucleus (416).

This model is largely in keeping with observations made in the context of UV stress. Here, PABPC1 and 4 were enriched in nuclei 15 hours after UV exposure, and this was accompanied by a nuclear accumulation of poly(A)+ RNA which appeared to occur faster than PABPC redistribution (417). Unlike in viral infection, PABPC1/4 knockdown did not affect the ability of stress to cause redistribution of mRNA and instead, blockage of mRNA export caused nuclear accumulation of PABPC (417).

How this co-accumulation of PABPC and hyperadenylated mRNA can be reconciled with earlier reports that PABP I (PABPC) in nuclear extracts inhibits polyadenylation is not clear, though it is perhaps possible that PABPC inhibits both PAP *and* some nuclear deadenylase activity, leading to a net increase in poly(A) tail length (350).

1.3.1.3 Non-canonical PABPs

Binding of the poly(A) tail is not only the remit of the canonical PABPs. Several novel PABPs have been identified, though only ZC3H14 has been well characterised (455, 473, 474). The mammalian zinc finger protein ZC3H14 is a recently identified poly(A) binding protein orthologous to Nab2 in yeast and Drosophila (421, 473). Nab2 also exhibits functional similarity to mammalian PABPN1 by stimulating PAP and protecting against hyperadenylation (419).

Four isoforms of ZC3H14 have been recorded, three of which are ubiquitously expressed and nuclear, and the fourth of which is expressed predominantly in the testes and brain and is cytoplasmic (475). ZC3H14 is particularly important for neuronal function; mutations in *ZC3H14* affecting isoforms 1-3 have been identified in intellectual disability and deletion of exon 13 (affecting all four isoforms) in mice leads to memory impairment (16, 476). The effect of exon 13 deletion on poly(A) length was compared between different areas of the brain, and the liver where

ZC3H14 is also highly expressed; curiously, there was little difference in bulk poly(A) length distribution in the liver or cortex samples, in contrast to a modest increase in bulk poly(A) length in the hippocampus (16). This is consistent with the observation that mutant Nab2 led to an increase in bulk poly(A) length in Drosophila heads, whereas overexpression of WT Nab2 caused shorter poly(A) tails, a phenotype which could be rescued by overexpression of the PABPN1 orthologue Pabp2 (476). These data suggest that in neuronal cells at least, PABPN1 and ZC3H14 may act antagonistically to achieve correct bulk poly(A) length. It is worth mentioning that the increases in poly(A) length observed in *Nab2* and *Zc3h14* mutations, rather than exhibiting a clear shift to longer tails, have a characteristic increase in abundance of long tails without a concomitant depletion of short tails (16, 476). This may be due either to an abrogation of regulation leading to more 'fuzzy' poly(A) dynamics, or to only a subset of mRNAs being affected. Importance of ZC3H14 for the poly(A) lengths of only a subset of mRNAs is consistent with the finding that depletion of ZC3H14 only affected the levels of 1 % of expressed transcripts compared to 17 % in PABPN1 depleted cells (477).

The non-canonical cytoplasmic PABPs, LARP4 and hnRNP-Q1 have also been identified (455). LARP4 can bind PABPC directly, in an interaction stabilised b y RNA binding (455, 474). Its main poly(A) binding-mediated effect seems to be to stabilise a subset of transcripts. Since LARP4 associates with the 40S ribosome subunit and with translating polysomes, it has been suggested that LARP4 binding predominantly affects translating mRNAs (455, 474). More recently, LARP4 was shown to specifically stabilise mRNAs with poly(A) tails between 30 and 75 nucleotides (478). hnRNP-Q1 is ubiquitously expressed and is involved in multiple post-transcriptional processes. hnRNP-Q1 possesses specific poly(A) binding activity which enable it to translationally repress target mRNAs by blocking PABPC binding (455).

1.3.2 Polyadenylation and mRNA biogenesis

Based solely on textbook descriptions, and indeed those presented earlier, it seems intuitive to visualise mRNA synthesis as a linear production line. Although some modifications precede others by necessity – a transcript lacking a 5' cap is unlikely to survive long enough to be polyadenylated – the actual timeline is less clear for others.

Of particular note is the mutually effectual relationship between polyadenylation and splicing (107, 161, 162, 429, 479–482). In higher eukaryotes where introns tend to be long, exon definition rather than intron definition is thought be the prevailing mode of splicing (159, 161, 479, 483). In this case, rather than the splice sites either end of a given intron interacting, splice sites (SS) interact across exons, such that the 3’ SS of the upstream intron contacts the 5’ SS of the downstream one. Because of this reliance on an adjacent intron, removal of the terminal introns become problematic, since for example the 3’ SS of the 3’- most intron would have no 5’ SS to interact with, short of looping round to interact with that of the first intron. Removal of these introns may instead rely on cooperation of the spliceosome with either the CBC or 3’ processing machinery (161–163, 429, 479, 481, 484).

Indeed, metabolic labelling and chromatin fractionation followed by nanopore sequencing showed that in human cells, only around 60 % of cleaved transcripts and <10 % of uncleaved transcripts have a spliced terminal intron, suggesting that splicing of this intron occurs alongside or after 3’ processing (121). The study also showed that in human K562 cells, very little splicing occurred until Pol II had transcribed at least 4 kb beyond the 3’ splice site (SS), whereas in Drosophila S2 cells most splicing was completed while Pol II was 1-2 kb downstream of the 3’ SS (121). Consistent with this, RNA-Seq of chromatin and nucleoplasmic fractions during the lipid A response in mouse macrophages showed that 3’ end processing occurs quickly after transcription, whereas splicing of within some genes is much slower (485). This delayed splicing is caused by slow ligation between certain exon pairs, and leads to retention on the chromatin of incompletely spliced, but cleaved and polyadenylated transcripts. In yeast on the other hand, splicing was shown to be 50 % complete by the time Pol II had transcribed only 45 nt past introns, perhaps due to shorter introns allowing more efficient removal (486).

The C-terminal domain of PAP stimulates splicing via U2AF65 and in turn, recognition of the final intron stimulates 3’ processing (107, 162, 480). Stimulation of 3’ processing may also involve interaction of U2AF65 with CFIm (429). In yeast, the poly(A) binding protein Nab2 is also required for splicing of a subset of mRNAs, particularly those encoding ribosomal proteins or containing large introns (482).

As well as acting synergistically, splicing machinery can inhibit polyadenylation. Binding of the U1 snRNP prevents premature cleavage and polyadenylation, and can influence alternative polyadenylation by a process known as telescripting (393–395). This may be achieved by U1 binding to the CPA machinery and preventing its interaction with the 68 kDa subunit of CFIm (395). Early work also suggested that U1 can interact with and inhibit PAP directly (487). Further linking splicing with polyadenylation, the splicing factors SRSF3 and 7 promote distal and proximal PAS use respectively and this may relate to their ability to interact with CFIm68 (237, 488).

3’ end processing may be also be linked to presence of a cap. Depletion of the CBC from HeLa cell nuclear extract greatly reduced cleavage of labelled pre-mRNA, but only slightly inhibited polyadenylation of a pre-cleaved transcript (489). This is thought to be due to the CBC stabilising complexes containing both CPSF and CstF (cleavage reaction), but not those lacking CstF (polyadenylation reaction).

1.3.3 Polyadenylation and mRNA export

Along with a 5’ cap and exon junction complexes (EJCs), a poly(A) tail is indicative of mRNA maturity and signals suitability of the mRNA for export (2, 5, 490). As touched on in discussion of PABPCs however, too long a poly(A) tail is associated with nuclear retention (415). In yeast the poly(A) binding protein Nab2 is required for poly(A)+ RNA export (491).

It is not clear exactly how the poly(A) tail signals for export since PABPN1 has no known interactions with the export machinery that are not shared by the CBC or EJCs (2, 490). It is possible that the process of polyadenylation rather than the tail itself is accountable, though no clear mechanism has been characterised. Nucleophosmin (NPM1) is deposited on the 3’UTRs of mRNAs unless polyadenylation is prematurely terminated by cordycepin (414, 422). Reduction in NPM1 leads to hyperadenylation and nuclear retention of poly(A)+ RNA and may therefore be an appropriate candidate for coupling the two processes (414). It has also been suggested that dissociating PAP may signal for nuclear export (490).

Several elements of the cleavage machinery also have reported links to export. In yeast a cleavage factor component, Pcf11, is required for recruitment of the export

factor Yra1 (492). In mammalian cells, CFIm68 shuttles between the nucleus and cytoplasm and promotes export through direct interactions with NXF1 (202). The TREX component THOC5 can interact with CPSF2 following serum stimulation of mouse cells and this may be required for recruitment of CPSF2 to a subset of immediate early genes (493).

1.3.4 Poly(A) tail and translation efficiency

Translation efficiency (TE) is often recited as one of the three main functions of the poly(A) tail, however the relationship appears to be less straightforward than originally thought. While a positive correlation exists during early development, a similar rule does not seem to hold true in somatic cells, with a few exceptions (11, 13, 62, 325, 326, 396, 397, 433, 434). Two high-throughput studies identified a moderate positive correlation during early development in vertebrates and Drosophila but found no correlation following onset of zygotic transcription (11, 13). Although these setups provide huge amounts of information, it is important to consider factors which may influence detection of a trend. For example, in one study the range of mean poly(A) tail lengths is much larger prior to zygotic transcription than afterwards, and this range importantly includes tails which would be too short to host a single PABPC (and may therefore not be highly translated) (11). That said, another study which drew the same conclusion in Drosophila found similar ranges and minimum values in immature and mature oocytes (13). Since PABPC is thought to enable formation of a closed loop which enhances translation, it is not surprising that mRNAs with tails too short to bind PABPC experience poor translation efficiency; indeed, other work has suggested that a correlation exists for tails below 20 nt in somatic cells (20, 21, 62).

Outside of early development, correlation between poly(A) length and translation efficiency has been identified in a number of dynamic processes in which transcriptional changes may present too great a delay in responding to signals (62, 66, 330, 434). These systems include examples both of up and down-regulation of TE through cytoplasmic polyadenylation and targeted deadenylation respectively.

In the context of unperturbed somatic cells, numerous studies have refuted the idea that TE is correlated with poly(A) tail length (11, 13, 325, 326, 433). In these systems however, a large proportion of the mRNA population is comprised of stable, highly

expressed mRNAs with medium length tails (326). Since these mRNAs already possess unexpectedly short tails for their stability, it is reasonable to consider that the mechanism conferring this stability may also enhance their translation efficiency in a tail length-independent manner. The PABPC-interacting protein LARP1 for example, affects translation of 5' TOP mRNAs. In one study LARP1 also promoted association with polysomes, however, the opposite effect on translation efficiency was observed by other groups (494–497). Even if this is not the case, these mRNAs appear to be a somewhat 'special case' when it comes to tail length and it would be prudent to examine the remaining pool in their absence.

High throughput studies often compare summary statistics such as the mean of medians of distributions, which may mask any trends at the level of individual genes. It would therefore be interesting to compare poly(A) tail length with TE over multiple transcripts from the same gene to see if there is any correlation. Furthermore, some groups, while focusing on accurate high throughput poly(A) measures, often re-use existing translation efficiency data or use codon optimality as a proxy rather than performing both measurements on the same biological samples (326, 433).

1.3.5 Poly(A) tail and mRNA stability

Since canonical mRNA turnover requires that deadenylation occurs prior to degradation, poly(A) tail length has long been considered an indicator of mRNA stability (269, 270). While long cytoplasmic poly(A) tails are protective, oligo(A) tails are an excellent substrate for 3' -5' degradation by the RNA exosome (498).

Recent work across several species has shown that many highly stable mRNAs only possess medium length rather than long tails, suggesting that a universal tail length-stability relationship may not be applicable (325, 326). More recently it was shown that once poly(A) tails are short, decay rates are still extremely variable (1000-fold variation in decay constants) (250). Decay rate and polyadenylation were not completely uncoupled though, since mRNAs which undergo rapid deadenylation are degraded faster than slowly deadenylated transcripts once their tails are short. Of perhaps more consequence for mRNA stability is the density of PABPC binding on the tail (465), though correlation was weak and additional factors may be involved.

For example, in addition to its effect on translation LARP1 specifically stabilises 5’ TOP mRNAs and was also shown to bind poly(A) tails (494, 496, 499).

While presence of a poly(A) tail is generally associated with cytoplasmic mRNA stability, eukaryotic poly(A) tails are not universally protective. Polyadenylation of some RNAs (eg. spliced transcripts from snoRNA host genes, some lncRNAs and premature termination products) in the nucleus instead promotes their degradation by the nuclear RNA exosome (319, 321, 322). Exosome-mediated degradation of mRNA can also occur in the cytoplasm and may be enhanced by the presence of oligo(A) tails (500).

1.4 Deadenylation

In the canonical model for bulk mRNA turnover, the first step is deadenylation in which poly(A) tails are removed (26, 269, 460). Below ~20 nt, cytoplasmic poly(A) binding protein (PABPC) is thought to no longer bind (28, 62, 501, 502). This disrupts the PABPC-eIF4G interaction and releases the mRNA from the closed loop formation, leaving the 5’ end vulnerable to decapping (19, 20). The cap is removed by DCP1-DCP2, which is activated by PAT1, DDX6 and the LSM1-7 complex, and the exposed mRNA is then degraded in the 5’-3’ direction by XRN1 (26, 280, 281, 283, 285).

Deadenylation can be performed by several different complexes within the eukaryotic cell. Although many confirmed and putative deadenylases have been identified, two complexes – PAN2-PAN3 and CCR4-NOT are thought to be responsible for bulk mRNA deadenylation (26, 50, 503). Other deadenylases affect subsets of transcripts, and at least one of these is specifically regulated according to internal and external stimuli (68, 504–506). Of the two major deadenylases, CCR4-NOT is considered dominant, though deletion of any single nuclease is not lethal in yeast (50, 503). PAN2-PAN3 is thought to operate early in deadenylation on long poly(A) tails since it requires stimulation by PABPC, and removal of the remaining tail by CCR4-NOT is suggested to follow (26). Recent cryo-EM of a reconstituted system corroborates this model since adjacent PABPCs interacted to fold the tail into a conformation recognisable by PAN2-PAN3 (442). Knockdown of PAN2-PAN3 had no effect on bulk mean poly(A) length in mammalian cells, suggesting that its action is easily compensated (50). CCR4-NOT contains two

nucleases (CCR4 and CAF1) which are differently affected by PABPC, and the complex is able to fully deadenylate synthetic constructs in the presence or absence of PABPC *in vitro* (28, 50). Depletion of CCR4-NOT components in mammalian cells caused bulk poly(A) lengthening (50).

Deadenylases fall into two superfamilies: DEDD (from conserved Asp and Glu residues) and EEP, exonuclease-endonuclease-phosphatase (507). CAF1, PAN2 and PARN are all members of the DEDD family, while CCR4 belongs to the EEP family.

1.4.1 CCR4-NOT

The CCR4-NOT complex is a eukaryotic-conserved ~600-730 kDa complex (observed mass of 659±4 kDa in humans) (31, 308, 508) thought to be responsible for most mRNA deadenylation. It is made up of 8 subunits, 6 of which (CNOT1, CAF1, CCR4, CNOT2, CNOT3 (Not5 in yeast) and CNOT9) are common between mammals and yeast (308, 509). Figure 1.4 gives an overview of CCR4-NOT subunits and their functions. It is not clear how stringently all 9 subunits are incorporated at any given time. In addition to its deadenylase activity, CCR4-NOT possesses an ever-growing list of additional functions (33–36, 39, 41–44, 143, 147, 307, 510, 511). This broad repertoire of functions may be mediated by the diversity of its constituents, which comprise two kinetically distinct nucleases (CCR4 and CAF1), and a number of other subunits organised into 3 non-nuclease modules: NOT, CNOT9 and CNOT10/11, which can contact a wide range of binding partners (31, 36, 42, 44, 46, 51, 256, 305, 511–516). These subunits are held together by the scaffold subunit, CNOT1 (27, 31, 517). In yeast the E3 ring ubiquitin ligase, Not4, is also incorporated whereas in Drosophila and mammals CNOT4 is not constitutively associated (308, 515). Human CNOT4 could partially compensate for *not4* deletion in yeast so it is thought that the function is conserved (518). In addition, the mammalian CNOT3 function seems to be provided by two subunits in yeast: Not3 and Not5, of which Not5 seems to be the most similar to CNOT3 (308). As well as small differences in the complex's composition between yeast and humans, the conserved subunits also display some variation in function. For example, unlike its yeast counterpart, the human NOT module (CNOT2, CNOT3 and the C-terminal domain of CNOT1) is unable to directly bind RNA and has been suggested to serve more in a stabilisation capacity (31, 514). Although the work presented here is predominantly mammalian-focused, much

research on the activities of CCR4-NOT has been carried out in yeast due to the comparative ease of genetic manipulation; it is therefore prudent to hold these differences in mind while examining the literature.

In mammalian cells, the NOT module comprises the C-terminal domain of CNOT1 along with CNOT2 and CNOT3, and in yeast it also includes Not5 (31, 46, 514). The NOT module has mainly been implicated in transcription initiation but its presence also seems important for deadenylation. Indeed, CNOT3 is required for the mRNA deadenylation necessary to maintain the pluripotent state of mouse epiblast cells (12). Interestingly, CNOT3 levels are further reduced in nutrient deprivation and this may occur as a result of reduced mTORC1 activity (60, 519). CNOT2 is downregulated in highly metastatic cells and it is therefore thought to have an

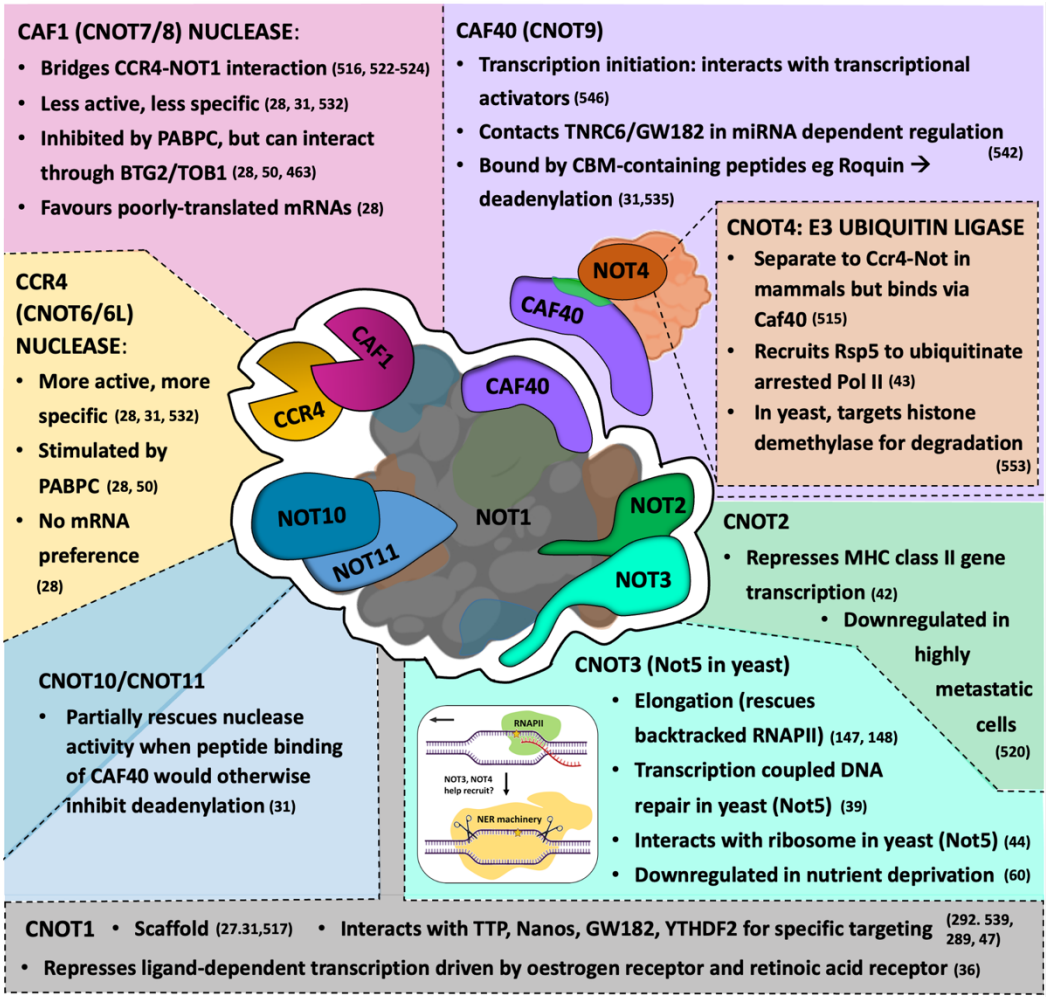


Figure 1.4 Schematic of the CCR4-NOT complex. Mammalian subunits shown.

antiproliferative role (520). Furthermore, in yeast, Not3 and Not5 promote The catalytic module comprises two nucleases: CCR4 and CAF1. In mammals, two isoforms of each subunit exist: with *CNOT6* and *CNOT6L* encoding CCR4a and CCR4b respectively, and *CNOT7* and *CNOT8* encoding CAF1 (521). The prevalence and relevance of each isoform/combination of isoforms *in vivo* does not seem to have been widely studied. Unless otherwise specified, CAF1 and CCR4 will be used here to refer to CNOT7/8 and CNOT6/6L respectively. CCR4 does not directly interact with CNOT1 and its interaction is mediated by CAF1 (516, 522–524). CNOT7 (CAF1) promotes metastasis in a manner which is dependent on both its deadenylase activity and interactions with CNOT1 and TOB1 (15). In another study, CNOT6 (CCR4a) and CNOT6L (CCR4b) were both required for cell proliferation and CNOT6L knockdown also decreased cell viability (525).

While the human NOT module promotes deadenylation by stabilising the complex, promotion of deadenylation by CNOT9 and CNOT10/11 *in vitro* involves direct binding of the RNA (31). CNOT9 also mediates interaction of CNOT4 with the rest of the complex in mammals (515). In yeast at least, Not4 or Not5 deletion causes hypoacetylation of histones H3 and H4 as well as reduced H3K4 tri-methylation (526, 527). Ubiquitylation of the eS7 ribosome subunit by Not4 is also required for downstream interaction of the Ccr4-Not complex with the ribosome in yeast (44). In addition, yeast Not4 indirectly promotes DNA damage-induced ubiquitylation of the Rbp1 subunit of Pol II following transcription arrest (43). Deletions of NOT module subunits in yeast produce more severe phenotypes than nuclease subunit deletions (308). Deletion of multiple subunits within either the NOT or nuclease modules is no worse than a single deletion, but combined deletion of components from both modules is lethal (308).

1.4.1.1 Deadenylase activity of CCR4-NOT

The deadenylase activity of both CCR4-NOT nuclease subunits was reported in yeast in 2001, though the importance of Pop2 (CAF1) *in vivo* is undetermined (460, 528–530). The importance of the rest of the CCR4-NOT complex for its deadenylase activity - outside of mediating specific targeting - was not made clear until recently. Raisch et al reconstituted the human CCR4-NOT complex and examined the contribution of each different module to deadenylase activity (31). Activity of the whole complex was greater than that of the exonuclease heterodimer alone, with

the minimum requirement for full activity being presence of the NOT module. Not only was deadenylase activity enhanced, but the specificity of the nucleases for the poly(A) tail was also increased when they were incorporated into the complex (31). In the context of the nuclease heterodimer (i.e. outside the whole complex) inactivating mutations showed that CCR4 is both more efficient and more specific for the poly(A) tail than CAF1. This at first seems to be in contrast to Yi et al. who showed that CAF1 was more active in the heterodimer than CCR4 (50), however this could be down to choice of CCR4 isoform since Raisch et al worked with CCR4a but Yi et al used CCR4b. In contrast to work showing that deadenylation could continue – albeit in some cases inefficiently – with only one active nuclease, an *in vitro* study of a human BTG2-CAF1-CCR4b trimeric nuclease showed that amino acid substitutions in *either* nuclease subunit completely inhibited deadenylation (531). While this result contradicts the two aforementioned studies, differences in substrate length and reaction conditions – such as pH and ratio of protein:RNA may be responsible. Cooperative action of the two nucleases was also suggested by a recent structural study of a recombinant human CCR4-CAF1 heterodimer (532), compatible with the observation that both nucleases are required for deadenylation (531). In this study, however, active site mutations in either nuclease in a CCR4-CAF1 heterodimer led only to slower deadenylation (at a 1:1 protein:RNA molar ratio) rather than complete loss of activity (532).

There is growing evidence of distinct differences in kinetics and substrate specificity between the two nucleases (28, 31, 50, 460, 530, 532). That said, moderate functional overlap seems to exist since both subunits are able to compensate to some degree for mutation in the other in yeast and cell culture, as measured by effect on bulk poly(A) length (28, 50). Notably, although median poly(A) length was unchanged in cultured cells with a single mutated nuclease, knockdown of CCR4, but not CAF1 produced a phasing pattern which the authors attribute slow removal of bound PABPCs (50). Recent work suggests that both nucleases can interact with PABPC, but where CCR4 is stimulated directly, CAF1 relies on bridging by TOB family proteins (28, 50, 440, 533). Assays with reconstituted *S. pombe* Ccr4-Not show conserved inhibition of Caf1 activity by Pab1 (PABPC orthologue) (28). Furthermore, isolated Caf1 nuclease - but not Ccr4 - was active in the absence of Pab1. Interestingly, Yi et al found by using immunopurified human subcomplexes, that at a

certain concentration of PABPC, CAF1 was not inhibited but CCR4 was still stimulated (50).

It was also shown that *S. pombe* Pab1, which includes 4 RNA recognition motifs (RRMs) can be dislodged from the tail one RRM at a time by Ccr4 (28). The authors suggest a model in which Ccr4 either forces Pab1 to peel away from the tail one RRM at a time, or pushes it along the tail, resulting in release of the 5'-most Pab1 first. The sliding model fits with an early observation that a single domain of *S. cerevisiae* Pab1 was sufficient for binding, with the authors suggesting this could allow Pab1 to 'transfer between poly(A) strands' (502). Yi et al propose a complimentary model in which CCR4 trims the PABPC-bound poly(A) segments and CAF1 degrades the stretches in between (50).

These differing reactions to PABPC may be important for determining the extent of deadenylation experienced by mRNAs since PABPC density varies between transcripts and exhibits weak positive correlation with mRNA stability and translation efficiency (28, 464). Indeed, Caf1 was more important for deadenylating mRNAs with low codon optimality, whereas Ccr4 was more general and showed no preference (28). Furthermore, Caf1 activity was undetectable if translation initiation was inhibited, whereas Ccr4 was unaffected. This is consistent with the observation that CNOT7 (CAF1) activity is stimulated by the DEAD-box helicase, DDX6 which has been shown to measure ribosome speed and thereby link translation rate with decay (301, 305). The above findings seem to naturally point to a model in which all transcripts are subject to deadenylation by CCR4, and inefficiently translated mRNAs are vulnerable to additional deadenylation by CAF1. The dependence of this additional deadenylation on translation may more easily allow for the storage of translationally dormant mRNAs without additional need for protection. As well as showing stimulation of CNOT7 (CAF1) by interaction of DDX6 with CNOT1, Meijer et al found that when eIF4A2 bound CNOT1 instead, CNOT7 activity was inhibited (305). Association of eIF4A2 with the CCR4-NOT complex may therefore be relevant to the storage of translationally dormant mRNAs.

There remains the question of whether CCR4 is itself constrained by additional regulation *in vivo* since many of its presumed substrates exhibit high stability

despite short tails (28, 326). One possibility is the presence of additional PABPs which prevent deadenylation.

1.4.1.2 Specific recruitment of CCR4-NOT

Specific recruitment of CCR4-NOT to mRNAs is well documented, including in miRNA mediated decay and nonsense mediated decay (30, 40, 45, 48, 51, 254, 256, 289, 292). Examples of specific CCR4-NOT recruitment to translation-competent mRNAs are depicted in figure 1.2. Of relevance to the serum response is targeting of the complex to mRNAs containing AU-rich elements (AREs) in their 3'UTRs via interaction of CNOT1 and CNOT9 with Tristetraprolin (TTP, encoded by *ZFP36*) (51, 292). HuR on the other hand, is thought to stabilise ARE-containing mRNAs by inhibiting degradation but not deadenylation (452). AREs can also be bound by KSRP, but the containing mRNAs are degraded by PARN and the exosome rather that via a CCR4-NOT mediated pathway (253, 534). CCR4-NOT can be targeted to a group of ~50 mRNAs enriched for developmental and inflammatory genes via recognition of a stem-loop constitutive decay element in their 3'UTRs by Roquin (45). In Drosophila at least, Roquin interacts via the Cnot9 subunit (535). In addition, the human Pumilio proteins can recruit CCR4-NOT, via CAF1, to a conserved UGUANAUA element in the 3'UTR of target mRNAs, including those involved in developmental signalling, fatty acid metabolism and transcription regulation (254, 536). The RNA binding protein TDP-43, which has roles in splicing, also recruits CCR4-NOT via the CAF1 subunit to destabilise mRNAs containing UG repeats in their 3'UTRs (537, 538). In vertebrates, Nanos recruits CCR4-NOT to Pumilio-bound mRNAs through a direct interaction with CNOT1, whereas in Drosophila the C-terminal domains of Not1-3 are involved (48, 539). The human NOT module also interacts in vitro with the HELZ helicase which may target a subset of mRNAs involved in nervous system development for translation repression or decay (30).

As well as being recruited to conserved sequence elements, the CCR4-NOT complex can be targeted to transcripts by miRNAs, which form ~6 bp matches with their target mRNAs. Translational repression and destabilisation by miRNAs relies on the incorporation of the mature miRNA into the RISC complex through association with the Argonaute (AGO) endonuclease subunit (258, 273, 274, 540, 541). Another RISC

subunit, GW182 recruits CCR4-NOT via an interaction with CNOT9 to promote deadenylation (542, 543).

In addition, presence of m^6A in the mRNA body can promote CCR4-NOT association and transcript degradation via interactions between CNOT1 and the m^6A reader YTHDF2 (47).

Although CCR4-NOT recruitment is often synonymous with mRNA degradation, it can in some cases lead only to translational silencing. The eIF4E-binding protein, 4E-T is enriched in P bodies (cytoplasmic decay hubs) and recruits CCR4-NOT to deadenylate ARE-containing mRNAs or miRNA targets (274, 544). 4E-T then redirects these mRNAs from decay to storage (274).

1.4.1.3 CCR4-NOT in other processes

CCR4-NOT has been historically implicated in both positive and negative transcriptional regulation (35, 36, 42, 147, 510, 526, 527, 545–551). It acts both at transcription initiation, either by direct binding of the promoter, interaction with transcription factors or influencing histone modifications, and in transcription elongation by rescuing backtracked Pol II. ChIP-Seq revealed that Pop2 (CAF1 homologue) was enriched across promoters, while Ccr4, Not5 and particularly Not3, were detectable in gene bodies but their effects were not investigated (552). Ccr4-Not was also suggested – under control of the Ras/cAMP pathway - to specifically repress Msn2/4p-dependent transcription of STRE (stress response element)-containing genes in yeast (510, 547). In human cells, CNOT2 and CNOT9 could both repress a Pol II promoter when fused to the Gal4 DBD (35). CNOT2 also repressed de novo transcription of MHC class II genes when tethered to a regulatory region (42). CNOT9 interacted with MYB to repress transcription of its target genes (546). CNOT1 meanwhile, could repress ligand-dependent transcriptional activation by oestrogen receptor (ER) α via a direct interaction in the presence of estrogen, as well as retinoic acid receptor (RXR)-mediated transcription (36).

As well as affecting specific promoters, CCR4-NOT affects recruitment of general transcription factors (547, 548). Deletion of *not5* in yeast caused enhanced presence of general transcription factor complex, TFIID components on heat shock promoters and a reduction on ribosomal protein promoters in unstimulated conditions (547).

Following heat shock, TFIID presence on ribosomal protein promoters was equal in WT and *Δnot5* cells, but was enhanced on heat shock promoters in the *Δnot5* strain. A mutation in the *not1* gene caused transcriptional activation which was associated with increased promoter binding by TFIID and SAGA (transcriptional coactivator) (548). In addition, some transcription factors/regulators are subject to the ubiquitin ligase activity of Not4 (308). Furthermore, both H3 and H4 acetylation as well as H3K4 tri-methylation are reduced in *Δnot4* or *Δnot5* yeast cells (527). This is thought to be due to interactions with the histone acetyltransferase, Gcn5 and the demethylase, Jhd2. In particular, Jhd2 is a target of Not4 ubiquitin ligase activity and its inhibited degradation in *Δnot4* cells presumably leads to continued demethylation of histone H3 (308, 526, 553). Overall, this suggests a general repressive effect of CCR4-NOT on transcription initiation.

In contrast, CCR4-NOT is thought to stimulate transcription elongation by rescuing backtracked Pol II in a mechanism distinct from that used by the elongation factor, TFIIS. Backtracked Pol II can occur for several reasons, including encountering roadblocks such as bound proteins, or incorrect nucleotide incorporation (144). Elongation by Pol II *in vitro* using yeast Pol II and a synthetic template was enhanced 1.5 – 2 fold by Ccr4-Not (147). Notably, transcriptional arrest by addition of a chain terminator could not be resolved by presence of Ccr4-Not, suggesting that rescue is not mediated by stimulating the intrinsic nuclease activity of Pol II as does TFIIS (147). Ccr4-Not interacts with the Rpb4/7 subunit of Pol II, possibly via Not5 since this subunit was separately found to bind Rpb4 (308, 554, 555). In a perhaps unrelated pathway, Ccr4-Not can also promote ubiquitylation of Pol II which has been stalled for too long and needs removing. Rather than using the ubiquitin ligase activity of Not4, the complex instead promotes recruitment of the E3 ubiquitin ligase Rsp5 (43). As well as aiding in transcription, interaction of Ccr4-Not with the transcription elongation complex may be important for mediating transcription-coupled DNA repair (39).

In addition, a growing number of links between CCR4-NOT and translation are being established (44, 307, 554, 556). Most recently, yeast Ccr4-Not was found to monitor the ribosome for codon optimality, eliciting degradation of the mRNA if the acceptor site is vacant for too long (44). Interaction with the ribosome is mediated by Not5,

but the complex's initial recruitment depends on Not4, which ubiquitylates the small subunit protein eS7. An earlier observation was that Not5 promotes translation of certain mRNAs, particularly those encoding ribosomal proteins (554, 557). Despite involvement of a common subunit, there is no immediately obvious mechanism which could link the two findings.

1.4.1.4 Regulation of CCR4-NOT

Given its involvement in so many aspects of gene expression, CCR4-NOT is -perhaps uniquely – poised to elicit widespread changes within the cell. Indeed, the complex has previously been implicated in nutrient sensing in yeast (510, 558–560); in conditions of low energy levels of Ccr4-Not can theoretically be reduced in order to elicit both slower transcription and slower degradation. Evidence of such a system is beginning to come to light in mice, where CNOT3 protein (but not mRNA) levels are downregulated in the liver and white adipose tissue after 24 hours of fasting (60). mTORC1 inhibition using rapamycin also caused a reduction in CNOT1 and CNOT3 protein levels in primary mouse cells, along with an increase in CNOT6L but not CNOT6 protein (519). Unlike in the fasted mice, this was accompanied by a reduction in Cnot1 and Cnot3 mRNA (as well as reduced Cnot6L but not Cnot6 mRNA).

Regulation of CCR4-NOT levels does not seem to be limited to nutrient sensing; the complex may also be leveraged to effect global changes to mRNA turnover in dynamic processes such as in development and B cell activation (12, 46, 58). This involvement in dynamic processes which require tight regulation and thus involve unstable transcripts is in keeping with the previous observation that STRE-containing mRNAs in yeast are more likely to be transcriptionally affected by Ccr4-Not subunits (547).

CCR4-NOT may also be regulated by changes to its localisation and post-translational modifications. Human CAF1 for example is nuclear during G0 and G1, but is mostly cytoplasmic in S phase (561). The yeast orthologue to CAF1, Pop2, is phosphorylated at Ser39 in glucose-rich conditions, causing repression of mRNAs encoding the small heat shock proteins Hsp12 and Hsp26 (560). In contrast, phosphorylation of Pop2 at Thr97 occurred within 2 minutes of glucose removal,

and substitution of Thr97 with alanine caused a failure of the cells to stop at G1 following glucose deprivation (558).

1.4.2 PAN2-PAN3

PAN2-PAN3, which is thought to trim long cytoplasmic poly(A) tails before CCR4-NOT takes over, is heterotrimeric, containing a single catalytic PAN2 along with an asymmetric PAN3 dimer (24, 562). It is thought to be mainly cytoplasmic but can cycle into the nucleus (26). PAN3 mediates interaction with PABPC, which stimulates PAN2 deadenylase activity (24, 563–565). As well as enhancing activity, interaction with PAN3 may improve the nuclease's poly(A) specificity (24). PAN2-PAN3 has a preference for long poly(A) tails since these in theory bind multiple PABPCs (26). Adjacent PABPC binding was suggested to coerce the poly(A) tail into forming multiple arches which are required for recognition of the substrate by PAN2-PAN3, since the complex itself has no nucleotide specificity (442). Alternatively, PAN2-PAN3 may recognise a helical structure with 9 nucleotides per turn which is formed uniquely by extended poly(A) tracts in solution (408, 566). Consistent with a requirement for PABPC binding, PAN (PAN2-PAN3) in yeast extracts did not efficiently remove the final 10-25 adenosines of an exogenous RNA substrate (567).

It is difficult to identify specific mRNA targets of PAN2-PAN3 since much of its activity can apparently be compensated - presumably by CCR4-NOT. Indeed, knockdown of PAN2/3 in human cells had no detectable effect on bulk poly(A) length as measured by TAIL-Seq (50). In yeast, *pan2* deletion results in increased bulk poly(A) length but does not impact viability (459). Although usually distributive, PAN2-PAN3 can act processively on the tails of specific transcripts and one could speculate that this could be due to differences in PABPC density (561, figure 4). Like CCR4-NOT, PAN2-PAN3 can be recruited by GW182 during miRNA-mediated degradation (568). Recruitment to mRNAs may also depend on PAN3 phosphorylation status which affect its interaction with PABPC (569).

1.4.3 PARN

PARN exists as a homodimer (570). Its main function is in maturation of the telomerase RNA component, with mutations in the PARN gene leading to the telomere diseases idiopathic pulmonary fibrosis and dyskeratosis congenita (571).

PARN targets are more limited than either PAN2-PAN3 or CCR4-NOT and is also able to bind the 5’ cap, an interaction which enhances its deadenylase activity (572). Exemplifying this limited pool of mRNA targets, knockdown of Parn in mouse myoblasts led to a stabilisation of only 40 mRNAs in a high throughput study (504). During Xenopus oocyte maturation, PARN is also thought to contribute to deadenylation-mediated translational silencing of maternal mRNAs (6). PARN can be targeted to ARE-containing mRNAs by KSRP and elicit degradation by either XRN1 or the RNA exosome (253, 534). Outside of mRNA deadenylation, PARN also functions as a nuclear deadenylase in maturation of small non-coding RNA species (573, 574). Included in its nuclear RNA targets are human telomerase RNAs (hTR), whose maturation require deadenylation by PARN (575). In addition, PARN is required for stabilisation of several mature and precursor miRNAs (576).

1.5 Crosstalk and mRNA homeostasis

Cells are exposed to frequently changing environments and must, for organisms to thrive, be able to respond effectively with minimal energy cost. Response to the environment and avoidance of disease requires tight regulation of gene expression, which is achieved by controlling synthesis and degradation of both mRNA and protein. Growing evidence is emerging of crosstalk between these processes (58, 148, 271, 415, 416, 470, 471, 577–586) and a summary of suggested crosstalk mechanisms is depicted in figure 1.5. This crosstalk may provide safeguards against inappropriate or excessive protein production or enable large and rapid gene expression changes through concerted regulation at multiple levels.

In order to execute a short-lived response to stimuli, some mRNAs are rapidly transcribed, but are unstable to avoid their inappropriate persistence (577, 584, 587). While some of these transcripts contain destabilising elements in their 3’UTRs (such as the ARE), there is also evidence that the promoter can influence mRNA stability (58, 251, 255, 256, 260, 261, 577, 578). Since the promoter sequence does not accompany the mRNA to the cytoplasm, it has been suggested that crosstalk is mediated by proteins being recruited to the promoter, enhancing transcription, then accompanying the mRNA to the cytoplasm and promoting degradation (577). The mitotic exit network protein Dbf2 is recruited co-transcriptionally to some cell cycle mRNAs and promotes their rapid decay during mitosis (578).

In mammalian cells, links between the promoter and mRNA stability were also established, however, in this case 'stronger' promoters were correlated with more stable transcripts (58). While this may reflect differences between species, it is also possible that the effect is not consistent across all genes, cell types or physiological conditions. In the mammalian study, the proposed mechanism did not depend on continued association with a co-transcriptionally recruited protein, but instead on variation in m^6A deposition. It was suggested that slow transcription leads to increased m^6A deposition and consequent increased susceptibility to deadenylation by CCR4-NOT (58). As well as being affected by the promoter, Pol II speed was reduced as it transcribed through IRES elements and this was sufficient to enhance m^6A deposition in the adjacent region, as well as cause shorter poly(A) tails. In an earlier publication from the same group, higher m^6A deposition caused by slower or paused Pol II also resulted in reduced translation efficiency (588).

Another factor which was recently suggested to couple transcription rate with mRNA stability in human cells is PHF3 (589). *PHF3* knockout cells, or those where PHF3 lacked its main Pol II-interacting domain, led to a bulk increase in mRNA level (via stabilisation) along with increased transcription of neuronal transcripts. In these cells the CCR4-NOT complex exhibited reduced association with Pol II, indicating that the usual destabilising effects of PHF3 may be mediated through CCR4-NOT. Increased transcription of neuronal transcripts was suggested to be due to reduced inhibition of TFIIS-mediated rescue of backtracked Pol II, since PHF3 could displace TFIIS in *in vitro* binding assays (589).

Crosstalk is not limited to gene-specific coupling of transcription and mRNA stability: inhibition of mRNA synthesis by multiple approaches in yeast resulted in widespread mRNA stabilisation (579, 583). Similarly, deletion of *nrg1,* which encodes a transcriptional repressor*,* resulted in global increases in both synthesis and decay rates (579). It has also been suggested that the Pol II subunits Rbp4 and Rbp7 remain attached to nascent transcripts and promote translation by interacting with eIF3 (590). Similarly, the Ccr4-Not complex may be recruited to a subset of mRNAs during transcription via its Not5 subunit, then remain attached in the cytoplasm and promote translation (557). Notably, transcriptional stress led to Not5-dependent enhanced translation. Since the Ccr4-Not complex associates with Pol II via its

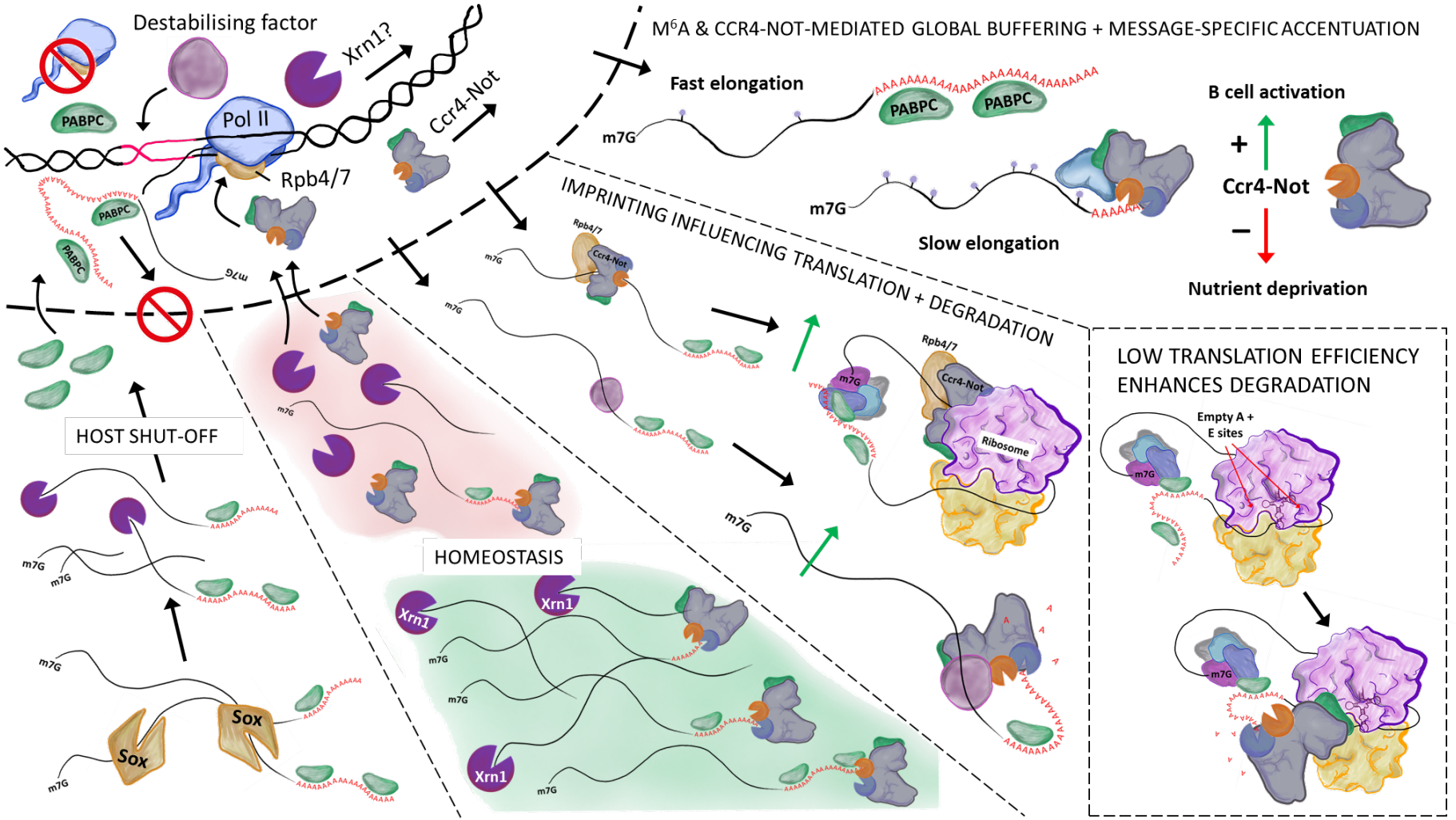


Figure 1.5 Overview of suggested crosstalk linking transcription, mRNA turnover and translation.

Rpb4/7 subunits to promote elongation it seems possible that their 'imprinting' of mRNAs is linked, at least in yeast (555).

In mammalian cells, prolonged impediment of Pol II elongation led to mass mRNA stabilisation despite the presence of short poly(A) tails and enhanced m^6A deposition (58). This general stabilisation was accompanied by a reduction in mRNAs encoding the degradation machinery, but it is not clear whether this reduction was a specific or a more widespread phenomenon as only one other mRNA was assayed.

In addition to 'forwards' communication, cells can alter transcription in response to changes in mRNA degradation and translation, and translation dynamics themselves can also influence mRNA stability (28, 44, 148, 269, 301, 579–581). In yeast, individual deletions of various decay machinery were accompanied by decreases in mRNA synthesis rate such that only minimal increases in global mRNA level were observed (579). The authors suggest that downregulation of degradation causes stabilisation of the mRNA encoding the transcription repressor Nrg1, which then enacts global repression of mRNA synthesis. The ability of cells to buffer transcript levels was abrogated following deletion of *xrn1*, which caused a 3.2-fold increase in global mRNA level, due to both a decreased decay rate and an increased synthesis rate. Though Xrn1 could therefore be suggested to directly repress transcription, in vitro transcription assays suggest this is not the case since addition of purified Xrn1 did not affect transcription activity (579).

A different group found that deletion of *xrn1*, like other decay factors, did result in decreased mRNA synthesis (580). Intriguingly, the same study found association of Xrn1 (and Lsm1 and Dcp2) with chromatin just upstream of TSSs, suggesting that Xrn1 could in some cases have a direct transcriptional role (580). A subsequent study by the same group showed that disruption of Xrn1 activity disproportionately affected the synthesis and decay rates of highly transcribed mRNAs such as those encoding ribosomal proteins, leading to an overall reduction in their mRNA levels (581). A direct nuclear role for Xrn1 is appealing since it allows easy detection of cytoplasmic degradation rate/mRNA levels: in the case of low Xrn1 engagement, excess Xrn1 is cycled back to the nucleus and promotes transcription. More recent work has suggested that both Xrn1 and Ccr4-Not promote elongation by Pol II, and notably only *ccr4* deletion inhibited transcription of ribosomal protein mRNAs (148).

In mammals, the extent of transcript buffering is less clear. While levels of CCR4-NOT may be altered by the cell to effect global changes in mRNA stability, knockout of *Cnot1* in mouse livers was not accompanied by obvious widespread transcript buffering (58, 60, 591). Although synthesis of mRNAs encoding metabolic enzymes were reduced, pre-mRNA levels of apoptosis-related and inflammatory genes increased (591). In addition, levels of other transiently expressed mRNAs were increased, suggesting that no concomitant decrease in synthesis occurred. It is difficult to draw a firm conclusion given that deletion of *Cnot1* in the liver may also have caused changes to cell composition.

While the effects of inhibited mRNA degradation have been thoroughly investigated in yeast, some studies of crosstalk in mammals have instead focused on promoting cytoplasmic mRNA degradation. This work largely originates from the Glaunsinger lab and involves expression of the gamma-herpesvirus SOX endonuclease (271, 415, 416, 470, 471). SOX cleaves host mRNAs, providing access points for degradation by XRN1, and this increased degradation is accompanied by a reduction in Pol II transcription as well as the accumulation of hyperadenylated transcripts and PABPC in the nucleus (415, 470, 471). Reduced Pol II recruitment in the presence of SOX was dependent on XRN1 catalytic activity, but immunofluorescent detection showed no change in XRN1 localisation, and XRN1 could also not be detected by chromatin immunoprecipitation at the affected promoters (471). Subsequent work showed that enhanced degradation instead released PABPC from poly(A) tails, and that high nuclear PABPC concentrations inhibited recruitment of Pol II and TATA binding proteins to promoters (470). UV irradiation also causes PABPC nuclear relocalisation and reduced mRNA synthesis, though in this case transcriptional effects are caused by phosphorylation and ubiquitination of Pol II (417, 592).

It is difficult to determine whether the apparent difference in endgame between yeast and mammalian cells represents a lack of conservation across species, or simply differences in experimental set up. It is possible that increased and decreased mRNA degradation are indicative of distinct stresses which both require downregulated transcription. Reduced degradation arising due to nutrient deprivation and downregulated CCR4-NOT should quite intuitively be accompanied by reduced transcription in order to minimise energy usage. On the other hand, in

the case of viral infection or DNA damage, it seems logical for cells to restrict available host machinery/limit growth for the benefit of the wider organism or colony.

A further example of crosstalk in yeast is the accelerated deadenylation and decay observed for inefficiently translated mRNAs (28, 44, 301). This link seems to be communicated by interaction of the ribosome with degradation machinery, in particular CCR4-NOT and DDX6 (Dhh1 in yeast). Deletion of *caf1* preferentially increased stability of mRNAs with low codon optimality, suggesting that Caf1 normally acts on poorly translated mRNAs (28). Importantly, the difference in half-life between codon optimal and non-optimal transcripts was abrogated upon translation inhibition, suggesting that Caf1 deadenylase activity is translation dependent. The decapping activator Dhh1 (DDX6 in mammals) also acts in a translation dependent manner to promote degradation of poorly translated mRNAs (301). Although DDX6 was recently shown to enhance CNOT7 activity, presence of Dhh1 was not required for the differences in deadenylation observed between codon optimal, and non-optimal mRNAs (28, 305). Indeed, Ccr4-Not may be recruited to inefficiently translated mRNAs independently of Dhh1 via interaction of Not5 with the empty ribosome E site (44).

Overall, this extensive crosstalk seems well placed both to enact mRNA homeostasis during reduced nutrition and on the flipside, to efficiently amplify transcriptional responses to stimuli by concurrent optimisation of downstream mRNA stability and translatability (58, 60, 510, 558–560, 588, 591, 593). Furthermore, the link between transcription rate and stability may serve as a further backstop to limit expression of mRNAs derived from 'leaky' transcription. Varied reports regarding feedback to the nucleus about cytoplasmic degradation rates suggest that several distinct mechanisms may be in play depending on the species and/or experimental conditions (148, 271, 415, 470, 471, 579–581). Near-central nodes in this cross talk seem to be CCR4-NOT and the poly(A) tail (28, 44, 58, 60, 148, 415, 510, 557–560, 591).

1.6 Aims of study

Given the abundant evidence of sub-200 nt length poly(A) tails and of extensive interactions between mRNA synthesis and degradation, this study aims to

investigate the existence of regulated nuclear polyadenylation, and the relationships this may have with mRNA turnover.

- First, using the well-characterised NIH 3T3 serum response, a method for precise tail length measurement (the poly(A) tail test, PAT) is validated and used to determine poly(A) tail lengths of mRNAs in different cellular compartments. Using the PAT assay, poly(A) tail lengths of certain mRNAs are shown to change throughout the serum response, with corresponding changes in mRNA stability.
- The potential of different deadenylases to affect early poly(A) length is then investigated, and the CCR4-NOT scaffold subunit, CNOT1 emerges as a regulator of early poly(A) tail length.
- Given the link between poly(A) length and stability, the effect of Cnot1 knockdown on mRNA abundance is investigated, and CNOT1 is found to also influence pre-mRNA level.
- The mechanism by which CNOT1 exerts control over both early poly(A) tail length and pre-mRNA level is probed further by additional siRNA knockdowns and small molecule inhibition of the CAF1 subunit.
- Finally, a new method for global poly(A) tail length measurement, PAT-Quant-Seq (PQ-Seq) is presented and is used to identify tail length regulation in the nuclear and cytoplasmic fractions of NIH 3T3 cells early and late in the serum response.

2 Materials and methods

2.1 Cell culture and treatment

2.1.1 Cell culture

NIH 3T3 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Sigma, 6429) supplemented with 10 % Newborn Calf Serum (NCS) (mainly HyClone, SH30401.01, Batch: DXJ0439, but also Gibco, 26010-074, Batch:1167012). HEK293 cells were cultured in DMEM (Sigma, 6429) supplemented with 10 % Foetal Bovine Serum (FBS) (Gibco, 10500-064). Cells were passed upon reaching 80 % confluency at ratios between 1:4 and 1:10. Cells were released from the growth surface using 1X Trypsin EDTA in PBS and gentle rocking by hand at 25 °C for 2-5 minutes. Two starting cultures of NIH 3T3 cells were used and serum response behaviour in the newer cells (purchased from the European Collection of Authenticated Cell Cultures) validated against the existing culture. NIH 3T3 experiments were performed between passages 5 and 30, numbering from receipt of the cells by our lab. HEK293 experiments were performed on passage 12-15 cells, continuing numbering from the p10 cells gifted to us by the Bushell lab (540). All cultures were maintained at 37 °C and 5 % CO_2.

2.1.2 Cryostorage

Cells were trypsinised and collected in 10 % serum DMEM, then centrifuged at 350 g for 5 minutes to pellet cells. Supernatant was removed, and the pellet resuspended in pre-chilled serum (NCS or FBS depending on cell line) containing 10 % DMSO. Suspended cells were aliquoted into cryovials and placed immediately on ice, then transferred quickly to a -80 °C freezer. After 1 day but within 7 days, vials were transferred to liquid nitrogen. Passage number on the tube was recorded as current passage + 1.

Cells were thawed in T75 flasks. 10 % serum DMEM was pre-warmed, and 9 mL transferred in advance to a 15 mL falcon tube. Vials were thawed quickly using either a 37 °C water bath or gloved hands and then transferred to the warm media in the falcon tube. Cells were centrifuged at 350 g for 5 minutes and the pellet resuspended in warm 10 % serum DMEM. Cells were checked the following day and media was changed if there were many dead/floating cells.

2.1.3 Serum stimulation

NIH 3T3 cells were seeded to reach ~80 % confluency (~40,000 cells/cm^2) 24 hours prior to serum stimulation. Seeding was carried out 24 (or a multiple of 24) hours prior to starvation to avoid potential unintended effects relating to circadian rhythm. The exact starting density varied slightly depending on culture vessel as doubling time did not seem to be completely consistent across all well/dish sizes. In particular, seeding density for immunofluorescence had to be optimised separately and is detailed in section 1.7.1. Cells then underwent serum starvation by removing the 10 % NCS media, washing once (gently) with PBS, then replacing with DMEM supplemented with 0.5 % NCS. The serum response was induced 24 hours later by adding pre-warmed NCS to 10 %. Depending on the culture dish set up and duration of downstream sample processing, serum stimulation was sometimes staggered by up to 100 minutes (in the case of plates containing both 20' and 120' serum stimulated samples). Stimulation was terminated by placing cells on ice and immediately removing media, followed by washing briefly with ice-cold PBS.

2.1.4 siRNA knockdown

NIH 3T3 cells were seeded at a density of between 0.61 and 0.85 x10^4 cells/cm^2 in 10 % NCS DMEM. This varied a little depending on a combination of dish/well size, anticipated time between seeding and transfection. There were also small differences in doubling time between batches of cells – perhaps due to passage number or possibly to serum batch number (though neither was investigated).
24 and 48 hours after seeding, siRNA (Dharmacon, ONTARGETplus SMARTPOOL) was transfected into cells to 10 nM final concentration using Lipofectamine RNAiMax (Invitrogen, 13778075) and complexed in Opti-MEM (Gibco, 31985062). Sequences of siRNAs are provided in table 2.1. Just before transfection, 10 % NCS DMEM was refreshed as this seemed to enhance transfection efficiency. Cells were starved as in the serum stimulation protocol 24 hours after the second transfection.

HEK293 cells underwent the same siRNA treatment (but with siRNAs directed against human mRNAs), however they were seeded at a density of 1.63 x 10^4 cells/cm^2 in 10 % FBS DMEM and were analysed at steady state without serum starvation or stimulation. 10 % FBS media was similarly refreshed directly before each transfection.

Dharmacon code	Target mRNA	Target sequence
D-001810-01	Non-targeting	Not disclosed
D-001810-02		Not disclosed
D-001810-03		Not disclosed
D-001810-04		Not disclosed
J-167425-05	Cnot1 (mouse)	CAAUAAGGUUCUCGGUAUA
J-167425-06		UGUUAGAGGCUUACGUUAA
J-167425-07		CCUACCAAGCCGAGCGGAU
J-167425-08		CUAUAUGGCCGUGGCGUUU
J-059073-05	Cnot7 (mouse)	GAUCAUAGCCAAAGAAUUU
J-059073-06		CUAUAGAGCUACUAACAAC
J-059073-07		GGACUGACCUUUAUGAAUG
J-059073-08		CAAAUACUGUGGUCACUUA
J-063124-05	Cnot8 (mouse)	UUUCGAAGCUCCAUAGAUU
J-063124-06		GGAGGAAGGGAUCGAUACA
J-063124-07		GUUCCCAGGUGUUGUUGUA
J-063124-08		GAUAUGUACUCCCAGGAUU
J-062986-09	Nxf1 (mouse)	GAUCAUGAGCAAACGAUAU
J-062986-10		AUGAAUUAUUUGUGCGGAA
J-062986-11		CUUCAAUCCCAUCGAGUUU
J-062986-12		UUGAAGACUGAGCGGGAAU
J-041138-05	Pan2 (mouse)	CAUCAAAUAUUCCAAGCUA
J-041138-06		GUGGAGGGCUCAUUAUAUU
J-041138-07		GUGCAUCUCUGGACUGAUU
J-041138-08		CGAAUGAUUUCCCUACGAU
J-040664-09	Parn (mouse)	UGAGAGAGCAAUUCGAUGA
J-040664-10		AAGUACAAAUUGCCGUUAA
J-040664-11		GUAGACACUUAACCAGAUA
J-040664-12		GCAAGUACGCUGAAAGUUA
J-041045-05	Zfp36 (mouse)	CCGAAGCUGUGGCUGGGUA
J-041045-06		UGUCGGACCUACUCAGAAA
J-041045-07		CGGAGGACUUUGGAACAUA
J-041045-08		AAACGGAACUCUGCCACAA
J-041703-05	Zfp36L1 (mouse)	UCAGCAGCCUUAAGGGUGA
J-041703-06		UCAAGACGCCUGCCCAUUU
J-041703-07		GGAGCUGGCGAGCCUCUUU
J-041703-08		CGAAUCCCCUCACAUGUUU
J-043119-05	Zfp36L2 (mouse)	UAACAACGCCUUCGCUUUC
J-043119-06		UGGCAAACCUCAAUCUGAA
J-043119-07		GCAAGUACGGCGAGAAGUG
J-043119-08		GAUAUCGACUUCUUGUGCA
J-015369-09	CNOT1 (human)	CUAUAAAGAGGGAACGAGA
J-015369-10		GGCCAAAUUGUCUCGAAUA
J-015369-11		CCAGAAACUUUGGCGACAA
J-015369-12		CAAGUUAGCACUAUGGUAA

Table 2.1 siRNA sequences

2.1.5 Caf1 inhibitor treatment

Cells were seeded to reach ~80 % confluency 24 hours prior to serum stimulation and then underwent starvation in 0.5 % NCS DMEM for 24 hours. Two hours before serum stimulation, Caf1 inhibitor 108 or 247 at a stock concentration of 10 mM in DMSO was added to the cells to a final concentration of 100 µM. An equal volume of DMSO was added to control wells. Other concentrations and incubation times were tested as indicated in the pilot experiment. Cells were stimulated with 10 % NCS for 60 minutes, since the poly(A) tails of mRNAs of interest are short at this time under control conditions. Inhibitor 108 has been characterised *in vitro* by Jadhav et al (2015) and Airhihen et al (2019) as compound 8j (594)/5 (595).

2.2 Subcellular fractionation

2.2.1 Nuclear/cytoplasmic fractionation

Buffer A: 10 mM Hepes pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1mM EGTA, 1 mM DTT, 0.5 mM PMSF.

Cells were placed on ice and washed twice briefly in ice-cold PBS. Cells were scraped and collected in PBS, then centrifuged at 700 g for 5 minutes. The pellet was resuspended in a smaller volume, moved to a 1.5 mL tube and spun for a further 5 minutes at 700 g. This pellet was suspended in 600 µL Buffer A per 15 cm dish and incubated on ice for 15 minutes. 37.5 µL/dish 10 % NP-40 was added and the tube vortexed then incubated for a further 3-5 minutes on ice to break the cell membrane. Centrifugation at 16,000 g was performed to pellet the nuclei. Around 50 % of the supernatant was removed and mixed with an equal volume of isopropanol and 1 µL GlycoBlue (Invitrogen, AM9515) to precipitate RNA from the cytoplasmic fraction. The nuclei were washed three times with 400 µL Buffer A then both the nuclei and the pelleted cytoplasmic RNA were mixed with BL+TG buffer (500 µL per 15 cm dish for nuclear fraction, 250 µL/dish for cytoplasmic) from the ReliaPrep RNA Cell Miniprep system (Promega, Z6012). Where required, nuclear suspensions were passed through a 21-gauge needle to break up the chromatin. The ReliaPrep kit was used both to isolate and DNase-treat the RNA. For protein isolation, supernatant was removed after nuclei were pelleted and used as the cytoplasmic fraction. Nuclei were resuspended in Buffer A following the three wash steps and used as the nuclear fraction. A small (and equal) volume of these samples

were used in Bradford assays and compared against BSA standards diluted containing an equal concentration of Buffer A. The remainder of each sample was diluted with 3X SDS loading buffer and stored at -20 °C. Despite efforts to thoroughly mix all lysates, subsequent western blots indicated that protein concentration of the nuclear fractions were often underestimated compared to cytoplasmic fractions, resulting in uneven gel loading. Additional western blots were therefore performed with loading volumes adjusted according to the original signal.

Since HEK293 cells do not spread out to the same extent as NIH 3T3 cells and therefore a larger number of cells grow on each plate, Buffer A and NP-40 volumes were increased to 1600 µL and 100 µL per 15 cm plate respectively.

2.2.2 Chromatin/nucleo-/cytoplasmic fractionation

Nuclei Resuspension buffer*:* 20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSF, 50 % Glycerol, 0.5 % GlycoBlue.

NUN buffer: 20 mM Hepes pH 7.6, 1 mM DTT, 7.5 mM MgCl2, 0.2 mM EDTA, 0.3 M NaCl, 1M Urea, 1 % NP-40, 0.5 % GlycoBlue.

Where a chromatin-associated fraction was also required, nuclei and a cytoplasmic fraction were isolated as above. Nuclei were resuspended in 40 µL Nuclei Resuspension buffer per 15 cm plate then fractionated further by adding 360 µL NUN buffer, vortexing, and incubating on ice for 5 minutes. Samples were centrifuged at 16,000 g for 3 minutes to collect the chromatin. 50 % of the supernatant was removed to form the nucleoplasmic fraction and RNA was precipitated by adding an equal volume of isopropanol and 1 µL GlycoBlue. The chromatin was washed three times in 500 µL 80 % ethanol. All 3 fractions were passed through the ReliaPrep RNA Cell Miniprep system following direct addition of BL+TG buffer to the washed chromatin pellet (500 µL BL+TG per 15 cm plate) and RNA pellets respectively (250 µL BL+TG per plate).

As in the nuclear/cytoplasmic fractionation for HEK293 cells, buffer volumes were increased 4-fold meaning that 160 µL Nuclei Resuspension Buffer and 1440 µL NUN buffer per 15 cm plate were used.

2.3 RNA isolation and treatments

2.3.1 RNA isolation from whole cells or cellular fractions

RNA was isolated from cells using the ReliaPrep RNA Cell Miniprep system (Promega, Z6012) according to the manufacturer's instructions, but with an extended DNase treatment step of 1 hour rather than 15 minutes. RNA was eluted in water into manufacturer-provided non-stick RNase-free tubes, quantified by nanodrop, where 260/280 ratios of 2.00-2.20 were considered good. Samples were not eliminated based on their 260/230 ratio, but the ratio was recorded in case of downstream problems. RNA was kept at -20 °C in the short term and moved to -80 °C for longer term storage.

2.3.2 RNA isolation after enzymatic treatments

A 25:24:1 phenol:chloroform:isoamyl (P:C:I) alcohol mixture was prepared using pH 4.5 citrate buffered phenol (Sigma-Aldrich, P4682), mixed well and allowed to separate overnight. After enzymatic treatment of RNA, an equal volume of P:C:I (25:24:1) was added to separate protein and RNA. RNA was isolated from the aqueous phase by ethanol precipitation (2.5 volumes ethanol, 0.1 volumes 3M sodium acetate pH 5.2, 1 µL GlycoBlue (Invitrogen, AM9515) if required). Depending on the downstream application, an additional chloroform wash of the aqueous phase was added prior to ethanol precipitation to remove any excess phenol that may inhibit reverse transcription.

2.3.3 DNase treatment

10X DNase I buffer: 10mM Tris HCl pH 7.5, 2.5mM $MgCl_2$, 0.5mM $CaCl_2$.

For applications in which trace genomic DNA was problematic, the standard on-column DNase treatment included in the ReliaPrep RNA isolation process was insufficient, even when increased to 1 hour. For these RNA samples, the on-column DNase treatment was carried out using twice the recommended volume of DNase mixture. For on-column DNase treatment of northern blot RNA, lyophilized DNase I (from the isolation kit, individual code Z358A-C) was reconstituted at 2X concentration but used at the recommended volume.

Eluted RNA underwent an additional DNase treatment step with Turbo DNase (Invitrogen, AM2238). For each 10 µg RNA, a 50 µL reaction was carried out using

1 µL DNase and homemade DNase I buffer. The reaction was incubated for 30 minutes at 37° C after which another 1 µL/10 µg RNA was added and the reactions incubated for a further 30 minutes (as recommended by the manufacturer for substantial DNA contamination). DNase was inactivated and the RNA isolated using Phenol:Chloroform extraction followed by ethanol precipitation.

2.3.4 RNase H treatment

RNA was mixed with specific oligo (table 2.2) and/or oligo dT (12) at a ratio of 5 µg RNA:4 µg oligo and heated to 90 °C for 5 minutes, then placed on ice. RNase H (5 U/5 µg RNA, NEB M0297 and Thermo Scientific EN0201), RNase H buffer and DTT where required were then added and incubated at 37 °C for 1 hour. Reaction was terminated by Phenol:Chloroform:Isoamyl (P:C:I) addition and RNA retrieved by ethanol precipitation as above. In order to remove any trace phenol, an additional chloroform wash (aqueous phase mixed with one volume chloroform) was sometimes performed prior to ethanol precipitation. For small volume RNase H reactions, water was added to increase volume prior to P:C:I addition and allow easier recovery of the aqueous phase.

Oligo name	Target mRNA	Sequence	Resulting fragment length
Egr1 168	Egr1	CCACAACACTCCAACTCCTG	168/193*
Rpl28 198	Rpl28	CATTCTTGTTGATGGT	198

Table 2.2 RNase H cleavage oligonucleotide sequences.

RNase H/Oligo(dT) treated RNA was also used to provide deadenylated controls for the RL2-PAT assay. Following poly(A) removal for 1 hour at 37 °C, an equal volume P:C:I was added and the RNA ethanol precipitated then fed into the PAT ligation and cDNA synthesis pipeline.

2.4 RNA detection

2.4.1 Agarose gel electrophoresis for RNA quality control

To check RNA integrity, RNA was separated by agarose gel electrophoresis. The highly abundant 28S and 18S ribosomal RNA bands were used to determine RNA integrity, with optimum quality appearing as a 28S band twice the intensity of 18S, and low background (representing little degradation). Gel tray and tanks were cleaned with a mild solution of Distel and rinsed thoroughly to remove any RNases,

then wiped with 70 % methylated spirits. 2 µL RNA samples were heated at 70 °C for 3 minutes with 10 µL deionised formamide loading dye containing bromophenol blue, xylene cyanol and 10 mM Tris pH 7.5. Samples were run at 80 V on a 1 % agarose TBE gel containing 1.5X SYBR safe (Invitrogen, S33102) for 1-2 hours and imaged using UV.

2.4.2 Glyoxal denaturing agarose gel

Glyoxal reaction mixture (to make 9.8 mL): 6 mL high grade DMSO, 2 mL molecular biology grade glyoxal solution (Sigma-Aldrich, 50649), 1.2 mL 10X BPTE buffer, 0.6 mL 80 % glycerol (in RNase-free water).

10 X BPTE buffer: 100 mM Piperazine-N,N'- bis(2-ethanesulfonic Acid) (PIPES), 300 mM Bis-Tris, 10mM EDTA pH 8.0. Incubated overnight with 0.1 % DEPC (Sigma-Aldrich, D5758) and autoclaved to remove RNases.

In order to directly and accurately measure poly(A) tail lengths of RNA, samples were treated with the denaturing agent glyoxal which forms adducts with RNA that are stable at pH <7. Up to 15 µg RNA in 4 µL H_20 was mixed with 20 µL glyoxal reaction mixture and incubated at 55 °C for 1 hour. Reactions were chilled on ice water for 10 minutes then mixed with 4 µL RNA gel loading buffer and run on a 20 cm long 1.8 – 2 % agarose BPTE gel at 120 V (~3.6 V/cm). Buffer was recirculated to avoid accumulation of buffer at pH >7 and consequent reforming of RNA secondary structure. This recirculation was originally performed manually every 30 minutes with a 60 mL syringe, but in later experiments was recirculated constantly using a peristaltic pump (Watson Marlow) to achieve a rate of 1.98 L/hr in a total volume of ~ 2.5 litres). Sample RNA was run alongside two different ssRNA markers (Low Range ssRNA Ladder - NEB, N0364S and RiboRuler Low Range RNA Ladder - Thermo Scientific, SM1831) which were precipitated and glyoxylated as above.

2.4.3 Northern blotting

dCTP labelling mix: 100 µm each dATP, dGTP, dTTP, 10 mM Tris pH 7.5.

Pre-hybridisation buffer: 0.5 M disodium phosphate pH 7.2, 1mM EDTA, 7 % (w/v) SDS, 1 % (w/v) BSA.

Phosphate wash buffer: 40 mM disodium phosphate pH 7.2, 1 % (w/v) SDS.

RNA was transferred from agarose gel to Hybond N neutral nylon membrane (GE Healthcare, RPN2020N) overnight via upward capillary transfer in 20X SSC then cross-linked using UV at 120 mJ/cm^2. To check transfer and visualise marker bands, membrane was stained with 0.02 % (w/v) methylene blue in 0.3 M sodium acetate pH 5.5 for around 5 minutes and excess dye washed off with nuclease-free water. Membranes were imaged with visible light, or photos taken with a phone camera

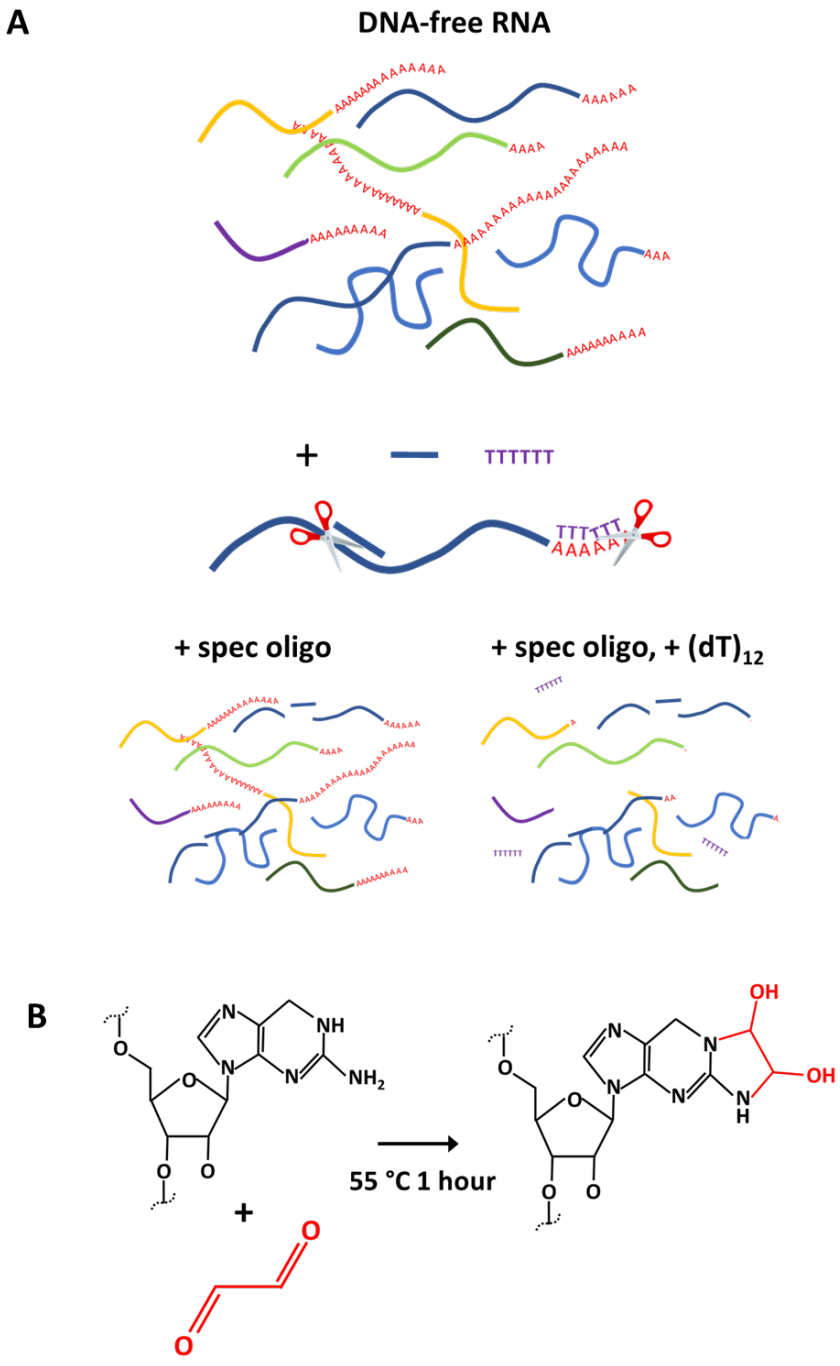


Figure 2.1 RNase H northern blot schematic. Continued overleaf.

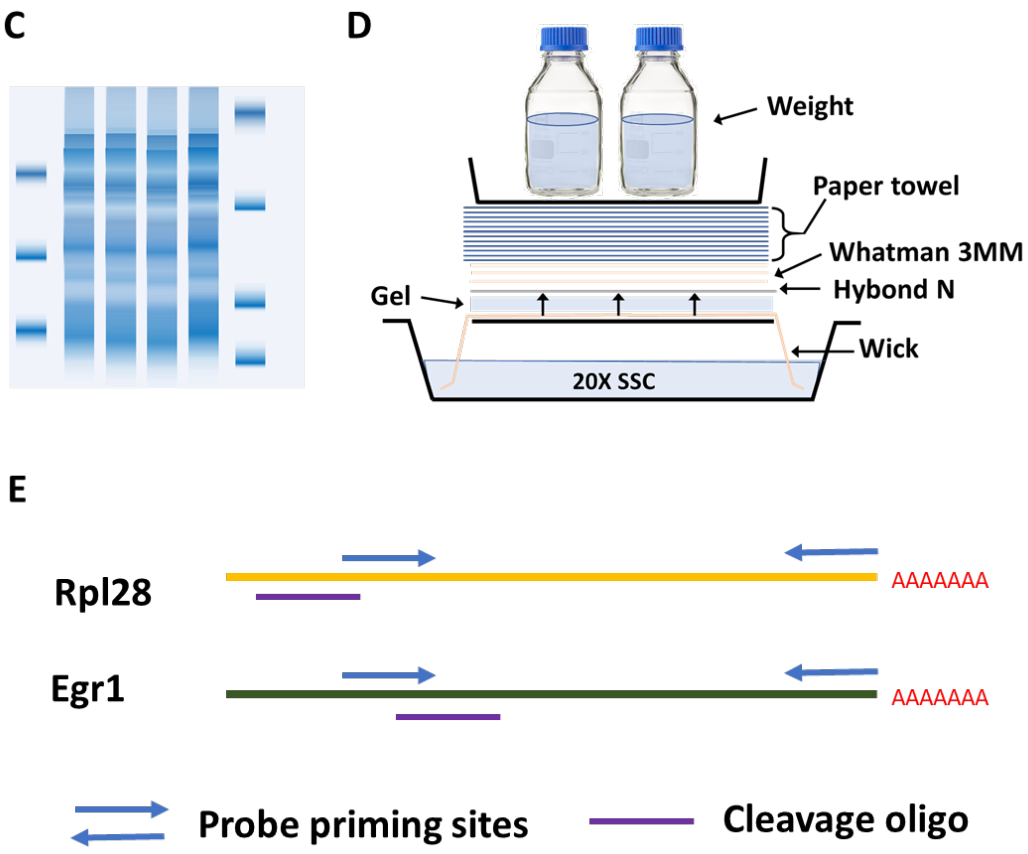


Figure 2.1 RNase H northern blot schematic. A) RNA was cleaved at sites of oligo binding by RNase H, targeting the 3' UTR of a specific mRNA and in a matched sample, the poly(A) tail as well. **B)** Glyoxal reacts with guanine (and to a lesser extent, adenine and cytidine), forming a cyclic structure and preventing base-pairing. Reaction summarised from (596). **C)** Glyoxylated RNA underwent agarose gel electrophoresis at pH < 7.0 alongside two ssRNA markers. **D)** RNA was transferred to Hybond N neutral nylon membrane via upwards capillary transfer with 20X SSC. **E)** Schematic of primer sites for probe template amplification and binding of the RNase H cleavage oligo for the two RNAs assayed.

then marker bands were traced with pencil for size determination later. The membrane was de-stained in nuclease-free water (though it was difficult to completely remove the dye) then heat sealed in plastic and stored at -20 °C until the hybridisation stage.

Probe template DNA was generated by PCR amplification of cDNA from a previous NIH 3T3 cell experiment using the primers detailed in table 2.3 and sequenced to

confirm the correct product. Where necessary, gel-isolated DNA underwent a second round of PCR to give a pure product of high enough yield.

Target mRNA	Name	Sequence	Tm (°C)	Fragment length
Egr1	Egr1 probe F Egr1 probe R	TATACATCTATTCAGGAGTTGGAGTG GAAGGATACACACCACATATCCC	57.82 58.36	200
Rpl28	Rpl28 Probe F Rpl28 Probe R	TGCTCGGGCTACCCTCAG GGAAGTCAGTGGACTCTTTATTGC	60.76 59.61	200

Table 2.3 Northern blot probe primer sequences.

Probes were labelled using the DNA polymerase I Klenow fragment (NEB, M0210L) with ^{32}P dCTP 3000Ci/mmol, 10 µCi/µL (Perkin Elmer, BLU013H250UC). 50 ng template DNA was mixed with 10 µL 1 mg/mL random hexamers in a 27 µL reaction and heated at 90 °C for 10 minutes. After placing on ice, 10 µL 5X dCTP labelling mix, 5 µL 10X Klenow buffer, 0.5 µL Klenow fragment (NEB, M0210S) and 7.5 µL α-^{32}P-dCTP was added and reactions incubated at 37 °C for 5 minutes. To separate labelled probe from free nucleic acids, reactions were separated by centrifugation using homemade 1 mL G50M spin columns. To assemble spin columns, plungers were removed from 1 mL syringes and rested in 15 mL falcon tubes. Autoclaved glass wool was added to the syringe and compressed using the plunger to a thickness of 3-5 mm. Autoclaved Sephadex beads in water were added, avoiding bubble formation, and centrifuged twice at 1500 rpm for 5 and then 10 minutes to remove excess water. Completed probe reaction was added to the prepared columns, followed by 50 µL nuclease-free water, and centrifuged at 1500 rpm for 10 minutes, resulting in elution of the probe, with any free label remaining in the column. Specific activity of probes was determined by measuring 1 µL (suspended in 10 mL EcoScint – Scientific Laboratory Supplies, NAT1392) in a scintillation counter.

Prior to hybridisation, membranes were incubated with 10-15 mL pre-hybridisation buffer and 200-300 µL freshly denatured salmon sperm DNA (UltraPure Salmon Sperm DNA Solution – Invitrogen, 15632011) for at least one hour at 65 °C to block non-specific binding. Probes were denatured at 90 °C for 10 minutes then added to between 0.5 and 2 million cpm/mL and incubated with the membranes at 65 °C

overnight in a hybridisation oven. The following morning, membranes were washed three times for 15 minutes at 65 °C in pre-heated phosphate wash buffer.

Washed blots were aligned next to paper strips on a clingfilm-covered cardboard backing board and lightly held in place using drops of wash buffer. Marker bands were indicated on the paper strips with marker pen soaked in hybridisation buffer or probe mix, taking care to avoid bleeding due to excess wash buffer. Blot and strips were covered in clingfilm and placed in a cassette to expose a (pre-erased) phosphor screen. For blots which required overnight exposure, the cassette was placed at -80 °C to reduce background excitation. Exposed screens were imaged using a Storm 825 system (GE Healthcare). In the event of high background, blots were washed twice more with phosphate wash buffer, placed on a clean backing-board and re-exposed.

2.5 cDNA synthesis and PCR

2.5.1 Quantitative real time (q-RT) PCR

Between 150 and 500 ng RNA was used for cDNA synthesis and the reaction scaled accordingly. For a 500 ng reaction, RNA was incubated for 5 minutes at 70 °C with 2 µL random primers (0.2 µg/µL, Invitrogen, 48190011) and 1 µL dNTPs (10 mM), and topped up to 14.5 µL with nuclease-free water. Mixtures were placed on ice for one minute, then 4 µL 5x First Strand Buffer, 1µL DTT (100 mM) and 0.5 µL SuperScript III (Invitrogen, 18080085) were added and the reactions incubated at 50 °C for one hour. The reaction was stopped by incubating at 70 °C for 15 minutes. The resulting cDNA was diluted 1 in 10 for use in qPCR unless RNA was of particularly low quality or abundance, in which case a lower dilution was used.

For each qPCR reaction, 0.5 µL each forward and reverse 20 µM primer, 2 µL H_20 and 5 µL 2X GoTaq qPCR Master Mix (Promega, A6002) were mixed with 2 µL diluted cDNA. Primer sequences can be found in table 2.4. Standard curves were performed for all new primer pairs (figure A.1). Reactions were carried out in a Qiagen Rotor-Gene Q , starting with an initial denaturation at 95 °C for 5 minutes, followed by 40 cycles of 10 sec 95 °C, 20 sec 60 °C, 20 sec 72 °C. Melt curves were generated between 72 and 95 °C. Three reactions were performed per sample.

Relative RNA abundance was calculated using the ΔΔCt method (597) using Gapdh as the reference gene unless otherwise stated. In knockdown time course experiments, all samples of a given biological replicate were normalised to the maximum value of the control set (usually 20' or 60') such that the effect of knockdown could best be compared between experiments. Unpaired t-tests were performed for each time point.

Mouse

Target RNA	Name	Sequence	Tm (°C)	Fragment length
Actb (mature)	Actb F	CTGTCGAGTCGCGTCCACCC	65.19	128
	Actb R	ACATGCCGGAGCCGTTGTCG	65.79	
Actb (unspliced)	un Actb2 Fw	AAGATCTGGCACCACACCTT	59.23	155
	un Actb2 Rv	TGAGAAGCTGGCCAAAGAGA	58.94	
Arl6ip5 (mature)	Arl6ip5 F	GGTTGTTGGGTTTCTGAGCCC	61.7	191
	Arl6ip5 R	TGACACCCCCGAACATGGAT	61.2	
Cnot1 (both)	Cnot1_all F	AGCATAGCAGCATGTCTTCC	58.04	147
	Cnot1_all R	GCAGAACCCCCTGGTCTATC	59.53	
Cnot7 (mature)	Cnot7 F	AATTTGACGGAGGACATGTATGC	59.37	158
	Cnot7 R	AGCCATTTGACCCCTTCACAA	60.13	
Cnot8 (mature)	Cnot8 F	TGGAGGCCTAAGGAATGATTGGAT	61.45	164
	Cnot8 R	CTGGCCCACACCTCACAGAT	61.85	
Cnot8 (mature)	Cnot8_KW F	CCGCCAGCCAGGATTGGA	62.11	172
	Cnot8_KW R	TTCCTCAAGATTGCTGGCCC	60.32	
Dld (mature)	Dld F	CGATGGCAGCACTCAGGTTA	60.11	193
	Dld R	AACTGAACCCAGTTCCACACC	60.41	
Egr1 (mature)	Egr1 F	AGTGATGAACGCAAGAGGCA	59.96	121
	Egr1 R	TAGCCACTGGGGATGGGTAA	59.95	
Egr1 (unspliced)	un Egr1 F	GGGTCTCATCGTCCAGTGAT	58.88	166
	un Egr1 R	GAAGCGGCCAGTATAGGTGA	59.25	
Egr2 (mature)	Egr2 F	GTAGCGAGGGAGTTGGGTCT	60.97	219
	Egr2 R	ATCATGCCATCTCCCGCCAC	62.69	
Fos (mature)	cFos Fw	GGGACAGCCTTTCCTACTACC	59.51	87
	cFos Rv	GATCTGCGCAAAAGTCCTGT	58.84	
Fos (unspliced)	cFos intron Fw	TGACCGGAATGCTTCTCTCT	58.44	165
	cFos intron Rv	TGTCACCGTGGGGATAAAGT	58.65	
Fosb (mature)	Fosb F	ACCCTCCGCCGAGTCTCAGT	64.94	128
	Fosb R	TTGCGGTGACCGTTGGCACG	66.51	
Fosb (unsplcied)	Fosb unspliced Fw	GGGGTCGGTGTGTGTTATGT	59.96	164
	Fosb unspliced Rv	GATCCTGGCTGGTTGTGATT	57.87	
Gapdh (mature)	Gapdh F	AAGAAGGTGGTGAAGCAGGC	60.54	114
	Gapdh R	ATCGAAGGTGGAAGAGTGGG	59.09	

Itga1 (mature)	Itga1 F	TGGCCAACCCAAAGCAAGAA	60.69	189
	Itga1 R	GGCCCACATGCCAGAAATCC	61.68	
Malat1	Malat1 F	GTGGGTGGGGGTGTTAGGTA	60.84	176
	Malat1 R	CAACCTTCCTTAGCTGCCCG	61.03	
Nxf1 (mature)	Nxf1 F	AAACGAATTGAAGACTGAGCGGG	61.66	167
	Nxf1 R	CCTCCCAGGGGTGAATGTCC	61.93	
Pan2 (mature)	Pan2 F	TCAACCTCATGGTGCCCAAGG	62.63	156
	Pan2 R	CTGTGCGGGCATCCTCAATG	61.72	
Parn (mature)	PARN F	CTGCTTTTGCTGCGGAACTC	60.39	181
	PARN R	TGCTGATTCCTGAAAACTCCCC	60.56	
Ppia (mature)	Ppia F	GCCGATGACGAGCCCTTG	60.89	169
	Ppia R	TAAAGTCACCACCCTGGCACAT	61.89	
Rpl28 (mature)	Rpl28 F	TACAGCACGGAGCCAAATAA	57.23	74
	Rpl28 R	ACGGTCTTGCGGTGAATTAG	58.28	
Rpl28 (unspliced)	Rpl28 Fw2	CATCGTGTACACCTATTCCC	55.36	88
	Rpl28 R	ACGGTCTTGCGGTGAATTAG	58.28	
Sqstm1 (mature)	Sqstm1 F	ACTACCGCGATGAGGATGG	58.96	142
	Sqstm1 R	CATGGTGGGCGATGTTCCC	61.12	
Sqstm1 (unspliced)	Sqstm1 us F	GAAAAGGCAACCAAGTCCCCA	61.03	137
	Sqstm1 us R	GCCTCTGCTGCATTTTAGCCT	61.29	
Zfp36 (mature)	Zfp36 F	TCTCTGCCATCTACGAGAGCC	61.09	111
	Zfp36 R	ACGGGATGGAGTCCGAGTTT	60.9	
Zfp36L1 (mature)	Zfp36L1 F	AGATCCTAGTCCTTGCCCCG	60.76	109
	Zfp36L1 R	GTTGAGCATCTTGTTACCCTTGC	60.37	
Zfp36L2 (mature)	Zfp36L2 F	GCCCTCGCCCGTTATTCATC	61.23	165
	Zfp36L2 R	CCAGGGATTTCTCCGTCTTGC	61.01	
Egr1 3' end	Egr1 probe F	TATACATCTATTCAGGAGTTGGAGTG	57.82	200
	Egr1 probe R	GAAGGATACACACCACATATCCC	58.36	
Rpl28 3' end	Rpl28 probe F	TGCTCGGGCTACCCTCAG	60.76	200
	Rpl28 probe R	GGAAGTCAGTGGACTCTTTATTGC	59.61	
Human				
Target RNA	**Name**	**Sequence**	**Tm (°C)**	**Fragment length**
CNOT1 (mature)	CNOT1 F	CAGCAGACGGACTTGTCTCA	59.68	121
	CNOT1 R	ACATGGTTGGATGTGGTGGAT	59.64	
RPL10A (mature)	RPL10A F	TCTCTCGCGACACCCTGT	60.28	220
	RPL10A R	TTAGCCTCGTCACAGTGCTG	60.04	
RPL10A (unspliced)	unspliced RPL10A F	AATGTGAACGCTCCGGGTAA	59.68	184
	unspliced RPL10A R	CTTAGCCTCGTCACAGTGCT	59.76	

Table 2.4 qPCR primer sequences.

2.5.2 Poly(A) tail tests (RL2-PAT)

In order to measure the poly(A) tail, poly(A) tail tests (PATs) were performed (figure 3.1 A). To confirm that changes observed were due to tail length rather than alternative polyadenylation, some RNA was first treated with RNase H and oligo(dT) to remove poly(A) tails and thereby serve as a deadenylated control. RNase H tail removal is described in section 1.3.4.

A 5' pre-adenylated DNA anchor was added to the 3' ends of mRNAs in the sample using T4 RNA Ligase truncated K227Q. This ligase does not possess the complete activity of wild type T4 RNA ligase and as such can only catalyse the second step of ligation which requires an adenylated 5' end. Its use in the PAT assay meant that only the anchor could be added to the 3' ends of mRNAs and that endogenous RNA fragments should not have been ligated together. The 3' end of the anchor oligo had a dideoxy modification to prevent self-ligation. 500 ng RNA was mixed with 1 µL PAT anchor (20 µM), 6 µL PEG 8000, 2 µL 10X T4 RNA ligase buffer and 1 µL T4 RNA ligase 2, tr K227Q (NEB, M0373L) in a 20 µL reaction and incubated overnight at 16 °C.

Reverse transcription was performed using the PAT-R1 primer, which was complementary to the anchor, but contained an additional five thymine residues to select only for polyadenylated 3' ends of RNAs. 23 µL H_20, 5 µL PAT-R1 (20 µM) and 2 µL dNTPs (10 mM) were added to the completed ligation reaction and incubated for 5 minutes at 65 °C. Samples were placed on ice for one minute and a further 23 µL H_20, 20 µL 5x First Strand Buffer, 5 µL DTT (0.1M) and 2 µL SuperScript III added. Reactions were incubated at 55 °C for one hour and the reaction stopped by heating to 70 °C for 15 minutes.

PCR was carried out using the PAT-R1 primer and a gene-specific forward primer. Each 50 µL reaction contained 1X GoTaq Flexi Buffer, 1.5 mM $MgCl_2$ (Promega, part A351B), 0.2 mM dNTPs, 0.5 µM PAT-R1, 0.5 µM Forward Primer, 1.25 units (0.25 µL) GoTaq G2 Flexi (Promega, M7805) and 1 µL PAT cDNA. Initial denaturation was carried out at 95 °C for 5 minutes, followed by 40 cycles of 1 min 95 °C, 1 min 58 °C, 2 mins 72 °C, then a final 10 minute extension at 72 °C. Primer details provided in table 2.5.

Mouse				
Target RNA	**Name**	**Sequence**	**Tm (°C)**	**Predicted deadenylated length**
Actb	Actb 3F1	AAACTTTCCGCCTTAATACTTC	56.4	219
Actb	Actb 3F2	GGAGGATGGTCGCGTCCAT	61.6	308
Egr1	Egr1 PAT	AGCTGAGCTTTCGGTCTCCA	60.5	339
Egr2	Egr2 PAT	GTGCTTCAATGTCACTGCCG	60.5	174
Fos	Fos 3F2/ Fos PAT	CTGACATTAACAGTTTTCCATG	56.4	221
Fosb	Fosb PAT	ATTGACTCCATAGCCCTCAC	58.4	184
Rpl28	Rpl28 3F1	GCCACTTCTTATGTGAGGAC	58.4	254
Sqstm1	Sqstm1 3F1	AAGAGGGGACTGTCCATAGT	58.4	255

Human				
Target RNA	**Name**	**Sequence**	**Tm (°C)**	**Predicted deadenylated length**
ACTB	ACTB PAT	GTCCTCTCCCAAGTCCACAC	62.5	288
GAPDH	GAPDH 1st	CCCCACCACACTGAATCTCC	62.5	137
RPL10A	RPL10A	TCACCTGGCTGTCAACTTCT	58.4	155

Worm				
Target RNA	**Name**	**Sequence**	**Tm (°C)**	**Predicted deadenylated length**
daf-21	daf-21 Fw1	AATCTCACGCTTCCCGCATC	60.5	292
	daf-21 Fw2	GACTGCTCTTCTCGCTTCCG	62.5	333
hsp-1	hsp-1 Fw1	ACCGCAGAGAAGGAGGAGTT	60.5	329
	hsp-1 FW2	AAGCAGACCATTGAGGACGAG	61.2	437
hsp-16.1	hsp-16 Fw1	ACCCGAAGATGTTGATGTTGGT	60.1	244
	hsp-16 Fw2	GTCTCGCAGTTCAAGCCAGA	60.5	372
hsp-70(C12C8.1)	hsp-70 C12C8.1 Fw1	TGAAGATACCCTTCGTTGGATGG	62.9	320
	hsp-70 C12C8.1 Fw2	ATCAGCAGAAGATGCAAAACGTG	60.9	356
hsp-70(F44E5.4)	hsp-70 F44E5.4 Fw1	TGGATTCCAACTCATTGGCTGAA	60.9	321
	hsp-70 F44E5.4 Fw2	CGTGCAAAAGAAGCTGTTGATGA	60.9	358
Y45F10D.4	Y45F10D.4 Inner	AAGGGTCGCAGTGGAGAATG	60.5	154
	Y45F10D.4 Outer	CCACGTCAGCAATTGGATTCG	61.2	252

RL2-PAT		
Oligo Name	**Sequence**	**Tm (°C)**
PAT anchor	rApp/GGTCACCTTGATCTGAAG/ddC	-
PAT-R1	GCTTCAGATCAAGGTGACCTTTTT	62
Dach anchor	rApp/GGT CAC CTT GAT CTG AAG CCA GCT GTA GCT ATG C/ddC	-
Dach1R	GGCATAGCTACAGCTGGC	58.4

Table 2.5 Primers and oligonucleotides used in the RL2=-PAT assay.

PCR products were run alongside 100bp ladder (NEB, N3231L) on a 1.2 – 1.5 % agarose TBE gel containing 1.5X SYBR Safe (Invitrogen, S33102) at 4 V/cm for 2.5-3 hours.

To determine exact lengths, sample and marker lanes, along with a minimum of one blank lane per gel image, were scanned using the Quantity One software. Quantity One denotes vertical position on the gel by rf (relative front) value, with 0 at the top of the selected frame and 1 at the bottom. For a vertical line down the centre of each lane, the intensity at each rf value was outputted into a .csv file. The intensity of the blank lane was subtracted from each sample lane, and all intensities scaled to the maximum value of that lane. Each sample lane at this stage had a maximum intensity value of 1, representing the point of greatest signal intensity. In order to convert gel position to fragment length, a relationship between rf value and fragment length was established using the 100 bp marker lanes. Rf value at the peak of each marker band was plotted on the x axis against $\log_2$ (fragment length) on the y axis, as the speed at which fragments migrate is inversely proportional to $\log_2$ of their length. A quartic polynomial equation was produced for each marker lane and used to calibrate the nearby sample lanes. Since SYBR safe intercalates with the DNA and staining is therefore proportional to fragment length, signal intensity for each sample lane was divided by estimated fragment length to give length-normalised signal intensity. Predicted deadenylated length was subtracted from estimated fragment length to give estimated poly(A) tail length, and peak length-normalised signal intensity value(s) for each sample lane were taken as the modal poly(A) length(s). Negative values were occasionally returned, likely resulting from slight differences in gel migration between lanes (especially if the ladder lane was not

nearby) or from deviation from the annotated 3' UTR. If significant differences arose (eg. in the case of Egr1), DNA was isolated from gel bands and sent for sequencing.

2.5.3 Sequencing of PCR products

To check identity of PCR amplicons and gel bands, DNA was isolated using either the Wizard SV Gel and PCR Clean-Up System (Promega, A9281) or the Monarch DNA Gel Extraction Kit (NEB, T1020S) according to the manufacturer's instructions. After imaging gels, bands were visualised with either a UV transilluminator or later with blue light, and the bands removed with a clean scalpel and forceps. In both systems, excess agarose was trimmed away before the remaining gel slice was melted in the relevant kit-specific solution. Isolated DNA was sent along with the appropriate primer for sanger sequencing with Source Bioscience. In the case of PAT PCR products, the gene specific forward primer was used for sequencing to avoid problems both of sequencing through a homopolymer and of entering the 3' UTR at different positions.

2.6 Protein work

2.6.1 Making lysates

RIPA buffer: 10 mM Tris HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 % NP-40, 0.1 % SDS, 0.1 % sodium deoxycholate, 1 mM PMSF (PMSF added directly before use).

3X SDS loading buffer: 195 mM Tris HCl pH 6.8, 9 % SDS, 30 % glycerol, 15 % beta-mercaptoethanol, 0.125 mg/mL bromophenol blue.

For western blotting of whole-cell protein, cells were lysed using RIPA buffer. A small amount of lysate was reserved for Bradford assay and 3X SDS buffer was added to the remainder. Generation of protein lysates from nuclear and cytoplasmic fractions was described previously in section 1.2.1.

2.6.2 Bradford assay

To determine protein concentration, samples were suspended in either RIPA buffer or Buffer A depending on the preceding experiment. Samples were diluted either 1 in 2 or 1 in 5 with water, depending on anticipated protein abundance. Standards of BSA ranging from 0.1-5 mg/mL, as well as a 0 mg control, were suspended in the same buffer and diluted either 1 in 4 or 1 in 10 as required with either water or a

water/buffer mixture such that the buffer concentration equalled that of the test samples. 5 µL diluted standard or sample was mixed with 200 µL Pierce Coomassie reagent (Thermo Scientific, 1856209) in triplicate in a 96 well plate, avoiding formation of bubbles. Absorbance at 595 nm was measured, and the absorbance from the 0 mg/mL sample subtracted from all other wells. Protein concentration in the samples were calculated using the calibration line, taking into account differences in dilution. Bradford-derived concentrations were multiplied by 0.667 to give concentration in the electrophoresis samples.

2.6.3 Western blotting

2.6.3.1 Gel casting

Running gel (10.5 mL): X mL 30 % acrylamide 37.5:1, 6.75 – X mL water, 3.45 mL 1.5 M Tris HCl pH 8.8, 105 µL 10 % SDS, 105 µL 10 % APS, 13.5 µL TEMED.

Stacking gel (3 mL): 2.1 mL water, 495 µL 30 % acrylamide 37.5:1, 375 µL 1 M Tris HCl pH 6.8, 37.5 µL 10 % SDS, 37.5 µL 10 % APS, 6 µL TEMED.

SDS polyacrylamide gel electrophoresis was performed using 30 % acrylamide/bis-acrylamide 37.5:1 solution (Sigma-Aldrich, A3699) at final acrylamide concentrations ranging from 6-12 %. Running gel was made up according to the recipe below and used to cast 1.5 mm thick mini-gels (measuring 8.3 x 7.3 cm, including wells). The surface of the running gel was levelled using ~ 1 mL water or isopropanol which was poured off once the gel was set, prior to addition of the stacking gel.

2.6.3.2 Sample preparation and electrophoresis

Western blot running buffer (10 X): 1.92 M glycine, 0.25 M Tris Base, 1 % SDS.

Western blot transfer buffer: 48 mM Tris Base, 39 mM glycine, 0.037 % SDS, 20 % (v/v) methanol.

Between 15 and 25 mg protein were loaded per lane - with the mass consistent across all lanes of any given gel – alongside a protein marker for the relevant molecular weight range (NEB, P7719S or Thermo Scientific, 26625). To achieve more uniform bands, the protein markers and all but the lowest concentration sample were diluted with 1X SDS such that an equal volume was loaded in each lane. Following dilution, samples (but not markers) were boiled at 90 °C for 10 minutes then placed on ice. Electrophoresis was carried out at a constant current

(20 – 25 mA for one gel, 40 – 50 mA for two) until the marker bands had reached the desired separation.

2.6.3.3 Transfer

TBST: 150 mM NaCl, 10 mM Tris HCl pH 8.0, 0.05 % Tween 20.

Protein was transferred to PVDF membrane (Thermo Scientific, 88518) by semi-dry transfer. PVDF membrane was wet first in methanol for at least 45 seconds, then in blotting buffer and assembled beneath the gel between two 3-layer thick stacks of buffer-soaked Whatman 3MM paper (GE Healthcare, 3030-917). Transfer was carried out at constant current of 0.8 mA/cm^2 for between 2 and 4 hours, depending on the size of the protein(s) of interest.

2.6.3.4 Protein detection

Membranes were blocked with 5 % milk for one hour at room temperature. Blots were incubated with primary antibody overnight at 4 °C then washed three times for 5 minutes in TBST before incubating with the relevant HRP-conjugated secondary antibody for 1 hour at room temperature. Details of antibodies and their dilutions are provided in table 2.6. Membranes were washed three to five times for 5 minutes in TBST before proceeding to chemiluminescent detection. ECL reagent (Amersham ECL Prime Western Blotting Detection Reagent, GE Healthcare, RPN2232) was removed from the fridge 15 minutes before use and the two components mixed immediately before application. Around 1.5 mL ECL solution was used per blot. Blots were imaged first under white light using a LAS-4000 (FujiFilm) at high resolution for 1/15th second to record position of the marker bands. The setting was then changed

PROTEIN OF INTEREST	kDA	HOST	DILUTION	SUPPLIER	PRODUCT CODE
α-TUBULIN	50	Mouse	1:1500-2000	Invitrogen	A11126
β-ACTIN	42	Mouse	1:1000	Cell Signalling	3700P
CNOT1	267	Rabbit	1:800	Proteintech	14276-1-AP
CNOT1	267	Rabbit	1:174-363	Novus biologicals	NBP2-31892
LAMIN A/C	74,63	Mouse	1:1000-1800	Cell signalling	4777T
Mouse IgG		Goat	1:1000-3000	Dako	P0447
Rabbit IgG		Swine	1:1000-3000	Dako	P0217

Table 2.6 Antibodies used for western blot.

to chemiluminescence and exposure time varied according to signal. The high-resolution setting was used where possible, however for low abundance proteins, either 'standard' or 'high' *sensitivity* were used if exposure time was otherwise in excess of 15 minutes. Signal was not quantified.

2.6.3.5 Re-use of membranes

Stripping buffer: 62.5 mM Tris HCl pH 6.8, 2 % SDS, 0.8 % β-mercaptoethanol.

Where necessary, membranes were stripped and re-probed. If this was planned rather than opportunistic, the less abundant protein was detected prior to stripping. The membrane to be stripped was placed in a 50 mL falcon tube, protein side facing in, and stripping buffer added such that it did not touch the membrane when the tube was upright (to avoid uneven stripping). As it was not possible to roll the tube at 50 °C, the tube was placed in a 55 °C water bath and rolled frequently by hand for 30 – 45 minutes. Stripping buffer was discarded and washed in TBST for 30 – 60 minutes, changing every 5 minutes, until β-ME was no longer detectable. Membranes were re-blocked then washed and imaged with ECL detection reagent to ensure removal of the previous antibodies.

2.7 Immunoflluorescence

2.7.1 Slide preparation

Cells were seeded in µ-Slide 8 well chambers (Ibidi, 80826) at densities of 0.7-2.0 x 10^4 cells/mL (depending on duration of experiment) with two wells assigned to each condition. Notably, cells grew more slowly than in larger culture dishes, and preferred to grow at the edges of the well. At the end of the relevant experiment, media was removed and cells were washed twice with cold PBS. Cells were fixed in 4 % paraformaldehyde (PFA) for 12 minutes then washed a further three times for 5 minutes with PBS. At this point, slides were stored for up to 14 days at 4 °C. Cells were next permeabilised in 0.1 % Triton X100 in PBS for 10 minutes, followed by three x 1 minute PBS washes. Blocking was carried out using 4 % BSA in PBS for one hour at room temperature. The primary antibody was then added in 4 % BSA and incubated overnight at 4 °C. Secondary only control wells remained in 4 % BSA. Following three x 1 minute PBS washes, slides were incubated with Alexa Fluorophore-conjugated secondary antibody for 1 hour at room temperature and washed again. Antibody details and dilutions are described in table 2.7. Where

Phalloidin (Cell Signalling, 8878) was used, stain was diluted 1:200 in PBS and slides incubated for 30 minutes. Nuclei were stained by incubating with Draq5 (Cell Signalling, 4084) diluted 1:4000 in PBS for 5 minutes. Prior to imaging, cells were washed three more times for 1 minute with PBS, then covered with 150 µL PBS to prevent drying. Following addition of fluorophores, slides were stored in foil.

2.7.2 Microscopy

Slides were imaged using a Zeiss LSM510 confocal microscope with 488 nm argon, 543 nm He-Ne and 633 nm He-Ne Lasers, under conditions of minimal ambient lighting. A drop of immersion oil was used to coat the 40x objective, and the slide placed on the platform above. One fluorescence colour (emitted colour) was chosen and used to find cells and focus the microscope through the eyepiece, using the microscope light rather than the lasers. Once cells were in focus, the microscope light was switched off, the 633 nm channel enabled and focus fine-tuned in the scanning window using the Draq5 staining. For each channel, the gain and offset adjusted were to ensure the images were neither over- nor underexposed, and to make sure any signal outside the cells was minimised. Where channels were detecting antibody staining, the relevant secondary-only well was imaged early on to check the level of background staining. Once settings were optimised, they were maintained across all images of that experiment, and immersion oil on the objective was refreshed regularly. Images were taken of at least two fields of view (covering 225 x 225 µm) per well, in some cases with an additional zoomed-in image taken (covering 83.33 x 83.33 µm).

PROTEIN OF INTEREST	HOST	DILUTION (IN 4 % BSA)	SUPPLIER	PRODUCT CODE	Fluorophore
CPSF4	Rabbit	1:200	Proteintech	15023-1-AP	-
PABPN1	Rabbit	1:200	Abcam	ab75855	-
WDR33	Mouse	1:200	Santa Cruz	sc-374466	-
Rabbit IgG	Goat	1:400	Invitrogen	A11034	Alexa 488
Mouse IgG	Goat	1:400	Invitrogen	A11004	Alexa 568
Mouse IgG	Goat	1:400	Invitrogen	A11003	Alexa 546

Table 2.7 Antibodies used for immunofluorescence.

2.7.3 Image analysis

2.7.3.1 Nuclear/cytoplasmic quantification

Images were analysed using Image J to determine relative fluorescence of nuclear:cytoplasmic staining (process illustrated in figure 2.2). First, nuclei were identified using the '3D objects counter' and 'threshold' functions on the Draq5

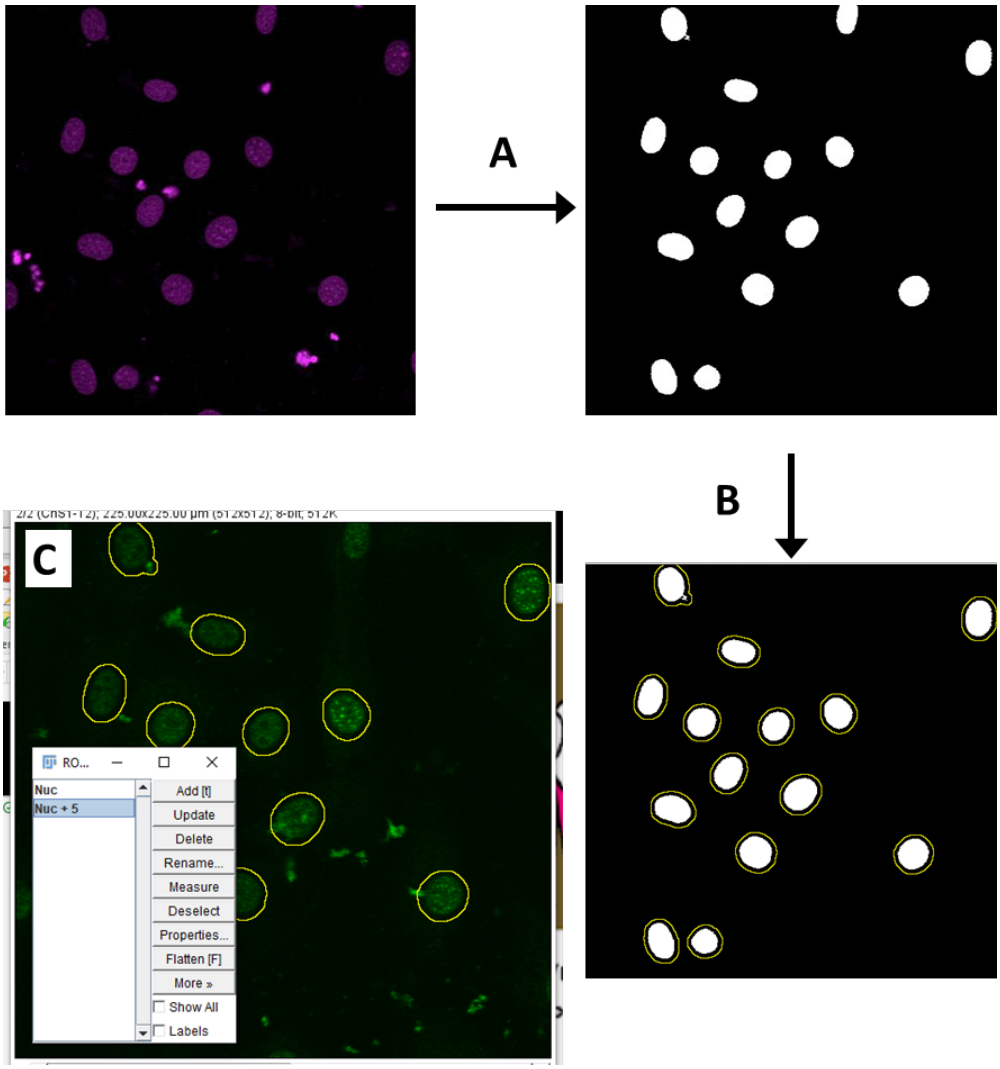


Figure 2.2 Schematic of nuclear and cytoplasmic signal quantification. A) The 3D objects counter tool was used on the Draq5 channel of the image to select bodies with an area >500 and an intensity > 15. 'Object' was set as the output. A selection was created then inverted to select nuclei rather than negative space, and added to the manager. **B)** The selection was expanded by 5 pixels and this selection also added to the manager, along with another for the nuclear/cytoplasmic boundary minus 2 pixels. **C)** The view was switched to the antibody stain channel and the fluorescence intensity and area measured for all three selections. To obtain mean fluorescence for the cytoplasmic regions, the total fluorescence and area of the nuclei (boundary, not -2) were first subtracted from the +5 selection.

channel of the image, using '15' as the fluorescence threshold and '500' as the minimum area. Three selections were then made: one along this nuclear/ cytoplasmic boundary ('B'), one increased in radius by 5 pixels to incorporate the cytoplasm ('+5'), and one reduced by 2 pixels ('-2') to ensure only nuclear stain was included. Each 'selection' was the sum across all nuclei in the image. The channel was changed to show antibody staining rather than Draq5, and measurements for total area and fluorescence were obtained for each selection in order to calculate mean fluorescence/area ('fluorescence intensity'). Figures for the cytoplasmic fraction were obtained by subtracting the boundary measurements from those of the +5 selection. A ratio of mean nuclear fluorescence/mean cytoplasmic fluorescence was calculated for each image. Zoomed in images were excluded from this analysis. Violin plots of the data from two replicates were generated using GraphPad Prism 8, and a t-test performed between control and Cnot1 knockdown at each time point.

2.8 Deep Sequencing

2.8.1 PQ-Seq Library preparation

RNA was isolated using the ReliaPrep RNA Cell Miniprep system (Promega, Z6012) according to the manufacturer's instructions, except for an extended DNase-treatment step of 60 rather than 15 minutes. 1 µg total RNA as determined by nanodrop was used as a starting point for each sample. RNA for the steady state experiment had undergone additional DNase treatment with Turbo DNase (Invitrogen, AM2238) and homemade DNase I buffer but this step was removed from the serum stimulation experiment as it was suspected this treatment may be causing RNA degradation. Anchor addition and library preparation were carried out by Sunir Malla (then at DeepSeq, University of Nottingham - both experiments), and Hilary Collins (GRRB, University of Nottingham - serum stimulation experiment). PQ-Seq is discussed in detail in chapter 7 *(PAT-Quant Seq: Poly(A) tail deep sequencing)* and is outlined in figure 7.1. In brief, an anchor was added to the 3' end of polyadenylated mRNAs using a biotinylated anchor template. RNA was then fragmented using RNase T1 (cleaves ssRNA after G) in order to enrich the sample for 3' end fragments which were a compatible length for Illumina short read sequencing. This step avoided removal of long RNAs during subsequent size

selection. The RNA:anchor template duplexes were pulled out of the solution using streptavidin beads and the RNA then eluted by disrupting the duplex with heat and formamide. The eluted RNA then entered the QuantSeq Flex pipeline in which Illumina-compatible libraries were synthesised. Fluorescence files were obtained from the sequencer and processed by tailseeker to give reliable poly(A) length calls.

2.8.1.1 Anchor addition

Klenow reaction mix (5 µL): 4 µL 5X First Strand buffer (Invitrogen, supplied with 18080093), 1 µL 0.1 M DTT, 1 µL 10 mM dNTPs, 1 µL Murine RNase Inhibitior (NEB, M0314).

Streptavidin beads (Dynabeads MyOne C1 - Invitrogen, 65001) were removed from the fridge 30 minutes prior to use then resuspended by vortexing for 30 seconds. For each sample, 50 µL beads were transferred to a 1.5 mL Eppendorf tube and the tube placed on a magnetic rack for one minute. Supernatant was removed, and the beads were washed first in 100 µL nuclease-free water at room temperature, then with 100 µL 1X RNase T1 digest buffer (Ambion, supplied with AM2283) which had been pre-warmed to 55 °C. The supernatant was removed and replaced with a fresh 100 µL digest buffer in which the beads were kept at room temperature until use.

In order to anneal the anchor template, RNA and 1 µL RA3-PAT-biotin primer (100 µM) were mixed in a total volume of 12 µL, heated to 80 °C for 5 minutes then cooled to 4 °C at a rate of 2 °C/min. 7 µL Klenow reaction mixture was added and mixed by pipetting, then warmed to 37 °C. 1 µL Klenow polymerase exo- (NEB, M0212L) was added and the reaction was incubated at 37 °C for 30 minutes.

2.8.1.2 Fragmentation

Prior to fragmentation, RNA was incubated at 80 °C to disrupt secondary structure. Incubation was for a total of 15 minutes, after 5 of which 80 µL digest buffer was added. The mixture was cooled to 22 °C in a PCR machine and 2 µL RNase T1 (Ambion, AM2283) was immediately added, mixing by pipetting. The reaction was incubated for 5 minutes at 22 °C, then vortexed for 15 seconds in a LoBind centrifuge tube with 200 µL 50:50 acid-phenol/chloroform and a further 100 µL digest buffer. After spinning at top speed for 5 minutes at room temperature, an additional chloroform wash was performed. The sample was spun again for

5 minutes at room temperature, and the aqueous phase transferred to a tube containing the prepared streptavidin beads.

2.8.1.3 Streptavidin purification

The DNA:RNA duplex was incubated with the beads for 15 minutes with shaking at 37 °C. Beads were concentrated on a magnetic rack for 2 minutes, after which the supernatant was removed and discarded. Care was taken to stabilise the tubes while opening the lids to avoid disruption of the beads. Beads were resuspended in 200 µL 0.2X SSC and incubated for 5 minutes at 65 °C, concentrated on the rack again for 2 minutes the supernatant removed, and the wash repeated. After removal of the second wash supernatant, beads were resuspended in 50 µL nuclease-free water and vortexed.

Resuspended beads were heated at 80 °C for 2 minutes, beads concentrated on a magnetic rack, and the supernatant moved to a fresh tube on ice. The beads were resuspended in 50 µL deionized formamide by vortexing and heated at 80 °C for a further 2 minutes. Beads were again concentrated for 2 minutes and the supernatant combined with the previous. One final elution was carried out using 100 µL nuclease-free water at 80 °C for 2 minutes, and the supernatant was again combined with that of the first and second elution steps.

RNA was precipitated from the eluate using 1 µL glycogen, 23 µL 3M sodium acetate pH 5.5, and 700 µL ethanol pre-cooled to -20 °C. The mixture was vortexed and the RNA precipitated at -80 °C for at least one hour. RNA was collected by spinning at 16,000 g at 4 °C for 30 minutes. The pellet was washed with 800 µL freshly prepared 80 % ethanol and spun again in a cold centrifuge at 16,000 g for 10 minutes. Supernatant was removed and the pellet allowed to air dry for 5 minutes before it was resuspended in 20 µL 1X RNA loading buffer in preparation for size selection. The RNA (still on the beads) along with 50 bp RNA ladder in a separate tube, were incubated at 95 °C for 2 minutes. The RNA was placed on a magnetic rack for 1 minute and the supernatant then transferred to a clean tube on ice. RNA was run alongside the marker on a Novex 6 % TBE-Urea gel (Invitrogen, EC6265BOX) that had been washed thoroughly with nuclease-free water. RNA was run half way to avoid dilution with too much gel. Gels were stained with SYBR Gold for 5 minutes then visualised on a clean blue light transilluminator. RNA fragments between 200 bp and

the top of the gel were excised (in a single strip) and transferred to a LoBind centrifuge tube. Gel slices were crushed with disposable RNase-free pestles and soaked in 300 µL TE buffer. 2 µL RNase inhibitor was added and the slurry rotated overnight at 4 °C. Gel solution was passed through a filtration column by centrifuging at 13,200 rpm for 2 minutes. Eluate then underwent ethanol precipitation, and the resulting pellet was resuspended in 55 µL nuclease-free water.

Before proceeding to library preparation, RNA was purified using 99 µL RNA Clean XP beads (Beckman Coulter, A63987). Cleanup was carried out loosely according to the manufacturer's instructions, though initial mixing was by vortexing rather than pipetting, and two rather than three 70 % ethanol washes were carried out. In addition, no evaporation step was expressly noted though the manufacturer suggests 10 minutes. Beads were incubated in 12 µL nuclease-free water at room temperature for 2 minutes to elute the RNA. After concentrating the beads on a magnetic rack for 2 minutes, 10 µL supernatant was transferred to a PCR tube to commence QuantSeq Flex library preparation.

2.8.1.4 QuantSeq Flex library preparation

Library preparation begins with cDNA synthesis using a primer complementary to the beginning of the anchor sequence. RNA is then degraded and a second DNA strand synthesised using random priming. These primers also contain Illumina linker sequences which enable later amplification. The resulting cDNA is purified then amplified by PCR, through which sample-specific index sequences are introduced via the reverse primer. After a final purification step, libraries are suitable for sequencing on an Illumina MiSeq or HiSeq. For PQ-Seq, second strand synthesis and PCR reagents are provided in the 3' mRNA-Seq FWD kit (Lexogen, 015.24), while the first strand synthesis reagents (which allow use of a custom primer) are supplied in an additional module (QuantSeq-Flex First Strand – Lexogen, 026). Primers are detailed in table 2.8.

The 10 µL purified RNA was mixed with 5 µL QuantSeq-Flex FS cDNA synthesis Mix1 (FS1x), 5 µL RTP (20 µM) and 5 µL nuclease-free water by gentle vortexing, then heated to 85 °C for 3 minutes. RTP matches the 5' end of the anchor template and

Name	Sequence
Anchor extension and cDNA synthesis	
RA3_PAT_biotin (anchor template)	(Bio)GCCTTGGCACCCGAGAATTCCANNNNNNNNNNNNGTCAGTTTTTTTTTTTTTTTT
RNA RT Primer (RTP) - serves as P5 adapter	GCCTTGGCACCCGAGAATTCCA
P7 adapter - added during random priming of 2nd strand	CACGACGCTCTTCCGATCT
Library amplification	
QuantSeq-Flex PCR primer (for libraries - included in PCR mix)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
RNA PCR Primer (RP1) - for spike ins	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA
RNA PCR Primer, Index 1 (RPI1)	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Spike in templates	
Poly A Spike N (0)	TCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNNNNNNNNNCTGACGAGCTACTGTTGGAATTCTCGGGTGCCA
Poly A Spike_16	TCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNNBAAAAAAAAAAAAAAAACTGACGAGCTACTGTTGGAATTCTCGGGTGCCA
Poly A Spike_32	TCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNNBAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTGACGAGCTACTGTTGGAATTCTCGGGTGCCA

Poly A Spike_64	TCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNBAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAACTGACGAGCTACTGTTGGAATTCTCGGGTGCCA
Poly A Spike_128	TCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNBAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAACTGACGAGCTACTGTTGG AATTCTCGGGTGCCA

Table 2.8 PQ-Seq sequences. Index sequences indicated in purple.

was used in first strand synthesis to prime reverse transcription from the 3' end, through the rest of the anchor and the poly(A) tail and into the gene body. After incubation, the RNA-primer mixture was cooled to 50 °C and 4.5 µL QuantSeq-Flex FS cDNA synthesis Mix2 (FS2x) was added along with 0.5 µL Enzyme mix. The reaction was mixed by pipetting and incubated at 50 °C for one hour.

RNA was degraded as per the QuantSeq manual. Random-primed second strand synthesis was then carried out according to the manufacturer's instructions, but with an extended incubation at 25 °C of 1 hour rather than 15 minutes. Library cDNA was purified using the magnetic bead-based purification module (Lexogen, 022.96). This was again performed according to the user manual, however a larger elution volume of 25 µL rather than 20 µL was used, and beads were only air dried for 3-5 rather than 5-10 minutes.

Samples from distinct cellular fractions have different starting compositions, and the extensive processing in the PQ-Seq protocol provided ample opportunity to introduce further variation. In addition, differences in primer efficiency may have existed. Use of a consistent cycle number for library amplification across samples may therefore have resulted in different library sizes and consequent unequal sequencing depth. Test amplifications were therefore carried out with a small volume of library cDNA before proceeding to library amplification. For each library, 5 µL cDNA was amplified in a 15 µL reaction under the following cycling conditions: Initial denaturation 98 °C 30 seconds, then 21 cycles of 98 °C 10 seconds, 65 °C 20 seconds, 72 °C 300 seconds. After 15 and 18 cycles 3 µL was removed from the reaction and stored. For each sample, PCR product from 15, 18 and 21 cycles was run on a bioanalyzer using the High Sensitivity DNA assay kit and chip (Agilent,

5067-4627, 5067-4626). Traces were compared to work out when products in the 200-700 bp size range started to appear, and this sample-specific cycle number was taken forwards for final library amplification.

Library amplification was carried out as in the QuantSeq manual, but with a 300 rather than 30 second extension at 72 °C. In addition, custom index primers which contained the RTP sequence (table 2.8) were used in place of the provided i7 indices, at a final concentration of 0.5 µM. The index used for each sample is indicated in table 2.9. Following amplification, libraries were purified once again using the purification module. After the 80 % ethanol wash, beads were again only air dried for 3-5 rather than the 5-10 minutes suggested by the QuantSeq manual. Quality of the purified libraries was verified by resolving 1 µL on the Bioanalyzer using the High Sensitivity DNA assay kit and chip (Agilent, 5067-4627, 5067-4626).

Steady state		
Sample	Index no.	Index Seq
Nuc	6	GCCAAT
Cyto	5	ACAGTG
Spike_N	2	CGATGT
Spike_16	4	TGACCA
Spike_32	7	CAGATC
Spike_64	8	ACTTGA
Spike_128	9	GATCAG

Serum stimulated		
Sample	Index no.	Index Seq
0' Nuc	1	ATCACG
15' Nuc	5	ACAGTG
60' Nuc	6	GCCAAT
0' Cyto	19	GTGAAA
15' Cyto	12	CTTGTA
60' Cyto	10	TAGCTT
Spike_N	2	CGATGT
Spike_16	4	TGACCA
Spike_32	7	CAGATC
Spike_64	8	ACTTGA
Spike_128	9	GATCAG

Table 2.9 Sample-index relationships for the steady state and serum-stimulated PQ-Seq experiments.

2.8.1.5 Spike in preparation

To ensure that tailseeker was generating reliable poly(A) length calls for endogenous mRNAs, synthetic spike-ins of known poly(A) length (0, 16, 32, 64 and 128 – table 2.8) were amplified and added to the libraries to be sequenced. The P5 adapter sequence of the spike ins differed from those introduced by the random primers during QuantSeq second strand synthesis and as such, amplification was not

performed with the QuantSeq reagents. Rather, 5 pmol spike in oligo were amplified using KAPA HiFi Master Mix (Roche, KK2103) with 0.4 µM RPI primer (table 2.8) and 0.4 µM index primer in a 25 µL reaction with the following program: 98 °C 3 minutes, then 10 cycles 98 °C 30 seconds, 62 °C 20 seconds, 71 °C 60 seconds, followed by a final extension at 71 °C for 2 minutes. Quality of the spike ins was assessed similarly to libraries using the Bioanalyzer. The index primers used for each spike in are given in table 2.9.

2.8.2 PQ-Seq bioinformatics analysis – tailseeker

Following sequencing on an Illumina MiSeq (using a MiSeq v2 500 cycle kit), the entire output (including fluorescence intensity .cif files) were fed into the tailseeker pipeline. tailseeker is a challenging software to set up and required much work by Daniel Zadik (DeepSeq, University of Nottingham) to get running. Reads were mapped against the GRCm38 mouse assembly. A pre-built package for tailseeker including genome indices, annotation and association databases is available at: https://zenodo.org/record/203939#.XzMB-ShKiUk (last accessed 08/01/2021).

tailseeker produces two metric summaries: single and multi. Data from the single mapping summary spreadsheet was used for analysis. The spreadsheet contained summary statistics (mean, median, high and low confidence intervals) rather than values for individual transcripts. To avoid inclusion of mRNAs with few reads and possibly distorted summary statistics, results were filtered for those mRNAs with a minimum value for poly(A) tag counts. In general, 30 tag counts were considered to give a reliable tail length call, however other values were also used as indicated in the relevant figures.

2.8.3 Analysis of a publicly available CNOT1 knockdown RNA-Seq dataset

Fastq.gz files generated by Roger Grand's lab (598) were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141496 . Experiments were carried out in HeLa cells 72 hours after transfection with control or CNOT1 siRNA. Two replicates each were available for untreated control or knockdown cells, and an additional one replicate for each condition was treated with the transcription elongation inhibitor DRB for 4 hours. The study also involved TAB182 knockdowns and treatments with DNA damage-inducing drugs but these samples were not included in this analysis. So far, these data do not seem to have been published.

Quality control for each read was assessed using FastQC (599). In this instance the FastQC v0.11.7 was run on command line using the provided script. Binaries were downloaded from www.bioinformatics.babraham.ac.uk/projects/fastqc/ .

2.8.3.1 Alignment with STAR

Mapping was performed using STAR version 2.7.4a (600). Genome indices were compiled to incorporate splice junctions using the Gencode GRCh38 primary assembly with the chromosomal comprehensive genome annotation (601). The sjdbOverhang option was set to 149 such that up to 149 nucleotides of a 150 nt read could be on one side of an exon junction. Reads successfully aligned using these indices should therefore comprise both spliced and unspliced transcripts and any DNA contamination should be evident by the extent of intergenic reads. During the mapping step, 8 nt adapter sequences (as indicated in the fastq.gz file names) were removed but no additional clipping was carried out. The libraries were reverse stranded, but strandedness options were absent from this STAR release so no strandedness information was defined. Duplicated fragments are not automatically flagged by STAR so were initially included in the analysis, however the 'bamRemoveDuplicatesType' option was later set to 'UniqueIdenticaNotMulti' (marks duplicate unique mappers, but not multimappers) or 'UniqueIdentical' (marks both duplicate unique mappers and multimappers) to assess the impact of duplicate removal. Using this option, all copies of a duplicated fragment (and multimappers if specified) are flagged and therefore ignored by featureCounts in downstream raw count determination.

The following script executes index assembly followed by mapping and adapter clipping.

```
# Maps paired end reverse stranded reads in a directory
using STAR
source /users/stxkw5/stxkw5/TailSeq-
Build/tailseeker_sourcefile

# Unzip and make list of fastq files
cd /users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Fastq
gunzip *.gz
ls ./*.fastq > run_list.txt # make list of fastq files

# For each pair, extract adapter sequences
n=1
```

```
N="$(grep -c ".fastq" run_list.txt)" # count entries in
run_list.txt
echo $N

while [ $n -le $(($N+1)) ]; do
cd /users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Fastq
awk '{print substr($n, 20, 7)}' run_list.txt > adapt_1.txt
# make list adapter 1 seqs
awk '{print substr($n, 29, 7)}' run_list.txt > adapt_2.txt
# make list adapter 2 seqs
awk '{print substr($n, 3, 16)}' run_list.txt > outname.txt
# make list outfile sample names
adapt_1n="$(cat adapt_1.txt | head -$n | tail -1)"
adapt_2n="$(cat adapt_2.txt | head -$n | tail -1)"
echo $adapt_1n
echo $adapt_2n

# For each pair, define input file paths and names
read1="/users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Fastq/$(cat
run_list.txt | head -$n | tail -1)"
read2="/users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Fastq/$(cat
run_list.txt | head -$(($n+1)) | tail -1)"

echo "read1: $read1"
echo "read2: $read2"

# For each pair, define output file path and name
outname="/users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Mapped2/$(cat
outname.txt | head -$n | tail -1)"

cd /users/stxkw5/stxkw5/RNA/STAR/STAR-
2.7.4a/bin/Linux_x86_64_static

#generate genome indices
./STAR --runThreadN 6 --runMode genomeGenerate --genomeDir
/users/stxkw5/stxkw5/RNA/Genome/Human/Human_indices_v2.7.4a
--genomeFastaFiles
/users/stxkw5/stxkw5/RNA/Genome/Human/GRCh38.primary_assemb
ly.genome.fa --sjdbGTFfile
/users/stxkw5/stxkw5/RNA/Genome/Human/gencode.v34.primary_a
ssembly.annotation.gtf --sjdbOverhang 149

# Run STAR spliced with adapter clipping
./STAR --runThreadN 4 --genomeDir
/users/stxkw5/stxkw5/RNA/Genome/Human/Human_indices_v2.7.4a
--readFilesIn $read1 $read2 --outFileNamePrefix $outname --
outSAMtype BAM SortedByCoordinate --clip3pAdapterSeq
$adapt_1n $adapt_2n
echo "completed spliced mapping for $(cat outname.txt |
head -$n | tail -1)"
echo "outfile: $outname"
```

2.8.3.2 Generating 'unspliced' alignments

Intron-containing reads were used as a proxy for unspliced RNA level. To select for these, the BEDtools (602) intersect function was used with an exon-only BED file to filter out reads which overlapped more than 95 % with exons. BED files were generated for the GRCh38 genome using the Table Browser program in UCSC genome browser and selecting exons as feature outputs. Removal of exons was used rather than selecting for intronic overlap in order to retain intergenic reads. The resulting 'unspliced' filtered alignments were used for downstream differential expression analysis and also to generate metaplots to assess bulk changes in mRNA level between samples.

Below is an example code for filtering out non-intron reads.

```
./intersectBed -abam
/users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Mapped2/BAMstore/Contro
l-repeat_1Aligned.sortedByCoord.out.bam -b
/users/stxkw5/stxkw5/RNA/Genome/Human/hg38_exons.bed -f
0.95 -v >
/users/stxkw5/stxkw5/RNA/HeLa_Cnot1/Mapped2/BAMstore/total_
Ctrl-rep1_minusE.bam
```

2.8.3.3 Creation of metaplots

To generate metaplots, the bamCoverage, computeMatrix and plotProfile tools from the deepTools suite were used (603). bamCoverage converts .bam files to .bw. During .bw file generation, one can elect to perform normalisation. Here, BPM (BEDtools TPM equivalent) was used to normalise both for total number of mapped reads and for gene length. The .bw files were used as input for the scaleRegions function of computeMatrix which scales all protein-coding regions to a set length (2000 nucleotides chosen here), divides the gene into bins of length ***n*** (10 nt chosen here) and counts the number of reads which map to each bin. The generated matrix was used to create metaplots using plotProfile. Additional options at the computeMatrix step allow inclusion of regions upstream of the TSS (transcription start site) and downstream of the TES (transcription end site). Here, plots were made to include regions both 2.0 and 10.0 kb upstream and downstream.

Sample code for matrix generation is below.

```
computeMatrix scale-regions -S Control-rep1_bpm.bw Control-
rep2_bpm.bw NOT1_KD-rep1_bpm.bw NOT1_KD-rep2_bpm.bw -R
```

```
/users/stxkw5/stxkw5/RNA/Genome/Human/gencode.v34.primary_a
ssembly.annotation.gtf -o control_NOT1_bpm_bothreps_matrix
--outFileSortedRegions control_NOT1_bpm_bothreps_names --
skipZeros -m 2000 -b 500 -a 5000
--samplesLabel Ctrl_1 Ctrl_2 NOT1_1 NOT1_2 -p 6
```

2.8.3.4 Obtaining raw counts

Differential analysis to quantify changes in gene expression between control and knockdown conditions were performed using DeSeq2 with the two available replicates (604). Raw counts required by DeSeq2 as input were generated from bam files using the featureCounts tool from the Subread package (605). FeatureCounts can be used either at feature (continuous range of positions e.g. exon or gene) or meta-feature (set of features which belong to some entity of interest e.g. set of exons belonging to a gene) level. In both modes, featureCounts counts fragments per feature, but can then sum these counts for all features within a meta-feature. By counting at meta-feature level using exons, reads originating mainly from introns will be ignored, whereas they would be included using feature level counting with the feature set to 'gene'. At feature level, a fragment will always be discarded if it overlaps two features however, at meta-feature level fragments are included if the features it overlaps belong to the same meta-feature eg. fragments crossing exon junctions. Feature level counting using genes, and meta-feature level counting using exons were performed separately, using the same annotation file as at the mapping step (chromosomal region comprehensive gene annotation for the GRCh38 assembly in GTF format).

The following featureCounts options were employed for counting at feature level by gene: -p (denotes paired end reads), -B (requires both ends of a fragment to be successfully aligned), -Q 10 (only counts fragments where at least one end has a mapping quality exceeding 10 so that only uniquely mapping fragments are used), -s 2 (reverse stranded) --ignoreDup (ignores fragments where either read is flagged as duplicated in the BAM file), -f (counts at feature level) -t gene ('gene' is the type of feature to count).

Since GTF files do not contain introns as features (and introns therefore cannot be counted this way), unspliced counts were again obtained from the exon-removed BAM files using featureCounts to count the number of fragments overlapping each

feature marked as ‘gene’ in the annotation file. Settings were identical to those detailed above.

2.8.3.5 Differential expression analysis

Online code for processing featureCounts output using DeSeq2 was used as a guide (606). In brief, featureCounts output of raw counts is first loaded into R as a dataframe, trimmed to exclude surplus information and converted to a matrix where each column corresponds to a sample and each row to a feature or meta-feature. An accompanying vector is also created to describe the condition (i.e. control or treated) associated with each column in the count matrix. Once this has been loaded into R, the DESeqDataSetFromMatrix function performs differential expression analysis, the results of which can be exported to a table and summarised graphically. Heatmaps were produced to display sample similarity, MA plots to show changes in expression vs average expression, and volcano plots to visualise extent of statistically significant fold changes.

The following script encodes the above.

```
R # opens R

raw_counts <- read.table("Ctrl_r1_r2_NOT1_r1_r2.txt",
header=TRUE, row.names=1) # imports txt file into R
dataframe and sets the row names of this data frame (row
names are an inherent and necessary feature of a data
frame) to the values of the 1st column in the imported file
i.e. the gene id in this case.

# This means in the data frame, column 1 is the chromosome,
2 is start position etc, so you only want the row names and
then columns 6 onwards (raw count data).

raw_counts <- raw_counts[ ,6:ncol(raw_counts)] # rewrites
raw_counts as only the 6th-to max column (i.e. ncol counts
number of columns)

# Currently the column titles still have .bam in them and
also the whole file path. These need removing.

colnames(raw_counts) <- c("Ctrl rep1", "Ctrl rep2", "NOT1
rep1", "NOT1 rep2")

# next convert data frame to a matrix

raw_counts <- as.matrix(raw_counts)

# create vector to describe conditions – i.e. control or
knockdown – of each sample so that DeSeq2 knows how to
compare them.

(sample_type <- factor(c(rep("ctrl", 2), rep("kd", 2))))
```

```
# Data now in suitable format to stick together.

library(DESeq2) # Preparing to organise data ready for
analysis (i.e. arranging the correctly formatted columns
into the right layout for DeSeq2 analysis)

# Set the following: countData = counts, colData = sample
names+associated condition and design = sample type

(coldata <- data.frame(row.names=colnames(raw_counts),
sample_type))   # this effectively transposes column names
and putting it next to sample names in a new dataframe

dds <- DESeqDataSetFromMatrix(countData=raw_counts,
colData=coldata, design=~sample_type) # creates the dataset

dds # prints info about number of entries etc.

dds <- DESeq(dds) # THIS IS THE DIFFERENTIAL EXPRESSION
ANALYSIS. 'dds' has been overwritten with the analysis.

res <- results(dds) # Outputs results of the analysis

resdata <- merge(as.data.frame(res),
as.data.frame(counts(dds, normalized=TRUE)),
by="row.names", sort=FALSE) # merges DE analysis results
with the normalised count data from the dds output

names(resdata)[1] <- "Gene" # names column 1

write.csv(resdata, file="diffexpr-results-
qualityAndStranded.csv") # writes out to csv format

# PLOTS #

# First, log transform data to make it easier to look at in
graphs. This uses the regularized log transformation
function of DESeq2 which transforms the count data to the
log2 scale.

RLT_dds <- rlogTransformation(dds)

head(assay(RLT_dds))

hist(assay(RLT_dds)) # creates histogram but I am not sure
where it goes.

library(RColorBrewer)

(mycols <- brewer.pal(8,
"Dark2")[1:length(unique(sample_type))]) # chooses some
nice colours from the RColorBrewer set, assigns colours
based on sample type

sampleDists <- as.matrix(dist(t(assay(RLT_dds)))) # makes
matrix of distances between log transformed counts for each
sample

library(gplots)

png("qc-heatmap-samples-qualityAndStranded.png", w=500,
h=500, pointsize=20) # makes empty plot ready for heatmap –
saving to a file which can be accessed later

heatmap.2(as.matrix(sampleDists), key=F, trace="none",
```

```
     col=colorpanel(100, "black", "white"),

     ColSideColors=mycols[sample_type],
RowSideColors=mycols[sample_type],

     margin=c(10, 10), main="Sample Distance Matrix")

dev.off() # finished creating

# create MA plot to look at how expression changes compared
to average expression (over all samples??) I inbuilt DESeq2
one.

png("MAplot_qualityAndStranded.png", 600, 500,
pointsize=20)

plotMA(dds, alpha = 0.05, main="MA Plot") # sets adjusted p
value limit to 0.05

dev.off()

# create volcano plot. For this I need to make the adjusted
p sorted results.

table(res$padj<0.05) # creates a table of results which are
significant based on adjusted p value

res <- res[order(res$padj), ] # orders by adjusted p value

resdata_padj <- merge(as.data.frame(res),
as.data.frame(counts(dds, normalized=TRUE)),
by="row.names", sort=FALSE)

names(resdata_padj)[1] <- "Gene"

head(resdata_padj)

# Preparing volcano plot

volcanoplot <- function (res, lfcthresh=2, sigthresh=0.05,
main="Volcano Plot", legendpos="topleft", labelsig=TRUE,
textcx=1, ...) {

  with(res, plot(log2FoldChange, -log10(pvalue), pch=20,
main=main, ...))

  with(subset(res, padj<sigthresh ), points(log2FoldChange,
-log10(pvalue), pch=20, col="red", ...))

  with(subset(res, abs(log2FoldChange)>lfcthresh),
points(log2FoldChange, -log10(pvalue), pch=20,
col="orange", ...))

  with(subset(res, padj<sigthresh &
abs(log2FoldChange)>lfcthresh), points(log2FoldChange, -
log10(pvalue), pch=20, col="green", ...))

  legend(legendpos, xjust=-2.3, yjust=1,
legend=c(paste("FDR<",sigthresh,sep=""),
paste("|LogFC|>",lfcthresh,sep=""), "both"), pch=20,
col=c("red","orange","green"))

}

# Creating plot
```

```
png("volcanoplot-
qualityAndStranded_legendMove_ydefined_points.png", 600,
500, pointsize=20)

volcanoplot(resdata_padj, lfcthresh=1, sigthresh=0.05,
textcx=.8, xlim=c(-2.3, 2.3), ylim=c(0, 30))

dev.off()
```

To explore whether significantly up- or downregulated genes were enriched for gene ontology (GO) terms, the tabular output of the DeSeq2 analysis was used. An adjusted p-value of 0.05 was used to select for significance, and a $\log_2$(fold change) of ± 1 to detect changes in mRNA level. For some conditions, the thresholds for adjusted p-value and $\log_2$(fold change) were decreased and increased respectively to avoid exceeding the 500 gene maximum for GO enrichment analysis (607–609). When graphically presenting GO terms, highly similar terms were pruned and the term with the highest fold enrichment was retained.

2.8.3.6 Manual calculation of TPM from raw counts

To look specifically at expression of genes which were of interest in NIH 3T3 cells, TPM (Transcripts Per Million) was calculated 'manually' using the raw count data produced by featureCounts. For total RNA, reads were counted per exon and transcript length was used to compute TPM, whereas for 'unspliced' RNA, reads were counted per gene and gene length was used to compute TPM. While the sum of intronic regions may have been a more appropriate length to normalise to, this information is not readily available as alternative splicing precludes annotation of introns. For each gene, relative expression was calculated by normalising to the mean expression of the two control replicates. TPM was simple to calculate for total fragments, but for the 'unspliced' fragments it was not clear whether it was more appropriate to normalise to total sequencing depth or to the total number of fragments remaining after filtering out exons. Both approaches seemed informative by different reasoning so a figure was created for each.

3 The poly(A) tail is regulated in multiple responses and subcellular locations

3.1 Introduction

The poly(A) tail is an important feature of eukaryotic mRNAs, mediating export, stability and in some cases translation efficiency, of the mRNA (1, 2, 10, 11, 13, 19, 23, 26, 62, 269–271, 274, 325, 398, 434). The poly(A) tail thereby provides another opportunity to modulate gene expression, through control of its length (13, 58, 60, 62, 269, 434, 587). Early experiments following the discovery of poly(A) tails concluded by radioactive labelling that bulk mammalian poly(A) tail length was heterogenous but peaked at around 150 – 200 nt (324, 610). The notion of a universal 200-250 nt tail was later compounded by *in vitro* experiments showing that poly(A) polymerase (PAP) loses processivity after adding 200-250 nt (52, 179, 363). More recent genome-wide studies based on reverse transcription and PCR have found that bulk poly(A) tail length is indeed very heterogenous, but that most tails are considerably shorter than the 200 nt, consistent with widespread cytoplasmic deadenylation (11, 325, 326). On top of the contribution from deadenylation, the pool of transcripts captured by these genome-wide studies differs from those of early work, measuring total cellular mRNA rather than labelling for between 12 minutes and 48 hours and isolating either nuclear or polysome-associated RNA. In particular, the 12 minute pulse labelling experiment will have favoured rapidly synthesised RNAs which may disproportionately possess longer tails. As well as these pools being unequally affected by deadenylation, the 12 minute pulse labelling experiment in particular will have favoured rapidly synthesised RNAs which may disproportionately possess longer tails.

In addition, previous work in our lab mathematically modelling poly(A) tail length in the serum response suggested that a consistent input of 200 nt poly(A) tails throughout the transcription pulse could not generate the poly(A) tail length distributions observed experimentally unless deadenylation/decay rates changed throughout the time course (611). Though a change in decay rate is also a possibility, the assumption that all mRNAs are created with 200 nt tails was tested first using nuclear/cytoplasmic fractionation and also by labelling of newly synthesised mRNA with 4-thiouridine. Surprisingly, tails of induced mRNAs produced

late in the serum response are short in the nuclear/4SU-labelled fractions as well as in the cytoplasmic/unlabelled fractions, indicating that poly(A) tails of these mRNAs are either synthesised short or are rapidly removed in the nucleus. Furthermore, housekeeping mRNAs possess a consistent medium length tail throughout their lifetime which is unexpected given their high stability (611).

Here, these results are recapitulated and the length of the poly(A) tail probed earlier in the mRNA lifetime to attempt to determine whether these short tails are a result of reduced synthesis or rapid deadenylation. In addition, the impact of poly(A) tail length on stability of transiently expressed mRNAs is confirmed and preliminary data in other organisms suggest that regulation of poly(A) tail length is widespread.

3.2 The poly(A) tail test (PAT)

To investigate poly(A) tail length in the chosen systems, poly(A) tail tests (PATs) were used. Following ligation of an anchor to the 3' end of the mRNA, cDNA was generated, forming the template for PCR which used a gene specific forward primer and reverse primer complementary to the 3' anchor. Using sequence data from NCBI and intensity scanning of agarose gel lanes by the Quantity One 1-D analysis software (Bio-Rad), highly specific, repeatable tail lengths for individual genes could be determined. An overview of the process is described in figure 3.1 A. For the detection of less abundant mRNAs, or those which share high sequence homology with other genes, nested PATs can be performed. Nested pats involve the ligation of a longer anchor such that two subsequent PCRs can be carried out with an outer and inner primer pair. Specificity is enhanced by using two gene-specific primers, and the extra round of PCR enables increased amplification for low abundance mRNAs.

One recurring problem of PAT gels is the appearance of larger than expected products which appear to track the changes of lower bands. The appearance of these bands was minimised by running in the cold room. A previous lab member demonstrated these to be artefacts of the PAT test caused by misalignment of poly(A):poly(T) stretches in the double stranded PCR products. Misalignment through repeated melting and re-annealing steps during PCR results in loops of unpaired A or T bases, which can pair with unpaired T or A stretches in other fragments, creating multimers (figure 3.1 C). Multimer bands were therefore excluded from size analysis to avoid falsely long poly(A) tail estimates. The chance of

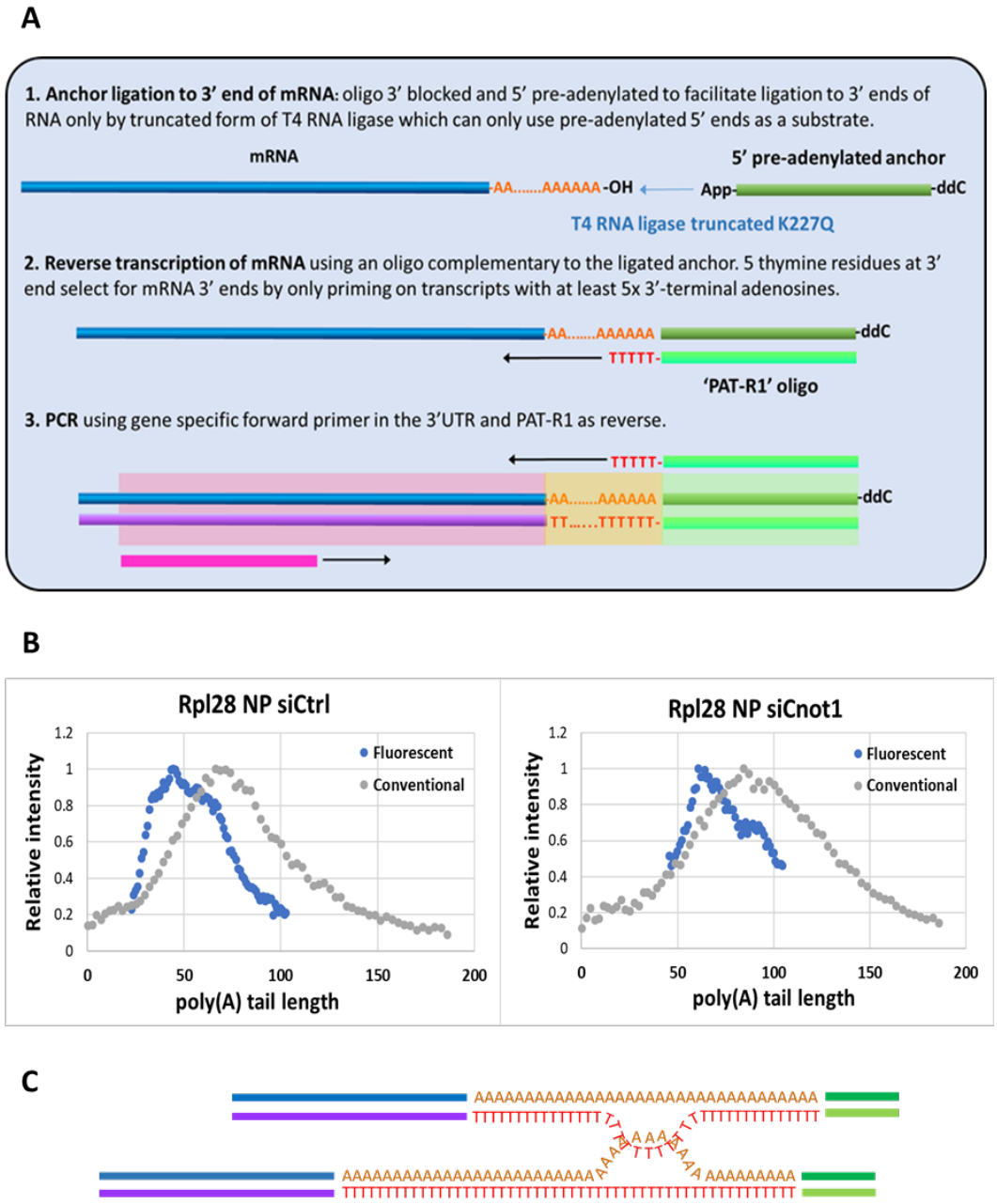


Figure 3.1 The poly(A) tail test. A) Schematic describing specific addition of a DNA anchor to the 3' ends of sample RNA, followed by a reverse transcription step which selects for ends with at least 5 adenosine residues to minimise contamination with internal regions of RNA. Reverse transcription is followed by PCR with a gene specific forward primer in the 3'UTR, which only differ through differences in polyadenylation, or rarely through 3'UTR variation. **B)** Comparison of tail lengths derived from image analysis of conventional PAT gels and fluorescent PAT PCR products subject to denaturing capillary gel electrophoresis. Data shown for Rpl28 mRNA in the nucleoplasmic fractions of control and Cnot1 knockdown cells. **C)** Schematic of multimer formation following repeated melting and re-annealing of fragments during PCR.

this hybridisation occurring could in theory increase with tail size, which may result in disproportionate multimerization of long-tailed mRNAs, however this has not previously been tested. If it were the case, multimerization could cause a concomitant removal of long-tailed species from the monomeric population. Since the PAT assay was routinely carried out use a native rather than denaturing gel, multimerisation could perhaps have led to an underestimate of tail length. To investigate this, PAT PCR was performed using a fluorescent reverse primer and the generated products were run on denaturing capillary gels (capillary gel electrophoresis performed by Matthew Carlisle at DeepSeq, University of Nottingham). Contrary to expectation, PAT PCR products separated by the denaturing gel resulted in a measurement of peak tail length that was approximately 20 nt shorter than when run on a native agarose gel (figure 3.1 B). Normalisation of signal to account for the linear relationship between SYBR Safe staining and fragment length was already performed during analysis of native agarose gels, so a bias in staining could not explain this discrepancy. This confirmed that exclusion of multimers did not remove long-tailed products from native PAT analysis. The difference in derived length between native and denaturing PAT gels may instead have stemmed from the use of a different size marker, or from the abrogation of any slight differences in migration efficiency between size marker and PCR product that may normally exist in native gels. Regardless, differences in estimated tail length between samples were consistent between electrophoresis methods, suggesting that formation of multimers did not qualitatively affect the conclusions drawn (figure 3.1 B).

Concerns have been raised about the reliability of PCR-based methods in determining the size of homopolymeric regions due to repeated dissociation and inaccurate re-joining of the polymerase. Although the likelihood of re-joining the tail sequence before and after the correct position are equally likely, the chance of these events occurring increase with region size. Longer tails would therefore be affected with higher frequency and shorter tails more likely to remain short, potentially leading to a bias towards short tails.

To test the validity of the RL2-PAT assay, RNase H northern blots were performed for representative mRNAs. As RNase H degrades RNA at DNA:RNA hybrids, gene

specific antisense oligos were used to cleave the RNA at a certain position in the 3'UTR, generating transcripts short enough to visualise differences in poly(A) length at sufficient resolution. Oligo(dT) was additionally added to some samples to produce deadenylated controls. RNA was denatured by glyoxylation and separated by agarose gel electrophoresis. Initial attempts using RNase H resulted in unintended degradation of poly(A) tails in reactions containing only gene-specific antisense oligo, with no oligo(dT) (data not shown). This problem was narrowed down to genomic DNA contamination in the sample and was rectified by extended DNA treatments, or by isolation of the cytoplasmic fraction which should have contained less DNA.

In the meantime, full length Rpl28 mRNA was assayed since the gene body is short enough (496 nt) that changes in poly(A) length should still have been visible at sufficient resolution. Comparison between methods applied to the same sample suggested that native PATs may indeed underestimate poly(A) tail length (figure 3.2). While there was only a small difference in peak length (47.4 PAT, 63.4 northern), the northern blot also appeared to contain a heterogenous population of longer tails. Comparison of the medians (63.5 PAT, 97.3 northern) and means (82.7 PAT, 107.1 northern) of the two distributions was perhaps more descriptive for this difference than comparing modal values, as the difference between mean and median was greater for the PAT, indicating a greater skew. Inconsistency between distributions suggested that some long tails may have been missed in the Rpl28 PAT. It is possible however, that overloading or incomplete denaturation caused inaccurate size determination in the northern blot; in particular, the buffer of the glyoxal gel system must be constantly recirculated to avoid localised occurrence of pH>7.0 and consequent loss of denaturation. Furthermore, the size markers themselves advised they were not suitable for size determination, however no alternatives were available.

Upon solving the DNA contamination problem, samples from 0, 20 and 60 minute serum stimulation time courses were cleaved with an antisense oligo for either Rpl28 or Egr1. For each timepoint, RNase H cleavage was carried out in the presence or absence of oligo(dT) to provide deadenylated controls. To avoid DNA contamination, the RNA assayed for Rpl28 poly(A) length was derived from

cytoplasmic fractions and the RNA for Egr1 tail length was DNase treated for 1 hour with Turbo DNase. In both cases, resistance of the poly(A) tail to degradation by RNase H alone was tested by PAT prior to carrying out the larger reactions for the northern blots (figures 3.3 A, 3.4 A). Although PAT primers were obtained which amplified within the region remaining after RNase H treatment, in practice these

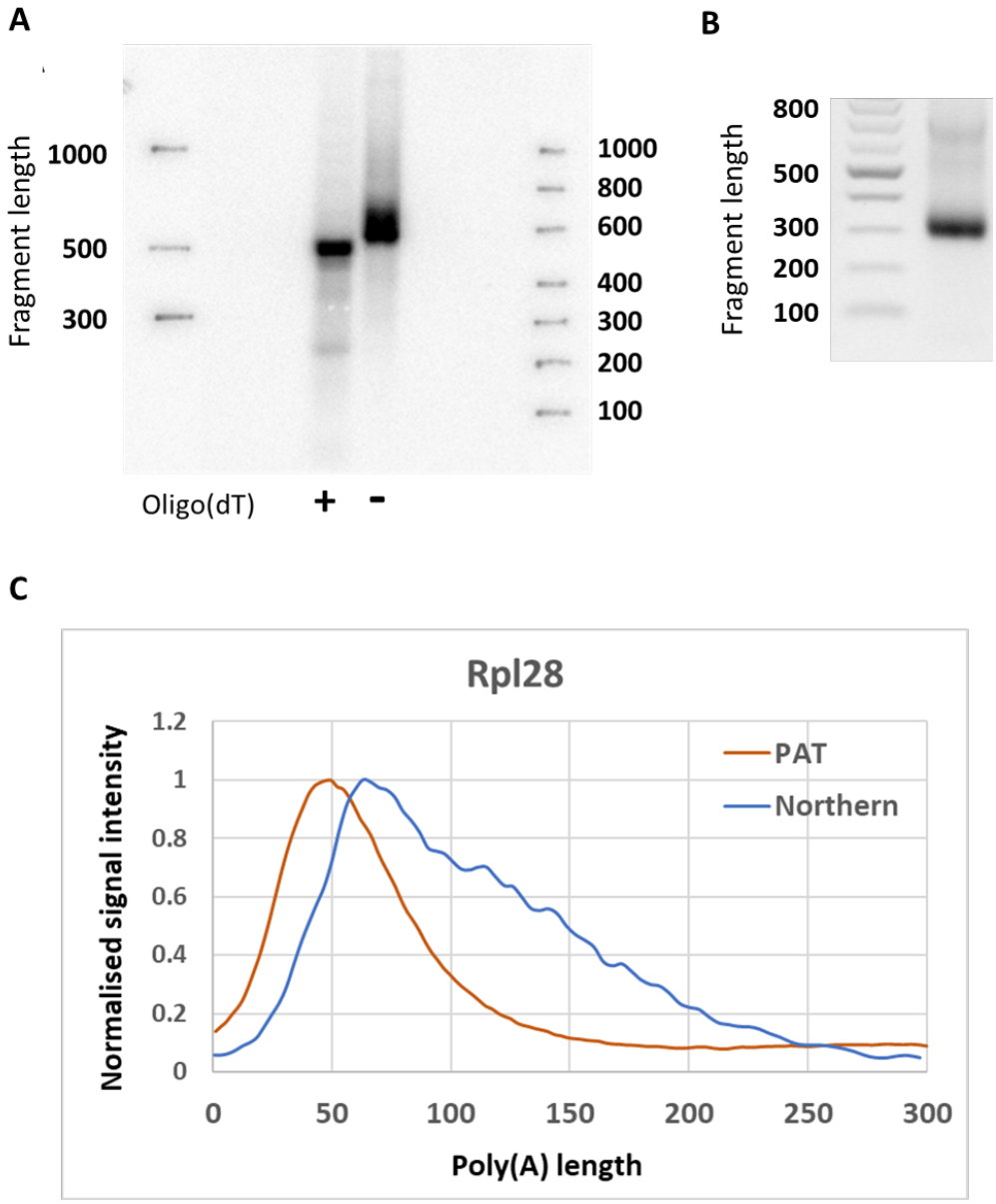


Figure 3.2 The PAT assay and northern blot give different estimated poly(A) lengths for full length Rpl28 mRNA. A) Northern blot of full-length Rpl28 mRNA (deadenylated length 496 nt) using ^{32}P probe. Samples with or without RNase H treatment with oligo dT are compared. **B)** Poly(A)+ sample from A, subjected to RL2-PAT using agarose gel stained with SYBR safe. **C)** Comparison of tail length distributions derived from scanning PAT or northern blot images of the same untreated sample.

only worked on untreated RNA; thus, these PATs also provided the closest data which could be obtained for comparing lengths estimated by each method.

Using RNase H with an oligo designed to bind 199 nt upstream of the cleavage site, modal estimated poly(A) tail length of cytoplasmic Rpl28 mRNA was remarkably similar between the PAT and northern blot methods for all three time points (figure 3.3 D). Since the northern gel experienced some frowning (figure 3.3 B), poly(A) length at each time was derived by subtracting the length value at the peak of the paired deadenylated sample, rather than an exact 198 nt. There were additional bands below the expected fragment sizes which were particularly evident in the RNA which underwent RNase H treatment with the specific oligo but not oligo(dT) (figure 3.3 C). These were likely to be off-target cleavage products which could perhaps have been minimised by shorter incubation with RNase H, or the use of less enzyme. Although the modal poly(A) lengths were similar between methods, this figure did not capture differences in the distributions. As with full length Rpl28, the northern blot indicated a less homogenous population than the PAT, with evidence of slightly longer tails in the northern blot, however this was diminished compared to the full length Rpl28 blot. This again suggested that some longer poly(A) tails could be missed in the Rpl28 PATs. Additional comparison of the means and medians was not possible due to the presence of additional RNase H digestion products at shorter lengths in the northern blots.

To validate the PAT for longer poly(A) tails, DNase-treated RNA was cleaved with RNase H and a specific antisense oligo for Egr1, which binds 169 nt upstream of the annotated cleavage site. Prior to treatment with the specific oligo, a test reaction was carried out with RNase H alone and showed that no DNA endogenous to the sample was causing unintended cleavage (figure 3.4 A). Methylene blue staining of the membrane revealed very slight frowning of the gel (figure 3.4 B). Unexpectedly, hybridization with a 200 nt probe for the 3’ end of the Egr1 mRNA returned two bands which were particularly clear in the oligo(dT) treated lanes (figure 3.4 C). The two bands may have represented different 3’ UTRs, or could have been caused by non-specific binding of oligo(dT) to A-rich regions in the 3’ UTR. Since the bands were either side of the predicted 168 nt deadenylated length, the two measurement methods were compared using both the long and the short band as the

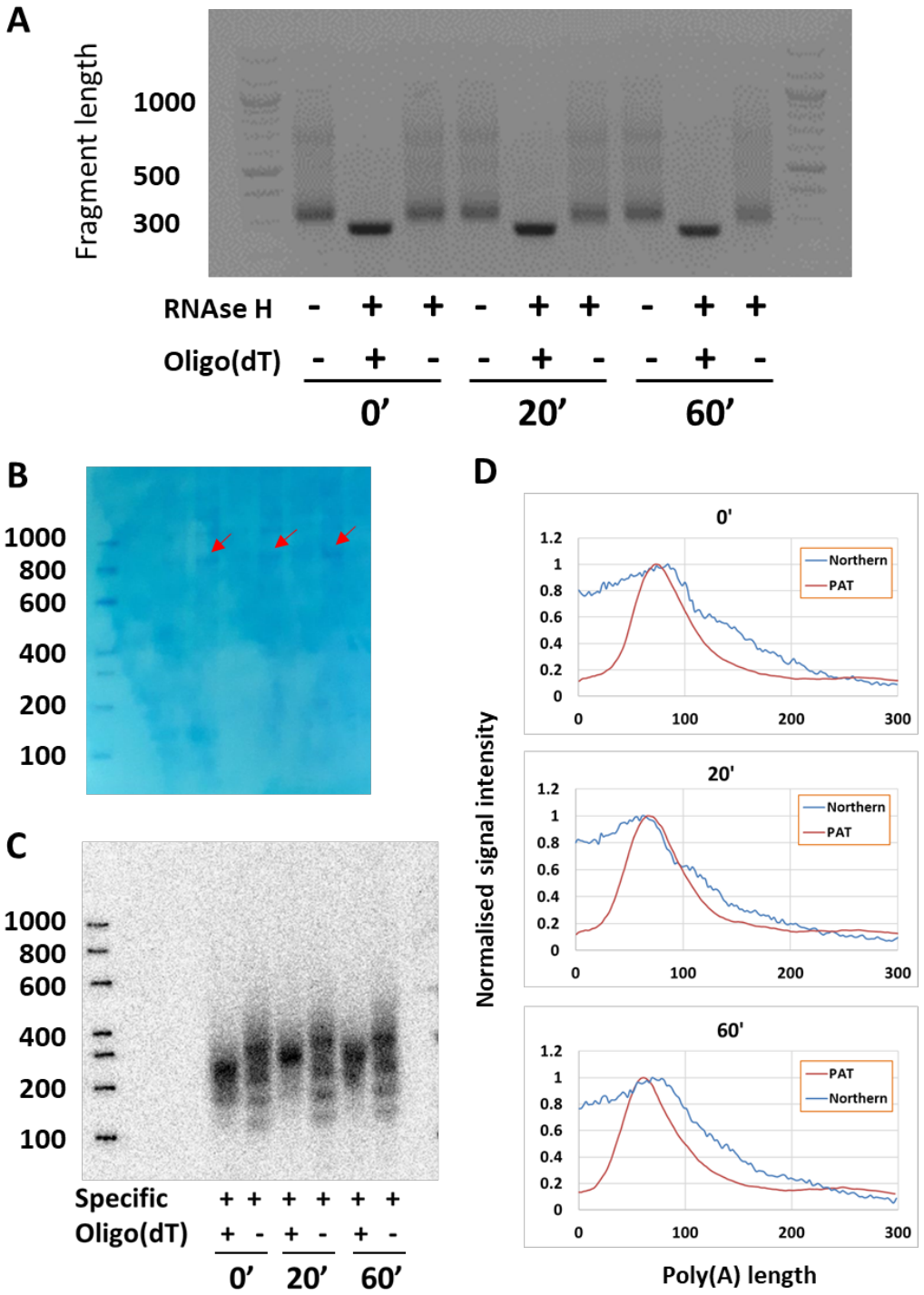


Figure 3.3 Comparison between PAT and RNase H northern blot measurements for Rpl28. A) RL2-PAT assay, using agarose gel and stained with SYBR safe, for RNA isolated from the cytoplasmic fraction of cells stimulated with serum for 0, 20 or 60 minutes, then treated with RNase H alone, or RNase H and oligo(dT). The untreated and RNase H only bands are similar, indicating no DNA contamination. **B)** Methylene blue staining of a section of the nitrocellulose membrane, showing slight frowning of the gel (compare position of bands indicated by red arrows). **C)** Blot incubated with a ^{32}P labelled probe for the 3’-terminal 200nt of the Rpl28 mRNA. All samples were cleaved with the same gene-specific oligo, and half of the samples were also deadenylated using oligo(dT). **D)** Poly(A) tail length distributions of the two methods for each time point.

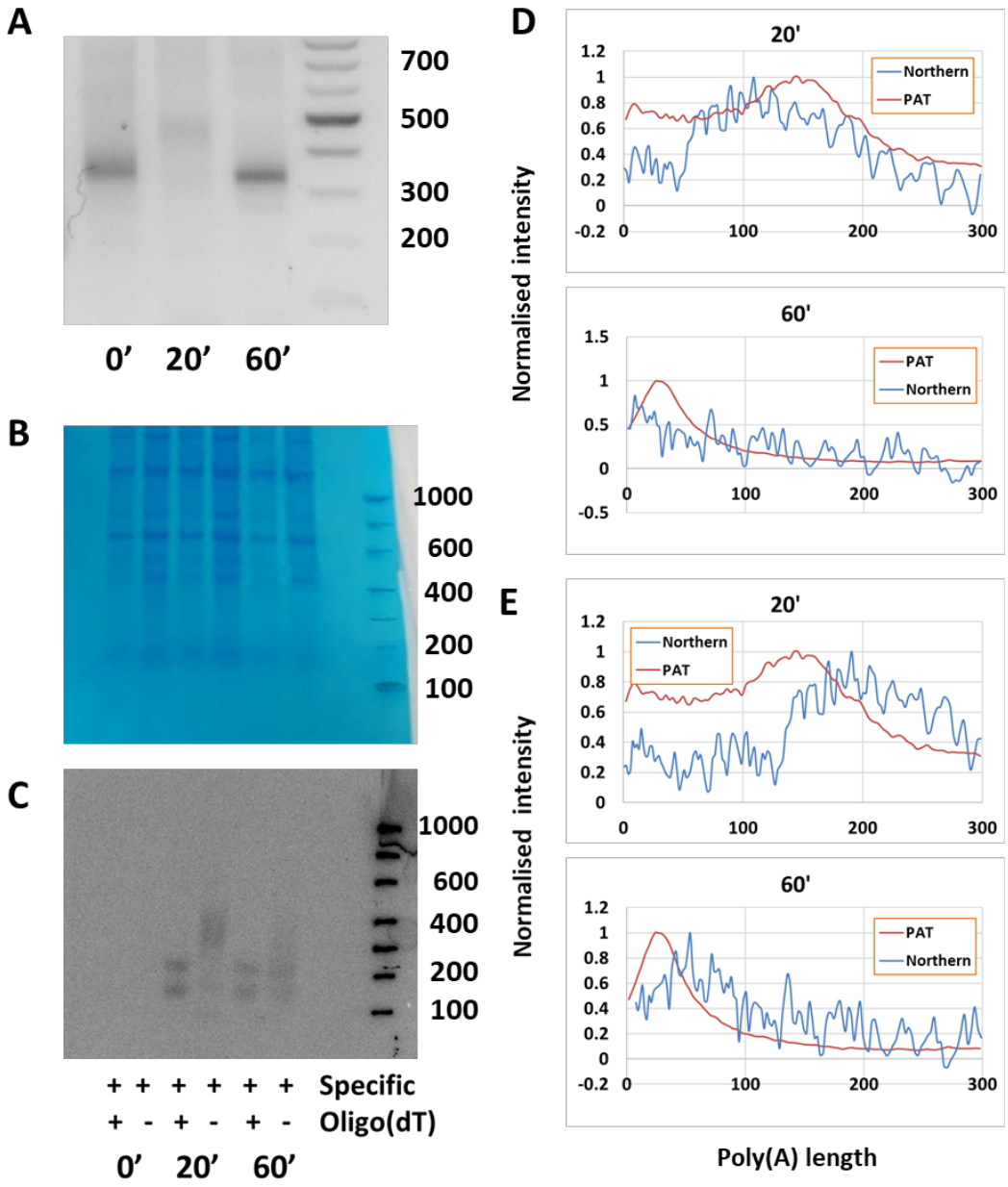
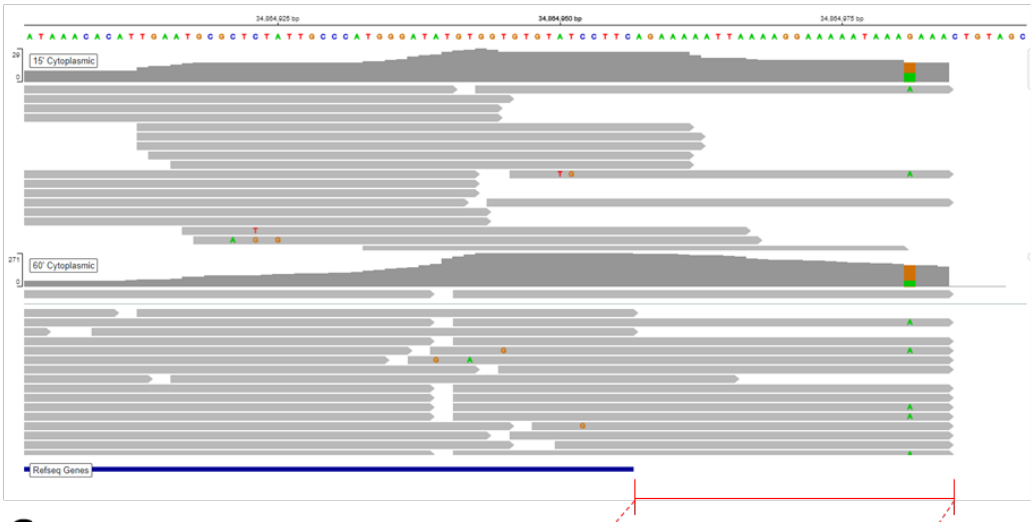
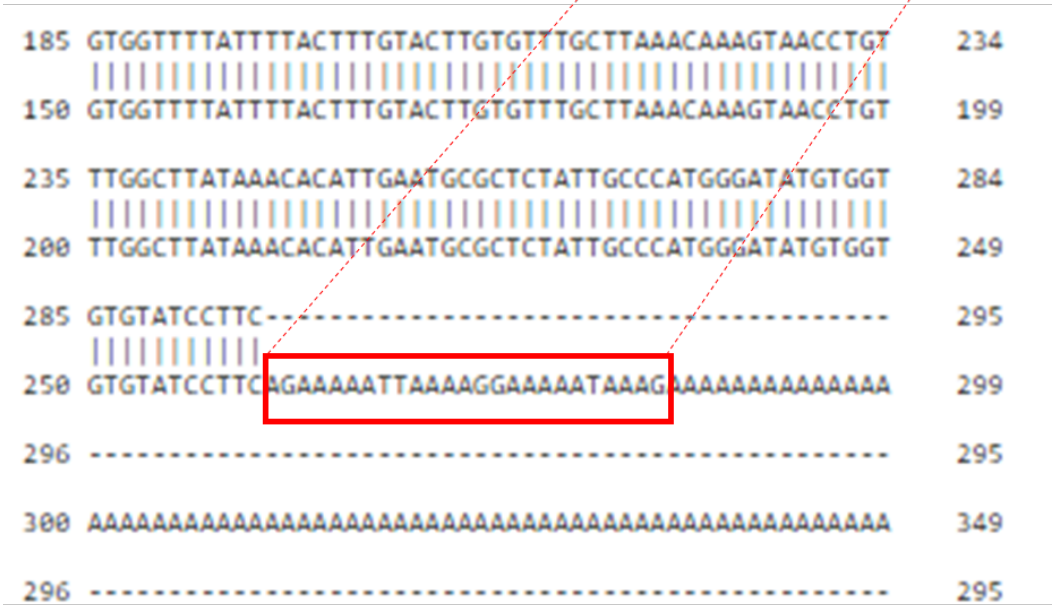


Figure 3.4 Comparison between PAT and RNase H northern blot measurements for Egr1 (continued overleaf). **A)** RL2-PAT assay, using agarose gel stained with SYBR safe, confirming the presence of long poly(A) tails on Egr1 mRNA following incubation with RNase H. Note: no oligos were included in treatment as this adversely affected the PAT assay. **B)** Methylene blue staining of northern blot showing slight gel frowning. **C)** Northern blot with probe against the 3’-most 200 nt of the Egr1 mRNA. **D, E)** Comparison of estimated poly(A) distributions between the PAT assay and northern blot using either the higher (D) or lower (E) band in the oligo(dT)+ lane as the deadenylated length. **F)** IGV viewer tracks for PQ-Seq 50 nt forward read alignments of 15 or 60 minute serum stimulated cytoplasmic fractions along the Egr1 gene. **G)** Sequencing results of the Egr1 PAT product confirming the same longer than expected 3’ UTR as in PQ-Seq data. **H)** Comparison of estimated poly(A) distribution from the PAT and northern blot methods, assuming a 3’UTR which is 25 nt longer than annotated for the PAT, and using the longer band in the oligo(dT) lanes as the deadenylated lengths for the northern.

F



G



H

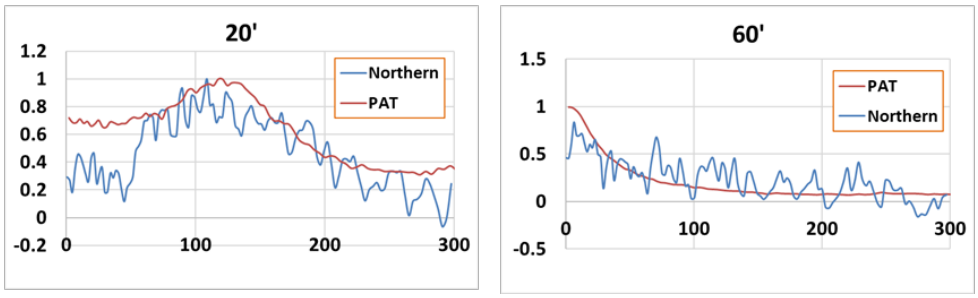


Figure 3.4 continued.

deadenylated lengths for the northern. While the shapes of poly(A) distributions were remarkably similar between methods, the PAT either under or over-estimated poly(A) length compared to the northern, depending on which of the +oligo(dT) bands was used as the deadenylated length (figure 3.4 D, E).

To see whether alternative 3'UTR isoforms were likely, forward read alignments for cytoplasmic fractions of 15 and 60 minute serum stimulated NIH 3T3 cells were visualised against the mouse GRCm38 genome in the IGV web viewer (612). This revealed reads extending approximately 25 nucleotides further than the annotated cleavage site (figure 3.4 F). This longer UTR was confirmed by sequencing of the PAT product (figure 3.4 G). Using the longer value for the deadenylated PAT product, estimated poly(A) distributions for the northern blot and the PAT assay much more closely resembled each other (figure 3.4 H). Weak signal from the northern blot and the consequent 'lumpy' distribution made it difficult to establish an accurate figure with which to calculate the differences in modal values between methods. The longer deadenylated length was taken forwards for subsequent analyses of Egr1 PATs, though frequent conversion of
G → A at the final G residue suggested that the exact cleavage site may in some cases have been 4 bases earlier (figure 3.4 F). Since the additional section of 3' UTR was very A rich, it was possible that the two bands in the oligo(dT) treated lanes were caused by inconsistent removal of this section.

Overall, PAT assays using agarose gels appear to detect differences in poly(A) length reliably in comparison with denaturing gel PATs and northern blotting, though there are subtle differences in distributions and absolute lengths obtained (figures 3.3 D, 3.4 H). Given their reproducibility and additional convenience compared with other methods, native agarose gel PATs were used to measure poly(A) tails for the rest of the study.

3.3 Poly(A) tail length is regulated in the nucleoplasm and on the chromatin

In the NIH 3T3 serum response, long poly(A) tails (~150-200 nt) are detectable on transcriptionally induced immediate early gene transcripts within 15-20 minutes of serum addition (587, 611). By about an hour after induction, transcription is much

reduced and poly(A) tails are short (<30 nt). In contrast, constitutively expressed mRNAs possess only medium length (~30 – 70 nt) poly(A) tails which do not change in response to serum addition (11, 325, 326, 423, 611). The observation that most mRNAs have tails shorter than the 200 nt which is thought to be added during synthesis is usually attributed to cytoplasmic deadenylation. Previous mathematical modelling work however, indicates that the poly(A) tail distributions observed late in the serum response cannot arise without either a change in deadenylation or degradation rate, or a change in initial poly(A) length (611). Examination of nuclear and 4SU-labelled mRNA confirmed that initial poly(A) tail is indeed regulated prior to export into the cytoplasm.

To probe the discovery of nuclear poly(A) tail regulation further, chromatin fractions were isolated. Fractionation efficiency was checked by comparing enrichment of unspliced over spliced mRNA in each fraction, and normalising to the value for the cytoplasm. DNA contamination was also tested by performing qPCR for the unspliced Rpl28 transcript on cDNA and no-reverse transcription controls. For both replicates, unspliced enrichment was around 100-fold greater in the nucleoplasmic fraction and at least 1000-fold greater in the chromatin-associated fraction when compared to that of the cytoplasm, indicating good separation (figure 3.5 B, E). Some DNA contamination was evident, mainly in the chromatin fractions (figure 3.5 C, F), however this should not have affected results of the PAT assay since the RNA ligase used can only ligate to the 3’ end of RNA (though it can use the pre-adenylated 5’ end of either RNA or DNA).

For the majority of mRNAs, tail length was defined before the RNA was released from the chromatin (figure 3.5 A,D). The exception seemed to be the Sqstm1 mRNA which possessed a long chromatin associated tail but medium length nucleoplasmic and cytoplasmic tails at all time points. This strongly indicated a case for nucleoplasmic deadenylation of a subset of mRNAs. There also seemed to be some evidence of this for Actb in the first replicate, but a mid-length tail was present on the chromatin also (figure 3.5 A). This could have been due to differences in fractionation purity between replicates, or could have been caused by slight differences in the time of harvest or quality of induction. Likewise, in the second replicate there appeared to be some long poly(A) tails in the 50 minute chromatin

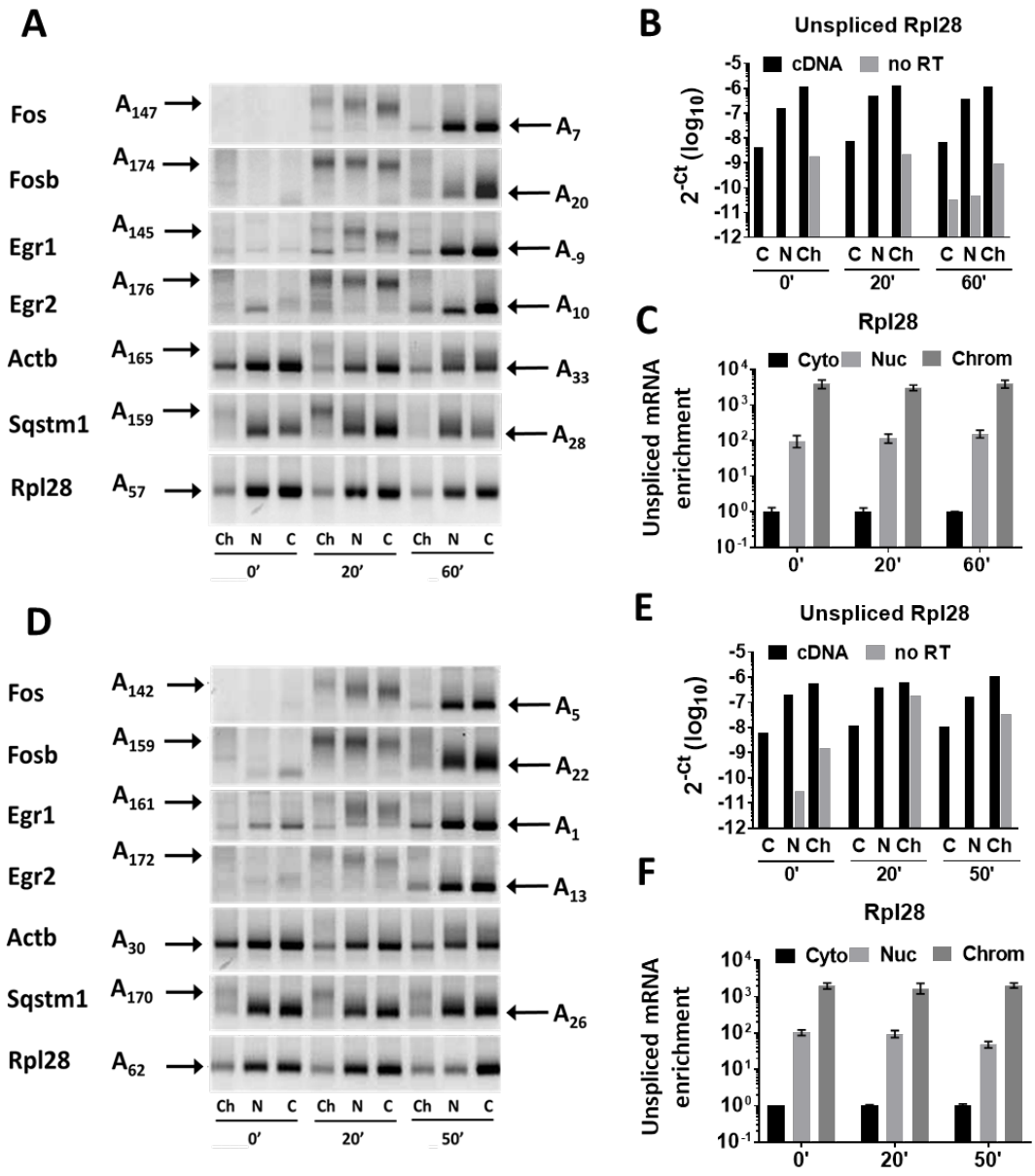


Figure 3.5 Poly(A) tail is regulated at multiple stages. Results shown for two biological replicates. **A, D)** Chromatin associated, nucleoplasmic and cytoplasmic poly(A) tails measured by RL2-PAT, using agarose gel and staining with SYBR safe, at the time points indicated. Arrows point to maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning . Notably, most tail length control seems to occur while the mRNA is chromatin associated, except in the case of Sqstm1 whose tail is trimmed in the nucleoplasm. **B, E)** qPCR of unspliced Rpl28 mRNA shows evidence of DNA contamination in the chromatin fractions, which is not detectable by the PAT assay. **C, F)** qPCR of mature and unspliced mRNA show enrichment of unspliced over spliced transcripts in the nucleoplasmic and chromatin fractions.

fraction for Fosb, suggestive either of the synthesis of both long and short tails, or the complete chromatin-associated deadenylation of some transcripts (figure 3.5 D). Rpl28 was the only transcript examined for which a long poly(A) tail was never observed.

3.4 Long poly(A) tails are associated with increased stability of unstable mRNAs

The poly(A) tail is thought to affect stability of the associated mRNA, however studies of endogenous transcripts in mammalian somatic cells are limited and suggest this relationship may not be universal (326, 587). To investigate the relationship between tail length and mRNA stability, transcription was inhibited either early or late in the serum response using 40 μg/mL Actinomycin D (AcD). 20 and 50 minutes were chosen since at these times, all induced mRNAs of interest tended to have either long or short poly(A) tails respectively. Samples were harvested at 10 minute intervals following AcD addition and mRNA levels measured by qPCR to assess degradation rate. Cessation of transcription was checked by measuring unspliced mRNA levels (figure 8.2). Unexpectedly, unspliced levels of Fosb did not decrease rapidly upon AcD treatment, even upon increasing the dose from 5 to 40 μg/mL. This could either have resulted from a failure to inhibit transcription or from delayed splicing. Since it was not clear which of these was the case, Fosb was excluded from the analysis.

If a relationship between tail length and stability existed, degradation should have been faster when AcD was added at 50 minutes than at 20. Such a pattern was clear for all three induced mRNAs tested, with very slow degradation of the long-tailed transcripts and rapid decay of the short-tailed transcripts (figure 3.6). As the poly(A) tail of constitutively expressed mRNAs did not change by any measurable degree, it was not possible to determine a relationship by this method, however the demonstrably stable Rpl28 mRNA was included as a control.

In dealing with endogenous mRNA populations rather than carefully controlled levels of some synthetic species, this method could not rule out a model in which the much higher amount of RNA present at 20 minutes saturated some specificity

factor responsible for targeting degradation machinery to the mRNA, thereby eliciting a proportionally slower degradation rate independently of poly(A) tail size.

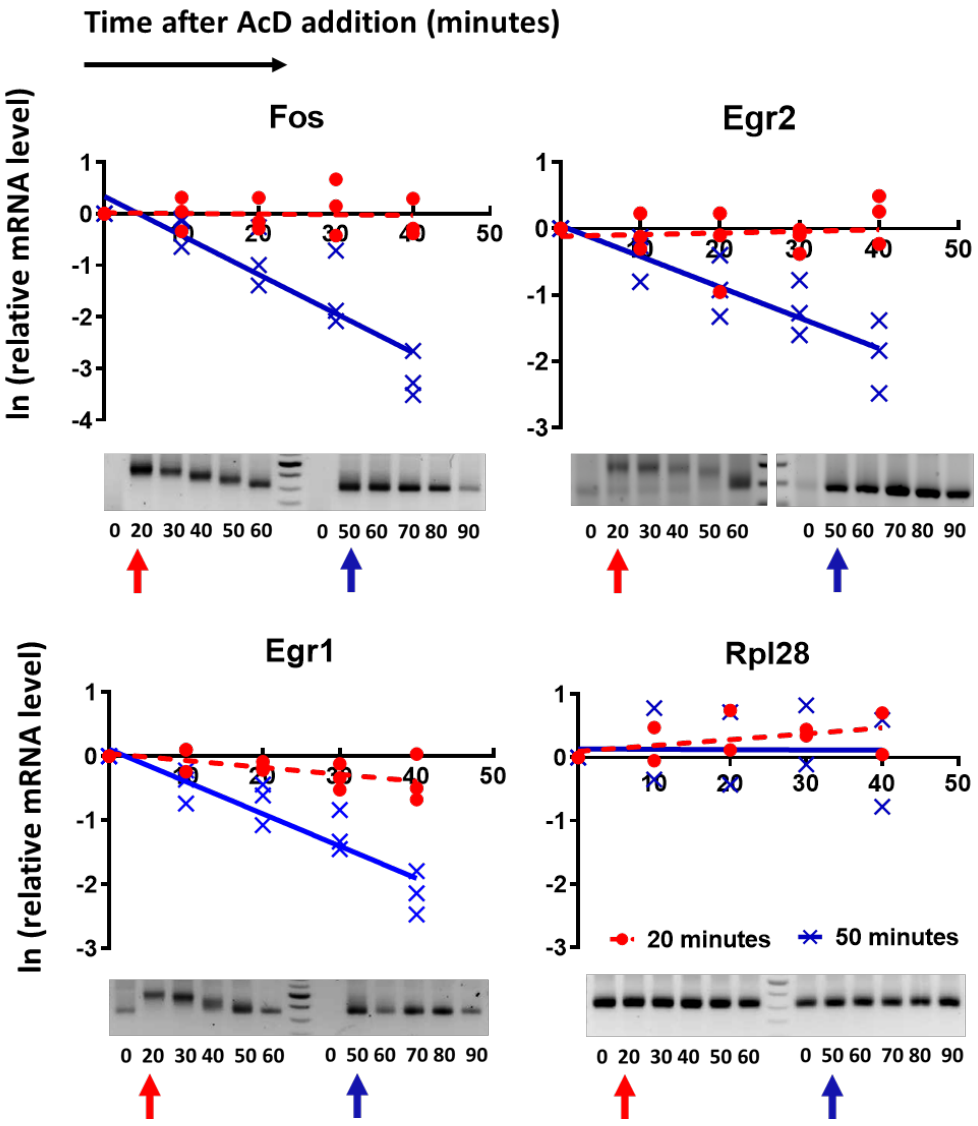


Figure 3.6 Long poly(A) tails are associated with greater stability of unstable mRNAs. Transcription was stopped by adding Actinomycin D after 20 (red dot) or 50 (blue cross) minutes, when poly(A) tails of induced mRNAs are long or short respectively. mRNA abundance at 10 minute intervals following transcription cessation was measured using qRT-PCR, analysed using the ΔΔCt method, with Gapdh as the reference gene. mRNA level was normalised to the level at either 20 or 50 minutes to give fold change, and this fold change transformed by the natural log, since decay kinetics are exponential. A line at x=0 would indicate no decay. 3 biological replicates are shown for the induced mRNAs, and 2 for Rpl28. Representative RL2-PAT gels (agarose, stained with SYBR safe) are shown below for each gene, where the number below indicates minutes after serum stimulation.

3.5 HEK293 cells also exhibit chromatin-associated poly(A) tail regulation

To determine whether the early poly(A) tail length regulation observed was an isolated feature of NIH 3T3 cells, chromatin fractionations were also performed on the HEK293 human kidney cell line. As there is no well characterised serum response in these cells, experiments were performed at steady state. Separation of fractions was again validated using unspliced mRNA enrichment, this time for RPL10A. The magnitude of enrichment was lower than in NIH 3T3 cells, but clear differences between fractions were visible in the PAT assay (figure 3.7). Although only two mRNAs were tested, both nucleoplasmic and chromatin associated poly(A) control were evident (figure 3.7 A). While ACTB mRNA had a medium length tail across all fractions, a mixed population of long and medium length tails was evident on chromatin-associated RPL10A mRNAs. This is different to NIH 3T3 cells where mixed long and medium tails were detected for Actb, while only medium length tails were observed on the ribosomal protein mRNA, Rpl28 (figure 3.5 A, D).

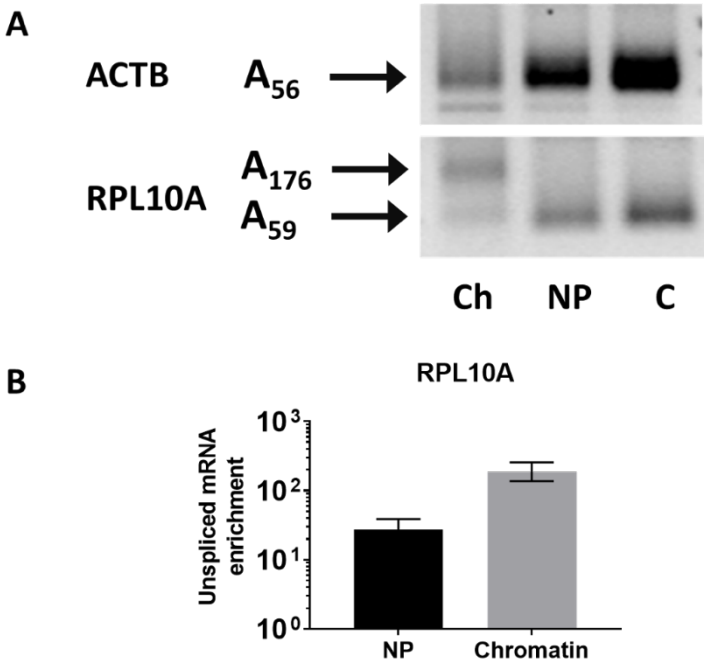


Figure 3.7 Preliminary data suggest varied locations of poly(A) length control in human cells. A) RL2-PAT assays using agarose gel stained with SYBR safe were carried out on RNA isolated from chromatin-associated, nucleoplasmic and cytoplasmic fractions of HEK293 cells. Arrows point to maximum and minimum modal poly(A) tail sizes determined using quantitative gel scanning. **B)** qPCR of mature and unspliced RPL10A mRNA shows enrichment of unspliced over spliced transcripts in the nucleoplasmic and chromatin-associated fractions.

3.6 Discussion

Despite possible inaccuracies in absolute tail length determination, PATs are able to quickly and easily detect small changes in poly(A) tail length between different conditions. Additionally, the tail lengths estimated by these PAT assays are in agreement with those observed by numerous other groups for cytoplasmic length (11, 325, 326), suggesting that these measurements are suitable for interpretation by the scientific community. On the other hand, most high throughput measures of poly(A) tail length also involve a PCR step and as such may be subject to similar biases. When compared to northern blots, the PAT assay underestimated the modal value of the Rpl28 poly(A) tail by 3-16 nucleotides (figures 3.2 B, 3.3 D). The PAT and northern distributions appeared more similar for the long and short tails present on Egr1 mRNA (figure 3.4 H), though the hybridization step should be repeated in order to obtain smoother curves.

Analyses with the PAT assay show that the poly(A) tail is regulated in both mouse and human cells (figures 3.5, 3.7). Furthermore, in both cell lines the poly(A) tails of some mRNAs were determined before their release from the chromatin. As the poly(A) tail enhances stability of some induced transcripts (figure 3.6), early control of tail length could enable tighter regulation of protein synthesis, i.e. by causing those transcripts produced at the end of the transcription pulse to be degraded more rapidly than those at the beginning.

Strikingly the poly(A) tail of Rpl28 exhibited only a medium length tail at all time points and in all fractions of NIH 3T3 cells (figure 3.5). This does not seem to be a conserved feature of ribosomal protein mRNAs across organisms since the RPL10A mRNA had a long poly(A) tail in the chromatin associated fraction of HEK293 cells. Common among housekeeping mRNAs of both organisms were their medium length nucleoplasmic and cytoplasmic poly(A) tails, which were shorter than what might be expected given their high stability. One possibility is that stability is conferred by some other factor; such a factor may not necessarily be poly(A)-independent, but may only require tails of a certain length. One possibility is that these mRNAs are protected by a higher density of PABPC binding, while another is that proteins other than PABPC also bind their tails and inhibit deadenylation either directly or by promoting closed loop formation (465, 494, 499).

The existence of dynamic chromatin-associated and nucleoplasmic poly(A) tail length regulation (figure 3.5) highlights the importance of the poly(A) tail for gene expression. While the decreases which occurred between fractions of a given time point are likely explained by nucleoplasmic and cytoplasmic deadenylase activity, it is not clear how poly(A) tail length of nascent mRNA is controlled.

For several mRNAs mixed long and short-tailed populations existed in the chromatin fraction. Such a pattern could have arisen either through rapid processive deadenylation of initially long poly(A) tails, or through delayed polyadenylation of successfully cleaved transcripts. Both scenarios are in some sense unprecedented since deadenylation is documented to occur much more slowly than would be required here, and the nuclear polyadenylation reaction is coupled to cleavage so would have to be inhibited in some way – perhaps through slow recruitment of PAP itself or PABPN1 (28, 31, 179, 347, 363, 613). A recent publication found evidence of chromatin-associated deadenylation by the Ccr4-Not complex in the Drosophila germline, though this was closely linked to decay of the target transcripts (33). In theory, simultaneous synthesis of long and short-tailed species would also produce a dual population, though this seems functionally irrelevant if poly(A) tail length ultimately converges in the nucleoplasm. While such an event could perhaps be considered at some transient switch, this pattern occurred at multiple time points. Although two bands were visible in several lanes of the Egr1 PAT, we propose that the deadenylated band present across all fractions represented a constitutively produced deadenylated form of the mRNA since it was also detectable prior to serum stimulation.

Whether nascent poly(A) length control is elicited through rapid deadenylation or control of tail synthesis is not yet clear. Given the variation in patterns between genes, it seems possible that the mode of regulation too is gene-specific.

4 CNOT1 is a key regulator of early poly(A) length in NIH 3T3 cells

4.1 Introduction

Two complexes are responsible for the majority of bulk mRNA deadenylation in mammalian cells: CCR4-NOT and PAN2-PAN3 (26, 503). The individual contributions of the complexes under normal conditions has not yet been delineated since they seem able to partially compensate for each other's depletion (529). While PABPC is required for PAN2-PAN3 deadenylase activity, earlier work suggested that it inhibited the CCR4-NOT complex (24, 459, 460). This, along with the observation of biphasic deadenylation dynamics, led to the prevailing model in which PAN2-PAN3 performs an initial trimming of long tails - presumably bound by multiple PABPCs - followed by CCR4-NOT mediated deadenylation resulting in complete tail removal (26). Recent work has confirmed selectivity of PAN2-PAN3 for long tails, but stimulation of the CCR4 nuclease by PABPC at permissible concentrations has also been observed (28, 50).

4.1.1 CCR4-NOT

CCR4-NOT is a large complex which was originally identified as a transcription factor and has been implicated in many processes besides deadenylation (35, 38, 39, 41, 42, 44, 147, 550). In addition to contributing to bulk mRNA deadenylation, CCR4-NOT can be targeted to subsets of mRNAs through interactions with proteins such as Tristetraprolin and Roquin which recognise 3'UTR AU-rich elements (AREs) and stem-loops respectively (45, 255). The complex is also involved in miRNA-induced silencing and some nonsense mediated decay pathways (256, 258, 266, 268).

Integral to the CCR4-NOT complex are two catalytic subunits, each having two isoforms encoded by two genes: CCR4 (*CNOT6*/*CNOT6L*) and CAF1 (*CNOT7*/ *CNOT8*). CCR4 is of the EEP (exonuclease-endonuclease-phosphatase) family, and CAF1 the DEDD type (named due to conserved Asp and Glu residues) (522, 614–616). *In vitro* experiments using reconstituted complexes/components, as well as knockdown and rescue experiments in HeLa cells, have shown that the two nucleases differ in several characteristics, summarised in table 4.1 (28, 31, 50).

CCR4	CAF1	Reference(s)
No[1]/low[2] activity of isolated nuclease	Isolated nuclease active	[1]**Webster (28)** (recombinant *S. pombe* protein, *in vitro* assay), [2]**Chen** (532) (recombinant human subcomplex, *in vitro* assay)
Higher activity in heterodimer and full complex.	Lower activity in heterodimer and full complex.	**Raisch** (31) (reconstituted human [sub]complex, *in vitro* assay)
Stimulated by PABPC	Inhibited by PABPC	**Webster (28)** (recombinant *S. pombe* protein, *in vitro* assays), **Yi** (50) (purified human subcomplex, *in vitro* assay)
More specific for poly(A)	Less specific for poly(A)	**Raisch** (31) (reconstituted human subcomplex, *in vitro* assay), **Chen** (532) (recombinant human subcomplex, *in vitro* assay)
Translation independent	Translation-dependent	**Webster (28)** (*S. cerevisiae, in vivo assay*)
Interacts with complex via Caf1	Directly interacts with complex (binds CNOT1)	**Basquin** (523), **Petit** (524), **Raisch** (31), **Chen** (532)

Table 4.1 Comparison of CCR4-NOT nuclease activities.

Importantly, CCR4 was recently shown to be stimulated by PABPC, whereas CAF1 was inhibited (28, 50). This has implications for their substrate preferences and indeed, deletion of *ccr4* in yeast stabilised the majority of transcripts (to different degrees), whereas deletion of *caf1* only affected low codon optimality mRNAs which are thought to be poorly translated and to exhibit relatively sparse Pab1 (PABPC) binding on their tails (28). Although this is an appealing model, CAF1 bridges the interaction of CCR4 with the rest of the complex, so it is surprising that *caf1* deletion had a more specific effect than deletion of *ccr4* (28, 523). CNOT7 (CAF1) activity can be further modulated depending on incorporation of other proteins into the complex (305). In addition, interaction of CAF1 with PABPC via BTG2 was found to stimulate rather than inhibit its activity *in vitro* (463).

4.1.2 Approach

Poly(A) tails observed late in the serum response - as well as those on constitutively expressed mRNAs - are much shorter than the canonical 200-250 nt initially discovered by pulse labelling (figure 3.5 A, D) (324). The simplest model to resolve this disparity is one in which mRNA is produced with long tails which are then

pruned in the nucleus by a deadenylase to some proper length. Repeated casual observation was that tail length and unspliced level of serum response mRNAs peak soon after induction (~15-20 minutes) and are greatly reduced after one hour,

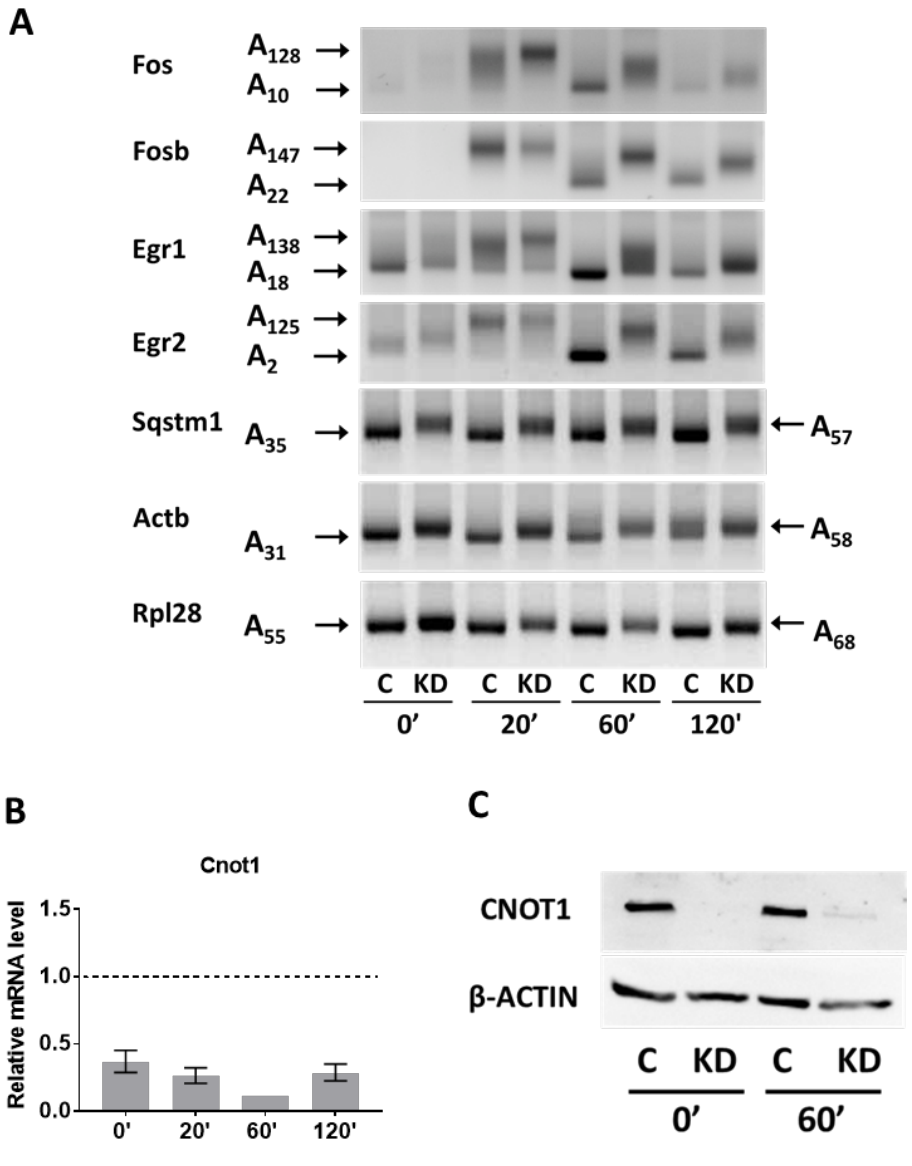


Figure 4.1 Cnot1 knockdown increases total poly(A) tail length in the serum response and this corresponds with slower decay of unstable mRNAs. NIH 3T3 cells treated with control or Cnot1 siRNA and stimulated with serum as indicated previously. Total RNA was isolated. **A)** RNA was subjected to RL2-PAT using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. C and KD denote control and Cnot1 knockdowns respectively. Results shown for one biological replicate. Times indicate minutes of serum stimulation. **B)** Validation of Cnot1 knockdown at the mRNA level for the time course shown in A).

suggesting that tail length may be coordinated with transcription rate for these genes. Since the CCR4-NOT complex has known roles in transcription, it was investigated first (147, 550). In order to reduce the complex's activity, mRNA

encoding the scaffold subunit CNOT1 was depleted using siRNA. The efficacy of the method was validated by western blot and routinely monitored by q-RT PCR. Effect on poly(A) tail length was measured at different points following serum stimulation or in different cell fractions.

4.2 Cnot1 knockdown results in slower deadenylation

Consistent with observations that *not1* deletion is lethal (517), Cnot1 proved slightly resistant to knockdown at the mRNA level, however the same knockdown protocol resulted in a large reduction in CNOT1 protein levels (figure 4.1 B, C). As expected, reduced CNOT1 level was concordant with slower reduction in poly(A) tail length of induced mRNAs following serum induction (figure 4.1 A). In addition, a modest increase in tail length of constitutively expressed mRNAs was observed which remained consistent throughout the response (figure 4.1 A). Despite this increase, tails of these abundant mRNAs still fell far short of the 200 nt figure first identified.

4.3 Cnot1 knockdown increases nuclear poly(A) tail length

Although deadenylases are thought to act in the cytoplasm, the CCR4-NOT complex has several documented nuclear functions (147, 617–619). One study subsequent to this work discovered chromatin-localised Ccr4-Not deadenylase activity, mediating Piwi-directed degradation of telomere and transposon-derived transcripts in the *Drosophila* germline (33).

To determine whether the dependence of poly(A) tail length on CNOT1 extended to earlier in the mRNA's lifespan, nuclear and cytoplasmic fractions of knockdown cells were examined. Cells were serum stimulated for 60 minutes as this is when the difference in tail length is largest for total RNA. Knockdown of Cnot1 at the mRNA level was confirmed by relative qRT-PCR for each replicate (figure 4.2 D). Validation of fractionation by unspliced Rpl28 mRNA enrichment is shown for each replicate in figure 4.2 E. Separation was confirmed by nuclear enrichment of the noncoding RNA Malat1 (620), and the method validated by western blotting for Lamin A/C and α-tubulin (figure 4.2 F, G). Analysis of RNA from these fractions by RL2-PAT showed

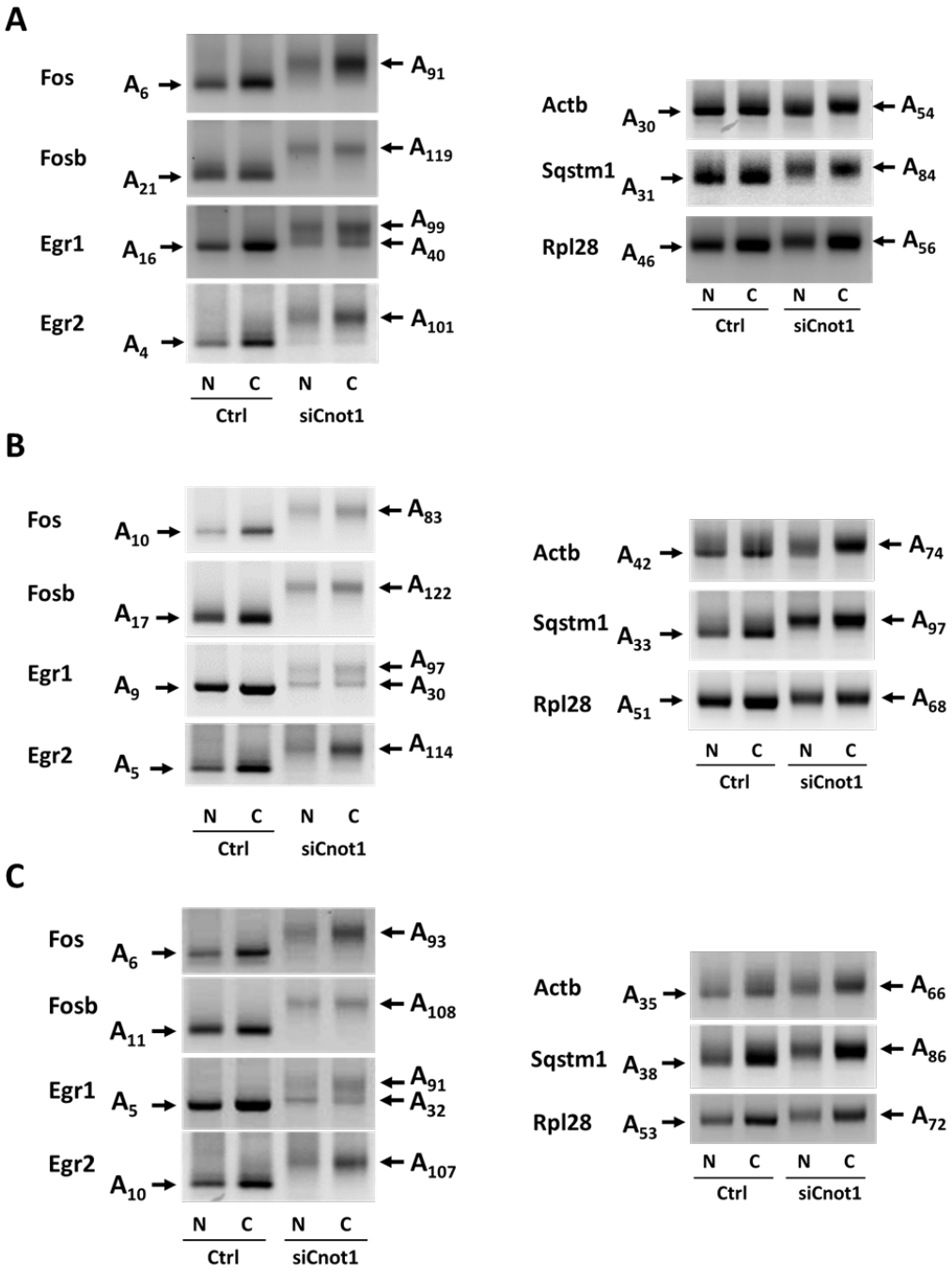


Figure 4.2 Nuclear poly(A) tail size is increased in Cnot1 knockdown (continued overleaf). NIH 3T3 cells were treated with control or Cnot1 siRNA and serum stimulated for 60 minutes. Nuclear (N) and cytoplasmic (C) RNA was isolated and subjected to RL2-PAT using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. 3 biological replicates (**A-C**) shown. **D)** Validation of knockdown at the mRNA level by qPCR for each replicate, using the ΔΔCt method with Gapdh as the reference gene. **E)** Validation of nuclear/cytoplasmic separation of each biological replicate through comparison of unspliced:spliced mRNA ratios. **F, G)** Further validation of fractionation method by F) nuclear enrichment of Malat1 RNA (qPCR, 4 biological replicates) and G) western blotting for Lamin A/C (nuclear) and α-tubulin (cytoplasmic).

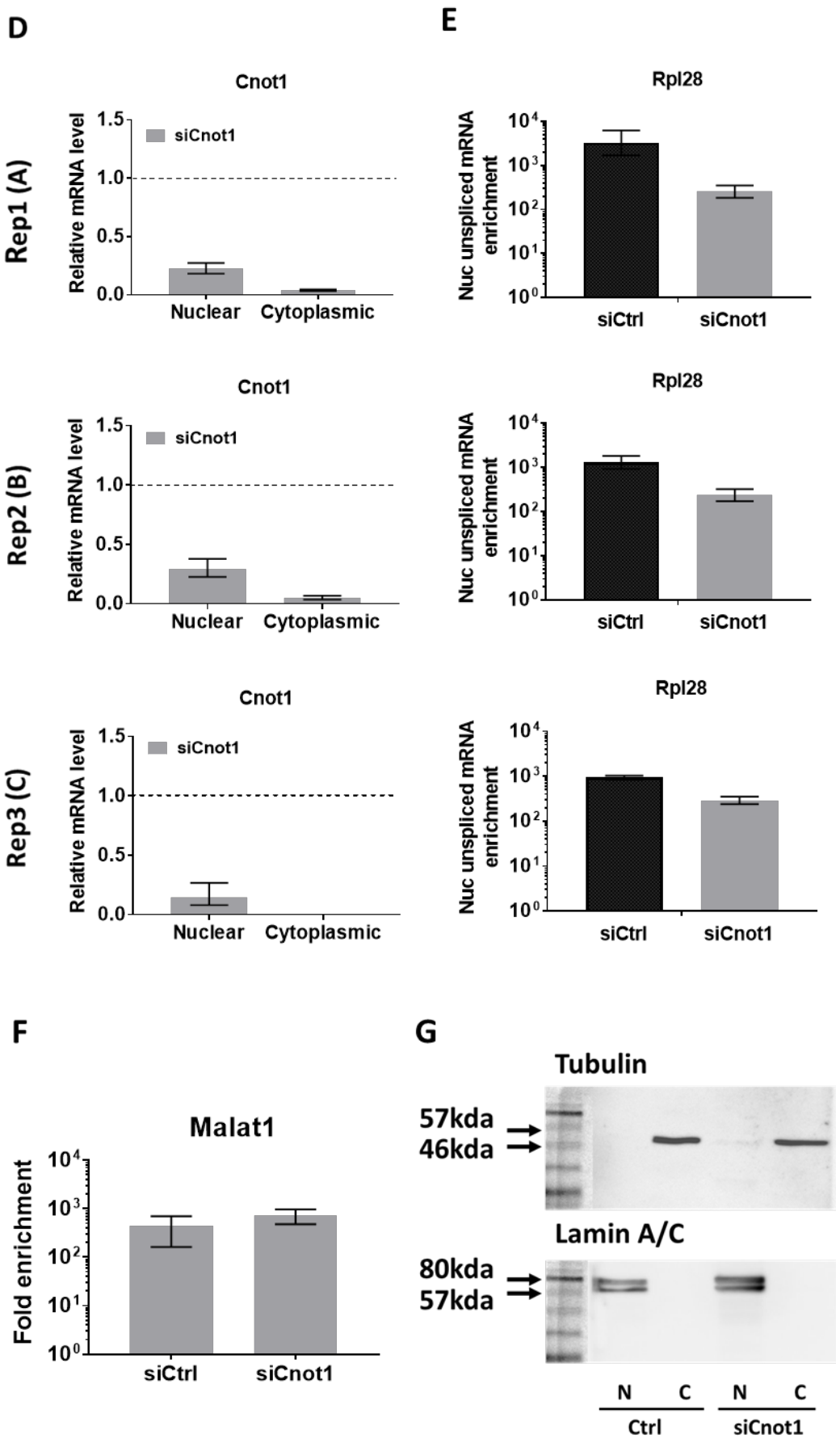


Figure 4.2. contd.

that the increases in tail length observed following depletion of CNOT1 were primarily nuclear rather than being a consequence of reduced cytoplasmic deadenylation (figure 4.2 A-C). Notably, the increase in tail length induced by Cnot1 knockdown was greater in serum-induced mRNAs than in housekeeping mRNAs, and the longer tails observed on serum-induced mRNAs following knockdown were shorter than those observed after 20 minutes in control cells (figure 4.1 A). Together, these results suggest that CCR4-NOT normally deadenylates transcripts in the nucleus and that incomplete depletion of CNOT1 allowed some nuclear tail shortening to still take place.

One might expect that reduction in available deadenylase machinery would lead to full deadenylation of some tails while others remain untouched (in the case of processive activity) or result in a range of partially digested tails of different lengths (in the case of distributive activity). Neither scenario appeared to arise here (figure 4.2 A-C); the bands themselves were relatively sharp rather than smeared, suggesting that trimming to a defined length was still occurring, but that this length was altered by Cnot1 knockdown. In addition, the longer tails seen for housekeeping mRNAs were particularly striking, since these were longer than the those observed at any point for the same genes in control cells (figure 4.2 A-C, figure 4.1 A).

4.4 Cnot1 knockdown affects Chromatin associated mRNA

To investigate how early CNOT1 acts to limit poly(A) tail length, chromatin fractionations were performed on Cnot1 knockdown cells. Cells were harvested after 50 rather than 60 minutes to try and increase the yield from the chromatin fraction, while maintaining short tails in the control cells. Separation of fractions was again confirmed by determining unspliced Rpl28 mRNA enrichment via qRT-PCR, and DNA contamination was measured using primers for Rpl28 or Egr1 pre-mRNA to compare amplification of cDNA with no-reverse transcription controls. Separation of fractions was good for both replicates (figure 4.3 C) and only minimal DNA contamination could be observed which was more pronounced in the first replicate, particularly in the knockdown samples (figure 4.3 D). Knockdown of Cnot1 at the mRNA level was successful in both replicates, with slightly greater efficiency in the second (figure 4.3 E).

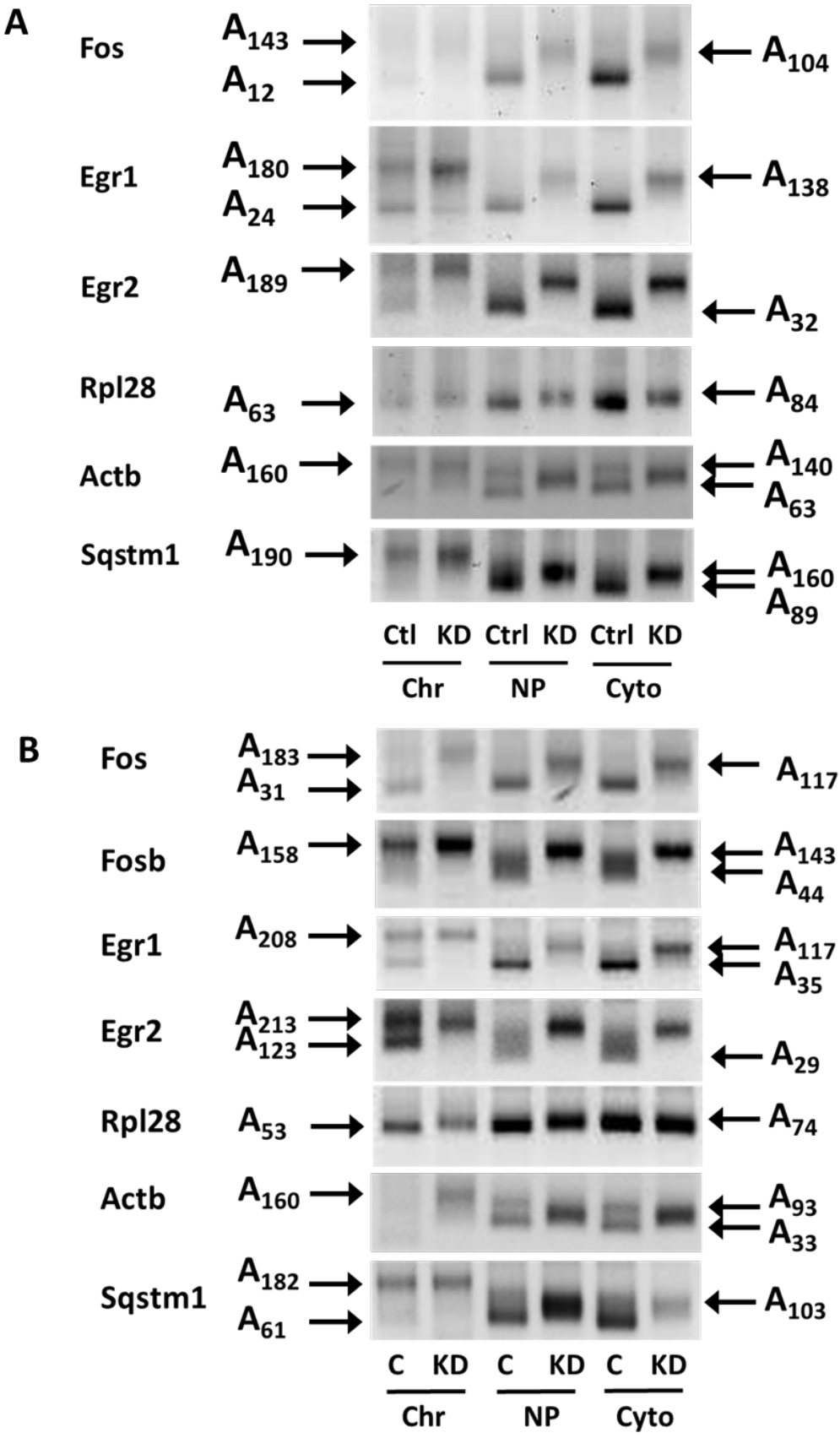


Figure 4.3 Cnot1 regulates poly(A) tail length at different stages (continued overleaf).

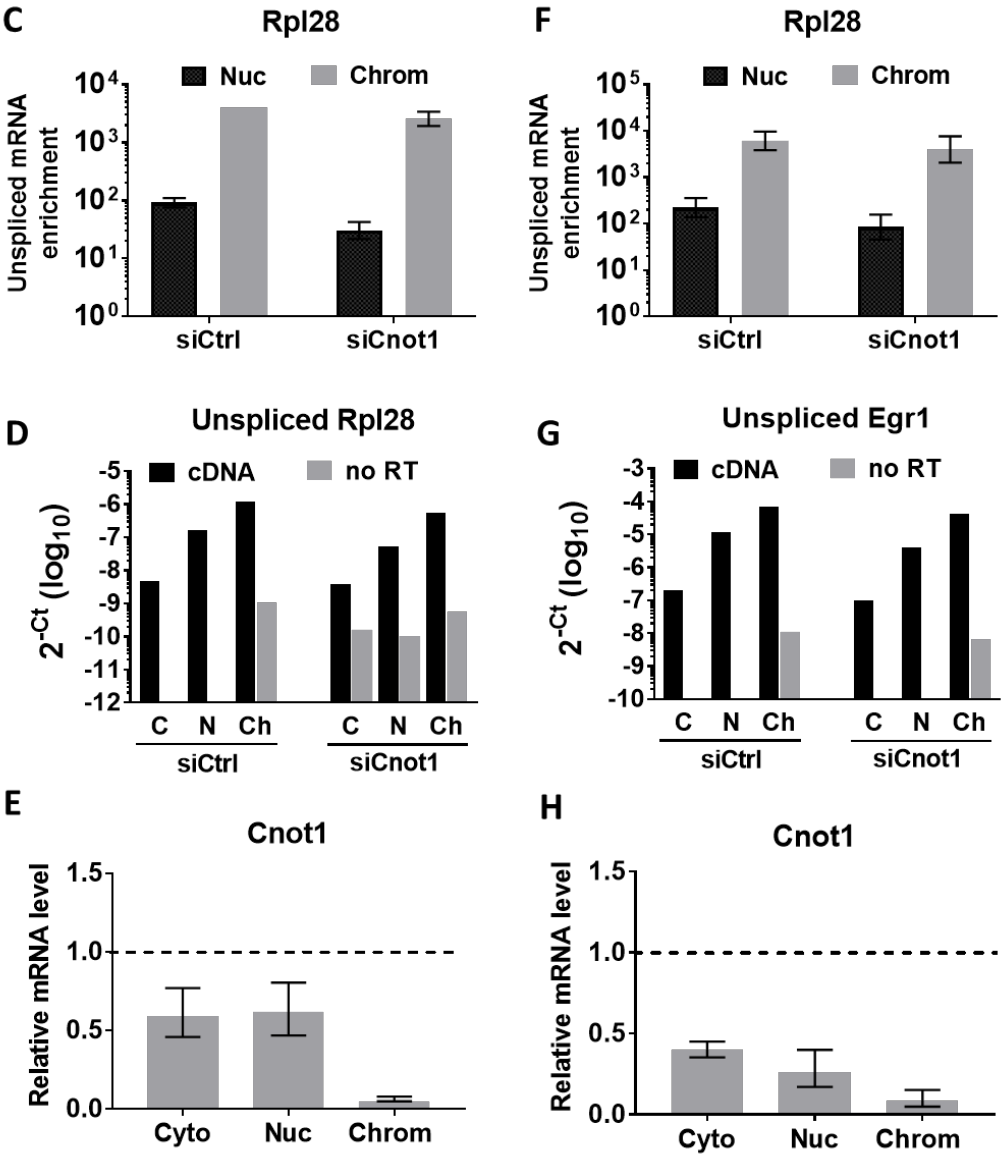


Figure 4.3 Cnot1 regulates poly(A) tail length at different stages. NIH 3T3 cells were treated with control or Cnot1 siRNA and serum stimulated for 50 minutes. Chromatin associates, nuclear and cytoplasmic RNA was isolated and subjected to RL2-PAT, using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. 2 biological replicates (**A, B**) shown. **C, F**) Validation of nuclear/cytoplasmic separation of each biological replicate through comparison of unspliced:spliced mRNA ratios. **D, G**) qPCR data showing low level of DNA contamination in some fractions of each replicate. **E, H**) Validation of knockdown at the mRNA level by qPCR for each replicate, using the ΔΔCt method with Gapdh as the reference gene.

The increased poly(A) length observed in Cnot1 knockdown varied between genes as to whether it was first observed on the chromatin or in the nucleoplasm (figure 4.3 A, B). Interestingly, at the earlier time point of 50 minutes rather than 60, two distinct bands appeared in the chromatin fraction for Egr1 and Egr2, suggesting that there may have been a switch from long to short chromatin associated tails for these mRNAs. Based on the gradual decrease in tail length across fractions in the knockdown cells, it seems likely that the mid-length tails observed previously for serum-induced mRNAs in nuclear fractions (figure 4.2 A-C) were a result of incomplete nuclear or chromatin-associated deadenylation. It is not clear how the residual deadenylase activity resulted in such a narrow range of tail lengths.

When chromatin associated lengths were considered, further divergence was observed within the housekeeping group of mRNAs. Rpl28 alone exhibited slightly longer poly(A) tails in all fractions (figure 4.3 A, B), despite showing no sign of similar lengths under any condition studied in control cells. This suggested either that the 'trimmed' form is usually so abundant that signals from long-tailed transcripts were undetectable, or that a distinct form of regulation exists in which tail synthesis itself is limited by some CNOT1-dependent mechanism. Actb and Sqstm1 on the other hand, had similar length long tails in the chromatin fractions in both conditions, but seemed to experience slower nuclear deadenylation in Cnot1 knockdown (figure 4.3 A, B).

Strangely, two bands were visible in the nuclear and cytoplasmic fractions for Actb in the control cells of both replicates. This distinct double band pattern was not visible in the 60' stimulated nuclear/cytoplasmic fractionations (figure 4.2 A – C) or in earlier serum time course chromatin fractionation experiments, though there was greater smearing in the nuclear and cytoplasmic fractions at the late time point than at 0 or 20 minutes (figure 3.5 A, D). The single band for the knockdown counterpart sat between the control bands rather than showing an enrichment for one or the other, suggesting that it may have corresponded to an incompletely deadenylated form of the shorter control band (figure 4.3 A, B).

4.5 CNOT1 knockdown also affects poly(A) tail length in HEK293 cells

To determine whether the early poly(A) tail-limiting effects of CNOT1 were conserved in other organisms, CNOT1 was knocked down in the HEK293 human

kidney cell line and RNA isolated from nuclear and cytoplasmic fractions. As a serum response has not been well characterised in these cells, the experiment was carried out using cells growing at steady state. Knockdown and separation of fractions were again validated using qRT-PCR data, with RPL10A being used for the latter. Modest knockdown was achieved at the mRNA level, with a much greater reduction in the cytoplasm (figure 4.4 B). Nuclear enrichment of unspliced mRNA was lower than in NIH 3T3 cells (figure 4.4 C, 4.2 E), possibly due to differences in turnover rates of the mRNAs assayed or variation in primer efficiency. It was equally possible that the fractionation method was less efficient for these cells.

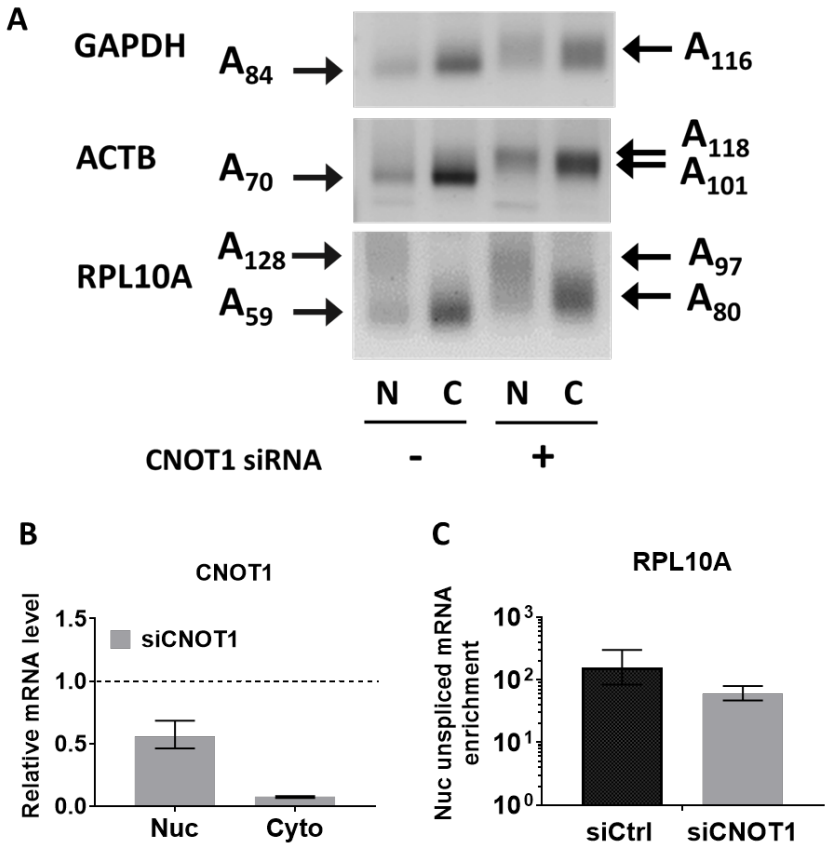


Figure 4.4 Preliminary data suggest nuclear poly(A) tail size is also regulated by CNOT1 in human cells. HEK293 cells were treated with control or CNOT1 siRNA. Nuclear and cytoplasmic RNA was isolated from cells at steady state and subjected to RL2-PAT, using agarose gel stained with SYBR safe, and qRT-PCR **A)** PAT gel. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. **B)** Validation of knockdown at the mRNA level by qPCR, using the ΔΔCt method with GAPDH as the reference gene. **C)** Validation of nuclear/cytoplasmic separation through comparison of unspliced:spliced RPL10A mRNA ratios.

Although only one replicate was performed, these preliminary data showed that CNOT1 knockdown also caused longer nuclear tails in HEK293 cells (figure 4.4 A), suggesting that this was not a phenomenon limited to mouse cells. In addition, the housekeeping mRNAs assayed here possessed medium-length tails similar to those in NIH 3T3 cells, and these appeared similar in the nucleus and cytoplasm.

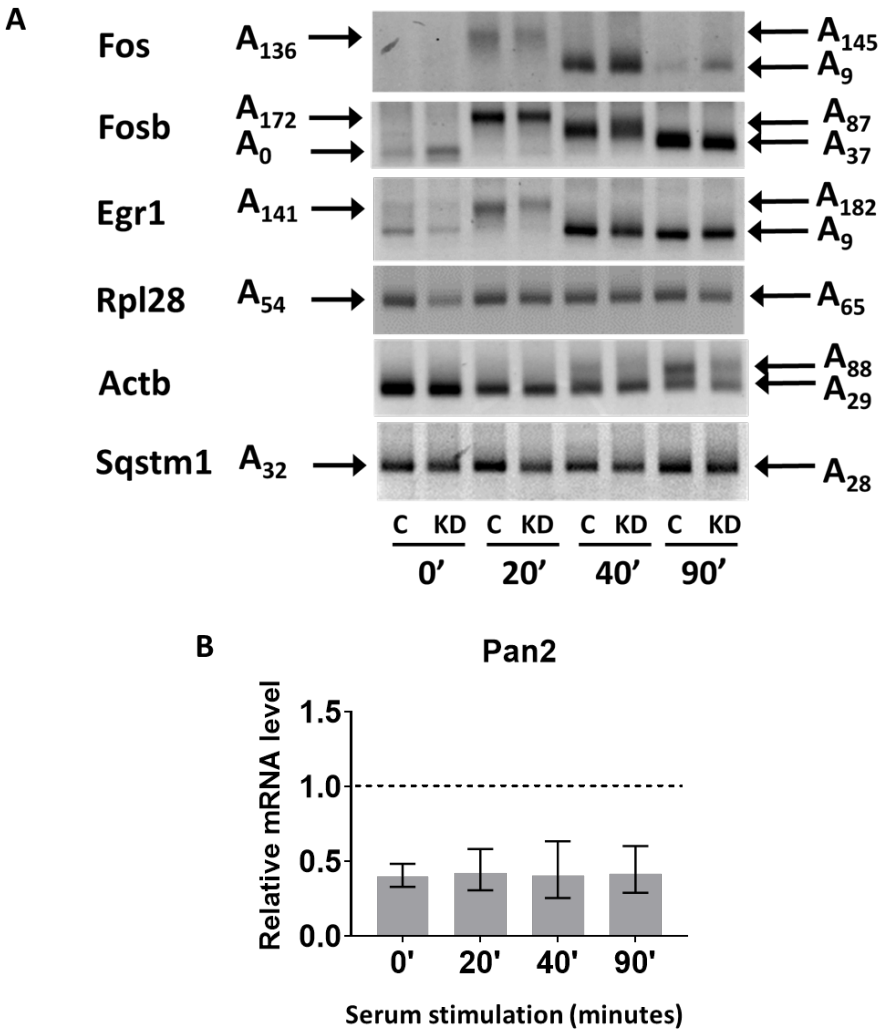


Figure 4.5 Pan2 knockdown causes a minor increase in poly(A) length of induced mRNAs early in the serum response. NIH 3T3 cells treated with control or Pan2 siRNA and stimulated with serum as indicated previously. Total RNA was isolated. **A)** RNA was subjected to RL2-PAT using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. C and KD denote control and Pan2 knockdowns respectively. Results shown for one representative biological replicate. Time indicates minutes of serum stimulation. **B**) Validation of Pan2 knockdown at the mRNA level for the time course shown in A).

4.6 Pan2 knockdown and deadenylation

CCR4-NOT is not the only contributor to deadenylation in mammals; PAN2-PAN3 is also thought to play a role in bulk turnover, primarily acting early in deadenylation on long PABPC-bound tails (26, 529). Furthermore, yeast PAN was previously suggested to mediate message-specific poly(A) trimming in the nucleus (430). To see whether effects on early poly(A) tail length were unique to the CCR4-NOT complex, or a general feature of reduced deadenylation capacity, the catalytic subunit of PAN2-PAN3 was depleted by RNAi. Knockdown was checked at the mRNA level by qPCR and showed a modest depletion (figure 4.5 C). There was little effect on total poly(A) length, except for Egr1 and perhaps Fosb at the 20 minute time point (figure 4.5 A). This was consistent with a preference of PAN2-PAN3 for long poly(A) tails but could also have resulted from incomplete Pan2 knockdown.

To determine whether these small differences extended to nuclear tail length control, Pan2 was knocked down and the poly(A) tail lengths in nuclear and cytoplasmic fractions examined after 50 or 20 minutes' serum stimulation. Validation of knockdown and nuclear/cytoplasmic separation was again validated at the mRNA level using qRT-PCR. Knockdown was much more efficient in the cytoplasm, and separation of fractions was similar in both control and knockdown cells (figure 4.6 D, E).

PAN2 did not seem to be required to limit tail lengths late in the serum response (figure 4.6 A), suggesting that reduced trimming of long tails could be compensated for by other deadenylases. Actb was the exception to this, showing a slight increase in nuclear poly(A) length in the Pan2 knockdown. This increase was also detectable in one of the two 20 minute experiments (figure 4.6 B, C) however, given such small differences it was hard to separate these conclusively from variation during gel running. An increase in poly(A) tail length could also be observed for Egr1 in the Pan2 knockdown cells, but in this case the difference seemed to be between the cytoplasmic fractions, in line with predominantly cytoplasmic localisation of PAN2-PAN3 (24, 26). Neither Rpl28 nor Sqstm1 seemed to be affected in either the total RNA time course or fractionation experiments (figures 4.5 A, 4.6 A-C).

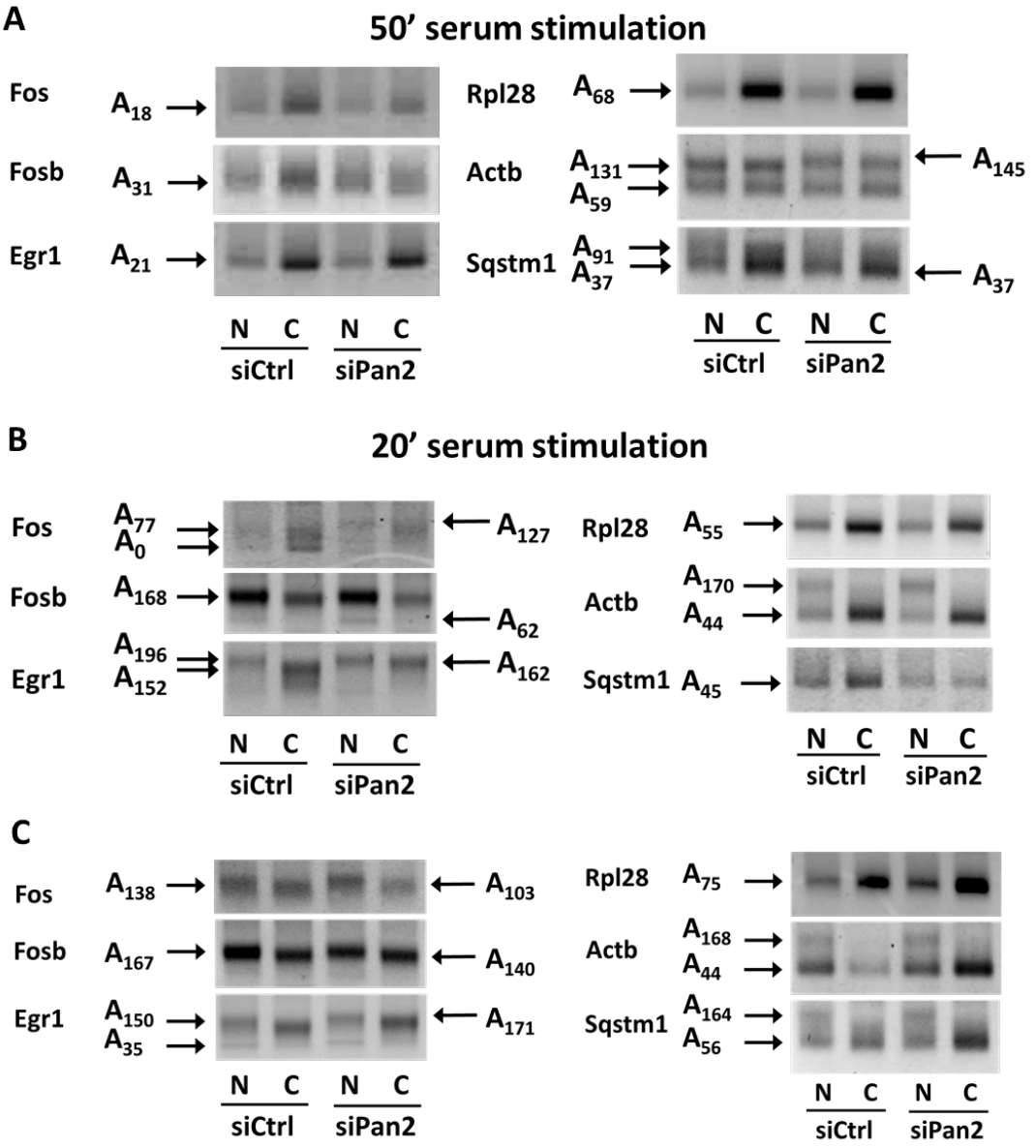


Figure 4.6 Nuclear poly(A) tail size is not markedly affected by Pan2 knockdown. (continued overleaf). NIH 3T3 cells were treated with control or Pan2 siRNA and serum stimulated for 50 **(A)** or 20 **(B,C)** minutes. Nuclear and cytoplasmic RNA was isolated and subjected to RL2-PAT using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. 3 biological replicates (A-C) shown. **D)** Validation of knockdown at the mRNA level by qPCR for each replicate, using the ΔΔCt method with Gapdh as the reference gene. **E)** Validation of nuclear/cytoplasmic separation of each biological replicate through comparison of unspliced:spliced mRNA ratios.

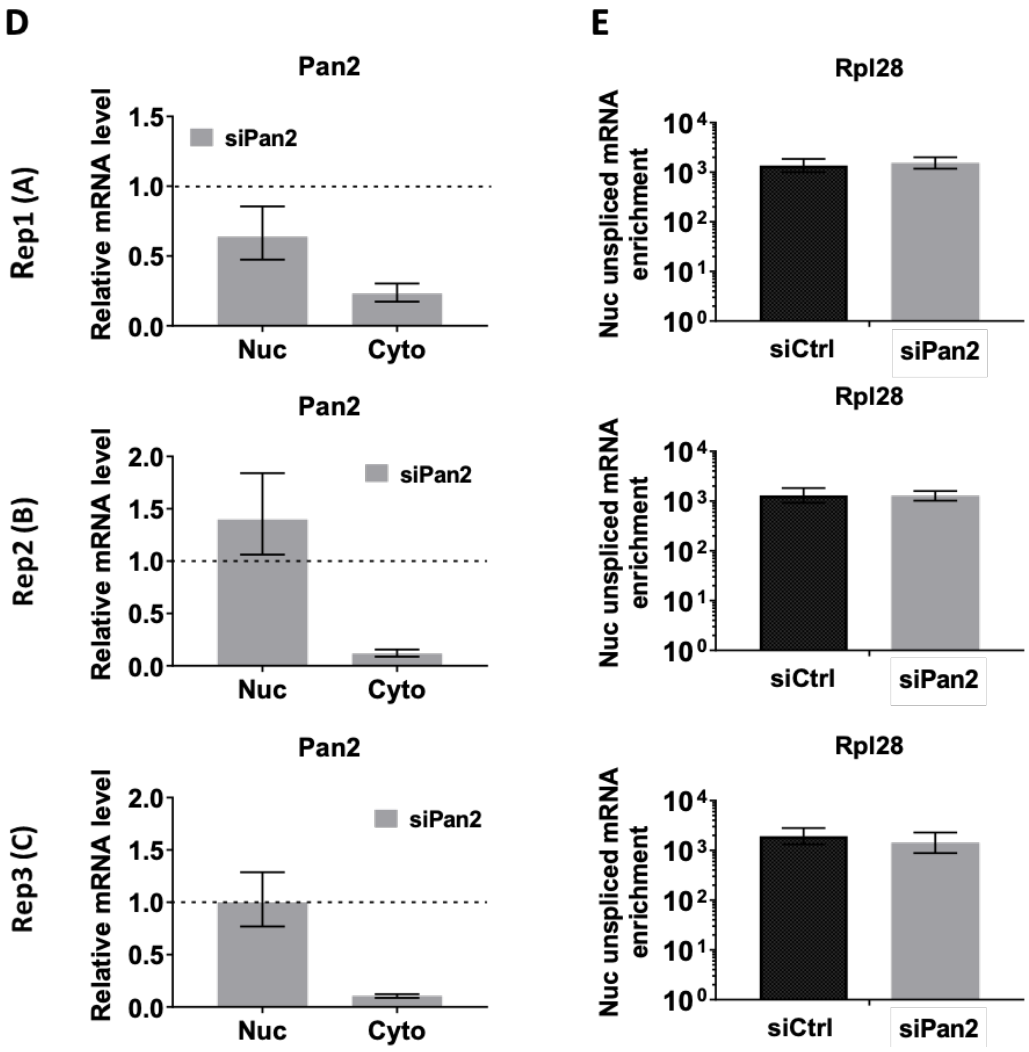


Figure 4.6 continued.

4.7 Discussion

Cnot1 depletion caused an increase in nuclear and in some cases, chromatin-associated tail length late in the serum response, implicating involvement of the CCR4-NOT complex in defining initial tail length (figures 4.2 A-C, 4.3 A,B). Preliminary data suggest this may be consistent in human cells (figure 4.4). Though nuclear poly(A) tail trimming by CCR4-NOT may seem intuitive, the kinetics implicated here would be far different to those previously established; deadenylation would need to occur substantially faster than previously recorded in order to remove ~150 nucleotides before dissociation of the nascent mRNA from the chromatin. Data presented earlier in this work showed that when transcription was stopped, removal of around 100-150 adenosines took 40 minutes (figure 3.6) – a much longer period than would be possible between transcription and export. While other reports of

nuclear and chromatin-associated deadenylation exist, these are limited to stem cells and germline cells respectively, and seem to be linked with targeted degradation/destabilization (33, 259). Nevertheless, nuclear deadenylation by the CCR4-NOT complex seems the simplest model to explain the observations presented here.

The control of early poly(A) tail length does not seem to be a property of deadenylases in general since Pan2 knockdown had only minor effects (figures 4.5, 4.6), though this may have been due to insufficient knockdown. This is consistent with recent TAIL-Seq data showing that PAN2/3 depletion had no effect on mean poly(A) length (50). As CCR4-NOT has functions outside deadenylation, it is not certain that the increases in initial poly(A) length observed following Cnot1 knockdown were due solely to inhibited deadenylation. Comparison of Pan2 knockdown with combined depletion of the CCR4-NOT nucleases (rather than CNOT1) may therefore be more appropriate for considering the relative roles of the two complexes.

Although all mRNAs tested had longer poly(A) tails an hour after serum stimulation in Cnot1 knockdown, the size of the increase and main location of control varied between genes. Serum induced mRNAs showed a greater increase in tail length than housekeeping genes, and this increase was elicited mainly in the chromatin fraction (except for Fosb). The longer nuclear tails observed were not as long as those produced after 20 minutes' serum stimulation in control cells, but this could have been due to remaining deadenylase activity (from incomplete knockdown, or perhaps PAN2-PAN3) trimming the tail in the nucleus (figure 4.3). Similar processes seemed to be at play on Sqstm1 and Actb mRNAs, however, these appeared to normally undergo nuclear rather than chromatin-associated deadenylation since differences in the knockdown cells were predominantly nuclear. Rpl28 alone did not ever seem to receive a long (~200 nt) tail, even in the chromatin fraction of Cnot1 knockdown cells, suggesting that synthesis of the tail itself may be regulated.

It is interesting that for all mRNAs, the increase in tail length observed in response to Cnot1 knockdown appeared as a discrete band rather than a smear. In a model where CCR4-NOT acts directly on nuclear poly(A) tails, insufficient deadenylase would be expected to result either in some untouched and some fully deadenylated

tails (in the case of processive activity), or in a range of partially digested poly(A) tails (in the case of distributive activity). This model is summarised in figure 4.7. Neither appeared to be the case here; tight bands were visible at intermediate poly(A) lengths, suggesting that trimming was still being controlled, but to a different length. This, along with the different 'final' poly(A) tail lengths observed for different genes in control cells, suggests that CCR4-NOT complex activity is both targeted and titratable. Gene-specificity could be explained by the numerous known examples of CCR4-NOT targeting (30, 45, 255), while the titratable effect may arise either from modulation of CCR4-NOT nuclease activity, or through binding of the tail by multiple copies of a protein such as PABPN1 or PABPC which could affect deadenylation.

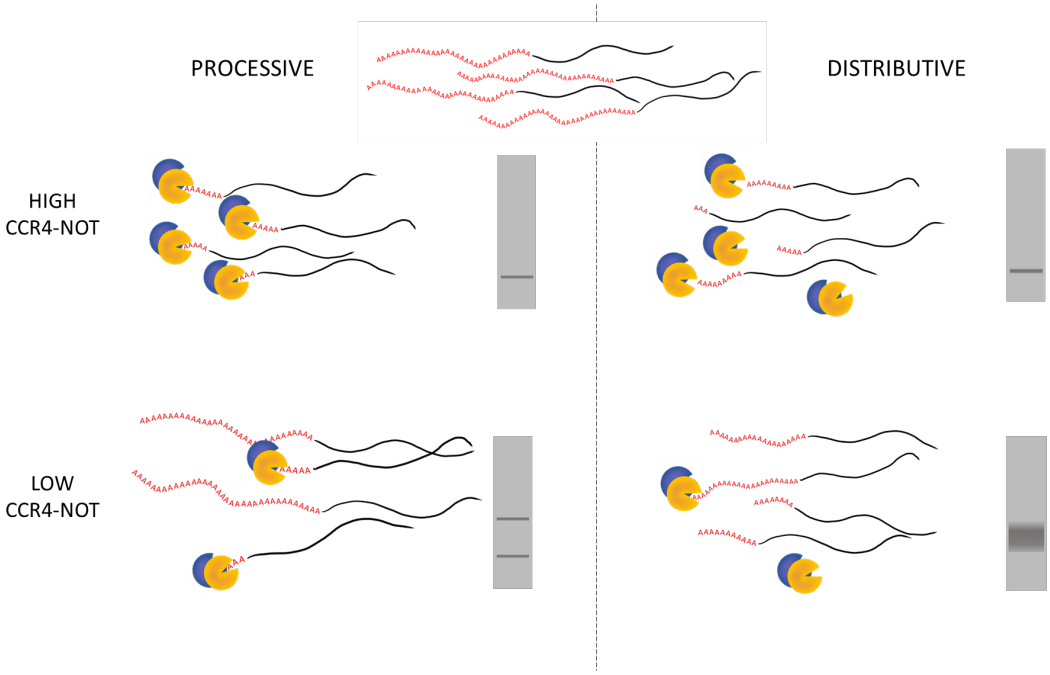


Figure 4.7 Schematic of expected PAT distributions with different CCR4-NOT activity and levels. In conditions of low CCR4-NOT, processive activity would lead to two bands: fully deadenylated mRNA, and mRNAs which are still fully polyadenylated. Low levels of CCR4-NOT with distributive activity would instead lead to stochastic poly(A) shortening and a smeared gel band.

5 CNOT1 is required for normal induction of the serum response

5.1 Introduction

Following stimulation, transcription of responsive genes is rapidly upregulated, then tails off (587, 611). Our earlier data show that poly(A) tail length of induced mRNAs seems to track with these changes to transcription, with long nuclear tails early in the serum response when transcription is rapid, and short nuclear tails later on when transcription is reduced (figure 3.5 A, D), (611).

Several groups have found evidence of crosstalk between mRNA degradation and synthesis, giving rise to models of transcript buffering and transcriptional imprinting (58, 148, 470, 471, 578–580, 585). In several yeast studies, depletion of various degradation machinery led to a reduction in synthesis in order to maintain mRNA homeostasis (579, 580, 582). The mammalian system is less well studied, but in addition to limited reports of homeostasis, there is evidence that crosstalk can instead enhance a particular state (58, 470, 471). For example, if enhanced cytoplasmic degradation is triggered by expression of the gamma-herpesvirus SOX nuclease, mammalian cells experience transcriptional repression rather than upregulated synthesis (470, 471).

The CCR4-NOT deadenylase complex was first identified as a transcription factor and has roles in Pol II transcription elongation, making it well placed to link synthesis with degradation (35, 38, 39, 42, 147, 148, 550). We were therefore curious as to whether the longer nuclear and chromatin-associated poly(A) tails observed previously in Cnot1 knockdown (figures 4.2 A-C, 4.3 A,B) were accompanied by changes to mRNA abundance.

5.2 Cnot1 knockdown dampens the transcriptional response to serum

To investigate whether the longer tails caused by Cnot1 knockdown were accompanied by changes to mRNA abundance, q-RT PCR of serum stimulation time courses was performed. Two siRNA transfections were carried out 24 hours

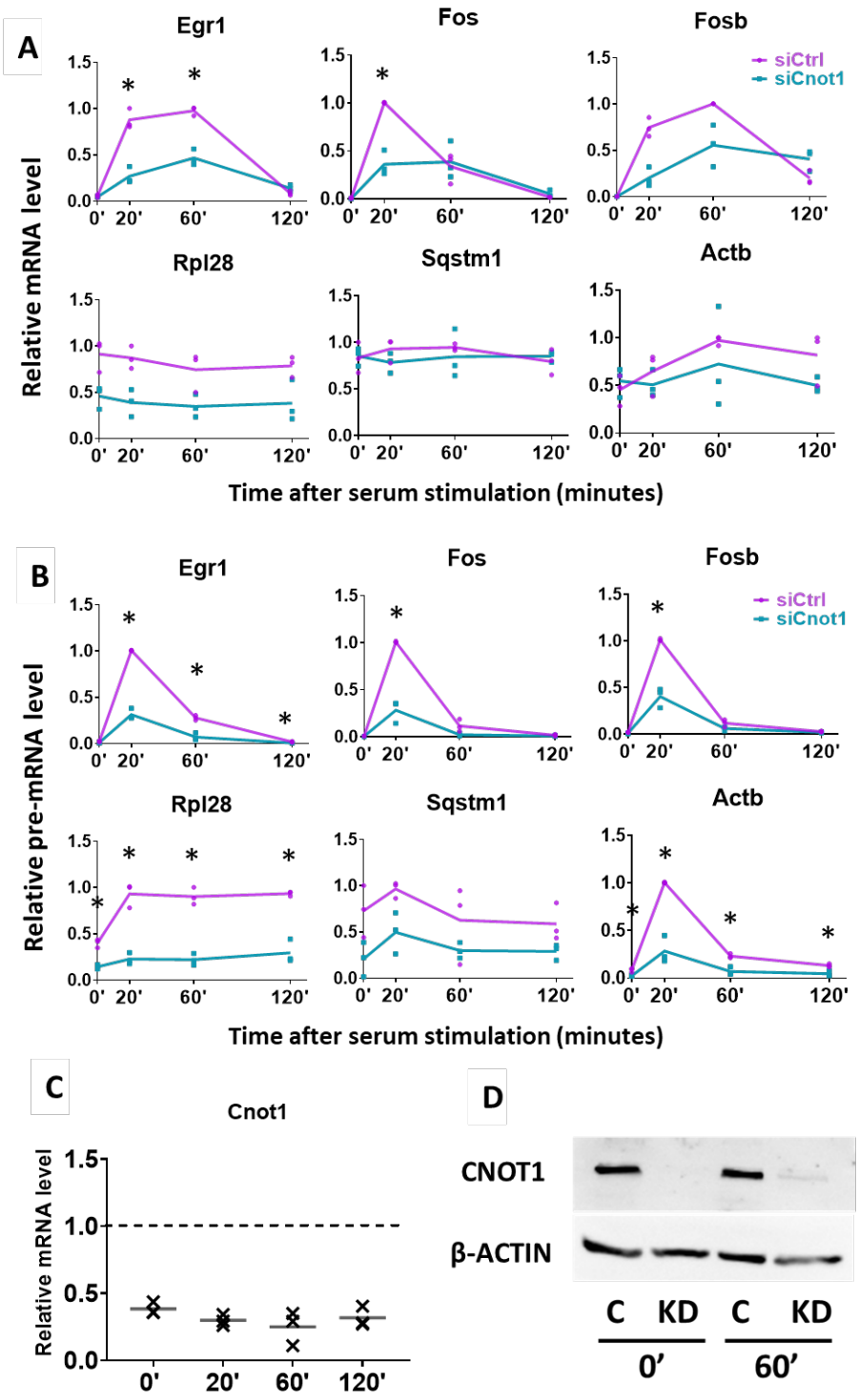


Figure 5.1 Cnot1 knockdown reduced mRNA abundance at the mature and unspliced level early in the serum response. A, B) qRT-PCR data showing mRNA abundance following serum stimulation at the mature (A) and unspliced (B) levels. Expression relative to Gapdh and normalised to the maximum value within the control set. Data collected from three biological replicates. T tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.01. **C)** Validation of knockdown at the mRNA level for three biological replicates. **D)** Validation of siRNA knockdown procedure at the protein level by western blot.

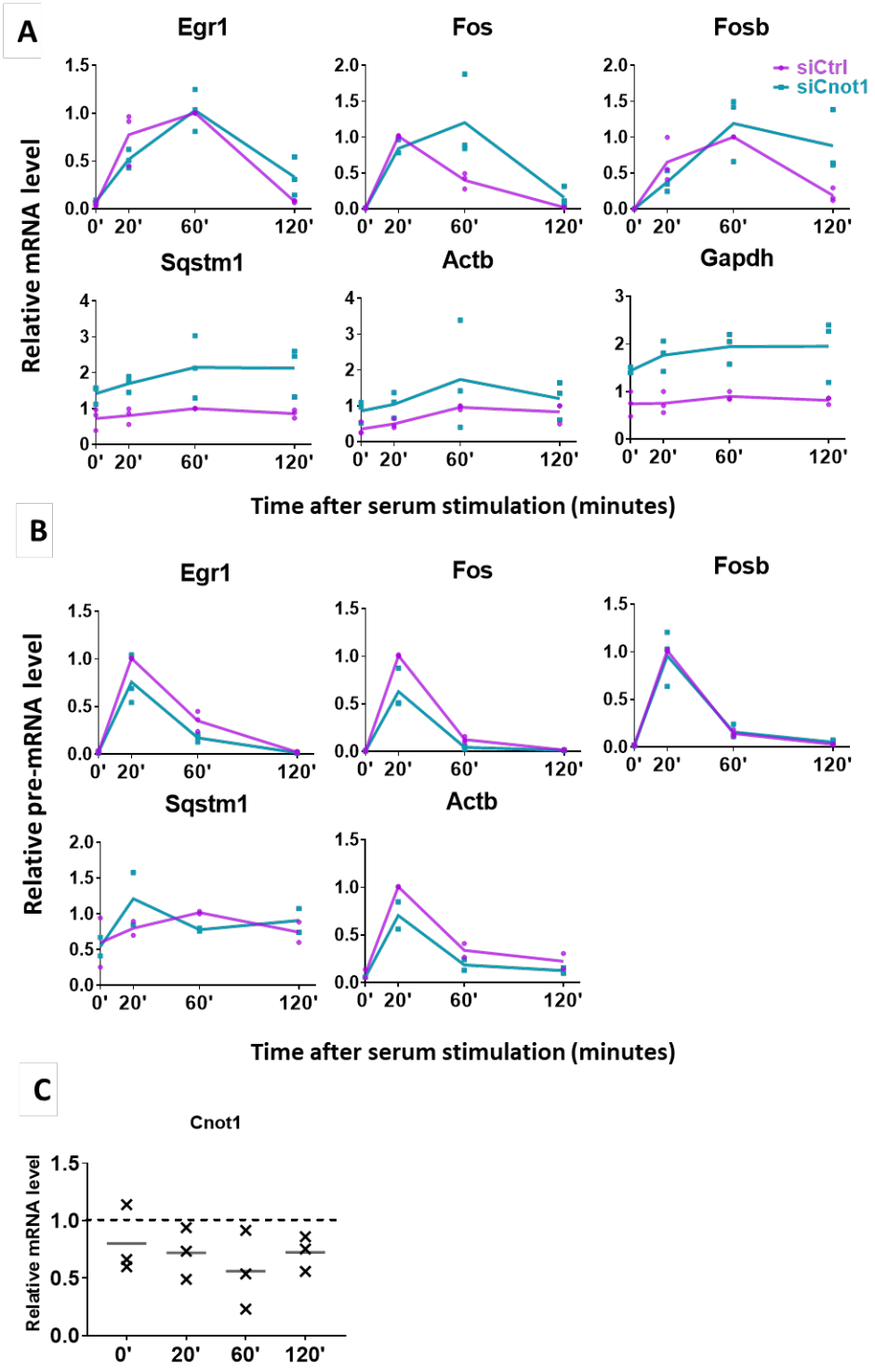


Figure 5.2 Normalisation to Rpl28 minimises differences between control and knockdown. Cnot1 knockdown has a smaller apparent effect when mRNA levels are normalised to Rpl28 instead of Gapdh. Presented are qPCR data from 3 biological replicates normalised using the ΔΔCt method. Rpl28 was used as a reference gene and in A-C, all data points are normalised to the maximum value within the control set. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.01. **A)** Mature mRNAs. **B)** Unspliced mRNAs. **C)** Levels of Cnot1 mRNA normalised to the control for each time point.

apart, resulting in cells which had been treated at 72 and 48 hours prior to the serum time course. Knockdown of Cnot1 was validated for all replicates at the mRNA level by qPCR (figure 5.1 C) and the method validated by western blot (figure 5.1 D).

Surprisingly, instead of sustained normal levels of serum response mRNAs, a reduction in mature mRNA levels of was observed in Cnot1 knockdown cells at all time points (figure 5.1 A). Given the role of deadenylation as a gatekeeper to decay and the early identification of CCR4-NOT as a regulator of transcription, it seemed more likely that this reduction was due to changes in mRNA synthesis rather than increased decay. Primers amplifying unspliced transcripts showed that pre-mRNA levels of these mRNAs were also reduced in the Cnot1 knockdown cells (figure 5.1 B). This suggested that reduction in CNOT1 both dampened the transcriptional response and theoretically extended the lifetime of the mRNAs produced through increased poly(A) length (figures 4.1, 4.2 A-C, 4.3 A,B). Hints of an extended response can be seen for the serum-induced mRNAs at the 2 hour time point.

In addition to changes in induced mRNA abundance, the relative levels of several housekeeping pre-mRNAs were reduced in Cnot1 knockdown. This reduction was also evident for mature Rpl28 mRNA, which caused a temporary complication as Rpl28 was routinely used as a control to validate the reference gene. In fact, based on raw Ct value, Rpl28 expression was more consistent than Gapdh across samples. Switching to Rpl28 as a reference gene showed much reduced effects of knockdown on pre-mRNA levels and showed a clearer persistence of transcripts at the end of the response (figures 5.2 A, C). Housekeeping mRNA transcripts were also elevated at the mature level. While this in many ways represents a more expected outcome, normalisation to Rpl28 showed very poor knockdown at the mRNA level, which is inconsistent with the western blot validating the knockdown method (figures 5.2 C, 5.1 D). Expression of most other 'housekeeping' mRNAs tested (derived from the literature and past nCounter experiments) did not vary significantly between control and knockdown conditions when normalised to Gapdh, suggesting it remains a suitable reference gene (figure 5.3).

To get a better idea whether the decrease in unspliced mRNA was a result of reduced transcription or increased degradation, qPCR was performed on the

chromatin fractionation experiments introduced previously (section 4.4). Chromatin fractionations were performed after 50 minutes of serum stimulation as this was where the most significant difference in poly(A) tail length was observed. No-reverse transcription controls were performed to ensure that genomic DNA contamination did not contribute significantly to the chromatin fraction reactions, and that any contamination was consistent between the control and knockdown samples (figure 4.3 D, G). Results of two biological replicates indicated that downregulation of induced mRNA levels occurred post-transcriptionally, with reduced abundance in nucleoplasmic and cytoplasmic fractions but not in the chromatin-associated fraction (figure 5.4). A reduction in both spliced and mature Rpl28 mRNA occurred in all fractions. In some cases, relative mRNA level was higher in the cytoplasmic fraction than the nucleoplasmic one; this may have been measurement error but could also have resulted from cytoplasmic stabilisation caused by longer poly(A) tails.

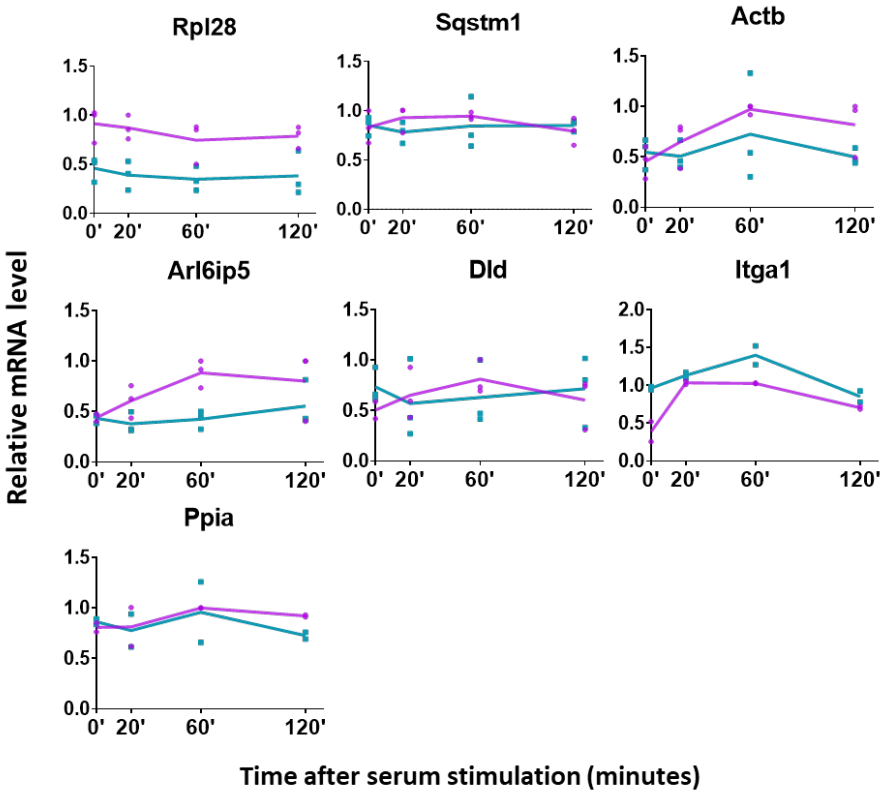


Figure 5.3 Other housekeeping mRNAs suggest Gapdh is a suitable reference gene. qPCR data from 3 biological replicates (or 2 for Itga1 and Ppia) normalised using the ΔΔCt method. Gapdh was used as a reference gene and all data points are normalised to the maximum value within the control set. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.01.

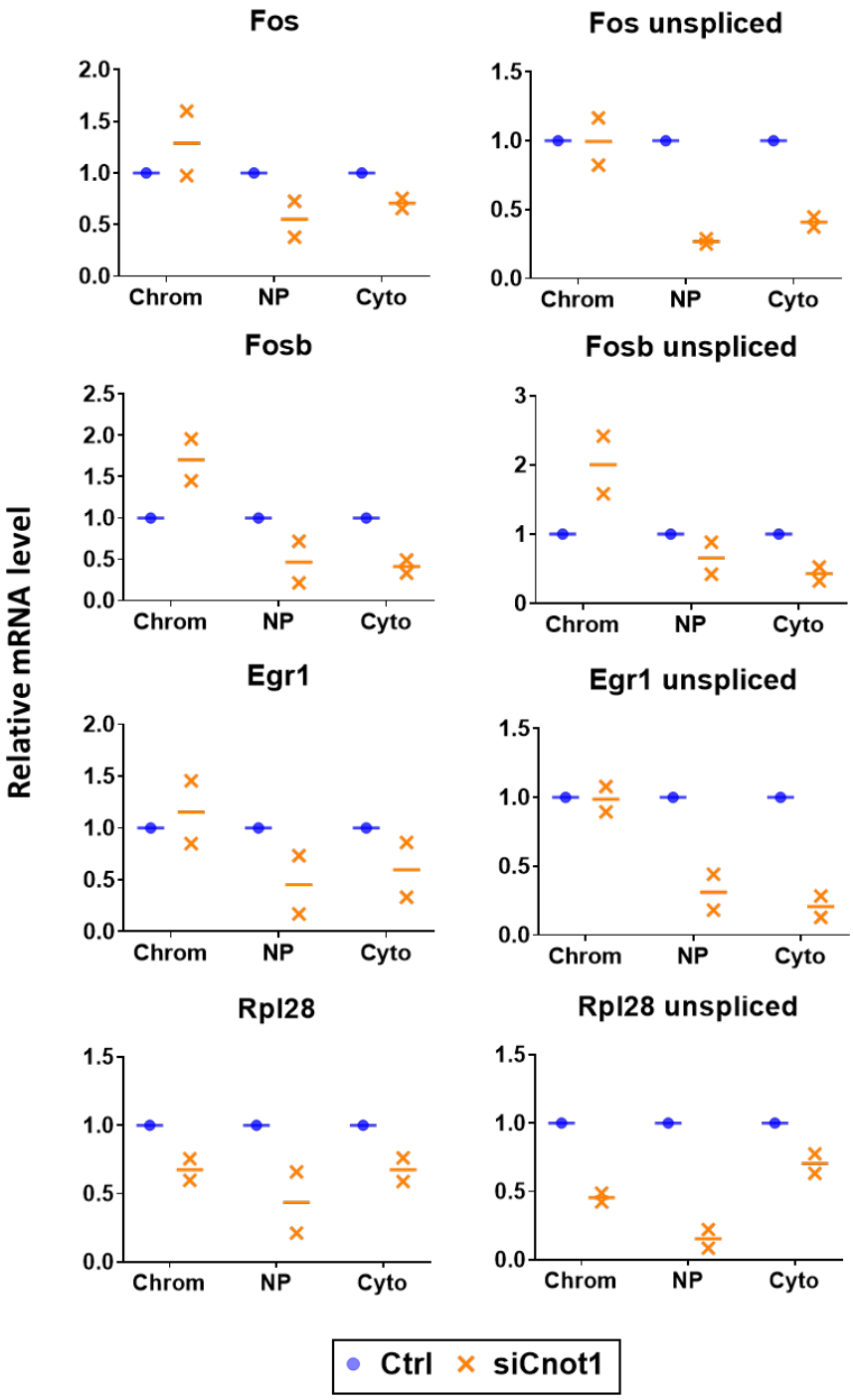


Figure 5.4 Reduction in mRNA abundance in Cnot1 knockdown occurs between the chromatin-associated and nucleoplasmic fractions for induced mRNAs. qPCR data for 2 biological replicates using the ΔΔCt method with Gapdh as a reference gene, and normalised to the control in each fraction.

Minimal change in chromatin-associated mRNA abundance could indicate that transcription of induced mRNAs was unaffected by Cnot1 knockdown, but that there was an increased rate of nuclear decay. Alternatively, release of mRNA from the

chromatin may have been slowed down, perhaps accompanied by slower transcription. Mature and unspliced profiles of each mRNA were similar, making altered splicing an unlikely explanation for the observed pattern (figure 5.4). More genes and replicates are required before this can be considered further.

5.3 CNOT1 knockdown in HeLa cells affects expression of genes involved in proliferation

To see whether Cnot1 knockdown had similar effects on mRNA abundance in another species, and to identify additional affected transcripts, a publicly available but unpublished CNOT1 knockdown RNA-Seq dataset was analysed (621). This dataset comprised two replicates of control and CNOT1 knockdowns in HeLa cells as well as an additional single replicate for each condition treated with the RNA Polymerase II (Pol II) transcription elongation inhibitor, DRB for 4 hours. Data for samples treated with inducers of DNA damage and with TAB182 siRNA were also available but were not deemed relevant to this study. Unlike the NIH 3T3 experiments presented previously, cells were captured at steady state rather than following stimulation. Sequencing samples were prepared from ribodepleted total RNA and libraries generated using the NEBNext Ultra Directional RNA library preparation kit, which involves RNA fragmentation followed by random primed first strand synthesis. dUTP is incorporated during second strand synthesis and following adapter ligation these are excised to cause effective degradation of the second strand, resulting in a reverse stranded library. Libraries are then amplified by PCR.

Quality of each fastq file was assessed using FastQC (599). All samples showed good per base sequencing quality throughout both reads, indicative of very reliable base calling (data not shown). All samples were also flagged as containing up to 10 % duplicated reads which were duplicated at least 10 times, and a lower percentage which were duplicated to a greater level (data not shown). This amount of duplication at first seemed cause for concern, but the consensus among the bioinformatic community is that reads registering as duplicates may actually represent increased abundance; since more abundant transcripts generate more reads there is greater likelihood of exact matches, particularly at high sequencing depths. This of course cannot rule out the possibility of PCR-generated duplicates, but in removing duplicates one also risks loss of information.

FastQC also returns a list of overrepresented sequences defined are those comprising > 0.1 % of total reads. Unexpectedly, all samples contained overrepresented sequences. A greater number of overrepresented sequences were detected in read 1 than read 2 in all samples, with read 2 of some samples not containing any significant overrepresentation. Furthermore, a greater number of sequences were identified in the DRB-treated samples, consistent with reduced Pol II transcription and therefore a lower overall mRNA abundance which would otherwise create diversity and dilute Pol I/Pol III transcripts or residual DNA. These overrepresented sequences seemed to originate predominantly from 7SL and 7SK RNAs.

Reads were mapped against the Gencode GRCh38 genome assembly using the splice-aware STAR (Spliced Transcripts Alignment to a Reference) alignment tool (600). A Gencode primary annotation file was provided during generation of genome indices in order to assemble a database of splice junctions to which reads could be mapped.

To gain a crude indication of whether there was a general change in mRNA abundance, metaplots were created using the deepTools package for the gene body and regions 10,000 bp upstream and downstream of transcription start sites (TSS) and transcription end sites (TES) respectively (603). Reads were normalised for sequencing depth and gene length. Unless there were changes to transcription termination or promoter upstream transcription (which could result in elevated levels of RNA outside the gene body), global changes in mRNA level should have been visible as an altered ratio of gene body to flanking reads.

For total RNA, there was some small variation in levels of RNA spanning the gene body between samples, however this was not associated with control or knockdown samples in particular (figure 5.5 A, left panel). A small peak in reads was present immediately upstream of the TSS in the majority of samples, which may have been due to upstream antisense transcription (622, 623). This peak was much more prominent in the DRB treated samples, perhaps through less total mRNA causing a greater effective sequencing depth (figure 5.5, left panel). Higher gene body reads in the DRB-treated cells would suggest that overall mRNA abundance was not lower than in the other samples; however, since the computeMatrix function (deepTools)

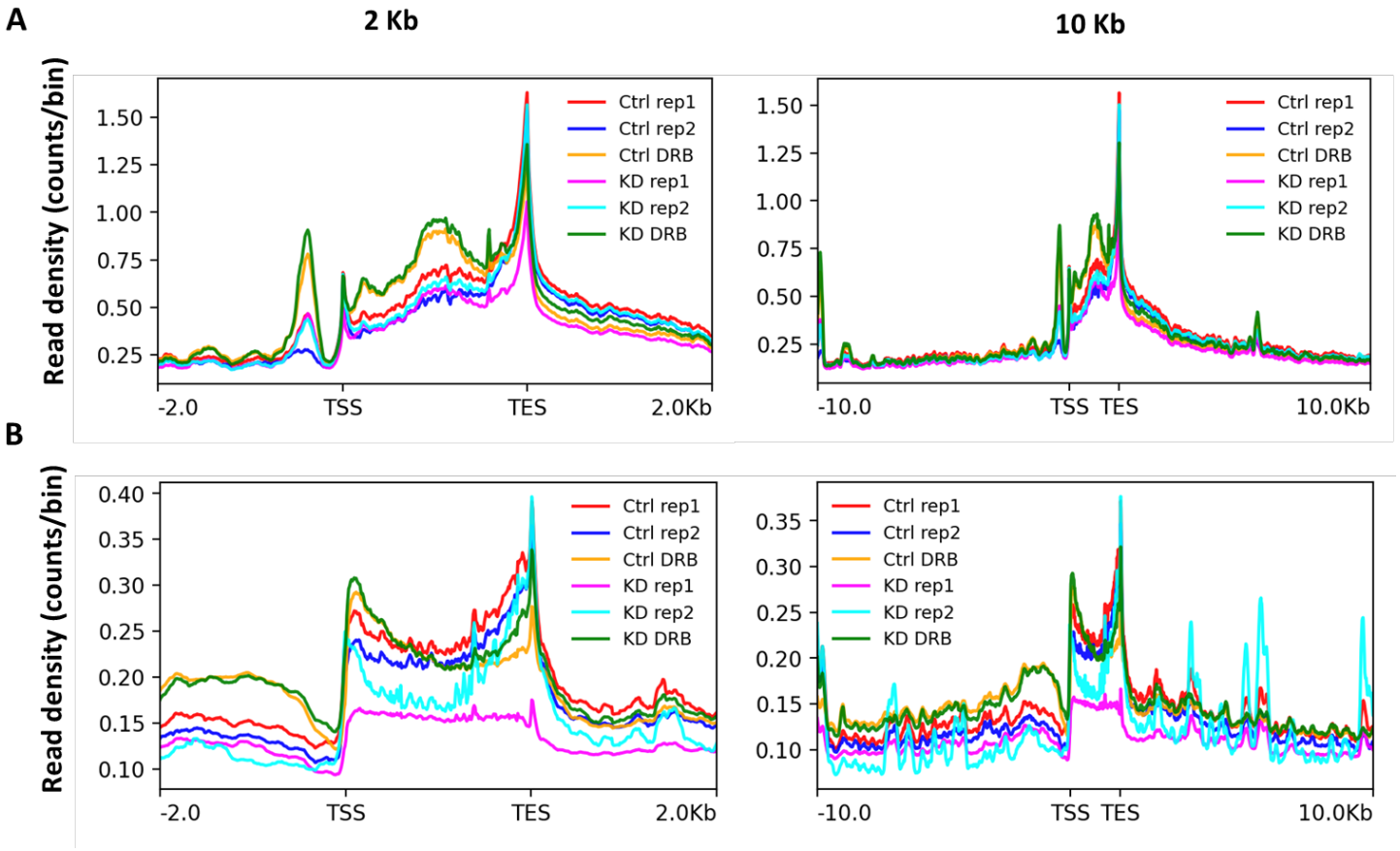


Figure 5.5 Bulk mRNA abundance is not affected by Cnot1 knockdown. A) Read distribution for all mapped reads (spliced and unspliced) using BPM normalisation and either 2 or 10 Kb flanking regions. **B)** Read distribution for exon-depleted reads, using BPM normalisation and either 2 or 10 Kb flanking regions. TSS = transcription start site, TES = transcription end site.

counts normalised read counts per *x* nt section of each gene (where *x* is the user-inputted bin size), it is possible that some pool of stable mRNAs received a higher normalised count due to less dilution by new transcription.

Both control and knockdown DRB treated samples exhibited a milder 3' bias than their untreated counterparts. The reason for this reduced bias was unclear since these reads must have originated from mRNAs which existed prior to DRB treatment – mRNAs which should not have differed substantially from those which continued to be synthesised, unless a large number were undergoing 3'-5' degradation. The appearance of peaks at the 3' and 5' ends was in itself surprising given the library synthesis method and was the opposite pattern to that shown on the NEB website (624).

To further understand differences observed in the DRB-treated samples, and to see whether CNOT1 knockdown affected transcription in general, information about unspliced mRNA level was required. As it was not possible to determine the exact molecule a read originated from, introns were used as a proxy for unspliced mRNA. This was a valid approximation since introns are thought to be rapidly degraded following excision (625). Though it was possible to filter for reads which only mapped to introns, this would also remove reads mapping to intergenic regions which served as a useful control. Instead, reads which overlapped more than 95 % with an exon were removed using the BEDTools intersect function (602). Alignments were spot checked in IGV viewer (612) to confirm removal of exons (figure 5.6). Metaplots of the 'unspliced' mRNA showed reduced read density in all samples (figure 5.5 B). At this level, gene body reads for DRB-treated cells fell back in line with the other samples, and were accompanied by higher signal in the flanking regions, suggesting the DRB treatment was at least partially effective. In addition, both CNOT1 knockdown replicates exhibited lower unspliced signal in the gene body than their control counterparts, consistent with a role for CCR4-NOT in stimulating transcription elongation (147, 148).

Count data was obtained from each alignment using featureCounts and counting by either exon (for total fragments) or gene (for exon-depleted fragments) (605). Differential expression analysis was then performed using DESeq2, which tests for differential expression based on a negative binomial distribution model (604).

Consistent with metaplot profiles, heatmaps showed higher similarity in gene expression between the control replicates than between the knockdown replicates, though this similarity was still lower than expected (figure 5.7 A, left panel). In fact, the second CNOT1 knockdown sample was as similar to the first control replicate as it was to the other knockdown. The lack of similarity between knockdown replicates

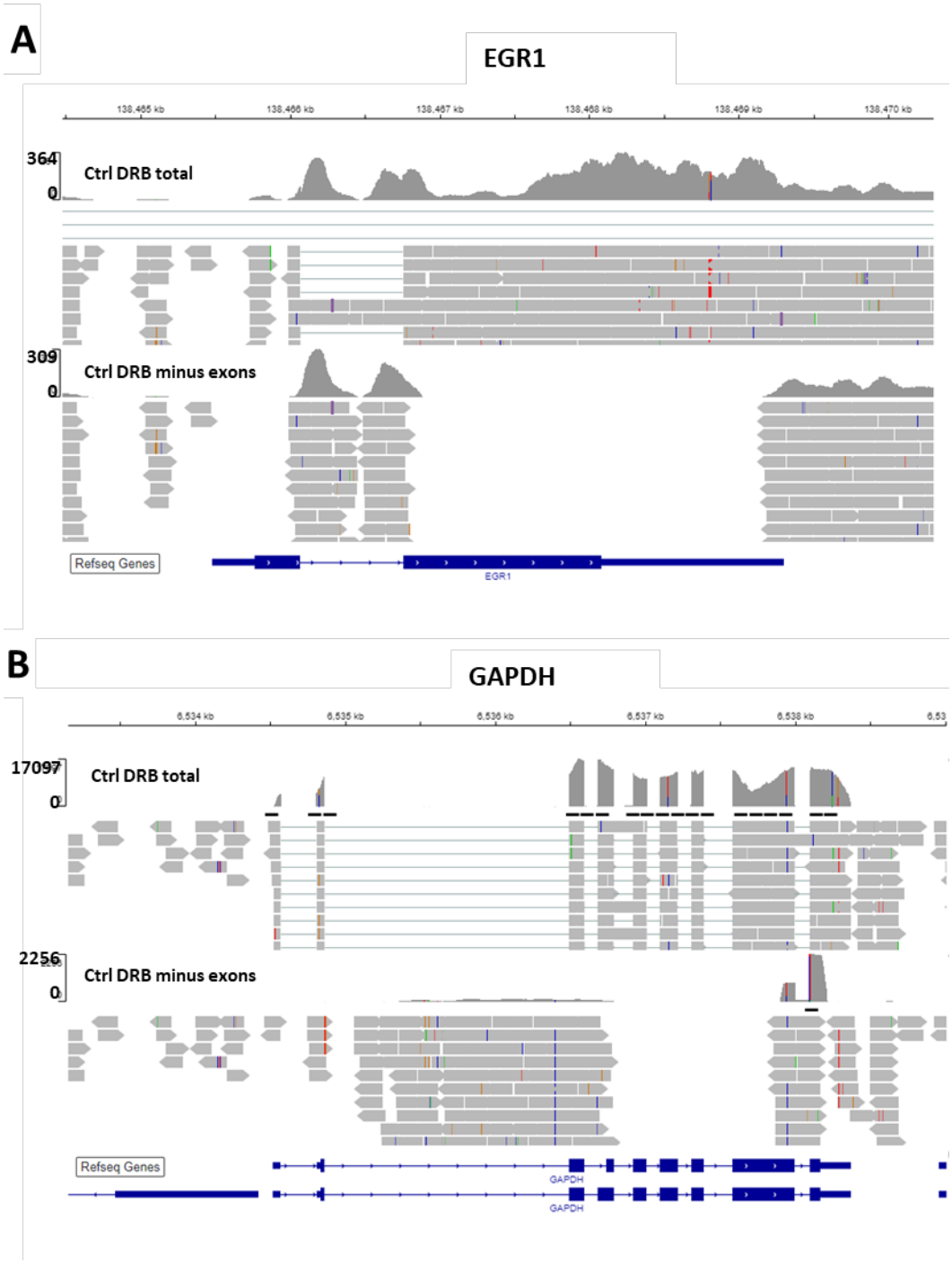


Figure 5.6 Verification of exon removal in IGV viewer. BAM alignment files viewed with the GRCh38 reference genome before and after exon removal for **A)** EGR1 and **B)** GAPDH.

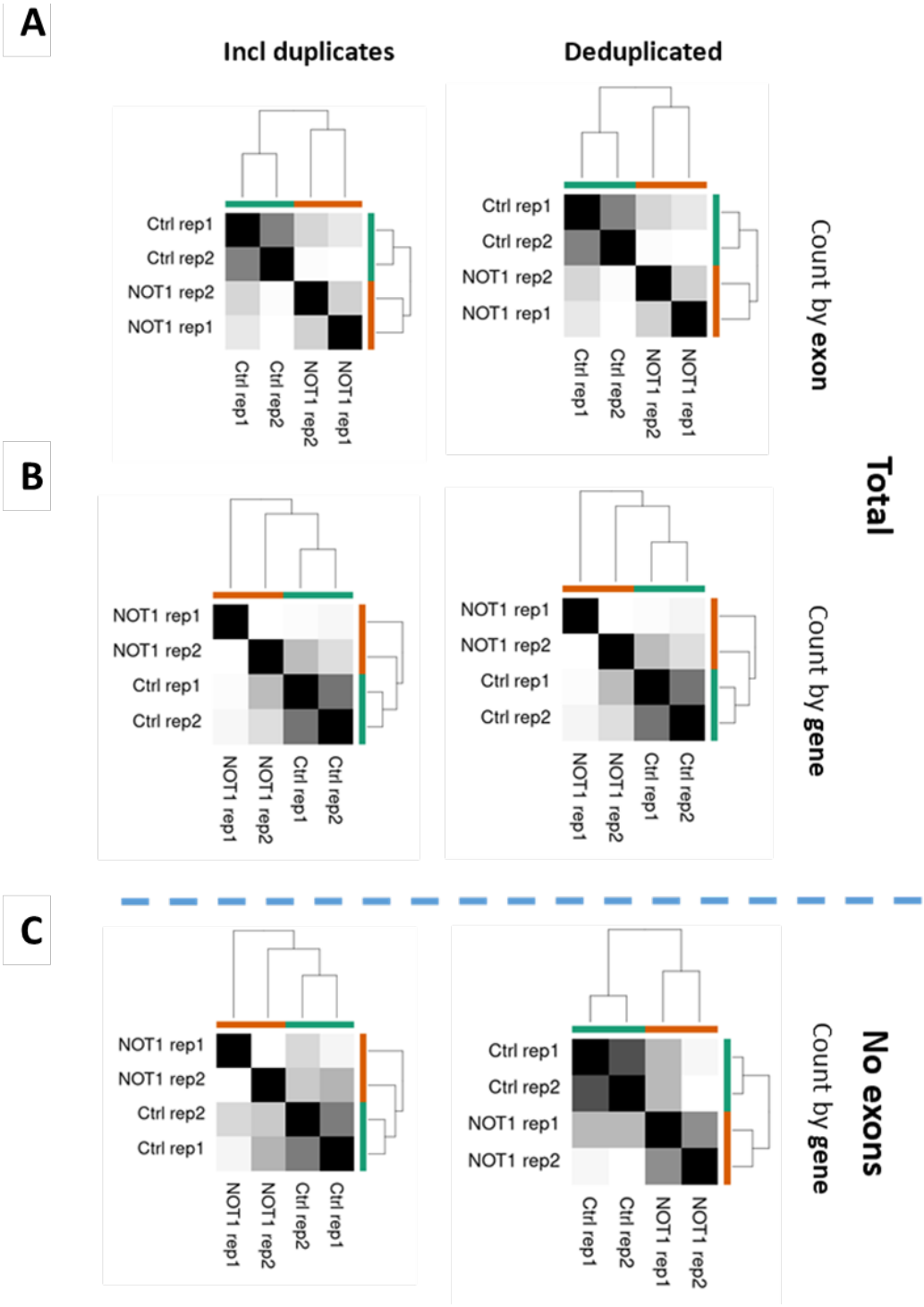


Figure 5.7 Alterations in counting options affects sample similarity. Total **(A, B)** or exon-depleted **(C)** reads from bam alignment files (in which duplicates had been flagged) were counted using the featureCounts program from the Subread package. FeatureCounts can filter out reads flagged as duplicates (right hand side) and can also count by reads per gene **(B, C)** or per exon **(A)**. Black = most similarity, white = least similarity.

was even greater in exon-depleted reads, which was consistent with the different profiles observed in the metaplots (figures 5.7 C, 5.5 B, left panel). To see if sample similarity between replicates could be improved, duplicates were filtered out. In addition, mode of counting was changed from 'exon' to 'gene' for total reads such that unspliced and total reads were treated consistently. Deduplication made a substantial improvement to sample similarity in the exon-depleted alignments (figure 5.7 C), but did not have an appreciable effect at the level of total reads (figure 5.7 A, B). Interestingly, similarity of the knockdown samples at the total reads level was worsened by counting by gene, while control samples were unaffected (figure 5.7 A vs B). This could perhaps have been suggestive of differences in unspliced level, since reads in introns would also have contributed to gene level counts.

For downstream analysis, duplicates were filtered out for both total and exon-depleted fragments. Since sample similarity was greater when counting by exons but counting by gene seemed more appropriate, both methods were taken forwards for total reads since differences could be informative.

When total reads were counted per gene, few mRNAs change expression significantly in CNOT1 knockdown, with more changes occurring in a positive direction (figure 5.8 A, B). This was consistent with increased stability due to less deadenylation. Upregulated genes were enriched for a variety of gene ontology (GO) terms including growth factor signalling, cell adhesion and response to various stimuli (figure 5.8 C).

Counting by exons resulted in a higher number of significant changes to expression in CNOT1 knockdown, again with more upregulation than downregulation (figure 5.9 A, B). Downregulation was more common in highly expressed mRNAs, whereas upregulation was relatively consistent across all mean expression levels (figure 5.9 A). In general, upregulation also seemed to be of a greater magnitude (figure 5.9 B). Two genes whose expression change was significant stood out as being more downregulated than the others; one of these was CNOT1 as expected, and the other was a little studied lncRNA, AC004148.1 which is antisense to the protein coding gene RPAIN. Gene lists for significant (FDR < 0.05), greater than 2 fold up- or downregulation were assayed for gene ontology enrichment. Upregulated RNAs

were overrepresented in a varied group of terms including several concerning cell adhesion and secretion, as well as signalling pathways involved in proliferation (figure 5.9 C). Downregulated RNAs were almost exclusively enriched for terms relating to mitosis (figure 5.9 D).

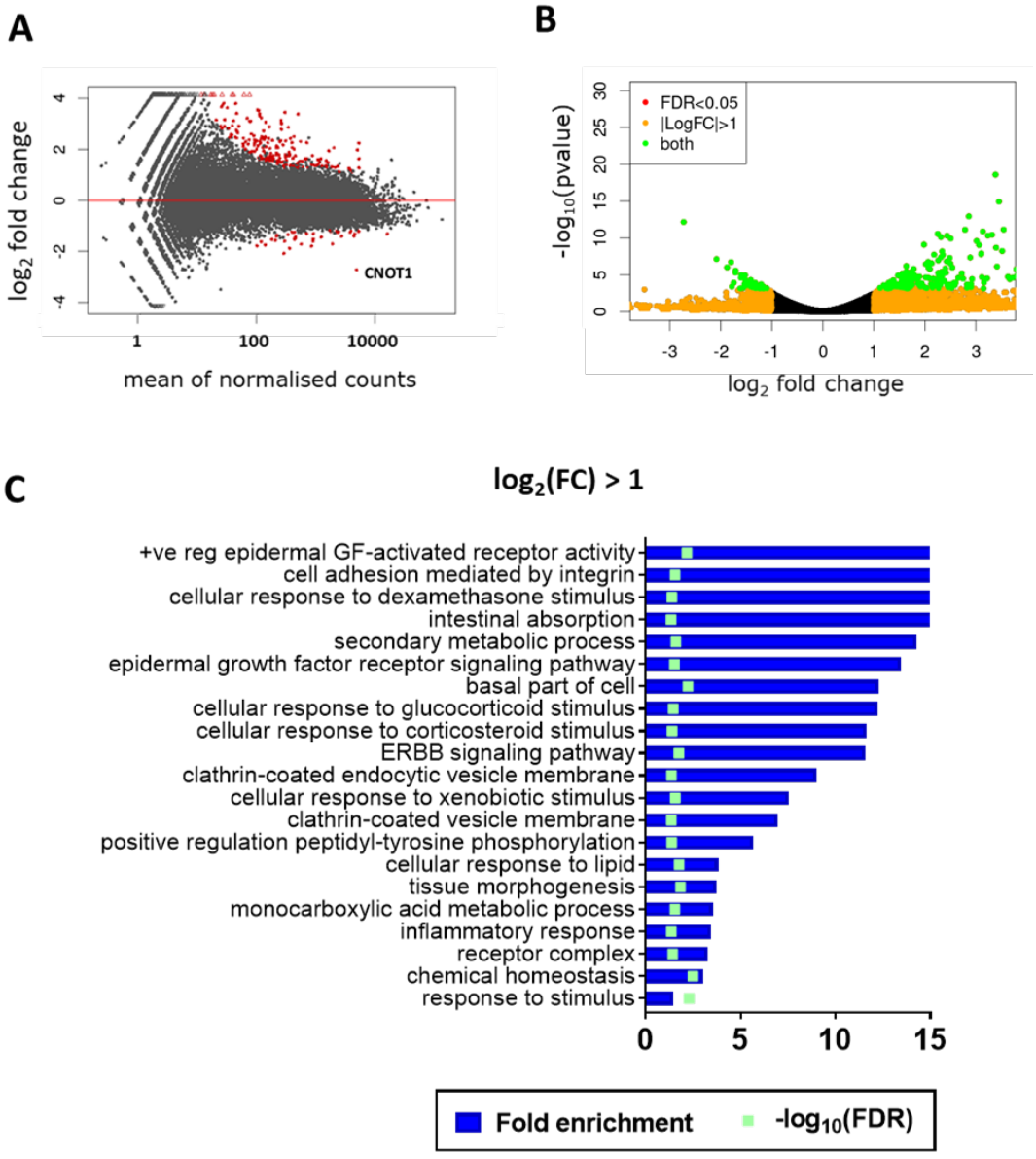


Figure 5.8 Differential expression analysis for total reads counted by gene. Raw counts were generated per gene and duplicates were filtered out. Count data were subjected to differential expression analysis with DESeq2. **A)** Fold change in CNOT1 knockdown compared with average expression across all samples. Red denotes adjusted p-value < 0.05. **B)** Volcano plot showing significant up and downregulated genes. **C)** GO term (+ve) enrichment for genes with a significant (p_{adj} < 0.05) log_2 fold change >1 in the CNOT1 knockdown condition. Terms with fold enrichment > 3 and FDR < 0.05 are displayed.

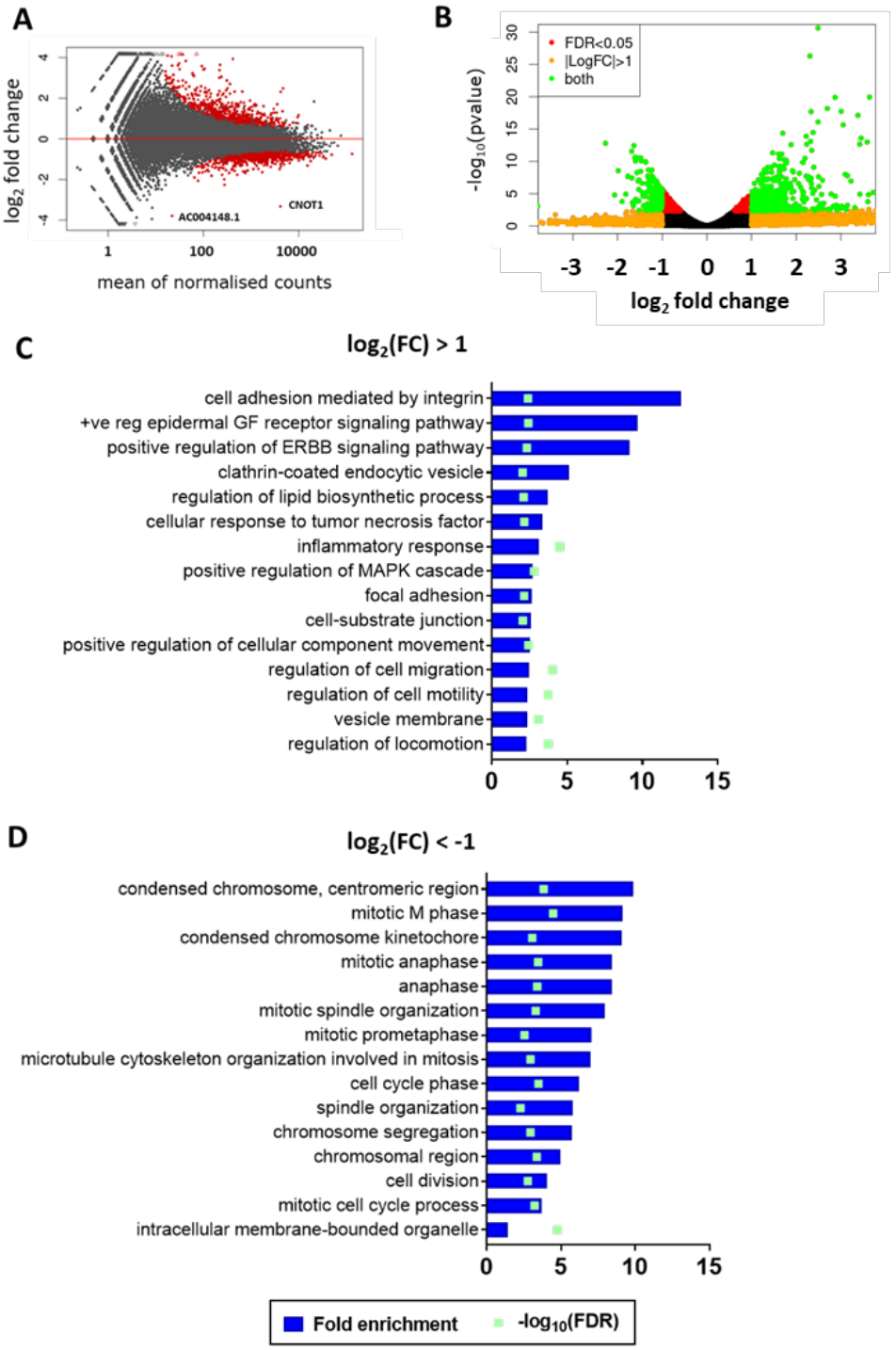


Figure 5.9 Differential expression analysis for total reads indicates roles for CNOT1 in proliferation and cell migration. Raw counts were generated per exon then grouped into genes, and duplicates were removed. Count data were subjected to differential expression analysis with DESeq2. **A)** Fold change in CNOT1 knockdown compared with average expression across all samples. Red denotes adjusted p-value < 0.05. **B)** Volcano plot showing significant up and downregulated genes. **C)** GO term (+ve) enrichment for genes with a significant (p_{adj} < 0.05) $\log_2$ fold change >1 in the CNOT1 knockdown condition. Top 15 most enriched terms with FDR < 0.01 are displayed. **D**) As in C, but for $\log_2$ fold change <-1 in the knockdown condition.

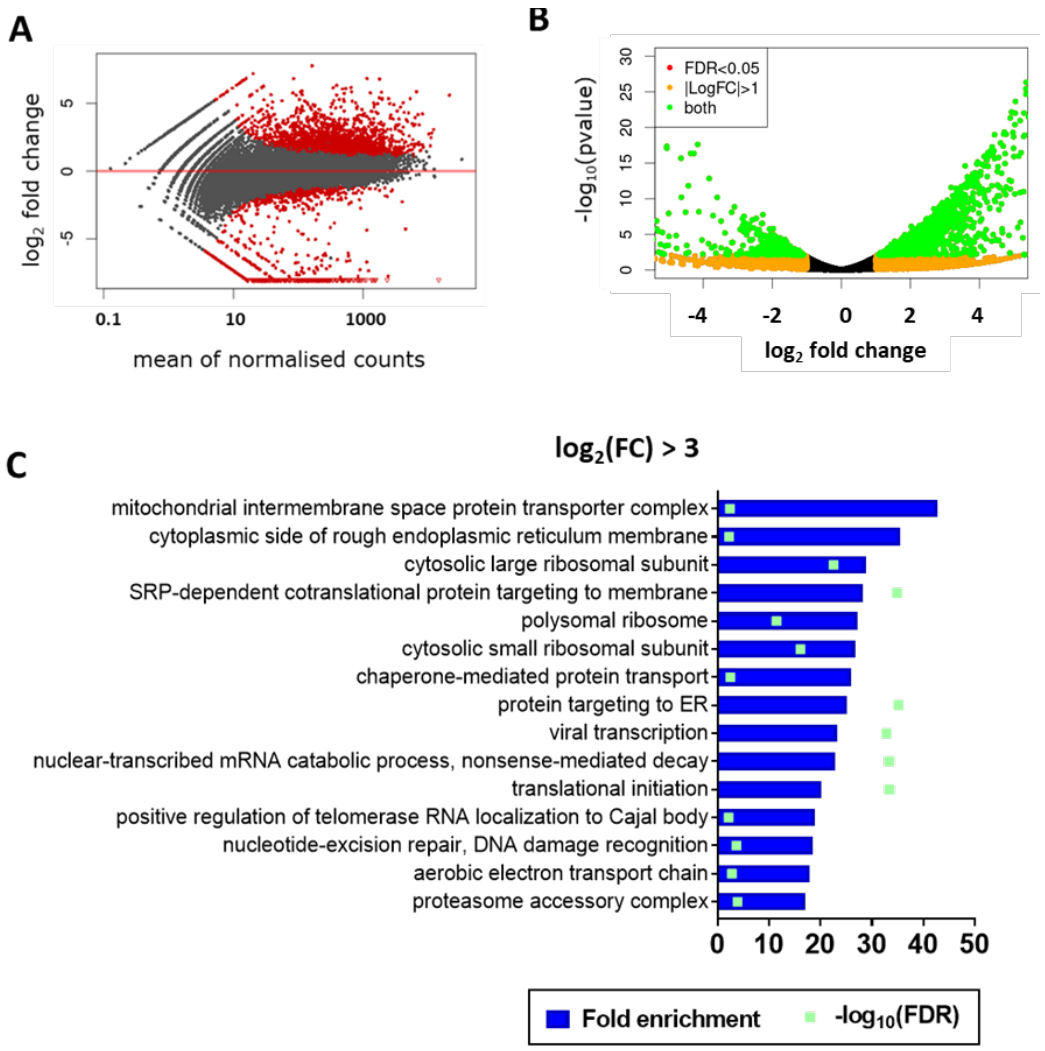


Figure 5.10 Differential expression analysis for exon-depleted reads indicates roles for CNOT1 in proliferation. Raw counts were generated per exon then grouped into genes, and duplicates were removed. Count data were subjected to differential expression analysis with DESeq2. **A)** Fold change in CNOT1 knockdown compared with average expression across all samples. Red denotes adjusted p-value < 0.05. **B)** Volcano plot showing significant up and downregulated genes. **C)** GO term (+ve) enrichment for genes with a significant ($\mathbf{p_{adj} < 0.005}$) $\log_2$ fold change > 3 in the CNOT1 knockdown condition. Top 15 most enriched terms with FDR < 0.01 are displayed.

Since changes to unspliced mRNA levels were previously observed (figure 5.1 B), differential expression analysis was also performed on exon-depleted reads. A high proportion of genes displayed significant changes in expression, though it is not clear how genuine these were (figure 5.10 A, B). A more stringent cut-off for false discovery rate (FDR) had to be used than for total reads (<0.005 instead of <0.05) and only genes with a fold change > 3 were assayed for GO term enrichment. Genes

upregulated at the 'unspliced' level were enriched for terms involving ribosomes/ translation and [protein targeting to] intracellular membranes, as well as for nonsense-mediated decay and DNA damage recognition (figure 5.10 C). A list of downregulated genes returned no significant GO term enrichment, but this could have been due to contamination with large numbers of what may be falsely significant points (figure 5.10 A).

GO term analysis seemed to indicate downregulation of genes involved in proliferation at the mature level, but this could not be confirmed for exon-depleted reads. Cursory appraisal of the significantly up- or downregulated genes did not obviously corroborate the earlier observation of reduced serum induction in Cnot1 knockdown NIH 3T3 cells (figure 5.1 B). Given the high level of variation between knockdown replicates (figure 5.7), it was possible that these differences were present but were not detected as significant by the DESeq2 model. Subtle changes in culture conditions or overlooked differences in experimental procedure or timings could also have introduced variation.

TPM (Transcripts Per Million - a normalised count measurement which takes into account sequencing depth and transcript length) was therefore calculated for the genes assayed previously using the raw count files generated by featureCounts. For total RNA, reads were counted by exon and TPM was calculated using transcript rather than gene length. Since pre-mRNA exists at a range of lengths depending on extent of splicing, TPM for exon-depleted reads was calculated using the theoretical maximum of gene length (and reads were counted per gene).

As in the sample similarity heatmap (figure 5.7 A, B) the two control replicates showed similar expression of all genes, whereas the CNOT1 knockdown samples were more disparate (figure 5.10). Rather than being overall more variable, the differences seemed to be consistent within replicates, with the second replicate exhibiting a reduction in serum response mRNAs at the mature and 'unspliced' (exon-depleted) levels (figure 5.10). This suggested a difference in serum exposure, for example due to passing cells or replacing medium at different times. Unspliced TPM was calculated using both the sum of total aligned fragments (i.e. normalising to sequencing depth, figure 5.10 B) or to the sum of exon-depleted aligned fragments (i.e. normalising to total pool of 'unspliced' mRNA, figure 5.10 C). Both

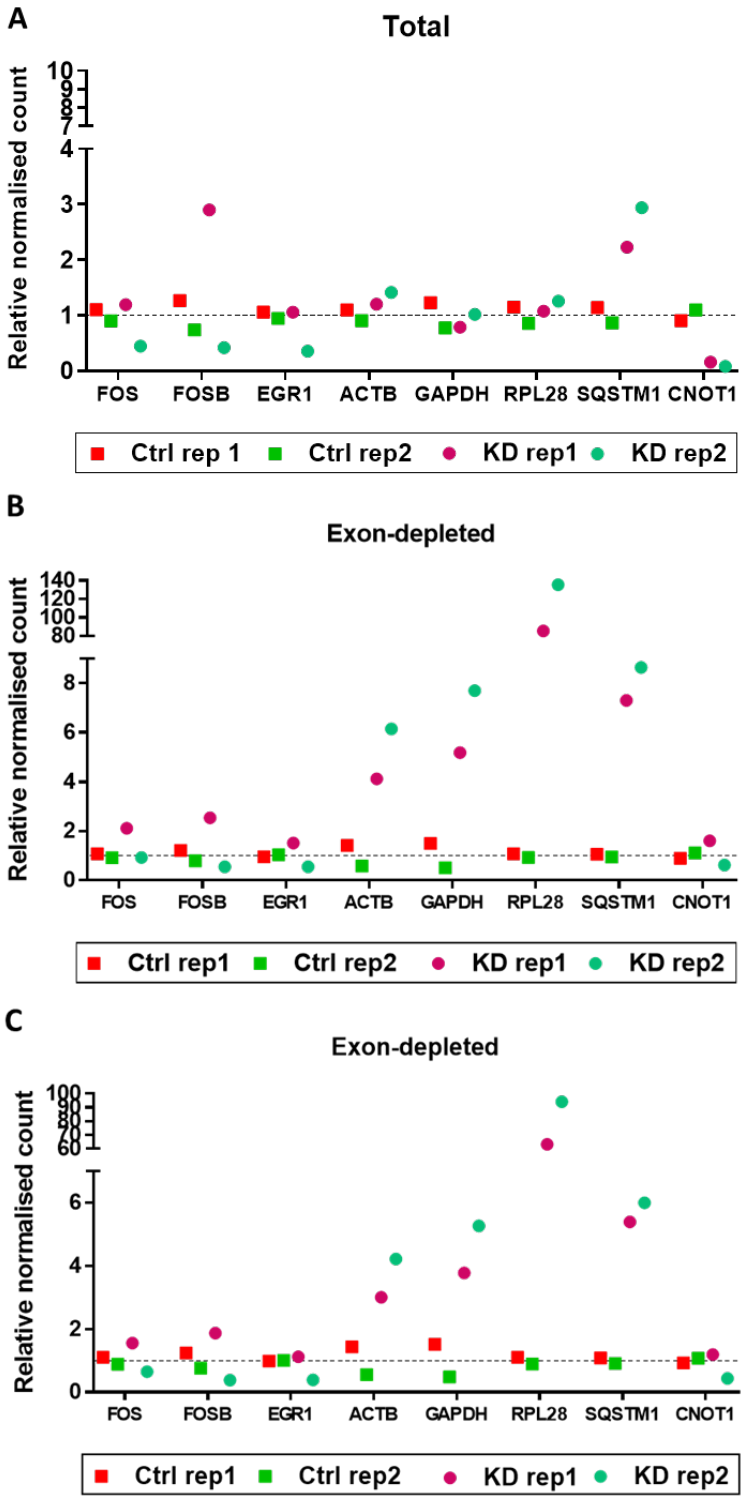


Figure 5.11 Housekeeping mRNAs are increased at the unspliced level in CNOT1 knockdown. RNA-Seq data normalised by TPM and made relative to the mean of the control counts. Duplicated fragments removed. **A)** Total RNA – aligned against genome and splice junction database. **B)** 'Unspliced RNA' / introns only - aligned against genome then fragments which overlap >95% with exons removed. TPM calculated using sum of raw counts for all aligned fragments in the sample. **C)** As in B, but TPM calculated using sum of aligned fragments following exon removal.

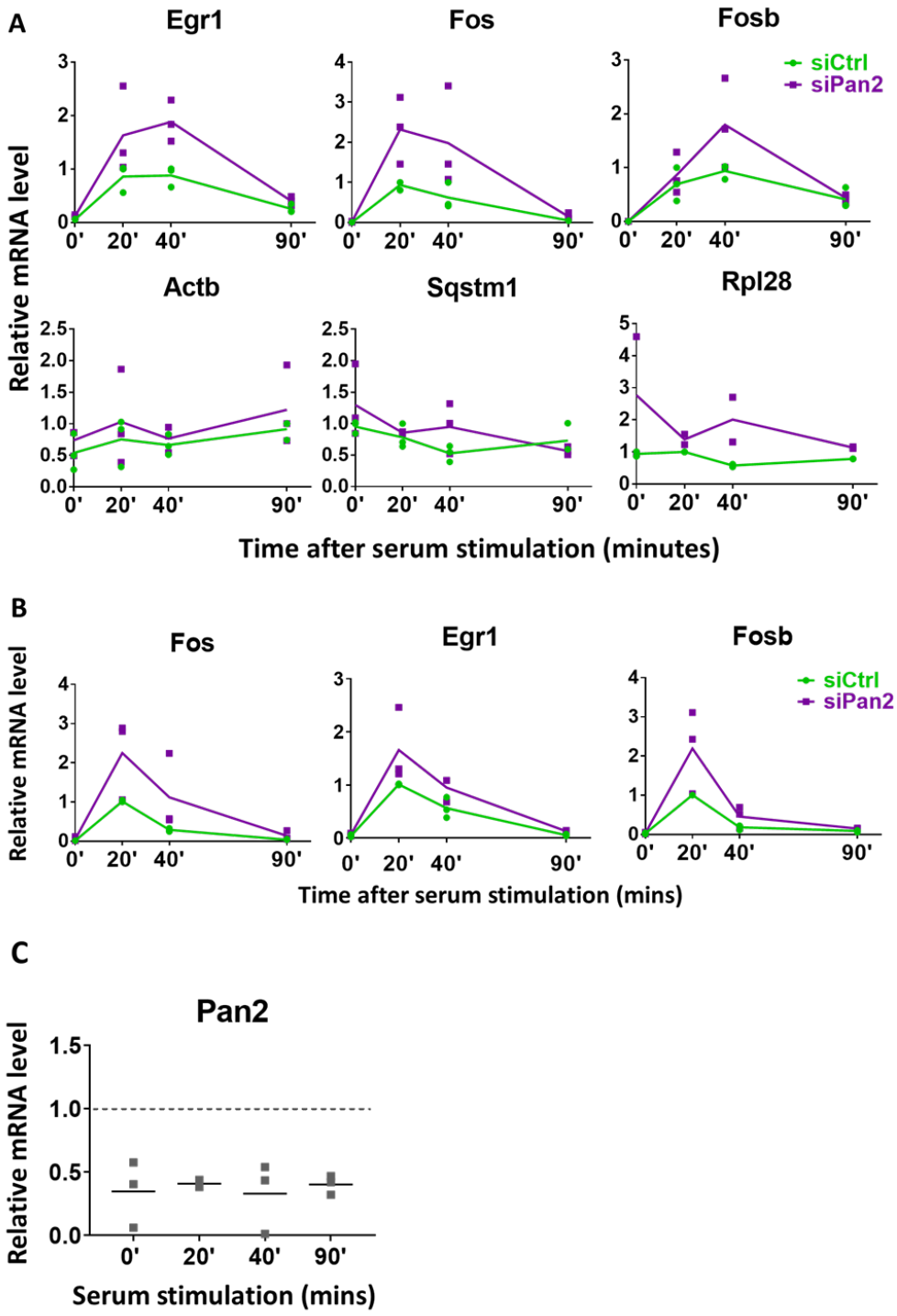


Figure 5.12 Pan2 knockdown caused a moderate increase in serum induced gene expression at the mature and unspliced levels. qPCR data for 3 biological replicates showing **A)** mature and **B)** unspliced mRNA abundance relative to Gapdh and normalised to the maximum value within the control time course. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.01. **C)** Validation of Pan2 knockdown at the mRNA level, relative to Gapdh and normalised to the control level for each time point.

choices resulted in a similar overall pattern. Notably, TPM values for previously used housekeeping mRNAs increased substantially at the 'unspliced' level, which is in contrast to observations by qPCR in NIH 3T3 cells (figure 5.1 B). A consistent reduction in serum response mRNAs was not observed across both knockdown replicates at either level, though this may in part be due to analysis at steady state rather than following serum induction. Overall the analysis of this dataset did not seem to allow clear conclusions.

5.4 Pan2 knockdown increases mRNA levels early on, but does not prolong gene expression

To explore whether reduction in mRNA abundance is a conserved response to deadenylase depletion in NIH 3T3 cells, the catalytic subunit of PAN2-PAN3, Pan2, was knocked down. PAN2-PAN3 is theorised to act on long tails which appear early in the serum response for the mRNAs studied here (26); indeed the poly(A) length measurements presented in chapter 4 showed that Pan2 knockdown caused a small increase in tail length at 20 minutes but not at any later time points (figure 4.5 A). Hence, Pan2 knockdown serum stimulation time courses were carried out over a slightly shorter period than the Cnot1 equivalents. Moderate knockdown was achieved at the mRNA level for all 3 replicates (figure 5.12 C).

Pan2 knockdown caused an increase in induced mRNA levels, consistent with a decreased degradation rate (figure 5.12 A). These decreased to near-control levels by 90 minutes, likely indicating compensation by other deadenylases such as CCR4-NOT. Unexpectedly, Pan2 knockdown also caused an increase in unspliced mRNA level for induced genes (figure 5.12 B). These data were much more variable than their Cnot1 counterparts (figure 5.1 A, B) and were largely not statistically significant, possibly due to modest and variable knockdown efficiencies. There was very little difference in housekeeping gene expression (figure 5.12 A).

5.5 Discussion

Quantitative real time PCR data showed that in addition to increasing early poly(A) tail length, Cnot1 knockdown dampened the transcriptional response to

serum induction (figure 5.1 A, B). This dampening of the response seemed to be unique to the CCR4-NOT complex, since knockdown of the PAN2-PAN3 nuclease had a smaller and opposite effect (figure 5.12 A, B).

Reduction in mRNA levels appeared to be at the level of unspliced mRNA, but preliminary data showed that chromatin associated pre-mRNA levels did not change (figure 5.4). This suggests that reduced expression was not occurring at the transcriptional level, though in the absence of metabolic labelling this conclusion is somewhat speculative. It is possible that newly made mRNA was more vulnerable to nuclear degradation in Cnot1 knockdown, perhaps due to the generation of aberrantly long poly(A) tails which could destabilise transcripts - though nuclear retention of hyperadenylated transcripts has also been recorded in mammalian cells (321, 415). Since both mature and unspliced levels decreased in the nucleoplasmic fractions, any degradation would have had to occur prior to splicing of the relevant introns, but subsequent to addition and detection of a long tail. This may have been possible, since it was recently shown by Oxford Nanopore sequencing that splicing of the 3' terminal exon primarily occurs after cleavage in mammalian cells (121). Given the role of CCR4-NOT in promoting Pol II elongation (147, 148), it is equally possible that transcription initiation was slower in the knockdown, but that concomitant slowing of elongation resulted in no detectable change in chromatin-associated mRNA abundance.

Analysis of a publicly available but not yet published CNOT1 knockdown dataset for HeLa cells found no obvious transcriptional downregulation of serum-induced transcripts (figure 5.11), however this may have been due to differences in experimental design since RNA was isolated from cells growing at steady state, or to variation between species. Equally, only a handful of genes were considered in NIH 3T3 cells and these may not be reflective of the wider transcriptome. At the level of total RNA (i.e. mainly mature transcripts), CNOT1 knockdown led to upregulation of transcripts involved in cell migration, secretion, and lipid synthesis and downregulation of genes related to mitosis (figure 5.9 C, D). This points to a role for CNOT1 in proper control of proliferation, which is consistent with observation of reduced proto-oncogene expression in NIH 3T3 cells. Indeed, a previous study showed that knockdown of CNOT1 or CNOT3 caused reduced proliferation of MCF7

cells but did not affect viability (525). It is not clear whether genes involved in cell growth and proliferation are the main targets of CNOT1 regulation or whether regulation occurs more widely on short-lived mRNAs of many processes; it is possible that the particular effects seen here resulted from the use of a cancer cell line where expression of oncogenes is higher and therefore more susceptible to downregulation.

Crosstalk between mRNA synthesis and decay has been relatively well explored in yeast, but less is known about the situation in higher eukaryotes (58, 148, 470, 471, 578–580, 582, 585). In yeast, crosstalk promotes mRNA homeostasis (with high reliance on both Ccr4-Not and Xrn1), however mammalian studies are sparse. Data presented here for NIH 3T3 cells indicate that disruption of mRNA decay through knockdown of a CCR4-NOT subunit may cause a decrease in mRNA production in order to maintain mRNA homeostasis. This at first seems incongruous with findings in mammalian cells that increased degradation caused by expression of the viral nuclease, SOX also led to reduced transcription (470, 471) but the two scenarios may represent distinct biological situations for which the same response may not be appropriate. A recent paper found links between transcription rate and poly(A) tail length in mammalian cells and also observed changes in expression of CCR4-NOT subunit mRNAs during B cell activation and glucose starvation, highlighting the importance of the complex in maintaining correct transcript levels in dynamic processes (58). This may simply be through its deadenylase activity, however, another study which knocked out *Cnot1* in mouse livers found that as well as broad transcript stabilisation, different groups of genes (which did not include transcription factors) were affected at the pre-mRNA level (591).

The extent of crosstalk between mRNA birth and decay is as yet unknown, and the end goal of such may vary between both species and scenarios. The majority of investigations have used knockout and knockdown experiments which make it difficult to assess how soon information is conveyed, and therefore hard to pinpoint the most likely mode of communication. Despite the extensive mechanistic exploration that remains, a central role for the CCR4-NOT complex seems certain.

6 Probing mechanism of CNOT1-mediated early poly(A) and pre-mRNA control

6.1 Introduction

The previous chapters showed that Cnot1 knockdown caused unusually long poly(A) tails on induced mRNAs late in the NIH 3T3 serum response, along with reduced induction of gene expression (figures 4.2, 4.3, 5.1). Constitutively expressed mRNAs also exhibited modestly increased poly(A) length in knockdown cells, and in some cases also reduced expression. It was not clear how Cnot1 knockdown reduced mRNA abundance while presumably increasing mRNA stability.

CNOT1 is an essential component of the CCR4-NOT complex, behaving as a scaffold for its several modules (513, 516, 517). CCR4-NOT structure and function is discussed in considerable depth in section 1.4.1. Some transcriptional effects of the complex are mediated by the Not module, whereas the specific subunits required for stimulation of Pol II elongation seem to include components of both the Not and deadenylase modules (147, 148, 512, 516). Deadenylation is mediated by the deadenylase module. While the former is a widely accepted nuclear function, deadenylation by CCR4-NOT is usually regarded as cytoplasmic.

The following work aimed to uncover – via several lines of investigation – whether the coupling observed between initial poly(A) tail length and mRNA production occurred through dual nuclear functions of CCR4-NOT (i.e. both transcription and nuclear deadenylation), or through a wider feedback mechanism sensing changes to cytoplasmic deadenylation/degradation (e.g. through cycling of some poly(A)/mRNA-bound factors back into the nucleus). It was also hoped these experiments would provide some insight into the varied magnitude of effect Cnot1 knockdown had on the poly(A) tails of different gene classes. Taking into account the differences in tail length observed on transiently expressed mRNAs across the serum time course and the differences observed between genes, one could envisage a model in which nuclear deadenylation is modulated both in time and gene-specifically by nuclear availability of some specific factor which either protects the poly(A) tail or targets it for degradation. An analogous system could be envisaged for regulation of PAP activity, but this would not require CCR4-NOT to exhibit

nuclear deadenylase activity. Possible mechanisms are summarised in figure 6.1. A handful of these possibilities are considered in the following chapter, prefaced where relevant by a short introduction to concepts which have not been discussed thus far.

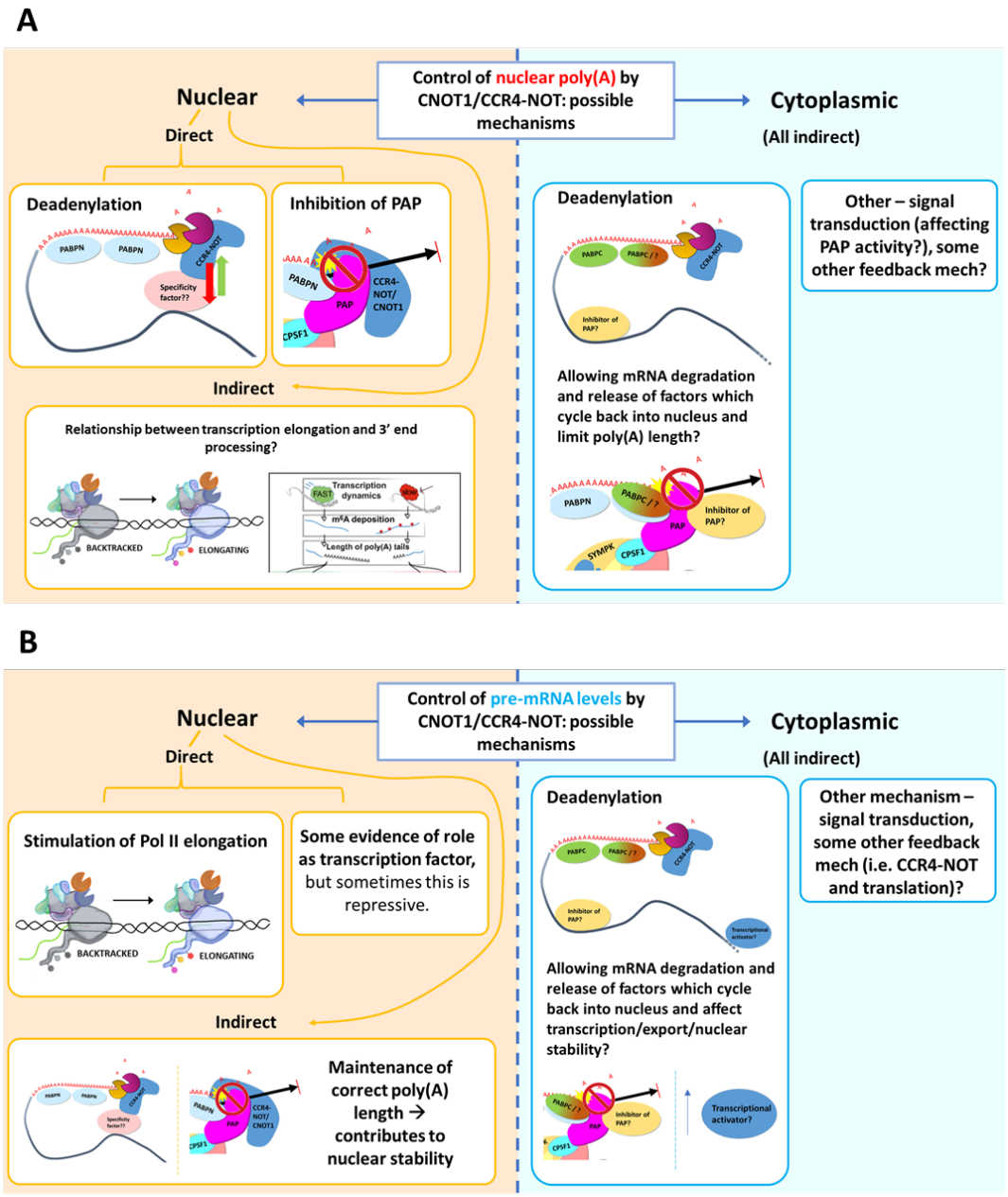


Figure 6.1 Schematics of possible mechanisms for CNOT1 involvement in **A)** nuclear poly(A) tail length control and **B)** pre-mRNA level. Nuclear CCR4-NOT could either directly remove poly(A) tails or perhaps inhibit their addition. CCR4-NOT could couple transcription elongation with 3' processing, or could be recruited by increased m^6A deposition in the case of slow elongation as proposed by Slobodin et al (58). Note: this was originally proposed as a cytoplasmic phenomenon, but the m^6A reader YTHDF2 with which CCR4-NOT interacts is also present in the nucleus. Cytoplasmic CCR4-NOT levels could affect release of factors from the mRNA which cycle back to the nucleus to influence poly(A) length/transcription.

6.2 CNOT1 is present in the nucleus at all stages of the serum response

Critical to understanding the mechanism through which Cnot1 knockdown elicited more extensive polyadenylation was the question of whether CNOT1 was present in the nucleus in the NIH 3T3 cells used, and therefore whether a model in which CCR4-NOT rapidly trims the tails of chromatin-associated mRNA was plausible. Study of this was first attempted using immunofluorescence, however, no CNOT1 antibody was specific enough (as judged by a failure to reduce signal in knockdown cells, data not shown). Instead, western blots were performed on nuclear/cytoplasmic fractionations at three time points of interest (0', 20', 60') and fractionation efficiency tested by probing for Lamin A/C (nuclear) and α-Tubulin (cytoplasmic). Both replicates showed good separation (figure 6.2 C, D). Preliminary data suggested that CNOT1 protein was not present in the nuclei of our NIH 3T3 cells (data not shown), but in subsequent experiments nuclear CNOT1 was easily detectable (figure 6.2 B-D).

The first anti-CNOT1 antibody used (from Proteintech) bound several proteins in the vicinity of the 250 kDa marker, none of which gave an obviously lower signal in the knockdown lanes (figure 6.2 A). To see if any of these were CNOT1, the blot was stripped and probed with an alternative anti-CNOT1 antibody (Novus). This gave a much cleaner signal, showing two main bands in both fractions of the control which were depleted in the knockdown, with a slightly greater reduction in the cytoplasmic fraction (figure 6.2 B). The higher band seemed more likely to be CNOT1 on the basis that its predicted molecular weight is 267 kDa, but staining was much stronger for the lower band, suggesting that perhaps CNOT1 did not run to its exact predicted size. Alternatively, since multiple CNOT1 isoforms exist around the same molecular weight (as well as others which are lower), it seemed possible that both bands were CNOT1. Having established efficacy of the newer antibody, western blots for serum stimulation fractionations were stripped of the original CNOT1 antibody and re-probed. In both replicates, CNOT1 was present in the nucleus at all time points (figure 6.2 C, D) and therefore direct nuclear regulation of poly(A) tails by CCR4-NOT cannot be ruled out. While the ratio of nuclear:cytoplasmic (N:C) CNOT1 protein seemed consistent across all time points of the second replicate, in

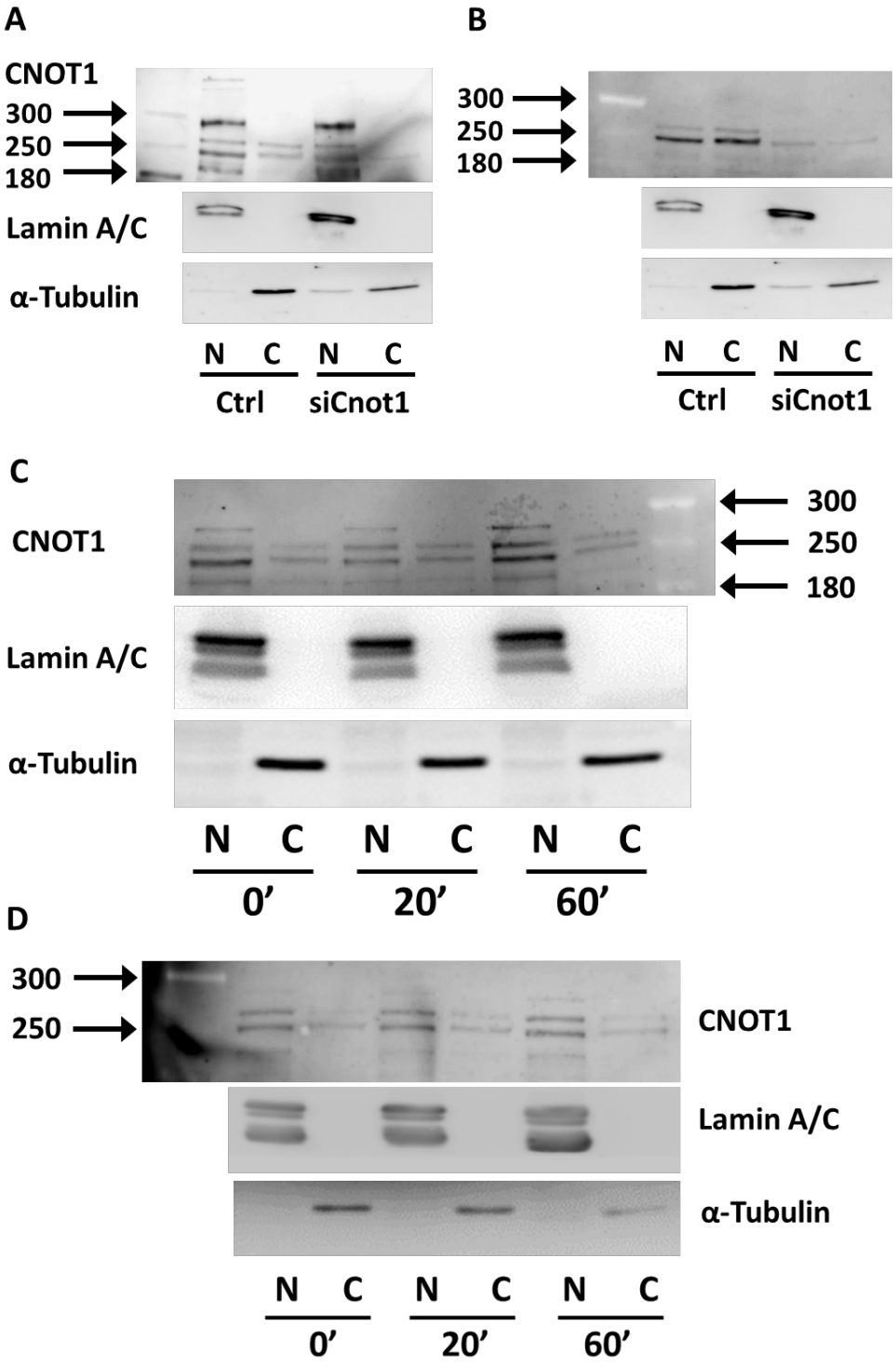


Figure 6.2 CNOT1 is present in the nucleus at all time points studied in the serum response. A) Fractionation of Cnot1 knockdown and control cells stimulated for 60 minutes, incubated with Proteintech anti-CNOT1 antibody. **B)** The same blot, stripped and re-probed with Novus biologicals anti-CNOT1 antibody which is more specific. **C)** and **D)** Two biological replicates showing CNOT1 protein localisation in cells fractionated at 0, 20 and 60 minutes following serum stimulation. In all cases, Lamin and Tubulin were used to both validate separation of nuclear and cytoplasmic fractions, and to confirm equal loading.

the first replicate the N:C ratio of the lower, darker band seemed to be reduced at the 20 minute time point compared with 0’ or 60’. Although this compliments the observation that nuclear poly(A) tails of induced mRNAs are long after 20 minutes stimulation, this was only a moderate effect in one replicate and western blots are not a reliable source of exact quantification. It therefore seems more likely that any nuclear deadenylation of these mRNAs is modulated by some specificity factor which itself changes localisation or activity, in order to achieve serum-induced temporal changes in poly(A) length.

6.3 Probing simple models of nuclear deadenylation

As CNOT1 was present in the nucleus at all time points tested, by far the simplest model to marry its effect on nuclear poly(A) tail length with the observations of shorter than expected nuclear tails in control cells was one of direct nuclear deadenylation by CCR4-NOT. In such a model, mRNAs would still be synthesised with 200 nucleotide tails which are then trimmed in the nucleus by the CCR4-NOT complex to some gene or time-specific proper length. Similar message-specific deadenylation by PAN (PAN2-PAN3) in yeast was suggested previously, and more recently a model was put forwards in which CCR4-NOT prunes the poly(A) tails of highly expressed mRNAs upon their entry to the cytoplasm (326, 430). While a tailored nuclear, or indeed chromatin-associated, trimming model could have been valid for the majority of mRNAs tested (where long tails were detectable in the chromatin fraction), it seemed inconsistent with the absence of long poly(A) tails in any location or condition for Rpl28 (figures 3.5 A,D, 4.3) unless this absence was caused by extremely rapid deadenylation. There were several possibilities for how the gene and time – specific control of deadenylation could have been mediated. Given that poly(A) tails on induced mRNAs tend to be longer early in the serum response, when transcription is rapid, it seemed possible that some factor(s) which targets CCR4-NOT to specific mRNAs was present either in a limiting quantity, or had yet to translocate to the nucleus. Conversely, some protective factor(s) with variable affinities for different mRNAs may have protected the tail initially but then be exported from the nucleus. Alternatively, nuclear deadenylation by CCR4-NOT could simply have been a function of nuclear dwell time, with those RNAs resident in the nucleus for longer allowing more time for their tails to be degraded.

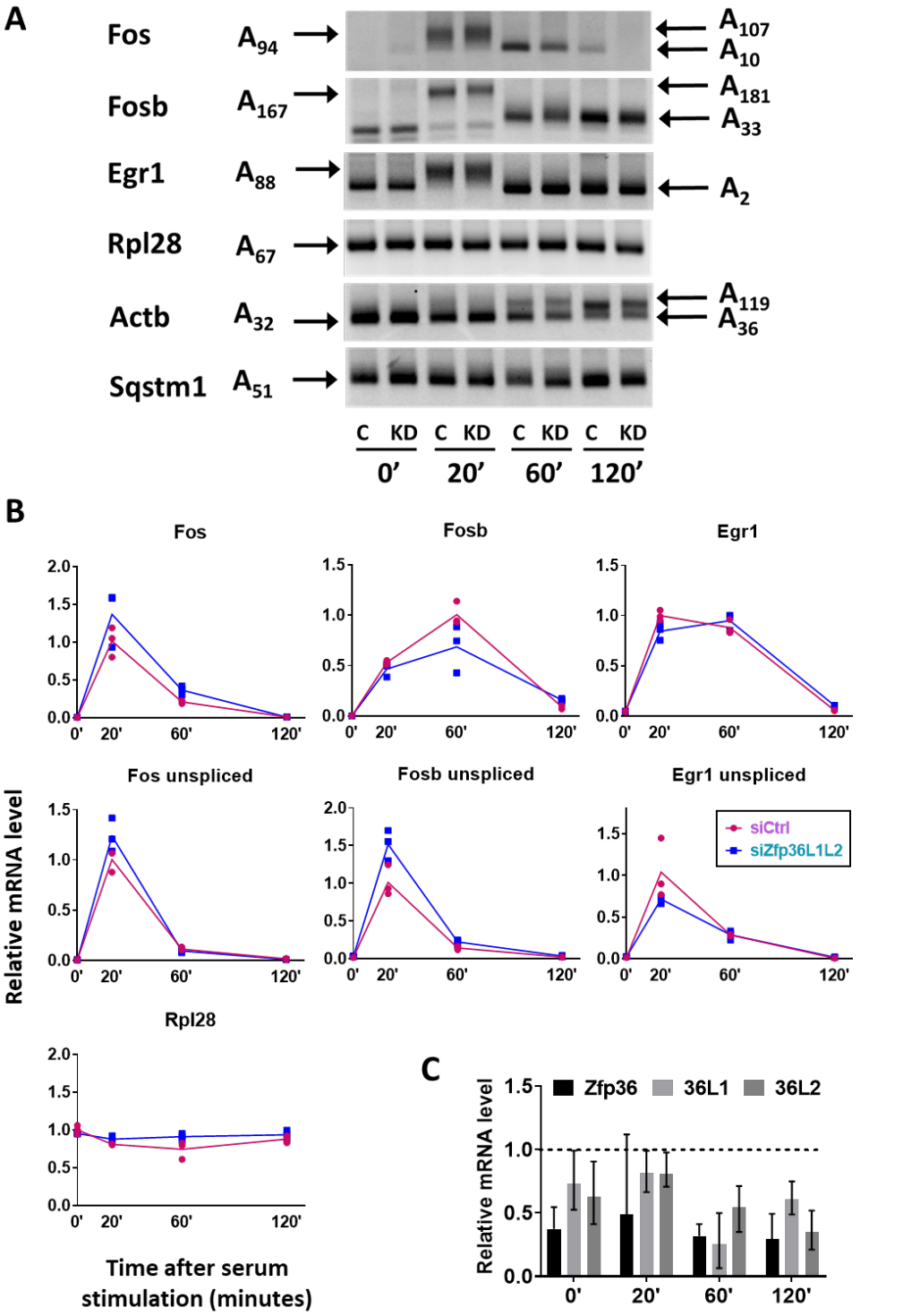


Figure 6.3 TTP knockdown has minimal effect on serum response mRNA poly(A) tail length. A) RL2-PAT, using agarose gel stained with SYBR safe, showing changes in poly(A) tail lengths over a serum time course in combined Zfp36,L1&L2 knockdown cells. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. Times shown indicate duration of serum stimulation in minutes. **B)** qPCR data relative to Gapdh showing mature and unspliced mRNA abundance across the serum response. Single biological (3 technical) replicate shown. **C)** qPCR data relative to Gapdh showing knockdown efficiency at the mRNA level. Single biological replicate.

6.3.1 TTP knockdown only minimally affects poly(A) tail lengths of serum response mRNAs

One factor known to specifically target a subset of mRNAs for deadenylation is the zinc finger protein, Tristetraprolin (TTP, Zfp36). TTP binds to AU-rich elements (AREs) in the 3' UTRs of target mRNAs - including those involved in the serum response - and interacts with the CNOT1 and CNOT9 subunits of CCR4-NOT to elicit their deadenylation (51, 290, 292, 626, 627). Though this is thought mainly to be a cytoplasmic phenomenon, TTP is also present in the nucleus and could therefore direct CCR4-NOT to specific mRNAs in the nucleus and thereby mediate their trimming (297, 431). Previous work also implicated TTP in limiting poly(A) tail synthesis on ARE-containing mRNAs in the nucleus via interaction with PABPN1 and PAP (431). mRNAs encoding TTP and its paralogues ZFP36L1 and ZFP36L2 were knocked down using RNAi, and effects on transcription and poly(A) length were measured.

As assessed by relative qPCR, the knockdown of TTP and its paralogues at the mRNA level was modest (figure 6.3 C). A minimal increase in poly(A) tail length was visible after 20 minutes of stimulation for Fos and after 20 and 60 minutes' stimulation for Fosb, likely due to slower cytoplasmic deadenylation (figure 6.3 A). There was no discernible change for the other genes. The effect was much smaller than in Cnot1 knockdown, suggesting either that TTP levels were not sufficiently depleted or that TTP had only a limited role in poly(A) control for these mRNAs. In addition, there were no clear changes in mRNA level at the mature or unspliced levels, though only one biological replicate was performed so a definitive conclusion cannot be drawn (figure 6.3 B). Based on observations of only minimal changes, it seemed unlikely that TTP was involved in the regulation of nuclear poly(A) length and its coupling with mRNA production for the mRNAs tested.

6.3.2 Nxf1 knockdown – does nuclear dwell time determine poly(A) length?

Failure to detect changes in localisation of CNOT1 at different time points left the mystery of temporal nuclear poly(A) regulation in the serum response unanswered. Since Cnot1 knockdown caused increased nuclear and chromatin associated poly(A) tail lengths at late time points (figures 4.2 A-C, 4.3 A,B) , a logical hypothesis could have been that these tails are normally subjected to deadenylation by CCR4-NOT for

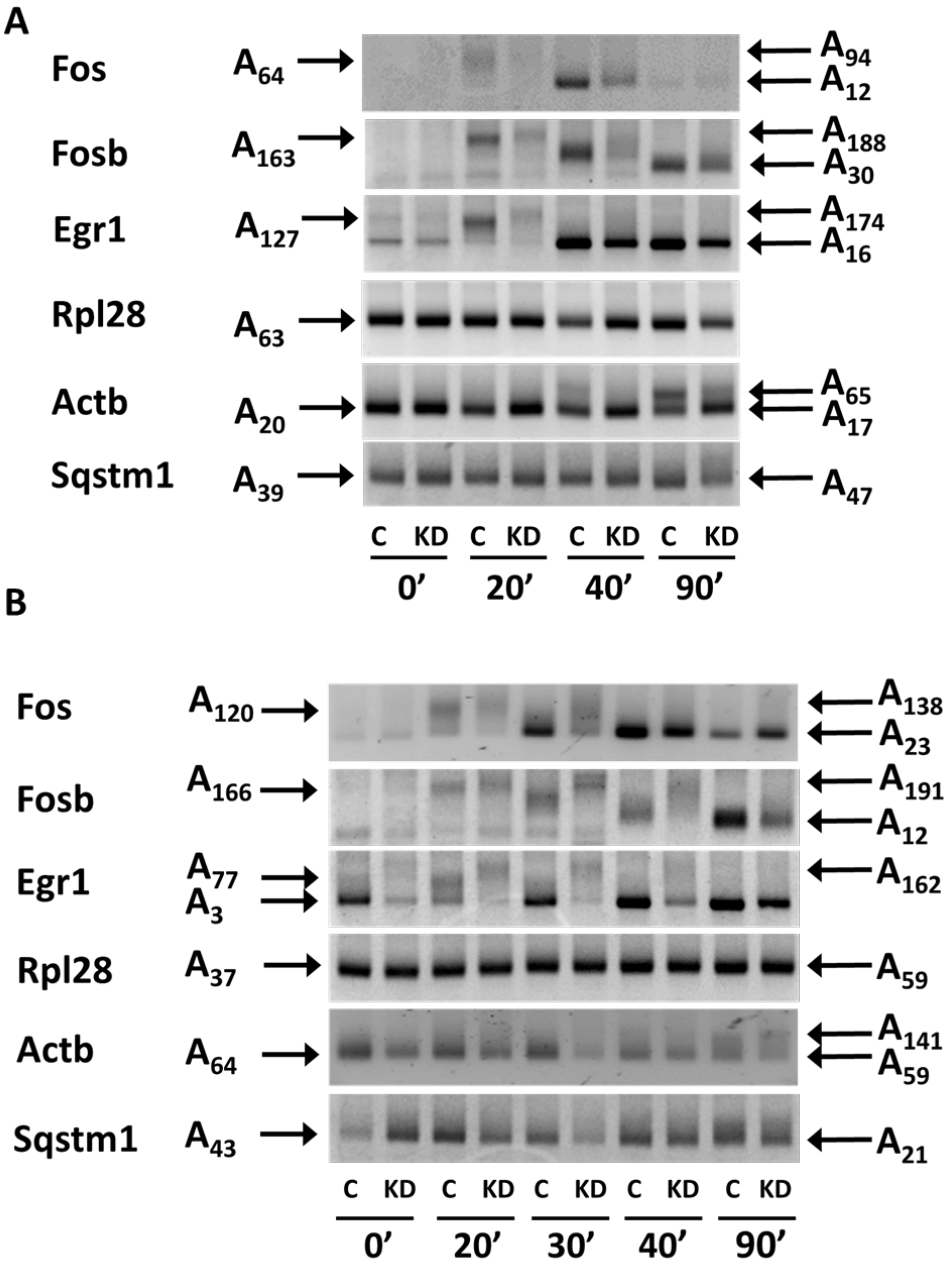


Figure 6.4 Inhibition of mRNA export through knockdown of Nxf1 results in longer poly(A) tails on induced mRNAs (continued overleaf). **A, B)** RL2-PAT, using agarose gel stained with SYBR safe, showing changes in poly(A) tail lengths across serum time courses in Nxf1 knockdown cells for two biological replicates. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. Times shown indicate duration of serum stimulation in minutes. **C)** qPCR data showing mRNA abundance throughout serum induction in Nxf1 knockdown, relative to Gapdh. 3 biological replicates are shown for most time points, except 30 minutes, which only has 2 due to its initial omission from the experiment. In addition, only two replicates were included for Actb. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.05, ** indicates FDR adjusted p <0.01. **D)** Validation of Nxf1 knockdown at the mRNA level.

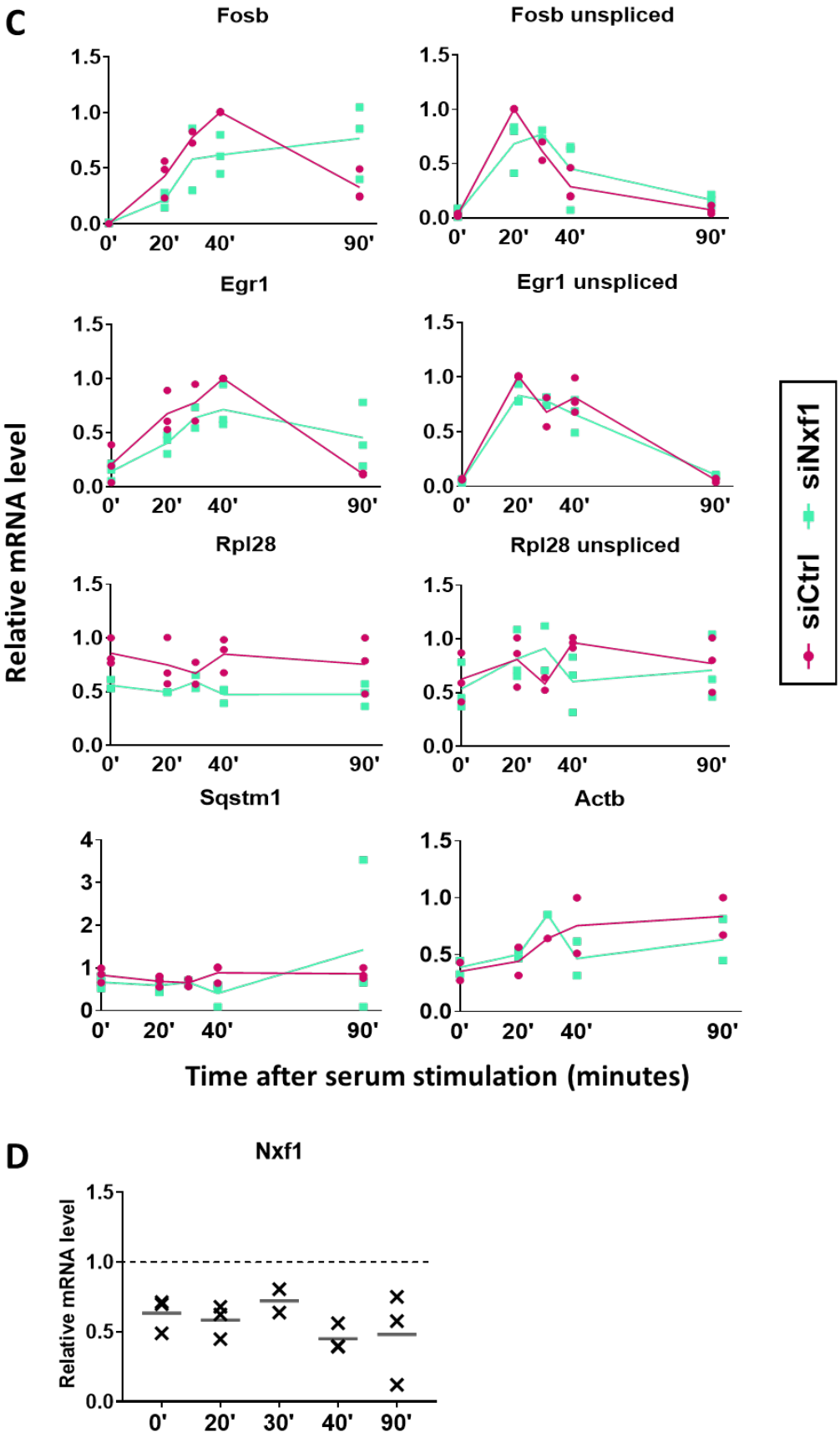


Figure 6.4 contd.

longer due to extended nuclear dwell times. In such a model, some of those mRNAs synthesised at the beginning of the serum response would be rapidly exported and so escape substantial deadenylation, whereas others may not be commandeered for export so early, and therefore be trimmed more extensively in the nucleus. To this end, Nxf1, a gene thought to be essential for bulk mRNA export, was knocked down to abrogate export (628, 629). If longer nuclear dwell time enabled more substantial trimming of poly(A) tails by CCR4-NOT, knocking down Nxf1 should have caused shorter tails earlier in the time course, though the magnitude of such an effect would have depended on the speed of CCR4-NOT deadenylation.

In all experiments, treatment with Nxf1 siRNA led to significant cell death on day 4 so the method was adjusted to perform serum starvation and the second siRNA transfection on the same day for the final replicate. Knockdown at the mRNA level in the surviving cells was poor but sufficient to elicit clear effects on tail length (figure 6.4 A, B, D). Knockdown of Nxf1 led to longer poly(A) tails, though the bands were considerably more diffuse than those observed in the Cnot1 knockdown. In addition, only unstable mRNAs seemed to experience an increase in poly(A) length (figure 6.4 A, B). Consistent with increased stability conferred by longer poly(A) tails, higher levels of mature serum response mRNAs were detectable in the knockdown cells late in the time course, despite no changes to unspliced levels (figure 6.4 C). Housekeeping mRNAs were not affected. Since only whole cell lysate was assayed, it is not clear whether the longer tailed transcripts were successfully exported or nuclear-retained.

6.4 Inhibition of CAF1 activity and knockdown of the encoding mRNAs have distinct effects

The CCR4-NOT complex has been implicated in several additional molecular functions, including transcription elongation, and depletion of the scaffold component CNOT1 likely disrupts the entire complex (39, 42, 44, 147, 516, 517). It was therefore important to determine whether the effects of Cnot1 knockdown on pre-mRNA level were due to loss of deadenylase function, or to loss of the complex's other activities – though of course integrity of the whole complex in the absence of the deadenylase module cannot be guaranteed. Two approaches were employed to investigate this: combined RNAi-mediated depletion of the Cnot7 and

Cnot8 mRNAs (two isoforms which encode the CAF1 nuclease subunit), and small molecule inhibition of CAF1 activity. Since CAF1 bridges the connection between CCR4 and the rest of the complex, and nuclease activity of isolated CCR4 is minimal – at least in vitro - depletion of CAF1 subunits should in theory have abrogated all activity of the affected complexes (28, 513, 523). Inhibition of CAF1 meanwhile, may have had less of an effect on CCR4 activity, and given the far shorter treatment time, should also have allowed less opportunity for buffering by the cell (531, 594).

6.4.1 Cnot7/8 depletion reproduces extended tails and reduced induction observed in Cnot1 knockdown

Combined knockdown of Cnot7 and Cnot8 was very efficient at the mRNA level (figure 6.5 C) and as expected, caused an increase in total poly(A) tail length after 60 and 120 minutes of serum stimulation on all mRNAs tested (figure 6.5 A). These increases were practically identical to those observed in Cnot1 knockdown (figure 4.1 A), consistent with Cnot1 and Cnot7/8 knockdowns having similar efficacy in restricting the complex's deadenylase activity. Furthermore, pre-mRNA levels of the four genes tested were again reduced to a remarkably similar level as in Cnot1 knockdown (figures 6.5 B, 5.1 B). This suggested either that the effects observed in Cnot1 knockdown were due predominantly to a loss in deadenylase activity, or that absence of the CAF1 subunit was sufficient to disrupt the whole complex, or to prevent its usual recruitment. It could have been suggested that transcription was unaffected, and that these patterns instead resulted from stabilisation of the reference gene mRNA, however earlier qPCRs of other housekeeping mRNAs following Cnot1 knockdown suggested this was not the case (figure 5.3).

Alongside one of the Cnot7/8 experiments, a Parn (separate deadenylase) knockdown time course was also performed which achieved a reasonable knockdown but had no effect on the poly(A) tails of the mRNAs assayed (figure 6.5 A, D).

6.4.2 CAF1 inhibition uncouples poly(A) length from mRNA production

To avoid possible adaptation of the cell, a CAF1 (CNOT7/8) inhibitor was used instead. Two compounds were available which, prior to these experiments, had not been used in mammalian cells (594, 595). Pilot experiments were therefore carried out to compare the two candidates and to optimise treatment time and dosage.

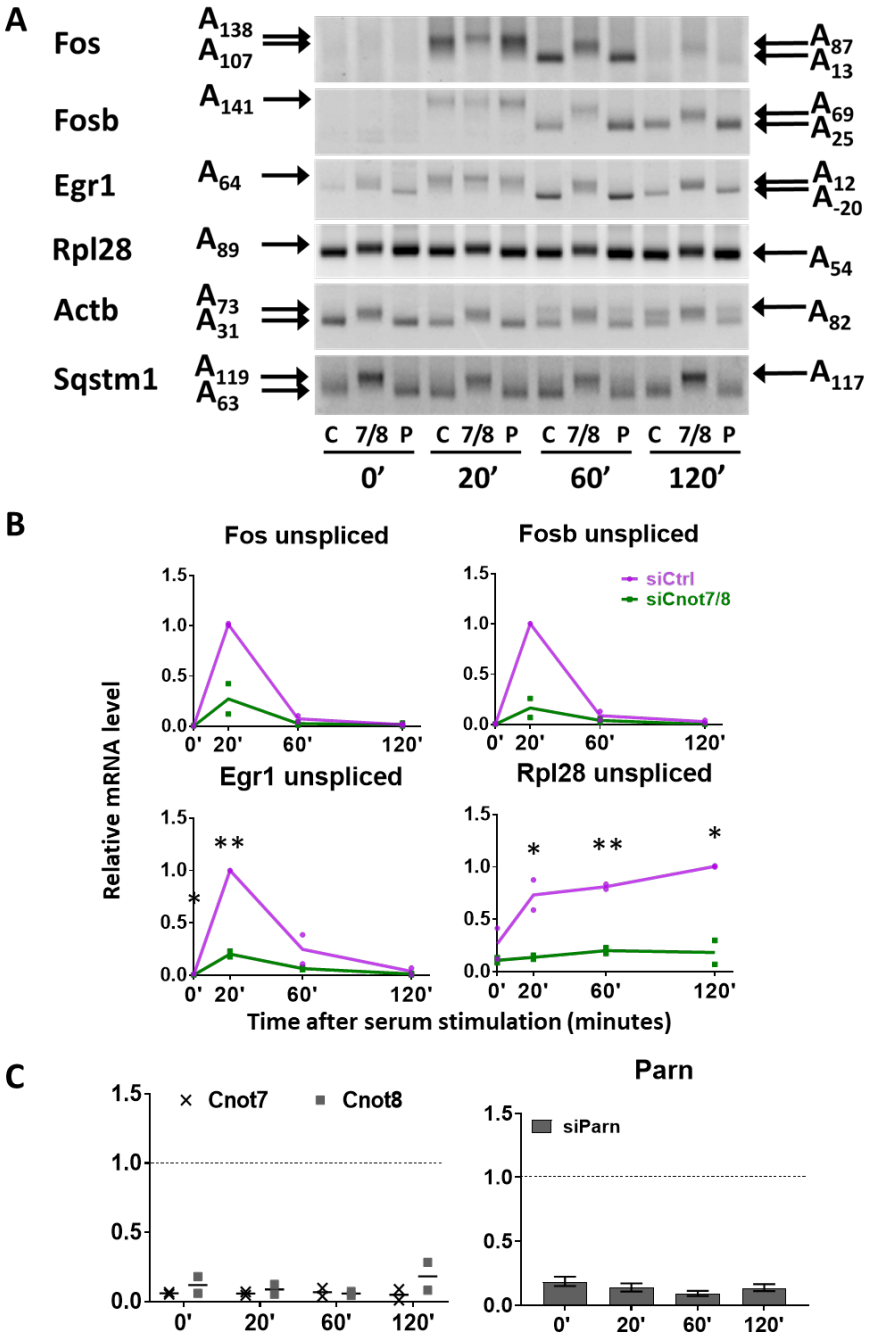


Figure 6.5 Effects of Caf1 subunit (Cnot7/8) knockdown closely resemble that of Cnot1 knockdown. A) RL2-PAT, using agarose gel stained with SYBR safe, showing changes in poly(A) tail lengths over a serum time course in Cnot7/8 and Parn knockdown cells**.** Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. Times shown indicate duration of serum stimulation in minutes. **B)** qPCR data showing mRNA abundance throughout serum induction in combined Cnot7/Cnot8 knockdown, relative to Gapdh and normalised to the maximum value in the control set. Data for 2 biological replicates included. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.05, ** indicates FDR adjusted p <0.01. **C)** Validation of knockdowns by qPCR at the mRNA level, relative to Gapdh. Cnot7/8: 2 biological replicates, Parn: 1 biological replicate.

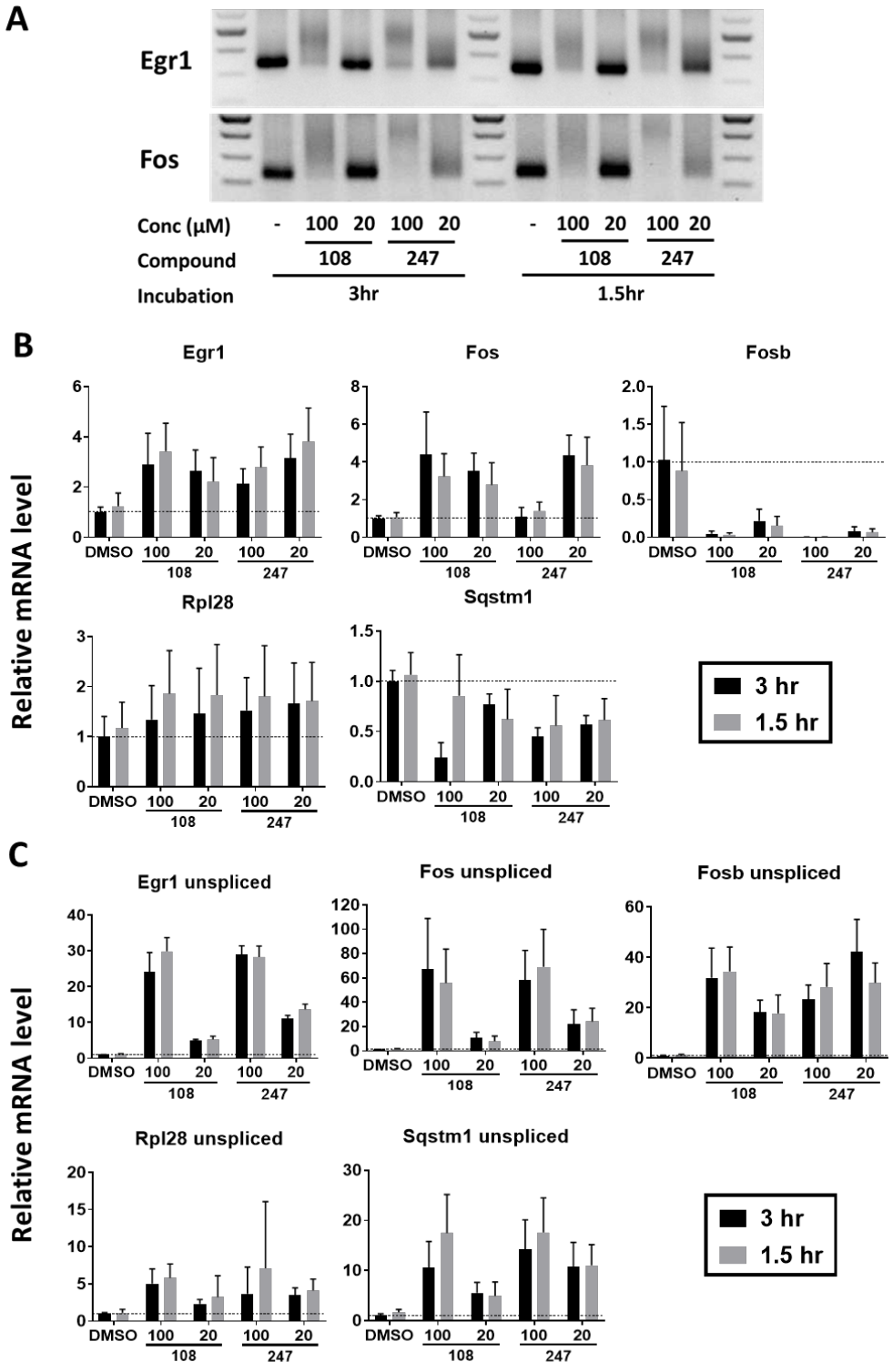


Figure 6.6 Inhibition of CAF1 activity in NIH 3T3 cells by two small molecule inhibitors. **A)** Two inhibitors, compounds 108 and 247 are compared in their ability to inhibit deadenylation of serum response mRNAs. Extent of deadenylation measured using RL2-PAT with agarose gel stained with SYBR safe. Different concentrations (μM) and pre-serum treatment times are indicated. **B, C)** qRT-PCR data showing mRNA abundance at the mature (B) and unspliced (C) levels after inhibitor treatment and 60 minute serum stimulation. Gapdh was used as a reference gene and all levels were normalised to 3 hour DMSO treatment. Error bars indicate standard deviation of 3 technical replicates.

In order to test inhibition efficiency, cells were serum stimulated for 60 minutes and PATs performed for serum response genes; if inhibition of deadenylation was successful – and if CAF1 made a significant contribution to deadenylase activity - long tails rather than short tails should have been present at this time point. Both compounds severely inhibited deadenylation and were able to enter the cell relatively quickly, with no discernible difference between incubating for 1.5 and 3 hours pre-serum stimulation (figure 6.6 A). Although compound 247 had superior activity – both leading to markedly longer tails than 108 and acting at a lower dose – compound 108 was taken forwards as it has been better characterised and published.

Serum stimulation time courses were then carried out following 2 hours of treatment with 100 μM compound 108. Predictably, total poly(A) tail length was longer for induced mRNAs at all time points tested (figure 6.7 A, B). Interestingly, the tails observed here seemed slightly longer than those resulting from Cnot1 or Cnot7/8 knockdown. This perhaps resulted from the inhibitor affecting a larger proportion of complexes or could also have related in some way to the shorter treatment time. There was no apparent effect of inhibitor treatment on the housekeeping mRNAs, likely due to the majority of these stable mRNAs existing before treatment (figure 6.7 A, B).

The patterns in mRNA abundance observed in CAF1 inhibition were very different to those in Cnot1 or Cnot7/8 knockdown (figure 6.7 C, D – compare with 5.1 A,B, 6.5 B). In particular, the mature Egr1 mRNA was more abundant in the treated cells than the control at all time points, and in general the peak of expression seemed to be obviously delayed for all serum-induced mRNAs. Though these mRNAs appeared at first glance to be stabilised, their unspliced levels were also increased, perhaps to a greater degree than their spliced counterparts, though differences in primer efficiency could contribute to the seeming disparity. Highlighting this difference further, mature levels of Fosb mRNA were lower when CAF1 was inhibited, despite unspliced levels increasing. Housekeeping mRNAs seemed largely unaffected by CAF1 inhibition at the mature level - likely due to their high stability – whereas unspliced levels were increased following CAF1 inhibition (figure 6.7 C, D). The data surprisingly therefore indicated that CAF1 (CNOT7/8) inhibition appeared to

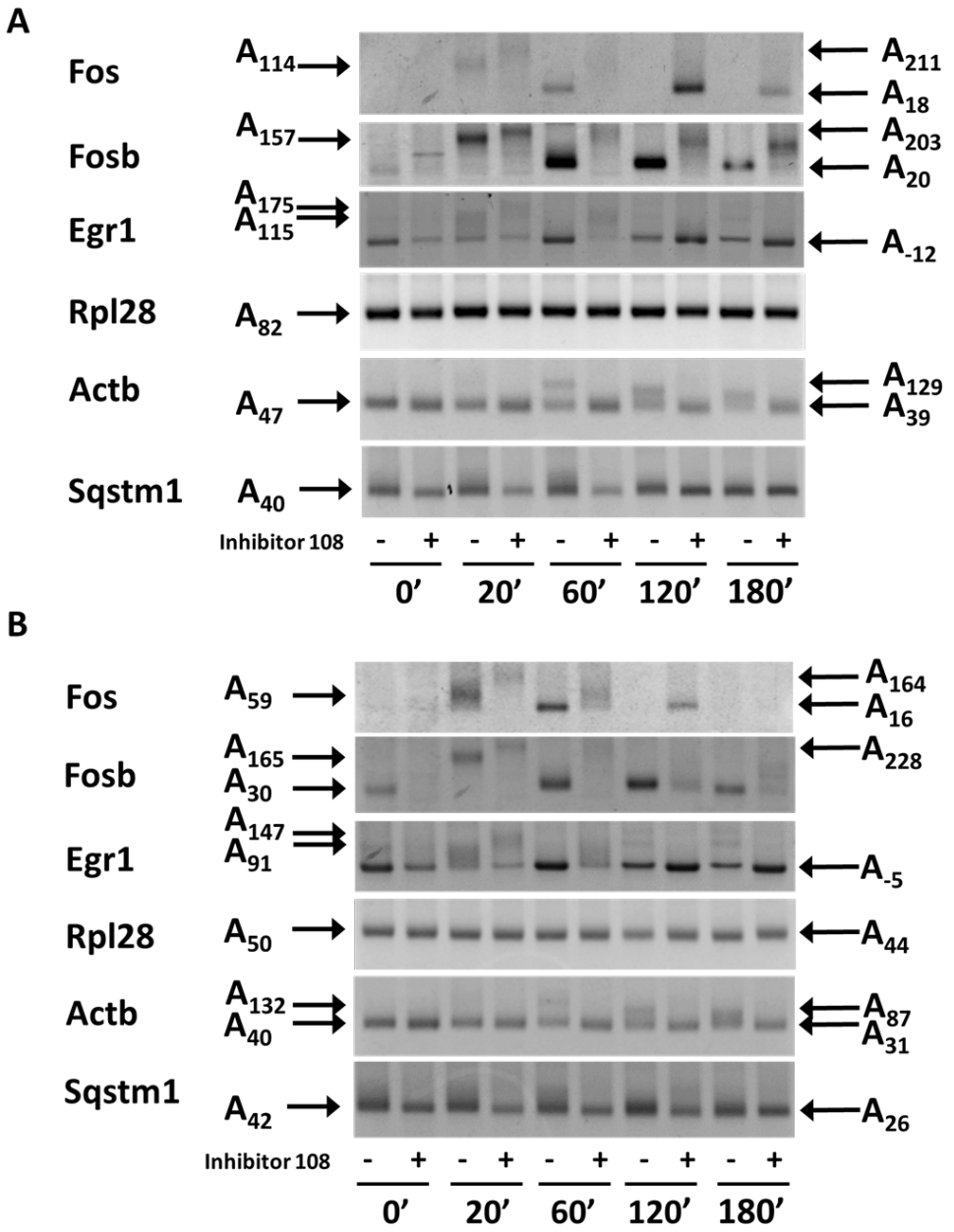


Figure 6.7 Effect of CAF1 inhibition on poly(A) tail length and mRNA abundance throughout the serum response (continued overleaf). **A, B)** Two biological replicates of RL2-PAT, using agarose gel stained with SYBR safe, showing poly(A) lengths in DMSO or inhibitor treated cells. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning. Times shown indicate duration of serum stimulation in minutes. **C, D)** qRT-PCR data showing mRNA abundance at the mature (C) and unspliced (D) levels after inhibitor treatment and stimulation with serum for the times indicated. Gapdh was used as a reference gene and all levels were normalised to the maximum value within the control set for each replicate. 3 biological replicates are shown for all time points except 180’ which only comprises two. t-tests were carried out between control and knockdown values for each time point. * indicates false discovery rate adjusted p-value < 0.05, ** indicates FDR adjusted p <0.01.

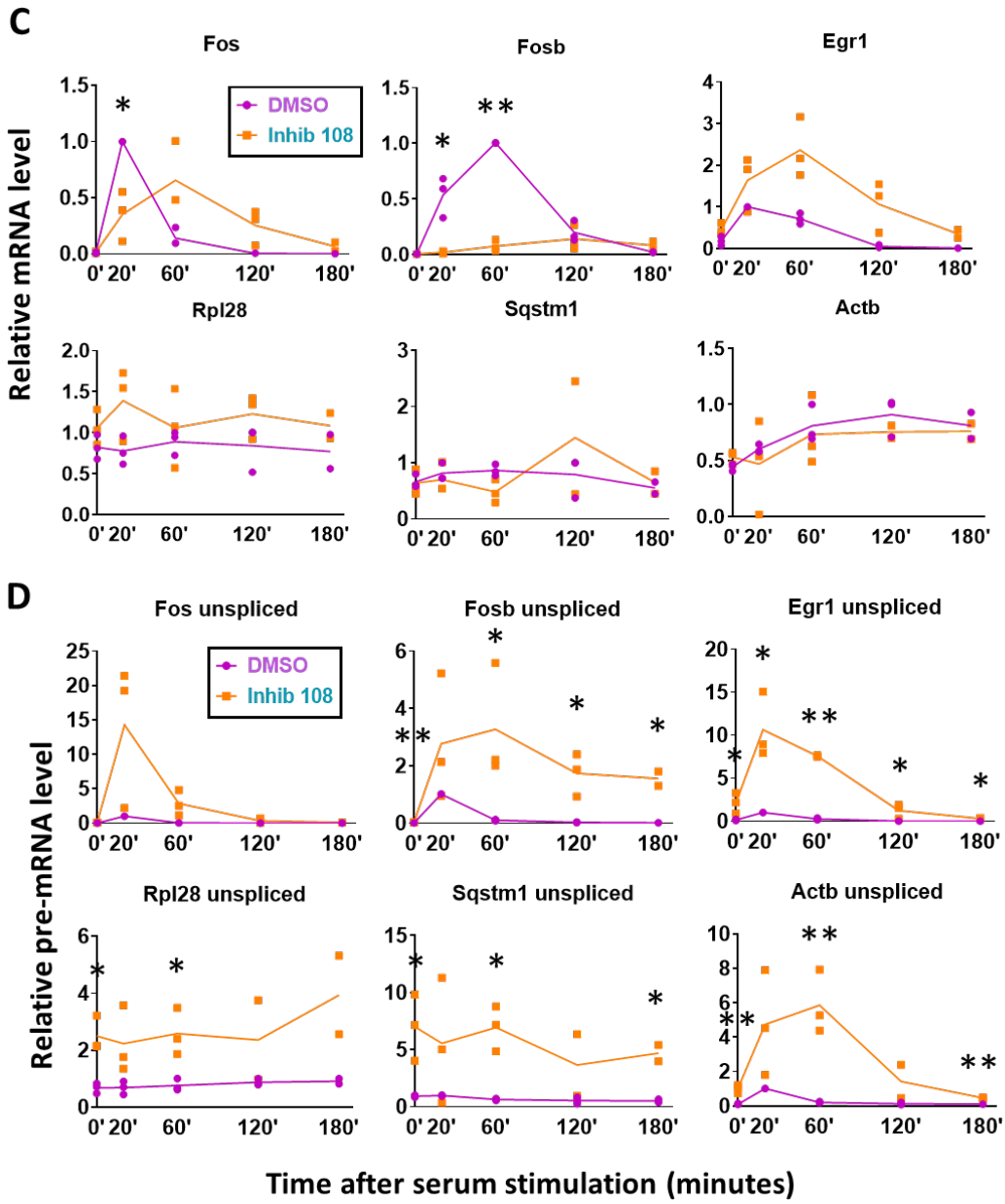


Figure 6.7 contd.

increase transcription, perhaps in combination with delaying splicing. This is in contrast to Caf1 (Cnot7/8) knockdown which reduced transcription.

To check whether this was a true effect of CAF1 inhibition or merely an artefact of the specific inhibitor, compound 108, RNA from the pilot experiment was repurposed. mRNA level after 60 minutes serum stimulation was compared between compounds 108 and 247, using 100 and 20 µM concentrations, and 3 and 1.5 hours treatment pre-stimulation. Strikingly, both inhibitors reproduced the varied effects on mature and pre-mRNA level seen at the 60 minute measurement in the time course (figure 6.6 B, C). Though it was possible that both compounds had

the same off-target effect, these data suggested that inhibition of CAF1 was the more likely explanation. There was no consistent difference in magnitude of effect between the two inhibitors which is surprising given the more pronounced effect of compound 247 on poly(A) tail length (figure 6.6 A). Interestingly, lower inhibitor concentration seemed to be associated with markedly smaller effect size in unspliced but not mature mRNA. The data clearly indicate that the effects on unspliced mRNA resulting from knocking down CCR4-NOT subunits or chemical inhibition of CAF1 are due to distinct effects on the intended targets rather than off-target effects.

6.5 Nuclear PABPN1 levels increase in Cnot1 knockdown

Cnot1 knockdown causes an increase in early poly(A) tail length and one possibility, aside from decreased deadenylation, is that poly(A) tail synthesis is enhanced. This may be particularly relevant for Rpl28 which never seemed to exhibit a long tail of which deadenylation could be inhibited. Changes to poly(A) synthesis – both in Cnot1 knockdown and normally over the serum response - could occur as a result of altered PAP activity; this could be achieved either through modification of PAP or through availability of PABPN1 or CPSF which are required for efficient poly(A) addition (52, 56, 179, 363). Polyadenylation factors were previously shown to become more nuclear following LPS stimulation and it seemed possible that a similar pattern could exist for stimulation with serum (354).

Nuclear poly(A) binding protein (PABPN1) promotes initial poly(A) tail synthesis by poly(A) polymerase (PAP) through its PAP-stimulating domain (55, 56, 321). PABPN1 stimulation of PAP is also required for targeting of certain transcripts for degradation by the nuclear RNA exosome (231, 321, 322). Given that Cnot1 knockdown is accompanied by longer poly(A) tails and lower levels of pre-mRNA (figures 4.2, 4.3, 5.1), changes to PABPN1 localisation could be well-placed to explain these effects. Localisation of PABPN1 was therefore investigated in Cnot1 knockdown using immunofluorescence with confocal microscopy. Knockdown and control cells were fixed at 0, 20 and 60 minutes following serum induction to also investigate the temporal changes in poly(A) tail length observed under normal circumstances. Nuclear enrichment of PABPN1 was enhanced in Cnot1 knockdown and appeared to form small granules (figure 6.8 A-D, F). It is not clear whether this

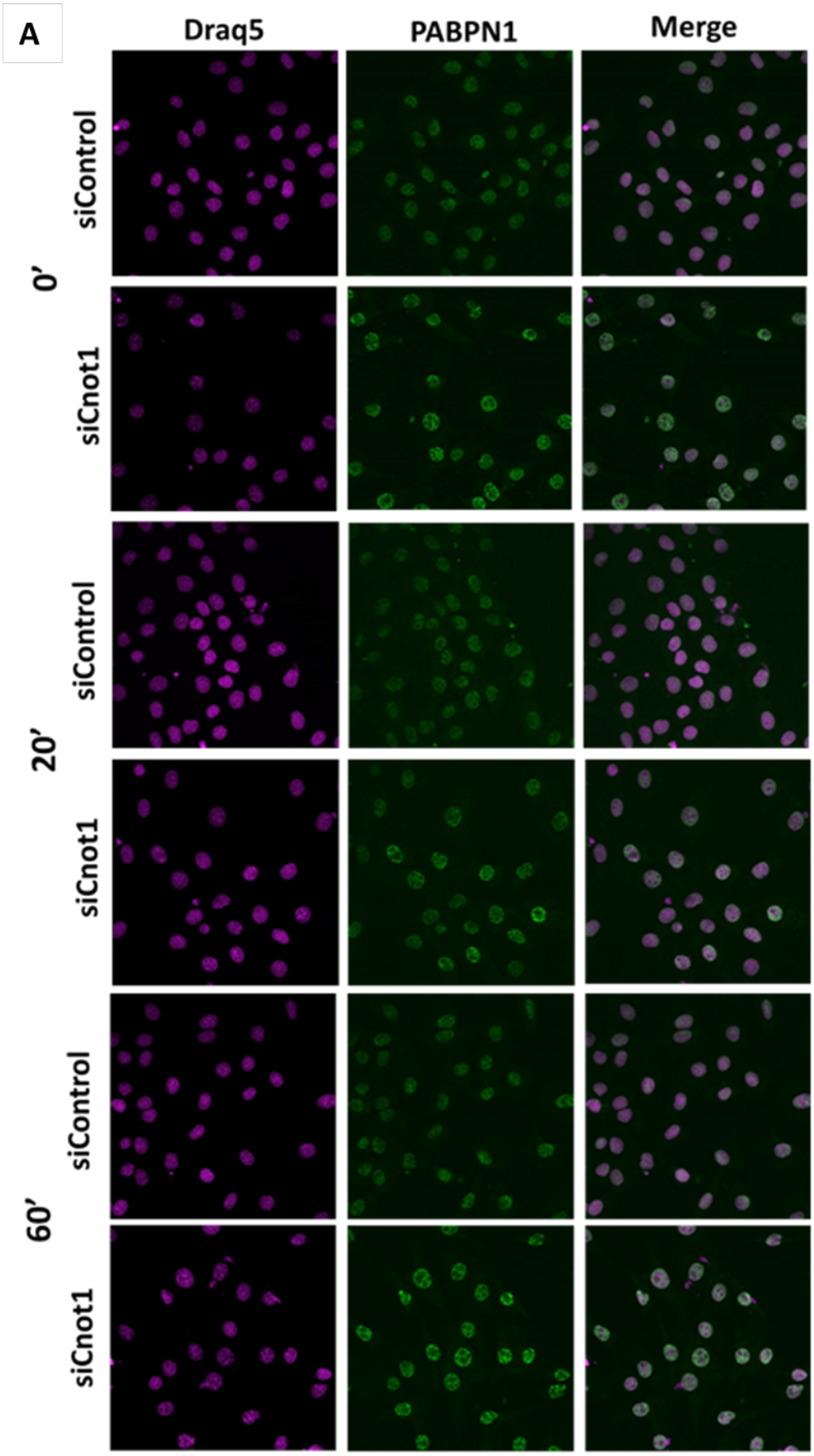


Figure 6.8 PABPN1 becomes more nuclear in Cnot1 knockdown but its localisation does not change over the serum response (continued overleaf).

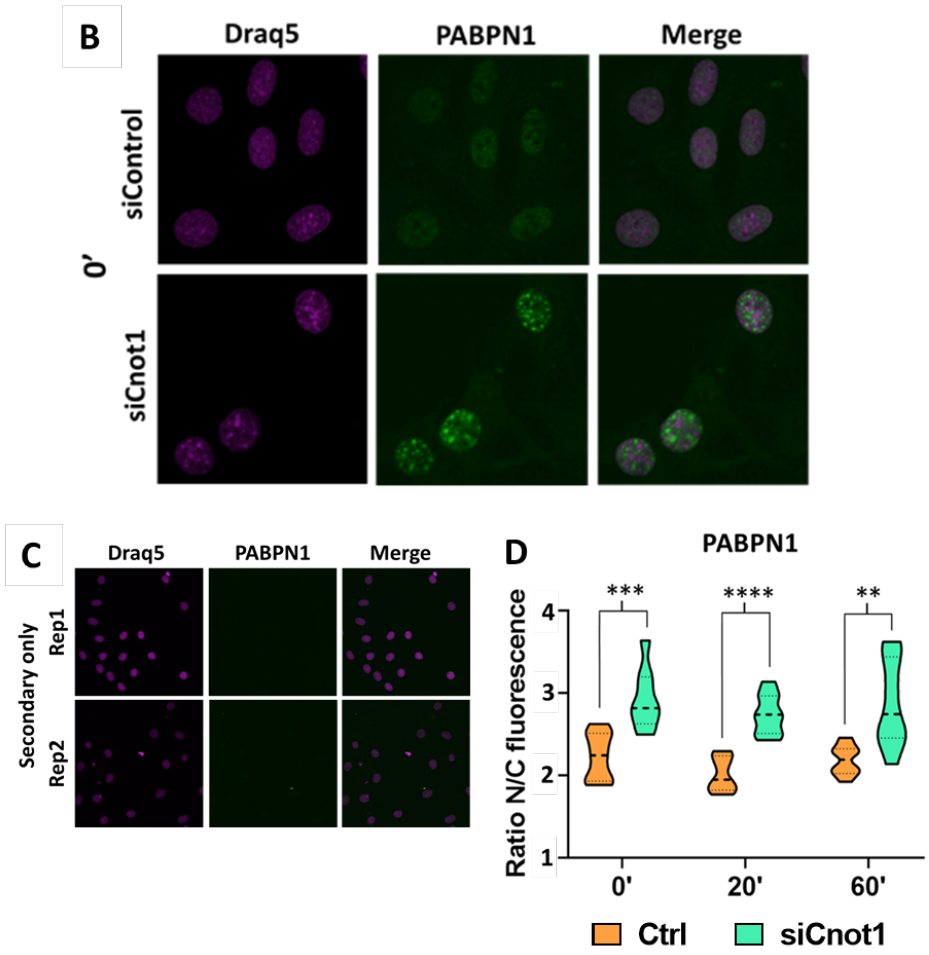


Figure 6.8 PABPN1 becomes more nuclear in Cnot1 knockdown but its localisation does not change over the serum response. A) Representative lower magnification confocal images of cells from the same biological replicate, stained with anti-PABPN1 and Draq5 (nuclei). **B)** Representative higher magnification image from the replicate shown in A. **C)** Secondary only controls for both replicates. **D)** Analysis of nuclear:cytoplasmic PABPN1 ratio based on fluorescence for two biological replicates. All lower magnification images were used. Stars indicate significance level of t-tests. ** $P < 0.005$, *** $p < 0.001$, **** $p < 0.0001$

was a direct effect of Cnot1 knockdown or whether PABPN1 may have been retained in the nucleus by hyperadenylated transcripts resulting from impaired deadenylation. Although specificity of the PABPN1 antibody was not validated by knockdown here, it has been used for immunofluorescence in several reputable publications (231, 431, 444, 630, 631).

6.6 Discussion

Initial poly(A) tails of induced mRNAs are regulated throughout the serum response in NIH 3T3 cells, while those of housekeeping mRNAs remain consistent. For most

genes, long poly(A) tails (> ~150 nt) are present in at least one cellular fraction under control conditions, loosely consistent with early reports of 200 nt tails, whereas long poly(A) tails could not be detected for Rpl28 (figure 3.5 A, D). Cnot1 knockdown caused increased nuclear or chromatin-associated mRNA poly(A) tails for all mRNAs tested, as well as reduced relative pre-mRNA level for many genes (figures 4.3 A, B, 5.1 B) ; this was suggestive of a coupling between early poly(A) length control and mRNA production which could act to maintain mRNA homeostasis. A link between elongation rate and mRNA stability was recently suggested by Slobodin et al, proposed to be regulated through degree of m^6A modification and poly(A) status (58). While this is an exciting idea, a more detailed appraisal of smaller gene groups is required, as well as greater precision in poly(A) measurement. In an attempt to understand this coupling, along with normal gene-specific and temporal regulation, several lines of enquiry were pursued.

Given the existence of both long and short chromatin-associated poly(A) tails, it was not possible to distinguish between populations of mRNAs which had yet to receive a poly(A) tail and those whose tails had been completely removed. Rapid chromatin-associated deadenylation by CCR4-NOT and inhibition of synthesis by PAP therefore both remained as possibilities.

Western blotting of nuclear and cytoplasmic fractions demonstrated presence of CNOT1 in both compartments, consistent with its roles in cytoplasmic deadenylation and stimulation of Pol II elongation. This also fits with the earlier observation of poly(A) shortening between the chromatin-associated and nuclear fractions for some mRNAs (figure 4.3 A, B). Localisation did not change throughout the serum response, indicating that nuclear availability of CNOT1 is unlikely to explain the usual temporal differences in poly(A) length (figure 6.1 C, D).

CCR4-NOT can be targeted to specific transcripts, a property which could be relevant to the gene specific differences in normal poly(A) length and magnitude of response to Cnot1 knockdown (29, 30, 45, 254, 255). As the serum response mRNAs tested here contain AU-rich elements (AREs) in their 3'UTRs, mRNAs encoding the TTP family of RNA binding proteins (Zfp36, Zfp36L1, Zfp36L2) were knocked down, but only minimal effects on poly(A) length were observed (figure 6.3). This either suggests that targeting of CCR4-NOT to these mRNAs by TTP does not play a role in

early poly(A) length control, or that knockdown was insufficient to cause detectable effects.

Another possibility was that tail length is determined by nuclear dwell time, and therefore the period of susceptibility to nuclear deadenylation by CCR4-NOT. Depletion of the main mRNA export factor, NXF1 had little effect on the poly(A) tails of constitutively expressed mRNAs and resulted in the appearance of longer poly(A) tails on serum induced mRNAs (figure 6.4 A, B), the opposite of what would have been expected if longer nuclear dwell time allowed more extensive nuclear deadenylation. In line with the longer poly(A) tails, qPCR data also seemed to show enhanced stability of the two serum-induced mRNAs assayed (figure 6.4 C).

If Nxf1 knockdown was successful in restricting mRNA export, it is possible that rather than allowing more extensive trimming of the poly(A) tail, nuclear retention allowed PAP to add a longer tail. In such a model, the availability of export factors and consequent speed of export could determine how soon after cleavage an mRNA is exported, and therefore what length tail it has. Equally, since NXF1 acts late in the mRNA export pathway (5, 210), its knockdown could have allowed mRNAs to escape from early association with CCR4-NOT, while still preventing their export to the cytoplasm and consequent cytoplasmic deadenylation. Were the experiment repeated, it would perhaps be pertinent to intervene at an earlier stage of export – for example through knockdown of a TREX component – so that the RNA is not commandeered from the transcription complex and rescued from any chromatin-associated deadenylation.

Alternatively, NXF1 may not be essential for export of all mRNAs, perhaps explaining why poly(A) tails of serum-induced but not housekeeping mRNAs were affected (628, 629). In a recent study, those mRNAs which were AU-rich or contained few exons showed greater dependence on NXF1 for export, suggesting that the greater effects observed here on the serum response mRNAs - which contain fewer exons than average – may be due to their heightened dependence on this export pathway (236). If it is the case that housekeeping transcripts are less susceptible to cytoplasmic deadenylation however, the absence of effect on these mRNAs could also have been explained by the idea that retention in the nucleus only caused protection from negligible cytoplasmic deadenylation. Another possibility is that

slower export led to a higher ratio of RNA to some factor which normally limits tail length of induced mRNAs (for example by promoting deadenylation), saturating this factor and resulting in longer poly(A) tails. Similarly, if long poly(A) tails are usually attained through protection by some factor which blocks deadenylation, inhibition of export could have allowed continued presence of this factor in the nucleus and consequent emergence of abnormally long poly(A) tails.

CNOT1 is the central scaffold of the CCR4-NOT complex and its depletion likely resulted in concomitant loss of the whole complex. To establish whether coupling of early poly(A) length and pre-mRNA level are achieved through the multifunctionality of CCR4-NOT, or through a more general sensing of poly(A) status or deadenylation rate, effects of knocking down and inhibiting the CAF1 (CNOT7/CNOT8) subunit were compared. While both treatments caused longer poly(A) tails on serum-induced mRNAs, only the knockdown affected those of housekeeping mRNAs (figures 6.5 A, 6.7 A, B). Cnot7/8 knockdown had a similar dampening effect to Cnot1 depletion on pre-mRNA level, whereas CAF1 inhibition resulted in large increases (figures 6.5 B, 6.7 D). These increases in unspliced mRNA level were accompanied by delayed peaks in mature mRNA levels for serum-induced mRNAs, which varied in height compared to the control (figure 6.7 C). Unaffected transcription in combination with delayed splicing and enhanced stability could perhaps cause such effects, though longer time courses would be required for some genes, and reports of CCR4-NOT involvement with splicing are not widespread (616). Interestingly, increases in unspliced level of housekeeping mRNAs were also observed, though these seemed to only minimally affect mature mRNA level, if at all (figure 6.7 C, D). This may be partially have been due to the very high ratio of mature:nascent mRNA for these genes making short term changes undetectable, or could have resulted from nuclear degradation aided by possession of longer tails or slowed splicing.

In keeping with its lack of stimulation (and inhibition at higher concentrations) by Pab1 (PABPC), a recent report suggested that the Caf1(Cnot7/8) nuclease has a preference for poorly translated mRNAs whereas Ccr4 is not selective (28, 50). Since CCR4 attaches to the complex via CAF1, knockdown of Cnot7 and Cnot8 should in theory remove both CAF1 and CCR4 activity whereas small molecule inhibition

selectively inactivates CAF1 (516, 523, 594). Hence, differences in effect on housekeeping mRNA poly(A) tails by the two treatments could have resulted from the continued presence of CCR4 nuclease activity in inhibitor treated cells (though both activities may be required for successful deadenylation (531)). It seems equally likely however, that changes taking place over 2 rather than 72 hours did not manifest as detectable differences for such stable transcripts. It is of course not possible to compare treatment times with any precision since the inhibitor works directly on the assembled complex, whereas siRNA treatment requires changes at the mRNA level to filter through.

Considering pre-mRNA level, three explanations seem possible for the differences between treatments. One is that CCR4 but not CAF1 could have a role in promoting transcription elongation and its presence enabled normal rates of transcription which were not balanced by CAF1 mediated deadenylation in the presence of the inhibitor. The second is that loss of CAF1 (CNOT7/8), but not its inhibition, destabilises the whole complex resulting in reduction of both deadenylase and transcription-stimulatory activities (which may be mediated by other CCR4-NOT modules). A third explanation is that differences in treatment time play a major role, with some feedback loop not yet in play after 2 hours of inhibitor treatment. Such a feedback loop could involve initial upregulation of transcription – perhaps through stabilisation of transcripts encoding transcription factors – followed by a compensatory decrease (eg. through consequent production of a transcriptional repressor). On the other hand, transcriptome-wide longer poly(A) tails could result in upregulation of CCR4-NOT at the protein level through enhanced mRNA stability – a similar idea to that suggested by Slobodin et. al (58), but this would in theory increase transcription further. Extended treatments with the inhibitor could perhaps shed further light on this point.

Data from the pilot experiment showed that pre-mRNA levels after 60 minutes of serum stimulation were affected to a much greater magnitude when a higher concentration of either inhibitor was used (figure 6.6 B, C). This may suggest that the inhibitor preferentially binds cytoplasmic CAF1 - affecting mature mRNA levels - and that higher concentrations are required to cause binding of nuclear CAF1 which

could mediate effects on pre-mRNA. If true, the use of different concentrations could prove a valuable investigative tool.

One curious feature is the appearance of long tails on Actb late in the serum response in the control cells, but not in those treated with the inhibitor (figure 6.7 A, B). These dual bands were also present in control sets of the knockdown experiments, but in the knockdowns themselves, either a long or an intermediate length tail is visible (figures 6.5 A, 4.3 A, B). Since Actb is transcriptionally induced by serum (figure 5.1 B), the longer band in the control cells likely resulted from new transcription, which may have been inhibited in Cnot1/Cnot7/8 knockdown and CAF1 (CNOT7/8) inhibition. Following this line of reasoning, the single band in the treated cells probably represented the constitutively expressed medium-tailed product. Variation in length between knockdown and inhibition could have resulted from the differing treatment times, meaning that different proportions of the mRNA pool had been produced since deadenylation was inhibited.

Taken together, these varied lines of enquiry indicate a global shift in the molecular landscape of the cell following Cnot1 knockdown, but do not clearly reveal the driving mechanism. More immediate depletion of individual subunits, for example via the Auxin inducible degron (AID) system, may help shed light on the early effects of deadenylase depletion. In addition, metabolic labelling and/or Pol II RNA immunoprecipitation, perhaps combined with depletion of nuclear RNA turnover machinery, would be important in delineating changes in transcription from altered splicing or nuclear degradation rate. In light of the seemingly pro-homeostasis effects of CCR4-NOT subunit knockdowns, the opposing transcriptional effects of CAF1 inhibition which indicate disruption of homeostasis may provide clues as to the underlying mechanism. Experiments to determine the abundance, localisation and composition of CCR4-NOT complexes following each treatment could be informative. Overall, these data point towards a role for CCR4-NOT as a downstream effector of mRNA metabolism which, according to the literature, may be titrated according to upstream signals as part of a wider cellular agenda (58, 60, 61).

7 PAT-Quant Seq: Poly(A) tail deep sequencing

7.1 Introduction

Recent increases in the availability of next generation sequencing technologies have paved the way for development of countless specialised methodologies, with more continuing to emerge. These techniques typically involve isolation of the genetic material of interest via a selection technique such as immunoprecipitation or metabolic labelling, followed by reverse transcription, where relevant, and sequencing on an established platform (11, 13, 120, 325, 326, 432, 433, 632–636). These platforms vary in their sample requirements and outputs for RNA sequencing, with some requiring reverse transcription and library amplification by PCR (Illumina, PacBio) and others able to sequence RNA directly (Oxford Nanopore) (637–639). Technologies also differ in their ability to capture long reads and their base-calling reliability, with Oxford Nanopore also able to detect base modifications (433, 636, 640–645). In general, Illumina short read sequencing offers superior sequencing quality (< 0.5 % overall error rate, dropping to < 0.1 % for quality scores > 30) than either PacBio or Oxford Nanopore (642–645).

Optimal sequencing reliability and quality is not always possible within financial constraints and sample abundance/purity. While financial limits are hard to mitigate, innovative approaches can address the problem of low sample yield. SLAM-Seq for example, uses an elegant orthogonal chemistry approach to communicate 4-thiouridine labelling of nascent RNA. Rather than cross linking thiouridine with biotin and losing material to inefficient streptavidin bead purification, the addition of a thiol-reactive compound results in alkylation of the thiol groups on the labelled RNA. These thiol groups cause T > C conversions upon reverse transcription which can then be detected bioinformatically (635).

Several groups have developed poly(A) deep-sequencing methods, largely around the Illumina sequencing by synthesis platform (11, 325, 432, 433, 435, 582, 632, 633, 636, 646, 647). Most methods use the poly(A) tail as selection criteria for generating 3’-enriched libraries (11, 433, 582, 632, 633, 636); while this circumvents some inefficient purification steps, it does restrict the investigation to transcripts with a minimum number of terminal adenosines. TAIL-Seq on the other hand, is able to detect non-poly(A) termini but requires a much greater RNA input (325). While

FLAM-Seq only captures poly(A) – terminal transcripts, its use of single molecule long read sequencing offered by PacBio enables the non-ambiguous linking of poly(A) lengths with individual messages (433). This enables higher resolution analysis, for example by looking at different mRNA isoforms which may be challenging to crystallize out of short read data, but sequencing at a depth sufficient to construct poly(A) profiles of rarer mRNAs is currently very expensive. Oxford Nanopore also allows for sequencing of full length mRNAs, meaning that no technical limit determines maximum detectable poly(A) length and that transcripts can be cap-selected, enabling retrieval of mRNAs while avoiding biases caused by poly(A) selection or inefficient ribodepletion techniques (636, 646, 647). Although Oxford Nanopore is able to directly sequence RNA - and thus avoids potential PCR-based artefacts - it requires a large amount of starting material (~100 µg) which is not easily attainable for some experiments (636).

Conventional base calling approaches are inaccurate for long homopolymers such as the poly(A) tail due to the merging of multiple identical fluorescence signals. PAL-Seq overcomes this by incorporating biotin-dUTP during sequencing (which should label sequencing clusters proportionally to poly(A) tail length), then adding fluorescent streptavidin after sequencing the poly(A)-proximal region (11). Although this removes the need to base call the tail, PAL-Seq requires modifications to the Illumina sequencer in a non-approved manner which leads to a loss of factory guarantees. TED-Seq also avoids direct base calling of the poly(A) region and instead calculates poly(A) tail lengths by subtracting sequenced 3'UTR and adapter length from the total (432). The tailseeker program developed by Hyeshik Chang remedies inaccurate calling of homopolymers with a machine learning approach, taking the raw fluorescence files from the sequencer as input (325). In FLAM-Seq on the other hand, repeated cycling of the polymerase over a single molecule during sequencing in theory gives higher confidence in base calling – including over long homopolymers. This means that the algorithm employed by FLAM-Seq for identifying poly(A) tails (and determining their length) is based on the more simple matter of setting a threshold for the acceptable number of non-T nucleotides within the putative poly(A) sequence (648). Reliability of Oxford Nanopore base calling also suffers over homopolymeric stretches, so the tailfindr algorithm instead calculates poly(A) length based on identifying boundaries of possible poly(A) stretches and

using read-specific translocation times of nucleotides through the pore to normalise these (647).

Most global poly(A) measures rely on cDNA synthesis and subsequent PCR which could in theory distort true poly(A) lengths through repeated dissociation and inaccurate re-joining of the polymerase during the homopolymeric stretch. The TAIL-Seq protocol includes synthetic spike ins of known poly(A) tail length which show minimal effect of repeated amplification, though due to limitations on synthesis, the longest spike in is 118 adenosines, half the purported tail length on mammalian mRNAs (325). The problem of repeat rounds of PCR can be negated by use of a direct RNA sequencing method such as Oxford Nanopore for which poly(A) length measurement software has already been developed (636, 647). Oxford Nanopore is only recently becoming widely accessible however, and poly(A) measurement has still only been tested with spike ins up to an input length of 150 adenosines (A_{150}). Notably, while measured modal poly(A) length matched the spike ins well between inputs of A_{30} and A_{60}, modal tail length was overestimated by 110 % for spike ins with 10 adenosines and underestimated by about 10 % for A_{100} and A_{150}.

Here we present the TAIL-Seq-based poly(A) deep sequencing technique, Pat-Quant-Seq (PQ-Seq). PQ-Seq generates a similar library to mTAIL-Seq (13), but the 3’ anchor through which polyadenylated transcripts are tagged and selected for is added by extension rather than ligation, as in several other existing poly(A) sequencing protocols (11, 325, 633, 636). Although PAT-Seq also uses Klenow extension to add the anchor and requires only 1 µg starting RNA, it cannot detect tails above 80 adenosines which make it unsuitable for use on the mammalian transcriptome (582). The PQ-Seq biotinylated anchor enables subsequent isolation of polyadenylated transcripts and removes the need for ribodepletion.

Captured RNA is then fed into the Lexogen QuantSeq pipeline for Illumina-compatible library generation. Rather than adding the 5’ sequencing adapter by ligation which is generally inefficient, QuantSeq introduces it during the random-primed second strand synthesis step. The higher efficiency steps included here permit acquisition of global mammalian poly(A) length information from expensive

or low yield experiments, requiring only ~1 µg RNA rather than 80 µg as in TAIL-Seq or 100 µg in Nanopore (325).

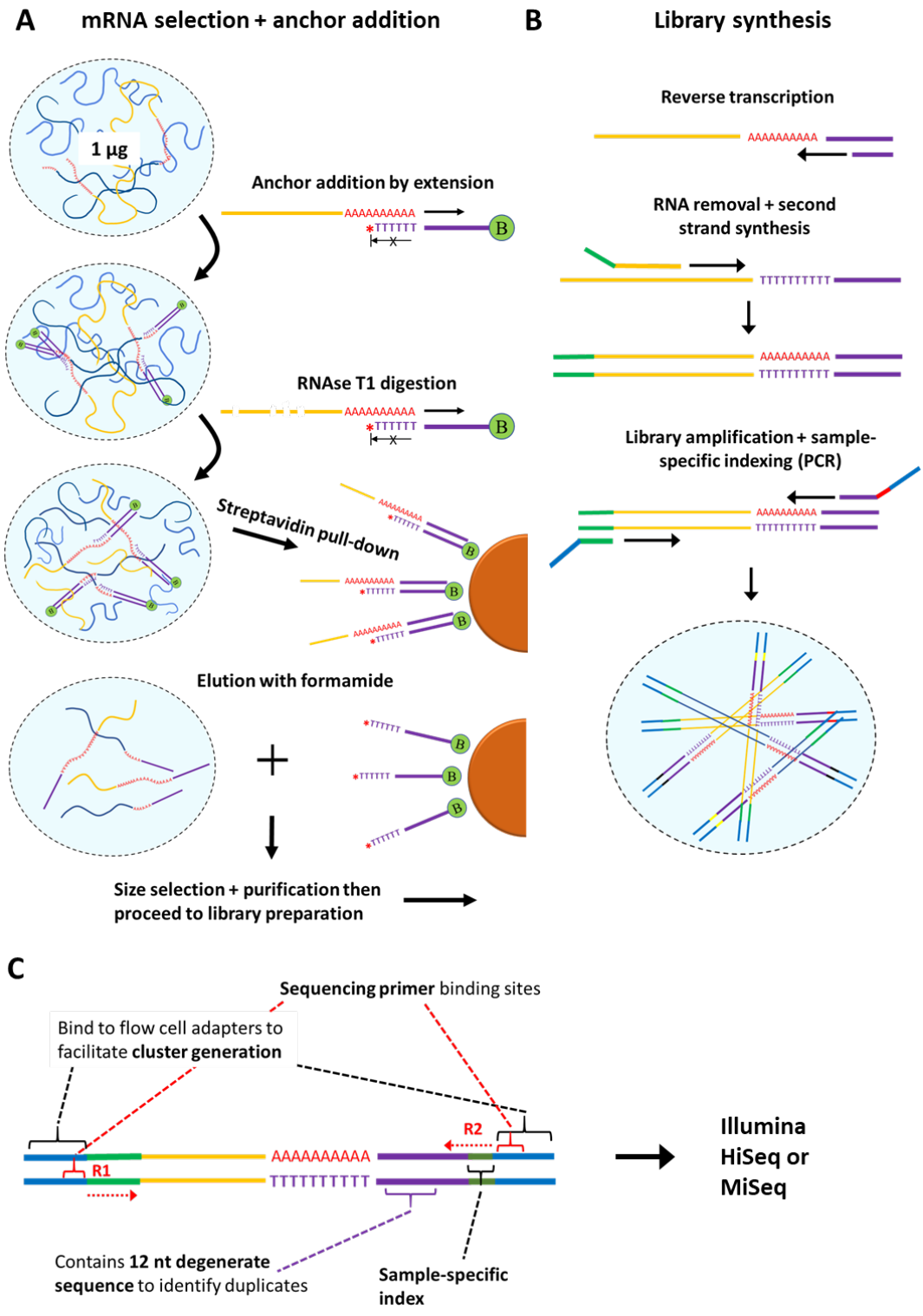


Figure 7.1 Schematic of PAT-Quant-Seq (PQ-Seq). Legend overleaf.

7.2 PQ-Seq workflow/pipeline

PQ-Seq used as little as 1 μg DNase-treated starting RNA, and an overview of the process is outlined in figure 7.1. In this, the fourth optimisation of the protocol, a biotinylated anchor template was added which contained 18 T residues in order to specifically isolate polyadenylated 3' ends, but not introduce a bias towards long tails. The anchor itself was then added by Klenow extension which could proceed only along the anchor template and not into the gene body since the 3' end of the template was blocked. The Klenow fragment used was negative for 3' - 5' exonuclease activity (exo-) to avoid inadvertent degradation of the RNA. To generate 3' end fragments of a suitable size for RNA-Seq, RNA was fragmented by RNase T1 digestion after anchor extension then incubated with streptavidin beads to capture biotinylated fragments (i.e. polyadenylated 3' ends). RNA was eluted from the beads using formamide and heat to denature the 3' duplex, then separated by a denaturing acrylamide gel electrophoresis. Fragments of over 200 nt were isolated.

Library preparation was carried out using the QuantSeq Flex kit. At the reverse transcription (RT) step, a small amount of Klenow (exo-) was added to the reaction as it was not known whether the QuantSeq reverse transcriptase could efficiently use the DNA in the anchor region as a template. After RT, the RNA was degraded, and a second DNA strand synthesised using random primers with attached Illumina linker sequences. These linker sequences enabled introduction of Illumina adapter sequences, and optional sample-specific 5' indexes (not used in PQ-Seq) during PCR. During PCR amplification of the PQ-Seq library cDNA, index sequences were instead added to the 3' end, between the anchor and Illumina adapter sequences, by use of

---

Figure 7.1 (previous page) **Schematic of PAT-Quant-Seq (PQ-Seq). A)** Polyadenylated transcripts are selected by an 18 nt oligo(dT) stretch of the biotinylated anchor template. Klenow (exo-) fragment synthesises the anchor. The template is blocked at its 3' end to prevent extension in the other direction. Transcripts bound by the anchor template are isolated using streptavidin beads, and the RNA eluted by melting in the presence of formamide. **B)** Anchor-tagged RNA is used as a template for Quant-Seq library synthesis. Sample specific index sequences are added at the 3' end during library amplification. **C)** Illumina sequences at 3' and 5' ends mediate annealing to flow cell and subsequent cluster generation, as well as providing priming sites for sequencing by synthesis. Random sequence in anchor allows detection of duplicates by Tailseeker software.

sample-specific reverse primers. 3' index sequences were also added via PCR to DNA spike ins containing poly(A) tails of different lengths (0, 16, 32, 64, 128). The resulting libraries underwent quality control on a Bioanalyzer (to check fragment size distribution and ensure minimal contamination with adapter dimers) and qPCR (to ensure adequate inclusion of Illumina priming sites within the sample). The resulting libraries were sequenced using an Illumina MiSeq.

Tail length was ascertained using the tailseeker software (649) which uses the fluorescence images outputted by the sequencer as a set of .cif files, rather than the fastq files also generated by inbuilt base-calling software. At the time of processing there were no other known pipelines for processing Illumina generated reads, and tailseeker had been successfully employed by other groups (13, 325, 326). From these, images 1-51 comprised read 1 and were used to map the fragment to the genome. Images 52-57 recorded the read 2 index sequence which was added during the PCR step of the library preparation and identified which sample the fragment originated from. Image 58 onwards (to 308) began with the 12 degenerate bases (which aided deduplication) followed by a 5 nt delimiter sequence which heralded entry to the poly(A) tail (read as poly(T)). This gave a theoretical detection limit of 234 adenosines.

Analysis using tailseeker is described in the supplementary material of the original TAIL-Seq paper (325). In brief, optional base-calling with the third-party software AYB was first carried out, which according to the tailseeker documentation, gives improved alignment of read 1 (5' end of fragment) to the genome. Read 1 was then filtered for common contaminants, and these clusters were removed from further analysis. Duplicates were then identified by checking for identical sequences in a representative section of read 1, and in the read 2 degenerate sequence, resulting in removal of all but the highest quality cluster. Degenerate sequences with low diversity were also removed at this point since this diversity is required to calibrate signal normalisation in the TAIL-Seq algorithm. Read 2 was then clipped of the degenerate sequence and 5 nt delimiter, leaving read 2 in theory beginning with poly(T).

Several strategies for determining poly(A) length are presented in Figure S1 of the TAIL-Seq supplementary data (325). The simplest and most restrictive technique ('Strategy I') is to count the number of residues called as 'T' in a row, however this would only account for perfect A stretches, which may not be the norm (433, 632). Strategy II allows up to around 10 % non-A bases and does not require a 3' terminal A. It does this by scoring possible start and end pairs, using the sum of the weighted value for each base in the interval (T = 1, N = -5 and A, C and G = -10). This Strategy II length call is fed into the TAIL-Seq algorithm where it is either accepted if it is ≤ 8 or undergoes further processing if it is > 8. In the current spot (i.e. location on the flow cell), a normalisation factor is calculated for each channel (i.e. A, C, G or T) based on signal throughout the degenerate region. The original signal for each base is divided by the normalisation factor for the spot to give normalised signal intensity for each position in the read 2 sequence. The relative T signal in each position is then calculated by dividing normalised T signal by the sum of normalised signal for the other bases. Relative T signal is assessed from the poly(A) start (obtained using Strategy II) until the end of read 2. A machine learning algorithm is trained on the expected distribution of relative T signal using poly(A) spike ins of different lengths. Finally, poly(A) length is called as the sum of the regions defined as 'poly(A) body' or 'poly(A) transitive' (rather then 3'UTR transitive or 3' UTR body) (325).

7.3 Preliminary results

Two NIH 3T3 datasets were trialled with the iteration of the PQ-Seq protocol described above. The first experiment compared nuclear and cytoplasmic fractions of cells at steady state, and the other at 3 time points (0, 20 and 60 minutes) following serum stimulation. For both datasets, tailseeker output was generated by Daniel Zadik (DeepSeq, University of Nottingham). Quality control data are detailed in table 7.1. For both experiments, more reads were obtained for the cytoplasmic fraction, though this varied between 1.38-fold greater in the steady state experiment, and 6-10 fold greater in the serum stimulation experiment. Between 61 and 90 % of the reads were mappable. Of those mapped, up to 59 % were non-mRNA species (0' stimulated nuclear fraction), though for the cytoplasmic fractions and steady state nuclear fraction, non-mRNA reads comprised only 36 % of those which were successfully mapped.

		No. fragments post-filtering	% reads mapped	% non-mRNA mappable reads
Steady state				
Nuc		1382571	82.55	36.11
Cyto		1911996	73.09	36.68
Serum stimulated				
Nuc	0'	225849	85.75	59.07
	15'	278064	89.17	52.95
	60'	408531	71.3	53.71
Cyto	0'	1373258	61.95	43.31
	15'	2697545	76.48	36.11
	60'	2457225	64.9	36.68

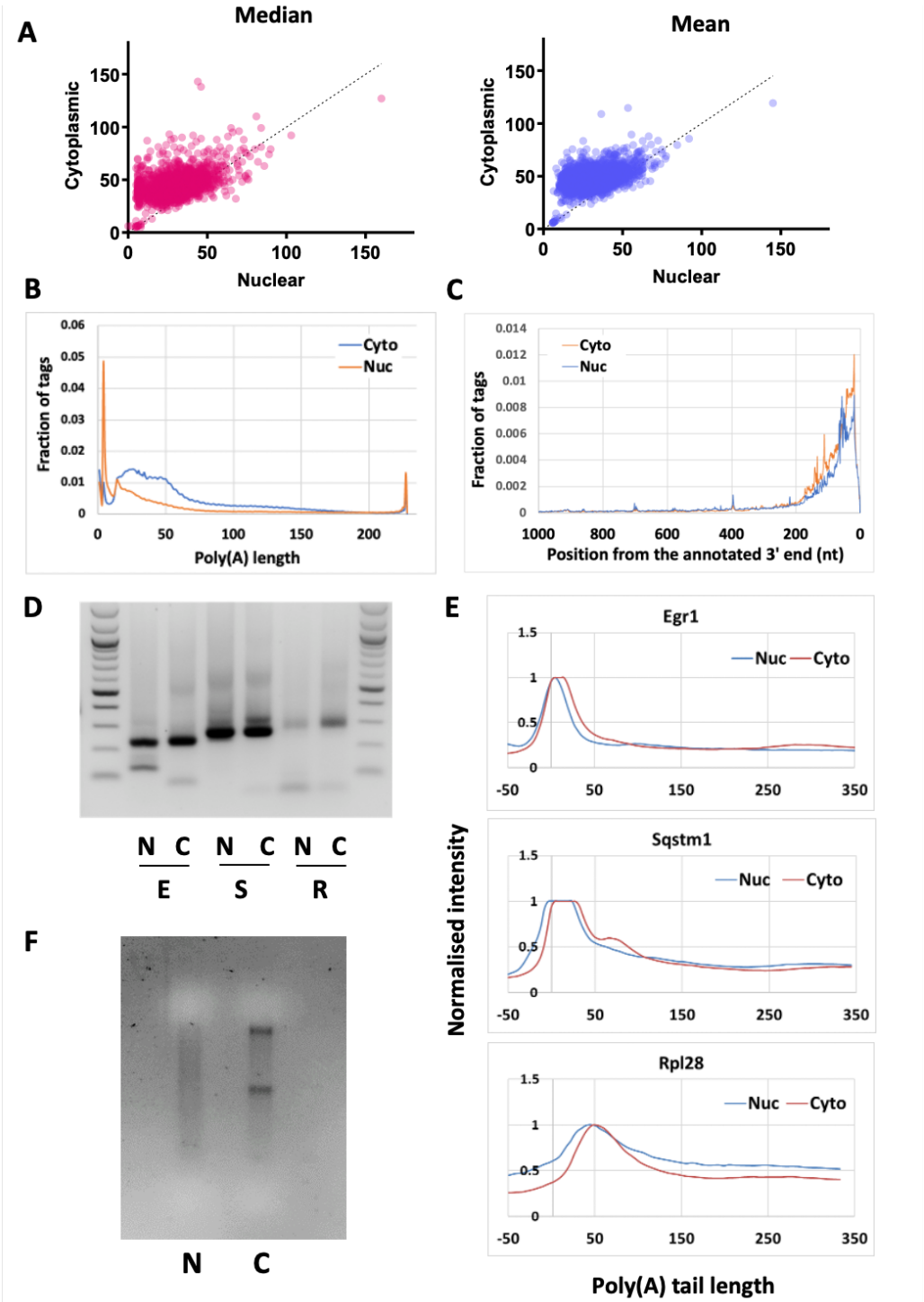
Table 7.1 Quality control data for steady state and serum-stimulated PQ-Seq experiments.

7.3.1 Poly(A) tails of nuclear/cytoplasmic RNA from cells at steady state

The earliest test of PQ-Seq was using RNA isolated from the nuclear and cytoplasmic fractions of NIH 3T3 cells at steady state rather than following serum stimulation. RNA was harvested and DNase treated, and quality control was carried out to test RNA integrity and fractionation efficiency. qPCR data showed good unspliced enrichment in the nuclear fraction (figure 7.2 G) but the RNA gel suggested possible degradation of the nuclear fraction (figure 7.2 F). It was difficult to judge integrity of the nuclear RNA with any certainty since using intensity of ribosomal RNA bands may not be appropriate.

Anchor addition and library preparation were performed by Sunir Malla, part of DeepSeq (University of Nottingham) at the time. Distribution of read 1 5' ends confirmed enrichment of the library towards the 3' ends of annotated transcripts (figure 7.2, C). Summary data returned from the tailseeker software included – for each gene - the mean and median poly(A) length with upper and lower confidence intervals, the number of poly(A)+ and non-poly(A) tag counts, as well as information on proportion of other residues in different regions of the tail. To generate an overview of poly(A) dynamics at steady state, median and mean poly(A) lengths were compared between the cytoplasmic and nuclear fractions for genes with at least 30 poly(A)+ tag counts in each fraction.

Median and mean poly(A) tail lengths were similar between the two fractions (figure 7.2 A), suggesting that the poly(A) tails of most mRNAs with sufficient counts were



G

		2^{-Ct}			2^{-Ct}
N mature	17.71	4.68019E-06	C mature	14.32	4.9E-05
N unspliced	19.52	1.33475E-06	C unspliced	25.4	2.3E-08
N us/m		**0.285190929**	**C us/m**		**0.00046**
N/C =	617.4				

Figure 7.2 Poly(A) tail length determination in the nuclei and cytoplasm of steady state NIH 3T3 cells. Legend overleaf.

controlled in the nucleus. Good enrichment of unspliced over spliced Rpl28 in the nuclear fraction suggested that this was not due to poor separation of nuclear and cytoplasmic RNA (figure 7.2 G). Surprisingly, both the mean and median metrics showed slightly longer cytoplasmic than nuclear poly(A) tail lengths. This contradicted the view that poly(A) tails are made long in the nucleus and then shortened in the cytoplasm. Taking all poly(A) tags (filtered of common contaminants, but not aligned) into account, the poly(A) length distributions of individual tags for the two fractions were starkly different; the cytoplasmic fraction contained a single broad peak, whereas the nuclear fraction contained additional sharper peaks at very short and long poly(A) lengths (figure 7.2 B). As these plots were generated prior to the mapping step, it was not clear whether the very short or very long nuclear tails were present in the subpopulation of mRNAs considered above (i.e. those which were required to have at least 30 poly(A) tags per fraction). Since the smaller median length in the nuclear fraction appeared to be a small but global effect (i.e. rather than large differences on known targets of cytoplasmic polyadenylation), other comparisons were carried out to try and resolve whether this was a true biological phenomenon or a methodological artefact. Individual mRNAs measured by PAT also displayed subtly longer peak tail lengths in the cytoplasm compared with the nucleus, suggesting that the shorter average poly(A) tails observed in the nucleus were genuine, though slight differences in ladder migration could have accounted for this (figure 7.2 D, E). Since both methods had a number of steps in common (addition of a 3' anchor, albeit by different means, reverse transcription and PCR) it was also possible that some artefact common to both techniques could have resulted in a shorter nuclear poly(A) tail readout.

Figure 7.2 (previous page). **Poly(A) tail length determination in the nuclei and cytoplasm of steady state NIH 3T3 cells. A)** Median and mean poly(A) tail lengths measured by PQ-Seq for genes with at least 30 poly(A) counts in both fractions. **Dotted line indicates x=y**. **B)** Tail length distribution of all poly(A) tags after removal of common contaminants but before alignment. **C)** Distribution of read 1 5' ends showing enrichment of PQ-Seq fragments at the 3' ends of annotated transcripts. **D)** RL2-PAT, using agarose gel stained with SYBR safe, for Egr1 (E), Sqstm1 (S) and Rpl28 (R) using the same RNA samples. **E)** Poly(A) distributions for the gel in D. **F)** Agarose gel electrophoresis to determine integrity of these RNA samples. **G)** qPCR data for spliced and unspliced Rpl28 mRNA confirming good separation of nuclear and cytoplasmic fractions.

It was possible that some RNAs displaying unexpected poly(A) tail dynamics could be noncoding RNAs. To see if these made a significant contribution, RNAs were filtered such than only protein-coding or only noncoding RNAs were plotted. Both groups

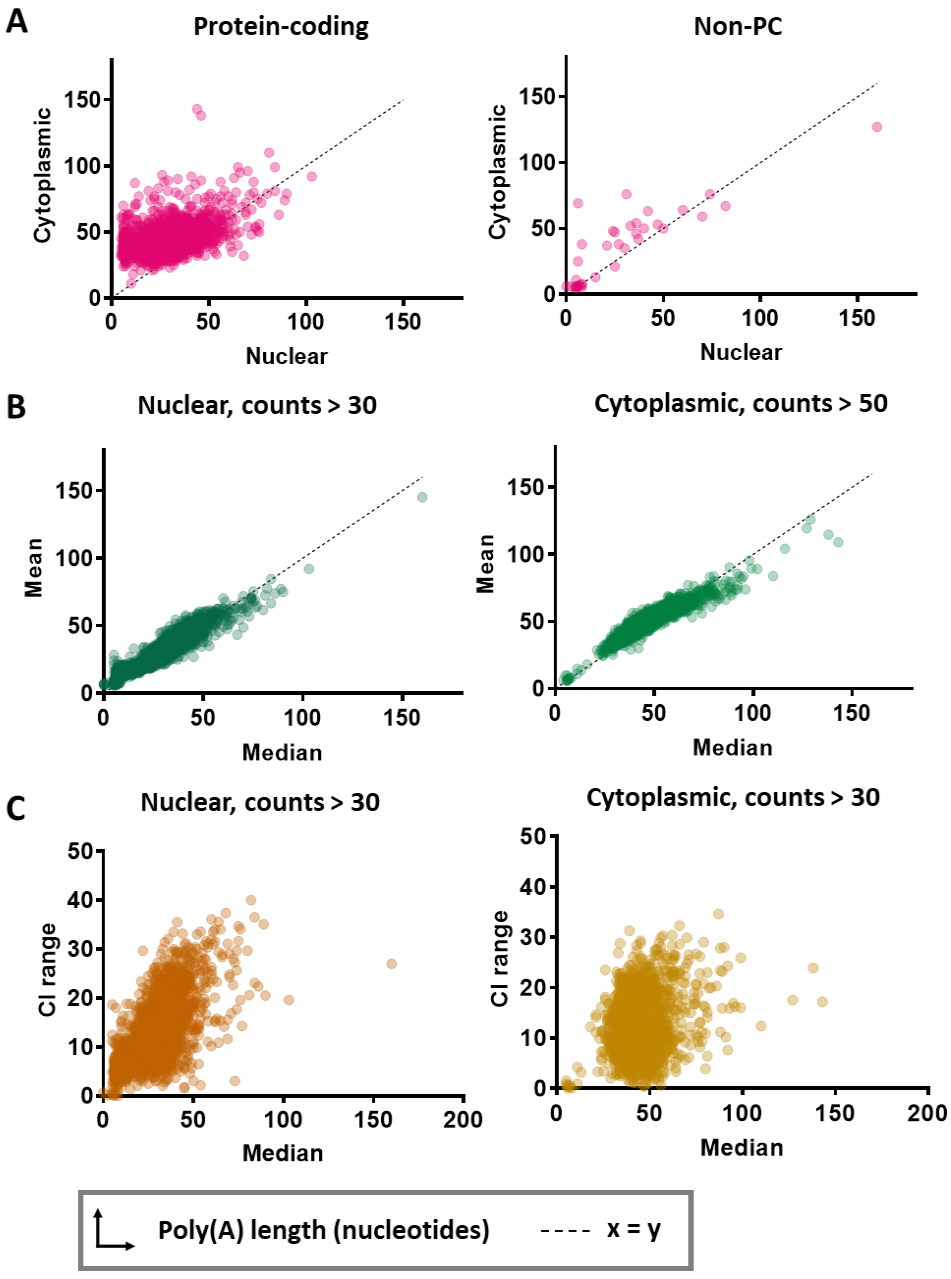


Figure 7.3 Poly(A) tail lengths in NIH 3T3 cells at steady state measured by PQ-Seq and analysed using tailseeker. A) Median poly(A) tail lengths for protein-coding and non protein-coding genes with at least 30 poly(A) counts in both fractions. **B)** Median vs mean for genes with at least 30 or 50 poly(A) counts in the nuclear or cytoplasmic fractions respectively. **C)** Range between min and max confidence interval plotted against median tail length in each fraction for genes with at least 30 poly(A) counts both fractions.

showed longer tails in the cytoplasm (figure 7.3 A). The mean and median do not take into account spread and therefore when used alone cannot describe the shape of the underlying data. This was relevant because the nuclear fraction will have contained a mix of RNAs at different stages of synthesis, likely including RNAs possessing incompletely synthesised poly(A) tails and perhaps some committed to a nuclear decay pathway. Since export from the nucleus is thought to be quick, the nucleus could have contained a lower proportion of mRNAs with full length tails, particularly in the case of stable mRNAs which can accumulate in the cytoplasm. Under steady state conditions, this may have had a particularly noticeable effect since at any given moment the number of inducible mRNAs would have been low compared to those involved in housekeeping.

The mean and median were compared for all genes with a poly(A) tag count greater than 30 (nuclear) or 50 (cytoplasmic) to get an idea of skew in the different fractions. Mean and median were approximately equal in both fractions, suggesting little skew to the data (figure 7.3 B). In the cytoplasm, the mean appeared slightly higher than the median for mRNAs with poly(A) tails < 50 nt, and slightly lower for mRNAs with longer tails. This suggested that a slight majority of poly(A) tails were shorter than the mean length for mRNAs with mostly short poly(A) tails – i.e. most poly(A) tails on these mRNAs were short, but there could have been a few longer tails which increased the mean. In theory this could have been caused by trimming of poly(A) tails to some short set length, with the few longer tails yet to be deadenylated. Since the opposite was true in mRNAs with mostly long tails, this perhaps suggested existence of a few almost fully deadenylated transcripts. It is important to note that in both cases, the effect size was small.

To investigate whether incomplete polyadenylation/processing was detectable as greater length variation in the nuclear fraction, the range between upper and lower confidence interval was plotted against median poly(A) length for genes with poly(A) tag count >30 in both fractions. While spread did not vary with median length for the cytoplasmic RNA, there appeared to be a slight increase in spread with tail length in the nuclear fraction (figure 7.3 C). Both the median vs mean and median vs CI range plots showed a large number of short poly(A) tails in the nuclear but not the cytoplasmic fraction (figure 7.3 B, C). While the presence of these shorter tails

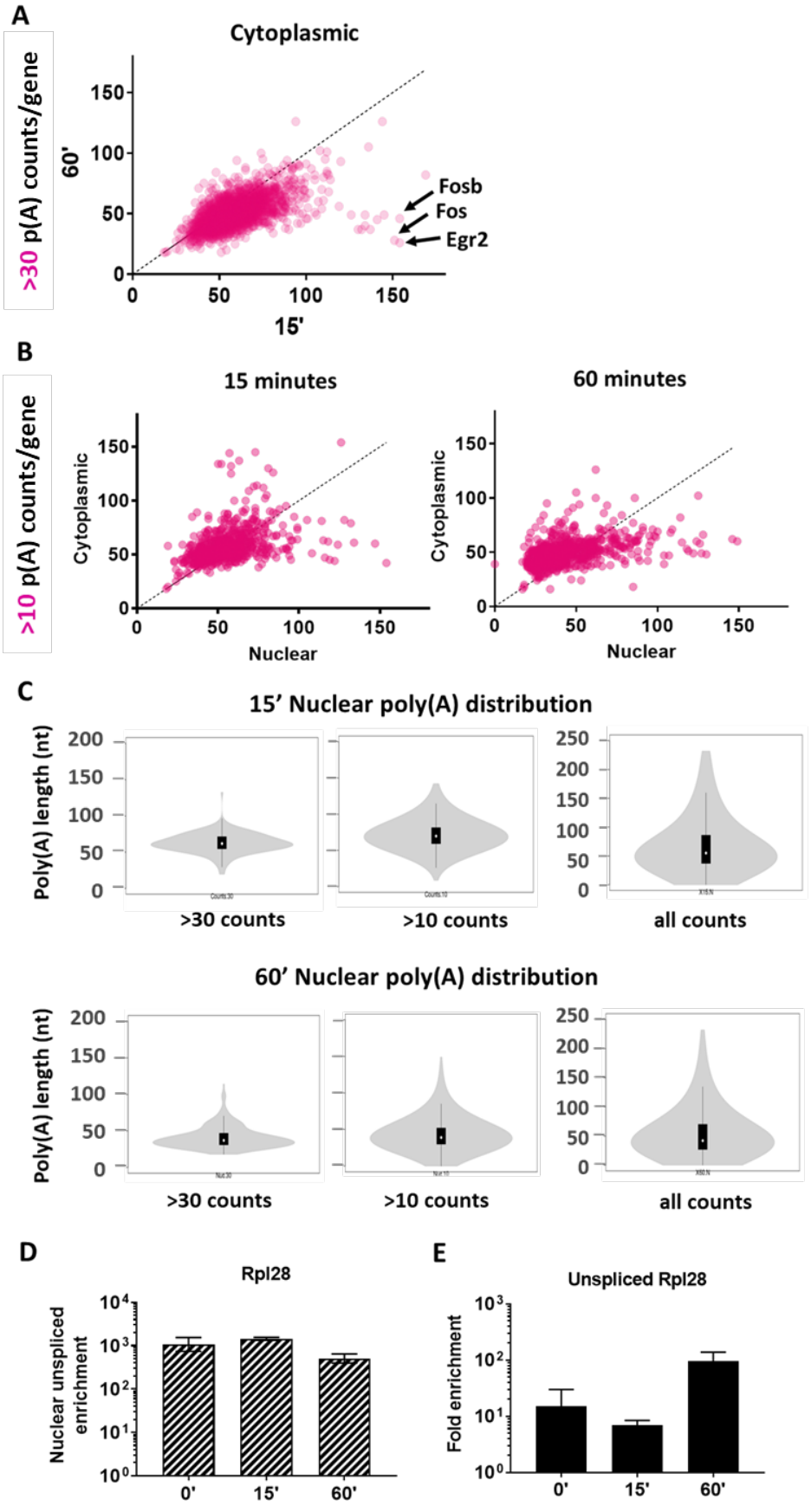


Figure 7.4 PQ-Seq nuclear/cytoplasmic serum stimulation preliminary experiment (legend overleaf).

may have been biologically significant, they could also have been caused by possible degradation in the nuclear fraction.

7.3.2 Poly(A) tails of nuclear/cytoplasmic fractions in the serum response

To see whether changes in polyadenylation that occur during the serum response were detectable by PQ-Seq, NIH 3T3 cells were stimulated with serum for 0, 15 or 60 minutes and separated into nuclear and cytoplasmic fractions. All time points exhibited good separation as measured by nuclear enrichment of unspliced mRNA (figure 7.4 D) and there was some low-level contamination with genomic DNA (figure 7.4 E). PQ-Seq libraries were prepared by Hilary Collins (GRRB, University of Nottingham) and Sunir Malla (DeepSeq, University of Nottingham), and poly(A) tail length data was generated by Daniel Zadik (DeepSeq, University of Nottingham) using the tailseeker pipeline (325, 649). Although PCR cycle number was optimised for each sample during library preparation and thus equal concentrations of each library were loaded onto the MiSeq chip, sequencing depth was superior in the cytoplasmic fractions. This may have been due to lower relative proportion of polyadenylated RNA in the nuclear fraction – perhaps due to unintended capture of DNA or lower purity starting material. As a consequence, poly(A) tail lengths could be determined for more genes with greater confidence in the cytoplasmic fractions compared with their nuclear counterparts.

As with the steady state experiment, median poly(A) length was similar between the two fractions after both 15 and 60 minutes of serum stimulation (figure 7.4 B). This time, points emerged showing long nuclear but medium length cytoplasmic poly(A) tails, consistent with cytoplasmic deadenylation of serum-induced mRNAs occurring. These differences suggested that the observation of highly similar nuclear and

---

Figure 7.4 (previous page) **PQ-Seq nuclear/cytoplasmic serum stimulation preliminary experiment. A)** Comparison of median poly(A) tail lengths in the cytoplasmic fraction of 15 and 60 minute stimulated cells. **Dashed line indicates x=y**. **B)** Comparison of nuclear and cytoplasmic median poly(A) tail lengths after 15 or 60 minutes of serum stimulation. **Dashed line indicates x=y. C)** Distributions of median poly(A) tail lengths in the nuclear fractions after 15 or 60 minutes of serum stimulation. Different thresholds for poly(A) tag count were compared. **D)** Ratio of unspliced:spliced mRNA in the nuclear fraction divided by that in the cytoplasmic fraction. **E)** qPCR data showing DNA enrichment in cDNA vs no RT control for nuclear fractions.

cytoplasmic poly(A) tails at steady state was not due to a technical glitch. It is possible, however, that the appearance of these points could also have resulted from the use of a different poly(A) count threshold (10 vs 30) to account for the lower sequencing depth of the serum response experiment. Median cytoplasmic poly(A) length was higher early in the serum response compared with the late time point for most mRNAs, consistent either with changes in synthesised poly(A) length or gradual cytoplasmic deadenylation (figure 7.4 A). At nuclear tail lengths below about 50 nt, cytoplasmic poly(A) tails tended to be longer, whereas the opposite was true of nuclear tails > 100 nt. The cluster of genes after 15 minutes stimulation which had median cytoplasmic tails >100 nt but median nuclear tails <100 nt did not show any GO term enrichment. Again, the shorter median tails in the nuclear fractions compared with their cytoplasmic counterparts may have been due to contributions from not yet polyadenylated transcripts compared with large numbers of stable polyadenylated transcripts in the cytoplasm. Long nuclear but not cytoplasmic tails on the other hand, may have represented transcripts which underwent rapid cytoplasmic deadenylation.

When results were restricted to those genes with at least 30 poly(A) counts, neither fraction contained 200-250 nt median poly(A) tails at any time point (figure 7.4 A – C, 0’ data not shown), though this could have been partially due to the PQ-Seq detection limit of 234 adenosines. Although tails slightly shorter than 200 nt were expected for the mRNAs studied by PAT, the absence of any long median tail lengths at all was initially puzzling. Of course, longer individual poly(A) tails may have been detected but not faithfully conveyed by use of the median value for the gene. If the threshold for the number of poly(A) counts per gene was reduced from 30 to 10, there was a considerable increase in the range of median tail lengths observed (figure 7.4 C). 200-250 nt median tails only appeared when fewer than 10 counts were required, but so few counts may not have been sufficient for a reliable median. In agreement with the trends observed by PAT in serum response mRNAs, the distribution of nuclear poly(A) tail lengths was more enriched for shorter tails in the 60 than the 15 minute time point (figure 7.4 C).

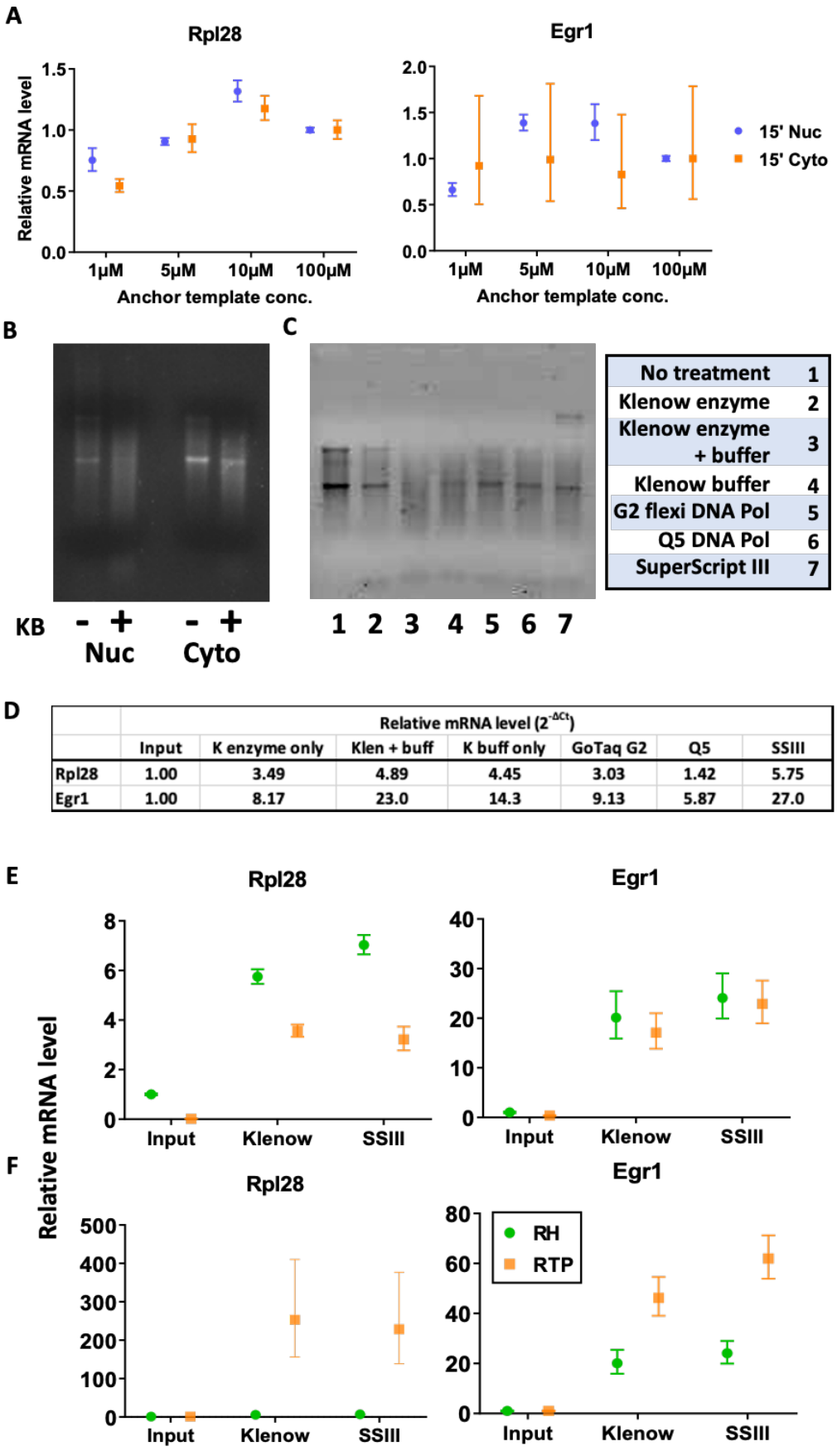
Aside from the possibility that long poly(A) tails were not present in these samples, there are several routes via which long tails could become depleted in the analysis.

One possibility, alluded to in the tailseeker documentation is the increased risk that long-tailed transcripts have of their forward read also being poly(A), and consequently being unmappable. As mentioned previously, anchor design combined with choice of sequencing platform limit detection to a maximum of 234 nt. Furthermore, since the above analyses only take into account the mean/median poly(A) tail length provided by the tailseeker output summary, it may be that some long tails are present and detectable but at the edges of a distribution so may not be well conveyed by summary statistics.

7.4 Optimisation for future libraries

Given the earlier observation that poly(A) tail length seems to be regulated before dissociation from the chromatin for a number of mRNAs (figure 3.5 A,D), the prospect of global chromatin-associated poly(A) measurements was enticing. To try and improve the PQ-Seq pipeline such that RNA from the even lower-yield chromatin fraction could be used – and to address the problem of low poly(A) sequencing depth in the nuclear fraction, the library preparation protocol was dissected to identify and optimise less efficient steps. At each stage, quick measurements were employed, with the aim of compiling a set of standards against which future libraries could be compared at regular stages in their synthesis.

There was speculation as to whether excess concentration of the anchor template could be competing with the RTP primer during first strand synthesis. Using nuclear and cytoplasmic RNA from 15 minute serum stimulated cells, four different stock concentrations of anchor template were trialled. The PQ-Seq protocol was followed from anchor extension, through streptavidin bead pull down to RNA cleanup using RNA Clean XP beads, but the T1 digestion and gel separation were omitted due to concerns over efficiency. Rather than proceeding to QuantSeq library preparation, RNA samples underwent reverse transcription with the RTP primer followed by quantitative PCR using primers designed for the 3' ends (since in the full protocol, this end of the mRNA should be enriched following pull-down with the anchor template). For the Egr1 transcript, 5 μM and 10 μM stock concentrations of anchor template both allowed for a moderately more efficient reverse transcription than the original 100 μM concentration for the nuclear fraction (figure 7.5 A). For Rpl28,



	Relative mRNA level ($2^{-\Delta Ct}$)						
	Input	K enzyme only	Klen + buff	K buff only	GoTaq G2	Q5	SSIII
Rpl28	1.00	3.49	4.89	4.45	3.03	1.42	5.75
Egr1	1.00	8.17	23.0	14.3	9.13	5.87	27.0

Figure 7.5 Optimisation of PQ-Seq anchor addition (continued overleaf).

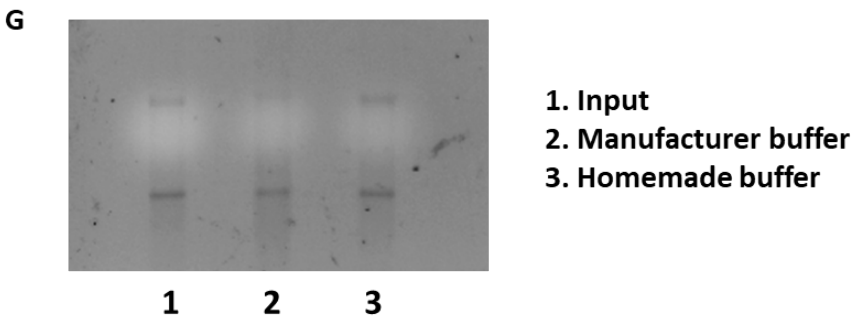


Figure 7.5 Optimisation of PQ-Seq anchor addition. A) qPCR data following reverse transcription with RTP primer to assess optimum RA3 anchor template concentration. CTs normalised to 100µM value for each fraction. **B, C)** RNA integrity following B) incubation with manufacturer-supplied Klenow buffer or C) incubation with different enzymes/ buffers. **D)** Relative mRNA level of 3' most 200nt determined by qPCR after anchor addition by different enzymes. Value given to 3 significant figures. **E)** Comparison of 3' end recovery following anchor extension using either klenow exo- fragment or SuperScript III compared with no extension (input). mRNA level normalised to that of the input reverse transcribed using random hexamers. 25' Cytoplasmic RNA was used. **F)** Same data as in E), but RTP data is normalised to input reverse transcribed with RTP rather than RH (i.e both input values = 1) in order to compare enrichment over input for each reverse transcription primer choice. **G)** Comparison of RNA integrity using homemade vs manufacturer Klenow buffer.

10 µM was clearly the best tested concentration, and 10 µM anchor template stock concentration was therefore taken forwards for the remaining optimisation steps.

It was also possible that the improvement seen using 10 µM (figure 7.5 A) meant that less unbound anchor associated with the streptavidin and allowed better recovery of mRNA. Although associated documentation for the beads indicated a much higher binding capacity than required, previous experience using a number of different streptavidin beads was that similar yields could be obtained by incubating a fresh set of beads with the original flow-through; this suggested that poor recovery via the biotin-streptavidin interaction could be responsible. The initial part of the PQ-Seq protocol (without the T1 digest and size separation) was carried out on new, 25 minute stimulated nuclear and cytoplasmic RNA using the amended 10 µM anchor concentration. RNA was recovered from each flow through and wash step. Strikingly, only a tiny proportion of the input RNA (i.e. RNA before anchor

addition) was recovered in the eluted fraction, while 10-50 % was present in the combined flow-through/first wash (W1) fraction, leaving between 50 and 90 % of the input RNA unaccounted for (figure 7.6 A). Interestingly, the proportion in the wash seemed to be higher in the cytoplasmic fraction, possibly reflecting a higher starting amount of RNA, a greater proportion of which may therefore fail to be captured by the beads or anchor template.

A number of hypotheses could explain the missing RNA/cDNA. One obvious possibility is that the RTP primer used for reverse transcription of the wash and eluted fractions was significantly less efficient than the random hexamers used for the input RNA. Subsequent experiments comparing reverse transcription primers showed a variable magnitude of effect with a maximum 2-fold reduction (figure 7.5 E). If there was a 2-fold reduction in RT efficiency using RTP compared with RH, the previous data for cDNA in each purification fraction suggested that up to ~35 % of the nuclear RNA, and almost all of the cytoplasmic RNA may have failed to be captured by the streptavidin beads and instead been lost in the flow-through and first wash. Another possibility was that some reagent introduced during the elution or RNA isolation steps (eg. Formamide or traces of Phenol) could have inhibited the reverse transcription or PCR, leading to a falsely low readout of mRNA abundance in this fraction, but this was not formally tested. A more recent consideration is that non-stick Eppendorf tubes were not used for the streptavidin purification step and could also have accounted for some loss which would not have been detected in the flow-through.

Another explanation may have been RNA degradation. Since all reagents in the pull down were RNase-free, the most likely point of degradation was during anchor addition. Some degradation was identified following incubation with manufacturer supplied Klenow buffer (figure 7.5 B). As exo(-) Klenow was used, it seemed unlikely that the enzyme's proofreading activity was responsible. To see whether the degradation could be circumvented using another enzyme, RNA quality was tested by agarose gel electrophoresis after incubation with the anchor template along with GoTaq G2 Flexi or Q5 DNA polymerases, or SuperScript III (SS III) reverse transcriptase, as well as different combinations of Klenow enzyme and buffer. All three alternative enzymes were less detrimental to RNA integrity than Klenow and

the manufacturer-supplied buffer (figure 7.5 C). The same RNA samples were reverse transcribed with random hexamers, and abundance of 3' ends of Rpl28 and Egr1 measured by qPCR. Despite the RNA degradation observed in the Klenow + buffer reaction, this treatment, along with SS III gave the highest levels of mRNA 3' ends (figure 7.5 D). Notably, all treatments showed an increase in cDNA level relative to input which is surprising since reverse transcription was carried out with random hexamers and should not have favoured RNA with anchor attached. One possibility is that all enzymes enhanced cDNA production during the reverse transcription stage. To test whether SS III could also use the DNA anchor template, Klenow and SS III samples were reverse transcribed using the RTP primer, and 3' end abundance was again tested by qPCR. Original incubation with SSIII was slightly superior to Klenow when comparing both random and RTP-primed reverse transcription reactions (figure 7.5 E, F) indicating its capability in using a DNA template. The superior performance of SS III–mediated anchor addition could have been due to reduced RNA degradation but may also have resulted from a higher concentration of reverse transcriptase causing more efficient reverse transcription. Since there was only minimal difference between the two enzymes, and Klenow was used successfully in previous iterations, use of Klenow was continued but a homemade Klenow buffer was trialled which elicited marginally less RNA degradation (figure 7.5 G).

A final avenue of exploration was efficiency of elution from the streptavidin beads. The original protocol entailed a series of short melt steps at 80° (water, formamide, then water again) which should disrupt base pairing between the anchor template and the captured RNA fragment. Two other strategies were tested to see whether RNA recovery in the eluted fraction could be improved: elution using BL+TG lysis buffer from the Promega ReliaPrep RNA isolation kit (which contained guanidine thiocyanate and so should have denatured the duplex) and treatment with 1 % SDS, which should have disrupted the biotin-streptavidin interaction and caused release of the whole duplex from the beads. Eluate from the SDS-treated beads was further divided into 'melt' or 'no-melt' treatments with the rationale that continued presence of anchor template may competitively inhibit downstream reverse transcription and PCR steps. In the 'melt' pathway, the eluted fraction was heated and passed over a fresh set of streptavidin beads to bind the biotinylated anchor

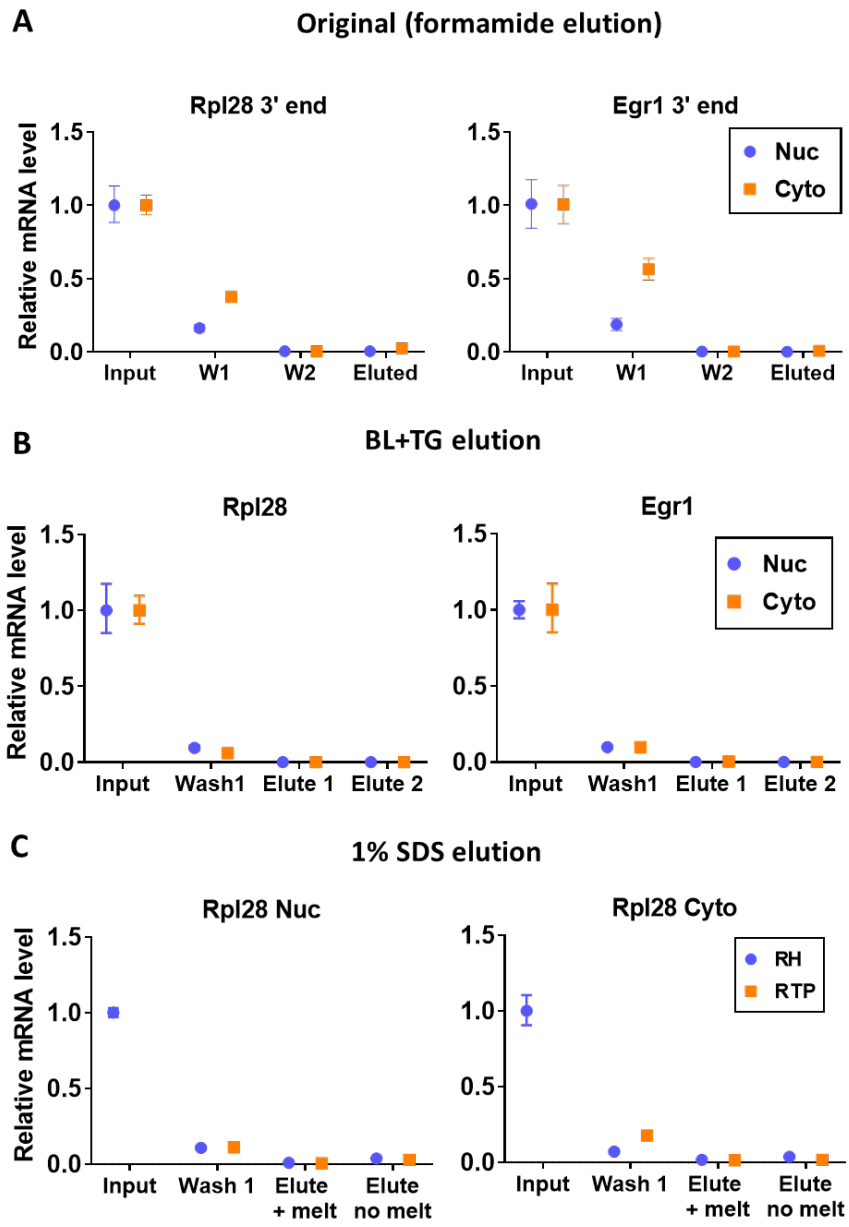


Figure 7.6 PQ-Seq optimisation: streptavidin bead binding. qPCR performed to assess abundance of the 3' ends of two mRNAs in different fractions of the streptavidin bead purification compared with original RNA. W1 indicates combined flow-through and washing of streptavidin beads. W2 represents flow through and wash during purification with RNAClean XP beads. Error bars indicate standard deviation. **A)** Original elution protocol. Reverse transcription performed using the RTP primer for W1, W2 and Eluted, and with random hexamers for input. mRNA level was normalised by ΔCt to the input RNA. **B**) As in A, but BL+TG lysis buffer used to ensure full elution of RNA, and reverse transcription was performed with random hexamers. Elute 1 is from streptavidin beads, Elute 2 is following subsequent cleanup with RNACLean XP beads. **C**) As in A,B but high SDS concentration used to disrupt the biotin-streptavidin interaction and only a single mRNA assayed. Eluted RNA representing the biotinylated template:mRNA duplex was melted to separate the strands and passed over another set of beads to remove the biotinylated template. Reverse transcription using random hexamers (RH) and the RTP primer are compared.

template but not the RNA. Neither the BL+TG nor the SDS elution methods noticeably improved RNA recovery (figure 7.6 B, C) suggesting that poor binding to the beads or poor binding of the anchor template to the RNA was responsible. Strangely, less cDNA was detected in the wash steps of these experiments, despite using the same batch of 25' serum-stimulated nuclear and cytoplasmic RNA as in the original protocol. This could have been due to variability in RT efficiency, RNA degradation and adherence of RNA to the Eppendorf tube. While it was not determined exactly where the RNA was lost, failure to bind the beads or RNA sticking to the microcentrifuge tube seemed the most likely explanations, and non-stick tubes should be employed for future rounds.

7.5 Discussion

In an ever-expanding landscape of sequencing techniques, PAT-Quant-Seq is able to measure poly(A) tail lengths at the global level from low input (1 μg) RNA. In combination with subcellular fractionation, PQ-Seq showed that median poly(A) lengths were much more similar in the nucleus and cytoplasm than would be expected if all poly(A) tails exited the nucleus with long poly(A) tails and were then deadenylated in the cytoplasm (figures 7.2 A, 7.4 B). This is suggestive of widespread nuclear poly(A) length regulation in NIH 3T3 cells. In agreement with other poly(A) deep-sequencing studies, PQ-Seq also confirmed the existence of a majority of mammalian RNAs with poly(A) tails which were significantly shorter than the widely recognised 200-250 nt figure (11, 13, 325, 326, 432, 433, 633, 636). Under steady state conditions, median poly(A) length was slightly higher in the cytoplasm than the nucleus (figure 7.2 A), which may have come about through the presence of transcripts in the nucleus which were not fully polyadenylated, or through cytoplasmic polyadenylation. Since the confidence intervals were approximately the same in each fraction, it seemed unlikely that a large proportion of nuclear polyadenylated RNA had long poly(A) tails. This is consistent with a model in which most poly(A) tails are rapidly deadenylated to their 'proper' lengths while still attached to the chromatin.

Although PQ-Seq is demonstrably accurate when compared with PAT measurements, the question remains as to why so few long tails were observed in the nuclear fraction when early pulse labelling experiments showed an abundance

of long poly(A) tails with a peak centred around 210 nt in the nucleus and 190 nt in the cytoplasm (324, 610). While the PAT assay has detected long tails in a minority of cases, this was often on low abundance mRNAs -some of which did not have enough nuclear counts to get a reliable read by PQ-Seq (eg. Fos) - or in the chromatin fraction which only comprises a minority of the nuclear RNA. Indeed, long median poly(A) lengths were reported in the nuclear fraction of both the 15' and 60' time points if genes with low tag counts were included (figure 7.4 C). Summary statistics reported by tailseeker will also have failed to convey information about lengths at the edges of the distribution for any given gene, meaning that long tails may still have existed on individual transcripts. This seemed to be the case in the original TAIL-Seq paper where distribution of poly(A) lengths extended to 230 nt for NIH 3T3 cells, but median poly(A) length was only between 0 and 120 nt (325). Long poly(A) tails may also have been disproportionately lost during the mapping step since there would be a higher chance of read 1 also containing poly(A) tail and therefore mapping poorly or failing to map. Another possibility could in theory have been the presence of non-adenosine 3' termini on long tails inhibiting anchor addition, however there is no basis for this in the literature, and uridylation for example is more common on short poly(A) tails (325).

The number of poly(A) counts were lower in general for the nuclear fractions, perhaps due to lower starting concentrations of RNA (due to contribution of DNA to the nanodrop read) or interference by other non-polyadenylated RNAs or DNA. Since the intention was to apply PQ-Seq to fractions of potentially even lower quality (i.e. chromatin-associated), the anchor addition and purification portion of the protocol underwent optimisation. It was determined that the anchor template concentration should be reduced tenfold, and that a homemade rather than manufacturer-supplied Klenow buffer should be used during anchor extension to minimise RNA degradation (figures 7.5 A – C, G). The streptavidin bead purification step also seemed to be a significant point of RNA loss, though RNA degradation and inefficiencies in reverse transcription could also have contributed to loss compared with the input (figure 7.6 A). Poor binding of beads by the duplex or loss of RNA through erroneous use of 'sticky' Eppendorf tubes are likely contributors.

8 Conclusions and discussion

This project investigated the existence of nuclear poly(A) length control during the NIH 3T3 serum response, and its relationship with mRNA production. The poly(A) tail test (PAT) was validated as a convenient and reproducible assay for identifying differences in poly(A) tail sizes and was the primary mode of tail length measurement used. Nuclear poly(A) tails of serum-induced mRNAs were long (120-180 nt) 20 minutes after serum induction, but were short by one hour post-stimulation, whereas housekeeping mRNAs had a consistent medium length (30-70 nt) tail throughout (figure 3.5), and the location of poly(A) regulation varied between the nucleoplasm and the chromatin. This contradicts early reports of a universal 200-250 nt poly(A) tail on newly made mRNA (179, 269, 324, 363, 411). Knockdown of mRNA encoding the CCR4-NOT scaffold subunit, CNOT1, in NIH 3T3 cells caused gene-specific increases in nuclear poly(A) tail length on all transcripts tested (figures 4.2, 4.3), suggesting that widespread regulated nuclear deadenylation by CCR4-NOT may normally take place. It was not possible to distinguish between rapid chromatin-associated deadenylation and limited initial poly(A) synthesis using the PAT assay, however, since short-term pharmacological inhibition of the CAF1 nuclease subunit caused long poly(A) tails on transiently expressed mRNAs late in the serum response (figure 6.7 A, B), deadenylation by CCR4-NOT seems the simplest model.

As well as increasing early poly(A) tail length, Cnot1 knockdown caused a reduction in pre- and mature serum response mRNA levels and in housekeeping pre-mRNA levels (figure 5.1), which is consistent with the CCR4-NOT complex's role as an elongation factor (39, 147, 148). This coupling of longer poly(A) tails with reduced mRNA production suggests a drive towards mRNA homeostasis when degradation is impaired. Finally, application of our global poly(A) length measurement technique, PQ-Seq, to nuclear and cytoplasmic NIH 3T3 fractions showed that for most mRNAs, median poly(A) length in the nucleus and cytoplasm were similar. This suggests that nuclear poly(A) length control is a major contributor to poly(A) tail regulation for most mRNAs.

8.1 Appraisal of the poly(A) tail test (PAT)

PAT analysis using native agarose gel electrophoresis gave clear bands at reproducible lengths throughout the project. Additional longer bands were regularly also visible and these were demonstrated by a previous lab member to be multimers, formed by the hybridisation of poorly aligned A and T stretches of different fragments. As such, these longer bands were excluded from length determination. It was unclear however, whether multimer formation was more likely to involve fragments with longer tails and therefore whether this would introduce a bias towards shorter poly(A) tails in the lower ('monomeric') band. To address this, PAT PCR products were separated by denaturing capillary gel electrophoresis since this should dismantle any multimers. Denaturing PATs produced a similar shape poly(A) distribution to native PATs, but with a slightly shorter peak length (figure 3.1 B), suggesting that multimer formation in the native PAT does not cause a loss of detection of long tails. It seems unlikely that shorter poly(A) tails are enriched in multimers since there is a lower chance of poor alignment between strands, and therefore less opportunity for hybridisation with other fragments. Thus, this small difference in peak length may have arisen from the use of different markers between systems, or from slight differences in how the marker ran compared with the samples in the non-denaturing gel.

While PCR-based techniques are more sensitive - and therefore enable higher resolution - than direct measurement of RNA, it has been suggested that the PCR step could introduce a bias towards short poly(A) tails. This is based on the propensity of the polymerase to repeatedly dissociate and re-join the template, and in homopolymeric regions, this re-joining may not occur at the same place it fell off. Although there is an equal probability of extending or reducing a poly(A) tract by this means, the chance of incorrect re-joining occurring is reduced in shorter tails, which could lead to an overall bias towards short poly(A) tails. Such an artefact would affect the majority of global poly(A) surveys published in the last 7 years. This deviation from 200 nt in whole cell mRNA has been thought to be explained by cytoplasmic deadenylation (325, 432, 433, 582, 632, 633). To check whether the PAT assay suffered from artificial poly(A) shortening, RNase H northern blots were performed on identical biological samples. As RNase H treatment generally yielded

poor PAT signal - even with primers that should bind the remaining fragment - PAT assays were performed on the original RNA (or DNase-treated RNA where relevant).

For Egr1 mRNA which usually has a long poly(A) tail after 20 minutes and a short tail after 60, the PAT and northern blot measures gave similar peak lengths distributions (figure 3.4 H). As the northern blot signal was too low to give a smooth curve it was not possible to compare exact figures for modal poly(A) length. Derivation of poly(A) lengths in the northern blot was initially complicated by the presence of a second band in 3 out of 4 lanes which was particularly prominent in oligo(dT) treated samples (figure 3.4 C), suggesting it was caused or enriched by additional degradation at some A-rich segment. Surprisingly, there was no A-rich region within the annotated 3'UTR which would generate shorter fragments that would still be detected by the labelled probe. Upon closer inspection of deep sequencing data from our NIH 3T3 cells, there appeared to be a 25 nt A-rich extension to the Egr1 3'UTR which was confirmed by Sanger sequencing of the Egr1 PAT product (figure 3.4 F, G). Oligo(dT) treatment likely caused occasional degradation of this additional section of 3'UTR, resulting in two different lengths for the deadenylated bands. Both UTRs may have been expressed in our cells, explaining the presence of additional faint bands in the no-oligo(dT) lanes.

For Rpl28 mRNA which has a consistent medium length poly(A) tail throughout the serum response, the PAT and northern blot measures gave similar modal tail length values, but had different distributions (figures 4.2 C, 4.3 D). Tail length profiles were more similar when the northern blot was carried out using the RNase H cleaved RNA than when full length RNA was used, though this cleavage also caused the appearance of additional shorter than expected bands. Although the modal tail length values were similar, the northern blots seemed to detect a greater number of longer tails than the PAT. While this could have been due to theoretical poly(A) shortening by PCR, the additional smearing in the northern blots could equally have resulted from sample over-loading or from less accurate size determination. Despite this, PAT assays and northern blots were in good general agreement in terms of detecting size differences, and the benefit of additional convenience and sensitivity with the PAT assay outweighed any concerns regarding absolute tail length determination.

8.2 Poly(A) tail length is regulated in different responses and locations

Eukaryotic mRNA poly(A) tails have been recorded to change length in a number of processes (13, 62, 66, 330, 331, 396, 397, 402, 413, 577, 584, 587). These changes affect gene expression by altering mRNA stability and/or translation efficiency (396, 397, 584, 587). While much of this control is thought to be mediated by cytoplasmic polyadenylation or deadenylation, message-specific nuclear poly(A) regulation has also been suggested (413).

PAT assays confirmed that following serum stimulation in NIH 3T3 cells, poly(A) tails of induced mRNAs are long early on (20 minutes) and short towards the end (60 minutes) of the transcriptional response (figure 3.5). The shorter poly(A) tails present at the end of the response corresponded to reduced mRNA stability (figure 3.6). Rather than being caused by cytoplasmic deadenylation, the short poly(A) tails late in the response appear to already be decided either in the nucleoplasm or while the mRNA is still attached to the chromatin (figure 3.5). Even early in the response, the longer poly(A) tails observed fell short - to varying degrees - of the classical 200-250 nt figure, though it cannot be ruled out that this length did transiently occur but was not detectable at the chosen time points.

Poly(A) tails of representative housekeeping mRNAs do not change throughout the serum response and are considerably shorter than expected for their stability, measuring between 30 and 70 adenosines which is consistent with observations in the literature (325, 326, 423). These tails too, seem to be controlled in the nucleoplasm or on the chromatin (figure 3.5). Long chromatin-associated tails could be observed for all the mRNAs tested apart from Rpl28, perhaps indicating separate control of ribosomal protein or 5' TOP mRNAs.

Deep sequencing of poly(A) tails using PQ-Seq revealed that in both steady state and serum stimulated NIH 3T3 cells, median nuclear and cytoplasmic poly(A) tail lengths were highly similar (figures 7.2 A, 7.4 B). This suggests that nuclear and perhaps chromatin-associated poly(A) regulation may be a transcriptome-wide phenomenon in these cells. In addition, very few genes had median poly(A) tail lengths

approaching 200-250 nt. It is unclear what caused the disparity between the PQ-Seq measured nuclear poly(A) tail lengths and those first obtained by pulse labelling.

One possibility is that the methods used to determine length in early experiments lacked accuracy and therefore that the 200 nt figure was incorrect. In one paper, ^{3}H-adenosine labelled poly(A) tails were separated by electrophoresis, alongside two size markers: 4S and 5S RNA, representing 80 and 122 nucleotides respectively (324). In both 12 minute and 48 hour labelled RNA, a large portion of signal for both nuclear and cytoplasmic RNA was apparently around 200 nt, with the nuclear population peaking at slightly longer poly(A) lengths than the cytoplasmic one. Although the absolute tail length is different, this pattern is in general agreement with the small differences in length across fractions in the PAT assays, and with the similarity in nuclear and cytoplasmic length as measured by PQ-Seq (figures 3.5, 7.2, 7.4). While the use of two markers may not be sufficient for reliable size determination, it is interesting to note that in the 12 minute labelled cytoplasmic RNA, there was an additional peak beginning just below the 80 nt marker (324). If this size determination were correct, this peak could perhaps represent highly expressed mRNAs which have been shown to possess only medium length tails (325, 326). An equivalent peak was not present in the nuclear fraction, however this fraction generated a peak of a similar size but at even lower poly(A) lengths. The magnitude of difference is too great to relate to the observation by PQ-Seq that median poly(A) tail length was greater in the cytoplasmic than the nuclear fraction at steady state (figure 7.2 A).

The other possibility for the difference in poly(A) length measurement is that PQ-Seq underestimates poly(A) length, and this could occur by several means. First, the apparent failure to detect long poly(A) tails could simply be a symptom of using summary statistics such as the median instead of considering distributions for each gene. Second, PQ-Seq may have suffered from tail shortening due to the PCR amplification step; however, comparison of PAT assays and northern blots suggest that any biases introduced by PCR would not be sufficient to explain such a large disparity. For example, comparing median poly(A) length between PAT assays and northern blot for full length Rpl28 (for which the PAT and northern distributions varied most) gave a 33.8 nt difference, which although substantial, is a long way

from the ~140 nt difference between a 200 nt tail and the ~60 nt tail usually measured for Rpl28 by PAT . Another possibility is that long poly(A) tails may be disproportionately lost during alignment as their forward read is more likely to also contain poly(A) tail and therefore not be mappable. Furthermore, since the PQ-Seq protocol relies on the presence of 3' terminal oligo(A) to hybridise with the anchor template, some tails could in theory be missed by both PQ-Seq and the PAT assay if they possessed non-adenosine termini. This seems an unlikely explanation for bulk tail differences, however, since a genome-wide study found that alternative termini were more common adjacent to otherwise very short poly(A) tails (325).

Evidence of chromatin-associated and nucleoplasmic poly(A) control was also present in HEK293 cells and interestingly, the ribosomal protein mRNA in this case did have a long chromatin-associated tail which was shortened in the nucleoplasm (figure 3.7 A).

8.3 Cnot1 knockdown causes increased poly(A) length alongside decreased pre-mRNA level

Most cytoplasmic deadenylation in mammals is carried out by the CCR4-NOT and PAN2-PAN3 complexes, with the former thought to play a dominant role (26, 50, 503). As well as catalysing poly(A) tail removal in the cytoplasm, CCR4-NOT can enter the nucleus and influence transcription both at the promoter (often repressive) and by promoting Pol II elongation (35, 36, 39, 42, 147, 148, 510, 526, 527, 545–551). As such, CCR4-NOT presented a good candidate for a nuclear effector of shorter than expected poly(A) tails.

Knockdown of the Cnot1 scaffold subunit of the CCR4-NOT deadenylase complex in NIH 3T3 cells led to increased nuclear poly(A) tail length late in the serum response on all mRNAs tested, and in some cases to increased chromatin-associated poly(A) tail length (figures 4.2 A-C, 4.3 A, B). The location in which CNOT1 seemed to affect length corresponded to the location where the first short poly(A) tails are usually seen for a given mRNA, suggesting that it has a role in shortening initially long tails within the nucleus. The increases in poly(A) tail length were larger for induced than housekeeping mRNAs (increases of between 60 and 100 nt compared to between 20 and 40 nt). A similar pattern was observed following *ccr4* deletion in yeast (582).

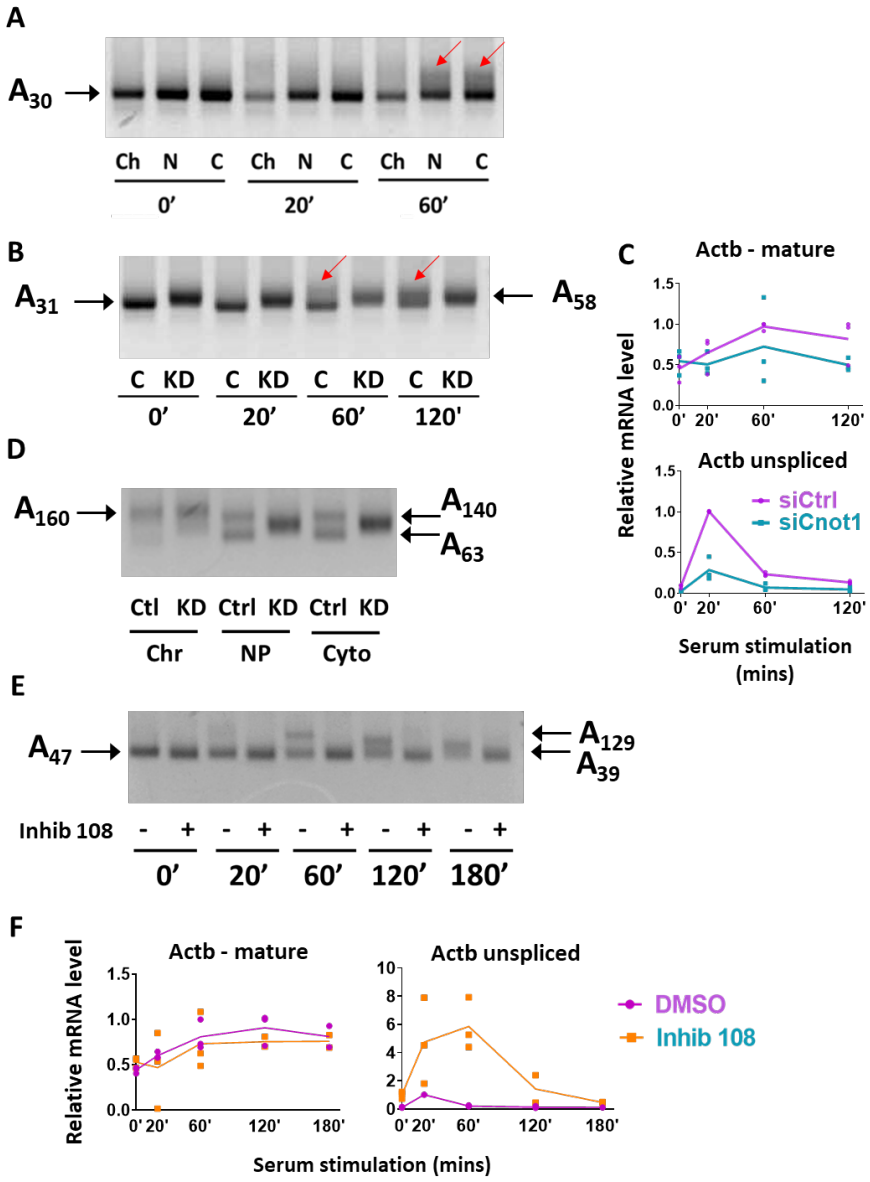


Figure 8.1 Collated Actb data showing instances of second bands in the PAT assay late in the serum response, as well as qPCR data showing transcriptional induction of Actb. **A)** RL2-PAT, using agarose gel stained with SYBR safe, of serum stimulated chromatin fractionation. Times indicate duration of serum stimulation (minutes). Black arrow indicates modal poly(A) length value and red arrows indicate increased smearing which may correspond to a second poorly resolved band. **B)** As in A, but for total RNA from control (C) and Cnot1 knockdown (KD) timecourses. **C)** qPCR data showing relative level of mature and unspliced Actb mRNA in 3 replicates of control and knockdown cells, of which the replicate shown in B was one. Relative expression was calculated using the ΔΔCt method with Gapdh as the reference gene, and normalising to the maximum value in the control set. **D)** As in A, but for control (C) and Cnot1 knockdown (KD) chromatin fractionation of cells serum stimulated for 50 minutes. Second bands here are more clearly resolved. *Note: bands in the KD lanes are between those in the control lanes.* **E)** As in A, but for total RNA from DMSO or CAF1 inhibitor treated cells. Again, the second band is more clearly resolved. **F)** As in C, but for 3 biological replicates of DMSO or CAF1 inhibitor treated cells, of which the replicate shown in E was one. *Note: bands in the Inhib 108 lanes are in line with the lower bands in the control lanes.*

Strikingly, in all cases the increases in tail length occurred as single discrete bands which were neither fully deadenylated nor untouched, and were too crisp to be the result of a reduced level of distributive activity (see figure 4.7 for illustration). This suggests either that CCR4-NOT could still perform some residual activity but was limited by some additional factor, or that PAN2-PAN3, which prefers long poly(A) tails, could also enter the nucleus and partially compensate for depletion of CCR4-NOT. Although PAN2-PAN3 is predominantly cytoplasmic, it was also shown to be capable of shuttling to the nucleus in NIH 3T3 cells (24, 26).

One unexpected phenomenon was the appearance of two bands in the PAT gel for nucleoplasmic and cytoplasmic Actb mRNA in control but not knockdown cells (figure 4.3). When chromatin fractionations were performed on control cells at different points after serum induction, Actb did not show two distinct bands in the nucleoplasm or cytoplasm at any point but did show much greater smearing in these fractions after 60 minutes. Two Actb bands were also visible using whole cell mRNA at 60 and 120 minutes post-serum stimulation (figure 6.3 A). Although mature Actb mRNA levels change only slightly following serum stimulation, Actb is transcriptionally induced (figure 5.1 A, B) and as such, the additional band could represent synthesis of a different 3'UTR or differently polyadenylated transcript. Alternatively, some switch in Actb poly(A) length control may take place at late time points. A summary of the relevant Actb data has been collated in figure 8.1.

CNOT1 knockdown in HEK293 cells yielded similar increases in nuclear poly(A) lengths at steady state (figures 4.4 A).

In addition to causing long poly(A) tails in NIH 3T3 cells, Cnot1 knockdown led to decreases in mature and pre-mRNA levels for serum induced genes as well as for Rpl28 (figure 5.1). Unspliced levels of Sqstm1 and Actb were also reduced but their mature mRNA levels were unaffected. The more substantial downregulation of Rpl28 compared to other housekeeping mRNAs is consistent with the observation that *ccr4* deletion in yeast caused a specific increase in the ratio of TFIIS to the Pol II subunit, Rpb3 (thought to indicate Pol II backtracking) along ribosomal protein genes (148).

A publicly available, unpublished CNOT1 knockdown RNA-Seq dataset from HeLa cells (621) was also examined but showed different effects on serum response mRNAs in different replicates, perhaps due to inconsistencies in growth conditions which may have seemed inconsequential to the original purposes of the study. CNOT1 knockdown caused more mature mRNAs to increase significantly in level than to decrease (figures 5.8 A, B, 5.9 A, B), consistent with its major role in mRNA decay. GO term analysis showed that these upregulated genes were enriched for a variety of processes including various signalling pathways, responses to stimuli and cellular adhesion (figure 5.8 C). By a slightly different method for generating raw count data, which ignored contributions from introns, the upregulated genes were again enriched for a variety of processes whereas the downregulated genes were enriched for involvement in mitosis (figure 5.9 C, D). Although analysis of this dataset highlighted the vast extent of the CCR4-NOT complex's influence, it was otherwise inconclusive.

The NIH 3T3 data indicate a coupling between poly(A) tail length and mRNA level such that mRNA homeostasis can be maintained, i.e. if the cell is unable to degrade the RNA as fast, not as much is produced to begin with. Other groups have shown, by deletion of degradation machinery, that yeast cells exhibit a similar buffering capacity (148, 579–581). Coupling of induction with poly(A) tail length in mammalian cells has been described in the literature, however, in that case, greater transcription rate was accompanied by long poly(A) tails (58).

To see whether this coupling was CCR4-NOT-specific or a general feature of depleting deadenylase activity, PAN2-PAN3 was also investigated. Only minimal effects on poly(A) length were observed following Pan2 depletion, though this could have been due to poor knockdown efficiency (figures 4.5, 4.6). In contrast to Cnot1 knockdown, both pre- and mature mRNA levels for serum response mRNAs were transiently increased in Pan2 knockdown, while levels of mature housekeeping mRNAs were unchanged (figure 5.12 A, B). Since knockdown was sufficient to cause detectable effects on mRNA level, it seems possible that differences in poly(A) length could have occurred, but were missed by the chosen time points. Return of serum response mRNAs to normal mature levels by 90 minutes post-stimulation suggests that PAN2-PAN3 depletion may have been compensated at later stages by

CCR4-NOT, and is in keeping with the idea that PAN2-PAN3 acts earlier in the cytoplasmic deadenylation pathway (26). How increases in pre-mRNA levels were elicited is unclear since we are unaware of any role for PAN2-PAN3 in transcription or splicing.

There are several mechanisms through which CCR4-NOT could mediate the connection between poly(A) length and pre-mRNA level. These mechanisms include direct nuclear effects as well as feedback loops caused by impaired cytoplasmic deadenylation which could, for example, result in reduced recycling of some mRNA-bound factor back to the nucleus, or in increased translation of some transcriptional effector. When considering possible mechanisms, it is useful to also hold in mind the temporal differences in polyadenylation and transcription rate of induced mRNAs that occur naturally during the serum response.

Western blots of fractionated NIH 3T3 cells showed that CNOT1 protein was present in both the nucleus and cytoplasm prior to serum stimulation, as well as at early and late time points after induction (figure 6.2 C, D). This suggests that nuclear deadenylation by CCR4-NOT as well as direct transcriptional effects are both realistic possibilities.

A simple model which incorporated CCR4-NOT involvement and could explain shorter poly(A) tails at the end of the serum response was that many of the mRNAs produced in the large initial burst of transcription are rapidly exported before extensive deadenylation by CCR4-NOT can occur; those exported later on may spend longer in the nucleus, allowing CCR4-NOT to remove more of the tail. To test this, siRNA knockdowns were performed for Nxf1, a factor thought to be required for bulk mRNA export (628, 629). If longer nuclear dwell time allowed more opportunity for deadenylation by CCR4-NOT then inhibiting nuclear export should have resulted in shorter poly(A) tails. Consistent with a central role in mRNA export, Nxf1 knockdown caused significant cell death and knockdown in the remaining cells was poor (figure 6.4 D). Nevertheless, this mild depletion had an obvious effect on poly(A) length of serum response mRNAs (figure 6.4 A, B). However, rather than having shorter poly(A) tails, serum response transcripts from the Nxf1 knockdown cells had longer tails with a more heterogenous distribution than those from the controls. Poly(A) tails of housekeeping mRNAs seemed unaffected (figure 6.4 A, B).

One explanation for this is that rather than causing extended nuclear deadenylation, nuclear retention protected these mRNAs from cytoplasmic deadenylation. It is difficult to fully dismiss a model of deadenylation-export competition since Nxf1 is involved relatively late in the export process (5, 210) and it is therefore possible that transcripts were still released from early association with CCR4-NOT but were then retained in the nucleus. Export may commandeer mRNAs from nuclear degradation pathways (231, 321). If nuclear retention was occurring, it would therefore be interesting to investigate how these transcripts escaped decay as only minimal changes in mRNA level were detected (figure 6.4 C).

The expectation was that depletion of Cnot1 would abrogate activity of the whole complex, and therefore that its knockdown may mimic conditions of low turnover, such during fasting or glucose starvation, when some CCR4-NOT subunits are downregulated (58, 60). Surprisingly, Cnot7/8 (Caf1) knockdown (which should remove both deadenylase activities from the complex since CAF1 also bridges the CCR4 interaction with the complex (516, 522–524)) yielded identical effects to Cnot1 knockdown on both total poly(A) length and on pre-mRNA level of serum response and Rpl28 mRNAs (figure 6.5 A, B). This suggested either that Cnot7/8 knockdown also led to disruption of the whole complex, or that loss of CCR4-NOT deadenylase activity was sufficient to drive a reduction in mRNA production. Given the long duration required for siRNA knockdown, it is also possible that reduction in transcription rate could have been caused indirectly by changes to the stability of mRNAs with transcriptional or signalling functions.

Treatment with a CAF1 (CNOT7/8) inhibitor for 2 hours caused an increase in poly(A) length for all of the serum-induced mRNAs tested, which unlike in the knockdowns was also visible after 20 minutes (figure 6.7 A, B). This suggests that the ~ 100 nt tails after 60 minutes in Cnot1 knockdown may have been mediated by residual CCR4-NOT activity rather than by some nuclear form of PAN2-PAN3; however, it is also possible that the long knockdown period led to upregulation of PAN2-PAN3, or that PAN2-PAN3 was also affected by the inhibitor in this experiment. Poly(A) tails of Rpl28 were unaffected by inhibitor treatment (figure 6.7 A, B), possibly due to the short treatment time and high stability of the mature mRNA meaning that newly made transcripts did not contribute markedly to the PAT signal. Intriguingly, Sqstm1

seemed to display slightly longer smeared bands in the untreated cells and Actb again exhibited a second band only in untreated cells (figure 6.7 A, B). The latter is incorporated into a summary of Actb data in figure 8.1. This time, the band in the deadenylation-inhibited lane of the Actb PAT was at the same length as the lower band in the control cells (rather than in between the two as in Cnot1 knockdown). It therefore seems likely that the lower band in the control lane represented a form of Actb which, like Rpl28, may have been stable enough that inhibition of deadenylation on transcripts made within the 2 hours of inhibitor treatment were not noticeable on the PAT. The additional band in the control cells was longest when it first appeared at 60 minutes and then gradually decreased (figure 8.1 E), suggesting it may have corresponded to some serum-induced form of Actb mRNA which was more extensively polyadenylated. The absence of this second band from both the Cnot1 knockdown and CAF1 inhibition experiments suggested that either its induction was impaired or that it was no longer polyadenylated distinctly from the other form.

Strikingly, CAF1 inhibition caused an increase in pre-mRNA level for all genes tested, whereas its effect on mature transcripts was variable (figure 6.7 C, D). It is not clear whether this represented an increase in transcription, a reduction in splicing or a combination of the two. For Fos and Fosb, delayed splicing (to varying degrees) seems most likely since there was a comparative lag in reaching the mature mRNA peak in the inhibitor set compared with the control. This explanation seems insufficient for Egr1, since both mature and unspliced levels increased. Reduced cytoplasmic degradation of mature transcripts due to CAF1 inhibition could have contributed to increased mature levels, however, it is unclear whether this reduced degradation alone could have accounted for a tenfold increase in unspliced levels. Since the CCR4 subunit may be involved in promoting transcription elongation (147, 148), one possibility is that inhibition of CAF1 (CNOT7/8) caused less CCR4-NOT to be engaged in deadenylase activity and to instead move to the nucleus and promote elongation. Excessive transcription could perhaps have saturated the splicing machinery and led to processing defects which may not have affected all mRNAs equally. Competition for splicing machinery may explain the extremely low levels of mature Fosb mRNA; following Actinomycin D addition in another experiment,

unspliced levels of Fosb decreased more slowly than for other transcripts (figure 8.2), suggesting Fosb may be more slowly spliced in these cells in general.

How differences in pre-mRNA level intersect with the shorter or missing bands in the inhibitor treated lanes of the Sqstm1 and Actb PATs is unclear. As the pre-mRNA

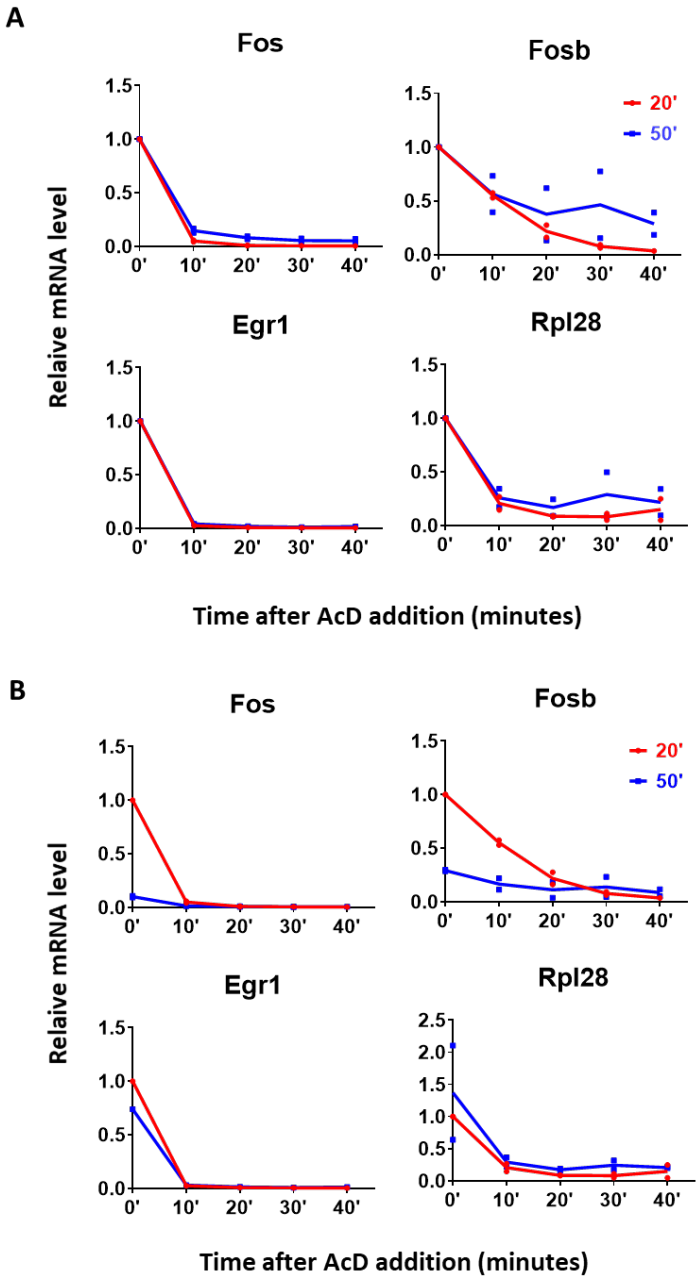


Figure 8.2 Pre-mRNA levels for most mRNAs decrease rapidly after actinomycin D treatment. qPCR data from 2 biological replicates (or one for Egr1) showing relative pre-mRNA levels following Actinomycin D treatment. Gapdh was used as a reference gene, and data were either **A)** normalised to the value at time of actinomycin treatment (i.e. after 20 or 50 minutes serum stimulation), or **B)** normalised to the value at the 20 minute time point.

level was increased for both genes following CAF1 inhibition (figure 6.7 D) but neither showed changes in mature level, it seems unlikely that failed induction at the level of transcription was responsible. Instead, it may be that slowed splicing caused a delay to polyadenylation for newly made mRNA. This cannot be a universal phenomenon, however, since serum-induced transcripts still received long tails.

Two possible factors could have been responsible for the different effects that CAF1 (CNOT7/8) inhibition had on mRNA level compared with Cnot1 or Cnot7/8 knockdown: integrity of the deadenylase module, and treatment time. Since CAF1 (CNOT7/8) bridges the interaction between CCR4 and the rest of the complex (516, 522–524), in the Cnot1 or Cnot7/8 knockdowns neither CAF1 nor CCR4 could theoretically have been present in the complex. That said, deletion of *ccr4* in yeast had a greater impact on bulk poly(A) tail length than deletion of *caf1* (28), suggesting that reliance of CCR4 on CAF1 for incorporation may not be absolute, or that CCR4 still has activity outside the CCR4-NOT complex, at least in yeast. Alternatively, deletion of *ccr4* alone could perhaps have allowed Ccr4-Not to still interact with poly(A) tails via Caf1, but to block access by Pan2-Pan3, potentially leading to effective protection of Pab1 (PABPC)-bound tails. During CAF1 inhibition, CCR4 activity should not have been significantly affected, though both subunits may be required for deadenylase activity (531, 594). It cannot be ruled out therefore that continued presence of CCR4 in the CCR4-NOT complex when CAF1 is inhibited may be significant for reasons unrelated to its deadenylase activity; for example, promotion of transcription elongation, which could be direct or could occur through stabilisation of the complex (147, 148). If presence of the CCR4 subunit was responsible for the differences observed between Cnot1/Cnot7/8 knockdown and CAF1 inhibition, it would point to the model mentioned earlier in which CCR4-NOT mediates crosstalk through its dual involvement in transcription and deadenylation. In conditions of reduced CCR4-NOT, both processes would be downregulated in a balanced manner, whereas, if deadenylation is inhibited but the complex is still intact, transcription may still be stimulated, uncoupling the two processes (figure 8.3). Furthermore, since the CAF1 inhibitor likely blocks RNA from entering the active site (594), this could have led to an increased cytoplasmic pool of unengaged CCR4-NOT. As in models of cytoplasmic mRNA sensing and transcript buffering (148, 580, 581), some of this free CCR4-NOT could cycle back to the nucleus to elicit

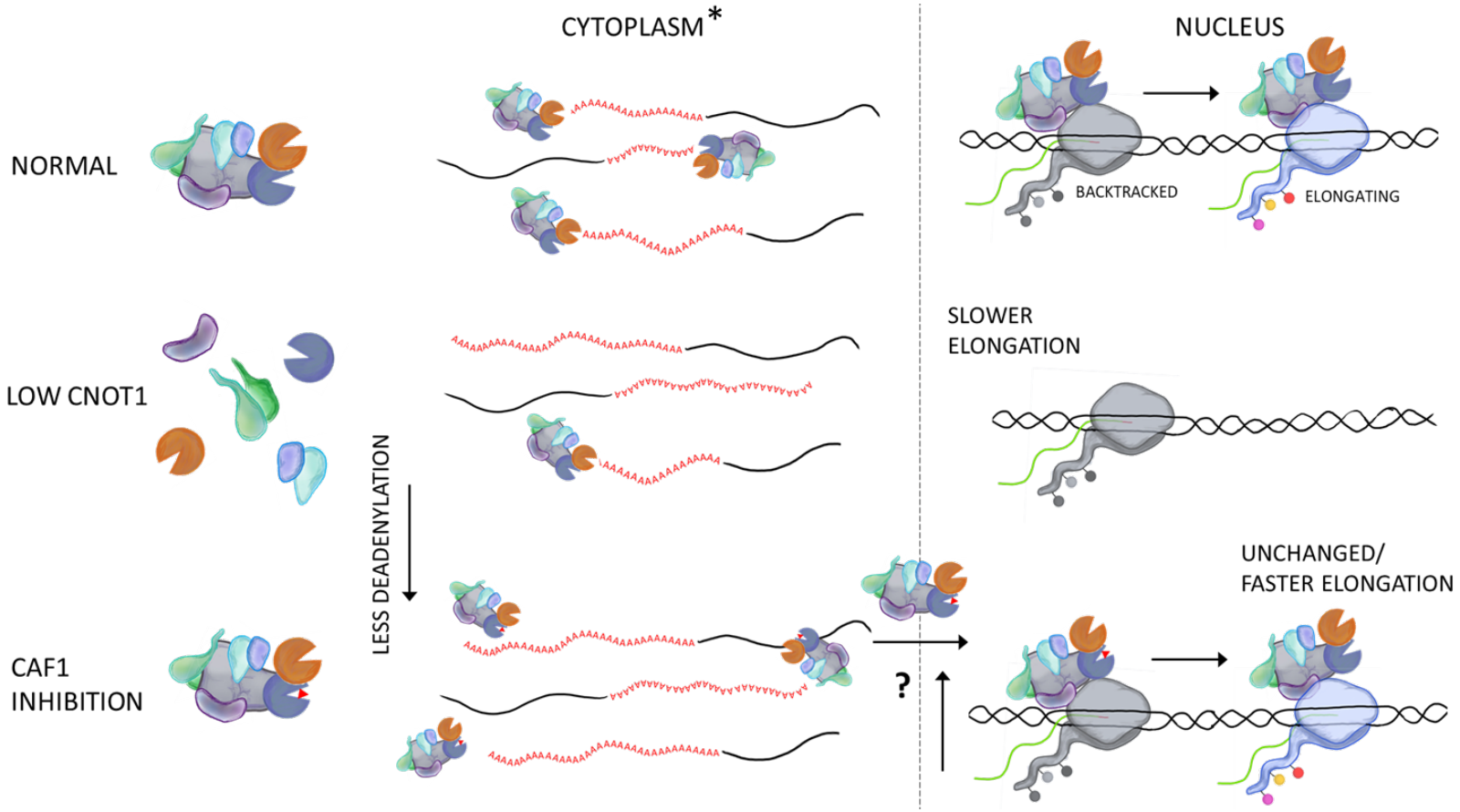


Figure 8.3 Model explaining uncoupling of transcription rate from poly(A) tail length. CCR4-NOT stimulates rescue of backtracked RNA Pol II in addition to functioning as a deadenylase. When Cnot1 or Cnot7/8 are depleted, both activities may be disrupted and mRNA homeostasis is maintained. When CAF1 (CNOT7/8) is inhibited, deadenylase activity is inhibited but stimulation of transcription can continue. In addition, excess unengaged CCR4-NOT may contribute further to promoting elongation. *In light of other work presented here, deadenylation may not be strictly cytoplasmic.

additional stimulation of transcription elongation, resulting in the abnormally high levels of unspliced transcripts observed.

Timings could also have played a role, in that the 72 hour treatment required by siRNA knockdown presents greater opportunity for adaptation/feedback mechanisms than the 2 hour inhibitor treatment. If different treatment times caused the different effects on transcript levels, this could point to a model in which CAF1 inhibition or depletion leads initially to stabilisation of transcripts which may encode transcriptional activators. These initially higher levels of transcription could perhaps induce a feedback loop to limit transcription which may become apparent within the timeframe of the knockdown experiments.

8.4 Concluding remarks

Gene expression was initially thought mainly to be enacted by transcriptional changes; however, it transpires that mRNA stability and translational efficiency also contribute substantially (298, 299, 584, 587). For example, while transcription is important for determining the level of the inflammatory response, mRNA stability is important for determining its duration (298). The presence and length of a poly(A) tail is often an integral determinant of these properties, and polyadenylation is therefore a useful subject of regulation for tuning gene expression. Furthermore, since polyadenylation of most mRNAs is connected with their synthesis, the poly(A) tail is perfectly placed to connect transcription with mRNA fate (52–55). Control of poly(A) length via deadenylation is therefore – unsurprisingly – also regulated, and examples exist both of gene-specific targeting, and of global shifts in mRNA turnover mediated by changes in deadenylase level (30, 45, 48, 58, 60, 62, 254, 289, 293, 430). Eukaryotic mRNA deadenylation was previously thought to be a cytoplasmic phenomenon, with recent examples of nuclear deadenylation eliciting imminent degradation (24–27, 33, 259). The work presented here in NIH 3T3 cells suggests that regulated nuclear deadenylation by the CCR4-NOT complex may be a widespread phenomenon (figures 3.5, 4.2, 4.3, 7.2, 7.4). In transient responses to stimuli, regulated nuclear deadenylation may benefit the cell by further sharpening the peak in mRNA levels produced by the transcriptional response. Specifically, in the serum response, induced transcripts produced early on receive longer poly(A) tails and are more stable than those produced late in the response (figures 3.5,

7.4 A). This could theoretically lead to a steeper drop in levels of translatable mRNA after transcription has peaked.

Although Cnot1 knockdown caused an increase in nuclear poly(A) length of all mRNAs tested after 60 minutes of serum stimulation (figure 4.2), some genes were affected with a greater magnitude than others, and this in general tracked with the maximum length usually observed during the serum response – that is, housekeeping mRNAs with consistently medium length nuclear tails were moderately affected by Cnot1 knockdown, whereas serum induced mRNAs experienced a much greater increase in poly(A) length. A similar pattern was observed in *Δccr4* yeast (582). When chromatin associated mRNA was assayed, Cnot1 knockdown caused a reversion to long (>~150 nt) poly(A) tails for all genes except Rpl28 (figure 4.3). A remaining question is what causes the gene-specific differences in poly(A) tail regulation, both in response to serum stimulation and following Cnot1 knockdown. Since evidence of specific targeting of CCR4-NOT exists in the cytoplasm (30, 45, 48, 254, 292), the possibility of a similar system for nuclear deadenylation was considered. The only potential specificity factor tested was TTP, which binds 3' UTR AREs to promote deadenylation and destabilisation of the containing transcript – often transcripts involved in the inflammatory or serum responses (51, 288, 290–296). Although a nuclear fraction was not isolated, PATs on total RNA showed only a minimal increase in poly(A) length of serum induced mRNAs following combined knockdown of TTP and its paralogues (figure 6.3 A) and this was after 20 rather than 60 minutes of stimulation. This could suggest that TTP is an unlikely candidate for delivering CCR4-NOT to these particular mRNAs in the nucleus, but it may also be that the knockdown was insufficient or could be compensated. A linked but alternative model for targeting may be that CCR4-NOT is recruited as default, but that in the case of long poly(A) tails, some other factor prevents its recruitment or inhibits its activity. Such a factor could be initially present in the nucleus but may become depleted by association with mRNAs exported early in the serum response. The reduced mRNA level which also occurs in Cnot1 knockdown could mean that, as well as less deadenylase activity being present, this protective factor may be depleted from the nucleus more slowly. Since initial poly(A) tail length seems to be scalable rather than simply long or short, one idea is that ratio of poly(A) binding proteins could be involved – indeed, relative

quantities of PABPs were suggested to be important for controlling poly(A) length in yeast (420). Following stimulation of cytoplasmic mRNA cleavage by the gamma-herpesvirus SOX nuclease, nuclear relocalisation of PABPC is accompanied by hyperadenylation and reduced transcription (415, 470, 471). It is easy to imagine that high nuclear concentrations of PABPC might inhibit early deadenylation by CCR4-NOT since high PABPC concentrations inhibited both Ccr4-Not nucleases *in vitro* (460). High concentrations of PABPC could perhaps protect mRNAs from nuclear deadenylation early in the serum response but fail to do so later on, after much PABPC has exited the nucleus bound to mRNA. Another possibility is that differences in transcription rate early and late in the response lead to differences in m^6A deposition as suggested by Slobodin et al (58) but that this affects nuclear rather than cytoplasmic deadenylation by CCR4-NOT.

Apparent exemption of Rpl28 from ever receiving a long chromatin associated tail, even in Cnot1 knockdown, may be suggestive of distinct regulation of a ribosomal protein or 5' TOP mRNAs. These mRNAs have been shown by others to possess unexpectedly short poly(A) tails and be disproportionately affected at the transcriptional and translational levels by CCR4-NOT (148, 325, 326, 557).

As well as mediating nuclear and chromatin-associated control of poly(A) length, the CCR4-NOT complex seems to couple poly(A) length with mRNA level to maintain mRNA homeostasis (figure 5.1). This is in keeping with evidence of transcript buffering in yeast and more recently in mammalian systems (58, 60, 148, 579–581, 591). How this crosstalk is mediated is unclear, however the simplest model is that the multifunctional nature of CCR4-NOT allows simultaneous control of mRNA synthesis and degradation. Further investigation of Pol II elongation in Cnot1 knockdown cells would be required to confirm this. Alternatively, new mRNA production may not be affected. Since Cnot1 knockdown caused an apparent increase in nuclear PABPN1 (figure 6.8 A, B), and PABPN1 was shown to be involved in nuclear degradation of polyadenylated RNAs (231, 321, 322) it may be that increased nuclear turnover rather than decreased transcription occurred in the knockdowns. This is consistent with the observation that most reductions in mRNA level resulting from Cnot1 knockdown occurred in the nucleoplasm and cytoplasm rather than on the chromatin (figure 5.4).

An attractive model, and a view that seems to be shared with others, is that CCR4-NOT acts to scale gene expression according to the physiological state of the cell (58, 60, 510). Changes in CCR4-NOT level have been shown to take place during B cell activation, following UVC irradiation, and according to nutrient availability (58, 60). It would be interesting to find out what other responses have this effect, and what determines an mRNA's susceptibility to this regulation. Liver-specific *Cnot1* knockout in mice showed diverse responses at the mRNA level which varied according to gene group, indicating that not all transcripts are equally affected, and that there may be variation in the mRNA substrates of the complex between tissues (591).

Given evidence of a role for Ccr4-Not in monitoring translation in yeast (44), it is tempting to consider the possibility of even greater coordination between levels of gene expression. Suggestive of such coordination is that inhibition of translation was previously shown to upregulate transcription of certain genes, while another group found that cycloheximide treatment caused a widespread reduction in mRNA synthesis (579, 650, 651). CCR4-NOT is directly involved in different levels of gene expression which largely occur in separate locations (26, 35, 36, 39, 42, 44, 147, 148, 460, 510, 526, 527, 547, 549, 550, 557), and varies in its interaction partners (including its CNOT4 subunit in mammals) (30, 40, 45, 48, 51, 254, 256, 289, 292, 515, 546). It may therefore be informative to investigate the localisation and composition of the CCR4-NOT complex under different conditions.

Overall this work has demonstrated the early and gene-specific definition of poly(A) tail length which could add even more finesse to the regulation of gene expression. The CCR4-NOT complex was identified as central to delivering this novel regulation. In addition, evidence is provided of transcript buffering, a phenomenon which had previously been described almost exclusively in yeast. This work has thereby uncovered another link in the vast interconnectedness of the mammalian cell.

References

1. Fuke,H. and Ohno,M. (2008) Role of poly (A) tail as an identity element for mRNA nuclear export. *Nucleic Acids Res.*, **36**, 1037–49.

2. 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.

3. Furuichi,Y., LaFiandra,A. and Shatkin,A.J. (1977) 5'-Terminal structure and mRNA stability. *Nature*, **266**, 235–239.

4. Galloway,A. and Cowling,V.H. (2019) mRNA cap regulation in mammalian cell function and fate. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1862**, 270–279.

5. Viphakone,N., Sudbery,I., Griffith,L., Heath,C.G., Sims,D. and Wilson,S.A. (2019) Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol. Cell*, **75**, 310–323.

6. Copeland,P.R. and Wormington,M. (2001) The mechanism and regulation of deadenylation: Identification and characterization of Xenopus PARN. *RNA*, **7**, 875–886.

7. Gowrishankar,G., Winzen,R., Dittrich-Breiholz,O., Redich,N., Kracht,M. and Holtmann,H. (2006) Inhibition of mRNA deadenylation and degradation by different types of cell stress. *Biol. Chem.*, **387**, 323–327.

8. Mazumder,B., Li,X. and Barik,S. (2010) Translation Control: A Multifaceted Regulator of Inflammatory Response. *J. Immunol.*, **184**, 3311–3319.

9. Cui,J., Sartain,C. V., Pleiss,J.A. and Wolfner,M.F. (2013) Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in Drosophila. *Dev. Biol.*, **383**, 121–131.

10. Barckmann,B. and Simonelig,M. (2013) Control of maternal mRNA stability in germ cells and early embryos. *Biochim. Biophys. Acta*, **1829**, 714–724.

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

12. Zheng,X., Yang,P., Lackford,B., Bennett,B.D., Wang,L., Li,H., Wang,Y., Miao,Y., Foley,J.F., Fargo,D.C., *et al.* (2016) CNOT3-Dependent mRNA Deadenylation Safeguards the Pluripotent State. *Stem Cell Reports*, **7**, 897–910.

13. Lim,J., Lee,M., Son,A., Chang,H. and Kim,V.N. (2016) mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev.*, **30**, 1671–82.

14. Yamaguchi,T., Suzuki,T., Sato,T., Takahashi,A., Watanabe,H., Kadowaki,A., Natsui,M., Inagaki,H., Arakawa,S., Nakaoka,S., *et al.* (2018) The CCR4-NOT deadenylase complex controls atg7-dependent cell death and heart function. *Sci. Signal.*, **11**, eaan3638.

15. Faraji,F., Hu,Y., Yang,H.H., Lee,M.P., Winkler,G.S., Hafner,M. and Hunter,K.W. (2016) Post-transcriptional Control of Tumor Cell Autonomous Metastatic Potential by CCR4-NOT Deadenylase CNOT7. *PLoS Genet.*, **12**, e1005820.

16. Rha,J., Jones,S.K., Fidler,J., Banerjee,A., Leung,S.W., Morris,K.J., Wong,J.C., Inglis,G.A.S., Shapiro,L., Deng,Q., *et al.* (2017) The RNA-binding protein, ZC3H14, is required for proper poly(A) tail length control, expression of synaptic proteins, and brain function in mice. *Hum. Mol. Genet.*, **26**, 3663–3681.

17. Gruber,A.J. and Zavolan,M. (2019) Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.*, **20**, 599–614.

18. Parras,A., de Diego-Garcia,L., Alves,M., Beamer,E., Conte,G., Jimenez-Mateos,E.M., Morgan,J., Ollà,I., Hernandez-Santana,Y., Delanty,N., *et al.* (2020) Polyadenylation of mRNA as a novel regulatory mechanism of gene expression in temporal lobe epilepsy. *Brain*, **143**, 2139–2153.

19. Gallie,D.R. (1998) A tale of two termini: A functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation. *Gene*, **216**, 1–11.

20. 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.

21. 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**, 104–13.

22. Archer,S.K., Shirokikh,N.E., Hallwirth,C. V, Beilharz,T.H. and Preiss,T. (2015) Probing the closed-loop model of mRNA translation in living cells. *RNA Biol.*, **12**, 248–54.

23. Brawerman,G. (1981) The role of the poly(a) sequence in mammalian messenger RNA. *Crit. Rev. Biochem. Mol. Biol.*, **10**, 1–38.

24. Uchida,N., Hoshino,S.I. and Katada,T. (2004) Identification of a Human Cytoplasmic Poly(A) Nuclease Complex Stimulated by Poly(A)-binding Protein. *J. Biol. Chem.*, **279**, 1383–1391.

25. Parker,R. and Song,H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.*, **11**, 121–127.

26. Yamashita,A., Chang,T.C., Yamashita,Y., Zhu,W., Zhong,Z., Chen,C.Y.A. and Shyu,A. Bin (2005) Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.*, **12**, 1054–1063.

27. Wahle,E. and Winkler,G.S. (2013) RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1829**, 561–570.

28. Webster,M.W., Chen,Y.H., Stowell,J.A.W., Alhusaini,N., Sweet,T., Graveley,B.R., Coller,J. and Passmore,L.A. (2018) mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol. Cell*, **70**, 1089–1100.

29. Lee,D., Ohn,T., Chiang,Y.C., Quigley,G., Yao,G., Liu,Y. and Denis,C.L. (2010) PUF3 acceleration of deadenylation in vivo can operate independently of CCR4 activity, possibly involving effects on the PAB1-mRNP structure. *J. Mol. Biol.*, **399**, 562–575.

30. Hanet,A., Räsch,F., Weber,R., Ruscica,V., Fauser,M., Raisch,T., Kuzuoğlu-Öztürk,D., Chang,C.-T., Bhandari,D., Igreja,C., *et al.* (2019) HELZ directly interacts with CCR4–NOT and causes decay of bound mRNAs. *Life Sci. Alliance*, **2**, e201900405.

31. Raisch,T., Chang,C.-T., Levdansky,Y., Muthukumar,S., Raunser,S. and Valkov,E. (2019) Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation. *Nat. Commun.*, **10**, 3173.

32. Webster,M.W., Stowell,J.A. and Passmore,L.A. (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife*, **8**, e40670.

33. Kordyukova,M., Sokolova,O., Morgunova,V., Ryazansky,S., Akulenko,N., Glukhov,S. and Kalmykova,A. (2020) Nuclear Ccr4-Not mediates the degradation of telomeric and transposon transcripts at chromatin in the Drosophila germline. *Nucleic Acids Res.*, **48**, 141–156.

34. Badarinarayana,V., Chiang,Y.C. and Denis,C.L. (2000) Functional interaction of CCR4-NOT proteins with TATAA-binding protein (TBP) and its associated factors in yeast. *Genetics*,

155, 1045–54.

35. Zwartjes,C.G.M., Jayne,S., Van Den Berg,D.L.C. and Timmers,H.T.M. (2004) Repression of Promoter Activity by CNOT2, a Subunit of the Transcription Regulatory Ccr4-Not Complex. *J. Biol. Chem.*, **279**, 10848–10854.

36. Winkler,G.S., Mulder,K.W., Bardwell,V.J., Kalkhoven,E. and Timmers,H.T.M. (2006) Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J.*, **25**, 3089–3099.

37. Jayne,S., Zwartjes,C.G.M., van Schaik,F.M.A. and Timmers,H.T.M. (2006) Involvement of the SMRT/NCoR-HDAC3 complex in transcriptional repression by the CNOT2 subunit of the human Ccr4-Not complex. *Biochem. J.*, **398**, 461–7.

38. Garapaty,S., Mahajan,M.A. and Samuels,H.H. (2008) Components of the CCR4-NOT complex function as nuclear hormone receptor coactivators via association with the NRC-interacting Factor NIF-1. *J. Biol. Chem.*, **283**, 6806–16.

39. Gaillard,H., Tous,C., Botet,J., González-Aguilera,C., Quintero,M.J., Viladevall,L., García-Rubio,M.L., Rodríguez-Gil,A., Marín,A., Ariño,J., *et al.* (2009) Genome-Wide Analysis of Factors Affecting Transcription Elongation and DNA Repair: A New Role for PAF and Ccr4-Not in Transcription-Coupled Repair. *PLoS Genet.*, **5**, e1000364.

40. Loh,B., Jonas,S. and Izaurralde,E. (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.*, **27**, 2125–38.

41. Laribee,R.N., Hosni-Ahmed,A., Workman,J.J. and Chen,H. (2015) Ccr4-not regulates RNA polymerase I transcription and couples nutrient signaling to the control of ribosomal RNA biogenesis. *PLoS Genet.*, **11**, e1005113.

42. Rodríguez-Gil,A., Ritter,O., Saul,V. V., Wilhelm,J., Yang,C.-Y., Grosschedl,R., Imai,Y., Kuba,K., Kracht,M. and Schmitz,M.L. (2017) The CCR4-NOT complex contributes to repression of Major Histocompatibility Complex class II transcription. *Sci. Rep.*, **7**, 3547.

43. Jiang,H., Wolgast,M., Beebe,L.M. and Reese,J.C. (2019) Ccr4–not maintains genomic integrity by controlling the ubiquitylation and degradation of arrested RNAPII. *Genes Dev.*, **33**, 705–717.

44. Buschauer,R., Matsuo,Y., Sugiyama,T., Chen,Y.H., Alhusaini,N., Sweet,T., Ikeuchi,K., Cheng,J., Matsuki,Y., Nobuta,R., *et al.* (2020) The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science (80-.).*, **368**, eaay6912.

45. Leppek,K., Schott,J., Reitter,S., Poetz,F., Hammond,M.C. and Stoecklin,G. (2013) Roquin promotes constitutive mrna decay via a conserved class of stem-loop recognition motifs. *Cell*, **153**, 869–881.

46. Collart,M.A., Panasenko,O.O. and Nikolaev,S.I. (2013) The Not3/5 subunit of the Ccr4-Not complex: A central regulator of gene expression that integrates signals between the cytoplasm and the nucleus in eukaryotic cells. *Cell. Signal.*, **25**, 743–751.

47. Du,H., Zhao,Y., He,J., Zhang,Y., Xi,H., Liu,M., Ma,J. and Wu,L. (2016) YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.*, **7**, 1–11.

48. Raisch,T., Bhandari,D., Sabath,K., Helms,S., Valkov,E., Weichenrieder,O. and Izaurralde,E. (2016) Distinct modes of recruitment of the CCR4-NOT complex by Drosophila and vertebrate Nanos. *EMBO J.*, **35**, 974–90.

49. Zhu,L., Kandasamy,S.K., Liao,S.E. and Fukunaga,R. (2018) LOTUS domain protein MARF1 binds CCR4-NOT deadenylase complex to post-transcriptionally regulate gene expression in oocytes. *Nat. Commun.*, **9**.

50. Yi,H., Park,J., Ha,M., Lim,J., Chang,H. and Kim,V.N. (2018) PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol. Cell*, **70**, 1081–1088.

51. Bulbrook,D., Brazier,H., Mahajan,P., Kliszczak,M., Fedorov,O., Marchese,F.P., Aubareda,A., Chalk,R., Picaud,S., Strain-Damerell,C., *et al.* (2018) Tryptophan-Mediated Interactions between Tristetraprolin and the CNOT9 Subunit Are Required for CCR4-NOT Deadenylase Complex Recruitment. *J. Mol. Biol.*, **430**, 722–736.

52. Bienroth,S., Keller,W., Wahle,E. and Keller,W. (1993) Assembly of a processive messenger RNA polyadenylation complex. *EMBO J.*, **12**, 585–594.

53. Manley,J.L. and Hirose,Y. (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, **395**, 93–96.

54. Bear,D.G., Fomproix,N., Soop,T., Björkroth,B., Masich,S. and Daneholt,B. (2003) Nuclear poly(A)-binding protein PABPN1 is associated with RNA polymerase II during transcription and accompanies the released transcript to the nuclear pore. *Exp. Cell Res.*, **286**, 332–44.

55. Kerwitz,Y., Kühn,U., Lilie,H., Knoth,A., Scheuermann,T., Friedrich,H., Schwarz,E. and Wahle,E. (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J.*, **22**, 3705–14.

56. Kühn,U., Gündel,M., Knoth,A., Kerwitz,Y., Rüdel,S. and Wahle,E. (2009) Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J. Biol. Chem.*, **284**, 22803–14.

57. Chen,C.-Y.A. and Shyu,A.-B. (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA*, **2**, 167–83.

58. Slobodin,B., Bahat,A., Sehrawat,U., Becker-Herman,S., Zuckerman,B., Weiss,A.N., Han,R., Elkon,R., Agami,R., Ulitsky,I., *et al.* (2020) Transcription Dynamics Regulate Poly(A) Tails and Expression of the RNA Degradation Machinery to Balance mRNA Levels. *Mol. Cell*, **78**, 434-444.e5.

59. Bond,G.L., Prives,C. and Manley,J.L. (2000) Poly(A) Polymerase Phosphorylation Is Dependent on Novel Interactions with Cyclins. *Mol. Cell. Biol.*, **20**, 5310–5320.

60. Morita,M., Oike,Y., Nagashima,T., Kadomatsu,T., Tabata,M., Suzuki,T., Nakamura,T., Yoshida,N., Okada,M. and Yamamoto,T. (2011) Obesity resistance and increased hepatic expression of catabolism-related mRNAs in Cnot3+/- mice. *EMBO J.*, **30**, 4678–4691.

61. Rodriguez-Gil,A., Ritter,O., Hornung,J., Stekman,H., Kröger,M., Braun,T., Kremmer,E., Kracht,M. and Lienhard Schmitz,M. (2016) HIPK family kinases bind and regulate the function of the CCR4-NOT complex. *Mol. Biol. Cell*, **27**, 1969–1980.

62. Park,J.-E., Yi,H., Kim,Y., Chang,H. and Kim,V.N. (2016) Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol. Cell*, **62**, 462–471.

63. Colgan,D.F., Murthy,K.G.K., Prives,C. and Manley,J.L. (1996) Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature*, **384**, 282–285.

64. Lee,S.-H., Choi,H.-S., Kim,H. and Lee,Y. (2008) ERK is a novel regulatory kinase for poly(A) polymerase. *Nucleic Acids Res.*, **36**, 803–813.

65. Laishram,R.S. and Anderson,R.A. (2010) The poly A polymerase Star-PAP controls 3-2-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. *EMBO J.*, **29**, 4132–4145.

66. Kojima,S., Sher-Chen,E.L. and Green,C.B. (2012) Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev.*, **26**, 2724–36.

67. Lin,H.-H., Lo,Y.-L., Wang,W.-C., Huang,K.-Y., I,K.-Y. and Chang,G.-W. (2020) Overexpression of FAM46A, a Non-canonical Poly(A) Polymerase, Promotes Hemin-Induced Hemoglobinization in K562 Cells. *Front. Cell Dev. Biol*, **8**, 414.

68. Green,C.B. and Besharse,J.C. (1996) Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 14884–14888.

69. Garbarino-Pico,E., Niu,S., Rollag,M.D., Strayer,C.A., Besharse,J.C. and Green,C.B. (2007) Immediate early response of the circadian polyA ribonuclease nocturnin to two extracellular stimuli. *RNA*, **13**, 745–755.

70. Niu,S., Shingle,D.L., Garbarino-Pico,E., Kojima,S., Gilbert,M. and Green,C.B. (2011) The circadian deadenylase nocturnin is necessary for stabilization of the iNOS mRNA in mice. *PLoS One*, **6**, e26954.

71. Conaway,R.C. and Conaway,J.W. (1997) General Transcription Factors for RNA Polymerase II. In *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, Vol. 56, pp. 327–346.

72. Yan,C., Dodd,T., He,Y., Tainer,J.A., Tsutakawa,S.E. and Ivanov,I. (2019) Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.*, **26**, 397–406.

73. Bartkowiak,B., Mackellar,A.L. and Greenleaf,A.L. (2011) Updating the CTD Story: From Tail to Epic. *SAGE-Hindawi Access to Res. Genet. Res. Int.*, **2011**, 623718–16.

74. Nilson,K.A., Guo,J., Turek,M.E., Brogie,J.E., Delaney,E., Luse,D.S. and Price,D.H. (2015) THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol. Cell*, **59**, 576–587.

75. Fitzgerald,M. and Shenk,T. (1981) The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*, **24**, 251–260.

76. Connelly,S. and Manley,J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.*, **2**, 440–452.

77. Eaton,J.D., Francis,L., Davidson,L. and West,S. (2020) A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.*, **34**, 132–145.

78. Chan,S.L., Huppertz,I., Yao,C., Weng,L., Moresco,J.J., Yates,J.R., Ule,J., Manley,J.L. and Shi,Y. (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev.*, **28**, 2370–2380.

79. Schönemann,L., Kühn,U., Martin,G., Schäfer,P., Gruber,A.R., Keller,W., Zavolan,M. and Wahle,E. (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.*, **28**, 2381–93.

80. Sun,Y., Zhang,Y., Hamilton,K., Manley,J.L., Shi,Y., Walz,T. and Tong,L. (2017) Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc. Natl. Acad. Sci.*, **115**, E1419–E1428.

81. Hilleren,P., McCarthy,T., Rosbash,M., Parker,R. and Jensen,T.H. (2001) Quality control of mRNA 3′-end processing is linked to the nuclear exosome. *Nature*, **413**, 538–542.

82. Milligan,L., Torchet,C., Allmang,C., Shipman,T. and Tollervey,D. (2005) A Nuclear Surveillance Pathway for mRNAs with Defective Polyadenylation. *Mol. Cell. Biol.*, **25**, 9996–10004.

83. Bousquet-Antonelli,C., Presutti,C. and Tollervey,D. (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell*, **102**, 765–775.

84. Kutter,C., Watt,S., Stefflova,K., Wilson,M.D., Goncalves,A., Ponting,C.P., Odom,D.T. and Marques,A.C. (2012) Rapid Turnover of Long Noncoding RNAs and the Evolution of Gene Expression. *PLoS Genet.*, **8**, e1002841.

85. Davidson,L., Kerr,A. and West,S. (2012) Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J.*, **31**, 2566–2578.

86. Mure,F., Corbin,A., Benbahouche,N.E.H., Bertrand,E., Manet,E. and Gruffat,H. (2018) The splicing factor SRSF3 is functionally connected to the nuclear RNA exosome for intronless mRNA decay. *Sci. Rep.*, **8**, 12901–15.

87. Jimeno-González,S., Haaning,L.L., Malagon,F. and Jensen,T.H. (2010) The Yeast 5'-3' Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II. *Mol. Cell*, **37**, 580–587.

88. Mendell,J.T., Sharifi,N.A., Meyers,J.L., Martinez-Murillo,F. and Dietz,H.C. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.*, **36**, 1073–1078.

89. Melero,R., Buchwald,G., Castaño,R., Raabe,M., Gil,D., Lázaro,M., Urlaub,H., Conti,E. and Llorca,O. (2012) The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3'2 end. *Nat. Struct. Mol. Biol.*, **19**, 498–505.

90. Van Hoof,A., Frischmeyer,P.A., Dietz,H.C. and Parker,R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science (80-.).*, **295**, 2262–2264.

91. Doma,M.K. and Parker,R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561–564.

92. Schmid,M. and Jensen,T.H. (2018) Controlling nuclear RNA levels. *Nat. Rev. Genet.*, **19**, 518–529.

93. Carter,R. and Drouin,G. (2009) Structural differentiation of the three eukaryotic RNA polymerases. *Genomics*, **94**, 388–396.

94. Bunch,H. (2018) Gene regulation of mammalian long non-coding RNA. *Mol. Genet. Genomics*, **293**, 1–15.

95. Hsin,J.-P. and Manley,J.L. (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.*, **26**, 2119–2137.

96. McCracken,S., Fong,N., Yankulov,K., Ballantyne,S., Pan,G., Greenblatt,J., Patterson,S.D., Wickens,M. and Bentley,D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357–361.

97. Rosonina,E. and Blencowe,B.J. (2004) Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *RNA*, **10**, 581–589.

98. Doamekpor,S.K., Sanchez,A.M., Schwer,B., Shuman,S. and Lima,C.D. (2014) How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes Dev.*, **28**, 1323–1336.

99. Guo,Y.E., Manteiga,J.C., Henninger,J.E., Sabari,B.R., Dall'Agnese,A., Hannett,N.M., Spille,J.H., Afeyan,L.K., Zamudio,A. V., Shrinivas,K., *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, **572**.

100. Egloff,S. and Murphy,S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet.*, **24**, 280–288.

101. Nag,A., Narsinh,K. and Martinson,H.G. (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat. Struct. Mol. Biol.*, **14**, 662–669.

102. McCracken,S., Fong,N., Rosonina,E., Yankulov,K., Brothers,G., Siderovski,D., Hessel,A., Foster,S., Shuman,S. and Bentley,D.L. (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes*

Dev., **11**, 3306–3318.

103. Gu,B., Eick,D. and Bensaude,O. (2013) CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.*, **41**, 1591–1603.

104. Davidson,L., Muniz,L. and West,S. (2014) 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev.*, **28**, 342–356.

105. Wong,K.H., Jin,Y. and Struhl,K. (2014) TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol. Cell*, **54**, 601–612.

106. Zhang,Z., Fu,J. and Gilmour,D.S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev.*, **19**, 1572–1580.

107. Bentley,D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–175.

108. Feng,Z., Chen,X., Wu,X. and Zhang,M. (2019) Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *J. Biol. Chem.*, **294**, 14823–14835.

109. Cho,W.K., Spille,J.H., Hecht,M., Lee,C., Li,C., Grube,V. and Cisse,I.I. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (80-.).*, **361**, 412–415.

110. Sabari,B.R., Dall'Agnese,A., Boija,A., Klein,I.A., Coffey,E.L., Shrinivas,K., Abraham,B.J., Hannett,N.M., Zamudio,A. V., Manteiga,J.C., *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).*, **361**, eaar3958.

111. Hengartner,C.J., Myer,V.E., Liao,S.M., Wilson,C.J., Koh,S.S. and Young,R.A. (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell*, **2**, 43–53.

112. Shiekhattar,R., Mermelstein,F., Fisher,R.P., Drapkin,R., Brian,D., Wessling,H.C., Morgan,D.O. and Reinberg,D. (1995) Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, **374**, 283–287.

113. Glover-Cutter,K., Larochelle,S., Erickson,B., Zhang,C., Shokat,K., Fisher,R.P. and Bentley,D.L. (2009) TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Mol. Cell. Biol.*, **29**, 5455–5464.

114. Sampathi,S., Acharya,P., Zhao,Y., Wang,J., Stengel,K.R., Liu,Q., Savona,M.R. and Hiebert,S.W. (2019) The CDK7 inhibitor THZ1 alters RNA polymerase dynamics at the 5' and 3' ends of genes. *Nucleic Acids Res.*, **47**, 3921–3936.

115. Bartkowiak,B., Liu,P., Phatnani,H.P., Fuda,N.J., Cooper,J.J., Price,D.H., Adelman,K., Lis,J.T. and Greenleaf,A.L. (2010) CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.*, **24**, 2303–2316.

116. Tellier,M., Zaborowska,J., Caizzi,L., Mohammad,E., Velychko,T., Schwalb,B., Ferrer-Vicens,I., Blears,D., Nojima,T., Cramer,P., *et al.* (2020) CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Res.*, **48**, 7712–7727.

117. Tietjen,J.R., Zhang,D.W., Rodríguez-Molina,J.B., White,B.E., Akhtar,M.S., Heidemann,M., Li,X., Chapman,R.D., Shokat,K., Keles,S., *et al.* (2010) Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.*, **17**, 1154–1161.

118. Hsin,J.-P., Xiang,K. and Manley,J.L. (2014) Function and Control of RNA Polymerase II C-Terminal Domain Phosphorylation in Vertebrate Transcription and RNA Processing. *Mol. Cell. Biol.*, **34**, 2488–2498.

119. Buratowski,S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol Cell*, **36**, 541–546.

120. Nojima,T., Gomes,T., Grosso,A.R.F., Kimura,H., Dye,M.J., Dhir,S., Carmo-Fonseca,M. and Proudfoot,N.J. (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, **161**, 526–540.

121. Drexler,H.L., Choquet,K. and Churchman,L.S. (2019) Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol. Cell*, **77**, 985–998.

122. Hsin,J.P., Sheth,A. and Manley,J.L. (2011) RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science (80-.).*, **334**, 683–686.

123. Nemec,C.M., Yang,F., Gilmore,J.M., Hintermair,C., Ho,Y.H., Tseng,S.C., Heidemann,M., Zhang,Y., Florens,L., Gasch,A.P., *et al.* (2017) Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E3944–E3953.

124. Harlen,K.M., Trotta,K.L., Smith,E.E., Mosaheb,M.M., Fuchs,S.M., Stirling,L. and Correspondence,C. (2016) Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *CellReports*, **15**, 2147–2158.

125. Stiller,J.W., McConaughy,B.L. and Hall,B.D. (2000) Evolutionary complementation for polymerase II CTD function. *Yeast*, **16**, 57–64.

126. Beggs,S., James,T.C. and Bond,U. (2012) The PolyA tail length of yeast histone mRNAs varies during the cell cycle and is influenced by Sen1p and Rrp6p. *Nucleic Acids Res.*, **40**, 2700–11.

127. Mayer,A., Heidemann,M., Lidschreiber,M., Schreieck,A., Sun,M., Hintermair,C., Kremmer,E., Eick,D. and Cramer,P. (2012) CTD Tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science (80-.).*, **336**, 1723–1725.

128. Descostes,N., Heidemann,M., Spinelli,L., Schüller,R., Maqbool,M.A., Fenouil,R., Koch,F., Innocenti,C., Gut,M., Gut,I., *et al.* (2014) Tyrosine phosphorylation of RNA Polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife*, **2014**, e02105.

129. Hsin,J.P., Li,W., Hoque,M., Tian,B. and Manley,J.L. (2014) RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife*, **2014**.

130. Soutourina,J. (2018) Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.*, **19**, 262–274.

131. Yamaguchi,Y., Shibata,H. and Handa,H. (2013) Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1829**, 98–104.

132. Izaurralde,E., Lewis,J., McGuigan,C., Jankowska,M., Darzynkiewicz,E. and Mattaj,I.W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, **78**, 657–668.

133. Patzelt,E., Thalmann,E., Hartmuth,K., Blass,D. and Kuechler,E. (1987) Assembly of pre-mRNA splicing complex is cap dependent. *Nucleic Acids Res.*, **15**, 1387–1399.

134. Ashfield,R., Patel,A.J., Bossone,S.A., Brown,H., Campbell,R.D., Marcu,K.B. and Proudfoot,N.J. (1994) MAZ-dependent termination between closely spaced human complement genes. *EMBO J.*, **13**, 5656–5667.

135. Mellon,I., Spivak,G. and Hanawalt,P.C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*, **51**, 241–249.

136. Yonaha,M. and Proudfoot,N.J. (1999) Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol. Cell*, **3**, 593–600.

137. Alexander,R.D., Innocente,S.A., Barrass,J.D. and Beggs,J.D. (2010) Splicing-Dependent RNA polymerase pausing in yeast. *Mol. Cell*, **40**, 582–593.

138. Cheung,A.C.M. and Cramer,P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, **471**, 249–253.

139. 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**, 950–3.

140. Saldi,T., Cortazar,M.A., Sheridan,R.M. and Bentley,D.L. (2016) Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J. Mol. Biol.*, **428**, 2623–2635.

141. Wang,W., Walmacq,C., Chong,J., Kashlev,M. and Wang,D. (2018) Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E2538–E2545.

142. Cortazar,M.A., Sheridan,R.M., Erickson,B., Fong,N., Glover-Cutter,K., Brannan,K. and Bentley,D.L. (2019) Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Mol. Cell*, **76**, 896–908.

143. Dutta,A., Babbarwal,V., Fu,J., Brunke-Reese,D., Libert,D.M., Willis,J. and Reese,J.C. (2015) Ccr4-Not and TFIIS Function Cooperatively To Rescue Arrested RNA Polymerase II. *Mol. Cell. Biol.*, **35**, 1915–1925.

144. Sheridan,R.M., Fong,N., D'Alessandro,A. and Bentley,D.L. (2019) Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5′ Pause Release, Termination, and Transcription Elongation Rate. *Mol. Cell*, **73**, 107–118.

145. Beaudenon,S.L., Huacani,M.R., Wang,G., McDonnell,D.P. and Huibregtse,J.M. (1999) Rsp5 Ubiquitin-Protein Ligase Mediates DNA Damage-Induced Degradation of the Large Subunit of RNA Polymerase II in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **19**, 6972–6979.

146. Lee,K.B., Wang,D., Lippard,S.J. and Sharp,P.A. (2002) Transcription-coupled and DNA damage-dependent ubiquitination of RNA polymerase II in vitro. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 4239–4244.

147. Kruk,J.A., Dutta,A., Fu,J., Gilmour,D.S. and Reese,J.C. (2011) The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.*, **25**, 581–593.

148. Begley,V., Corzo,D., Jordán-Pla,A., Cuevas-Bermúdez,A., Miguel-Jiménez,L. de, Pérez-Aguado,D., Machuca-Ostos,M., Navarro,F., Chávez,M.J., Pérez-Ortín,J.E., *et al.* (2019) The mRNA degradation factor Xrn1 regulates transcription elongation in parallel to Ccr4. *Nucleic Acids Res.*, **47**, 9524–9541.

149. 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.

150. Rasmussen,E.B. and Lis,J.T. (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 7923–7.

151. Jove,R. and Manley,J.L. (1984) In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *J. Biol. Chem.*, **259**, 8513–8521.

152. 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.*, **18**, 637–650.

153. Anna,A. and Monika,G. (2018) Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.*, **59**, 253–268.

154. Wang,Y., Liu,J., Huang,B., XuU,Y.-M., Li,J., Huang,L.-F., Lin,J., Zhang,J., Min,Q.-H., Yang,W.-M., *et al.* (2015) Mechanism of alternative splicing and its regulation. *Biomed. Reports*, **3**, 152–158.

155. Berglund,J.A., Abovich,N. and Rosbash,M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.*, **12**, 858–867.

156. Hesselberth,J.R. (2013) Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA*, **4**, 677–691.

157. Morgan,J.T., Fink,G.R. and Bartel,D.P. (2019) Excised linear introns regulate growth in yeast. *Nature*, **565**, 606–611.

158. Guo,M., Lo,P.C. and Mount,S.M. (1993) Species-specific signals for the splicing of a short Drosophila intron in vitro. *Mol. Cell. Biol.*, **13**, 1104–1118.

159. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–4.

160. Fox-Walsh,K.L., Dou,Y., Lam,B.J., Hung,S.-P., Baldi,P.F. and Hertel,K.J. (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing.

161. Dye,M.J. and Proudfoot,N.J. (1999) Terminal Exon Definition Occurs Cotranscriptionally and Promotes Termination of RNA Polymerase II. *Mol. Cell*, **3**, 371–378.

162. Vagner,S., Vagner,C. and Mattaj,I.W. (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3′-end processing and splicing. *Genes Dev.*, **14**, 403–13.

163. Rappsilber,J., Ryder,U., Lamond,A.I. and Mann,M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, **12**, 1231–1245.

164. Rigo,F. and Martinson,H.G. (2008) Functional Coupling of Last-Intron Splicing and 3′-End Processing to Transcription In Vitro: the Poly(A) Signal Couples to Splicing before Committing to Cleavage. *Mol. Cell. Biol.*, **28**, 849–862.

165. Proudfoot,N.J. and Brownlee,G.G. (1976) 3′ non-coding region sequences in eukaryotic messenger RNA. *Nature*, **263**, 211–4.

166. Tian,B., Hu,J., Zhang,H. and Lutz,C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.

167. Gil,A. and Proudfoot,N.J. (1984) A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. *Nature*, **312**, 473–474.

168. Rüegsegger,U., Beyer,K. and Keller,W. (1996) Purification and characterization of human cleavage factor Im involved in the 3′ end processing of messenger RNA precursors. *J. Biol. Chem.*, **271**, 6107–6113.

169. Hart,R.R., Mcdevitt,M.A. and Nevins,J.R. (1985) Poly(A) Site Cleavage in a HeLa Nuclear Extract Is Dependent on Downstream Sequences. *Cell*, **43**, 677–663.

170. Conway,L. and Wickens,M. (1985) A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3′ termini in frog oocytes (mRNA processing/polyadenylylation). *Biochemistry*, **82**, 3949–3953.

171. Darmon,S.K. and Lutz,C.S. (2012) Novel upstream and downstream sequence elements contribute to polyadenylation efficiency. *RNA Biol.*, **9**, 1255–65.

172. Clerici,M., Faini,M., Muckenfuss,L.M., Aebersold,R. and Jinek,M. (2018) Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat. Struct. Mol. Biol.*, **25**, 135–138.

173. MacDonald,C.C., Wilusz,J. and Shenk,T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol. Cell. Biol.*, **14**, 6647–6654.

174. Beyer,K., Dandekar,T. and Keller,W. (1997) RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J. Biol. Chem.*, **272**, 26769–26779.

175. Kaufmann,I., Martin,G., Friedlein,A., Langen,H. and Keller,W. (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J.*, **23**, 616–26.

176. Brown,K.M. and Gilmartin,G.M. (2003) A Mechanism for the Regulation of Pre-mRNA 3 Processing by Human Cleavage Factor I m.

177. Ryan,K., Calvo,O. and Manley,J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, **10**, 565–73.

178. Winters,M.A. and Edmonds,M. (1973) A Poly (A) Polymerase from Calf Thymus. *J. Biol. Chem.*, **248**, 4763–4768.

179. Sheets,M.D. and Wickens,M. (1989) Two phases in the addition of a poly(A) tail. *Genes Dev.*, **3**, 1401–12.

180. Martin,G. and Keller,W. (1996) Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and a catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J.*, **15**, 2593–2603.

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

182. Logan,J., Falck-Pedersen,E., Darnell,J.E. and Shenk,T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci. U. S. A.*, **84**, 8306–8310.

183. Dengl,S. and Cramer,P. (2009) Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in Vitro. *J. Biol. Chem.*, **284**, 21270–21279.

184. Osheim,Y.N., Proudfoot,N.J. and Beyer,A.L. (1999) EM visualization of transcription by RNA polymerase II: Downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell*, **3**, 379–387.

185. Osheim,Y.N., Sikes,M.L. and Beyer,A.L. (2002) EM visualization of Pol II genes in Drosophila: Most genes terminate without prior 3 end cleavage of nascent transcripts. *Chromosoma*, **111**, 1–12.

186. West,S., Gromak,N. and Proudfoot,N.J. (2004) Human 5' → 3' exonuclease Xm2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522–525.

187. Fong,N., Brannan,K., Erickson,B., Kim,H., Cortazar,M.A., Sheridan,R.M., Nguyen,T., Karp,S. and Bentley,D.L. (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell*, **60**, 256–267.

188. Sansó,M., Levin,R.S., Lipp,J.J., Wang,V.Y.-F., Greifenberg,A.K., Quezada,E.M., Ali,A., Ghosh,A., Larochelle,S., Rana,T.M., *et al.* (2016) P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev.*, **30**, 117–131.

189. Eaton,J.D., Davidson,L., Bauer,D.L.V., Natsume,T., Kanemaki,M.T. and West,S. (2018) Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity.

Genes Dev., **32**, 127–139.

190. Plotch,S.J., Bouloy,M., Ulmanen,I. and Krug,R.M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, **23**, 847–858.

191. Bauer,D.L.V., Tellier,M., Martínez-Alonso,M., Nojima,T., Proudfoot,N.J., Murphy,S. and Fodor,E. (2018) Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Rep.*, **23**, 2119–2129.

192. Shi,Y., Di Giammartino,D.C., Taylor,D., Sarkeshik,A., Rice,W.J., Yates,J.R., Frank,J. and Manley,J.L. (2009) Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. *Mol. Cell*, **33**, 365–376.

193. Licatalosi,D.D., Geiger,G., Minet,M., Schroeder,S., Cilli,K., McNeil,J.B. and Bentley,D.L. (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol. Cell*, **9**, 1101–1111.

194. Meinhart,A. and Cramer,P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, **430**, 223–226.

195. Volanakis,A., Kamieniarz-Gdula,K., Schlackow,M. and Proudfoot,N.J. (2017) Wnk1 kinase and the termination factor PCF11 connect nuclear mRNA export with transcription. *Genes Dev.*, **31**, 2175–2185.

196. Zhang,Z. and Gilmour,D.S. (2006) Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell*, **21**, 65–74.

197. Kamieniarz-Gdula,K., Gdula,M.R., Panser,K., Nojima,T., Monks,J., Wiśniewski,J.R., Riepsaame,J., Brockdorff,N., Pauli,A. and Proudfoot,N.J. (2019) Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. *Mol. Cell*, **74**, 158–172.

198. Li,Y., Junod,S.L., Ruba,A., Kelich,J.M. and Yang,W. (2019) Nuclear export of mRNA molecules studied by SPEED microscopy. *Methods*, **153**, 46–62.

199. Williams,T., Ngo,L.H. and Wickramasinghe,V.O. (2018) Nuclear export of RNA: Different sizes, shapes and functions. *Semin. Cell Dev. Biol.*, **75**, 70–77.

200. Speese,S.D., Ashley,J., Jokhi,V., Nunnari,J., Barria,R., Li,Y., Ataman,B., Koon,A., Chang,Y.T., Li,Q., *et al.* (2012) Nuclear envelope budding enables large ribonucleoprotein particle export during synaptic Wnt signaling. *Cell*, **149**, 832–846.

201. 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.

202. Ruepp,M.D., Aringhieri,C., Vivarelli,S., Cardinale,S., Paro,S., Schümperli,D. and Barabino,S.M.L. (2009) Mammalian pre-mRNA 3' end processing factor CF Im68 functions in mRNA export. *Mol. Biol. Cell*, **20**, 5211–5223.

203. Chi,B., Wang,Q., Wu,G., Tan,M., Wang,L., Shi,M., Chang,X. and Cheng,H. (2013) Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.*, **41**, 1294–1306.

204. 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.

205. Viphakone,N., Hautbergue,G.M., Walsh,M., Chang,C. Te, Holland,A., Folco,E.G., Reed,R. and Wilson,S.A. (2012) TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat. Commun.*, **3**, 1006.

206. Fribourg,S., Braun,I.C., Izaurralde,E. and Conti,E. (2001) Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. *Mol. Cell*, **8**, 645–656.

207. Heath,C.G., Viphakone,N. and Wilson,S.A. (2016) The role of TREX in gene expression and disease. *Biochem. J.*, **473**, 2911–2935.

208. Stutz,F., Bachi,A., Doerks,T., Braun,I.C., Séraphin,B., Wilm,M., Bork,P. and Izaurralde,E. (2000) REF, an evolutionarily conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA*, **6**, 638–650.

209. Stäßer,K., Masuda,S., Mason,P., Pfannstiel,J., Oppizzi,M., Rodriguez-Navarro,S., Rondón,A.G., Aguilera,A., Struhl,K., Reed,R., *et al.* (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, **417**, 304–308.

210. Hautbergue,G.M., Hung,M.L., Golovanov,A.P., Lian,L.Y. and Wilson,S.A. (2008) Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 5154–5159.

211. Hautbergue,G.M., Hung,M.L., Walsh,M.J., Snijders,A.P.L., Chang,C. Te, Jones,R., Ponting,C.P., Dickman,M.J. and Wilson,S.A. (2009) UIF, a New mRNA Export Adaptor that Works Together with REF/ALY, Requires FACT for Recruitment to mRNA. *Curr. Biol.*, **19**, 1918–1924.

212. Viphakone,N., Cumberbatch,M.G., Livingstone,M.J., Heath,P.R., Dickman,M.J., Catto,J.W. and Wilson,S.A. (2015) Luzp4 defines a new mRNA export pathway in cancer cells. *Nucleic Acids Res.*, **43**, 2353–2366.

213. Luo,M.J., Zhou,Z., Magni,K., Christoforides,C., Rappsilber,J., Mann,M. and Reed,R. (2001) Pre-mrna splicing and mRNA export linked by direct interactions between UAP56 and aly. *Nature*, **413**, 644–647.

214. Taniguchi,I. and Ohno,M. (2008) ATP-Dependent Recruitment of Export Factor Aly/REF onto Intronless mRNAs by RNA Helicase UAP56. *Mol. Cell. Biol.*, **28**, 601–608.

215. Dufu,K., Livingstone,M.J., Seebacher,J., Gygi,S.P., Wilson,S.A. and Reed,R. (2010) ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes Dev.*, **24**, 2043–2053.

216. Li,Y., Lin,A.W., Zhang,X., Wang,Y., Wang,X. and Goodrich,D.W. (2007) Cancer cells and normal cells differ in their requirements for Thoc1. *Cancer Res.*, **67**, 6657–6664.

217. Mancini,A., Niemann-Seyde,S.C., Pankow,R., El Bounkari,O., Klebba-Färber,S., Koch,A., Jaworska,E., Spooncer,E., Gruber,A.D., Whetton,A.D., *et al.* (2010) THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. *BMC Biol.*, **8**, 1.

218. Rehwinkel,J., Herold,A., Gari,K., Köcher,T., Rode,M., Ciccarelli,F.L., Wilm,M. and Izaurralde,E. (2004) Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. *Nat. Struct. Mol. Biol.*, **11**, 558–566.

219. Katahira,J., Inoue,H., Hurt,E. and Yoneda,Y. (2009) Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *EMBO J.*, **28**, 556–567.

220. Guria,A., Tran,D.D.H., Ramachandran,S., Koch,A., Bounkari,O.E.L., Dutta,P., Hauser,H. and Tamura,T. (2011) Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. *RNA*, **17**, 1048–1056.

221. Tran,D.D., Saran,S., Dittrich-Breiholz,O., Williamson,A.J., Klebba-Färber,S., Koch,A., Kracht,M., Whetton,A.D. and Tamura,T. (2013) Transcriptional regulation of immediate-early gene response by THOC5, a member of mRNA export complex, contributes to the M-CSF-induced macrophage differentiation. *Cell Death Dis.*, **4**, e879.

222. Kumar,R., Gardner,A., Homan,C.C., Douglas,E., Mefford,H., Wieczorek,D., Lüdecke,H.-J., Stark,Z., Sadedin,S., Nowak,C.B., *et al.* (2018) Severe neurocognitive and growth disorders due to variation in *THOC2* , an essential component of nuclear mRNA export machinery. *Hum. Mutat.*, **39**, 1126–1138.

223. Kumar,R., Corbett,M.A., Van Bon,B.W.M., Woenig,J.A., Weir,L., Douglas,E., Friend,K.L., Gardner,A., Shaw,M., Jolly,L.A., *et al.* (2015) THOC2 Mutations Implicate mRNA-Export Pathway in X-Linked Intellectual Disability. *Am. J. Hum. Genet.*, **97**, 302–310.

224. Huertas,P. and Aguilera,A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell*, **12**, 711–721.

225. 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.

226. Abruzzi,K.C., Lacadie,S. and Rosbash,M. (2004) Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J.*, **23**, 2620–2631.

227. Masuda,S., Das,R., Cheng,H., Hurt,E., Dorman,N. and Reed,R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.*, **19**, 1512–1517.

228. Fleckner,J., Zhang,M., Valcárcel,J. and Green,M.R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.*, **11**, 1864–1872.

229. Gatfield,D., Le Hir,H., Schmitt,C., Braun,I.C., Köcher,T., Wilm,M. and Izaurralde,E. (2001) The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr. Biol.*, **11**, 1716–1721.

230. Macmorris,M., Brocker,C. and Blumenthal,T. (2003) UAP56 levels affect viability and mRNA export in Caenorhabditis elegans. *RNA*, **9**, 847–857.

231. Silla,T., Karadoulama,E., Mąkosa,D., Lubas,M. and Jensen,T.H. (2018) The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep.*, **23**, 2199–2210.

232. Morris,K.J. and Corbett,A.H. (2018) The polyadenosine RNA-binding protein ZC3H14 interacts with the THO complex and coordinately regulates the processing of neuronal transcripts. *Nucleic Acids Res.*, **46**, 6561–6575.

233. Uranishi,H., Zolotukhin,A.S., Lindtner,S., Warming,S., Zhang,G.M., Bear,J., Copeland,N.G., Jenkins,N.A., Pavlakis,G.N. and Felber,B.K. (2009) The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1. *J. Biol. Chem.*, **284**, 26106–26116.

234. Lesbirel,S., Viphakone,N., Parker,M., Parker,J., Heath,C., Sudbery,I. and Wilson,S.A. (2018) The m6A-methylase complex recruits TREX and regulates mRNA export. *Sci. Rep.*, **8**, 13827.

235. Katahira,J., Okuzaki,D., Inoue,H., Yoneda,Y., Maehara,K. and Ohkawa,Y. (2013) Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. *Nucleic Acids Res.*, **41**, 7060–7072.

236. Zuckerman,B., Ron,M., Mikl,M., Segal,E. and Ulitsky,I. (2020) Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex. *Mol. Cell*, **79**, 251–267.

237. Müller-McNicoll,M., Botti,V., de Jesus Domingues,A.M., Brandl,H., Schwich,O.D., Steiner,M.C., Curk,T., Poser,I., Zarnack,K. and Neugebauer,K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.*, **30**, 553–566.

238. 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.

239. Grüter,P., Tabernero,C., Von Kobbe,C., Schmitt,C., Saavedra,C., Bachi,A., Wilm,M., Felber,B.K. and Izaurralde,E. (1998) TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell*, **1**, 649–659.

240. Culjkovic-Kraljacic,B. and Borden,K.L.B. (2013) Aiding and abetting cancer: MRNA export and the nuclear pore. *Trends Cell Biol.*, **23**, 328–335.

241. Hutten,S. and Kehlenbach,R.H. (2007) CRM1-mediated nuclear export: to the pore and beyond. *Trends Cell Biol.*, **17**, 193–201.

242. Kimura,T., Hashimoto,I., Nagase,T. and Fujisawa,J.I. (2004) CRM1-dependent, but not ARE-mediated, nuclear export of IFN-α1 mRNA. *J. Cell Sci.*, **117**, 2259–2270.

243. Culjkovic,B., Topisirovic,I., Skrabanek,L., Ruiz-Gutierrez,M. and Borden,K.L.B. (2005) eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'UTR. *J. Cell Biol.*, **169**, 245–256.

244. Culjkovic,B., Topisirovic,I., Skrabanek,L., Ruiz-Gutierrez,M. and Borden,K.L.B. (2006) eIF4E is a central node of an RNA regulon that governs cellular proliferation. *J. Cell Biol.*, **175**, 415–426.

245. Fan,X.C. and Steitz,J.A. (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J.*, **17**, 3448–3460.

246. Brennan,C.M., Gallouzi,I.E. and Steitz,J.A. (2000) Protein ligands to HuR modulate its interaction with target mRNAs in vivo. *J. Cell Biol.*, **151**, 1–13.

247. Jang,B.C., Muñoz-Najar,U., Paik,J.H., Claffey,K., Yoshida,M. and Hla,T. (2003) Leptomycin B, an inhibitor of the nuclear export receptor CRM1, inhibits COX-2 expression. *J. Biol. Chem.*, **278**, 2773–2776.

248. Gallouzi,I.E., Brennan,C.M. and Steitz,J.A. (2001) Protein ligands mediate the CRM1-dependent export of HuR in response to heat shock. *RNA*, **7**, 1348–1361.

249. Brennan,C.M. and Steitz,J.A. (2001) HuR and mRNA stability. *Cell. Mol. Life Sci.*, **58**, 266–277.

250. Eisen,T.J., Eichhorn,S.W., Subtelny,A.O., Lin,K.S., McGeary,S.E., Gupta,S. and Bartel,D.P. (2020) The Dynamics of Cytoplasmic mRNA Metabolism. *Mol. Cell*, **77**, 786–799.

251. Shaw,G. and Kamen,R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659–667.

252. Ford,L.P., Watson,J., Keene,J.D. and Wilusz,J. (1999) ELAV proteins stabilize deadenylated intermediates in a novel in vitro mRNA deadenylation/degradation system. *Genes Dev.*, **13**, 188–201.

253. Gherzi,R., Lee,K.-Y., Briata,P., Wegmüller,D., Moroni,C., Karin,M. and Chen,C.-Y. (2004) A KH Domain RNA Binding Protein, KSRP, Promotes ARE-Directed mRNA Turnover by Recruiting the Degradation Machinery. *Mol. Cell*, **14**, 571–583.

254. Van Etten,J., Schagat,T.L., Hrit,J., Weidmann,C.A., Brumbaugh,J., Coon,J.J. and Goldstrohm,A.C. (2012) Human pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.*, **287**, 36370–36383.

255. Brooks,S.A. and Blackshear,P.J. (2013) Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta*, **1829**, 666–79.

256. 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.

257. Yokoshi,M., Li,Q., Yamamoto,M., Okada,H., Suzuki,Y. and Kawahara,Y. (2014) Direct Binding of Ataxin-2 to Distinct Elements in 3′ UTRs Promotes mRNA Stability and Protein Expression. *Mol. Cell*, **55**, 186–198.

258. Kuzuoğlu-Öztürk,D., Bhandari,D., Huntzinger,E., Fauser,M., Helms,S. and Izaurralde,E. (2016) mi RISC and the CCR 4– NOT complex silence mRNA targets independently of 43S ribosomal scanning . *EMBO J.*, **35**, 1186–1203.

259. Sarshad,A.A., Juan,A.H., Correa,A.I., Haase,A.D., Sartorelli,V., Correspondence,M.H., Correa Muler,A.I., Anastasakis,D.G., Wang,X., Genzor,P., *et al.* (2018) Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol. Cell*, **71**, 1040–1050.

260. Geisberg,J. V., Moqtaderi,Z., Fan,X., Ozsolak,F. and Struhl,K. (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*, **156**, 812–824.

261. Wu,X. and Bartel,D.P. (2017) Widespread Influence of 3′-End Structures on Mammalian mRNA Processing and Stability. *Cell*, **169**, 905–917.

262. Moqtaderi,Z., Geisberg,J. V. and Struhl,K. (2018) Extensive Structural Differences of Closely Related 3′ mRNA Isoforms: Links to Pab1 Binding and mRNA Stability. *Mol. Cell*, **72**, 849–861.

263. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G., *et al.* (2014) N 6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117–120.

264. Arango,D., Sturgill,D., Alhusaini,N., Dillman,A.A., Sweet,T.J., Hanson,G., Hosogane,M., Sinclair,W.R., Nanan,K.K., Mandler,M.D., *et al.* (2018) Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*, **175**, 1872–1886.

265. 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.

266. Chen,C.-Y.A. and Shyu,A.-B. (2003) Rapid Deadenylation Triggered by a Nonsense Codon Precedes Decay of the RNA Body in a Mammalian Cytoplasmic Nonsense-Mediated Decay Pathway. *Mol. Cell. Biol.*, **23**, 4805–4813.

267. Garneau,N.L., Wilusz,J. and Wilusz,C.J. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **8**, 113–126.

268. Kurosaki,T., Popp,M.W. and Maquat,L.E. (2019) Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **20**, 406–420.

269. Sheiness,D., Puckett,L. and Darnell,J.E. (1975) Possible Relationship of Poly(A) Shortening to mRNA Turnover. *Proc. Natl. Acad. Sci.*, **72**, 1077–1081.

270. Zheng,D., Ezzeddine,N., Chen,C.Y.A., Zhu,W., He,X. and Shyu,A. Bin (2008) Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. *J. Cell Biol.*, **182**, 89–101.

271. Lee,Y.J. and Glaunsinger,B.A. (2009) Aberrant Herpesvirus-Induced Polyadenylation Correlates With Cellular Messenger RNA Destruction. *PLoS Biol.*, **7**, e1000107.

272. Brengues,M., Teixeira,D. and Parker,R. (2005) Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science (80-.).*, **310**, 486–489.

273. Bhattacharyya,S.N., Habermacher,R., Martine,U., Closs,E.I. and Filipowicz,W. (2006) Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell*, **125**, 111–1124.

274. Räsch,F., Weber,R., Izaurralde,E. and Igreja,C. (2020) 4E-T-bound mRNAs are stored in a silenced and deadenylated form. *Genes Dev.*, **34**, 847–860.

275. 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.

276. Baer,B.W. and Kornberg,R.D. (1983) The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *J. Cell Biol.*, **96**, 717–21.

277. Smith,B.L., Gallie,D.R., Le,H. and Hansma,P.K. (1997) Visualization of poly(A)-binding protein complex formation with poly(A) RNA using atomic force microscopy. *J. Struct. Biol.*, **119**, 109–117.

278. Chowdhury,A., Mukhopadhyay,J. and Tharun,S. (2007) The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA*, **13**, 998–1016.

279. Song,M.-G., Li,Y. and Kiledjian,M. (2010) Multiple mRNA decapping enzymes in mammalian cells. *Mol. Cell*, **40**, 423–32.

280. Parker,R., Aebi,M., Kirchner,G., Chen,J.Y., Vijayraghavan,U., Jacobson,A., Alexandrov,A., Chernyakov,I., Gu,W., Hiley,S.L., *et al.* (2012) RNA degradation in Saccharomyces cerevisae. *Genetics*, **191**, 671–702.

281. Tuck,A.C., Rankova,A., Arpat,A.B., Liechti,L.A., Hess,D., Iesmantavicius,V., Castelo-Szekely,V., Gatfield,D. and Bühler,M. (2020) Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Mol. Cell*, **77**, 1222–1236.

282. Bouveret,E., Rigaut,G., Shevchenko,A., Wilm,M. and Séraphin,B. (2000) A Sm-like protein complex that participates in mRNA degradation. *EMBO J.*, **19**, 1661–1671.

283. Coller,J.M., Tucker,M., Sheth,U., Valencia-Sanchez,M.A. and Parker,R. (2001) The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA*, **7**, 1717–1727.

284. Ozgur,S., Chekulaeva,M. and Stoecklin,G. (2010) Human Pat1b Connects Deadenylation with mRNA Decapping and Controls the Assembly of Processing Bodies. *Mol. Cell. Biol.*, **30**, 4308–4323.

285. Wu,D., Muhlrad,D., Bowler,M.W., Jiang,S., Liu,Z., Parker,R. and Song,H. (2013) Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res.*, **24**, 233–246.

286. Alhusaini,N. and Coller,J. (2016) The deadenylase components Not2p, Not3p, and Not5p promote mRNA decapping. *RNA*, **22**, 709–721.

287. Braun,J.E., Truffault,V., Boland,A., Huntzinger,E., Chang,C. Te, Haas,G., Weichenrieder,O., Coles,M. and Izaurralde,E. (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. *Nat. Struct. Mol. Biol.*, **19**, 1324–1331.

288. Chen,C.Y.A. and Shyu,A. Bin (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.*, **20**, 465–470.

289. 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.

290. Blackshear,P.J. (2002) Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. In *Biochemical Society Transactions*. Portland Press, Vol. 30, pp. 945–952.

291. Lykke-Andersen,J. and Wagner,E. (2005) Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1.

Genes Dev., **19**, 351–61.

292. Fabian,M.R., Frank,F., Rouya,C., Siddiqui,N., Lai,W.S., Karetnikov,A., Blackshear,P.J., Nagar,B. and Sonenberg,N. (2013) Structural basis for the recruitment of the human CCR4–NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.*, **20**, 735–739.

293. Lai,W.S., Carballo,E., Strum,J.R., Kennington,E.A., Phillips,R.S. and Blackshear,P.J. (1999) Evidence that Tristetraprolin Binds to AU-Rich Elements and Promotes the Deadenylation and Destabilization of Tumor Necrosis Factor Alpha mRNA. *Mol. Cell. Biol.*, **19**, 4311–4323.

294. Ogilvie,R.L., Abelson,M., Hau,H.H., Vlasova,I., Blackshear,P.J. and Bohjanen,P.R. (2005) Tristetraprolin Down-Regulates IL-2 Gene Expression through AU-Rich Element-Mediated mRNA Decay . *J. Immunol.*, **174**, 953–961.

295. Tiedje,C., Diaz-Muñoz,M.D., Trulley,P., Ahlfors,H., Laaß,K., Blackshear,P.J., Turner,M. and Gaestel,M. (2016) The RNA-binding protein TTP is a global post-transcriptional regulator of feedback control in inflammation. *Nucleic Acids Res.*, **44**, 7418–40.

296. Wilson,T. and Treisman,R. (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3′ AU-rich sequences. *Nature*, **336**, 396–399.

297. Phillips,R.S., Ramos,S.B.V. and Blackshear,P.J. (2002) Members of the tristetraprolin family of tandem CCCH zinc finger proteins exhibit CRM1-dependent nucleocytoplasmic shuttling. *J. Biol. Chem.*, **277**, 11606–11613.

298. Hao,S. and Baltimore,D. (2009) The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.*, **10**, 281–288.

299. Elkon,R., Zlotorynski,E., Zeller,K.I. and Agami,R. (2010) Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics*, **11**, 259.

300. Presnyak,V., Alhusaini,N., Chen,Y.H., Martin,S., Morris,N., Kline,N., Olson,S., Weinberg,D., Baker,K.E., Graveley,B.R., *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.

301. Radhakrishnan,A., Chen,Y.H., Martin,S., Alhusaini,N., Green,R. and Coller,J. (2016) The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell*, **167**, 122–132.

302. Wu,Q., Medina,S.G., Kushawah,G., Devore,M.L., Castellano,L.A., Hand,J.M., Wright,M. and Bazzini,A.A. (2019) Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife*, **8**, e45396.

303. Hoekema,A., Kastelein,R.A., Vasser,M. and de Boer,H.A. (1987) Codon replacement in the PGK1 gene of Saccharomyces cerevisiae: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.*, **7**, 2914–2924.

304. Andreev,D.E., O'connor,P.B.F., Loughran,G., Dmitriev,S.E., Baranov,P. V and Shatsky,I.N. (2016) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.*, **45**, 513–526.

305. Meijer,H.A., Schmidt,T., Gillen,S.L., Langlais,C., Jukes-Jones,R., de Moor,C.H., Cain,K., Wilczynska,A. and Bushell,M. (2019) DEAD-box helicase eIF4A2 inhibits CNOT7 deadenylation activity. *Nucleic Acids Res.*, **47**, 8224–8238.

306. Mugridge,J.S., Coller,J. and Gross,J.D. (2018) Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. *Nat. Struct. Mol. Biol.*, **25**, 1077–1085.

307. Preissler,S., Reuther,J., Koch,M., Scior,A., Bruderek,M., Frickey,T. and Deuerling,E. (2015) Not4-dependent translational repression is important for cellular protein homeostasis in yeast. *EMBO J.*, **34**, 1905–1924.

308. Collart,M.A. (2016) The Ccr4-Not complex is a key regulator of eukaryotic gene expression.

Wiley Interdiscip. Rev. RNA, **7**, 438–54.

309. Preker,P., Nielsen,J., Kammler,S., Lykke-Andersen,S., Christensen,M.S., Mapendano,C.K., Schierup,M.H. and Jensen,T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science (80-.).*, **322**, 1851–1854.

310. Davidson,L., Francis,L., Cordiner,R.A., Eaton,J.D., Estell,C., Macias,S., Cáceres,J.F. and West,S. (2019) Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. *Cell Rep.*, **26**, 2779–2791.

311. Lykke-Andersen,S., Žumer,K., Molska,E.Š., Rouvière,J.O., Wu,G., Demel,C., Schwalb,B., Schmid,M., Cramer,P. and Jensen,T.H. (2020) Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell*, **81**, 514–529.

312. Bergeron,D., Pal,G., Beaulieu,Y.B., Chabot,B. and Bachand,F. (2015) Regulated Intron Retention and Nuclear Pre-mRNA Decay Contribute to PABPN1 Autoregulation . *Mol. Cell. Biol.*, **35**, 2503–2517.

313. Yap,K., Lim,Z.Q., Khandelia,P., Friedman,B. and Makeyev,E. V. (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.*, **26**, 1209–1223.

314. Garland,W., Comet,I., Wu,M., Radzisheuskaya,A., Rib,L., Vitting-Seerup,K., Lloret-Llinares,M., Sandelin,A., Helin,K. and Jensen,T.H. (2019) A Functional Link between Nuclear RNA Decay and Transcriptional Control Mediated by the Polycomb Repressive Complex 2. *Cell Rep.*, **29**, 1800–1811.

315. Wyers,F., Rougemaille,M., Badis,G., Rousselle,J.C., Dufour,M.E., Boulay,J., Régnault,B., Devaux,F., Namane,A., Séraphin,B., *et al.* (2005) Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, **121**, 725–737.

316. Lykke-Andersen,S., Brodersen,D.E. and Jensen,T.H. (2009) Origins and activities of the eukaryotic exosome. *J. Cell Sci.*, **122**, 1487–1494.

317. Drązkowska,K., Tomecki,R., Stoduś,K., Kowalska,K., Czarnocki-Cieciura,M. and Dziembowski,A. (2013) The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro. *Nucleic Acids Res.*, **41**, 3845–3858.

318. Garland,W. and Jensen,T.H. (2020) Nuclear sorting of RNA. *Wiley Interdiscip. Rev. RNA*, **11**, e1572.

319. Wu,G., Schmid,M., Rib,L., Polak,P., Meola,N., Sandelin,A., Heick,T. and Correspondence,J. (2020) A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome. **30**, 2387–2401.

320. Mattout,A., Gaidatzis,D., Padeken,J., Schmid,C.D., Aeschimann,F., Kalck,V. and Gasser,S.M. (2020) LSM2-8 and XRN-2 contribute to the silencing of H3K27me3-marked genes through targeted RNA decay. *Nat. Cell Biol.*, **22**, 579–590.

321. Bresson,S.M. and Conrad,N.K. (2013) The Human Nuclear Poly(A)-Binding Protein Promotes RNA Hyperadenylation and Decay. *PLoS Genet.*, **9**, e1003893.

322. Meola,N., Domanski,M., Karadoulama,E., Chen,Y., Gentil,C., Pultz,D., Vitting-Seerup,K., Lykke-Andersen,S., Andersen,J.S., Sandelin,A., *et al.* (2016) Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell*, **64**, 520–533.

323. Fan,J., Kuai,B., Wu,G., Wu,X., Chi,B., Wang,L., Wang,K., Shi,Z., Zhang,H., Chen,S., *et al.* (2017) Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J.*, **36**, 2870–2886.

324. Sheiness,D. and Darnell,J.E. (1973) Polyadenylic acid segment in mRNA becomes shorter with age. *Nat. New Biol.*, **241**, 265–268.

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

326. Lima,S.A., Chipman,L.B., Nicholson,A.L., Chen,Y.-H., Yee,B.A., Yeo,G.W., Coller,J. and Pasquinelli,A.E. (2017) Short poly(A) tails are a conserved feature of highly expressed genes. *Nat. Struct. Mol. Biol.*, **24**, 1057–1064.

327. Marzluff,W.F., Wagner,E.J. and Duronio,R.J. (2008) Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nat. Rev. Genet.*, **9**, 843–854.

328. Sullivan,K.D., Steiniger,M. and Marzluff,W.F. (2009) A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol. Cell*, **34**, 322–32.

329. Barnard,D.C., Ryan,K., Manley,J.L. and Richter,J.D. (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell*, **119**, 641–651.

330. Jae,E.K., Drier,E., Barbee,S.A., Ramaswami,M., Yin,J.C.P. and Wickens,M. (2008) GLD2 poly(A) polymerase is required for long-term memory. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 14644–14649.

331. Novoa,I., Gallego,J., Ferreira,P.G. and Mendez,R. (2010) Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat. Cell Biol.*, **12**, 447–456.

332. Takagaki,Y. and Manley,J.L. (2000) Complex Protein Interactions within the Human Polyadenylation Machinery Identify a Novel Component. *Mol. Cell. Biol.*, **20**, 1515–1525.

333. Bienroth,S., Wahle,E., Suter-Crazzolara,C. and Keller,W. (1991) Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J. Biol. Chem.*, **266**, 19768–19776.

334. Venkataraman,K., Brown,K.M. and Gilmartin,G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.*, **19**, 1315–27.

335. De Vries,H., Rüegsegger,U., Hübner,W., Friedlein,A., Langen,H. and Keller,W. (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.*, **19**, 5895–5904.

336. Marzluff,W.F. (1992) Histone 3' ends: essential and regulatory functions. *Gene Expr.*, **2**, 93–97.

337. Wilus,J., Pettine,S.M. and Shenk,T. (1989) Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. *Nucleic Acids Res.*, **17**, 3899–3908.

338. Beaudoing,E., Freier,S., Wyatt,J.R., Claverie,J.M. and Gautheret,D. (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res.*, **10**, 1001–1010.

339. Mandel,C.R., Bai,Y. and Tong,L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.*, **65**, 1099–122.

340. Yang,Q., Coseno,M., Gilmartin,G.M. and Doublié,S. (2011) Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure*, **19**, 368–377.

341. Tian,B. and Graber,J.H. (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA*, **3**, 385–96.

342. Dantonel,J.C., Murthy,K.G.K., Manjey,J.L. and Tora,L. (1997) Transcription factor TFIID

recruits factor CPSF for formation of 3 end of mRNA. *Nature*, **389**, 399–402.

343. Heidemann,M., Hintermair,C., Voß,K. and Eick,D. (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1829**, 55–62.

344. Parua,P.K., Kalan,S., Benjamin,B., Sansó,M. and Fisher,R.P. (2020) Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II. *Nat. Commun.*, **11**, 4338.

345. Nemeroff,M.E., Barabino,S.M.L., Li,Y., Keller,W. and Krug,R.M. (1998) Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′ end formation of cellular pre-mRNAs. *Mol. Cell*, **1**, 991–1000.

346. Fong,N., Saldi,T., Sheridan,R.M., Cortazar,M.A. and Bentley,D.L. (2017) RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Mol. Cell*, **66**, 546–557.

347. Takagaki,Y., Ryner,L.C. and Manley,J.L. (1988) Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell*, **52**, 731–742.

348. Terns,M.P. and Jacob,S.T. (1989) Role of poly(A) polymerase in the cleavage and polyadenylation of mRNA precursor. *Mol. Cell. Biol.*, **9**, 1435–1444.

349. Topalian,S.L., Kaneko,S., Gonzales,M.I., Bond,G.L., Ward,Y. and Manley,J.L. (2001) Identification and Functional Characterization of Neo-Poly(A) Polymerase, an RNA Processing Enzyme Overexpressed in Human Tumors. *Mol. Cell. Biol.*, **21**, 5614–5623.

350. Wahle,E., Lustig,A., Jeno,P. and Maurer,P. (1993) Mammalian poly(A)-binding protein II. Physical properties and binding to polynucleotides. *J. Biol. Chem.*, **268**, 2937–2945.

351. Nemeth,A., Krause,S., Blank,D., Jenny,A., Jenö,P., Lustig,A. and Wahle,E. (1995) Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II. *Nucleic Acids Res.*, **23**, 4034–4041.

352. Huang,C., Shi,J., Guo,Y., Huang,W., Huang,S., Ming,S., Wu,X., Zhang,R., Ding,J., Zhao,W., *et al.* (2017) A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.*, **45**, 8647–8660.

353. Noah,D.L., Twu,K.Y. and Krug,R.M. (2003) Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3′ end processing of cellular pre-mRNAS. *Virology*, **307**, 386–395.

354. Ashraf,S., Radhi,M., Gowler,P., Burston,J.J., Gandhi,R.D., Thorn,G.J., Piccinini,A.M., Walsh,D.A., Chapman,V. and de Moor,C.H. (2019) The polyadenylation inhibitor cordycepin reduces pain, inflammation and joint pathology in rodent models of osteoarthritis. *Sci. Rep.*, **9**, 4696.

355. Chen,W., Guo,W., Li,M., Shi,D., Tian,Y., Li,Z., Wang,J., Fu,L., Xiao,X., Liu,Q.Q., *et al.* (2013) Upregulation of cleavage and polyadenylation specific factor 4 in lung adenocarcinoma and its critical role for cancer cell survival and proliferation. *PLoS One*, **8**, e82728.

356. Lee,Y.J., Lee,Y. and Chung,J.H. (2000) An intronless gene encoding a poly(A) polymerase is specifically expressed in testis. *FEBS Lett.*, **487**, 287–292.

357. Kyriakopoulou,C.B., Nordvarg,H. and Virtanen,A. (2001) A Novel Nuclear Human Poly(A) Polymerase (PAP), PAPγ. *J. Biol. Chem.*, **276**, 33504–33511.

358. Martin,G., Doublié,S. and Keller,W. (2008) Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1779**,

206–216.

359. Yang,Q., Nausch,L.W.M., Martin,G., Keller,W. and Doublié,S. (2014) Crystal structure of human poly(A) polymerase gamma reveals a conserved catalytic core for canonical poly(A) polymerases. *J. Mol. Biol.*, **426**, 43–50.

360. Kashiwabara,S.I., Tsuruta,S., Okada,K., Yamaoka,Y. and Baba,T. (2016) Adenylation by testis-specific cytoplasmic poly(A) polymerase, PAPOLB/TPAP, is essential for spermatogenesis. *J. Reprod. Dev.*, **62**, 607–614.

361. Murthy,K.G.K. and Manley,J.L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev.*, **9**, 2672–2683.

362. Zhu,Y., Wang,X., Forouzmand,E., Jeong,J., Qiao,F., Sowd,G.A., Engelman,A.N., Xie,X., Hertel,K.J. and Shi,Y. (2018) Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol. Cell*, **69**, 62–74.

363. Wahle,E. (1991) A Novel Poly(A)-Binding Protein Acts As a Specificity Factor in the Second Phase of Messenger RNA Polyadenylation. *Cell*, **66**, 759–768.

364. Keller,R.W., Kühn,U., Aragón,M., Bornikova,L., Wahle,E. and Bear,D.G. (2000) The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *J. Mol. Biol.*, **297**, 569–583.

365. Mouland,A.J., Coady,M., Yao,X.-J. and Cohen,E.´ R.A. (2002) Hypophosphorylation of Poly(A) Polymerase and Increased Polyadenylation Activity Are Associated with Human Immunodeficiency Virus Type 1 Vpr Expression. *Virology*, **292**, 321–330.

366. Rose,K.M., Roe,F.J. and Jacob,S.T. (1977) Two functional states of poly(adenylic acid) polymerase in isolated nuclei. *BBA Sect. Nucleic Acids Protein Synth.*, **478**, 180–191.

367. Bresson,S.M., Hunter,O. V., Hunter,A.C. and Conrad,N.K. (2015) Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. *PLoS Genet.*, **11**, e1005610.

368. Rissland,O.S., Mikulasova,A. and Norbury,C.J. (2007) Efficient RNA Polyuridylation by Noncanonical Poly(A) Polymerases. *Mol. Cell. Biol.*, **27**, 3612–3624.

369. Lim,J., Ha,M., Chang,H., Kwon,S.C., Simanshu,D.K., Patel,D.J. and Kim,V.N. (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, **159**, 1365–1376.

370. Yu,S. and Kim,V.N. (2020) A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing. *Nat. Rev. Mol. Cell Biol.*, **21**, 542–556.

371. Mullen,T.E. and Marzluff,W.F. (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.*, **22**, 50–65.

372. Rissland,O.S. and Norbury,C.J. (2009) Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat. Struct. Mol. Biol.*, **16**, 616–623.

373. Mellman,D.L., Gonzales,M.L., Song,C., Barlow,C.A., Wang,P., Kendziorski,C. and Anderson,R.A. (2008) A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature*, **451**, 1013–1017.

374. Sudheesh,A.P., Mohan,N., Francis,N., Laishram,R.S. and Anderson,R.A. (2019) Star-PAP controlled alternative polyadenylation coupled poly(A) tail length regulates protein expression in hypertrophic heart. *Nucleic Acids Res.*, **47**, 10771–10787.

375. Trippe,R., Sandrock,B. and Benecke,B.J. (1998) A highly specific terminal uridylyl transferase modifies the 3'-end of U6 small nuclear RNA. *Nucleic Acids Res.*, **26**, 3119–3126.

376. Trippe,R., Guschina,E., Hossbach,M., Urlaub,H., Lührmann,R. and Benecke,B.J. (2006) Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA*, **12**, 1494–1504.

377. Kandala,D.T., Mohan,N., A,V., A P,S., G,R. and Laishram,R.S. (2016) CstF-64 and 3'-UTR cis-element determine Star-PAP specificity for target mRNA selection by excluding PAPα. *Nucleic Acids Res.*, **44**, 811–23.

378. Diener,S., Bayer,S., Sabrautzki,S., Wieland,T., Mentrup,B., Przemeck,G.K.H., Rathkolb,B., Graf,E., Hans,W., Fuchs,H., *et al.* (2016) Exome sequencing identifies a nonsense mutation in Fam46a associated with bone abnormalities in a new mouse model for skeletal dysplasia. *Mamm. Genome*, **27**, 111–121.

379. Doyard,M., Bacrot,S., Huber,C., Di Rocco,M., Goldenberg,A., Aglan,M.S., Brunelle,P., Temtamy,S., Michot,C., Otaify,G.A., *et al.* (2018) FAM46A mutations are responsible for autosomal recessive osteogenesis imperfecta. *J. Med. Genet.*, **55**, 278–284.

380. Sandberg,R., Neilson,J.R., Sarma,A., Sharp,P.A. and Burge,C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (80-.).*, **320**, 1643–1647.

381. Takagaki,Y., Seipelt,R.L., Peterson,M.L., Manley,J.L., Alt,F.., Bothwell,A.L.., Knapp,M., Siden,E., Mather,E., Koshland,M., *et al.* (1996) The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. *Cell*, **87**, 941–952.

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

383. Lianoglou,S., Garg,V., Yang,J.L., Leslie,C.S. and Mayr,C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.*, **27**, 2380–2396.

384. Ji,Z. and Tian,B. (2009) Reprogramming of 3' Untranslated Regions of mRANs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. *PLoS One*, **4**, e8419.

385. Mayr,C. and Bartel,D.P. (2009) Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, **138**, 673–684.

386. Tang,H.W., Hu,Y., Chen,C.L., Xia,B., Zirin,J., Yuan,M., Asara,J.M., Rabinow,L. and Perrimon,N. (2018) The TORC1-Regulated CPA Complex Rewires an RNA Processing Network to Drive Autophagy and Metabolic Reprogramming. *Cell Metab.*, **27**, 1040–1054.

387. Lackford,B., Yao,C., Charles,G.M., Weng,L., Zheng,X., Choi,E.-A., Xie,X., Wan,J., Xing,Y., Freudenberg,J.M., *et al.* (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.*, **33**, 878–889.

388. Li,W., You,B., Hoque,M., Zheng,D., Luo,W., Ji,Z., Park,J.Y., Gunderson,S.I., Kalsotra,A., Manley,J.L., *et al.* (2015) Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. *PLoS Genet.*, **11**, e1005166.

389. Jenal,M., Elkon,R., Loayza-Puch,F., Van Haaften,G., Kühn,U., Menzies,F.M., Vrielink,J.A.F.O., Bos,A.J., Drost,J., Rooijers,K., *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, **149**, 538–553.

390. Hardy,J.G., Tellier,M., Murphy,S. and Norbury,C.J. (2017) The RS Domain of Human CFIm68 Plays a Key Role in Selection Between Alternative Sites of Pre-mRNA Cleavage and Polyadenylation. *bioRxiv*, 10.1101/177980.

391. Takagaki,Y., Manley,J.L., Ashe,H.L., Monks,J., Wijgerde,M., Fraser,P., Proudfoot,N.J.,

Beyer,K., Dandekar,T., Keller,W., *et al.* (1998) Levels of Polyadenylation Factor CstF-64 Control IgM Heavy Chain mRNA Accumulation and Other Events Associated with B Cell Differentiation. *Mol. Cell*, **2**, 761–771.

392. Shell,S.A., Hesse,C., Morris,S.M. and Milcarek,C. (2005) Elevated Levels of the 64-kDa Cleavage Stimulatory Factor (CstF-64) in Lipopolysaccharide-stimulated Macrophages Influence Gene Expression and Induce Alternative Poly(A) Site Selection. *J. Biol. Chem.*, **280**, 39950–39961.

393. Berg,M.G., Singh,L.N., Younis,I., Liu,Q., Pinto,A.M., Kaida,D., Zhang,Z., Cho,S., Sherrill-Mix,S., Wan,L., *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, **150**, 53–64.

394. Oh,J.M., Di,C., Venters,C.C., Guo,J., Arai,C., So,B.R., Pinto,A.M., Zhang,Z., Wan,L., Younis,I., *et al.* (2017) U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat. Struct. Mol. Biol.*, **24**, 993–999.

395. So,B.R., Di,C., Cai,Z., Venters,C.C., Guo,J., Oh,J.M., Arai,C. and Dreyfuss,G. (2019) A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. *Mol. Cell*, **76**, 590–599.

396. Vassalli,J.D., Huarte,J., Belin,D., Gubler,P., Vassalli,A., O'Connell,M.L., Parton,L.A., Rickles,R.J. and Strickland,S. (1989) Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes Dev.*, **3**, 2163–71.

397. McGrew,L.L., Dworkin-Rastl,E., Dworkin,M.B. and Richter,J.D. (1989) Poly(A) elongation during Xenopus oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, **3**, 803–815.

398. Paris,J. and Richter,J.D. (1990) Maturation-Specific Polyadenylation and Translational Control: Diversity of Cytoplasmic Polyadenylation Elements, Influence of Poly(A) Tail Size, and Formation of Stable Polyadenylation Complexes. *Mol. Cell. Biol.*, **10**, 5634–5645.

399. Benoit,P., Papin,C., Kwak,J.E., Wickens,M. and Simonelig,M. (2008) PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. *Development*, **135**, 1969–79.

400. Juge,F., Zaessinger,S., Temme,C., Wahle,E. and Simonelig,M. (2002) Control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in Drosophila. *EMBO J.*, **21**, 6603–6613.

401. Eichhorn,S.W., Subtelny,A.O., Kronja,I., Kwasnieski,J.C., Orr-Weaver,T.L. and Bartel,D.P. (2016) mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. *Elife*, **5**, e16955.

402. Yamagishi,R., Tsusaka,T., Mitsunaga,H., Maehata,T. and Hoshino,S.-I. (2016) The STAR protein QKI-7 recruits PAPD4 to regulate post-transcriptional polyadenylation of target mRNAs. *Nucleic Acids Res.*, **44**, 2475–2490.

403. Cioni,J.M., Koppers,M. and Holt,C.E. (2018) Molecular control of local translation in axon development and maintenance. *Curr. Opin. Neurobiol.*, **51**, 86–94.

404. 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.

405. YS,H., MC,K., CL,L. and JD,R. (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J.*, **25**, 4865–4876.

406. Wang,X.-P. and Cooper,N.G.F. (2010) Comparative in silico analyses of cpeb1-4 with functional predictions. *Bioinform. Biol. Insights*, **4**, 61–83.

407. Lim,J., Kim,D., Lee,Y.S., Ha,M., Lee,M., Yeo,J., Chang,H., Song,J., Ahn,K. and Kim,V.N. (2018) Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science (80-.).*, **361**, 701–704.

408. Tang,T.T.L., Stowell,J.A.W., Hill,C.H. and Passmore,L.A. (2019) The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases. *Nat. Struct. Mol. Biol.*, **26**, 433–442.

409. Kim,D., Lee,Y. suk, Jung,S.J., Yeo,J., Seo,J.J., Lee,Y.Y., Lim,J., Chang,H., Song,J., Yang,J., *et al.* (2020) Viral hijacking of the TENT4–ZCCHC14 complex protects viral RNAs via mixed tailing. *Nat. Struct. Mol. Biol.*, **27**, 581–588.

410. Mroczek,S., Chlebowska,J., Kuliński,T.M., Gewartowska,O., Gruchota,J., Cysewski,D., Liudkovska,V., Borsuk,E., Nowis,D. and Dziembowski,A. (2017) The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat. Commun.*, **8**, 619.

411. Mendecki,J., Lee,S.Y. and Brawerman,G. (1972) Characteristics of the Polyadenylic Acid Segment Associated with Messenger Ribonucleic Acid in Mouse Sarcoma 180 Ascites Cells. *Biochemistry*, **11**, 792–798.

412. Wahle,E. (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.*, **270**, 2800–8.

413. Moore,J.P., Weber,M. and Searles,C.D. (2010) Laminar shear stress modulates phosphorylation and localization of RNA polymerase ii on the endothelial nitric oxide synthase gene. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 561–567.

414. Sagawa,F., Ibrahim,H., Morrison,A.L., Wilusz,C.J. and Wilusz,J. (2011) Nucleophosmin deposition during mRNA 3' end processing influences poly(A) tail length. *EMBO J.*, **30**, 3994–4005.

415. Kumar,G.R. and Glaunsinger,B.A. (2010) Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. *Mol. Cell. Biol.*, **30**, 4996–5008.

416. Kumar,G.R., Shum,L. and Glaunsinger,B.A. (2011) Importin -Mediated Nuclear Import of Cytoplasmic Poly(A) Binding Protein Occurs as a Direct Consequence of Cytoplasmic mRNA Depletion. *Mol. Cell. Biol.*, **31**, 3113–3125.

417. Burgess,H.M., Richardson,W.A., Anderson,R.C., Salaun,C., Graham,S. V and Gray,N.K. (2011) Nuclear relocalisation of cytoplasmic poly(A)-binding proteins PABP1 and PABP4 in response to UV irradiation reveals mRNA-dependent export of metazoan PABPs. *J. Cell Sci.*, **124**, 3344–55.

418. Viphakone,N., Voisinet-Hakil,F. and Minvielle-Sebastia,L. (2008) Molecular dissection of mRNA poly(A) tail length control in yeast. *Nucleic Acids Res.*, **36**, 2418–2433.

419. Hector,R.E., Nykamp,K.R., Dheur,S., Anderson,J.T., Non,P.J., Urbinati,C.R., Wilson,S.M., Minvielle-Sebastia,L. and Swanson,M.S. (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J.*, **21**, 1800–10.

420. Schmid,M., Poulsen,M.B., Olszewski,P., Pelechano,V., Saguez,C., Gupta,I., Steinmetz,L.M., Moore,C. and Jensen,T.H. (2012) Rrp6p Controls mRNA Poly(A) Tail Length and Its Decoration with Poly(A) Binding Proteins. *Mol. Cell*, **47**, 267–280.

421. Kelly,S.M., Leung,S.W., Pak,C., Banerjee,A., Moberg,K.H. and Corbett,A.H. (2014) A conserved role for the zinc finger polyadenosine RNA binding protein, ZC3H14, in control of poly(A) tail length. *RNA*, **20**, 681–8.

422. Palaniswamy,V., Moraes,K.C.M., Wilusz,C.J. and Wilusz,J. (2006) Nucleophosmin is selectively deposited on mRNA during polyadenylation. *Nat. Struct. Mol. Biol.*, **13**, 429–435.

423. Morrison,M.R., Merkel,C.G. and Lingrel,J.B. (1973) Size of the poly(A) region in mouse globin messenger RNA. *Mol. Biol. Rep.*, **1**, 55–60.

424. Rao,M.N., Chernokalskaya,E. and Schoenberg,D.R. (1996) Regulated nuclear polyadenylation of Xenopus albumin pre-mRNA. *Nucleic Acids Res.*, **24**, 4078–83.

425. Das Gupta,J., Gu,H., Chernokalskaya,E., Gao,X. and Schoenberg,D.R. (1998) Identification of two cis-acting elements that independently regulate the length of poly(A) on Xenopus albumin pre-mRNA. *RNA*, **4**, 766–776.

426. Gu,H., Gupta,J. Das and Schoenberg,D.R. (1999) The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. *Biochemistry*, **96**, 8943–8948.

427. Gu,H. and Schoenberg,D.R. (2003) U2AF modulates poly(A) length control by the poly(A)-limiting element. *Nucleic Acids Res.*, **31**, 6264–6271.

428. Choi,Y.H. and Hagedorn,C.H. (2003) Purifying mRNAs with a high-affinity eIF4E mutant identifies the short 3′ poly(A) end phenotype. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 7033–7038.

429. Millevoi,S., Loulergue,C., Dettwiler,S., Karaa,S.Z., Keller,W., Antoniou,M. and Vagner,S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.*, **25**, 4854–64.

430. Brown,C.E. and Sachs,A.B. (1998) Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. *Mol. Cell. Biol.*, **18**, 6548–59.

431. Su,Y.-L., Wang,S.-C., Chiang,P.-Y., Lin,N.-Y., Shen,Y.-F., Chang,G.-D., Chang,C.-J., Chen,C., Shyu,A., Shaw,G., *et al.* (2012) Tristetraprolin Inhibits Poly(A)-Tail Synthesis in Nuclear mRNA that Contains AU-Rich Elements by Interacting with Poly(A)-Binding Protein Nuclear 1. *PLoS One*, **7**, e41313.

432. Woo,Y.M., Kwak,Y., Namkoong,S., Kristjánsdóttir,K., Lee,S.H., Lee,J.H. and Kwak,H. (2018) TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Rep.*, **24**, 3630–3641.

433. Legnini,I., Alles,J., Karaiskos,N., Ayoub,S. and Rajewsky,N. (2019) FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat. Methods*, **16**, 879–886.

434. Chorghade,S., Seimetz,J., Emmons,R., Yang,J., Bresson,S.M., De Lisio,M., Parise,G., Conrad,N.K. and Kalsotra,A. (2017) Poly(A) tail length regulates PABPC1 expression to tune translation in the heart. *Elife*, 10.7554/eLife.24139.001.

435. Eisen,T.J., Eichhorn,S.W., Subtelny,A.O. and Bartel,D.P. (2020) MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. *Mol. Cell*, **77**, 775–785.

436. Apponi,L.H., Leung,S.W., Williams,K.R., Valentini,S.R., Corbett,A.H. and Pavlath,G.K. (2010) Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Hum. Mol. Genet.*, **19**, 1058–65.

437. Flamand,M.N., Wu,E., Vashisht,A., Jannot,G., Keiper,B.D., Simard,M.J., Wohlschlegel,J. and Duchaine,T.F. (2016) Poly(A)-binding proteins are required for microRNA-mediated silencing and to promote target deadenylation in C. elegans. *Nucleic Acids Res.*, **44**, 5924–5935.

438. Borman,A.M., Michel,Y.M. and Kean,K.M. (2000) Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5′-end. *Nucleic Acids Res.*, **28**, 4068–75.

439. Dunn,E.F., Hammell,C.M., Hodge,C.A. and Cole,C.N. (2005) Yeast poly(A)-binding protein,

Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev.*, **19**, 90–103.

440. Ezzeddine,N., Chang,T.-C., Zhu,W., Yamashita,A., Chen,C.-Y.A., Zhong,Z., Yamashita,Y., Zheng,D. and Shyu,A.-B. (2007) Human TOB, an antiproliferative transcription factor, is a poly(A)-binding protein-dependent positive regulator of cytoplasmic mRNA deadenylation. *Mol. Cell. Biol.*, **27**, 7791–801.

441. Fasken,M.B., Stewart,M. and Corbett,A.H. (2008) Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. *J. Biol. Chem.*, **283**, 27130–43.

442. Schäfer,I.B., Yamashita,M., Schuller,J.M., Schü,S., Reichelt,P., Strauss,M., Correspondence,E.C., Schü Ssler,S. and Conti,E. (2019) Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylase. *Cell*, **177**, 1619–1631.

443. Hosoda,N., Lejeune,F. and Maquat,L.E. (2006) Evidence that poly(A) binding protein C1 binds nuclear pre-mRNA poly(A) tails. *Mol. Cell. Biol.*, **26**, 3085–97.

444. Bhattacharjee,R.B. and Bag,J. of N.P.B.P.P.P. a C.R. by C.P. and P. in C.H.C. (2012) Depletion of Nuclear Poly(A) Binding Protein PABPN1 Produces a Compensatory Response by Cytoplasmic PABP4 and PABP5 in Cultured Human Cells. *PLoS One*, **7**, e53036.

445. Krause,S., Fakan,S., Weis,K. and Wahle,E. (1994) Immunodetection of poly(A) binding protein II in the cell nucleus. *Exp. Cell Res.*, **214**, 75–82.

446. Lee,Y.J., Lee,J., Yang,I.C., Hahn,Y., Lee,Y. and Chung,J.H. (1998) Genomic structure and expression of murine poly(A) binding protein II gene. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **1395**, 40–46.

447. Ishigaki,Y., Li,X., Serin,G. and Maquat,L.E. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell*, **106**, 607–617.

448. Sato,H. and Maquat,L.E. (2009) Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin β. *Genes Dev.*, **23**, 2537–2550.

449. Brais,B., Bouchard,J.P., Xie,Y.G., Rochefort,D.L., Chrétien,N., Tomé,F.M.S., Lafrenière,R.G., Rommens,J.M., Uyama,E., Nohira,O., *et al.* (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.*, **18**, 164–167.

450. Apponi,L.H., Corbett,A.H. and Pavlath,G.K. (2013) Control of mRNA stability contributes to low levels of nuclear poly(A) binding protein 1 (PABPN1) in skeletal muscle. *Skelet. Muscle*, **3**, 23.

451. Phillips,B.L., Banerjee,A., Sanchez,B.J., Marco,S. Di, Gallouzi,I.-E., Pavlath,G.K. and Corbett,A.H. (2018) Post-transcriptional regulation of Pabpn1 by the RNA binding protein HuR. *Nucleic Acids Res.*, **46**, 7643–7661.

452. Peng,S.S.Y., Chen,C.Y.A., Xu,N. and Shyu,A. Bin (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.*, **17**, 3461–3470.

453. Cammas,A., Sanchez,B.J., Lian,X.J., Dormoy-Raclet,V., Van Der Giessen,K., De Silanes,I.L., Ma,J., Wilusz,C., Richardson,J., Gorospe,M., *et al.* (2014) Destabilization of nucleophosmin mRNA by the HuR/KSRP complex is required for muscle fibre formation. *Nat. Commun.*, **5**, 4190.

454. Anantharaman,A., Tripathi,V., Khan,A., Yoon,J.H., Singh,D.K., Gholamalamdari,O., Guang,S., Ohlson,J., Wahlstedt,H., Öhman,M., *et al.* (2017) ADAR2 regulates RNA stability by modifying access of decay-promoting RNA-binding proteins. *Nucleic Acids Res.*, **45**, 4189–4201.

455. Wigington,C.P., Williams,K.R., Meers,M.P., Bassell,G.J. and Corbett,A.H. (2014) Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley Interdiscip. Rev. RNA*, **5**, 601–22.

456. Aslam,M., Ullah,A., Paramasivam,N., Kandasamy,N., Naureen,S., Badshah,M., Khan,K., Wajid,M., Abbasi,R., Eils,R., *et al.* (2019) Segregation and potential functional impact of a rare stop-gain PABPC4L variant in familial atypical Parkinsonism. *Sci. Rep.*, **9**, 13576.

457. Wu,Y.Q., Ju,C.L., Wang,B.J. and Wang,R.G. (2019) PABPC1L depletion inhibits proliferation and migration via blockage of AKT pathway in human colorectal cancer cells. *Oncol. Lett.*, **17**, 3439–3445.

458. Sladic,R.T., Lagnado,C.A., Bagley,C.J. and Goodall,G.J. (2004) Human PABP binds AU-rich RNA via RNA-binding domains 3 and 4. *Eur. J. Biochem.*, **271**, 450–457.

459. Boeck,R., Tarun,S., Rieger,M., Deardorff,J.A., Müller-Auer,S. and Sachs,A.B. (1996) The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.*, **271**, 432–8.

460. Tucker,M., Staples,R.R., Valencia-Sanchez,M.A., Muhlrad,D. and Parker,R. (2002) Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. *EMBO J.*, **21**, 1427–1436.

461. Viswanathan,P., Ohn,T., Chiang,Y.C., Chen,J. and Denis,C.L. (2004) Mouse CAF1 can function as a processive deadenylase/3'-5' -exonuclease in vitro but in yeast the deadenylase function of CAF1 is not required for mRNA poly(A) removal. *J. Biol. Chem.*, **279**, 23988–23995.

462. Ezzeddine,N., Chen,C.-Y.A. and Shyu,A.-B. (2012) Evidence Providing New Insights into TOB-Promoted Deadenylation and Supporting a Link between TOB's Deadenylation-Enhancing and Antiproliferative Activities. *Mol. Cell. Biol.*, **32**, 1089–1098.

463. Stupfler,B., Birck,C., Séraphin,B. and Mauxion,F. (2016) BTG2 bridges PABPC1 RNA-binding domains and CAF1 deadenylase to control cell proliferation. *Nat. Commun.*, **7**, 10811.

464. 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.

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

466. Brook,M., McCracken,L., Reddington,J.P., Lu,Z.-L., Morrice,N.A. and Gray,N.K. (2012) The multifunctional poly(A)-binding protein (PABP) 1 is subject to extensive dynamic post-translational modification, which molecular modelling suggests plays an important role in co-ordinating its activities. *Biochem. J.*, **441**, 803–812.

467. Afonina,E., Stauber,R. and Pavlakis,G.N. (1998) The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm. *J. Biol. Chem.*, **273**, 13015–13021.

468. Harb,M., Becker,M.M., Vitour,D., Baron,C.H., Vende,P., Brown,S.C., Bolte,S., Arold,S.T. and Poncet,D. (2008) Nuclear Localization of Cytoplasmic Poly(A)-Binding Protein upon Rotavirus Infection Involves the Interaction of NSP3 with eIF4G and RoXaN. *J. Virol.*, **82**, 11283–11293.

469. Ma,S., Bhattacharjee,R.B. and Bag,J. (2009) Expression of poly(A)-binding protein is upregulated during recovery from heat shock in HeLa cells. *FEBS J.*, **276**, 552–570.

470. Gilbertson,S., Federspiel,J.D., Hartenian,E., Cristea,I.M. and Glaunsinger,B. (2018) Changes in mRNA abundance drive shuttling of RNA binding proteins, linking cytoplasmic RNA degradation to transcription. *Elife*, **7**, e37663.

471. Abernathy,E., Gilbertson,S., Alla,R. and Glaunsinger,B. (2015) Viral nucleases induce an mRNA degradation-transcription feedback loop in mammalian cells. *Cell Host Microbe*, **18**, 243–253.

472. 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.

473. Kelly,S.M., Pabit,S.A., Kitchen,C.M., Guo,P., Marfatia,K.A., Murphy,T.J., Corbett,A.H. and Berland,K.M. (2007) Recognition of polyadenosine RNA by zinc finger proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 12306–12311.

474. Yang,R., Gaidamakov,S.A., Xie,J., Lee,J., Martino,L., Kozlov,G., Crawford,A.K., Russo,A.N., Conte,M.R., Gehring,K., *et al.* (2011) La-Related Protein 4 Binds Poly(A), Interacts with the Poly(A)-Binding Protein MLLE Domain via a Variant PAM2w Motif, and Can Promote mRNA Stability. *Mol. Cell. Biol.*, **31**, 542–556.

475. Leung,S.W., Apponi,L.H., Cornejo,O.E., Kitchen,C.M., Valentini,S.R., Pavlath,G.K., Dunham,C.M. and Corbett,A.H. (2009) Splice variants of the human ZC3H14 gene generate multiple isoforms of a zinc finger polyadenosine RNA binding protein. *Gene*, **439**, 71–78.

476. Pak,C., Garshasbi,M., Kahrizi,K., Gross,C., Apponi,L.H., Noto,J.J., Kelly,S.M., Leung,S.W., Tzschach,A., Behjati,F., *et al.* (2011) Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 12390–5.

477. Wigington,C.P., Morris,K.J., Newman,L.E. and Corbett,A.H. (2016) The Polyadenosine RNA-binding Protein, Zinc Finger Cys3His Protein 14 (ZC3H14), Regulates the Pre-mRNA Processing of a Key ATP Synthase Subunit mRNA. *J. Biol. Chem.*, **291**, 22442–22459.

478. Mattijssen,S., Iben,J.R., Li,T., Coon,S.L. and Maraia,R.J. (2020) Single molecule poly(A) tail-seq shows larp4 opposes deadenylation throughout mrna lifespan with most impact on short tails. *Elife*, **9**, e59186.

479. Cooke,C., Hans,H. and Alwine,J.C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol. Cell. Biol.*, **19**, 4971–9.

480. Lu,S. and Cullen,B.R. (2003) Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA*, **9**, 618–30.

481. Awasthi,S. and Alwine,J.C. (2003) Association of polyadenylation cleavage factor I with U1 snRNP. *RNA*, **9**, 1400–1409.

482. Soucek,S., Zeng,Y., Bellur,D.L., Bergkessel,M., Morris,K.J., Deng,Q., Duong,D., Seyfried,N.T., Guthrie,C., Staley,J.P., *et al.* (2016) Evolutionarily Conserved Polyadenosine RNA Binding Protein Nab2 Cooperates with Splicing Machinery To Regulate the Fate of Pre-mRNA. *Mol. Cell. Biol.*, **36**, 2697–2714.

483. Lam,B.J. and Hertel,K.J. (2002) A general role for splicing enhancers in exon definition. *RNA*, **8**, 1233–41.

484. Lewis,J.D., Izaurralde,E., Jarmolowski,A., Mcguigan,C. and Mattaj^,I.W. (1996) A nuclear cap-binding complex facilitates association of Ul snRNP with the cap-proximal 5' splice site. *Genes Dev.*, **10**, 1683–1698.

485. Pandya-Jones,A., Bhatt,D.M., Lin,C.-H., Tong,A.-J., Smale,S.T. and Black,D.L. (2013) Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA*, **19**, 811–27.

486. Carrillo Oesterreich,F., Herzel,L., Straube,K., Hujer,K., Howard,J. and Neugebauer,K.M. (2016) Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell*,

165, 372–381.

487. Gunderson,S.I., Polycarpou-Schwarz,M. and Mattaj,I.W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell*, **1**, 255–264.

488. Dettwiler,S., Aringhieri,C., Cardinale,S., Keller,W. and Barabino,S.M.L. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J. Biol. Chem.*, **279**, 35788–35797.

489. Flaherty,S.M., Fortes,P., Izaurralde,E., Mattaj,I.W. and Gilmartin,G.M. (1997) Participation of the nuclear cap binding complex in pre-mRNA 3′ processing. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 11893–11898.

490. Stewart,M. (2019) Polyadenylation and nuclear export of mRNAs. *J. Biol. Chem.*, **294**, 2977–2987.

491. Green,D.M., Marfatia,K.A., Crafton,E.B., Zhang,X., Cheng,X. and Corbett,A.H. (2002) Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. *J. Biol. Chem.*, **277**, 7752–7760.

492. Johnson,S.A., Cubberley,G. and Bentley,D.L. (2009) Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3′ End Processing Factor Pcf11. *Mol. Cell*, **33**, 215–226.

493. Duy,D., Tran,H., Saran,S., Williamson,A.J.K., Pierce,A., Dittrich-Breiholz,O., Wiehlmann,L., Koch,A., Whetton,A.D. and Tamura,T. (2014) THOC5 controls 3 end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100). *Nucleic Acids Res.*, **42**, 12249–12260.

494. Aoki,K., Adachi,S., Homoto,M., Kusano,H., Koike,K. and Natsume,T. (2013) LARP1 specifically recognizes the 3′ terminus of poly(A) mRNA. *FEBS Lett.*, **587**, 2173–2178.

495. Tcherkezian,J., Cargnello,M., Romeo,Y., Huttlin,E.L., Lavoie,G., Gygi,S.P. and Roux,P.P. (2014) Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5′TOP mRNA translation. *Genes Dev.*, **28**, 357–371.

496. Fonseca,B.D., Zakaria,C., Jia,J.J., Graber,T.E., Svitkin,Y., Tahmasebi,S., Healy,D., Hoang,H.D., Jensen,J.M., Diao,I.T., *et al.* (2015) La-related protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR complex 1 (mTORC1). *J. Biol. Chem.*, **290**, 15996–16020.

497. Philippe,L., Vasseur,J.-J., Debart,F. and Thoreen,C.C. (2018) La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region. *Nucleic Acids Res.*, **46**, 1457–1469.

498. Eckmann,C.R., Rammelt,C. and Wahle,E. (2011) Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA*, **2**, 348–361.

499. Al-Ashtal,H.A., Rubottom,C.M., Leeper,T.C. and Berman,A.J. (2019) The LARP1 La-Module recognizes both ends of TOP mRNAs. *RNA Biol.*, **33**, 236–252.

500. Harnisch,C., Cuzic-Feltens,S., Dohm,J.C., Götze,M., Himmelbauer,H. and Wahle,E. (2016) Oligoadenylation of 3′ decay intermediates promotes cytoplasmic mRNA degradation in Drosophila cells. *RNA*, **22**, 428–42.

501. Baer,B.W. and Kornberg,R.D. (1980) Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.*, **77**, 1890–1892.

502. Sachs,A.B., Davis,R.W. and Kornberg,R.D. (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol. Cell. Biol.*, **7**, 3268–3276.

503. Wolf,J. deadenylation by P.-P. and Passmore,L.A. (2014) mRNA deadenylation by Pan2-Pan3. *Biochem. Soc. Trans.*, **42**, 184–7.

504. Lee,J.E., Lee,J.Y., Trembly,J., Wilusz,J., Tian,B. and Wilusz,C.J. (2012) The PARN Deadenylase Targets a Discrete Set of mRNAs for Decay and Regulates Cell Motility in Mouse Myoblasts. *PLoS Genet.*, **8**, e1002901.

505. Godwin,A.R., Kojima,S., Green,C.B. and Wilusz,J. (2013) Kiss your tail goodbye: the role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. *Biochim. Biophys. Acta*, **1829**, 571–9.

506. Laothamatas,I., Gao,P., Wickramaratne,A., Quintanilla,C.G., Dino,A., Khan,C.A., Liou,J. and Green,C.B. (2020) Spatiotemporal regulation of NADP(H) phosphatase Nocturnin and its role in oxidative stress response. *Proc. Natl. Acad. Sci.*, **117**, 993–999.

507. Goldstrohm,A.C. and Wickens,M. (2008) Multifunctional deadenylase complexes diversify mRNA control. *Nat. Rev. Mol. Cell Biol.*, **9**, 337–44.

508. Ukleja,M., Cuellar,J., Siwaszek,A., Kasprzak,J.M., Czarnocki-Cieciura,M., Bujnicki,J.M., Dziembowski,A. and M. Valpuesta,J. (2016) The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. *Nat. Commun. 2016 71*, **7**, 1–11.

509. Miller,J.E., Zhang,L., Jiang,H., Li,Y., Pugh,B.F. and Reese,J.C. (2018) Genome-Wide Mapping of Decay Factor–mRNA Interactions in Yeast Identifies Nutrient-Responsive Transcripts as Targets of the Deadenylase Ccr4. *G3 Genes Genomes Genet.*, **8**, 315–330.

510. Lenssen,E., Oberholzer,U., Labarre,J., De Virgilio,C. and Collart,M.A. (2002) Saccharomyces cerevisiae Ccr4-Not complex contributes to the control of Msn2p-dependent transcription by the Ras/cAMP pathway. *Mol. Microbiol.*, **43**, 1023–1037.

511. Kassem,S., Villanyi,Z. and Collart,M.A. (2017) Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA. *Nucleic Acids Res.*, **45**, 1186–1199.

512. Cui,Y., Ramnarain,D.B., Chiang,Y.C., Ding,L.H., McMahon,J.S. and Denis,C.L. (2008) Genome wide expression analysis of the CCR4-NOT complex indicates that it consists of three modules with the NOT module controlling SAGA-responsive genes. *Mol. Genet. Genomics*, **279**, 323–337.

513. Bawankar,P., Loh,B., Wohlbold,L., Schmidt,S. and Izaurralde,E. (2013) NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA Biol.*, **10**, 228–244.

514. Bhaskar,V., Roudko,V., Basquin,J., Sharma,K., Urlaub,H., Séraphin,B. and Conti,E. (2013) Structure and RNA-binding properties of the Not1-Not2-Not5 module of the yeast Ccr4-Not complex. *Nat. Struct. Mol. Biol.*, **20**, 1281–1288.

515. Keskeny,C., Raisch,T., Sgromo,A., Igreja,C., Bhandari,D., Weichenrieder,O. and Izaurralde,E. (2019) A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4-NOT complex. 10.1101/gad.320952.118.

516. Bai,Y., Salvadore,C., Chiang,Y.C., Collart,M.A., Liu,H.Y. and Denis,C.L. (1999) The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.*, **19**, 6642–51.

517. Maillet,L., Tu,C., Hong,Y.K., Shuster,E.O. and Collart,M.A. (2000) The essential function of Not1 lies within the Ccr4-Not complex. *J. Mol. Biol.*, **303**, 131–143.

518. Albert,T.K., Lemaire,M., Van Berkum,N.L., Gentz,R., Collart,M.A. and Timmers,H.T.M. (2000) Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. *Nucleic Acids Res.*, **28**, 809–817.

519. Dominick,G., Bowman,J., Li,X., Miller,R.A. and Garcia,G.G. (2017) mTOR regulates the expression of DNA damage response enzymes in long-lived Snell dwarf, GHRKO, and PAPPA-KO mice. *Aging Cell*, **16**, 52–60.

520. Faraji,F., Hu,Y., Wu,G., Goldberger,N.E., Walker,R.C., Zhang,J. and Hunter,K.W. (2014) An integrated systems genetics screen reveals the transcriptional structure of inherited predisposition to metastatic disease. *Genome Res.*, **24**, 227–40.

521. Doidge,R., Mittal,S., Aslam,A. and Winkler,G.S. (2012) Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. *Biochem. Soc. Trans.*, **40**, 896–901.

522. Dupressoir,A., Morel,A.P., Barbot,W., Loireau,M.P., Corbo,L. and Heidmann,T. (2001) Identification of four families of yCCR4- and Mg2+ -dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAFI/hPOP2 binding. *BMC Genomics*, **2**, 9.

523. Basquin,J., Roudko,V. V., Rode,M., Basquin,C., Séraphin,B. and Conti,E. (2012) Architecture of the nuclease module of the yeast ccr4-Not complex: The not1-caf1-ccr4 interaction. *Mol. Cell*, **48**, 207–218.

524. Petit,A.P., Wohlbold,L., Bawankar,P., Huntzinger,E., Schmidt,S., Izaurralde,E. and Weichenrieder,O. (2012) The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res.*, **40**, 11058–11072.

525. Mittal,S., Aslam,A., Doidge,R., Medica,R. and Winkler,G.S. (2011) The Ccr4a (CNOT6) and Ccr4b (CNOT6L) deadenylase subunits of the human Ccr4-Not complex contribute to the prevention of cell death and senescence. *Mol. Biol. Cell*, **22**, 748–758.

526. Mulder,K.W., Brenkman,A.B., Inagaki,A., van den Broek,N.J.F. and Marc Timmers,H.T. (2007) Regulation of histone H3K4 tri-methylation and PAF complex recruitment by the Ccr4-Not complex. *Nucleic Acids Res.*, **35**, 2428–2439.

527. Peng,W., Togawa,C., Zhang,K. and Kurdistani,S.K. (2008) Regulators of cellular levels of histone acetylation in Saccharomyces cerevisiae. *Genetics*, **179**, 277–289.

528. Daugeron,M.-C., Mauxion,F. and Séraphin,B. (2001) The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.*, **29**, 2448–2455.

529. Tucker,M., Valencia-Sanchez,M.A., Staples,R.R., Chen,J., Denis,C.L. and Parker,R. (2001) The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell*, **104**, 377–386.

530. Chen,J., Chiang,Y.-C. and Denis,C.L. (2002) CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. *EMBO J.*, **21**, 1414–26.

531. Maryati,M., Airhihen,B. and Winkler,G.S. (2015) The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4–Not nuclease module. *Biochem. J*, **469**, 169–176.

532. Y,C., E,K., E,I. and O,W. (2021) Crystal structure and functional properties of the human CCR4-CAF1 deadenylase complex. *Nucleic Acids Res.*, **49**, 6489–6510.

533. Horiuchi,M., Takeuchi,K., Noda,N., Muroya,N., Suzuki,T., Nakamura,T., Kawamura-Tsuzuku,J., Takahasi,K., Yamamoto,T. and Inagaki,F. (2009) Structural basis for the antiproliferative activity of the Tob-hCaf1 complex. *J. Biol. Chem.*, **284**, 13244–13255.

534. Chou,C.-F., Mulky,A., Maitra,S., Lin,W.-J., Gherzi,R., Kappes,J. and Chen,C.-Y. (2006) Tethering KSRP, a Decay-Promoting AU-Rich Element-Binding Protein, to mRNAs Elicits mRNA Decay. *Mol. Cell. Biol.*, **26**, 3695–3706.

535. Sgromo,A., Raisch,T., Bawankar,P., Bhandari,D., Chen,Y., Kuzuoglu-Öztürk,D., Weichenrieder,O. and Izaurralde,E. (2017) A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat. Commun.*, **8**, 14307.

536. Bohn,J.A., Van Etten,J.L., Schagat,T.L., Bowman,B.M., McEachin,R.C., Freddolino,P.L. and Goldstrohm,A.C. (2018) Identification of diverse target RNAs that are functionally regulated by human pumilio proteins. *Nucleic Acids Res.*, **46**, 362–386.

537. Fukushima,M., Hosoda,N., Chifu,K. and Hoshino,S. (2019) TDP-43 accelerates deadenylation of target mRNAs by recruiting Caf1 deadenylase. *FEBS Lett.*, **593**, 277–287.

538. Colombrita,C., Onesto,E., Megiorni,F., Pizzuti,A., Baralle,F.E., Buratti,E., Silani,V. and Ratti,A. (2012) TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *J. Biol. Chem.*, **287**, 15635–15647.

539. Bhandari,D., Raisch,T., Weichenrieder,O., Jonas,S. and Izaurralde,E. (2013) Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes Dev.*, **28**, 888–901.

540. Wilczynska,A., Gillen,S.L., Schmidt,T., Meijer,H.A., Jukes-Jones,R., Langlais,C., Kopra,K., Lu,W.-T., Godfrey,J.D., Hawley,B.R., *et al.* (2019) eIF4A2 drives repression of translation at initiation by Ccr4-Not through purine-rich motifs in the 5'UTR. *Genome Biol.*, **20**, 262.

541. Bartel,D.P. (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, **136**, 215–233.

542. Chen,Y., Boland,A., Kuzuoğlu-Öztürk,D., Bawankar,P., Loh,B., Chang,C.-T., Weichenrieder,O. and Izaurralde,E. (2014) A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. *Mol. Cell*, **54**, 737–750.

543. 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**, 751–65.

544. Luo,Y., Na,Z. and Slavoff,S.A. (2018) P-Bodies: Composition, Properties, and Functions. *Biochemistry*, **57**, 2424–2431.

545. Deluen,C., James,N., Maillet,L., Molinete,M., Theiler,G., Lemaire,M., Paquet,N. and Collart,M.A. (2002) The Ccr4-Not Complex and yTAF1 (yTafII130p/yTafII145p) Show Physical and Functional Interactions. *Mol. Cell. Biol.*, **22**, 6735–6749.

546. Haas,M., Siegert,M., Schürmann,A., Sodeik,B. and Wolfes,H. (2004) c-Myb protein interacts with Rcd-1, a component of the CCR4 transcription mediator complex. *Biochemistry*, **43**, 8152–8159.

547. Lenssen,E., James,N., Pedruzzi,I., Dubouloz,† Frédérique, Cameroni,E., Bisig,R., Maillet,L., Werner,‡ Michel, Roosen,J., Petrovic,K., *et al.* (2005) The Ccr4-Not Complex Independently Controls both Msn2-Dependent Transcriptional Activation-via a Newly Identified Glc7/Bud14 Type I Protein Phosphatase Module-and TFIID Promoter Distribution. *Mol. Cell. Biol.*, **25**, 488–498.

548. James,N., Landrieux,E. and Collart,M.A. (2007) A SAGA-independent function of SPT3 mediates transcriptional deregulation in a mutant of the Ccr4-not complex in Saccharomyces cerevisiae. *Genetics*, **177**, 123–135.

549. Collart,M.A. and Struhl,K. (1994) NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. *Genes Dev.*, **8**, 525–537.

550. Liu,H.-Y., Badarinarayana,V., Audino,D.C., Rappsilber,J., Mann,M. and Denis,C.L. (1998) The

NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively.

551. Oberholzer,U. and Collart,M.A. (1998) Characterization of NOT5 that encodes a new component of the Not protein complex. *Gene*, **207**, 61–69.

552. Venters,B.J., Wachi,S., Mavrich,T.N., Andersen,B.E., Jena,P., Sinnamon,A.J., Jain,P., Rolleri,N.S., Jiang,C., Hemeryck-Walsh,C., *et al.* (2011) A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces. *Mol. Cell*, **41**, 480–492.

553. Mersman,D.P., Du,H.N., Fingerman,I.M., South,P.F. and Briggs,S.D. (2009) Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. *Genes Dev.*, **23**, 951–962.

554. Villanyi,Z., Ribaud,V., Kassem,S., Panasenko,O.O., Pahi,Z., Gupta,I., Steinmetz,L., Boros,I. and Collart,M.A. (2014) The Not5 Subunit of the Ccr4-Not Complex Connects Transcription and Translation. *PLoS Genet.*, **10**, e1004569.

555. Babbarwal,V., Fu,J. and Reese,J.C. (2014) The Rpb4/7 module of RNA polymerase II is required for carbon catabolite repressor protein 4-negative on TATA (Ccr4-Not) complex to promote elongation. *J. Biol. Chem.*, **289**, 33125–33130.

556. Cooke,A., Prigge,A. and Wickens,M. (2010) Translational repression by deadenylases. *J. Biol. Chem.*, **285**, 28506–28513.

557. Gupta,I., Villanyi,Z., Kassem,S., Hughes,C., Panasenko,O.O., Steinmetz,L.M. and Correspondence,M.A.C. (2016) Translational Capacity of a Cell Is Determined during Transcription Elongation via the Ccr4-Not Complex. *Cell Rep.*, **15**, 1782–1794.

558. Moriya,H., Shimizu-Yoshida,Y., Omori,A., Iwashita,S., Katoh,M. and Sakai,A. (2001) Yak1p, a DYRK family kinase, translocates to the nucleus and phosphorylates yeast Pop2p in response to a glucose signal. *Genes Dev.*, **15**, 1217–1228.

559. Norbeck,J. (2008) Carbon source dependent dynamics of the Ccr4-Not complex in Saccharomyces cerevisiae. *J. Microbiol.*, **46**, 692–696.

560. Lien,P.T.K., Viet,N.T.M., Mizuno,T., Suda,Y. and Irie,K. (2019) Pop2 phosphorylation at S39 contributes to the glucose repression of stress response genes, HSP12 and HSP26. *PLoS One*, **14**, e0215064.

561. Morel,A.P., Sentis,S., Bianchin,C., Le Romancer,M., Jonard,L., Rostan,M.C., Rimokh,R. and Corbo,L. (2003) BTG2 antiproliferative protein interacts with the human CCR4 complex existing in vivo in three cell-cycle-regulated forms. *J. Cell Sci.*, **116**, 2929–2936.

562. Jonas,S., Christie,M., Peter,D., Bhandari,D., Loh,B., Huntzinger,E., Weichenrieder,O. and Izaurralde,E. (2014) An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. *Nat. Struct. Mol. Biol.*, **21**, 599–608.

563. Brown,C.E., Tarun,S.Z., Boeck,R. and Sachs,A.B. (1996) PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **16**, 5744–53.

564. Mangus,D.A., Evans,M.C., Agrin,N.S., Smith,M., Gongidi,P. and Jacobson,A. (2004) Positive and negative regulation of poly(A) nuclease. *Mol. Cell. Biol.*, **24**, 5521–33.

565. Siddiqui,N., Mangus,D.A., Chang,T.C., Palermino,J.M., Shyu,A. Bin and Gehring,K. (2007) Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *J. Biol. Chem.*, **282**, 25067–25075.

566. Saenger,W., Riecke,J. and Suck,D. (1975) A structural model for the polyadenylic acid single helix. *J. Mol. Biol.*, **93**, 529–534.

567. Lowell,J.E., Rudner,D.Z. and Sachs,A.B. (1992) 3'-UTR-dependent deadenylation by the yeast

poly(A) nuclease. *Genes Dev.*, **6**, 2088–99.

568. Christie,M., Boland,A., Huntzinger,E., Weichenrieder,O. and Izaurralde,E. (2013) Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. *Mol. Cell*, **51**, 360–373.

569. Huang,K.L., Chadee,A.B., Chen,C.Y.A., Zhang,Y. and Shyu,A. Bin (2013) Phosphorylation at intrinsically disordered regions of PAM2 motif-containing proteins modulates their interactions with PABPC1 and influences mRNA fate. *RNA*, **19**, 295–305.

570. Wu,M., Reuter,M., Lilie,H., Liu,Y., Wahle,E. and Song,H. (2005) Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.*, **24**, 4082–93.

571. Moon,D.H., Segal,M., Boyraz,B., Guinan,E., Hofmann,I., Cahan,P., Tai,A.K. and Agarwal,S. (2015) Poly(A)-specific ribonuclease (PARN) mediates 3'-end maturation of the telomerase RNA component. *Nat. Genet.*, **47**, 1482.

572. Dehlin,E., Wormington,M., Körner,C.G. and Wahle,E. (2000) Cap-dependent deadenylation of mRNA. *EMBO J.*, **19**, 1079–86.

573. Berndt,H., Harnisch,C., Rammelt,C., Stöhr,N., Zirkel,A., Dohm,J.C., Himmelbauer,H., Tavanez,J.P., Hüttelmaier,S. and Wahle,E. (2012) Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA*, **18**, 958–972.

574. Son,A., Park,J.-E. and Kim,V.N. (2018) PARN and TOE1 Constitute a 3' End Maturation Module for Nuclear Non-coding RNAs. *Cell Rep.*, **23**, 888–898.

575. Nguyen,D., Grenier St-Sauveur,V., Bergeron,D., Dupuis-Sandoval,F., Scott,M.S. and Bachand,F. (2015) A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA. *Cell Rep.*, **13**, 2244–57.

576. Shukla,S., Bjerke,G.A., Muhlrad,D., Yi,R. and Parker,R. (2019) The RNase PARN Controls the Levels of Specific miRNAs that Contribute to p53 Regulation. *Mol. Cell*, **73**, 1204–1216.

577. Bregman,A., Avraham-Kelbert,M., Barkai,O., Duek,L., Guterman,A. and Choder,M. (2011) Promoter elements regulate cytoplasmic mRNA decay. *Cell*, **147**, 1473–83.

578. Trcek,T., Larson,D.R., Moldó,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.

579. Sun,M., Schwalb,B.R., Pirkl,N., Maier,K.C., Schenk,A., Failmezger,H., Tresch,A. and Cramer,P. (2013) Global Analysis of Eukaryotic mRNA Degradation Reveals Xrn1-Dependent Buffering of Transcript Levels. *Mol. Cell*, **52**, 52–62.

580. Haimovich,G., Medina,D.A., Causse,S.Z., Garber,M., Millá N-Zambrano,G., Barkai,O., Chá,S.N., Pé Rez-Ortín,J.E., Darzacq,X. and Choder,M. (2013) Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis. *Cell*, **153**, 1000–1011.

581. Medina,D.A., Jordán-Pla,A., Millán-Zambrano,G., Chávez,S., Choder,M. and Pérez-Ortín,J.E. (2014) Cytoplasmic 5'-3' exonuclease Xrn1p is also a genome-wide transcription factor in yeast. *Front. Genet.*, **5**, 1–10.

582. Harrison,P.F., Powell,D.R., Clancy,J.L., Preiss,T., Boag,P.R., Traven,A., Seemann,T. and Beilharz,T.H. (2015) PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA*, **21**, 1502–10.

583. Rodríguez-Molina,J.B., Tseng,S.C., Simonett,S.P., Taunton,J. and Ansari,A.Z. (2016) Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. *Mol. Cell*, **63**, 433–444.

584. Shalem,O., Dahan,O., Levo,M., Martinez,M.R., Furman,I., Segal,E. and Pilpel,Y. (2008)

Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.*, **4**, 223.

585. Goler-Baron,V., Selitrennik,M., Barkai,O., Haimovich,G., Lotan,R. and Choder,M. (2008) Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev. Dev.*, **22**, 2022–2027.

586. Shalem,O., Groisman,B., Choder,M., Dahan,O. and Pilpel,Y. (2011) Transcriptome Kinetics Is Governed by a Genome-Wide Coupling of mRNA Production and Degradation: A Role for RNA Pol II. *PLoS Genet.*, **7**, e1002273.

587. Greenberg,M.E., Shyu,A.B. and Belasco,J.G. (1990) Deadenylylation: A mechanism controlling c-fos mRNA decay. *Enzyme*, **44**, 181–192.

588. Slobodin,B., Han,R., Calderone,V., Vrielink,J.A.F.O., Loayza-Puch,F., Elkon,R. and Agami,R. (2017) Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell*, **169**, 326–337.

589. Appel,L.-M., Franke,V., Bruno,M., Grishkovskaya,I., Kasiliauskaite,A., Schoeberl,U.E., Puchinger,M.G., Kostrhon,S., Beltzung,E., Mechtler,K., *et al.* (2020) PHF3 regulates neuronal gene expression through the new Pol II CTD reader domain SPOC. *bioRxiv*, 10.1101/2020.02.11.943159.

590. Harel-Sharvit,L., Eldad,N., Haimovich,G., Barkai,O., Duek,L. and Choder,M. (2010) RNA Polymerase II Subunits Link Transcription and mRNA Decay to Translation. *Cell*, **143**, 552–563.

591. Takahashi,A., Suzuki,T., Soeda,S., Takaoka,S., Kobori,S., Yamaguchi,T., Mohamed,H.M.A., Yanagiya,A., Abe,T., Shigeta,M., *et al.* (2020) The CCR4-NOT complex maintains liver homeostasis through mRNA deadenylation. *Life Sci. alliance*, **3**, e201900494.

592. Heine,G.F., Horwitz,A.A. and Parvin,J.D. (2008) Multiple mechanisms contribute to inhibit transcription in response to DNA damage. *J. Biol. Chem.*, **283**, 9555–9561.

593. Castells-Roca,L., García-Martínez,J., Moreno,J., Herrero,E., Bellí,G. and Pérez-Ortín,J.E. (2011) Heat Shock Response in Yeast Involves Changes in Both Transcription Rates and mRNA Stabilities. *PLoS One*, **6**, e17272.

594. Jadhav,G.P., Kaur,I., Maryati,M., Airhihen,B., Fischer,P.M. and Winkler,G.S. (2015) Discovery, synthesis and biochemical profiling of purine-2,6-dione derivatives as inhibitors of the human poly(A)-selective ribonuclease Caf1. *Bioorg. Med. Chem. Lett.*, **25**, 4219–4224.

595. Airhihen,B., Pavanello,L., Jadhav,G.P., Fischer,P.M. and Winkler,G.S. (2019) 1-Hydroxy-xanthine derivatives inhibit the human Caf1 nuclease and Caf1-containing nuclease complexes via Mg 2+ -dependent binding. *FEBS Open Bio*, **9**, 717–727.

596. Mitchell,D., Ritchey,L.E., Park,H., Babitzke,P., Assmann,S.M. and Bevilacqua,P.C. (2018) Glyoxals as in vivo RNA structural probes of guanine base-pairing. *RNA*, **24**, 114–124.

597. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 -ΔΔC T Method. *METHODS*, **25**, 402–408.

598. Chalabi Hagkarim,N., Ryan,E.L., Byrd,P.J., Hollingworth,R., Shimwell,N.J., Agathanggelou,A., Vavasseur,M., Kolbe,V., Speiseder,T., Dobner,T., *et al.* (2018) Degradation of a Novel DNA Damage Response Protein, Tankyrase 1 Binding Protein 1, following Adenovirus Infection. *J. Virol.*, **92**, e02034-17.

599. Andrews,S. (2010) Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available at www.bioinformatics.babraham.ac.uk/projects/fastqc. Last accessed 23/02/2021.

600. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and

Gingeras,T.R. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

601. GENCODE - Human Release 36. Available at www.gencodegenes.org/human/. Last accessed 23/02/2021.

602. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

603. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.

604. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

605. Liao,Y., Smyth,G.K. and Shi,W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

606. Turner,S. Template for analysis with DESeq2. Available at gist.github.com/stephenturner/f60c1934405c127f09a6 . Last accessed 23/02/2012.

607. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene ontology: Tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.

608. Carbon,S., Douglass,E., Good,B.M., Unni,D.R., Harris,N.L., Mungall,C.J., Basu,S., Chisholm,R.L., Dodson,R.J., Hartline,E., *et al.* (2021) The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.

609. Mi,H., Muruganujan,A., Ebert,D., Huang,X. and Thomas,P.D. (2019) PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.*, **47**, D419–D426.

610. Edmonds,M., Vaughan,M.H. and Nakazato,H. (1971) Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc. Natl. Acad. Sci. U. S. A.*, **68**, 1336–40.

611. Singhania,R., Thorn,G.J., Williams,K., Gandhi,R.D., Daher,C., Barthet-Barateig,A., Parker,H.N., Utami,W., Al-Siraj,M., Barrett,D.A., *et al.* (2019) Nuclear poly(A) tail size is regulated by Cnot1 during the serum response. *bioRxiv*, 10.1101/773432.

612. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

613. Moore,C.L. and Sharp,P.A. (1985) Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell*, **41**, 845–855.

614. Xu,K., Bai,Y., Zhang,A., Zhang,Q. and Bartlam,M.G. (2014) Insights into the structure and architecture of the CCR4-NOT complex. *Front. Genet.*, **5**, 137.

615. Aslam,A., Mittal,S., Koch,F., Andrau,J.C. and Sebastiaan Winkler,G. (2009) The Ccr4-not deadenylase subunits CNOT7 and CNOT8 have overlapping roles and modulate cell proliferation. *Mol. Biol. Cell*, **20**, 3840–3850.

616. Lau,N.-C., Kolkman,A., van Schaik,F.M.A., Mulder,K.W., Pijnappel,W.W.M.P., Heck,A.J.R. and Timmers,H.T.M. (2009) Human Ccr4–Not complexes contain variable deadenylase subunits. *Biochem. J.*, **422**, 443–453.

617. Wagner,E., Clement,S.L. and Lykke-Andersen,J. (2007) An Unconventional Human Ccr4-Caf1 Deadenylase Complex in Nuclear Cajal Bodies. *Mol. Cell. Biol.*, **27**, 1686–1695.

618. Azzouz,N., Panasenko,O.O., Colau,G. and Collart,M.A. (2009) The CCR4-NOT complex physically and functionally interacts with TRAMP and the nuclear exosome. *PLoS One*, **4**,

e6760.

619. Kerr,S.C., Azzouz,N., Fuchs,S.M., Collart,M.A., Strahl,B.D., Corbett,A.H. and Laribee,R.N. (2011) The Ccr4-Not Complex Interacts with the mRNA Export Machinery. *PLoS One*, **6**, e18302.

620. Hutchinson,J.N., Ensminger,A.W., Clemson,C.M., Lynch,C.R., Lawrence,J.B. and Chess,A. (2007) A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, **8**, 39.

621. Chalabi,N., CH,M. and Grand,R.J. (2019) GEO Accession viewer: CNOT1 and Transcriptomic Landscape of a HeLa Cell Line. Available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141496. Last accessed 23/02/2021.

622. Preker,P., Almvig,K., Christensen,M.S., Valen,E., Mapendano,C.K., Sandelin,A. and Jensen,T.H. (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res.*, **39**, 7179–7193.

623. Flynn,R.A., Almada,A.E., Zamudio,J.R. and Sharp,P.A. (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 10460–10465.

624. NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® | NEB. Available at https://www.neb.com/products/e7420-nebnext-ultra-directional-rna-library-prep-kit-for-illumina#Product Information. Last accessed 23/02/2021.

625. Clement,J.Q., Qian,L., Kaplinsky,N. and Wilkinson,M.F. (1999) The stability and fate of a spliced intron from vertebrate cells. *RNA*, **5**, 206–220.

626. Winstall,E., Gamache,M. and Raymond,V. (1995) Rapid mRNA degradation mediated by the c-fos 3' AU-rich element and that mediated by the granulocyte-macrophage colony-stimulating factor 3' AU-rich element occur through similar polysome-associated mechanisms. *Mol. Cell. Biol.*, **15**, 3796–3804.

627. Mou,Z., You,J., Xiao,Q., Wei,Y., Yuan,J., Liu,Y., Brewer,G. and Ma,W.J. (2012) HuR posttranscriptionally regulates early growth response-1 (Egr-1) expression at the early stage of T cell activation. *FEBS Lett.*, **586**, 4319–4325.

628. Herold,A., Klymenko,T. and Izaurralde,E. (2001) NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. *RNA*, **7**, 1768–1780.

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

630. Hett,A. and West,S. (2014) Inhibition of U4 snRNA in human cells causes the stable retention of polyadenylated pre-mRNA in the nucleus. *PLoS One*, **9**, e96174.

631. Klein,P., Oloko,M., Roth,F., Montel,V., Malerba,A., Jarmin,S., Gidaro,T., Popplewell,L., Perie,S., St Guily,J.L., *et al.* (2016) Nuclear poly(A)-binding protein aggregates misplace a pre-mRNA outside of SC35 speckle causing its abnormal splicing. *Nucleic Acids Res.*, **44**, 10929–10945.

632. Liu,Y., Nie,H., Liu,H. and Lu,F. (2019) Poly(A) inclusive RNA isoform sequencing (PAIso−seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nat. Commun.*, **10**, 5292.

633. Yu,F., Zhang,Y., Cheng,C., Wang,W., Zhou,Z., Rang,W., Yu,H., Wei,Y., Wu,Q. and Zhang,Y. (2020) Poly(A)-seq: A method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS One*, **15**, e0234696.

634. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (80-.).*, **324**, 218–223.

635. Herzog,V.A., Reichholf,B., Neumann,T., Rescheneder,P., Bhat,P., Burkard,T.R., Wlotzka,W., von Haeseler,A., Zuber,J. and Ameres,S.L. (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods*, **14**, 1198–1204.

636. Workman,R.E., Tang,A.D., Tang,P.S., Jain,M., Tyson,J.R., Razaghi,R., Zuzarte,P.C., Gilpatrick,T., Payne,A., Quick,J., *et al.* (2019) Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods*, **16**, 1297–1305.

637. Bentley,D.R., Balasubramanian,S., Swerdlow,H.P., Smith,G.P., Milton,J., Brown,C.G., Hall,K.P., Evers,D.J., Barnes,C.L., Bignell,H.R., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.

638. Eid,J., Fehr,A., Gray,J., Luong,K., Lyle,J., Otto,G., Peluso,P., Rank,D., Baybayan,P., Bettman,B., *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science (80-.).*, **323**, 133–138.

639. Garalde,D.R., Snell,E.A., Jachimowicz,D., Sipos,B., Lloyd,J.H., Bruce,M., Pantic,N., Admassu,T., James,P., Warland,A., *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods*, **15**, 201–206.

640. MiSeq Reagent Kit V2 (Information on currently available read lengths). www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/miseq-reagent-kit-v2.html. Last accessed 23/02/2021.

641. Tan,G., Opitz,L., Schlapbach,R. and Rehrauer,H. (2019) Long fragments achieve lower base quality in Illumina paired-end sequencing. *Sci. Rep.*, **9**, 2856.

642. Schirmer,M., D'Amore,R., Ijaz,U.Z., Hall,N. and Quince,C. (2016) Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*, **17**, 125.

643. Illumina sequencing quality scores. Available at www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html. Last accessed 23/02/2021.

644. Buck,D., Weirather,J.L., de Cesare,M., Wang,Y., Piazza,P., Sebastiano,V., Wang,X.J. and Au,K.F. (2017) Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000 Res.*, **6**, 100.

645. Cui,J., Shen,N., Lu,Z., Xu,G., Wang,Y. and Jin,B. (2020) Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the Arabidopsis transcriptome. *Plant Methods*, **16**, 85.

646. Bilska,A., Kusio-Kobiałka,M., Krawczyk,P.S., Gewartowska,O., Tarkowski,B., Kobyłecki,K., Nowis,D., Golab,J., Gruchota,J., Borsuk,E., *et al.* (2020) Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. *Nat. Commun.*, **11**, 2032.

647. Krause,M., Niazi,A.M., Labun,K., Torres Cleuren,Y.N., Müller,F.S. and Valen,E. (2019) TailFindR: Alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA*, **25**, 1229–1241.

648. Hebert,P.D.N., Braukmann,T.W.A., Prosser,S.W.J., Ratnasingham,S., DeWaard,J.R., Ivanova,N. V., Janzen,D.H., Hallwachs,W., Naik,S., Sones,J.E., *et al.* (2018) A Sequel to Sanger: Amplicon sequencing that scales. *BMC Genomics*, **19**, 219.

649. Chang,H. (2017) GitHub - hyeshik/tailseeker: Software for measuring poly(A) tail length and 3'-end modifications using a high-throughput sequencer. Available at

https://github.com/hyeshik/tailseeker. Last accessed 23/02/2021.

650. Elder,P.K., Schmidt,L.J., Ono,T. and Getz,M.J. (1984) Specific stimulation of actin gene transcription by epidermal growth factor and cycloheximide. *Proc. Natl. Acad. Sci. U. S. A.*, **81**, 7476–7480.

651. Dittman,W.A., Kumada,T. and Majerus,P.W. (1989) Transcription of thrombomodulin mRNA in mouse hemangioma cells is increased by cycloheximide and thrombin. *Proc. Natl. Acad. Sci. U. S. A.*, **86**, 7179–7182.

652. Hsin,H. and Kenyon,C. (1999) Signals from the reproductive system regulate the lifespan of C. elegans. *Nature*, **399**, 362–366.

653. Arantes-Oliveira,N., Apfeld,J., Dillin,A. and Kenyon,C. (2002) Regulation of life-span by germ-line stem cells in Caenorhabditis elegans. *Science (80-.).*, **295**, 502–505.

654. Labbadia,J. and Morimoto,R.I. (2015) Repression of the Heat Shock Response Is a Programmed Event at the Onset of Reproduction. *Mol. Cell*, **59**, 639–650.

9 Appendix

9.1 qPCR primer validation

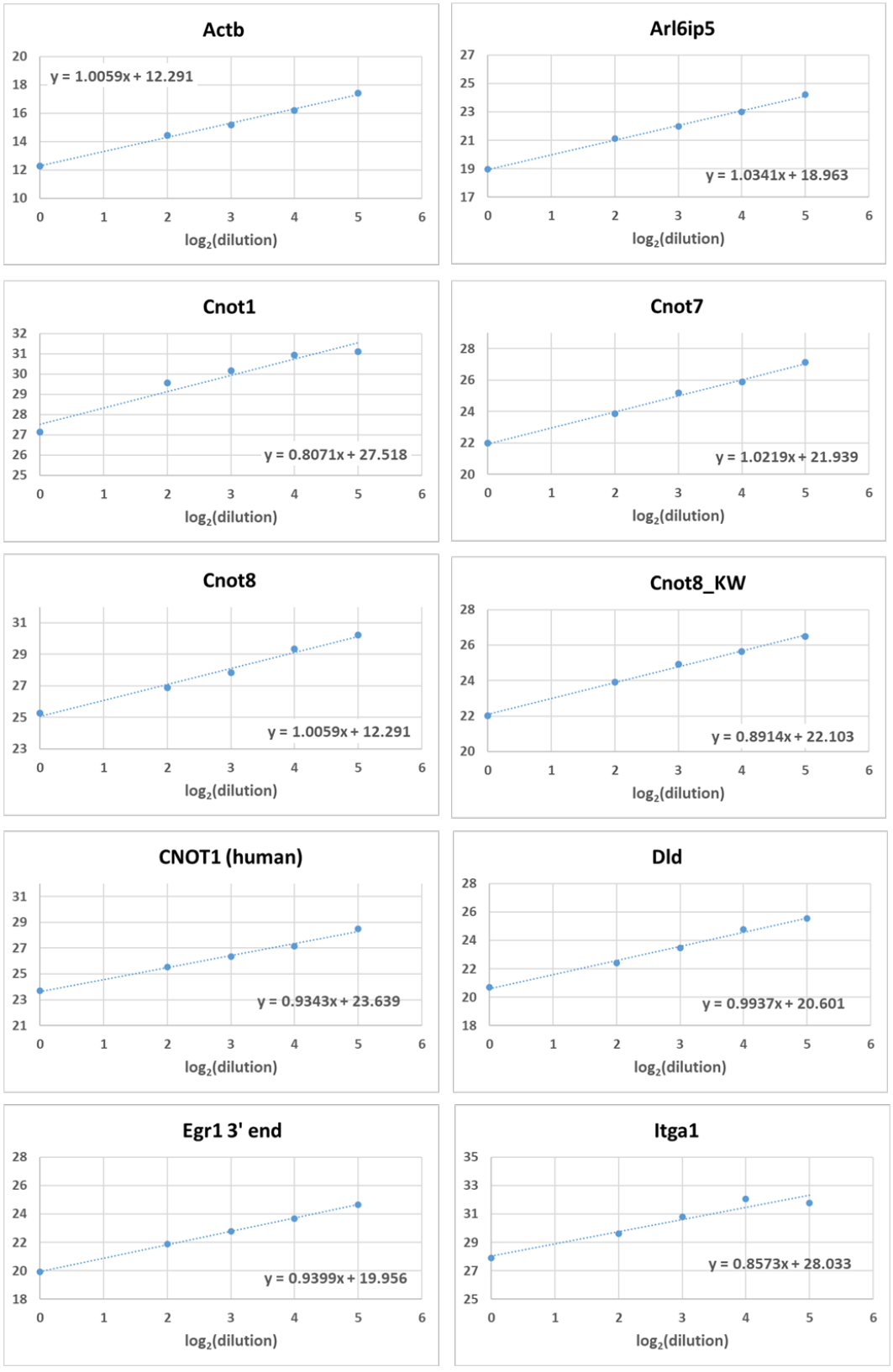


Figure A.1 Validation of qPCR primer pairs. Continued overleaf.

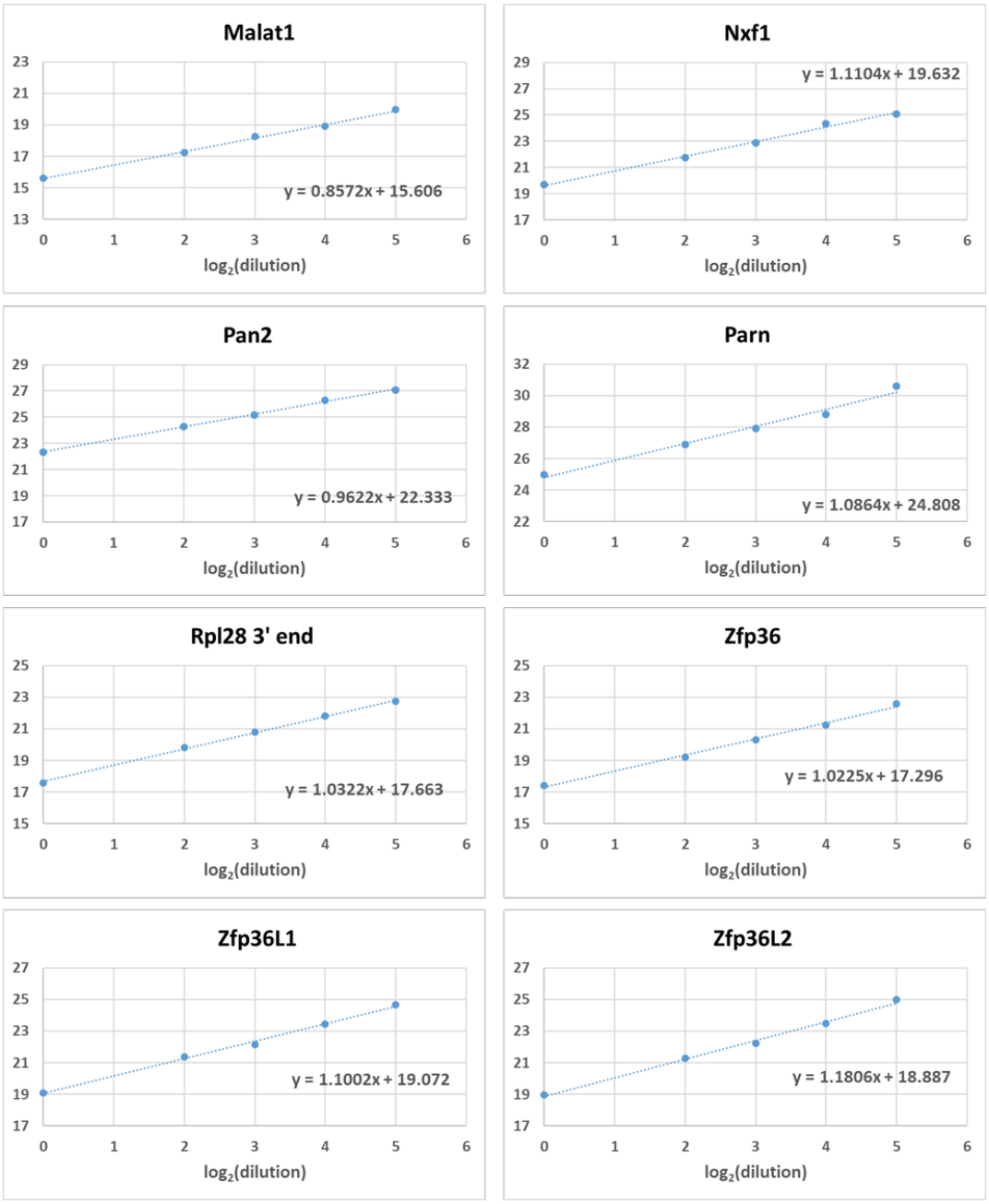


Figure A.1 continued.

9.2 Poly(A) tails are regulated in the *C. elegans* heat shock response

Related to the observations of dynamic poly(A) regulation in NIH 3T3 and HEK293 cells in chapter 3, the opportunity arose to also examine poly(A) tail lengths of mRNA from heat shocked *C. elegans* worms through a collaboration with Olivia Casanueva's group at the Babraham Institute. Two strains of worm were used, gon-2 and glp-1; while both worms lacked gonads (to avoid contamination of somatic tissues with material known to experience large changes during aging), glp -1 worms were completely devoid of a proliferating germline. This additional absence of a germline prevents production of signals which promote *C. elegans* ageing; glp-1 worms therefore exhibit extended lifespans and may retain a more agile heat shock response (652–654).

Heat shock experiments and RNA isolation were performed by Olivia Casanueva's group. Young (1 day old) or old (2 day) worms were transferred from 25 to 34 °C for 30 minutes, with samples taken before (nHs) and during (-15) heat shock as well as at a number of points following return to normal growth conditions. Nested PAT assays showed an increase in poly(A) tail length in response to heat shock for most transcripts in both worm strains, with a return to pre-stimulation poly(A) lengths after around 2 hours (figure A.2 A. B). Y45F10D.4 is not transcriptionally induced by heat shock and showed little variation across the heath shock response in either strain. Return to short poly(A) tails was slower in 2 day old worms in both strains. Since longer tails of induced mRNAs may confer greater stability (figure 3.6), this was indicative of a less tightly controlled heat shock response. These data suggest that the poly(A) tail is regulated in other organisms and during other transcriptional responses, and point to a role for polyadenylation control in maintaining the agility of such responses.

Figure A.2 (overleaf) Poly(A) tail length is regulated in the C. elegans response to heat shock. Nested PAT assays using agarose gel stained with SYBR safe were carried out on RNA isolated following heat shock of **A)** gon-2 worms which were sterile but retained a germline and **B)** glp-1 worms which did not contain a proliferating germline and were long-lived. Young (1 day) or old (2 day) old worms were treated, with the **young worm samples on the left for each time point**.

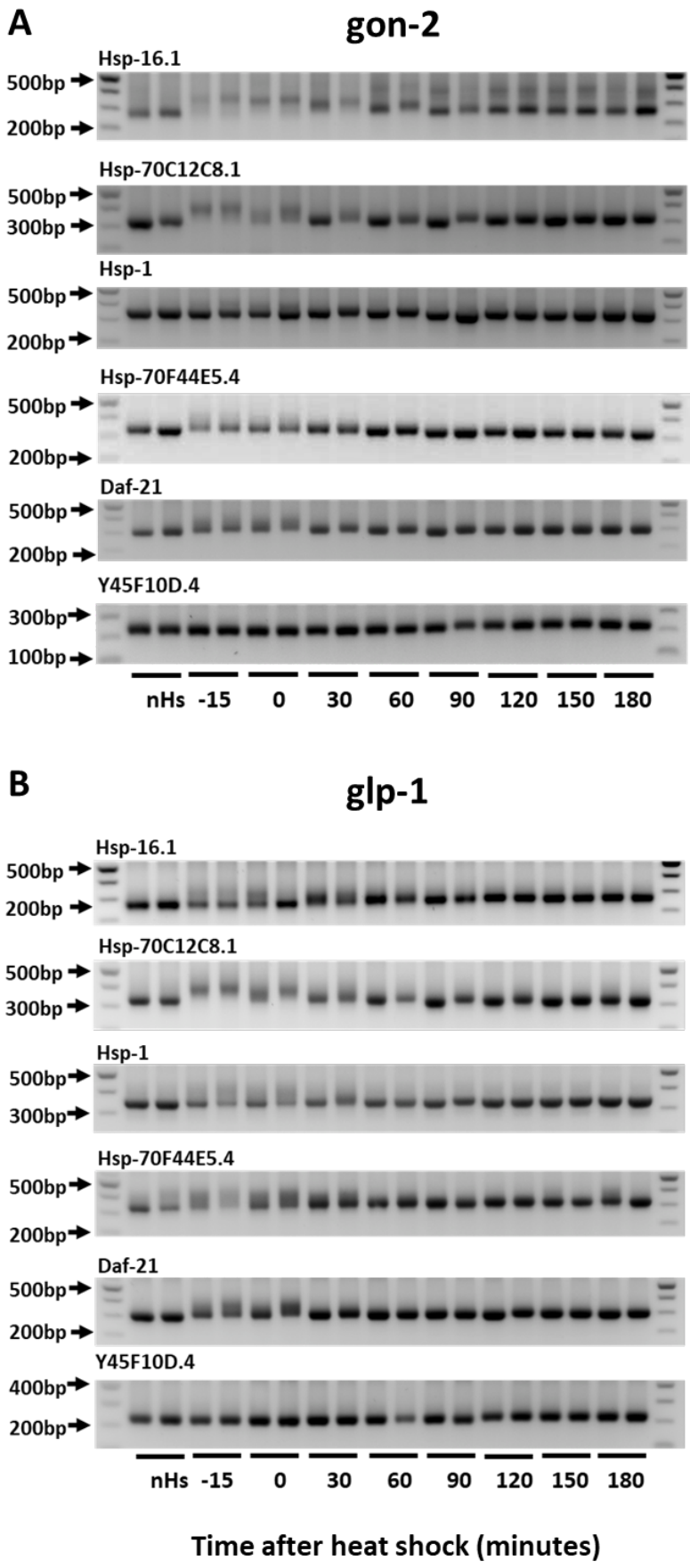


Figure A.2 Poly(A) tail length is regulated in the C. elegans response to heat shock.

In chapter 4, Cnot1 knockdown was shown to increase initial mRNA poly(A) length in NIH 3T3 cells. To see whether this extended to other metazoans, the effect of knocking down the Cnot1 orthologue, ntl-1, on poly(A) tail lengths in *C.elegans* worms was investigated. glp-1 worms, which lacked proliferating germline cells and were consequently resistant to ageing, were subject to ntl-1 knockdown or a control transfection. They later underwent heat shock at 34 °C for 30 minutes (from growth at 25 °C), and RNA was isolated either immediately or at *n* hours after the end of this heat shock. While slower deadenylation was detectable, it was hard to draw a conclusion on whether poly(A) tail length immediately after heat shock differed between control and knockdown worms (figure A.3). In contrast to findings in NIH 3T3 cells (figure 4.1), early poly(A) length appeared shorter in the ntl-1 knockdown worms for both hsp-1 and daf-21, and to a lesser extent in hsp-16. Interestingly, two of the genes thought not to be transcriptionally induced by heat shock still responded to heat shock with increases in polyadenylation (figure A.3).

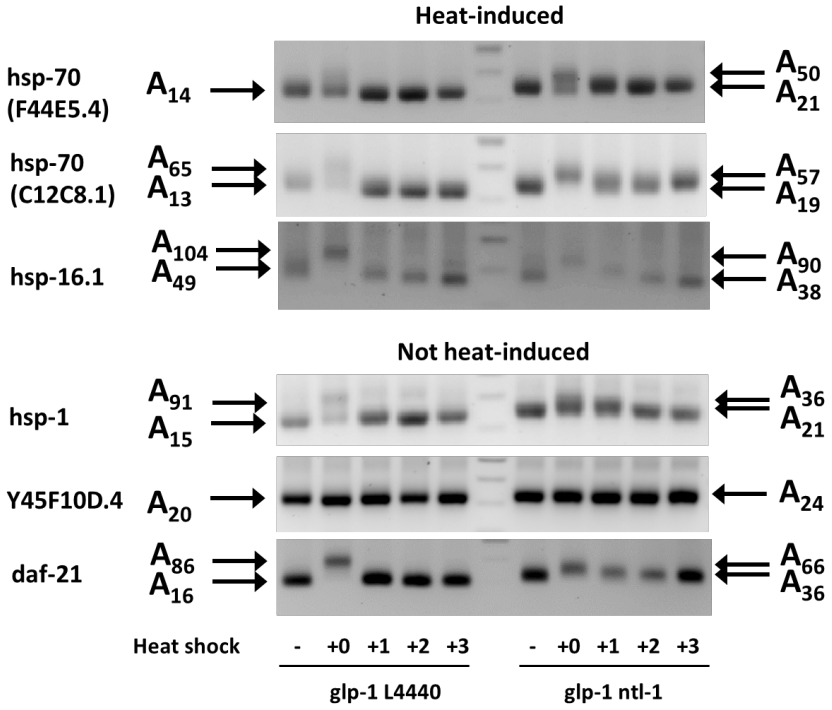


Figure A.3 Preliminary data suggest that the C. elegans CNOT1 orthologue, ntl-1, may regulate initial poly(A) tail length in the heat shock response. *C. elegans* worms were grown in *E. Coli* expressing an empty vector control (L4440) or ntl-1 RNAi (ntl-1), heat shocked for 30 minutes then returned to standard growth conditions. Worms were harvested at the times indicated (hours), where +0 indicates harvesting immediately after heat shock. RNA was prepared by Juan Rodriguez and subjected to the PAT assay using agarose gel stained with SYBR safe. Arrows indicate maximum and minimum modal poly(A) tail sizes as determined using quantitative gel scanning.

9.3 PIPS reflective statement

Note to examiners:

This statement is included as an appendix to the thesis in order that the thesis accurately captures the PhD training experienced by the candidate as a BBSRC Doctoral Training Partnership student.

The Professional Internship for PhD Students is a compulsory 3-month placement which must be undertaken by DTP students. It is usually centred on a specific project and must not be related to the PhD project. This reflective statement is designed to capture the skills development which has taken place during the student's placement and the impact on their career plans it has had.

PIPS Reflective Statement

During my placement I worked on *Cupriavidus necator* and *metallidurans*, with the aim of producing a monomer of interest to Biome Bioplastics. Although the placement was with Biome, the majority of the placement was undertaken in the Centre for Biomolecular Sciences (University of Nottingham) under the supervision of Dr Samantha Bryan.

The project involved using heterologous recombination to knock out two enzymes ('G' and 'H') that allowed the precursor to enter central metabolism, and heterologously expressing an enzyme ('X') that converts the precursor to the monomer of interest. I created plasmids for expression of X behind two different promoters, as well as producing one for knockout of G and H in *C.necator* using the tetA-sacB selection system. TetA confers tetracycline resistance after the first crossover event and SacB produces a toxic compound in the presence of sucrose, stimulating a second crossover event. I also attempted to use a previously constructed plasmid to knock out G and H in *C.metallidurans* using the pyrF selection system. PyrF is required for synthesis of uracil if cells are grown on minimal media, but also produces a toxic compound when presented with FOA. This plasmid also contained a chloramphenicol resistance gene.

I successfully expressed the enzyme X in two different *C.necator* strains and was able to produce the desired monomer in very small quantities (~1 mg/mL).

When trying to knock out G and H in *C.necator*, I managed to confirm a single cross, but was not able to force the second crossover event in any of the colonies screened before my placement finished. For *C.metallidurans* the Δ*pyrF* strain was too resistant to

chloramphenicol to force the second crossover, and when I tried to grow it on minimal media to force use of the *pyrF* gene, no cells survived.

Towards the end of my placement I also visited the facility in Southampton where Biome create and test new bioplastic formulations. I learnt about product development, testing procedures and some of the challenges faced when trying to create a product that meets the customer's needs, is not too expensive, and passes composting regulations.

Through my placement I added cloning to my molecular biology skillset and worked with organisms I was previously unfamiliar with. I faced new challenges and frustrations associated with trying to create a particular product, but enjoyed the experience of aiming to create something rather than trying to find something out. It was also interesting to experience working in a different research lab.

In terms of effects on my career direction, it has made me consider staying in a different area of academic research when I was previously considering abandoning research altogether. It also made me aware of the ability to change research area after the PhD, and it was encouraging to hear of the different routes people had taken into roles at Biome.