

From the Department of Oncology-Pathology
Karolinska Institutet, Stockholm, Sweden

Coordination of gene expression programs

Julie Lorent



**Karolinska
Institutet**

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Eprint AB 2020

© Julie Lorent, 2020

ISBN 978-91-7831-697-7

Coordination of gene expression programs

THESIS FOR DOCTORAL DEGREE (Ph.D.)

BioClinicum, J3:11 Birger & Margareta Blombäck, Karolinska University Hospital,
Solnavägen 30, Solna

Friday, February 14th, 2020 at 9:00

By

Julie Lorent

Principal Supervisor:

Associate Professor Ola Larsson
Karolinska Institutet
Department of Oncology-Pathology

Co-supervisor:

Professor Janne Lehtiö
Karolinska Institutet
Department of Oncology-Pathology

Opponent:

Professor Jack D. Keene
Duke University Medical Center
Department of Molecular Genetics and
Microbiology

Examination Board:

Associate Professor Sven Nelander
Uppsala University
Department of Immunology, Genetics and
Pathology

Professor Cecilia Williams
Kungliga Tekniska Högskolan
Department of Protein Science

Associate Professor Marc Friedländer
Stockholm University
Department of Molecular Biosciences

ABSTRACT

Most cellular processes depend on the activity and interactions of proteins. The proteome, i.e. the entire set of proteins in a specific condition, is shaped by regulation of transcription, mRNA-degradation, -processing, -storage, -translation and protein degradation. Cancer cells are known to hijack gene expression processes, including the translation machinery, for their growth and survival. This occurs as a result of converging oncogenic signaling pathways which impinge on translation factors to selectively modulate synthesis of cancer-related proteins.

Our understanding of mechanisms by which oncogenic pathways dynamically control their targets' translational activity is limited and could be extended by transcriptome-wide studies of changes in translation efficiency. In **Paper I**, we developed anota2seq which allows for statistical analysis of such data. Using a simulation approach, we showed that anota2seq constitutes an improvement compared to other methods for identification of genes under translational regulation.

The relative contribution of transcriptional and translational regulation to proteome modulation has been extensively debated. This raises the interest in studies integrating data on multiple levels of gene expression regulation. In **Paper II**, we study the role of estrogen receptor alpha (ER α), a transcription factor that is commonly targeted in hormone-dependent cancers, in coordinating transcriptional alterations with control at the level of translation. Upon ER α depletion in a prostate cancer model, we observed massive translational offsetting whereby the translational output remains unchanged despite changes in mRNA levels. To characterize mechanisms underlying translational offsetting, we extended the scope of the anota2seq method (**Paper I**) to also identify genes regulated by this underappreciated mode of gene expression regulation. Next, our detailed mechanistic study revealed that upon ER α depletion, mRNAs whose levels are reduced but translationally offset have less structured 5'UTRs and are devoid of miRNA target sites and thus cannot be influenced by such translational repressors. In contrast, transcripts which were upregulated but offset at the level of translation are enriched in codons requiring U34-modified tRNAs for their translation. We finally demonstrated that ER α impacts the levels of such modified tRNAs.

Cancer is a highly heterogeneous disease. In our studies of translational control, we are reaching the limits of reasonable inference when extending conclusions from experiments in cell lines into clinical settings. However, experimental methods to quantify translomes such as polysome-profiling, are not suitable for samples with low RNA input such as tissue samples from cancer patients. **Paper III** presents an optimization of the polysome-profiling method, compares it with the classical approach and validates that this new approach is suitable to study novel mechanisms regulating mRNA translation in large collections of tissue samples.

LIST OF SCIENTIFIC PAPERS

- I. Oertlin C, Lorent J, Murie C, Furic L, Topisirovic I[§], Larsson O[§].
Generally applicable transcriptome-wide analysis of translation using anota2seq.
Nucleic Acids Res. 2019 Jul 9;47(12):e70.
doi: <https://doi.org/10.1093/nar/gkz223>

- II. Lorent J[#], Kusnadi EP[#], van Hoef V, Rebello RJ, Leibovitch M, Ristau J, Chen S, Lawrence MG, Szkop KJ, Samreen B, Balanathan P, Rapino F, Close P, Bukczynska P, Scharmann K, Takizawa I, Risbridger GP, Selth LA, Leidel SA, Lin Q, Topisirovic I[§], Larsson O[§], Furic L[§].
Translational offsetting as a mode of estrogen receptor α -dependent regulation of gene expression.
EMBO J. 2019 Dec 2;38(23):e101323. Publisher: John Wiley and Sons.
doi: <https://doi.org/10.15252/embj.2018101323>

- III. Liang S[#], Bellato HM[#], Lorent J[#], Lupinacci FCS, Oertlin C, van Hoef V, Andrade VP, Roffé M, Masvidal L[§], Hajj GNM[§], Larsson O[§].
Polysome-profiling in small tissue samples.
Nucleic Acids Res. 2018 Jan 9;46(1):e3.
doi: <https://doi.org/10.1093/nar/gkx940>

[#] Equal contributions

[§] Corresponding authors

PUBLICATIONS NOT INCLUDED IN THESIS

Mao Y, van Hoef V, Zhang X, Wennerberg E, Lorent J, Witt K, Masvidal L, Liang S, Murray S, Larsson O[§], Kiessling R[§], Lundqvist A[§].

IL-15 activates mTOR and primes stress-activated gene expression leading to prolonged antitumor capacity of NK cells.

Blood. 2016 Sep 15;128(11):1475-89.

Ramón Y Cajal S, Capdevila C, Hernandez-Losa J, De Mattos-Arruda L, Ghosh A, Lorent J, Larsson O, Aasen T, Postovit LM, Topisirovic I[§].

Cancer as an ecomolecular disease and a neoplastic consortium.

Biochim Biophys Acta Rev Cancer. 2017 Dec;1868(2):484-499.

William M, Leroux LP, Chaparro V, Lorent J, Graber TE, M'Boutchou MN, Charpentier T, Fabié A, Dozois CM, Stäger S, van Kempen LC, Alain T, Larsson O, Jaramillo M[§].

eIF4E-binding proteins 1 and 2 limit macrophage anti-inflammatory responses through translational repression of IL-10 and cyclooxygenase-2.

J Immunol. 2018 Jun 15;200(12):4102-4116.

Leroux LP, Lorent J[#], Graber TE[#], Chaparro V, Masvidal L, Aguirre M, Fonseca BD, van Kempen LC, Alain T, Larsson O, Jaramillo M[§].

The protozoan parasite toxoplasma gondii selectively reprograms the host cell translatoe.

Infect Immun. 2018 Aug 22;86(9).

Equal contributions

§ Corresponding authors

Contents

1 Prolegomenon	1
1.1 Translational control	1
1.1.1 Introduction about the central dogma of molecular biology . . .	1
1.1.2 Translation of an mRNA	2
1.1.2.1 Translation initiation	2
1.1.2.2 Translation elongation	4
1.1.2.3 Translation termination	5
1.1.3 Efficiency of translation of an mRNA	5
1.1.4 Experimental methods to measure differential translational ef- ficiency	7
1.1.4.1 Polysome-profiling	7
1.1.4.2 Ribosome-profiling	9
1.1.5 Analytical methods for transcriptome-wide analysis of differen- tial translation	10
1.2 Coordination of gene expression programs	14
1.2.1 Regulation of gene expression at multiple levels	15
1.2.1.1 When mRNA synthegadosome and translation act in concert	16
1.2.1.2 When mRNA translation compensates modulations at the mRNA level	17
1.2.1.3 Role of estrogen receptor alpha in breast and prostate cancer	18
1.2.2 Coordinated regulation of sets of mRNAs	21
1.2.2.1 Regulatory RNA elements and trans-acting factors . .	22
1.2.2.1.1 microRNA and RNA-binding proteins	22
1.2.2.1.2 5'UTR elements	22
1.2.2.1.3 Codon-dependent elongation rates and the role of tRNA availability	23
1.2.2.2 Regulatory pathways impinging on translational out- puts of subsets of mRNAs	24
1.2.2.2.1 mTOR-sensitive mRNAs	24
1.2.2.2.2 Regulation of selective translation by MAPK signaling and eIF4E phosphorylation	26
2 Present investigations	27
2.1 Aims of the thesis	27
2.2 Results and discussion	28
2.2.1 Assessing differential translation and buffering using anota2seq	28

2.2.2	Transcriptome-wide analysis of mRNA translation in tissue samples	31
2.2.3	Investigating the mechanisms by which the transcription factor estrogen receptor alpha affects gene expression at the translational level	33
2.2.3.1	Reflections on ethical considerations in bioinformatics projects	36
3	Conclusion	41
	Acknowledgements	43
	References	45

List of abbreviations

4E-BP	eIF4E binding protein.
a.u.	arbitrary units.
ALKBH8	alkB homolog 8, tRNA methyltransferase.
ATP	adenosine triphosphate.
AUC	area under the receiver operating characteristic curve.
BCL-XL	B-cell lymphoma extra large.
BRAF	B-Raf proto-oncogene, serine/threonine kinase.
CCNA2	cyclin-A2.
CCND3	cyclin D3.
CDK5	cyclin-dependent-like kinase 5.
CTU	cytosolic thiouridylase subunit.
DEK	DEK Proto-Oncogene.
DNA	deoxyribonucleic acid.
eEF	eukaryotic elongation factor.
eIF	eukaryotic initiation factor.
ELP3	elongator acetyltransferase complex subunit 3.
EMT	epithelial-to-mesenchymal transition.
ER	estrogen receptor.
eRF	eukaryotic peptide chain release factor.
ERK	extracellular signal-regulated kinase.
ER α	estrogen receptor alpha.
ER β	estrogen receptor beta.
FC	fold change.
FDR	false discovery rate.
GCN2	general control non-derepressible 2.
GDP	guanosine diphosphate.
GLM	generalized linear model.

GTP	guanosine triphosphate.
HER2	human epidermal growth factor receptor 2.
HRI	heme-regulated inhibitor.
HuR	human antigen R.
m7G	7-methylguanosine 5'-cap.
MAPK	mitogen-activated protein kinase.
MAPKKK	mitogen-activated protein kinase kinase kinase.
mcm ⁵ s ² U	5-methoxycarbonyl-methyl-2-thiouridine.
MDM2	Mouse Double Minute 2.
MEF	mouse embryonic fibroblast.
MEK	MAPK/ERK kinase (also known as MAPKK).
MET	mesenchymal epithelial transition factor.
miRNA	microRNA.
MNK	MAPK-interacting kinase.
mRNA	messenger RNA.
mRNP	messenger ribo-nucleoprotein particle.
mTOR	mammalian/mechanistic target of rapamycin.
mTORC	mTOR complex.
MYC	MYC proto-oncogene.
p53	tumor protein p53.
PABP	poly(A)-binding protein.
PDC4	programmed cell death 4.
PDK1	3-phosphoinositide-dependent protein kinase 1.
PERK	PKR-like endoplasmic reticulum kinase.
PI3K	phosphatidylinositol-3-OH kinase.
PIC	pre-initiation complex.
PIP2	phosphatidylinositol-4,5-bisphosphate.
PIP3	phosphatidylinositol-3,4,5-trisphosphate.
PKR	double-stranded RNA-dependent protein kinase.
Pol II	RNA polymerase II.
Pol III	RNA polymerase III.
pRB	retinoblastoma protein.
PTEN	phosphatase and tensin homologue.
RBP	RNA-binding protein.
RFM	ribosome flow model.
RHEB	RAS homologue enriched in brain.

RNA	ribonucleic acid.
RNase	ribonuclease.
RNAseq	RNA sequencing.
RPF	ribosome-protected fragment.
rRNA	ribosomal RNA.
RT-qPCR	quantitative reverse transcription polymerase chain reaction.
S6K	ribosomal S6 kinase.
TC	ternary complex.
TE	translational efficiency.
TOP	terminal oligopyrimidine.
tRNA	transfer RNA.
TSC	tuberous sclerosis.
TSS	transcription start site.
U34	Uridine 34.
uORF	upstream open reading frame.
UTR	untranslated region.
VEGF	vascular endothelial growth factor.
WB	Western blotting.

1 Prolegomenon

1.1 Translational control

1.1.1 Introduction about the central dogma of molecular biology

An organism's genetic information is encoded in the deoxyribonucleic acid (DNA) located in its cells' nuclei. Depending on its cell type and on which state a cell is in, specific locations (genes) along the DNA will be transcribed (expressed). This entails synthesizing a necessary amount of temporary copies of the DNA region called transcripts. About 62% of the human genome is transcribed and processed further (Djebali et al. 2012). Of these transcripts, 2-5% are messenger RNAs (mRNAs) which encode proteins (Carninci et al. 2005). The process of "converting" an mRNA template into a protein is called mRNA translation (**Figure 1**); this thesis will mainly be focused around studying regulation of gene expression at the level of mRNA translation.

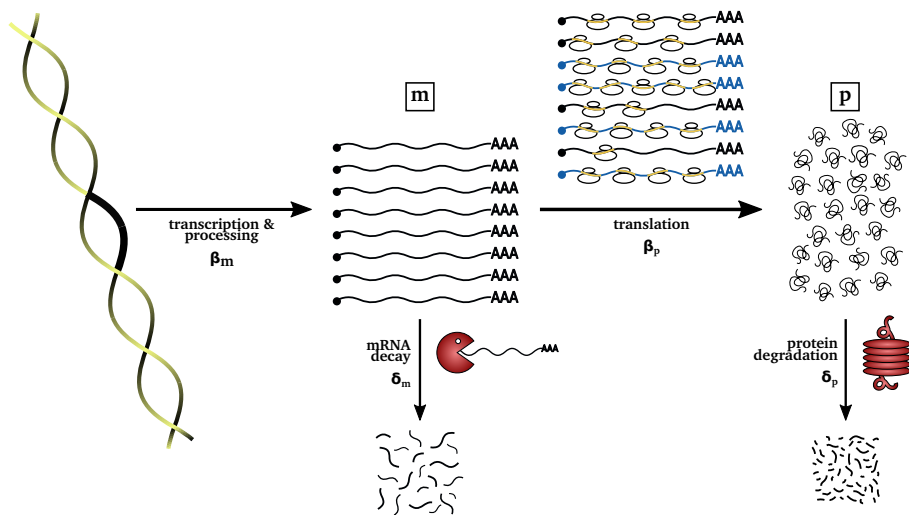


Figure 1: Main steps of gene expression regulation. This figure illustrates, for one gene, the processes of the central dogma of molecular biology. The protein level (p) is shaped by regulation of the rates of different steps including rate of transcription and processing (β_m), of mRNA-decay (δ_m), -storage, -translation (β_p) and protein degradation (δ_p). In the figure, m is the mRNA abundance. At a given time t , $p(t) = (\beta_m(t) - \delta_m(t)) \cdot \beta_p(t) - \delta_p(t)$. When measuring translational efficiency, some methods such as polysome-profiling (see section 1.1.4.1) estimate the proportion of efficiently translated mRNAs (mRNAs associated with many ribosomes, colored in blue) among all mRNAs. Other methods, such as ribosome-profiling (section 1.1.4.2), count the number of ribosome footprints (colored in yellow).

A cell's activity is characterized by quantity and interactions of proteins. The abundance of a protein is determined by the rate of transcription of its corresponding gene,

translation of the mRNA intermediate as well as rate of mRNA and protein degradation (**Figure 1**). These processes constitute the so-called central dogma of molecular biology (Crick 1970). At steady state, from one gene product to another, the range of transcription and translation rates is wide. Between 0.1 and 100 mRNAs are synthesized every hour while translation rates range from 10 to 10 000 proteins per mRNA per hour (Schwanhäusser et al. 2011; Li et al. 2014; Hausser et al. 2019). mRNA and protein median half-lives have been measured to be around 11 and 35.5 hours respectively and vary over a 10-fold range (Cambridge et al. 2011; Gregersen et al. 2014; Hausser et al. 2019). In terms of cellular energy consumption, mRNA translation is the most demanding step (~28% of the total adenosine triphosphate (ATP) production) in normal proliferating cells (Rolfe and Brown 1997). In uncontrolled proliferation contexts such as in cancer cells, control over the translational machinery therefore becomes essential (Robichaud et al. 2018).

1.1.2 Translation of an mRNA

Once transcribed, pre-mRNAs are processed (5'capping, 3' polyadenylation, intron splicing) and exported from the nucleus to the cytoplasm of the cell. During translation, mRNA messages consisting of nucleotide sequences are converted into protein sequences consisting of amino acid sequences (Jacob and Monod 1961). On an mRNA, the part which codes for the protein is called coding sequence and is enclosed between the 5' and the 3' untranslated regions (UTRs). These regions typically include regulatory sequences which are used for translational control as will be described later.

Before protein synthesis starts, the ribosome which will translate the mRNA and allow for the assembly of the amino acid sequence, scans the 5' UTR until recognition of the beginning of the coding sequence (Kozak 1989). This step is called translation initiation. Each triplet of nucleotides (codon) along the coding sequence will then be paired to the anticodon of a transfer RNA (tRNA) which will have previously been charged with the corresponding amino acid to be added to the polypeptide chain (Crick 1958; Chapeville et al. 1962). This occurs from the AUG start codon until the recognition of a stop codon (UAG, UAA or UGA) which terminates translation.

1.1.2.1 Translation initiation More specifically, for translation to be initiated, several translation factors have to coordinate and the two subunits of the ribosome have to assemble at the start codon (**Figure 2**). On one hand a TC consisting of eIF2, initiator tRNA and GTP is formed. On the other hand, eIF4E binds the mRNA-cap and together with eIF4G and eIF4A form the eIF4F complex. This allows for recruitment of the 43S pre-initiation complex (40S small ribosome subunit, TC and additional initiation factors including eIF3) to the mRNA template. 5' UTR scanning starts towards the

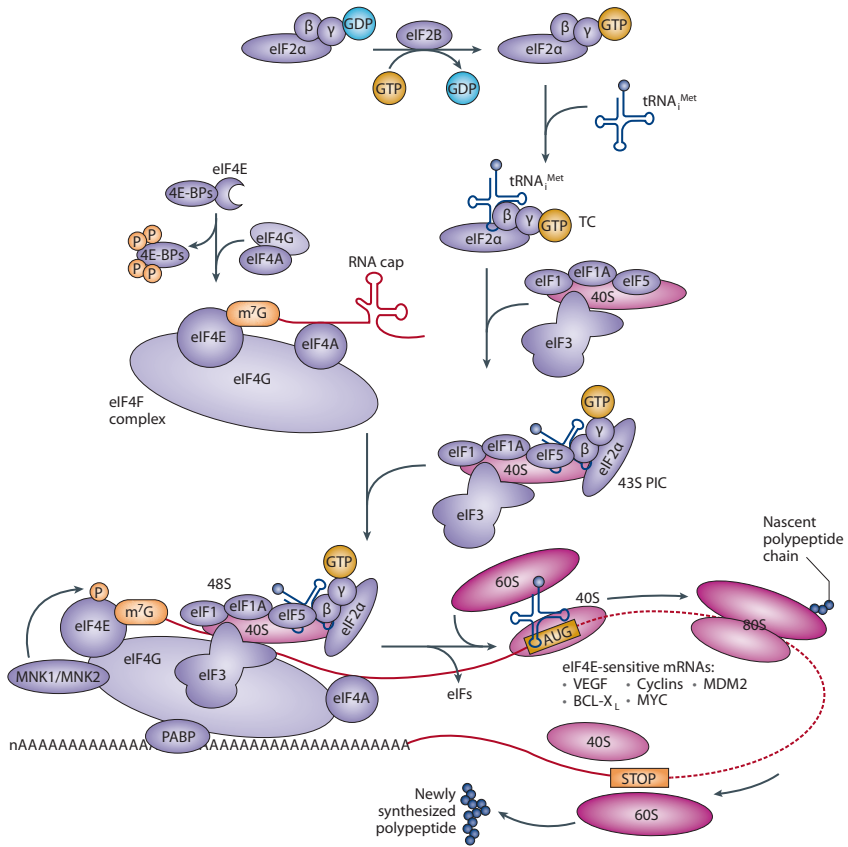


Figure 2: Translation initiation. Initiation is the rate-limiting phase of translational regulation. The formation of the 43S pre-initiation complex (PIC) assembles the 40S ribosomal subunit with eukaryotic initiation factor (eIF)1, eIF1A, eIF3, eIF5 and the ternary complex (TC) (eIF2 (containing α -, β - and γ -subunits), initiator methionyl tRNA and guanosine triphosphate (GTP)). eIF4E binds the 5'-cap of mRNAs and associates with eIF4G, a large scaffolding protein and eIF4A, a DEAD box ribonucleic acid (RNA) helicase. Assembly of such an eIF4F complex (eIF4E, eIF4G, eIF4A) facilitates the recruitment of ribosomes on the mRNA and forms a 48S PIC. Circularization of mRNAs via interaction between eIF4G and the poly(A)-binding protein (PABP) helps stability of the complex and increases translation efficiency.

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature Reviews Drug Discovery. Bhat M et al. Targeting the translation machinery in cancer. *Nat Rev Drug Discov.* 2015 Apr;14(4):261-78. doi: 10.1038/nrd4505. ©2015

start codon where the 60S big ribosomal subunit joins (Gingras et al. 1999; Jackson et al. 2010; Hinnebusch and Lorsch 2012; see also **Figure 2**).

1.1.2.2 Translation elongation Once the 80S ribosome is assembled, translation elongation and the formation of the polypeptide chain can start. For each codon, eukaryotic elongation factor (eEF)1A:aminoacyl tRNA:GTP are presented to the ribosome to its A site (site for incoming amino-acid charged tRNA). Upon codon recognition, GTP is hydrolyzed, the tRNA is accommodated to the A-site and a peptide bond is formed between the amino acid of the A-site tRNA and the amino acid of the tRNA in the peptidyl (P) site. The polypeptide chain is transferred to the A-site before translocation is assured by eEF2 binding and GTP hydrolysis (Dever and Green 2012). The uncharged tRNA which was in the P-site is then moved to the exit E-site and the charged tRNA and polypeptide chain are moved to the P-site.

Before tRNAs can be efficiently used in translation, they undergo a tightly controlled biosynthesis involving multiple steps: RNA polymerase III (Pol III) transcription, removal of the 5' leader and 3' trailer sequences, addition of CCA on the 3' end, splicing, modification and aminoacylation (Phizicky and Hopper 2010). tRNAs are ubiquitous and very abundant (15% of the total RNA) which facilitated their identification (Hoagland et al. 1958) even before understanding the concept of mRNA and the machinery of protein synthesis (Brenner et al. 1961). Notwithstanding major advances in the characterization of many RNAs notably thanks to breakthrough in sequencing technologies, a complete map of heavily modified RNAs such as tRNAs is still lacking (Juhling et al. 2009; Czerwoniec et al. 2009). Among the important tRNA modifications which are currently extensively studied, are the modifications which enable wobbling (non Watson-Crick base pairing between the first position of the tRNA anticodon (position 34) and the third position of the codon). Because of the wobbling rules, the same tRNA can decode several codons. In theory, a minimum of 30 tRNAs would be absolutely required. Yet, in human, tRNAs with 49 different anticodons are represented; these tRNAs are encoded by 513 genes (Chan and Lowe 2009). Not all tRNA genes are expressed at the same level and ribosomal speed at a specific codon will depend on abundance of the corresponding tRNA(s), the cellular demand for this/these tRNA species as well as the nature of the pairing (Watson-Crick or not). Along an mRNA, the ribosome transiently pauses when low abundance tRNAs are required which facilitates co-translational folding (Zhang and Ignatova 2011). The tRNA abundance correlates strongly with codon usage in prokaryotes and unicellular eukaryotes but not in higher eukaryotes (Novoa and Ribas de Pouplana 2012; Plotkin and Kudla 2011; Novoa et al. 2012).

New technological and analytical findings are allowing further understanding of the essential role of tRNAs and their modifications in translation (El Yacoubi et al. 2012; Sarin et al. 2018; Cozen et al. 2015). Concomitantly, more and more evidence have shown that dysfunctional tRNAs associates with the development of specific dis-

eases (Abbott et al. 2014; Torres et al. 2014; Kirchner and Ignatova 2015). In this thesis, the effect of the modulation of specific tRNA-modifying enzymes will be studied in prostate and breast cancer cell lines (**Paper II**).

1.1.2.3 Translation termination When a stop codon enters the A site, the eukaryotic peptide chain release factor (eRF)1:eRF3:GTP complex binds, GTP is hydrolyzed and the polypeptide is released leading to ATP hydrolysis and subunit dissociation (Zhouravleva et al. 1995). Recycling can occur by ribosome splitting, release of tRNA and mRNA for reuse in synthesis of additional proteins (Hellen 2018).

1.1.3 Efficiency of translation of an mRNA

For a specific protein, the synthesis rate will depend on the amount of available corresponding mRNA and on the rate of translation of each of these mRNA molecules. It has been known since the 1960s that protein synthesis occurs in polysomes meaning that mRNAs undergoing translation are typically associated with multiple ribosomes (Warner et al. 1963). Several models exist to describe ribosomal dynamics along an mRNA but the basic principle is illustrated in **Figure 3**.

The factors which can influence rate of initiation, elongation or termination include the number of available ribosomes, aminoacylated tRNAs, translation factors in the cell, the presence of specific structure or binding sites of the mRNA, codon usage along the transcript and amino acid charge of the growing polypeptide chain (Tuller 2014; **Figure 4**). Regulation of global players (ribosomes, tRNAs, translation factors) can impact translational efficiency of all mRNAs in the cell. In contrast, reduction or

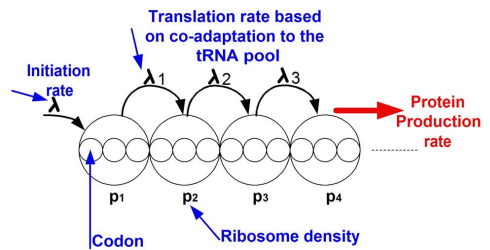


Figure 3: Dynamics of translational efficiency regulation. The efficiency at which an mRNA is translated will depend on the initiation rate λ (rate at which ribosomes start elongating), the transition rate at each codon ($\lambda_1, \lambda_2, \dots, \lambda_n$ for an mRNA with n codons) and the termination rate. In yeast, initiation rates have been estimated to be around 0.01-1.9 ribosomes per second while elongation rates would range from 1 to 20 codons per second (Riba et al. 2019). The average elongation rate in mouse embryonic stem cells was approximated around 5.2 codons per second (Sharma et al. 2019). More recent methods have included in this model the rate and position of ribosome drop-off (Bonnin et al. 2017).

Reprinted with permission from Reuveni, S et al. (2011). "Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model". PLOS Computational Biology 7(9). e1002127. doi:10.1371/journal.pcbi.1002127

overexpression of some translation factors are known to affect specific subsets of mRNAs more than others (Koromilas et al. 1992; Rubio et al. 2014; Wolfe et al. 2014). Moreover, even at steady-state and in the absence of any specific stress or treatment, the range of translational efficiencies from one mRNA to another is widespread (Mathews et al. 2007; Hausser et al. 2019). Finally, translation of mRNAs holding particular RNA or protein-binding sites may be affected according to the level, availability or likelihood of the interaction with corresponding partners (Gebauer et al. 2012; Hentze et al. 2018). As such, inherent characteristics of translational efficiencies and differential sensitivities are encoded in the mRNA's sequence and structures (cis-regulatory elements) while modulation of trans-acting factors (e.g. microRNAs (miRNAs), RNA-binding proteins (RBPs)) will mediate translational changes on the mRNA subsets that they target (**Figure 4**; Hinnebusch et al. 2016; Leppke et al. 2018). The interplay between RNA elements and trans-acting factors will be further described in section 1.2 about *Coordination of gene expression programs*.

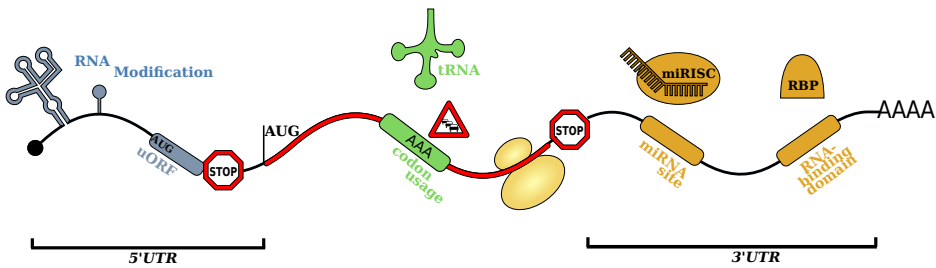


Figure 4: Cis- and trans-regulation of mRNA translation. Primary and higher-order structures along the mRNA influence the efficiency at which it is translated. For instance, complex structures in the 5'UTR, presence of upstream open reading frames (uORFs) or miRNA binding sites are mostly associated with down-regulation of translation efficiency of the main open reading frame; specific RNA modification such as m6A may help ribosome scanning while RBP can act either as enhancers or repressors of translation. Finally, presence of suboptimal codons may reduce the speed of elongation especially in relation to availability of their corresponding tRNAs.

Initiation is the rate limiting step of translation in most conditions (Sonnenberg and Hinnebusch 2009). This implies that small changes in elongation or termination rates are likely to have limited effect on the rate of protein output whereas modulation at the initiation step typically have more impact. Accordingly, regulatory mechanisms will affect assembly of initiation complexes or scanning along the 5'UTR more often than other steps of the translation process.

1.1.4 Experimental methods to measure differential translational efficiency

In cases where translation initiation is rate-limiting (which is the most common scenario), the number of ribosomes associated with mRNAs is a good proxy for efficiency of translation. mRNAs associated with few ribosomes are then considered inefficiently translated while heavy polysomes (mRNAs associated with many ribosomes) would have higher protein output per mRNA molecule and time unit.

1.1.4.1 Polysome-profiling The polysome-profiling method is based on the principle that heavier polysomes will contain more efficiently translated mRNAs. A gradient of sucrose is prepared with density increasing linearly from 5% to 50%. Treatment with an inhibitor of translation elongation such as cycloheximide immobilizes ribosomes on mRNAs and cytoplasmic lysates are loaded on the gradient. This allows, after ultracentrifugation, for differential sedimentation of mRNAs depending on the number of ribosomes they are associated with (Gandin et al. 2014; **Figure 5A**).

Polysome-profiling can be used to assess the overall quantity of translating mRNAs under different conditions. A decrease in global translation, i.e. in translation of most mRNAs, would be identified by a higher 80S peak and lower polysome peaks as exemplified in **Figure 5B** upon inhibition of mTOR translation by torin1 (pink tracing) as compared to insulin-stimulated MCF-7 cells (orange tracing). Quantification of specific transcripts along the gradient fractions allows for assessment of transcript-level regulation of mRNA translation. This potentially permits identification of subsets of transcripts sharing common features leading to their co-regulation at the level of translation. For a specific mRNA, in conditions where it is efficiently translated, its ribosome occupancy mode (most common value) will typically be higher than 3 while in inefficient translation conditions, it would usually be associated with fewer than 3 ribosomes. This "3-ribosome-cutoff" (**Figure 5A**, orange dotted line) for differentiation between efficient and inefficient translation applies to most mRNAs (Larsson et al. 2013; Gandin et al. 2016). Therefore, for transcriptome-wide analysis of differential translation, a cutoff is set at 3 ribosomes and fractions corresponding to mRNAs with more ribosomes than this cutoff are pooled (later referred to as polysome-associated mRNA). This efficiently translated mRNA pool is then extracted and quantified using DNA microarrays or RNAseq. At the single gene level, a change in abundance within the polysome-associated mRNA pool can either be caused by a shift in the number of bound ribosomes for that mRNA (i.e. change in translational efficiency, **Figure 5B**) or a change in the steady state mRNA level (i.e. change in transcription and/or mRNA stability, **Figure 5C**). To distinguish between changes in translation and changes in steady state mRNA levels, cytoplasmic mRNA is collected and quantified in parallel. In the absence of regulation via translation, changes in cytoplasmic mRNA levels should be reflected in corresponding changes in polysome-associated mRNA.

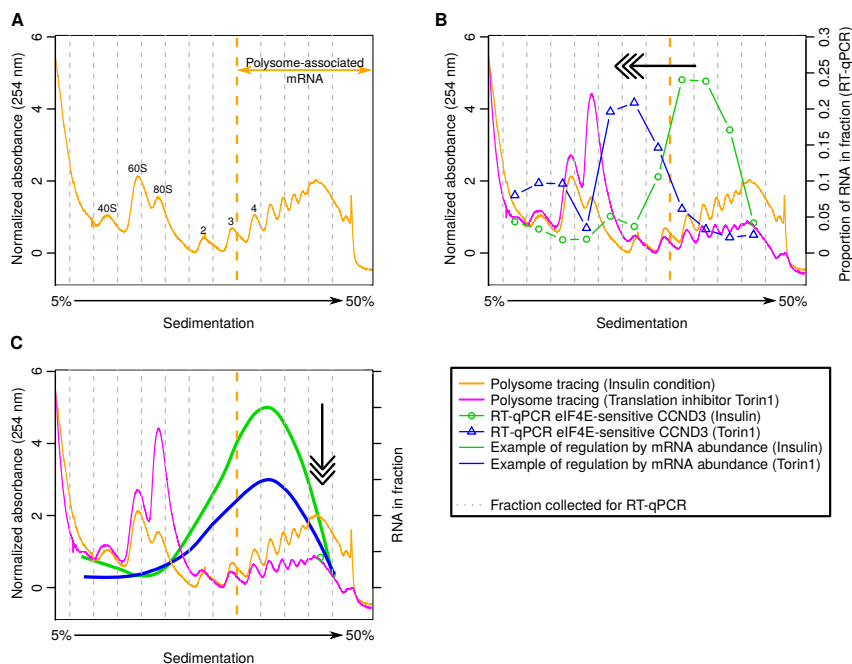


Figure 5: Examples of regulation by mRNA translation or abundance measured by polysome-profiling. After starvation, MCF7 cells were stimulated with insulin or insulin in the presence of torin1, an inhibitor of mammalian/mechanistic target of rapamycin (mTOR) translation (see also section 1.2.2.2.1). Polysome-profiling was performed on insulin and insulin+torin1 treated cells. **(A)** The profile illustrates separation of ribosomal subunits 40S and 60S, the 80S monosome peak as well as peaks corresponding to mRNAs associated with 2, 3, 4, ribosomes etc. For gene level quantification by DNA-microarrays or RNA sequencing (RNAseq), fractions corresponding to mRNAs associated to more than 3 ribosomes are pooled and quantified in parallel of the cytoplasmic mRNA input. **(B)** A known torin1-sensitive mRNAs (cyclin D3 (CCND3)) was quantified in each fraction by quantitative reverse transcription polymerase chain reaction (RT-qPCR) and illustrates a shift in translational efficiency towards lighter fractions upon inhibition of mTOR translation (blue and green lines). **(C)** In contrast, suppression of the cytoplasmic level of an mRNA i.e. by down-regulation of its synthesis or stability, leads to a vertical shift without reduction of the average number of associated ribosomes. However, polysome-associated mRNA is reduced between the green and blue tracings both in **(B)** and **(C)**.

Modified from Gandin, V et al. (2016). "nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs." In: *Genome research* 26(5), pp. 636-648. doi: <https://doi.org/10.1101/gr.197566.115>. ©2016 Gandin et al.; Published by Cold Spring Harbor Laboratory Press. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>

This standard polysome-profiling protocol is often adapted to specific cases: when studying specifically short mRNAs, a cutoff at 2 ribosomes can be deemed more ap-

propriate (Aspden et al. 2014); when assessing usage of alternative isoforms, more fractions and deeper sequencing may be required (Floor and Doudna 2016). **Paper III** of this thesis will present an optimization of the polysome-profiling technique allowing to expand its applicability to samples with low RNA-amounts such as most tissue samples.

In comparison with other methods measuring differential translation (see 1.1.4.2 Ribosome-profiling below), the polysome-profiling technique has the advantage to directly assess, on each mRNA molecule, whether it was associated with low or high amounts of ribosomes. However, it does not provide information about ribosome positioning along the mRNA.

1.1.4.2 Ribosome-profiling Ribosome-profiling was developed more recently than polysome-profiling. It entails blocking translation elongation before degrading RNA not protected by ribosomes using ribonuclease (RNase) (Ingolia et al. 2009). Sequencing libraries are built based on the remaining ribosome-protected fragments (RPFs) providing a quantitative profile of ribosome occupancy after alignment to the transcriptome. For reasons already mentioned in the previous section, cytoplasmic mRNA input (which can be randomly fragmented in order to obtain RNA fragments of similar sizes as RPFs) is collected and quantified in parallel. On one hand, because this method is based on RNase digested samples, information about the number of ribosomes associated with each mRNA is lost. On the other hand, due to its ability to locate ribosomes, it has been providing valuable insights into regulatory mechanisms that control elongation speed, ribosome pausing and translation of alternative open reading frames (Ingolia et al. 2018). Therefore, polysome- and ribosome-profiling methods are complementary.

Ribosome-profiling has however been suffering from experimental artefacts. For instance, it has been shown that cycloheximide, which can be used to inhibit translation elongation, may modify ribosome distribution along the mRNA near specific sequences leading to spurious footprints (Brar and Weissman 2015). Furthermore, the RNase treatment which is used in order to digest RNA between translating ribosomes, also tends to digest ribosomal RNA (rRNA) from ribosomes leading to partial loss of the footprints. However, tools to better assess the quality of ribosome-profiling data, alternative ribonucleases (Gerashchenko and Gladyshev 2017) and *in silico* methods to flag spurious RPFs have been made available (O'Connor et al. 2016; Brar and Weissman 2015; Kiniry et al. 2019).

One limitation which is shared by polysome- and ribosome-profiling arises in the context of specific mRNAs and/or conditions where translation elongation would be

the rate-limiting step, instead of initiation. Indeed, even if initiation speed would typically control protein synthesis rate, assuming that it is the case in any condition and for any gene may be oversimplifying. In occasions where elongation would be rate-limiting, an increased number of ribosomes along an mRNA would indicate a reduced rate of elongation and, provided that translation initiation remained unchanged, a reduced efficiency of translation. As such, when it is unknown which step is rate-limiting, a change in polysome-associated mRNA or RPF, even in the absence of corresponding change in cytoplasmic mRNA, cannot be attributed with certainty to a change in the same direction of the translational efficiency (Mathews et al. 2007). In order to overcome this limitation, other methods to measure translational output, for instance by labeling of newly synthesized polypeptides can be used. Puromycin incorporation-based methods are typically used in this context (Iwasaki and Ingolia 2017) and can be combined with protein-specific antibodies in order to assess translation output of a given protein (Söderberg et al. 2006; Tom Dieck et al. 2015). In order to assess in greater details the translational metrics, such as initiation and elongation rates, one would need to calculate them from the polysome size and ribosome transit time (defined as the time for the ribosome to traverse the mRNA) which can be measured based on radioactivity kinetics from nascent polypeptides to released polypeptides (Fan and Penman 1970; Gehrke et al. 1981). Even if they do not allow the level of precision offered by these methods, polysome- and ribosome-profiling have the advantage to provide a transcriptome-wide perspective of translational control.

1.1.5 Analytical methods for transcriptome-wide analysis of differential translation

Development and improvements of experimental methods highlighted the need for appropriate statistical methods to detect genes showing differential translational efficiencies within transcriptome-wide pools. This section will focus on reviewing analytical methods testing for regulation at the level of translation between 2 or more conditions. Such results can be obtained from both polysome-profiling data (provided that both efficiently translated mRNA and cytoplasmic mRNA were quantified) and ribosome-profiling data (providing quantification of RPFs and cytoplasmic mRNA). The main challenge of such methods is to accurately "correct" polysome-associated mRNA or RPF changes for modulations in steady-state mRNA. As mentioned previously, for one gene, a difference between conditions in its polysome-associated mRNA expression or its RPF expression, can mirror a similar difference of expression observed in its cytoplasmic mRNA expressions which were quantified in parallel; this would primarily be the case of genes under regulation by transcription or mRNA degradation. Translational regulation would then be defined as changes in polysome-associated mRNA or RPF which are independent from fluctuations in cytoplasmic mRNA levels.

It is important to note that even though similar computational methods are typically used downstream of polysome- or ribosome-profiling, data coming from these methods are inherently different. Indeed, while in polysome-profiling data, gene-level differences in translational efficiencies will be measured as horizontal shifts (exemplified in **Figure 5B**), they will be measured as differences in total numbers of ribosomes (coming from any mRNA synthesized from this gene) in ribosome-profiling experiments (see also **Figure 1**). The relationship between the log number of ribosomes associated with an mRNA and the sedimentation distance along a sucrose gradient is very robust (as observed in Gandin et al. (2016) and **Paper III**). However, on a gene-level, differences in RPFs (which "counts" ribosomes) can be very distinct from differences in polysome-associated mRNA (which quantifies copies of efficiently-translated mRNAs). As such, even when studying similar biological mechanisms, studies using polysome- and ribosome-profiling data have proven to sometimes lead to conflicting conclusions (Larsson et al. 2012; Hsieh et al. 2012; Thoreen et al. 2012; Gandin et al. 2016; Masvidal et al. 2017) and may require deviating interpretations of their results.

Most of the methods which will be described below have been designed for ribosome profiling data while **Paper I** involves a method which was developed for polysome-profiling. However, in theory, all these tools could be applied to any data source where the intention is to identify changes in an RNA subset that is independent of a background (e.g. total or cytoplasmic RNA). For this reason, in the next paragraphs, "translated mRNA" will be used as a generic term referring to polysome-associated mRNA or RPFs.

When most mRNA quantification was performed on DNA-microarrays, few methods for analysis of changes in translational efficiencies were available. They have been described and reviewed previously (Larsson et al. 2010; Larsson et al. 2013). Typically, statistical methods which were applicable to DNA-microarrays data cannot be directly used on RNAseq data because of inherent differences in the properties of these two data types (Robinson and Smyth 2008). DNA-microarrays quantify gene expression by measuring intensities or intensity summaries whereas RNAseq quantifies it by measuring numbers of short reads. Empirically, this results in data from DNA-microarrays which, after log-transformation, fulfils normality requirements of most statistical frameworks and data from RNAseq remaining non-normal even after classical transformations and rather following negative binomial-like distributions (the variance of read counts between replicates is generally higher than their mean).

Babel was one of the first method specifically developed for RNAseq based measurements of translation (Olshen et al. 2013) and was designed for data from ribosome-

profiling. It took into account the count nature of the data and used an errors-in-variables regression model between RPF and cytoplasmic mRNA in order to account for the fact that gene expression can only be measured with some level of uncertainty. This is particularly the case of non-replicated experiments which are not uncommon and for which Babel can be used. However, in later benchmark studies, this method showed to perform relatively poorly in terms of control of type I error (Xiao et al. 2016).

edgeR (Robinson et al. 2010) and DESeq2 (Love et al. 2014) are two methods which were initially developed for identification of differential gene expression from RNAseq experiments where only cytoplasmic mRNA was considered. These two methods differ in the specific methods for dispersion estimation, information sharing across genes and normalization but they share similar principles. They are both parametric methods assuming a negative binomial distribution for the read counts and use generalized linear models (GLMs) for differential expression testing. The use of GLMs makes these methods very flexible and allow for analyzes of complex study designs. As such, they can easily be adapted towards analysis of differential translation. Reads counts K_{gj} are modelled as negative binomial distributions with mean μ_{gj} and dispersion σ_g for gene g and sample j . The mean is considered as the product of q_{gj} (parameter proportional to the expected true concentration of fragments from gene g for sample j) by a size factor accounting for differences in sequencing depth between samples. For classical differential expression analysis between, for instance, a treated and a control condition the GLM would be as follows:

$$\log(q_{gj}) = \beta_g^{Trt} . X_j^{Trt} + \epsilon_{gj}$$

with coefficients β_g^{Trt} representing the \log_2 fold changes for gene g for each column of the model matrix X and ϵ_{gj} the error term. For analysis of differential translation, two additional parameters can be incorporated into the model: a parameter for the RNA type (i.e. cytoplasmic or translated mRNA), and an interaction parameter between treatment and RNA type.

$$\log(q_{gj}) = \beta_g^{Trt} . X_j^{Trt} + \beta_g^{RNAtype} . X_j^{RNAtype} + \beta_g^{RNAtype:Trt} . X_j^{interaction} + \epsilon_{gj}$$

The interaction term can be interpreted as the differential effect of treatment in translated mRNA compared to cytoplasmic mRNA i.e. the translational control effect. This "GLM with interaction term" is the principle used by several recent methods including Riborex (Li et al. 2017b), Ribodiff (Zhong et al. 2017) and deltaTE (Chothani et al. 2019). Riborex is directly available either as "DESeq2 with interaction term" or "edgeR with interaction term" whereas Ribodiff uses slightly different estimation

methods and allows for different dispersion estimations from cytoplasmic and translated mRNA data (Zhong et al. 2017). Although estimating different dispersions for data coming from different experimental procedures seems relevant, Ribodiff seemed to underperform in simulation studies compared to Riborex (Li et al. 2017b). In the implementation of Riborex as an R package, its use is restricted to simple designs whereas the deltaTE implementation of the same methods maintains their availability to more advanced applications including the possibility to correct for batch effects in the model (Chothani et al. 2019). Xtail (Xiao et al. 2016) uses DESeq2 for estimation of mean and dispersion while allowing for analysis of experiments with only one replicate per condition (which only Babel allowed so far). Xtail outperforms other methods in terms of specificity and sensitivity when the sample size is very low (1 or 2 replicates) but further analyzes included in **Paper I** show that this is at the cost of detection of high amounts of non-differentially translated genes when tested under a NULL model (no true differences in gene expression) especially when the sample size is low. Xtail was also shown to be strongly impacted when a batch effect was added to the data even though systematic differences between replicates are commonly observed in transcriptome data (Chothani et al. 2019).

The importance of dysregulated translation for cancer progression has been increasingly recognized over the last 30 years (Lazaris-Karatzas et al. 1990; De Benedetti and Graff 2004) such that the translation initiation apparatus is presently a target in oncology (Chu and Pelletier 2018). It has also been associated with other diseases such as neurodegenerative diseases (Moreno et al. 2012) or metabolic disorders (Shi et al. 2003). Consequently, efforts have been put to understand mechanisms dysregulating mRNA translation and specifically to improve experimental methods for quantification of the transcriptome as well as computational methods to detect differential translational control. Two of the constituent papers of this thesis (**Paper I** and **Paper III**) are extensions of classical methods: **Paper III** presents an optimization of polysome-profiling allowing it to become scalable to large sets of small samples (low RNA input as in biobanked tissue samples) and **Paper I** provides an improvement of the computational method anota (Larsson et al. 2010) which was only available for quantification by DNA-microarrays data. Anota2seq (**Paper I**) extends its use to RNAseq-quantified polysome- or ribosome-profiling data. Furthermore, we demonstrated that anota2seq shows appropriate control of type I error, can include batch-effect correction and allows identification of genes whose mRNA levels are buffered at the level of translation (this gene expression mode of regulation will be presented in the next sections).

1.2 Coordination of gene expression programs

In general, the objective of omics studies (e.g. transcriptomics which measure the full set of transcribed RNA molecules, proteomics which measure the full set of proteins, metabolomics measuring any metabolites, translomics measuring translated mRNAs, etc.) is to understand which cellular pathway or mechanism is influencing a specific phenotype. For instance, in an attempt to study which regulatory events or mechanisms lead to the development of resistance to BRAF inhibitors in melanoma (a classical treatment in this disease which is known to be efficient for a certain time until the patient stop responding and the tumor relapses), one would extract biomolecules of interest (RNA, proteins, metabolites, translated mRNAs, etc.) from a cell line which is sensitive to the treatment and one which has developed resistance. From inferred differences in abundance of specific species of these biomolecules, one would generate hypotheses regarding essential cellular events driving the resistance.

Because proteins are arguably the most important players in many biological processes, a majority of such research projects would favor proteomics over other omics approaches. Transcriptomics may be considered a less expensive approach which benefits from well-established technologies and which, in some specific contexts, would be deemed to provide an appropriate proxy for protein levels. The question of whether RNA and protein levels strongly correlate, of whether RNA levels is a good surrogate for protein levels is controversial. In other words; among transcription, mRNA translation or decay, which step has the most important contribution to shaping the proteome has been extensively debated (Schwanhäusser et al. 2011; Vogel and Marcotte 2012; Jovanovic et al. 2015; Li and Biggin 2015; Li et al. 2017a; Liu et al. 2016). A consensus has nonetheless been reached on one aspect of this discussion: the dynamic contribution of each process between conditions is context- and biological system-dependent with mRNA translation having a high contribution upon severe stress (Liu and Aebersold 2016) such as endoplasmic reticulum stress (Baird et al. 2014; Guan et al. 2017; Cheng et al. 2016). Thus, changes in mRNA levels cannot be generally considered as a good proxy for changes in protein levels.

Notwithstanding major improvements in coverage and precision of recently developed proteomics technologies (Branca et al. 2014; Orre et al. 2019), observing changes in protein levels upon a specific perturbation only informs on the impact on the output product of the gene expression pathway. In systems where it could be assumed that the intervention directly affects this output (i.e. the protein abundance), this would provide sufficient evidence for a broad understanding of the mechanisms affecting the phenotype. However, cellular pathways, including those active in cancer, are often very complex containing feedback loops and alternative branches. This is a

reason for the increased interest in multiple omics studies (i.e. where several levels of gene expression are measured on the same samples). In the next section, control at the translational level will be considered in the general context of gene expression regulation.

1.2.1 Regulation of gene expression at multiple levels

In eukaryotes, because mRNA translation occurs in the cytoplasm and transcription in the nucleus, these processes are often thought to be controlled independently. However, this view is probably over-simplistic and mechanistic examples of crosstalk between separate gene expression processes are common. Transcription and splicing are predominantly coupled (Beyer and Osheim 1988) and it has been hypothesized that co-transcriptional splicing could impact the integrity of transcription (Komili and Silver 2008). The development of methods to efficiently sequence long nascent RNA molecules will soon unravel dynamics and order of intron removal (Drexler et al. 2019). In the cytoplasm, advances in other sequencing-based technologies allowed 5' to 3' mRNA co-translational decay to be characterized by observing 3-nucleotide periodicity in mRNA degradation intermediates. These were interpreted as products of exonucleases following the last translating ribosome (Pelechano et al. 2015). Co-translational decay is general and conserved which indicates that interplay between mRNA translation and decay is more complex than the classical view stating that translating mRNAs are protected from decay (Roy and Jacobson 2013).

Concerning crosstalks between different cellular localizations, Haimovich et al. (2013) have introduced the concept of synthegradosome where mRNA synthesis is linked to degradation by decay factors that were shown to shuttle from the cytoplasm to the nucleus to associate with transcription start sites and regulate transcription initiation. Furthermore, few examples exist of coordination between transcription and translation. For instance, Rpb4p and Rpb7p which are subunits of the yeast RNA polymerase II (Pol II) and also mediate mRNA decay (Choder 2004; Lotan et al. 2005; Lotan et al. 2007) have later been associated with regulation of translation initiation by interacting with eIF3 (Harel-Sharvit et al. 2010).

Interestingly, when conducting ribosome- and polysome-profiling experiments, cytoplasmic mRNA samples always need to be quantified in parallel of translated mRNA. Thus, data from these measurements provide more information than only translational regulation. Differences in mRNA abundance and potentially coordination between different layers of gene expression regulation can be also be analyzed. This could unravel whether the mRNA synthegradosome (transcription and mRNA decay) act in concert with mRNA translation or if translation seems to compensate or counteract modula-

tions in mRNA abundance. Examples associated to both cases exist in the literature and will be explained below.

1.2.1.1 When mRNA synthegradosome and translation act in concert Gingold et al. (2014) gave an interesting example where regulation at the level of transcription/mRNA decay and mRNA translation may be coordinated to both enhance or both repress cellular functions. Namely, in proliferating or cancer cells vs. differentiated or normal cells, opposing tRNA signatures are expressed (**Figure 6**). Notably, tRNAs translating different codons for the same amino acid showed opposite trends. Furthermore, when analyzing the codon usage of genes expressed in proliferation and differentiation, this study demonstrated that proliferation genes are enriched in codons (**Figure 6**, red colored codons) matching the "proliferation-tRNA-signature" (**Figure 6**, red colored anticodons) and inversely differentiation mRNAs require the "differentiation-signature" tRNAs for their translation (**Figure 6**, blue). They conclude that if in translation, the codon usage of an mRNA is the demand and the available tRNAs correspond to the supply; proliferation and differentiation are cellular states in which supply and demand are well matched.

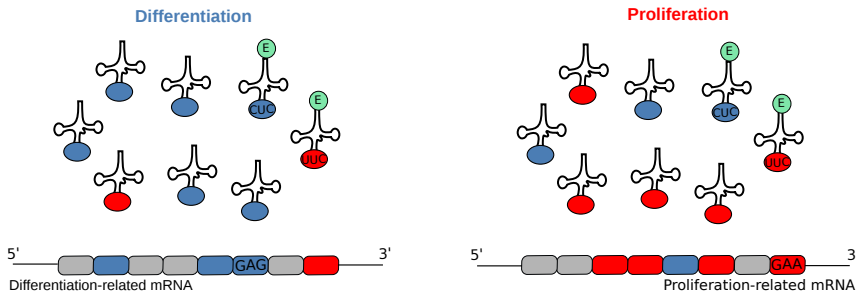


Figure 6: Coordination between tRNA supply and demand in proliferation and differentiation. Genes belonging to proliferation-, respectively differentiation-, related ontology genesets are enriched for a specific set of codons (red, respectively blue). Red codons preferably have, at their third nucleotide position, A or U while blue codons preferably have C or G at this position. Interestingly, method development in tRNA quantification revealed differences in tRNA pools between cell models of proliferation vs. differentiation as well as between cancer and normal tissue samples. Gingold et al. (2014) observed in each cellular state, a match between tRNA supply and codon usage.

Inspired by Gingold, H et al. (2014). "A dual program for translation regulation in cellular proliferation and differentiation." In: Cell 158(6), pp. 1281-1292.

This interpretation has however been challenged. Rudolph et al. (2016) has argued that although codon-driven differences in translational efficiency between conditions have been extensively observed in prokaryotes and single-cell eukaryotes (Man

and Pilpel 2007; Drummond and Wilke 2008), the level of evidence for similar mechanisms in mammals remains low. When studying several steps of tissue development in mice, they did not detect overall differences in tRNA supply nor demand (Schmitt et al. 2014). Deciphering such contradicting results is complex: established methods to measure modified and charged tRNAs are lacking as well as trustworthy quantification of global differences in tRNA levels; assessing whether a specific tRNA pool can equally well translate different transcriptome-weighted codon usages is difficult; translational efficiency correlates to some extent with codon usage and with transcriptomic GC content but the causality between these events is less trivial to address. However, the conclusion which has a high level of confidence is that mammalian systems do not depend on codon bias as a translational regulatory mechanism as much as prokaryotes (Rudolph et al. 2016).

1.2.1.2 When mRNA translation compensates modulations at the mRNA level

Situations where a gene is strongly up-regulated at the mRNA level while the corresponding protein is strongly down-regulated are uncommon and sometimes attributed to data normalization issues (Albert et al. 2014). However translational buffering where protein levels remain stable between conditions despite alterations in the RNA levels has been described in several organisms (Lalanne et al. 2018; McManus et al. 2014; Artieri and Fraser 2014; Genik et al. 2015). For instance, Lalanne et al. (2018) compare mRNA levels and protein synthesis rates between divergent bacteria species which have evolved independently for several billion years. They take the example of 4 genes (*rpsP*, *rplS* encoding two ribosomal proteins, *rimM* an rRNA-maturation factor and *trmD* a tRNA modification enzyme) which are expressed as a polycistronic mRNA in *E. coli* whereas in *B. subtilis*, they have a wide range of expression levels. Three of these genes show big differences in expression between the two bacterial species (from 5 to more than 30 fold). However, the function of these homologous proteins are similar and the relative levels of RPFs between the species are similar (less than 2-fold differences for all 4 mRNAs). Thus, transcriptional differences between divergent bacterial species are translationally compensated. Mechanistically, this is achieved by translationally suppressive mRNA secondary structures which are present in *E. coli* and absent in *B. subtilis*. Consequently, while regulation of these genes occurs at the transcriptional level in *B. subtilis*, it occurs by mRNA translation in *E. coli*.

Other mechanisms mediating translational buffering have been characterized in transcriptome-wide studies across yeast species (McManus et al. 2014; Artieri and Fraser 2014) and between human patients (Genik et al. 2015).

Notably, several definitions of "translational buffering" co-exist in the literature. First of all, in some publications, steady-state co-expression of proteins encoded by

non co-expressed mRNAs (or vice versa) is also termed translational buffering or translational compensation (Kustatscher et al. 2017; Dassi et al. 2015) but this is not the definition used herein. Indeed, this thesis will focus on changes in gene expression observed under dynamic settings (i.e. between conditions) and therefore needs to be distinguished from when mRNA and protein levels are compared under a single condition. The different modes of regulation of gene expression which can be analyzed from transcriptome-wide analyzes of differential translation (i.e. from polysome- or ribosome-profiling data for instance) are described in **Figure 7** between 2 hypothetical conditions (treatment T and control C). As explained earlier, a change in translated mRNA between conditions can be due to a change in mRNA abundance (**Figure 7A** top left; **Figure 7B** bottom) or a change in translational efficiency (i.e. a difference in ribosome occupancy resulting in a shift along the polysome gradient; **Figure 7A** top right; **Figure 7B** top; see also **Figure 5**). Finally, translational buffering is defined as a change in cytoplasmic mRNA levels which is not reflected in a change in translated mRNA (**Figure 7A** bottom left; **Figure 7B** middle).

It remains to be fully characterized whether specific mechanisms or specific biological processes are associated with situations where mRNA level modulations are buffered or not at the level of translation. An example of such a mechanism will be discussed in **Paper II**. However, this is largely unexplored area of research and to further the knowledge of mechanisms coordinating gene expression, availability of analytical methods distinguishing between changes in translational efficiencies leading to altered proteins levels or buffering is essential. This is an important feature of the anota2seq methodology which will be presented in **Paper I** of this thesis.

1.2.1.3 Role of estrogen receptor alpha in breast and prostate cancer Most studies analyzing transcriptome-wide changes in translational efficiencies are focused on the role of translation factors or on mechanisms known to impact mRNA translation such as stress responses. In contrast, in **Paper II**, we studied the role of a transcription factor, namely estrogen receptor alpha ($ER\alpha$), in coordinating transcriptional and translational output. Indeed, in addition to its role as a transcription factor, $ER\alpha$ has been shown to potentially have non-nuclear functions and to affect the mTOR and mitogen-activated protein kinase (MAPK) pathways which impinge on mRNA translation (see section 1.2.2.2). Further details about $ER\alpha$ -dependent gene expression regulation will be provided in **Paper II** while this section will expose some background information regarding $ER\alpha$'s position as therapy target in breast cancer and co-driver of tumorigenesis in prostate cancer.

$ER\alpha$ is mainly studied in breast cancer which is the most common cancer type in women accounting for almost one in 4 cancers. In March 2019, the reported world-

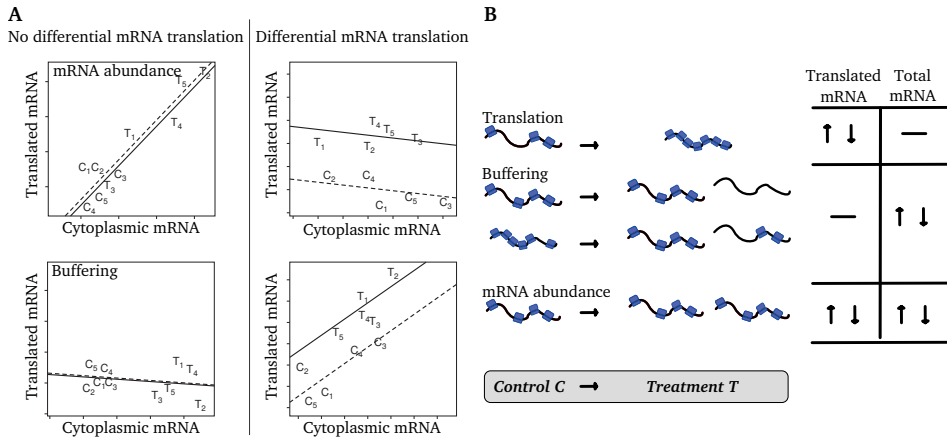


Figure 7: Modes of regulation of gene expression. (A) Cytoplasmic and translated mRNA quantification are provided for 4 hypothetical genes and 5 biological replicates (Treatment condition T and Control condition C). Four scenarios are illustrated: changes in mRNA abundance (top left); changes in translational efficiencies leading to altered protein levels (top right) or buffering (bottom left); bigger changes in translational efficiencies compared to changes in mRNA abundance (bottom right). (B) Schematic representation of the different modes of regulation of gene expression.

(A) Modified from Larsson, O et al. (2013). "Toward a genome-wide landscape of translational control." In: *Cold Spring Harbor Perspectives in Biology* 5(1), a012302. doi: <https://doi.org/10.1101/cshperspect.a012302>. ©2013 Cold Spring Harbor Laboratory Press. (B) Modified from Oertlin C et al. "Generally applicable transcriptome-wide analysis of translation using anota2seq." *Nucleic Acids Res.* 2019 Jul 9;47(12):e70. doi: <https://doi.org/10.1093/nar/gkz223>

wide age standardized incidence was 46.3 per 100 000 with a mortality rate of 13/100 000 (*Cancer fact sheets - Breast*). In this disease, ER α is both a prognostic marker, i.e. a characteristic of the disease which is associated with patients' survival independently of treatment, and a predictive factor for endocrine therapies, i.e. it characterizes how responsive to this treatment tumors are. Namely, patients whose breast tumor expresses the receptor (estrogen receptor (ER)+), have a better prognosis than ER- patients, at least in the first 5 years after their cancer diagnosis (Knight et al. 1977). Furthermore, ER+ tumors have been shown to benefit from endocrine therapy (Tamoxifen) whereas little to no such effect was seen in ER- patients (EBCTCG 1998).

Hormonal treatment is also the most common strategy against prostate cancer but in this case targeting the androgen signaling or receptor. Prostate cancer is the second most common cancer in men worldwide. Its age standardized incidence and mortality rates are 29.3 and 7.6 per 100 000 respectively (*Cancer fact sheets - Prostate*). Prostate cancer is thought to be an androgen-dependent disease making androgen-deprivation therapy its most classical treatment strategy (Parker et al. 2015). However,

evidence that androgens alone could induce sustained proliferation and carcinogenesis in prostate cells is remarkably lacking (Morgentaler and Traish 2009). In combination with other carcinogens or upon chronic treatment of high doses of testosterone (androgens), prostate cancer could be induced (Pour and Stepan 1987; Bosland 2013). Strikingly, estrogens, in combination with testosterone result in prostate cancer in 90 to 100% of cases in animal models (Bosland et al. 1995); they can also stimulate cancer progression and metastasis (Ricke et al. 2006). Estradiol (estrogens) are synthesized from testosterone by the aromatase enzyme which has been shown to be expressed in the prostate (Ellem et al. 2004). A landmark study showed that aromatase knock-out mice which were treated with androgens and estrogens developed pre-malignant prostate disease whereas androgen alone did not result in such development. Consistent with a leading role for ER signaling, the combination of androgens and estrogens did not lead to the disease in ER α knockout mice (Ricke et al. 2008). Understanding of the role of estrogens in prostate cancer is further complicated by the fact that estrogen receptor beta (ER β) may, in contrast to ER α , have beneficial effects on cancer development, although this is under debate (Nelson et al. 2014). In the study mentioned above (Ricke et al. 2008), knockdown of ER β led to development of pre-malignant prostate cancer upon treatment with both hormones.

In humans, even though aromatase inhibitors have proven their efficacy in postmenopausal breast cancer (i.e. in women whose ovaries no longer produce significant amounts of estrogens but in whom estrogens are still produced from androgens; Dowsett et al. 2010, Cardoso et al. 2019), clinical trials of similar drugs concluded that they should not be advised for prostate cancer treatment (Smith et al. 2002; Santen et al. 2001).

Although androgen-deprivation therapies are efficient against most prostate tumors, resistance to this treatment is not uncommon and more than 15% of men initially diagnosed with prostate cancer die from advanced incurable disease. A recent study showed that ER α is specifically expressed in the epithelium of advanced prostate tumors (Takizawa et al. 2015). The role of this receptor is still poorly understood in this disease but advances in this field could lead to promising therapeutic strategies. As mentioned previously, ER α is an extensively studied transcription factor in breast cancer; yet, its transcriptional targets can be tissue-specific which extends the need for transcriptome-wide studies of ER α -dependent gene expression in prostate tissue. Furthermore, Simoncini et al. (2000) showed that ER α can associate with phosphatidylinositol-3-OH kinase (PI3K) which is upstream of important regulators of mRNA translation. Thus, both transcriptional and translational outputs of ER α will be investigated in **Paper II**.

1.2.2 Coordinated regulation of sets of mRNAs

A well-accepted perspective is to consider that mRNAs are not linear nucleic acids but rather, in addition to harboring secondary structure and carrying multiple modifications, exist most often in complexes with proteins and other RNAs, so called messenger ribo-nucleoprotein particles (mRNPs) (Gebauer et al. 2012). If each mRNA should be viewed in regard to the activity of each of its interacting partners, each trans-acting factors, e.g. RBP or miRNA, typically binds to multiple mRNA targets thus coordinating entire RNA operons (Keene 2007). These many-to-many relationships define complex interactomes. Furthermore, multiple post-transcriptional regulatory features typically co-occur in mRNAs with the activity of one RNA feature also potentially depending on the proximity of other features (Cottrell et al. 2018). Recent advances in technology have facilitated better characterization of RNA features and trans-regulatory elements required for selective post-transcriptional regulation (Truitt and Ruggero 2016). Among others, experimental challenges have been resolved to better assess which mRNAs are selectively translated (Brar and Weissman 2015), binding and activity of RBPs (Van Nostrand et al. 2016), regulatory patterns of miRNAs (Katchy and Williams 2016) and tRNAs (Sarin et al. 2018), usage of alternative transcription start sites (Poulain et al. 2017; Gandin et al. 2016; Masvidal et al. 2017) but additional studies of the translome under multiple conditions will be essential before a complete understanding of how post-transcriptional regulation is manifested can be reached.

As mentioned in section 1.1.5, computational methods for transcriptome-wide analysis of differential translation will distinguish genes regulated at the translational level from changes in mRNA abundance. Previously annotated sets of mRNAs co-regulated at the translational level constitutes essential tools helping to extract relevant biological findings from lists of genes regulated at different steps of the gene expression pathway in new studies. Let us consider the following example: eIF4A is a DEAD-box RNA helicase which helps unwinding secondary structures along 5'UTRs (Parsyan et al. 2011). Consequently, upon treatment with an eIF4A-inhibitor, the set of mRNAs harboring long and highly structured 5'UTRs will show reduced translational efficiency (Wolfe et al. 2014; Rubio et al. 2014; Gandin et al. 2016). It has also been specified recently that the position of the structures along the UTR impacts the sensitivity to eIF4A (Waldron et al. 2019). Therefore, if, in a new study, it can be proven that mRNAs having complex structures close to the start codon are specifically downregulated at the translational level, a reasonable hypothesis could be that eIF4A suppression may drive the studied phenotype. Similarly, as a second example, if in another study, enrichment for a specific RNA-element is observed among translationally modulated genes, one might deduce that a factor binding to this element may be

regulated between the studied conditions. Several examples of known mechanisms by which subsets of mRNAs may be translationally co-regulated will be described in the next paragraphs.

1.2.2.1 Regulatory RNA elements and trans-acting factors

1.2.2.1.1 microRNA and RNA-binding proteins Regulatory mechanisms of translation commonly involve interactions between specific structure or sequences along the UTRs and RNA or proteins which bind to it (**Figure 4**). miRNAs for instance typically bind to 3'UTR of specific subsets of mRNAs thereby inhibiting their translation or inducing their degradation (Jonas and Izaurralde 2015). RBPs commonly affect mRNA stability, translation as well as processing or modification (Hentze et al. 2018). About 1400 RBPs have been identified in humans; some of them, like PABPs target widespread RNA sequences while others act on small groups of mRNAs. For instance, human antigen R (HuR) is a well-characterized RBP which binds U- or AU-rich sequences on 3'UTRs (Grammatikakis et al. 2017). This RBP has attracted special interest because many of its targets are involved in cellular functions such as proliferation, apoptosis and differentiation (Abdelmohsen and Gorospe 2010). HuR function is mainly regulated by posttranslational modifications. For instance, phosphorylation of HuR by cyclin-dependent-like kinase 5 (CDK5) was shown to specifically prevent binding to cyclin-A2 (CCNA2) mRNA, one of its target leading to its reduced translation and to cell cycle arrest (Filippova et al. 2012).

1.2.2.1.2 5'UTR elements 5'UTR length, presence of secondary structure or of uORFs play critical roles in translational control (Gebauer et al. 2012; Hinnebusch et al. 2016; Gandin et al. 2016). Studies by Kozak in the 80s suggested that an optimal 5'UTR length for efficiency scanning and translation initiation could be found around 80 nucleotides implying that both shorter and longer UTRs would be less efficiently translated (Kozak 1987). Subsets of long and structured 5'UTR mRNAs have been shown to be particularly sensitive to eIF4E availability i.e. reduction of eIF4E primarily affect these so-called "weak mRNAs" (Graff et al. 2008). They represent a classical example of oncogenic translational control. Indeed, these weak mRNAs are enriched in cellular functions such as survival, proliferation, angiogenesis (e.g. MYC proto-oncogene (MYC), Survivin, Cyclin D1, vascular endothelial growth factor (VEGF)). Most cancers overexpress eIF4E thereby enhancing translation of these oncogenic signatures (De Benedetti and Graff 2004).

Another example where translation is selectively controlled by differential sensitivity to a translation factor was described above (eIF4A-sensitive subset); in this case, the regulation was also mediated by structural features in the 5'UTR.

1.2.2.1.3 Codon-dependent elongation rates and the role of tRNA availability In addition to modulation of translation impinging on scanning and initiation, dysregulated translation depending on codon composition of mRNAs was recently described during tumorigenesis and therapy resistance (Rapino et al. 2018). In this context, tRNAs can be thought of as trans-acting factors whose regulation may impact translational elongation speed. In case of a global reduction of the levels of all tRNA species, the potential impact on protein synthesis rate is predicted to also be general. tRNA synthesis is for instance known to be stimulated by mTOR via suppression of the Pol III-inhibitor MAF1 (Michels et al. 2010). However, as already discussed, it has been debated whether regulation of specific tRNAs may have an impact on translation in a codon-dependent manner.

In the example of the study by Gingold et al. (2014) described above (**Figure 6**), differential expression of tRNAs corresponded to differences in codon usage between proliferation and differentiation conditions. Such codon-specific regulation may however occur at any level of tRNA biogenesis. Alterations of tRNA function by post-transcriptional modifications has recently raised interest (El Yacoubi et al. 2012). The 5-methoxycarbonyl-methyl-2-thiouridine (mcm^5s^2U) modification will, for instance, be of particular interest in **Paper II**. mcm^5s^2U strengthen the interaction between $tRNA^{UUC}_{Glu}$, $tRNA^{UUU}_{Lys}$, $tRNA^{UUG}_{Gln}$ and the codons they can pair with, namely GAA and GAG (coding for Glutamic acid), AAA and AAG (coding for Lysine), CAA and CAG (coding for Glutamine). This modification is necessary for wobbling (decoding GAG, AAG and CAG) but the cognate tRNAs of these G-ending codons ($tRNA^{CUC}_{Glu}$, $tRNA^{CUU}_{Lys}$, $tRNA^{CUG}_{Gln}$) are also expressed in humans. Ribosome-profiling studies performed in yeast upon loss of the mcm^5s^2U modification showed increased ribosome occupancy at AAA, CAA and, to a lesser extent, GAA codons which can be interpreted as a reduction of elongation rates at these codons (Zinshteyn and Gilbert 2013; Nedialkova and Leidel 2015). Interestingly, the increases in occupancy at AAG, CAG and GAG codons were unconvincingly small compared to those of A-ending codons. Therefore, mcm^5s^2U modification impact translation of A-ending codons to a higher extent than G-ending codons. The estimated differences in ribosome dwell time at A-ending codons upon loss of the modification did not lead to a significant reduction of global protein synthesis (indicating that translation initiation remains rate-limiting in these conditions; Zinshteyn and Gilbert 2013). When focusing on the set of proteins encoded by mRNAs which are enriched for AAA, CAA and GAA codons, Rezgui et al. (2013) observed a significant decrease in protein abundance upon loss of the modification. However, in the study by Nedialkova and Leidel (2015), these differences were also observed on cytoplasmic levels of AAA, CAA and GAA enriched mRNAs thus disproving the hypothesis that such regulation would be due to reduced translation elongation rate at specific codons. Instead, they showed that increased ribosomal

pausing perturbs co-translational folding and solubility of proteins.

The mcm^5s^2U modification and the tRNA species it impacts is well defined in yeast (Johansson et al. 2008) but remains to be fully characterized in mammalian systems. In melanoma, it has recently been shown that modulating expression and/or activity of the enzymes catalyzing this modification (so called Uridine 34 (U34) modifying enzymes; **Figure 8**) promotes survival and resistance to therapy (Rapino et al. 2018). Similarly, these enzymes has been described to have a key role in breast cancer metastasis (Delaunay et al. 2016) and Wnt-driven intestinal cancers (Ladang et al. 2015). These phenotypes were proposed to rely on differential requirement for tRNAs modified at the U34 position across mRNA-species. Thus, a higher level and/or activity of U34 modification enzymes would allow more efficient translation of mRNAs with high frequency of codons requiring U34 modified tRNAs.

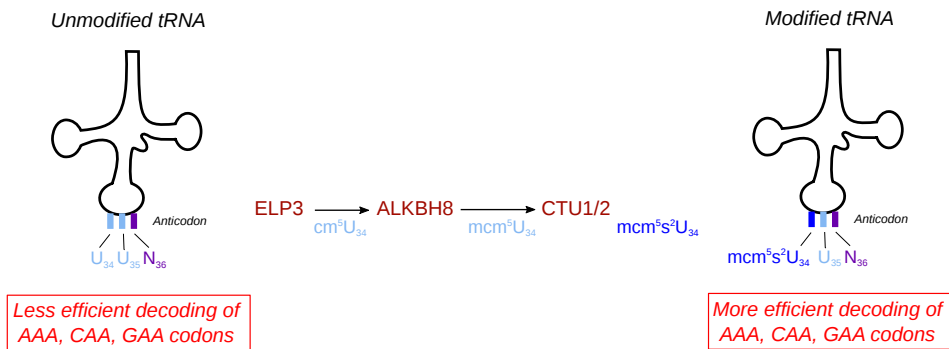


Figure 8: U34 tRNA modification pathway. The mcm^5s^2U modification is catalyzed by a series of enzymes: elongator acetyltransferase complex subunit 3 (ELP3), alkB homolog 8, tRNA methyltransferase (ALKBH8) and cytosolic thiouridylase subunit (CTU).

1.2.2.2 Regulatory pathways impinging on translational outputs of subsets of mRNAs Translational co-regulation of subsets of mRNAs also typically occurs downstream of specific regulatory pathways. A few examples will be explained below. The features allowing co-regulation can be found, for instance, along their 5'UTRs (presence of terminal oligopyrimidine (TOP) motifs, complex secondary structures (see section 1.2.2.2.1)).

1.2.2.2.1 mTOR-sensitive mRNAs Activation of mTOR, more specifically of its complex 1 (mTOR complex (mTORC)1), leads to phosphorylation of eIF4E binding proteins (4E-BPs) which are inhibitory proteins for translation initiation. Phosphorylated 4E-BPs cannot bind eIF4E freeing the cap-binding effector for assembly of the

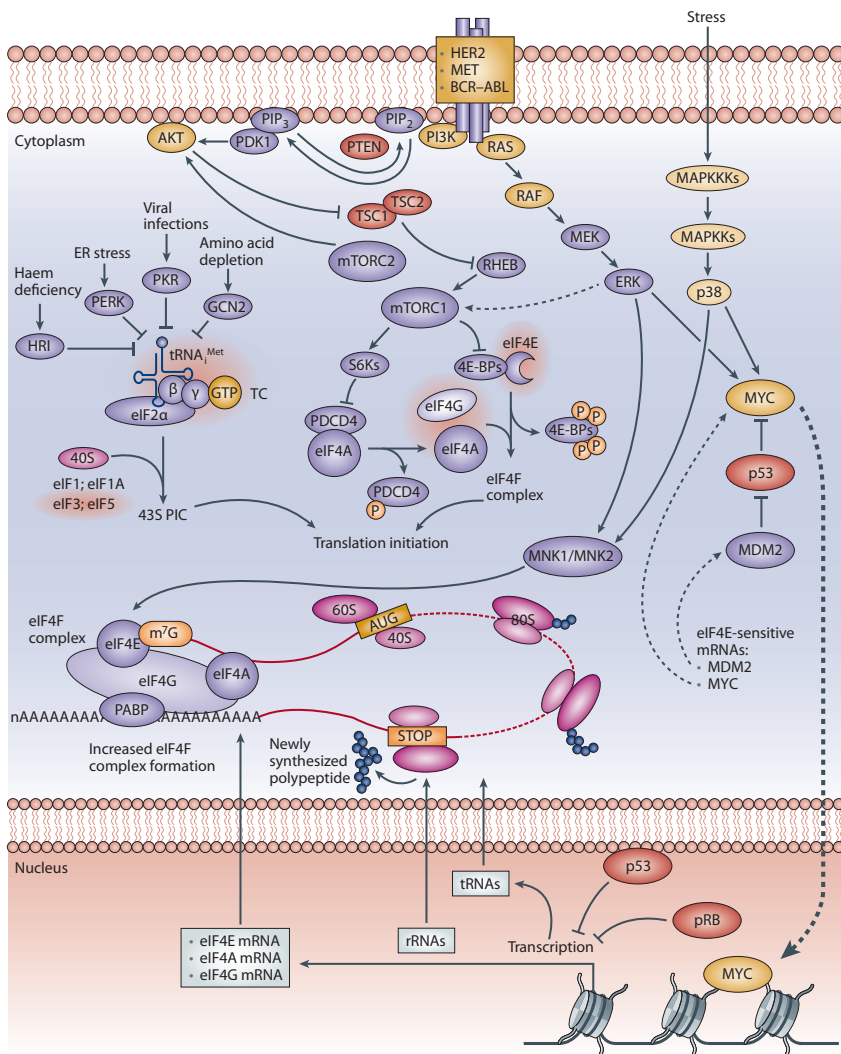


Figure 9: Regulation of selective translation by oncogenic pathways. Hyperactive PI3K-mTOR and MAPK pathways impinge on eIF4F complex assembly.

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature Reviews Drug Discovery. Bhat M et al. Targeting the translation machinery in cancer. *Nat Rev Drug Discov.* 2015 Apr;14(4):261-78. doi: 10.1038/nrd4505. ©2015

eIF4F complex and initiation of translation (Figure 9; Brunn et al. 1997). Specific sets of mRNAs are known to be more sensitive than others to alterations in the mTOR pathway. For instance, mRNAs harboring a TOP motif (cysteine after the 5' cap followed by 4-15 pyrimidines), corresponding mainly to mRNAs involved in the translational

machinery itself such as ribosomal proteins, show coordinated drastic reduction in their translational efficiency upon mTOR inhibition (Hsieh et al. 2012; Thoreen et al. 2012; Larsson et al. 2012). Another subset of mRNAs, those having long and structured 5'UTRs, shows specific sensitivity to mTOR inhibition (Hinnebusch et al. 2016; Larsson et al. 2012; Gandin et al. 2016). This effect depends on signaling via 4E-BPs.

mTOR is a regulator of responses to nutrient and growth factor stimuli. It can be part of several protein complexes including mTORC1 which regulates cell growth and proliferation (by phosphorylating ribosomal S6 kinase (S6K) and 4E-BPs) and mTORC2 which regulates survival mainly via the AKT (also known as Protein kinase B) pathway (Dowling et al. 2010; Sarbassov et al. 2005). mTOR is commonly hyperactive in human cancers (Vivanco and Sawyers 2002; Taylor et al. 2010) and attempts to selectively and efficiently target its pathway as an oncological therapy are still ongoing (Rodrik-Outmezguine et al. 2016; Fan et al. 2017).

ER α has been shown to associate with PI3K, an upstream activator of mTOR and directly or indirectly influence this oncogenic pathway in several tissues (Simoncini et al. 2000; Levin 2009). A recent study linked resistance to estrogen receptor antagonists with hyperactivation of mTOR as well as increased phosphorylation of eIF4E by MAPK-interacting kinase (MNK)1 (Geter et al. 2017).

1.2.2.2.2 Regulation of selective translation by MAPK signaling and eIF4E phosphorylation eIF4E activity and initiation of cap-dependent mRNA translation can also be modulated by phosphorylation of eIF4E at Serine 209 (Furic et al. 2010). This is mediated via MAPK or MAPK/ERK kinase (also known as MAPKK) (MEK)-extracellular signal-regulated kinase (ERK) signaling (**Figure 9**). The role of eIF4E phosphorylation is still poorly understood but has been associated with preferential translation of selected mRNAs (Robichaud et al. 2015; Furic et al. 2010). In mouse embryonic fibroblasts (MEFs), translational activation through increased phosphorylation of eIF4E is necessary for epithelial-to-mesenchymal transition (EMT) induction. Therefore, targets of the MAPK pathway constitutes another example of coordinated translational control.

2 Present investigations

2.1 Aims of the thesis

- To extend the methodological support allowing to study transcriptome-wide differences in translational efficiency:
 - Expand the polysome-profiling experimental method to make it available for large sets of small samples while maintaining the possibility to quantitatively measure translational control
 - Expand the anota computational method for analysis of changes in translational efficiencies to generalize it to polysome- and ribosome- profiling data quantified using DNA-microarrays or RNA sequencing and to adapt it to classical characteristics of RNA sequencing data
- To unravel a novel role for estrogen receptor alpha in cancer: namely to investigate the mechanisms by which this transcription factor also affects gene expression's translational output

2.2 Results and discussion

2.2.1 Assessing differential translation and buffering using anota2seq

For translatoemes quantified by DNA-microarrays, the anota algorithm had been developed to correct changes in polysome-associated mRNA for changes in cytoplasmic mRNA levels by using a linear regression method (in **Figure 7A**, the anota models are plotted for the treatment (solid line) and control conditions (dotted line); Larsson et al. 2010). This approach was shown to outperform the translational efficiency (TE) score approach (also called log-ratio method) which compares the $\log(\text{translated mRNA data}/\text{cytoplasmic mRNA data})$ between conditions (Larsson et al. 2010). The first objective of **Paper I** was to develop anota2seq to allow similar modelling for RNA-sequencing data (accounting for the count nature of such data with a mean variance dependency which cannot be corrected by log-transformation alone) and to compare anota2seq to other methods for transcriptome-wide analysis of differential translation.

In the early days of next-generation sequencing, several studies used classical DNA-microarrays methods on $\log(\text{RNAseq counts}/\text{total library size})$ (Perkins et al. 2009; Cloonan et al. 2008). It was shown later that RNAseq data should be analyzed using specific methods for count data. However, another strategy was taken by Law et al. (2014) when they established a new data transformation technique which, when applied to RNAseq data, "gives" properties of DNA-microarrays data. It permits that methods optimized for DNA-microarrays can be used on transformed RNAseq data. Using a similar strategy, anota2seq is a method which was adapted from the "DNA-microarrays anota method" (Larsson et al. 2010; Larsson et al. 2011). Indeed, anota2seq allows for transformation or normalization of RNAseq data prior to applying similar models as in anota.

As explained earlier, from RNA quantification of polysome- or ribosome-profiling, one can detect changes in translational efficiency leading to changes in protein levels (a change in the amount of translated mRNA that is not explained by a change in cytoplasmic mRNA) or buffering which opposes changes in protein levels despite alterations in corresponding cytoplasmic mRNA. A third mode of regulation of gene expression corresponds to having similar changes in cytoplasmic mRNA levels and levels of translated mRNA i.e. change in mRNA abundance. **Paper II** (regulation of gene expression upon depletion of ER α in a prostate cancer cell line) and **Paper III** (differences between serum starved HCT-116 cells (a colon-cancer cell line) with or without tumor protein p53 (p53)) are two examples where a subset of transcripts show increased or decreased abundance that were buffered at the level of translation. To allow for characterization of this mode of regulation, anota2seq includes the option to perform analysis of translational buffering and categorizes regulated genes in one

of three modes of regulation: changes in translational efficiencies leading to altered protein levels (in short, "translation"), changes in translational efficiencies leading to buffering ("buffering") or changes in mRNA abundance.

Similarly as in anota, the model fitted in anota2seq for analysis of changes in translational efficiency leading to altered protein levels consists of a linear regression with translated mRNA as dependent variable and cytoplasmic mRNA and the sample class variable as independent variables. A common slope for all sample categories is considered and the translational effect is defined as a difference in intercepts between conditions (Larsson et al. 2010). On the other hand, translational buffering is defined as changes in cytoplasmic mRNA level that are not paralleled by changes in levels of translated mRNA (Figure 10). As such, performing analysis of buffering considers cytoplasmic mRNA as dependent variable and translated mRNA as independent variable (together with the sample class). The "mRNA abundance" mode of regulation, corresponds to significant changes in both cytoplasmic and translated mRNA with changes in the same direction.

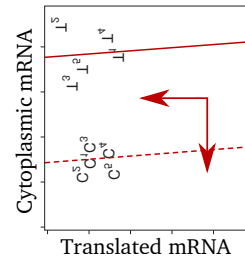


Figure 10: Anota2seq model for analysis of differential buffering. Same gene and condition as in Figure 7A (bottom left) but for analysis of differential buffering in anota2seq, cytoplasmic and translated mRNA have been reversed.

In order to compare the ability of different methods to identify genes regulated at the level of mRNA translation, we designed a simulation study. RNAseq counts were simulated for cytoplasmic and translated mRNA under two conditions for 4 categories of genes:

- truly unregulated genes (simulated from the same theoretical distributions for both conditions)
- genes regulated by differential translation (cytoplasmic mRNA counts were simulated from the same distribution but a positive or negative fold change (FC) was applied between the 2 conditions for the translated mRNA simulation)
- genes regulated by differential buffering (translated mRNA counts were simulated from the same distribution but a positive or negative FC was applied between the 2 conditions for the cytoplasmic mRNA simulation)
- genes regulated by differences in mRNA abundance (The same FC was applied

between conditions for both cytoplasmic and translated mRNA theoretical distributions)

We concluded from this study that anota2seq outperforms current methods notably by allowing distinction between changes in translational efficiency affecting protein levels and buffering. Indeed, anota2seq showed higher area under the receiver operating characteristic curve (AUC) and precision than other methods for identification of genes in the "differential translation" category (i.e. true differential translation). All methods except anota2seq, are based on the principle that a gene will be identified as differentially translated if the translated mRNA change between conditions is significantly different from the change in cytoplasmic mRNA levels, regardless of whether the difference is higher in translated mRNA or cytoplasmic mRNA. The inability of most methods to distinguish between differential translation leading to altered protein levels and buffering has led in the past to incorrect biological conclusions as explained in Larsson et al. (2010).

As mentioned earlier, regulation by translational buffering has been observed in other contexts (McManus et al. 2014; Artieri and Fraser 2014; Cenik et al. 2015; Lalanne et al. 2018). When discussing these different studies, we realized that translational buffering could be categorized in different contexts. The first context would be when translational buffering occurs in the context of transcript-dosage compensation (McManus et al. 2014; Artieri and Fraser 2014; Lalanne et al. 2018) or between different individuals (Cenik et al. 2015; Dassi et al. 2015) where differences between conditions are static. We envisaged that mechanisms at play in such cases are likely to differ from mechanisms mediating translational buffering upon adaptive responses which can be reverted (as when observing translational buffering upon depletion of ER α which can be a therapy target). In this second context (exemplified in **Paper II**), we would use the term "translational offsetting". Finally, response to acute stress (Tebaldi et al. 2012) or delayed synthesis between mRNA and protein can cause transient translational buffering, called equilibration at the translational level.

As any model, our simulation study has limitations and its applicability scope is unclear. However, the parameters used for the control condition were estimated from a real polysome-profiling dataset (Guan et al. 2017) to maximize the confidence in inferred conclusions from this study. Furthermore, using such a simulation study allowed us to test additional parameters: for instance, how robust the different algorithms for translome analyses are against increased variance and reduced sequencing depth. All algorithms perform well under increased variance and reduced sequencing depths as long as the number of mapped reads to mRNAs was higher than 5 million. As our simulation data are public, such explorations could easily be ex-

tended to verify for instance the impact of batch effects with increasing sizes as did Chothani et al. (2019). It would also be interesting to evaluate the impact of strong down-regulation of global translation. Indeed, as discussed in (Gandin et al. 2016), assessment of gene-level relative differences in translation between conditions can be influenced by reductions in translation of most genes quantified in parallel (as would be observed for instance upon mTOR inhibition).

Finally, even if the differences between the methods were smaller, anota2seq was shown to outperform current methods for statistical analysis of translational efficiency even in the absence of translational buffering. When further exploring the reason why some methods seemed to underperform, we observed that Xtail (Xiao et al. 2016) typically detects high amounts of non-differentially translated genes when tested under a NULL model (no true differences in gene expression) and that Babel poorly controls type I error (this was also noticed by Xiao et al. (2016) in their method comparison).

Thus, anota2seq allows efficient analysis of translomes quantified using DNA-microarrays or RNAseq which is essential to further our knowledge of the role of translational control in cancer and other diseases.

2.2.2 Transcriptome-wide analysis of mRNA translation in tissue samples

Dysregulation of translation contributes to both initiation and progression of cancer by inducing global changes in protein synthesis rates and changes in the translational activity of specific mRNAs encoding cancer related proteins (Truitt and Ruggero 2016). Dysregulated mRNA translation can also mediate resistance towards targeted therapies (Boussemart et al. 2014). This occurs notably via activation of eIF4E which is hyper-active in most cancers, including those of the breast (De Benedetti and Graff 2004). Furthermore, preliminary analyzes performed in our research group on a small subsets of breast cancers showed that regulation at the translational level could unravel molecular subgroups of tumors which could not have been distinguished when only looking at the transcriptomic level (unpublished).

Breast cancer is a highly heterogeneous disease at the molecular level. A first indication for this heterogeneity is the wide range of ER expression in population-based cohorts (Osborne 1998; Wenger et al. 1993). Furthermore, despite ER expression correlating with proliferation and tumor aggressiveness (Wenger et al. 1993), overall ER-tumors usually show poorer prognosis than ER+ ones (Knight et al. 1977). This constituted an initial proof that breast tumors can be driven by many different molecular mechanisms. Intra-tumor heterogeneity is also common in this disease (Teixeira et al. 1995). Thus, studying breast cancer biology nowadays requires using patient material

from large cohorts that are representative of the variability of this disease. A cohort of 161 biobanked breast cancer tissue samples was identified with the aim to perform translatome analysis.

Analyzing gene expression from tumor material implies the use of optimized experimental methods for low input of RNA. Thus, the polysome-profiling technique first needed to be adapted for large sample sizes with low RNA amounts. Indeed, in the classical linear gradient approach, mRNAs are separated by sedimentation and the entire volume is fractionated along the gradient in 26 fractions (Figure 11A; red numbers). Collecting the polysome-associated mRNA entails pooling a large volume (>3 mL) across many fractions (16 to 25 on Figure 11A) which is labor-intensive and can cause sample loss. Our optimization allowed to concentrate the polysome-associated mRNA in one or two fractions instead (fractions 16-17 on Figure 11B).

Using two conditions of a cell-line model (HCT116 p53^{-/-} and HCT116 p53^{+/+}), we performed polysome profiling using both methods in parallel and compare the results in terms of RNA quantity, quality and output of the analysis of differential translation using an-

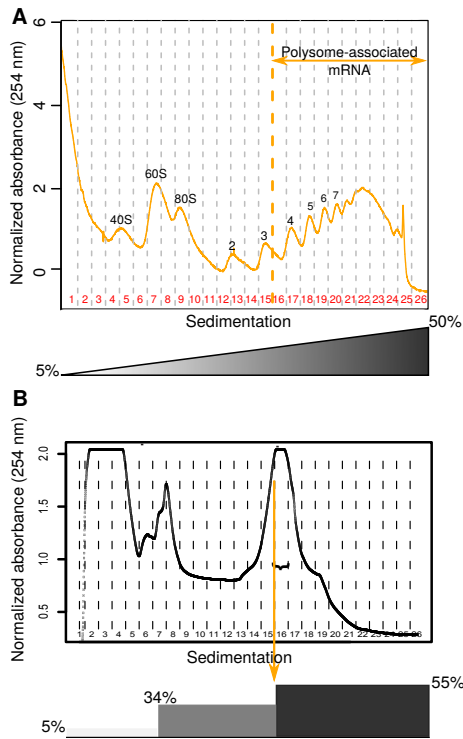


Figure 11: Polysome-profiling using the linear and the optimized non-linear gradient methods. Example of a polysome profile when using the classical method (A) or the optimized non-linear gradient (B). Fractions separations are shown as dotted lines. Fractions numbers are provided at the bottom of each plot.

(A) Modified from Gandin, V et al. (2016). "nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs." In: *Genome research* 26(5), pp. 636-648. doi: <https://doi.org/10.1101/gr.197566.115>. ©2016 Gandin et al.; Published by Cold Spring Harbor Laboratory Press. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>. (B) Modified from Liang C et al. "Polysome-profiling in small tissue samples." *Nucleic Acids Res.* 2018 Jan 9;46(1):e3. doi: <https://doi.org/10.1093/nar/gkx940>

ota2seq. High RNA quality and similar amounts were obtained from both methods. Next, we performed a differential expression analysis on polysome-associated mRNA obtained from both methods to assess gene-level differences between HCT116 p53+/+ and p53-/-. When comparing FCs, a high Spearman correlation (0.74) was observed between those of the linear gradient polysome-associated mRNA data and those from the optimized non-linear gradient data. After adjustment for multiple testing using Benjamini-Hochberg false discovery rates (FDRs), p-values obtained from the optimized non-linear method appeared globally lower. Taken together, these results are consistent with both methods resulting in similar effect sizes and different variability (the optimized method providing data with lower variance among biological replicates).

Finally, we also validated the feasibility for the optimized non linear gradient method to isolates efficiently translated mRNA from biobanked tissues. For this purpose, we selected 5 breast cancer tissue samples from the large cohort mentioned above (with 161 patient samples) and applied our new polysome-profiling method. RNAseq libraries were prepared for polysome-associated and cytoplasmic mRNA using Smartseq2 which has been developed for low input samples such as single-cell RNAseq (Picelli et al. 2014). High coverage of the transcriptome could be reached from breast cancer tissue samples and the sequencing depth was above 5 million reads mapped to protein coding mRNA for 3 samples and just below this threshold (between 4.5 and 5 million) for the remaining 2. In **Paper I**, we showed that statistical analysis of differential translation is not influenced by sequencing depth if it is above 5 million reads mapped to protein coding mRNAs. Anot2seq results were also quite stable in cases where 25% of samples have a sequencing depth of 2.5 million reads mapped to mRNAs and 75% have at least 5 million reads.

Thus studying novel mechanisms of regulation of mRNA translation in large collection of tissue samples is feasible using the optimized non-linear polysome-profiling method and anota2seq.

2.2.3 Investigating the mechanisms by which the transcription factor estrogen receptor alpha affects gene expression at the translational level

As described in section 1.2.1.3, the aim of **Paper II** was to study the role of ER α in coordinating gene expression at multiple levels. Indeed, in a previous publication (Takizawa et al. 2015), potential non-nuclear functions of ER α were illustrated by showing that its depletion partially inhibited members of the mTOR and MAPK pathways. Therefore, we used BM67 cells (a prostate cancer cell-line derived from phosphatase and tensin homologue (PTEN)-deficient mice) and performed polysome-profiling fol-

lowed by quantification of the translome upon depletion of ER α . As expected from perturbation of a transcription factor, major differences were observed on the cytoplasmic mRNA level of many transcripts. However, at the translational level, observed regulations did not seem consistent with reduction in activity of important translation pathways; as seen by the limited amount of genes classified in anota2seq's translation mode (i.e. changes in translational efficiency leading to altered protein levels) and by the absence of significant translationally suppressed cellular functions upon ER α depletion. Instead, alterations in mRNA levels were largely translationally offset. This unexpected result was carefully validated using independent technologies: the initial quantification was performed using DNA-microarrays; RNAseq resulted in very similar measurements of cytoplasmic and polysome-associated mRNA in both conditions; a subset of targets (selected in all different modes of regulation of gene expression) were validated using Nanostring; at the protein level, Western blotting (WB) confirmed regulation observed by polysome association for all tested targets. Thus, ER α up- and downregulates the mRNA abundance of many targets but a majority of these alterations are offset at the level of translation. We validated that, at least for a subset of targets, such translational offsetting leads to maintained protein levels despite mRNA modulations.

Next, we sought to pinpoint mRNA characteristics which could mediate translational offsetting upon ER α depletion. For this purpose, we explored features in different parts of the mRNA. Firstly, we performed nanoCAGE on shER α which allowed to precisely determine the position of transcription start sites (TSSs) and thereby analyzing characteristics of the 5'UTRs. Secondly, we explored presence of miRNA target sites on offset vs. non-offset mRNAs. We sequenced small RNAs in order to investigate whether ER α -dependent miRNA regulation could reconcile translational offsetting alterations. Lastly, we observed major differences in codon usage along mRNAs which are induced but translationally offset upon ER α depletion as compared to non-offset mRNAs.

mRNAs which are downregulated but translationally offset appeared to have shorter and less structured 5'UTRs than non-offset transcripts. Their median mRNA length is close to what has been hypothesized to be the "optimal" length for efficient translation (Kozak 1987) and they lack complex structures which could limit their translational efficiency.

Regarding analysis of miRNAs, we initially observed an enrichment of target sites (from the targetScan database (Lewis et al. 2005)) among mRNAs whose upregulation is translationally offset. Therefore, we hypothesized that ER α may upregulate specific miRNAs which would target these offset mRNAs thus mediating translational offset-

ting. Interestingly, our small RNA sequencing revealed that ER α depletion leads to more downregulation than upregulation of miRNAs and that no sign of upregulation was seen for miRNAs targeting mRNAs whose levels are induced but translationally offset. Accordingly, we concluded that ER α -regulated miRNAs do not seem to mediate offsetting at the level of translation. Yet, we detected that mRNAs whose levels are reduced upon ER α depletion but translationally offset (i.e. those lacking complex 5'UTR structures) generally lack miRNA target sites which could also limit their translational efficiency.

Intriguingly, when analyzing characteristics of transcripts whose levels are induced by ER α depletion but translationally offset, a major distinction with non-offset mRNAs is their codon usage. Specifically, ER α -induced but translationally offset transcripts are enriched in codons depending, for their translation, on tRNAs modified at their U34 position.

Initially, we noticed a substantial overlap between codons enriched in upregulated but translationally offset mRNAs and codons enriched in proliferation-related mRNAs in the study by Gingold et al. (see section 1.2.1.2 and **Figure 6**). Among these codons, the strongest over-representation was seen for AAA and GAA which require U34-modified tRNAs for efficient translation (**Figure 8**). Consistently, we showed that ER α depletion downregulates protein levels of ELP3 which is one of the enzymes catalyzing the modification. Moreover, translational offsetting of DEK Proto-Oncogene (DEK) (an offset target which we extensively validated) could be reversed by re-expression of ELP3 in BM67 cell-lines where ELP3 had previously been knocked out. Finally, a strong reduction of these mcm⁵s²U modifications was quantified by mass spectrometry upon ER α inhibition in a breast cancer cell line. In conclusion, ER α regulates expression of U34-modifications and selectively offsets mRNA abundance at the translational level.

Studying mechanisms of translational offsetting allowed to formulate new perspectives in the role that ER α plays in cancer. Functionally, this implies that upon ER α depletion, transcripts enriched in "proliferation signature codons" will not be synthesized despite high levels of mRNA molecules. Furthermore, we showed that the anti-proliferative effect of ER α -inhibitors was strongly reduced in cells where the U34-modifying pathway had been inactivated. In melanoma, perturbation of the U34-modifying pathway was associated with resistance to B-Raf proto-oncogene, serine/threonine kinase (BRAF) inhibitors (Rapino et al. 2018). In breast cancer, resistance to endocrine therapies is a major concern which limits new survival benefits (Clarke et al. 2015). Studying the U34-modifying pathway as a potential target to overcome treatment resistance in ER α -dependent cancers may be a promising strategy.

Mechanistically, an intriguing remaining question is whether and to which extent reduced abundance of U34-modified tRNAs would influence translation elongation vs. initiation rates. However, this might be challenging to decipher as reduced elongation rates can themselves limit initiation by accumulation of ribosomes towards the 5' end of the coding sequence. In our study, we observed translational offsetting of a large amount of mRNAs from polysome-profiling data which implies that for such transcripts, the same amount of mRNA is associated with >3 ribosomes (we used this threshold to isolate efficiently translated mRNA) following ER α depletion whereas the mRNA level is increased. We showed that this is mediated by reduced availability of U34-modified tRNAs. Ranjan and Rodnina (2017) proposed that U34 enzymes cause ribosome pausing during translation elongation and suggested that lack of thiolation of tRNA^{UUU}_{Lys} increases the residence time of ribosomes on AAA codons by about 40% in a prokaryotic translation system. We investigated further whether such an increase in translocation time at codons requiring U34-modified tRNAs could explain the translational offsetting observed in **Paper II**. For this purpose, we selected one target which was induced but translationally offset upon ER α depletion: DEK. We chose this target because it has an extreme frequency of AAA, CAA and GAA codons compared to other mouse transcripts (**Figure 12A**). We validated its FC on cytoplasmic mRNA using qPCR (**Figure 12B**). We used ribosome flow model (RFM)s (Reuveni et al. 2011) to estimate translational output of the DEK mRNA upon several hypothetical reductions of translation elongation rate at codons decoded by U34-modified tRNAs. As input of the RFM, we needed to give the coding sequence of DEK, a range of initiation rates and codon-specific residence times. We sought to assess if the combination of increased residence time coupled with increased mRNA level could lead to maintained protein synthesis as compared to the control condition (**Figure 12C**). The RFMs indicated that ribosome residence time on codons requiring U34-modified tRNAs cannot alone explain observed translational offsetting under any initiation rate (**Figure 12D**). Therefore, ribosome drop-off could be a potential mechanism restricting translation under conditions where tRNAs are hypomodified at the U34 position but additional experiments are required to explore this further and eventually to understand how translational control can be used to maintain homeostasis.

2.2.3.1 Reflections on ethical considerations in bioinformatics projects My doctoral studies have focused on statistical method development and bioinformatics analyses of transcriptomes from experiments mainly performed on cell lines. Reflections on ethical considerations instinctively arise in research projects for instance when experiments need to be performed on animals and these reflections are then of utmost importance. However, in other contexts such as data analysis, ethical discussions are often overlooked. **Paper II** provided an example of interesting ethical considerations in data analysis which will be described here.

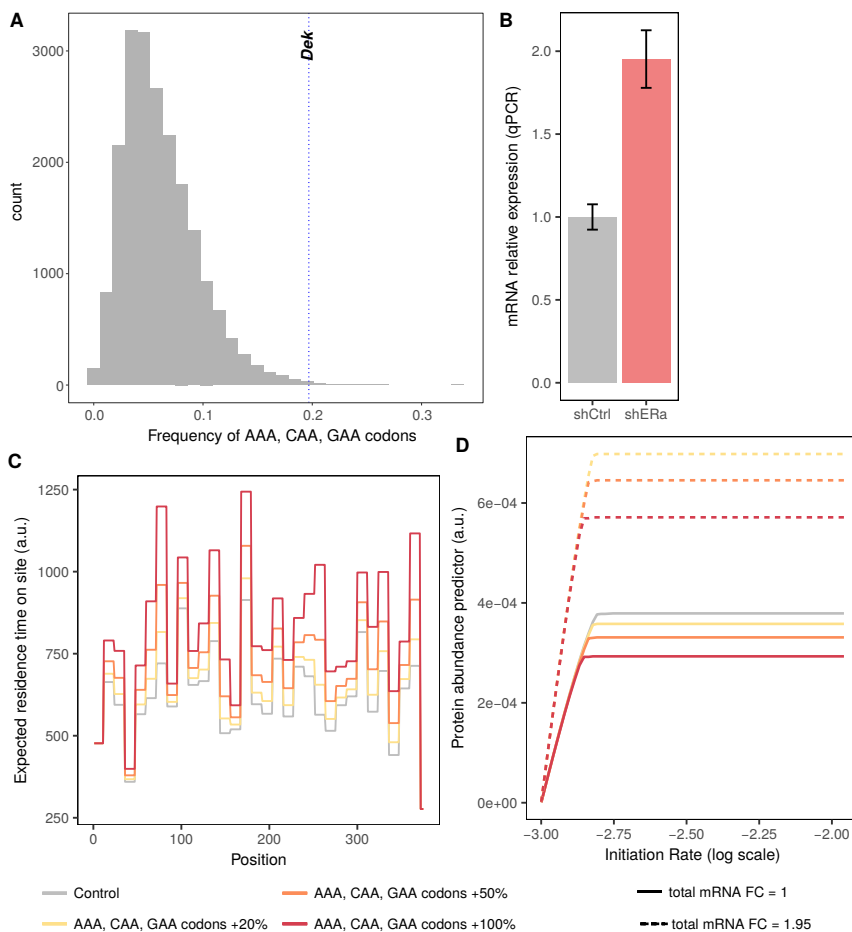


Figure 12: Increased ribosome residence time at AAA, CAA and GAA codons cannot alone explain DEK translational offsetting. We use the RFM to assess whether an increase in ribosome residence time on codons requiring U34 tRNA modifications can alone explain such a translational offsetting. We fit 4 RFMs: one with default residence codon times (calculated based on mouse tRNA copy numbers) and three other models with residence time on AAA, CAA and GAA codons increased by 20%, 50%, 100%. **(A)** Distribution of frequencies of AAA, CAA and GAA in coding sequences of the mouse transcriptome (from consensus coding sequence (CCDS) database). **(B)** DEK mRNA relative expression from shER α and control cells measured by qPCR. **(C)** Residence time (used as input in the RFM) of the ribosome at each position along the DEK coding sequence for baseline condition (gray) and when the residence time on AAA, CAA, GAA codons is increased by 20% (yellow), 50% (orange), 100% (red). Residence time is averaged across chunks of 12 codons. **(D)** Estimated protein abundance predictor as defined in Reuveni et al. (2011) showing the impact of increased residence times in case of unchanged abundance (plain lines) and when DEK mRNA levels is increased by 1.95 fold (dotted lines).

One classical example which raises questions related to scientific reasoning and ethics comes when the initial hypothesis of a study cannot be confirmed by the analysis. More specifically, a study has been designed based on a scientific hypothesis; it was conducted according to this design and then led to a negative result. Let's set apart the discussion about what should be considered as a negative result for now and focus on the following question: should one publish what would be considered as a negative result or should other options be considered? A common "excuse" for not trying to publish negative results is that it would be refused by journal editors themselves. Publication bias, also called positive-outcome bias, is an unfortunate reality in medical research. In clinical trial research for instance, studies with statistically significant results are definitely more likely to be published and more likely to get published early. This kind of bias can have deleterious consequences, notably because meta-analyses can only be based on published results. This issue has recently begun to be addressed by some journals which are dedicated to publication of negative results. Another question is: would it be unethical to publish these results as the "positive" results of another post-hoc hypothesis? Should this alternative be considered as a misleading way to make one's research more interesting than what is it in reality? Or should it be considered as the starting point of a potentially promising new hypothesis? Firstly, the answer to these questions depends on whether the research question is to be considered as having a "hypothesis confirmation" or "hypothesis generation" goal. The former would imply that a lot of research has previously been performed to lead to the hypothesis. In this case, publishing the negative result should probably be the reasonable option. An example of this would be phase III clinical trials, which are confirmatory trials and may lead to approval of new drugs or indications. In this case, a statistical analysis plan should be pre-defined and any additional analyzes should be interpreted as having lower levels of evidence. There should however be more freedom in research that is further away to applications for patients and research that is performed to generate new hypotheses. In this case, it should be acceptable to consider a post-hoc analysis as promising even if its level of evidence is not as high as the initial hypothesis.

In **Paper II**, the initial hypothesis was that ER α would not only regulate transcription of specific targets but also translational output. Indeed, ER α has been shown to interact with classical pathways which impinge on translational control such as the mTOR pathway. We therefore wanted to test whether ER α 's transcriptional output would also be regulated at the translational level notably via this pathway. As described above, we could however not confirm from our transcriptome-wide studies of changes in translation that ER α depletion impacted the mTOR pathway. However, the analysis of data from these experiments revealed other interesting mechanisms upon

ER α depletion (see **Paper II**). We published our results in the context of the second hypothesis, stating that this was not the initial hypothesis. Furthermore, additional validation experiments have been performed in order to give more strength to this post-hoc hypothesis. Several reflections were raised before writing the manuscript to decide in which order the "story" should be told and whether to omit the initial negative result. In the final version of this paper, the initial hypothesis is discussed in the introduction. Of course, ethics was not the only consideration when designing the manuscript but even in projects where the context of the research offers freedom in the way the analysis can be done, it is interesting to reflect upon the integrity and validity of scientific results from post-hoc hypotheses. Having stated that more freedom should be allowed in the data analysis of basic research, this paragraph should not be concluded before adding that data dredging is of course not to be advised.

3 Conclusion

Because of its tight connection with uncontrolled proliferation (Faller et al. 2015), immune response (Piccirillo et al. 2014), altered metabolism (Cunningham et al. 2014), invasion and metastasis (Robichaud et al. 2015); translational dysregulation is now itself considered as a hallmark of cancer (Vaklavas et al. 2017). This thesis provides methodological advances allowing to study further mechanisms by which translation influences cancer cell phenotypes. Indeed, our optimized gradient polysome-profiling method allows to isolate translatomers from tissue samples. Furthermore, we demonstrated that the anota2seq algorithm outperforms current methods for statistical identification of differences in translational efficiencies.

Furthermore, this thesis shines a light on an underappreciated mode of gene expression regulation whereby translation acts as an offsetting process which opposes protein levels despite fluctuations in corresponding mRNA abundance. Common mechanisms regulating translational offsetting are yet to be discovered but upon depletion of ER α in prostate cancer, the most influential regulatory pathway impinged on modification of specific tRNAs. This result could have implications in our understanding of how the proteome of hormone-dependent diseases is controlled as well as in strategies to tackle resistance to ER α -targeted therapies.

Acknowledgements

The work of all contributors to the research which led to the constituent papers presented in this thesis should be duly attributed. Below are provided the author lists of the articles in the order in which they have been cited here:

- *Generally applicable transcriptome-wide analysis of translation using anota2seq*: Christian Oertlin, JL, Carl Murie, Luc Furic, Ivan Topisirovic and Ola Larsson.
- *Translational offsetting as a mode of estrogen receptor α -dependent regulation of gene expression*: JL, Eric P. Kusnadi, Vincent van Hoef, Richard J. Rebello, Matthew Leibovitch, Johannes Ristau, Shan Chen, Mitchell G. Lawrence, Krzysztof J. Szkop, Baila Samreen, Preetika Balanathan, Francesca Rapino, Pierre Close, Patricia Bukczynska, Karin Scharmann, Itsuhiro Takizawa, Gail P. Risbridger, Luke A. Selth, Sebastian A. Leidel, Qishan Lin, Ivan Topisirovic, Ola Larsson and Luc Furic.
- *Polysome-profiling in small tissue samples*: Shuo Liang, Hermano M. Bellato, JL, Fernanda C.S. Lupinacci, Christian Oertlin, Vincent van Hoef, Victor P. Andrade, Martín Roffé, Laia Masvidal, Glaucia N.M. Hajj and Ola Larsson.

For Papers I and II, the research was driven by the close collaboration between Ola Larsson, Luc Furic and Ivan Topisirovic who all three have been essential to most of the research I have been involved within the last years. For Paper I, Carl Murie performed an initial method exploration while Christian Oertlin was the main contributor to the simulation study and comparison between the different methods. In addition to Luc Furic, Ola Larsson and Ivan Topisirovic who followed the work within Paper II from the beginning to the end, many people have contributed over the years: Eric Kusnadi, Richard Rebello, Vincent van Hoef, Matthew Leibovitch performed and analyzed most of the experimental work; the expertise of Sebastian Leidel, Pierre Close, Qishan Lin and Luke Selth has also been essential for this project. Paper III came to be from a close collaboration between Ola Larsson and Glaucia Hajj; Shuo Liang and Hermano Bellato performed most of the experimental optimization of the new polysome-profiling method and also applied it to biobanked tissue samples.

I am very grateful that Jack Keene, Sven Nelander, Cecilia Williams and Marc Friedländer have accepted to be the opponent and examiners of my PhD studies.

I have been supervised on all my doctoral work by Ola Larsson who has always been available for guidance and support and I want to greatly thank him for that.

I would also like to thank my co-supervisor Janne Lehtiö as well as everyone who spent shorter or longer periods of their professional paths around Ola's lab and with

whom I have been lucky to work: Carl Murie, Vincent van Hoef, Laia Masvidal, Shuo Liang, Christian Oertlin, Fernanda Lupinacci, Hermano Bellato, Johannes Ristau, Álvaro Serrano, Margarita Bartish, Hui Liu, Krzysztof Szkop, Visnu Chaparro, Martin Roffé, Baila Samreen, Dongmei Tong, Shan Chen, Thor Wahlestedt, Charlotte Rolny, Sabrina Ferreira. Vincent, I am really grateful for every, unfortunately too rare, interactions we had; Laia, your involvement in your work has always been inspiring; Shuo, thanks for the nice time as officemates and for sharing your experience of the PhD path; Christian, thanks for your refreshing views on science; Hermano, I have learned more from you than what you could ever realize and I want to specifically thank you for that; Johannes, I just feel so lucky to have gone through my PhD with you around; Margarita, I admire the way you drive your research and have greatly appreciated our discussions; Hui, thanks for your scientific curiosity; Krzysztof, your supportive attitude has always been greatly appreciated; Baila, thanks for your tenacity during the work on Paper II of this thesis. [fr] Et aussi, 1000 mercis à Florent Comte pour l'aide pour les figures ! [/fr].

On a more personal note, I want to address 3 special thanks: to Claus Mayer for encouraging me to work in research and for your trust in my scientific beginnings, to John Lövrot for getting me to realize how fun and creative science can be and for the mentoring that you offered at times when I needed it and to Johannes Ristau for liking to talk science through just as much as me (hopefully) and for your support.

On an even more personal note, I should maybe say that I would not claim that the PhD path has always been fun for me (yes, it mostly should!) and I want to take this as an opportunity to thank people whose presence along the way has been so essential: *[in FrenGLISH in the text]* les sis', les couz' (& assimilés), les cops' (from around here, d'ailleurs, de Génération.s Stockholm), les parents (& beau-) and most particularly Robin!

References

- Abbott, JA, Francklyn, CS and Robey-Bond, SM (2014). "Transfer RNA and human disease". In: *Frontiers in Genetics* 5(JUN).
- Abdelmohsen, K and Gorospe, M (2010). "Posttranscriptional regulation of cancer traits by HuR." In: *Wiley interdisciplinary reviews. RNA* 1(2), pp. 214–29.
- Alain, T et al. (2012). "eIF4E/4E-BP Ratio Predicts the Efficacy of mTOR Targeted Therapies". In: *Cancer Research* 72(24), pp. 6468–6476.
- Albert, FW, Muzzey, D, Weissman, JS and Kruglyak, L (2014). "Genetic influences on translation in yeast". In: *PLoS genetics* 10(10), e1004692.
- Artieri, CG and Fraser, HB (2014). "Evolution at two levels of gene expression in yeast". In: *Genome Research* 24(3), pp. 411–421.
- Aspden, JL, Eyre-Walker, YC, Phillips, RJ, Amin, U, Mumtaz, MAS, Brocard, M and Couso, JP (2014). "Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq". In: *eLife* 3.
- Avdulov, S et al. (2004). "Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant phenotype in human mammary epithelial cells." In: *Cancer cell* 5(6), pp. 553–63.
- Baird, TD, Palam, LR, Fusakio, ME, Willy, JA, Davis, CM, McClintick, JN, Anthony, TG and Wek, RC (2014). "Selective mRNA translation during eIF2 phosphorylation induces expression of IBTK α ". In: *Molecular Biology of the Cell* 25(10), pp. 1686–1697.
- Beyer, AL and Osheim, YN (1988). "Splice site selection, rate of splicing, and alternative splicing on nascent transcripts." In: *Genes & development* 2(6), pp. 754–65.
- Bhat, M, Robichaud, N, Hulea, L, Sonenberg, N, Pelletier, J and Topisirovic, I (2015). "Targeting the translation machinery in cancer". In: *Nature Reviews. Drug Discovery* 14(4), pp. 261–278.
- Bonnin, P, Kern, N, Young, NT, Stansfield, I and Romano, MC (2017). "Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile." In: *PLoS computational biology* 13(5), e1005555.
- Bosland, MC (2013). "A perspective on the role of estrogen in hormone-induced prostate carcinogenesis". In: *Cancer Letters* 334(1), pp. 28–33.
- Bosland, MC, Ford, H and Horton, L (1995). "Induction at high incidence of ductal prostate adenocarcinomas in NBL/Cr and Sprague-Dawley Hsd:SD rats treated with a combination of testosterone and estradiol-17 β or diethylstilbestrol". In: *Carcinogenesis* 16(6), pp. 1311–1317.
- Boussemart, L et al. (2014). "eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies". In: *Nature* 513(7516), pp. 105–109.
- Branca, RMM, Orre, LM, Johansson, HJ, Granholm, V, Huss, M, Pérez-Bercoff, Å, Forshed, J, Käll, L and Lehtiö, J (2014). "HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics". In: *Nature Methods* 11(1), pp. 59–62.
- Brar, GA and Weissman, JS (2015). "Ribosome profiling reveals the what, when, where and how of protein synthesis". In: *Nature Reviews Molecular Cell Biology* 16(11), pp. 651–664.

- Brenner, S, Jacob, F and Meselson, M (1961). "An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis". In: *Nature* 190(4776), pp. 576–581.
- Brunn, GJ, Hudson, CC, Sekulić, A, Williams, JM, Hosoi, H, Houghton, PJ, Lawrence, JC and Abraham, RT (1997). "Phosphorylation of the translational repressor PHAS-I by the mammalian target of rapamycin." In: *Science (New York, N.Y.)* 277(5322), pp. 99–101.
- Cambridge, SB, Gnad, F, Nguyen, C, Bermejo, JL, Krüger, M and Mann, M (2011). "Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover". In: *Journal of Proteome Research* 10(12), pp. 5275–5284.
- Cardoso, F, Kyriakides, S, Ohno, S, Penault-Llorca, F, Poortmans, P, Rubio, IT, Zackrisson, S and Senkus, E (2019). "Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up". In: *Annals of Oncology* 30(10), pp. 1674–1674.
- Carninci, P et al. (2005). "The Transcriptional Landscape of the Mammalian Genome". In: *Science* 309(5740), pp. 1559–1563.
- Genik, C et al. (2015). "Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans". In: *Genome Research* 25(11), pp. 1610–1621.
- Chan, PP and Lowe, TM (2009). "GtRNAdb: A database of transfer RNA genes detected in genomic sequence". In: *Nucleic Acids Research* 37(SUPPL. 1).
- Chapeville, F, Lipmann, F, Von Ehrenstein, G, Weisblum, B, Ray, WJ and Benzer, S (1962). "On the role of soluble ribonucleic acid in coding for amino acids." In: *Proceedings of the National Academy of Sciences of the United States of America* 48(6), pp. 1086–92.
- Cheng, Z, Teo, G, Krueger, S, Rock, TM, Koh, HW, Choi, H and Vogel, C (2016). "Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress". In: *Molecular Systems Biology* 12(1), p. 855.
- Choder, M (2004). "Rpb4 and Rpb7: subunits of RNA polymerase II and beyond." In: *Trends in biochemical sciences* 29(12), pp. 674–81.
- Chothani, S, Adami, E, Ouyang, JF, Viswanathan, S, Hubner, N, Cook, SA, Schafer, S and Rackham, OJL (2019). "deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Riboseq and RNAseq Data". In: *Current Protocols in Molecular Biology* 129(1).
- Chu, J and Pelletier, J (2018). "Therapeutic Opportunities in Eukaryotic Translation." In: *Cold Spring Harbor perspectives in biology* 10(6), a032995.
- Clarke, R, Tyson, JJ and Dixon, JM (2015). "Endocrine resistance in breast cancer - An overview and update". In: *Molecular and Cellular Endocrinology* 418, pp. 220–234.
- Cloonan, N et al. (2008). "Stem cell transcriptome profiling via massive-scale mRNA sequencing". In: *Nature Methods* 5(7), pp. 613–619.
- Cottrell, KA, Chaudhari, HG, Cohen, BA and Djuranovic, S (2018). "PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs." In: *Nature communications* 9(1), p. 301.
- Cozen, AE, Quartley, E, Holmes, AD, Hrabeta-Robinson, E, Phizicky, EM and Lowe, TM (2015). "ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments." In: *Nature Methods* 12(9), pp. 879–884.
- Crick, FH (1958). "On protein synthesis." In: *Symposia of the Society for Experimental Biology* 12, pp. 138–63.
- Crick, F (1970). "Central Dogma of Molecular Biology". In: *Nature* 227(5258), pp. 561–563.

- Cunningham, JT, Moreno, MV, Lodi, A, Ronen, SM and Ruggero, D (2014). “Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, PRPS2, to drive cancer”. In: *Cell* 157(5), pp. 1088–1103.
- Czerwoniec, A, Dunin-Horkawicz, S, Purta, E, Kaminska, KH, Kasprzak, JM, Bujnicki, JM, Grosjean, H and Rother, K (2009). “MODOMICS: a database of RNA modification pathways. 2008 update.” In: *Nucleic acids research* 37(Database issue), pp. D118–21.
- Dassi, E, Greco, V, Sidarovich, V, Zuccotti, P, Arseni, N, Scaruffi, P, Tonini, GP and Quattrone, A (2015). “Translational compensation of genomic instability in neuroblastoma”. In: *Scientific Reports* 5(1), p. 14364.
- De Benedetti, A and Graff, JR (2004). “eIF-4E expression and its role in malignancies and metastases.” In: *Oncogene* 23(18), pp. 3189–99.
- Delaunay, S et al. (2016). “Elp3 links tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis in breast cancer.” In: *The Journal of experimental medicine* 213(11), pp. 2503–2523.
- Dever, TE and Green, R (2012). “The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes”. In: *Cold Spring Harbor Perspectives in Biology* 4(7), a013706–a013706.
- Djebali, S et al. (2012). “Landscape of transcription in human cells”. In: *Nature* 489(7414), pp. 101–108.
- Dowling, RJO et al. (2010). “mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs.” In: *Science (New York, N.Y.)* 328(5982), pp. 1172–6.
- Dowsett, M et al. (2010). “Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen.” In: *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28(3), pp. 509–18.
- Drexler, HL, Choquet, K and Churchman, LS (2019). “Human co-transcriptional splicing kinetics and coordination revealed by direct nascent RNA sequencing”. In: *bioRxiv*, p. 611020.
- Drummond, DA and Wilke, CO (2008). “Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution”. In: *Cell* 134(2), pp. 341–352.
- EBCTCG (1998). “Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists’ Collaborative Group.” In: *Lancet (London, England)* 351(9114), pp. 1451–67.
- El Yacoubi, B, Bailly, M and Crécy-Lagard, V de (2012). “Biosynthesis and function of posttranscriptional modifications of transfer RNAs”. In: *Annual Review of Genetics* 46, pp. 69–95.
- Ellem, SJ, Schmitt, JF, Pedersen, JS, Frydenberg, M and Risbridger, GP (2004). “Local aromatase expression in human prostate is altered in malignancy.” In: *The Journal of clinical endocrinology and metabolism* 89(5), pp. 2434–41.
- Faller, WJ et al. (2015). “mTORC1-mediated translational elongation limits intestinal tumour initiation and growth.” In: *Nature* 517(7535), pp. 497–500.
- Fan, H and Penman, S (1970). “Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis.” In: *Journal of molecular biology* 50(3), pp. 655–70.
- Fan, Q et al. (2017). “A Kinase Inhibitor Targeted to mTORC1 Drives Regression in Glioblastoma.” In: *Cancer cell* 31(3), pp. 424–435.

- Filippova, N, Yang, X, King, P and Nabors, LB (2012). “Phosphoregulation of the RNA-binding protein Hu antigen R (HuR) by Cdk5 affects centrosome function.” In: *The Journal of biological chemistry* 287(38), pp. 32277–87.
- Floor, SN and Doudna, JA (2016). “Tunable protein synthesis by transcript isoforms in human cells.” In: *eLife* 5.
- Furic, L et al. (2010). “eIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression.” In: *Proceedings of the National Academy of Sciences of the United States of America* 107(32), pp. 14134–9.
- Gandin, V, Sikström, K, Alain, T, Morita, M, McLaughlan, S, Larsson, O and Topisirovic, I (2014). “Polysome fractionation and analysis of mammalian translatoemes on a genome-wide scale”. In: *Journal of Visualized Experiments: JoVE* (87).
- Gandin, V et al. (2016). “nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs.” In: *Genome research* 26(5), pp. 636–648.
- Gebauer, F, Preiss, T and Hentze, MW (2012). “From cis-regulatory elements to complex RNPs and back”. In: *Cold Spring Harbor Perspectives in Biology* 4(7), a012245.
- Gehrke, L, Bast, RE and Ilan, J (1981). “An analysis of rates of polypeptide chain elongation in avian liver explants following in vivo estrogen treatment.” In: *The Journal of biological chemistry* 256(5), pp. 2514–21.
- Gerashchenko, MV and Gladyshev, VN (2017). “Ribonuclease selection for ribosome profiling”. In: *Nucleic Acids Research* 45(2), e6–e6.
- Geter, PA, Ernlund, AW, Bakogianni, S, Alard, A, Arju, R, Giashuddin, S, Gadi, A, Bromberg, J and Schneider, RJ (2017). “Hyperactive mTOR and MNK1 phosphorylation of eIF4E confer tamoxifen resistance and estrogen independence through selective mRNA translation reprogramming.” In: *Genes & development* 31(22), pp. 2235–2249.
- Gingold, H et al. (2014). “A dual program for translation regulation in cellular proliferation and differentiation.” In: *Cell* 158(6), pp. 1281–1292.
- Gingras, AC, Raught, B and Sonenberg, N (1999). “eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation.” In: *Annual review of biochemistry* 68(1), pp. 913–63.
- Graff, JR, Konicek, BW, Carter, JH and Marcusson, EG (2008). “Targeting the eukaryotic translation initiation factor 4E for cancer therapy.” In: *Cancer research* 68(3), pp. 631–4.
- Grammatikakis, I, Abdelmohsen, K and Gorospe, M (2017). “Posttranslational control of HuR function”. In: *Wiley Interdisciplinary Reviews: RNA* 8(1), e1372.
- Gregersen, LH, Schueler, M, Munschauer, M, Mastrobuoni, G, Chen, W, Kempa, S, Dieterich, C and Landthaler, M (2014). “MOV10 Is a 5' to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs”. In: *Molecular Cell* 54(4), pp. 573–585.
- Guan, BJ et al. (2017). “A Unique ISR Program Determines Cellular Responses to Chronic Stress.” In: *Molecular cell* 68(5), 885–900.e6.
- Haimovich, G, Medina, DA, Causse, SZ, Garber, M, Millán-Zambrano, G, Barkai, O, Chávez, S, Pérez-Ortín, JE, Darzacq, X and Choder, M (2013). “Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis.” In: *Cell* 153(5), pp. 1000–11.

- 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." In: *Cell* 143(4), pp. 552–63.
- Hausser, J, Mayo, A, Keren, L and Alon, U (2019). "Central dogma rates and the trade-off between precision and economy in gene expression." In: *Nature communications* 10(1), p. 68.
- Hellen, CU (2018). "Translation Termination and Ribosome Recycling in Eukaryotes". In: *Cold Spring Harbor Perspectives in Biology* 10(10), a032656.
- Hentze, MW, Castello, A, Schwarzl, T and Preiss, T (2018). "A brave new world of RNA-binding proteins". In: *Nature Reviews Molecular Cell Biology* 19(5), pp. 327–341.
- Hinnebusch, AG and Lorsch, JR (2012). "The mechanism of eukaryotic translation initiation: new insights and challenges." In: *Cold Spring Harbor perspectives in biology* 4(10), a011544–a011544.
- Hinnebusch, AG, Ivanov, IP and Sonenberg, N (2016). "Translational control by 5'-untranslated regions of eukaryotic mRNAs". In: *Science (New York, N.Y.)* 352(6292), pp. 1413–1416.
- Hoagland, MB, Stephenson, ML, Scott, JF, Hecht, LI and Zamecnik, PC (1958). "A soluble ribonucleic acid intermediate in protein synthesis." In: *The Journal of biological chemistry* 231(1), pp. 241–57.
- Hsieh, AC et al. (2012). "The translational landscape of mTOR signalling steers cancer initiation and metastasis." In: *Nature* 485(7396), pp. 55–61.
- Ingolia, NT, Ghaemmaghami, S, Newman, JRS and Weissman, JS (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling". In: *Science (New York, N.Y.)* 324(5924), pp. 218–223.
- Ingolia, NT, Hussmann, JA and Weissman, JS (2018). "Ribosome Profiling: Global Views of Translation". In: *Cold Spring Harbor Perspectives in Biology*, a032698.
- International Agency for Research on Cancer - WHO - Global Cancer Observatory. *Cancer fact sheets - Breast*.
- International Agency for Research on Cancer - WHO - Global Cancer Observatory. *Cancer fact sheets - Prostate*.
- Iwasaki, S and Ingolia, NT (2017). "The Growing Toolbox for Protein Synthesis Studies." In: *Trends in biochemical sciences* 42(8), pp. 612–624.
- Jackson, RJ, Hellen, CUT and Pestova, TV (2010). "The mechanism of eukaryotic translation initiation and principles of its regulation". In: *Nature Reviews Molecular Cell Biology* 11(2), pp. 113–127.
- Jacob, F and Monod, J (1961). "Genetic regulatory mechanisms in the synthesis of proteins." In: *Journal of molecular biology* 3, pp. 318–56.
- Johansson, MJO, Esberg, A, Huang, B, Björk, GR and Byström, AS (2008). "Eukaryotic wobble uridine modifications promote a functionally redundant decoding system". In: *Molecular and Cellular Biology* 28(10), pp. 3301–3312.
- Jonas, S and Izaurralde, E (2015). "Towards a molecular understanding of microRNA-mediated gene silencing." In: *Nature reviews. Genetics* 16(7), pp. 421–33.
- Jovanovic, M et al. (2015). "Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens". In: *Science (New York, N.Y.)* 347(6226), p. 1259038.

- Juhling, F, Morl, M, Hartmann, RK, Sprinzl, M, Stadler, PF and Putz, J (2009). "tRNADB 2009: compilation of tRNA sequences and tRNA genes". In: *Nucleic Acids Research* 37(Database), pp. D159–D162.
- Katchy, A and Williams, C (2016). "Expression Profiles of Estrogen-Regulated MicroRNAs in Breast Cancer Cells". In: *Methods in molecular biology (Clifton, N.J.)* Vol. 1366, pp. 373–393.
- Keene, JD (2007). "RNA regulons: coordination of post-transcriptional events." In: *Nature reviews. Genetics* 8(7), pp. 533–43.
- Kiniry, SJ, Michel, AM and Baranov, PV (2019). "Computational methods for ribosome profiling data analysis." In: *Wiley interdisciplinary reviews. RNA*, e1577.
- Kirchner, S and Ignatova, Z (2015). "Emerging roles of tRNA in adaptive translation, signalling dynamics and disease". In: *Nature Reviews Genetics* 16(2), pp. 98–112.
- Knight, WA, Livingston, RB, Gregory, EJ and McGuire, WL (1977). "Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer." In: *Cancer research* 37(12), pp. 4669–71.
- Komili, S and Silver, PA (2008). "Coupling and coordination in gene expression processes: a systems biology view." In: *Nature reviews. Genetics* 9(1), pp. 38–48.
- Koromilas, AE, Lazaris-Karatzas, A and Sonenberg, N (1992). "mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E." In: *The EMBO journal* 11(11), pp. 4153–8.
- Kozak, M (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs." In: *Nucleic acids research* 15(20), pp. 8125–48.
- Kozak, M (1989). "The scanning model for translation: an update." In: *The Journal of Cell Biology* 108(2), pp. 229–241.
- Kustatscher, G, Grabowski, P and Rappsilber, J (2017). "Pervasive coexpression of spatially proximal genes is buffered at the protein level." In: *Molecular systems biology* 13(8), p. 937.
- Ladang, A et al. (2015). "Elp3 drives Wnt-dependent tumor initiation and regeneration in the intestine." In: *The Journal of experimental medicine* 212(12), pp. 2057–75.
- Lalanne, JB, Taggart, JC, Guo, MS, Herzog, L, Schieler, A and Li, GW (2018). "Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry." In: *Cell* 173(3), 749–761.e38.
- Larsson, O, Sonenberg, N and Nadon, R (2010). "Identification of differential translation in genome wide studies". In: *Proceedings of the National Academy of Sciences of the United States of America* 107(50), pp. 21487–21492.
- Larsson, O, Sonenberg, N and Nadon, R (2011). "anota: Analysis of differential translation in genome-wide studies". In: *Bioinformatics (Oxford, England)* 27(10), pp. 1440–1441.
- Larsson, O, Morita, M, Topisirovic, I, Alain, T, Blouin, MJ, Pollak, M and Sonenberg, N (2012). "Distinct perturbation of the translome by the antidiabetic drug metformin". In: *Proceedings of the National Academy of Sciences of the United States of America* 109(23), pp. 8977–8982.
- Larsson, O, Tian, B and Sonenberg, N (2013). "Toward a genome-wide landscape of translational control". In: *Cold Spring Harbor Perspectives in Biology* 5(1), a012302.

- Law, CW, Chen, Y, Shi, W and Smyth, GK (2014). “Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts”. In: *Genome Biology* 15(2), R29.
- Lazaris-Karatzas, A, Montine, KS and Sonenberg, N (1990). “Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap”. In: *Nature* 345(6275), pp. 544–547.
- Leppek, K, Das, R and Barna, M (2018). “Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them”. In: *Nature Reviews Molecular Cell Biology* 19(3), pp. 158–174.
- Levin, ER (2009). “Membrane oestrogen receptor alpha signalling to cell functions.” In: *The Journal of physiology* 587(Pt 21), pp. 5019–23.
- Lewis, BP, Burge, CB and Bartel, DP (2005). “Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.” In: *Cell* 120(1), pp. 15–20.
- Li, GW, Burkhardt, D, Gross, C and Weissman, J (2014). “Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources”. In: *Cell* 157(3), pp. 624–635.
- Li, JJ and Biggin, MD (2015). “Gene expression. Statistics requantitates the central dogma”. In: *Science (New York, N.Y.)* 347(6226), pp. 1066–1067.
- Li, JJ, Chew, GL and Biggin, MD (2017a). “Quantitating translational control: mRNA abundance-dependent and independent contributions and the mRNA sequences that specify them”. In: *Nucleic Acids Research* 45(20), pp. 11821–11836.
- Li, W, Wang, W, Uren, PJ, Penalva, LOF and Smith, AD (2017b). “Riborex: fast and flexible identification of differential translation from Ribo-seq data”. In: *Bioinformatics* 33(11), pp. 1735–1737.
- Liu, Y and Aebersold, R (2016). “The interdependence of transcript and protein abundance: new data-new complexities”. In: *Molecular Systems Biology* 12(1), p. 856.
- Liu, Y, Beyer, A and Aebersold, R (2016). “On the Dependency of Cellular Protein Levels on mRNA Abundance”. In: *Cell* 165(3), pp. 535–550.
- Lotan, R, Bar-On, VG, Harel-Sharvit, L, Duek, L, Melamed, D and Choder, M (2005). “The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs”. In: *Genes & Development* 19(24), pp. 3004–3016.
- Lotan, R, Goler-Baron, V, Duek, L, Haimovich, G and Choder, M (2007). “The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms”. In: *The Journal of Cell Biology* 178(7), pp. 1133–1143.
- Love, MI, Huber, W and Anders, S (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15(12), p. 550.
- Man, O and Pilpel, Y (2007). “Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species”. In: *Nature Genetics* 39(3), pp. 415–421.
- Masvidal, L, Hulea, L, Furic, L, Topisirovic, I and Larsson, O (2017). “mTOR-sensitive translation: Cleared fog reveals more trees”. In: *RNA Biology* 14(10), pp. 1299–1305.
- Mathews, MB, Sonenberg, N and Hershey, JW (2007). “Origins and Principles of Translational Control”. In: *Translational Control in Biology and Medicine*. Vol. 48. 0, pp. 1–40.

- McManus, CJ, May, GE, Spealman, P and Shteyman, A (2014). “Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast”. In: *Genome Research* 24(3), pp. 422–430.
- Michels, AA, Robitaille, AM, Buczynski-Ruchonnet, D, Hodroj, W, Reina, JH, Hall, MN and Hernandez, N (2010). “mTORC1 Directly Phosphorylates and Regulates Human MAF1”. In: *Molecular and Cellular Biology* 30(15), pp. 3749–3757.
- Moreno, JA et al. (2012). “Sustained translational repression by eIF2 α -P mediates prion neurodegeneration.” In: *Nature* 485(7399), pp. 507–11.
- Morgentaler, A and Traish, AM (2009). “Shifting the paradigm of testosterone and prostate cancer: the saturation model and the limits of androgen-dependent growth.” In: *European urology* 55(2), pp. 310–20.
- Nedialkova, DD and Leidel, SA (2015). “Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity”. In: *Cell* 161(7), pp. 1606–1618.
- Nelson, AW, Tilley, WD, Neal, DE and Carroll, JS (2014). “Estrogen receptor beta in prostate cancer: friend or foe?” In: *Endocrine-Related Cancer* 21(4), T219–T234.
- Novoa, EM and Ribas de Pouplana, L (2012). “Speeding with control: codon usage, tRNAs, and ribosomes.” In: *Trends in genetics : TIG* 28(11), pp. 574–81.
- Novoa, EM, Pavon-Eternod, M, Pan, T and Ribas de Pouplana, L (2012). “A Role for tRNA Modifications in Genome Structure and Codon Usage”. In: *Cell* 149(1), pp. 202–213.
- O’Connor, PBF, Andreev, DE, Baranov, PV, O’Connor, PBF, Andreev, DE and Baranov, PV (2016). “Comparative survey of the relative impact of mRNA features on local ribosome profiling read density”. In: *Nature Communications* 7, p. 12915.
- Olshen, AB, Hsieh, AC, Stumpf, CR, Olshen, RA, Ruggero, D and Taylor, BS (2013). “Assessing gene-level translational control from ribosome profiling.” In: *Bioinformatics (Oxford, England)* 29(23), pp. 2995–3002.
- Orre, LM, Vesterlund, M, Pan, Y, Arslan, T, Zhu, Y, Fernandez Woodbridge, A, Frings, O, Fredlund, E and Lehtiö, J (2019). “SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization”. In: *Molecular Cell* 73(1), 166–182.e7.
- Osborne, CK (1998). “Steroid hormone receptors in breast cancer management”. In: *Breast Cancer Research and Treatment* 51(3), pp. 227–238.
- Parker, C, Gillessen, S, Heidenreich, A, Horwich, A and ESMO Guidelines Committee (2015). “Cancer of the prostate: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.” In: *Annals of oncology : official journal of the European Society for Medical Oncology* 26 Suppl 5, pp. v69–77.
- Parsyan, A, Svitkin, Y, Shahbazian, D, Gkogkas, C, Lasko, P, Merrick, WC and Sonenberg, N (2011). “mRNA helicases: the tacticians of translational control.” In: *Nature reviews. Molecular cell biology* 12(4), pp. 235–45.
- Pelechano, V, Wei, W and Steinmetz, L (2015). “Widespread Co-translational RNA Decay Reveals Ribosome Dynamics”. In: *Cell* 161(6), pp. 1400–1412.
- Perkins, TT et al. (2009). “A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*.” In: *PLoS genetics* 5(7), e1000569.
- Phizicky, EM and Hopper, AK (2010). “tRNA biology charges to the front”. In: *Genes & Development* 24(17), pp. 1832–1860.

- Piccirillo, CA, Bjur, E, Topisirovic, I, Sonenberg, N and Larsson, O (2014). “Translational control of immune responses: from transcripts to translatoemes”. In: *Nature Immunology* 15(6), pp. 503–511.
- Picelli, S, Faridani, OR, Björklund, ÅK, Winberg, G, Sagasser, S and Sandberg, R (2014). “Full-length RNA-seq from single cells using Smart-seq2”. In: *Nature Protocols* 9(1), pp. 171–181.
- Plotkin, JB and Kudla, G (2011). “Synonymous but not the same: the causes and consequences of codon bias.” In: *Nature reviews. Genetics* 12(1), pp. 32–42.
- Poulain, S, Kato, S, Arnaud, O, Morlighem, JÉ, Suzuki, M, Plessy, C and Harbers, M (2017). “NanoCAGE: A Method for the Analysis of Coding and Noncoding 5'-Capped Transcripts”. In: pp. 57–109.
- Pour, PM and Stepan, K (1987). “Induction of prostatic carcinomas and lower urinary tract neoplasms by combined treatment of intact and castrated rats with testosterone propionate and N-nitrosobis(2-oxopropyl)amine.” In: *Cancer research* 47(21), pp. 5699–706.
- Ranjan, N and Rodnina, MV (2017). “Thio-Modification of tRNA at the Wobble Position as Regulator of the Kinetics of Decoding and Translocation on the Ribosome”. In: *Journal of the American Chemical Society* 139(16), pp. 5857–5864.
- Rapino, F et al. (2018). “Codon-specific translation reprogramming promotes resistance to targeted therapy”. In: *Nature* 558(7711), pp. 605–609.
- Reuveni, S, Meilijson, I, Kupiec, M, Ruppim, E and Tuller, T (2011). “Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model”. In: *PLoS Computational Biology* 7(9), e1002127.
- Rezgui, VAN, Tyagi, K, Ranjan, N, Konevega, AL, Mittelstaet, J, Rodnina, MV, Peter, M and Pedrioli, PGA (2013). “tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding.” In: *Proceedings of the National Academy of Sciences of the United States of America* 110(30), pp. 12289–94.
- Riba, A, Di Nanni, N, Mittal, N, Arhné, E, Schmidt, A and Zavolan, M (2019). “Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates.” In: *Proceedings of the National Academy of Sciences of the United States of America* 116(30), pp. 15023–15032.
- Ricke, WA, Ishii, K, Ricke, EA, Simko, J, Wang, Y, Hayward, SW and Cunha, GR (2006). “Steroid hormones stimulate human prostate cancer progression and metastasis”. In: *International Journal of Cancer* 118(9), pp. 2123–2131.
- Ricke, WA, McPherson, SJ, Bianco, JJ, Cunha, GR, Wang, Y and Risbridger, GP (2008). “Prostatic hormonal carcinogenesis is mediated by in situ estrogen production and estrogen receptor alpha signaling.” In: *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 22(5), pp. 1512–20.
- Robichaud, N et al. (2015). “Phosphorylation of eIF4E promotes EMT and metastasis via translational control of SNAIL and MMP-3.” In: *Oncogene* 34(16), pp. 2032–42.
- Robichaud, N, Sonenberg, N, Ruggero, D and Schneider, RJ (2018). “Translational Control in Cancer”. In: *Cold Spring Harbor Perspectives in Biology*, a032896.
- Robinson, MD and Smyth, GK (2008). “Small-sample estimation of negative binomial dispersion, with applications to SAGE data.” In: *Biostatistics (Oxford, England)* 9(2), pp. 321–32.

- Robinson, MD, McCarthy, DJ and Smyth, GK (2010). “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics (Oxford, England)* 26(1), pp. 139–140.
- Rodrik-Outmezguine, VS et al. (2016). “Overcoming mTOR resistance mutations with a new-generation mTOR inhibitor.” In: *Nature* 534(7606), pp. 272–6.
- Rolfe, DF and Brown, GC (1997). “Cellular energy utilization and molecular origin of standard metabolic rate in mammals”. In: *Physiological Reviews* 77(3), pp. 731–758.
- Roy, B and Jacobson, A (2013). “The intimate relationships of mRNA decay and translation.” In: *Trends in genetics : TIG* 29(12), pp. 691–9.
- Rubio, CA, Weisburd, B, Holderfield, M, Arias, C, Fang, E, DeRisi, JL and Fanidi, A (2014). “Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation.” In: *Genome biology* 15(10), p. 476.
- Rudolph, KL, Schmitt, BM, Villar, D, White, RJ, Marioni, JC, Kutter, C and Odom, DT (2016). “Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States”. In: *PLoS Genetics* 12(5).
- Santen, RJ, Petroni, GR, Fisch, MJ, Myers, CE, Theodorescu, D and Cohen, RB (2001). “Use of the aromatase inhibitor anastrozole in the treatment of patients with advanced prostate carcinoma”. In: *Cancer* 92(8), pp. 2095–2101.
- Sarbassov, DD, Guertin, DA, Ali, SM and Sabatini, DM (2005). “Phosphorylation and Regulation of Akt/PKB by the Rictor-mTOR Complex”. In: *Science* 307(5712), pp. 1098–1101.
- Sarin, LP et al. (2018). “Nano LC-MS using capillary columns enables accurate quantification of modified ribonucleosides at low femtomol levels”. In: *RNA* 24(10), pp. 1403–1417.
- Schmitt, BM, Rudolph, KLM, Karagianni, P, Fonseca, NA, White, RJ, Talianidis, I, Odom, DT, Marioni, JC and Kutter, C (2014). “High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface.” In: *Genome research* 24(11), pp. 1797–807.
- Schwanhäusser, B, Busse, D, Li, N, Dittmar, G, Schuchhardt, J, Wolf, J, Chen, W and Selbach, M (2011). “Global quantification of mammalian gene expression control”. In: *Nature* 473(7347), pp. 337–342.
- Sharma, AK, Sormanni, P, Ahmed, N, Ciryam, P, Friedrich, UA, Kramer, G and O'Brien, EP (2019). “A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data”. In: *PLOS Computational Biology* 15(5), e1007070.
- Shi, Y, Taylor, SI, Tan, SL and Sonenberg, N (2003). “When Translation Meets Metabolism: Multiple Links to Diabetes”. In: *Endocrine Reviews* 24(1), pp. 91–101.
- Simoncini, T, Hafezi-Moghadam, A, Brazil, DP, Ley, K, Chin, WW and Liao, JK (2000). “Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase”. In: *Nature* 407(6803), pp. 538–541.
- Smith, MR, Kaufman, D, George, D, Oh, WK, Kazanis, M, Manola, J and Kantoff, PW (2002). “Selective aromatase inhibition for patients with androgen-independent prostate carcinoma”. In: *Cancer* 95(9), pp. 1864–1868.
- Söderberg, O et al. (2006). “Direct observation of individual endogenous protein complexes in situ by proximity ligation.” In: *Nature methods* 3(12), pp. 995–1000.

- Sonenberg, N and Hinnebusch, AG (2009). "Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets". In: *Cell* 136(4), pp. 731–745.
- Takizawa, I et al. (2015). "Estrogen receptor alpha drives proliferation in PTEN-deficient prostate carcinoma by stimulating survival signaling, MYC expression and altering glucose sensitivity". In: *Oncotarget* 6(2), pp. 604–616.
- Taylor, BS et al. (2010). "Integrative Genomic Profiling of Human Prostate Cancer". In: *Cancer Cell* 18(1), pp. 11–22.
- Tebaldi, T, Re, A, Viero, G, Pegoretti, I, Passerini, A, Blanzieri, E and Quattrone, A (2012). "Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells". In: *BMC Genomics* 13(1), p. 220.
- Teixeira, MR, Pandis, N, Bardi, G, Andersen, JA, Mitelman, F and Heim, S (1995). "Clonal heterogeneity in breast cancer: Karyotypic comparisons of multiple intra- and extra-tumorous samples from 3 patients". In: *International Journal of Cancer* 63(1), pp. 63–68.
- Thoreen, CC, Chantranupong, L, Keys, HR, Wang, T, Gray, NS and Sabatini, DM (2012). "A unifying model for mTORC1-mediated regulation of mRNA translation". In: *Nature* 485(7396), pp. 109–113.
- Tom Dieck, S et al. (2015). "Direct visualization of newly synthesized target proteins in situ." In: *Nature methods* 12(5), pp. 411–4.
- Torres, AG, Batlle, E and Ribas de Pouplana, L (2014). "Role of tRNA modifications in human diseases". In: *Trends in Molecular Medicine* 20(6), pp. 306–314.
- Truitt, ML and Ruggero, D (2016). "New frontiers in translational control of the cancer genome". In: *Nature Reviews Cancer* 16(5), pp. 288–304.
- Tuller, T (2014). "Challenges and obstacles related to solving the codon bias riddles." In: *Biochemical Society transactions* 42(1), pp. 155–9.
- Vaklavas, C, Blume, SW and Grizzle, WE (2017). "Translational dysregulation in cancer: Molecular insights and potential clinical applications in biomarker development". In: *Frontiers in Oncology* 7(JUL).
- Van Nostrand, EL et al. (2016). "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)." In: *Nature methods* 13(6), pp. 508–14.
- Vivanco, I and Sawyers, CL (2002). "The phosphatidylinositol 3-Kinase AKT pathway in human cancer." In: *Nature reviews. Cancer* 2(7), pp. 489–501.
- Vogel, C and Marcotte, EM (2012). "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses". In: *Nature Reviews Genetics* 13(4), pp. 227–232.
- Waldron, JA, Tack, DC, Ritchey, LE, Gillen, SL, Wilczynska, A, Turro, E, Bevilacqua, PC, Assmann, SM, Bushell, M and Le Quesne, J (2019). "MRNA structural elements immediately upstream of the start codon dictate dependence upon eIF4A helicase activity". In: *Genome Biology* 20(1).
- Warner, JR, Knopf, PM and RICH, A (1963). "A multiple ribosomal structure in protein synthesis". In: *Proceedings of the National Academy of Sciences of the United States of America* 49, pp. 122–129.
- Wenger, CR, Beardslee, S, Owens, MA, Pounds, G, Oldaker, T, Vendely, P, Pandian, MR, Harrington, D, Clark, GM and McGuire, WL (1993). "DNA ploidy, S-phase, and steroid receptors in

- more than 127,000 breast cancer patients". In: *Breast Cancer Research and Treatment* 28(1), pp. 9–20.
- Wolfe, AL et al. (2014). "RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer." In: *Nature* 513(7516), pp. 65–70.
- Xiao, Z, Zou, Q, Liu, Y and Yang, X (2016). "Genome-wide assessment of differential translations with ribosome profiling data." In: *Nature communications* 7, p. 11194.
- Zhang, G and Ignatova, Z (2011). "Folding at the birth of the nascent chain: Coordinating translation with co-translational folding". In: *Current Opinion in Structural Biology* 21(1), pp. 25–31.
- Zhong, Y, Karaletsos, T, Drewe, P, Sreedharan, VT, Kuo, D, Singh, K, Wendel, HG and Rättsch, G (2017). "RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints." In: *Bioinformatics (Oxford, England)* 33(1), pp. 139–141.
- Zhouravleva, G, Frolova, L, Le Goff, X, Le Guellec, R, Inge-Vechtomov, S, Kisselev, L and Philippe, M (1995). "Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3." In: *The EMBO journal* 14(16), pp. 4065–72.
- Zinshteyn, B and Gilbert, WV (2013). "Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling". In: *PLoS Genetics* 9(8), e1003675.