

Université de Montréal

An Analysis of Translation Heterogeneity in Ribosome Profiling Data

Par  
Pedro do Couto Bordignon

Département de Biochimie et Médecine Moléculaire, Université de Montréal,  
Faculté de Médecine

Mémoire présenté en vue de l'obtention du grade de M.Sc.  
en Bio-informatique

Décembre 2019

© Pedro Bordignon, 2019

Ce mémoire intitulé:

An Analysis of Translation Heterogeneity in Ribosome Profiling Data

présenté par Pedro do Couto Bordignon

a été évalué par un jury composé des personnes suivantes :

Dre Gertraud Burger  
Présidente-rapporteure

Dr Sebastian Pechmann  
Directeur de recherche

Dre Sylvie Hamel  
Membre du jury

## Résumé

Les protéines sont responsables de pratiquement toutes les fonctions performées au sein du corps cellulaire et de ses alentours. Le contrôle de l'expression génique détermine l'abondance, la localisation et le moment de la production de protéines dans la cellule. Il s'agit de l'un des processus centraux à la régulation de la physiologie et du fonctionnement cellulaire. La moindre perte de balance dans ce complexe système engendre des conséquences majeures sur l'intégrité cellulaire, menant au développement de plusieurs maladies parfois incurables. La traduction de l'ARN messager en produit protéique constitue la dernière étape de l'expression génique. Elle est régulée de plusieurs façons, intrinsèques et extrinsèques à la séquence. Il s'agit également du processus cellulaire le plus coûteux en termes d'énergie. Le profilage des ribosomes (Ribo-Seq) figure parmi les récentes et prometteuses technologies ayant permis une meilleure étude des mécanismes de régulation de la traduction. Ces résultats contiennent toutefois la présence de variabilité et de bruits de nature infondée. Ce travail présente la mise en place d'une stratégie permettant la dissociation de signaux d'origine biologique de ceux ayant une origine technique. Ceci est effectué au travers de la mise en place de profils consensus de densité ribosomale extrait d'une analyse comparative de plusieurs expériences de Ribo-Seq chez la levure (*Saccharomyces cerevisiae*). Les signaux biologiques dérivés par les profils consensus correspondent avec les signatures de pauses ribosomales connues, telles que les scores de repliements de l'ARNm et la charge des acides aminés. Épatamment, notre stratégie a également permis l'identification de séquences différentiellement transcrites (DT). Ces dernières jouent un rôle sur la cinétique de la phase d'élongation de la traduction, elles comportent notamment une surreprésentation de codons associés aux modifications des ARNs de transfert (tRNAs). Elles se retrouvent d'ailleurs impliquées dans le maintien de l'homéostasie cellulaire, ayant une présence marquée chez des gènes prenants part aux mécanismes de biosynthèse de la macromolécule ribosomale ainsi que chez les ARNms aux sublocalisations cellulaires précises, notamment chez les mitochondries et le réticulum endoplasmique (ER). En plus de démontrer les possibilités de découvertes offertes par la technique du Ribo-Seq, cette étude présente une évidence de la nature dynamique et hétérogène du processus de traduction chez la cellule eucaryote. Elle démontre également le rôle de l'information directement encodée dans la séquence dans l'optimisation générale de l'homéostasie cellulaire.

**Mots-clés :** homéostasie, protéostasie, expression génétique, ARNm, traduction, élongation, régulation, *Saccharomyces*, Ribo-Seq, ribosome, ARNt, séquence, eucaryote, hétérogénéité

## Abstract

Proteins are responsible for virtually all functions performed within and in the surroundings of a cell. The control of gene expression, which determines the amount, localisation and timing of protein production in the cell, is the central processes in the regulation of cellular physiology and function. Any disturbance in this complex system can generate important consequences on cellular integrity, sometimes leading to incurable diseases. The translation of messenger RNA into a protein product is the last step of the gene expression mechanism. It can be regulated in manifold ways, both intrinsically and extrinsically to the transcript sequence. It is also the costliest cellular process in terms of energy. Ribosome profiling (Ribo-Seq) is one of the recent and promising technologies making it possible to better study the mechanisms of translation regulation. Its results have however been shown to display variability in reproducibility and to contain noise of uncharted sources. This work presents the implementation of a strategy for dissociating signals of biological origin from those of technical origin. This is performed by the computation of a consensus profile of ribosomal density derived from a comparative analysis of several Ribo-Seq experiments in yeast (*Saccharomyces cerevisiae*). The biological signals derived by the consensus profiles correspond with signatures of known ribosomal pauses, such as mRNA folding strength and amino acid charges. Amazingly, our strategy also enabled the identification of differentially transcribed (DT) sequences. The latter have shown an over-representation of codons associated with modifications of transfer RNAs (tRNAs). They are also involved in the control of cellular homeostasis, exhibiting a marked presence in genes involved in ribosome biosynthesis as well as in mRNAs with precise translation sub-localization, particularly in mitochondria and the endoplasmic reticulum (ER). In addition to demonstrating the possibilities of discovery offered by the Ribo-Seq technique, this study also presents evidence of the dynamic and heterogeneous nature of the translation process in the eukaryotic cell. It also showcases its diverse regulatory mechanisms and the role of information directly encoded in the sequence in the general optimization of cellular homeostasis.

**Keywords :** homeostasis, proteostasis, gene expression, mRNA, translation, elongation, regulation, *Saccharomyces cerevisiae*, Ribo-Seq, ribosome, tRNA, sequence, eukaryotic, heterogeneity

## Table des matières

Résumé.....	i
Abstract .....	ii
Table des matières.....	iii
Liste des figures .....	vi
Liste des tableaux.....	vii
Liste des sigles .....	viii
Liste des abréviations.....	x
Remerciements.....	xii
1. Introduction.....	1
1.1 L'expression génique .....	1
1.1.1 L'importance fonctionnelle de l'expression des gènes.....	1
1.1.2 La complexité de la régulation de l'expression génique .....	2
1.2 L'étape de la traduction.....	3
1.2.1 L'importance de l'étape de traduction.....	3
1.2.2 Les étapes de la traduction.....	5
1.2.3 Le contrôle de l'élongation de la traduction.....	5
1.2.4 Les modifications des ARNs de transfert .....	8
1.2.5 Le repliement protéique co-translationel.....	11
1.2.6 Les ribosomes spécialisés.....	13
1.3 Approches expérimentales et computationnelles à l'étude de la régulation de la traduction à haut débit.....	15
1.3.1 Le séquençage de nouvelle génération et ses dérivés (NGS).....	15
1.3.2 Le séquençage du transcriptome (RNA-Seq).....	16
1.3.3 Le profilage de ribosome (Ribo-Seq).....	18

1.3.4 L'analyse de signaux et la détection de pics.....	24
1.3.5 L'annotation fonctionnelle des gènes.....	26
1.4 Hypothèse et objectif.....	28
2. Article .....	29
Inferring translational heterogeneity from ribosome profiling data.....	29
Contribution of authors .....	29
Keywords .....	29
2.1 Abstract .....	30
2.2 Introduction .....	30
2.3 Results.....	33
2.3.1 Consensus RD profiles represent individual experiments.....	33
2.3.2 Consensus profiles are biologically meaningful.....	35
2.3.4 Differentially translated (DT) sequences are rare.....	37
2.3.5 DT sequences may be important for protein biogenesis.....	40
2.3.6 No systematic role of DT sequences in protein folding .....	42
2.3.7 DT sequences link to protein and organelle biogenesis .....	44
2.4 Discussion .....	48
2.5 Materials and Methods.....	51
2.5.1 Code and data availability .....	51
2.5.2 Data sources and processing.....	51
2.5.3 Consensus RD profiles .....	52
2.5.4 Validation of consensus profiles.....	53
2.5.5 Identification of differentially translated (DT) sequences.....	53
2.5.6 Analysis of DT sequences .....	54
2.6 Acknowledgments.....	54

3. Conclusions et perspectives .....	55
Références bibliographiques.....	58

## Liste des figures

Introduction:

**Figure 1.1:** Illustration de différents facteurs *cis* (affiché au-dessus du transcrit) et *trans* (au-dessous du transcrit) et leurs impacts sur la cinétique de la phase d'élongation de la traduction. Provenance : (Stein & Frydman, 2019).....7

**Figure 1.2:** Lieux des principales positions de modifications sur la structure de l'ARNt. Provenance: (Torres *et al.*, 2014).....9

**Figure 1.3:** Sources de différences structurales chez les ribosomes eucaryotes. Provenance: (Guo, 2018).....13

**Figure 1.4:** Illustration du principe du Ribo-Seq et la dérivation des DRs correspondantes. Provenance: (Ingolia, 2014).....19

Article:

**Figure 2.1:** Consensus ribosome density profiles.....34

**Figure 2.2:** Consensus profiles identify nonoptimal codons and stalling sequences.....36

**Figure 2.3:** Identification of differentially translated (DT) subsequences.....38

**Figure 2.4:** Characteristics of DT sequences.....41

**Figure 2.5:** DT sequences are not systematically coordinating protein folding on the ribosome.43

**Figure 2.6:** Gene Ontology (GO) analysis of DT sequences suggests DT sequences link to protein and organelle biosynthesis.....46



## Liste des tableaux

Article:

<b>Table 2.1:</b> Ribosome profiling datasets analyzed in this study.....	52
---	----

## Liste des sigles

ADN : Acide désoxyribonucléique (*DNA* en anglais)

ARN : Acide ribonucléique (*RNA* en anglais)

ARNm : ARN messenger (*mRNA* en anglais)

ARNr : ARN ribosomique (*rRNA* en anglais)

ARNt : ARN de transfert (*tRNA* en anglais)

cDNA : ADN complémentaire (*Complementary DNA* en anglais)

CDS : Région codante (*Coding sequence* en anglais)

CHX : Cycloheximide

DAG : Graphe orienté acyclique (*Directed acyclic graph* en anglais)

DT : Différentiellement traduite, soit différents RD pour une même séquence (*Differentially Translated* en anglais)

EF : Facteur d'élongation de la traduction (*Translation elongation factor* en anglais)

ER : Réticulum endoplasmique (*Endoplasmic reticulum* en anglais)

EST : Marqueur de séquence exprimée (*Expressed sequence tag* en anglais)

GO : Ontologie des gènes (*Gene ontology* en anglais)

GTP : Guanosine triphosphate

HSP : Protéine de choc thermique (*Heat shock protein* en anglais)

IF : Facteur d'initiation de la traduction (*Translation initiation factor* en anglais)

IRES : Sites internes d'entrée du ribosome (*Internal ribosome entry site* en anglais)

LSU : Grande sous-unité du ribosome (*Large ribosomal subunit* en anglais)

miARN : micro-ARN (*miRNA* en anglais)

ncRNA : ARN non codant (*non-coding RNA*)

NGS : Séquençage de nouvelle génération (*Next-Generation Sequencing* en anglais)

PCR : Réaction en chaîne par polymérase (*Polymerase chain reaction* en anglais)

RD : Densité ribosomale (*Ribosomal density* en anglais)

RNase : Ribonucléases

RP : Protéines ribosomales (*Ribosomal proteins* en anglais)

RPB : Facteur ribosomal de biosynthèse protéique (*ribosome-associated protein biogenesis factors* en anglais)

SAGE : L'analyse en série de l'expression des gènes (*Serial analysis of gene expression* en anglais)

SLA: Sclérose latérale amyotrophique (*ALS* en anglais)

TPM : Transcrits par millions (*Transcripts per million* en anglais)

TR : Facteur d'activation (*Trigger Factor* en anglais)

uORF : Cadre de lecture ouvert en amont (*Upstream open reading frame* en anglais)

UTR : Les régions non traduites, (*untranslated regions* en anglais)

## Liste des abréviations

*c. elegans* : *Caenorhabditis elegans*

Ribo-Seq : profilage des ribosomes (*Ribosome Sequencing* ou *Ribosome profiling* en anglais)

RNA-Seq : Next Generation RNA Sequencing

scRNA-seq: RNA-SEQ par Séquençage de cellule unique (*single-cell RNA-Seq* en anglais)

*S. coelicolor* : *Streptomyces coelicolor*

*S. cerevisiae* : *Saccharomyces cerevisiae*

*À l'humanité et à la science.*

## Remerciements

Je voudrais remercier mon directeur de maîtrise le Professeur Sebastian Pechmann, d'avoir cru en moi et m'avoir donné une opportunité, pour avoir partagé avec moi ses grandes connaissances scientifiques et sa façon intègre de voir le monde. Je le remercie également pour les conversations, les blagues et les nombreux apprentissages, autant sur le monde scientifique que sur mon propre développement professionnel.

Je remercie aussi Madame Élane Meunier pour son aide et support inusité, presque maternel au cours de ma maîtrise, mais aussi de mon baccalauréat.

Je remercie les professeurs que j'ai eu la chance de croiser au cours de ma maîtrise: Dr François Major, Dre Gertraud Burger, Dr Gerardo Ferbeyre et Dr Matthieu Willems pour m'avoir transmis une part de leurs connaissances.

Je remercie également tous mes collègues de laboratoire et du département que j'ai eu la chance de côtoyer et qui m'ont eux aussi beaucoup appris et supporter. Sans ordre particulier : Léonard Sauvé, Léah Pollet, Yasmine Draceni, Savandara Besse, Musa Ozboyaci, Mélissa Nelson, Nazli Kocatuğ, Amruta Sahoo, Christopher Kooistra, Vithuya Shanmugalingam, Matt Sarrasin, Louis Gendron, Shamim Hasan, Peniel Bustamante, Alexandra Léveillé-Kunst, Matus Valach et Paul Stet.

Je remercie également ma famille et mes amis proches pour leur support éternel.

Ce mémoire présente un travail destiné à mieux comprendre le mécanisme de la traduction, l'un des plus importants processus de la biologie cellulaire. Ce chapitre d'introduction présente une revue des connaissances entourant la traduction, et les divers mécanismes de régulation qui contribuent à sa complexité et sa flexibilité.

Il est suivi d'une deuxième partie présentant un article de recherche portant sur l'hétérogénéité de la traduction. Ce dernier relate les observations que nous avons pu effectuer en analysant des données résultantes de la technique de profilage de ribosome. Il démontre à la fois la puissance de la technique pour l'étude de ce mécanisme ainsi que les possibilités de découvertes encore nécessaires afin d'élucider le fonctionnement de la traduction.

## **1. Introduction**

Le fonctionnement, la vitalité ainsi que la spécialisation des cellules sont ultimement définis par la quantité et la fidélité de la biosynthèse protéique. Par conséquent, mieux comprendre et prédire les produits de ce mécanisme nous permet de mieux le contrôler, d'en anticiper les défaillances possibles et de pouvoir les corriger. Pour ce faire, il faut étudier et comprendre les multiples mécanismes qui régissent la régulation de l'expression génétique.

### **1.1 L'expression génique**

L'expression génétique est un mécanisme hautement complexe duquel dépend la survie cellulaire. Pour cette raison, elle est minutieusement régulée par une multitude de processus distincts qui agissent en unissons (Komili & Silver, 2008). Elle se divise en quatre étapes classiques; la transcription, la traduction, la dégradation des ARNm et la dégradation protéique (Schwanhäusser *et al.*, 2011). Ces dernières comportent une cascade d'événements et d'étapes intermittentes, à leurs tours contrôlés sur plusieurs niveaux, chacun assez élaborés pour être le sujet de nombreux domaines d'études.

#### **1.1.1 L'importance fonctionnelle de l'expression des gènes**

Les protéines jouent un rôle crucial dans le déroulement de pratiquement toutes les activités au sein de la cellule et dans ses environs. Elles agissent, par exemple, comme catalyseurs de réactions chimiques, prennent des rôles de structure et d'organisation cellulaire, constituent et catalysent les moteurs moléculaires permettant la mobilité et le transport, s'occupent de l'entretien cellulaire et de l'élimination de déchets, opèrent sur le conditionnement et l'isolement de l'ADN,

parmi d'autres (Voet & Voet, 2005). Elles se retrouvent également au sein de mécanismes d'importance capitale à la cellule, tels que la régulation de l'expression génétique, la signalisation cellulaire et réponse intracellulaire, la spécialisation cellulaire, les cycles de contrôle cellulaire et la mort cellulaire (Voet & Voet, 2005).

L'abondance, la localisation et les propriétés physiques des protéines, telles que la charge et la conformation tridimensionnelle, sont les principaux facteurs qui gouvernent leurs effets dans le milieu cellulaire (Berg *et al.*, 2002). Il est couramment bien établi qu'une défaillance de ses propriétés mène à des perturbations dans l'homéostasie des protéines, la protéase (Balch *et al.*, 2008). Essentielle au maintien de l'ordre, du développement et du cycle de vie cellulaire, la protéase assure également la capacité de la cellule à réagir optimalement à différentes sources de stress, tels que dans le cas d'infections par des organismes étrangers (Taylor *et al.*, 2014). Celles-ci sont fortement reliés à l'apparition de maladies graves telles que l'Alzheimer, le Parkinson ou la sclérose latérale amyotrophique (SLA), ainsi que dans le diabète de type 2 et le cancer (Cheng *et al.*, 2018; Chiti & Dobson, 2017; Hetz & Glimcher, 2011). Les mécanismes de régulation de l'expression des gènes forment un extraordinaire complexe d'interactions entre gènes, molécules d'ARN, protéines et autres composantes qui déterminent ultimement le moment et la localisation spécifiques de l'activation ainsi que de l'abondance de protéines ou de molécules d'ARN produites dans la cellule (García-Sánchez & Marqués-García, 2016). La vitalité cellulaire dépend donc d'une étroite régulation des nombreux aspects du processus de l'expression génique.

### **1.1.2 La complexité de la régulation de l'expression génique**

Comme décrit plus haut, l'expression génétique est un processus fondamental à la cellule, exigeant en ressources et extrêmement complexe. Pour cette raison, chacune de ses étapes est parsemée de mécanismes de régulation, couramment répertoriés en différentes sections; soit l'épigénétique; qui englobe la méthylation et autres modifications des histones, la transcription de l'ADN en ARN, les modifications post-transcriptionnelles; telles que l'épissage alternatif et l'action des miARNs, le transport des différents ARNs, la traduction localisée (Priesnitz & Becker, 2018), la régulation de la traduction de l'ARNm, les modifications post-traductionnelles ainsi que les processus de dégradation des ARNs et des produits protéiques (Schwanhäusser *et al.*, 2011) et plus récemment l'épitranscriptomique, soit le regroupement des modifications sur les nucléosides de l'ARN (Roundtree *et al.*, 2017). Chacune de ces étapes alimente d'elle-même de vastes



domaines d'études et s'avère critique à une meilleure compréhension du fonctionnement et du dérèglement cellulaire ainsi qu'à la prévention et au traitement des infirmités qui peuvent s'ensuivre. Cela dit, de récentes avancées technologiques apportent une nouvelle puissance et précision dans l'étude de ces mécanismes. Cela inclut des méthodes structurales telles que la cristallographie par rayons X, la spectroscopie NMR, la microscopie par *cryo-electron* (Cryo-EM) et des méthodes de mesures chimiques telles que le *Bulk Kinetics* et les principes de *Single-molecule*. L'apparition de méthodes basées sur le séquençage de nouvelle génération (NGS) (e.g. : le Ribo-Seq), la spectrométrie de masse ainsi que l'émergence de la puissance analytique des données par la bio-informatique apportent également d'importantes contributions (Choi *et al.*, 2018). De nouvelles études démontrent la forte disparité entre les niveaux d'ARNm et les taux protéiques observés chez la cellule. Effectivement, la taille d'un génome codant n'est pas représentative de la quantité de protéines qu'il peut générer, suggérant l'existence de différents mécanismes régulant l'expression des gènes. En guise d'exemple, le transcriptome humain est composé de plus de 80 000 transcrits codant pour des protéines et il est estimé qu'ils synthétisent de 250 000 à 1 million de protéines différentes (Klerk & Hoen, 2015). Cela provoque donc de l'intérêt croissant pour l'étude la régulation au niveau de l'étape de la traduction (Hanson & Collier, 2018).

## **1.2 L'étape de la traduction**

### **1.2.1 L'importance de l'étape de traduction**

Le processus de traduction est défini par la conversion des nucléotides de l'ARNm sous sa forme mature en une chaîne d'acides aminés concordants. Afin de pouvoir exercer leurs fonctions, ces polypeptides doivent atteindre leur forme native, soit en étant reliés et repliés en une structure tridimensionnelle et parfois même multicomplexée, maintenue ensemble au travers de liaisons peptidiques, hydrogènes et ioniques, de ponts disulfures et d'interactions hydrophiles et hydrophobes (Kim *et al.*, 2013). La traduction représente l'étape finale et irréversible du dogme central de la biologie moléculaire (Crick, 1970). Elle est un des processus le plus fondamental et moins bien compris du système de l'expression génétique. Le coût énergétique de la traduction pour la cellule, sa capacité de rapidement et habilement réguler l'homéostasie protéique ainsi que l'incompatible quantité de molécules d'ARNm comparativement à celle de leurs protéines

résultantes, témoignent de la présence d'une importante régulation de ce mécanisme et justifient l'intérêt pour son étude (Neuman, 2017).

La traduction représente le processus le plus demandant énergétiquement chez la cellule. En effet, il a été démontré qu'une majorité du guanosine triphosphate (GTP) cellulaire est utilisé pour la polymérisation des acides aminés (Russell & Cook, 1995). Cela est principalement dû au coût de production de la macromolécule ribosomale ainsi qu'à l'énorme quantité d'instances du processus de traduction se déroulant de façon simultanée dans le corps cellulaire (Granneman & Tollervey, 2007). La cellule eucaryote peut contenir jusqu'à  $10^8$  ribosomes. En plus de l'ARN et des composants protéiques du ribosome lui-même, plus de 200 protéines et ARN non codants supplémentaires participent à la série de réactions complexes de transformation, de modification, d'assemblage et de transport membranaire nécessaires à produire cette macromolécule (Lempiäinen & Shore, 2009). Les protéines ribosomales (RP) sont les polypeptides les plus exprimés dans la cellule. Les gènes de l'ARN ribosomal de *S. cerevisiae* représentent 10% du génome entier alors que la transcription de l'ARNr par l'ARN polymérase I (Pol I) représente près de 60% de la transcription totale dans la cellule (Warner, 1999). Cela dit, face à ce gargantuesque besoin en apport énergétique, la cellule requiert d'importants mécanismes de régulation afin de conserver le plus de ressources possible.

Par ailleurs, les molécules d'ARNms ont une demi-vie plus courte et sont moins nombreuses que les protéines, ces dernières étant en moyenne 5 fois plus stables et 2800 fois plus abondantes (Raghavan *et al.*, 2002; Schwanhäusser *et al.*, 2011). De plus, la corrélation entre les demi-vies de ses deux catégories de molécules n'est pas significative et la corrélation de leurs abondances est faible. Cela suggère que la cellule puisse adapter les niveaux de produit protéiques en temps réel par la rapide production de différents ARNms et d'ainsi modifier son patron d'expression génique de manière plus rapide et efficace en réponse à différents stimuli internes ou externes (Barbosa *et al.*, 2013). Il en va de même lors d'événements de développement ou de différenciation (Chekulaeva & Landthaler, 2016; Pechmann, 2018). Par exemple, la dépendance sur la traduction et l'utilisation de molécules d'ARNm et de protéines préstockées à défaut de performer de la transcription *de novo*, permettent de soutenir des phases de développement rapide (Susor *et al.*, 2015).

Finale­ment, les quan­ti­tés dispa­ra­tes d'ARNms et de pro­te­ines dans la cel­lule ain­si que la pa­no­plie de méca­nismes régu­lateurs de la tra­duc­tion qui ne cessent d'être dé­cou­verts té­moi­gnent de l'uti­lité et de l'im­por­tance cru­ciale de la régu­la­tion de ce pro­ces­sus (Maier *et al.*, 2009). Ceci per­met non seu­le­ment de main­te­nir l'homéostasie en pé­ri­ode de stress aléa­toire et/ou de plus haute in­ten­si­té, mais aus­si à la mise en place de chan­ge­ments plus per­ma­nents dans la phy­si­o­logie cel­lulaire. Comme tout méca­nisme cel­lulaire com­plexe et fon­da­men­tal, la tra­duc­tion se com­pose de plusie­urs éta­pes rigou­reusement régu­lées.

### **1.2.2 Les éta­pes de la tra­duc­tion**

La tra­duc­tion se di­vise en qua­tre prin­ci­pales éta­pes; l'ini­tiation, l'élongation, la ter­mi­na­ison et le re­cy­clage du ri­bo­some (Sonen­berg & Hinnebusch, 2009). L'ini­tiation est re­con­nue comme étant l'éta­pe li­mi­tante dans la régu­la­tion de la tra­duc­tion, no­ta­ment due à la com­ple­xi­té et la dis­poni­bi­lité des dif­fé­ren­tes com­po­san­tes né­ces­saire à l'as­sem­blage du com­plexe d'ini­tiation (Shah *et al.*, 2013). Ce der­nier re­qui­ert l'in­terac­tion d'au moins 11 fac­teurs de tran­scrip­tions dis­tin­cts, eux-mêmes su­jets à des forces régu­la­trices, telle que la sé­que­stra­tion de eIF4E par 4E-BPs ou en­core la phos­phorylation de eIF2 $\alpha$  (Richter & Col­ler, 2015). Elle peut éga­le­ment être con­trô­lée par de l'in­for­ma­tion pré­sen­te dans les ré­gions 5' UTR, soit en a­mont de l'ARNm, tel que la pré­sen­ce d'une coiffe en 5' et des sites IRES. Son pro­ces­sus peut, tout au­tant, être com­ple­xi­fié par la pré­sen­ce de struc­tures se­con­daire de l'ARNm ou de pro­te­ines en­vi­ron­nan­tes (Poulin & Sonen­berg, 2013).

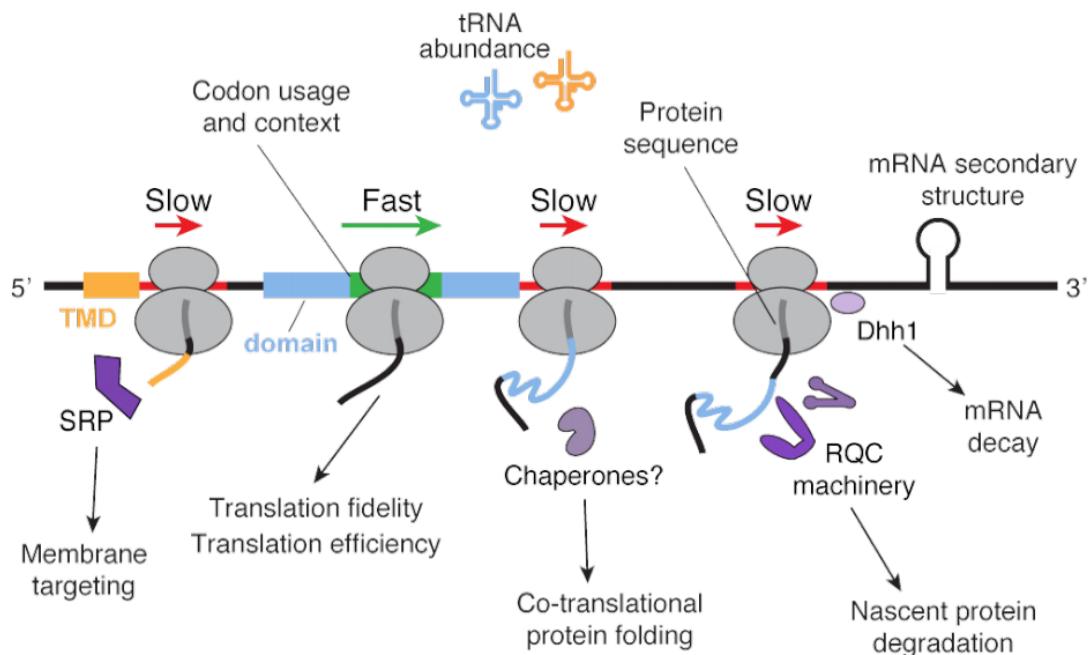
Mal­gré les nom­breuses ex­pé­ri­ences ay­ant dé­mon­tré le rôle de l'ini­tiation, comme l'éta­pe li­mi­tante de la tra­duc­tion, de nom­breuses études ré­centes, per­mises no­ta­ment par les progrès tech­no­logiques et ana­lytiques dans le do­maine, ont dé­mon­tré que l'élongation de la tra­duc­tion est éga­le­ment for­te­ment régu­lée et de fa­çon mul­ti­di­men­sion­nelle. Cer­taines études sont même allées jus­qu'à af­fir­mer que l'ef­fi­cacité de l'élongation se­rait en fait le meilleur pré­dic­teur des ni­veaux pro­te­iques dans la cel­lule (Richter & Col­ler, 2015).

### **1.2.3 Le con­trôle de l'élongation de la tra­duc­tion**

Inté­res­se­ment et de fa­çon con­traire à l'ini­tiation de la tra­duc­tion, presque toutes les éta­pes de l'élongation de la tra­duc­tion sont con­ser­vées à tra­vers les dif­fé­ren­tes espèces des do­maines du vi­vant, sug­gé­rant une im­por­tance cru­ciale de ce méca­nisme à la sur­vie cel­lulaire. Cela dit, l'élongation est rela­ti­ve­ment moins com­plexe que l'éta­pe d'ini­tiation, au­tant dans le nombre de

facteurs d'élongation (EFs) nécessaires à son fonctionnement que dans le temps nécessaire à l'assemblage de son complexe catalytique (Dever *et al.*, 2018; Richter & Coller, 2015). Toutefois, de nombreuses recherches ont démontré que contrairement à l'assomption classique, soit que la séquence de ARNm se traduit directement en séquence de polypeptides, l'ARNm ne prédit pas toujours la séquence ni la structure de sa protéine résultante (Choi *et al.*, 2018). Sachant cela, il est évident que la traduction soit affectée par de nombreux mécanismes de régulation et que ces derniers agissent pendant la phase d'élongation (Pechmann, 2018; Schwanhäusser *et al.*, 2011). Plusieurs facteurs exercent un contrôle sur l'élongation de la traduction. Ils peuvent être répartis en deux principaux groupes ; i.e. : extrinsèque (*trans*) ou intrinsèque (*cis*) à la séquence du transcrit d'ARNm (Figure 1.1) (Guo, 2018; Stein & Frydman, 2019).

Les mécanismes de type *trans* comprennent la modification des protéines, des facteurs d'élongation et des protéines ribosomales par méthylation, phosphorylation et ubiquitination (Jakobsson *et al.*, 2018; Matsuo *et al.*, 2017). Ils impliquent également les interactions de la chaîne polypeptidique naissante avec d'autres structures moléculaires environnantes telles que le tunnel de sortie du ribosome, traversant la LSU (Lu & Deutsch, 2008), les chaperons et autres facteurs d'associations ribosomales (Shalgi *et al.*, 2013; Stein *et al.*, 2019), le repliement co-translationnel des protéines (Pechmann & Frydman, 2013; Sharma & O'Brien, 2018), l'assemblage de complexes protéiques (Shiber *et al.*, 2018), les complexes de reconnaissance de signal (SRP) (Pechmann *et al.*, 2014; Richter & Coller, 2015) et même la dégradation co-translationnelle des protéines (Duttler *et al.*, 2013). De plus, l'entrave ou la dégradation des ARNms par la liaison d'ARN régulateurs non codants tels que les miARNs jouent également un rôle important dans la régulation de la traduction (Catalanotto *et al.*, 2016; Morris & Mattick, 2014). Plus récemment encore, la spécialisation des ribosomes et les ARNts modifiables, i.e. sujets à des modifications chimiques de certains nucléosides composants leur structure, par des enzymes spécifiques (El Yacoubi *et al.*, 2012), ont également été observés comme ayant un impact sur la cinétique de traduction (Guo, 2018; Xue & Barna, 2012).



**Figure 1.1:** Illustration de différents facteurs *cis* (affiché au-dessus du transcrit) et *trans* (au-dessous du transcrit) et leurs impacts sur la cinétique de la phase d'élongation de la traduction. Provenance : (Stein & Frydman, 2019).

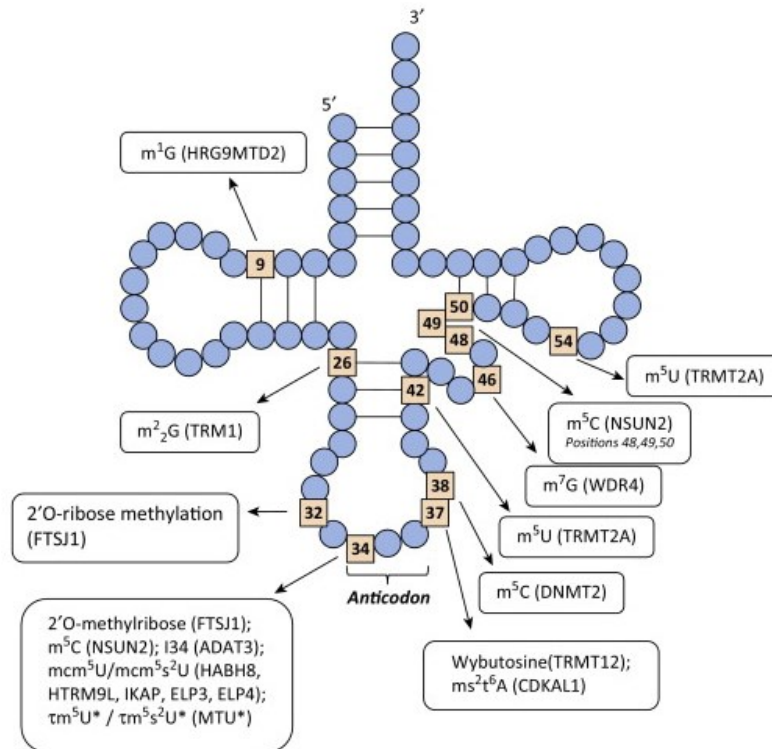
D'un autre côté, les mécanismes à action *cis* incluent la présence de paires de bases *wobble* (Wang et al., 2016), les variations dans la charge des acides aminés (Charneski & Hurst, 2013), les positions en tandem de codons structurellement complexes tels que la proline ou l'arginine (Sabi & Tuller, 2015) ainsi que la présence de structures secondaires d'ARNm dans la séquence du transcrit (Schuller & Green, 2018). Ils existent aussi différents signaux codés dans l'ARNm pouvant modifier la lecture du message qu'il contient en reprogrammant le ribosome, un phénomène appelé recodage traductionnel. Trois types de recodage ont été classifiés jusqu'à présent ; la lecture des codons stop, le décalage du cadre ribosomique et le contournement translationnel (Rodnina *et al.*, n.d.). Cependant, c'est principalement le choix particulier de codon qui s'avère être l'une des principales forces de régulation de la cinétique de la traduction. En effet, les codons présentent différentes propriétés d'optimalité, lesquelles sont liées, entre autres à la disponibilité de leur ARNt concordant dans l'environnement traductionnel. Même s'il est synonyme, le choix du codon ou le regroupement de codons rares ne serait pas fait de manière aléatoire, mais plutôt issue de pressions évolutives, offrant à la cellule un élégant mécanisme additionnel pour réguler la vitesse de la traduction et par le fait même contribuant à l'optimisation

des ressources de la cellule (Chaney & Clark, 2015; Dana & Tuller, 2014; Negrutskii *et al.*, 2018; Torrent *et al.*, 2018). À cela s'ajoute également l'épitranscriptomique, soit les modifications possibles des nucléosides de différentes molécules d'ARN, rendant le processus de traduction d'autant plus flexible (Pereira *et al.*, 2018). Bien que la plupart de ces modifications ont, jusqu'à présent, principalement été identifiées chez les ARNs ribosomique et de transfert, certaines sont également présentes sur le transcrit de l'ARNm, par exemple la modification methyl-6-adenosine (m<sup>6</sup>A), observée chez une multitude d'organismes (Saletore *et al.*, 2012). La traduction est également exécutée dans des localisations distinctes et souvent spécifiques à un ARNm, de la cellule (Martin & Ephrussi, 2009). Ce système de traduction localisé de l'ARNm représente une couche additionnelle de régulation des gènes sous un contrôle temporel et spatial. Ce dernier réalisé par des signaux de la séquence (Giorgi & Moore, 2007), des protéines de ciblage et de transport (Paquin & Chartrand, 2008), les granules de transport d'ARN (Krichevsky & Kosik, 2001), le cytosquelette cellulaire (Reck-Peterson *et al.*, 2001) ainsi que les ribosomes spécialisés (Guo, 2018), parmi d'autres mécanismes peut-être encore inconnus. Il existe une panoplie d'exemples marquants de cela, notamment, la localisation des ARNms dans l'embryon lors du développement d'un organisme (Weil *et al.*, 2008), la plasticité synaptique, la croissance et la communication des neurones dans le cerveau qui dépendent de la traduction localisée d'ARNm sélectionnés dans les dendrites et axones (Krichevsky & Kosik, 2001). De plus, la localisation de l'ARNm est importante pour la traduction spécifique chez différentes organelles, par exemple la traduction dans le ER (Schmid *et al.*, 2006) pour l'exportation de protéines ou encore dans la mitochondrie (Shi *et al.*, 2017). En somme, le fait que notre compréhension des mécanismes régulant la traduction soit encore élémentaire, ouvre l'excitante opportunité de les découvrir et les étudier, le discernement des mécanismes de contrôle de la régulation traduction étant cruciale à la compréhension de la biosynthèse protéique.

#### **1.2.4 Les modifications des ARNs de transfert**

Les ARNs sont les principaux participants du mécanisme de traduction (ARNm, ARNt et ARNr). Des modifications de leurs constituants ont par conséquent un impact direct sur la régulation de la traduction. Les ARNs de transfert (ARNt) sont les molécules responsables de l'acheminement d'un acide aminé, qui leur est conféré par une molécule d'aminoacyl-ARNt synthétase, au complexe de traduction afin qu'il puisse contribuer au polypeptide naissant (Phizicky & Hopper, 2010). Ils exécutent cela par la liaison de l'anticodon dans leur structure à un

codon spécifique de l'ARNm. Ce sont de courtes molécules, repliées sur elles-mêmes, normalement longues de 70-100 nucléotides (Kanai, 2014). Elles sont présentes dans tous les organismes vivants et plusieurs d'entre elles sont conservées au travers de l'évolution dans tous les domaines de la vie (Chen *et al.*, 2010). Dans leur fabrication, les ARNts subissent plusieurs formes de régulations; lors de leur transcription, du traitement des transcrits, leur sous-localisation ainsi que des modifications de leur base ribonucléotidiques (Bednářová *et al.*, 2017). À ce jour, plus de 80 modifications de l'ARNt ont été signalées, dont certaines sont hautement conservées au travers de l'évolution (Björk *et al.*, 2001). Une molécule d'ARNt contient en moyenne 13 bases modifiées (Pan, 2018). Les bases de données *The tRNA Modification Database* (Crain & McCloskey, 1997) et *MODOMICS* (Boccaletto *et al.*, 2018) en détiennent un inventaire exhaustif et grandissant.



**Figure 1.2:** Lieux des principales positions de modifications sur la structure de l'ARNt.

Provenance: (Torres *et al.*, 2014)

Récemment les chercheurs ont pu davantage démontrer l'importance des ARN de transfert notamment dans la régulation de la traduction et le maintien de l'homéostasie cellulaire. Il est très

clair que les ARNs représentent une importante couche de contrôle de l'information génétique. (Bednářová *et al.*, 2017). Ils font partie du nouveau domaine de l'épitranscriptomique, dénomination attribuée aux modifications des ribonucléosides dans la cellule, soit par l'ajout de groupes méthyle et hydroxyle à des nucléosides des molécules d'ARN. Les modifications nucléotidiques chez l'ARNt affectent généralement la région de la boucle de l'anticodon (les bases 34-35-36), particulièrement les nucléosides en position *wobble* (base 34) et la base 37 (Figure 1.2). Ces positions sont en liens directs avec la sensibilité et la stabilité de l'interaction avec le codon correspondant (Pereira *et al.*, 2018). La base 34 est responsable de la position *wobble* et confère donc la possibilité à l'ARNt de décoder plus d'un codon, dit synonyme. La majorité des modifications aux ARNs connus sont apportées à cette base (Ranjan & Rodnina, 2016). Les modifications sur la position 37, quant à elles, aident à maintenir la stabilité de l'interaction codon-anticodon (Tuorto & Lyko, 2016). Ces deux modifications contribuent au maintien de l'efficacité et de la fidélité de la traduction, au respect du cadre de lecture ainsi qu'à l'équilibre de l'homéostasie protéique. Les modifications des ARNs sont également impliqués dans le modèle des changements de cinétique de translocation de la traduction selon le choix de codon de la séquence traduite. Un exemple de cela a été observé chez une souche de *C. elegans*. Déficiante en U34 2-thiolation, cette dernière présentait un ralentissement prononcé des ribosomes aux codons AAA, CAA et GAA. Ces mutants comportaient également un taux d'agrégation de protéines 2.5 fois plus élevé lorsque comparé aux contrôles, ainsi qu'une augmentation de l'expression des protéines de choc thermique (Hsps), ceci démontre la perturbation qu'une hypomodification de l'ARNt à la position 34 peut avoir sur la protéostase des nématodes (Nedialkova & Leidel, 2015). Comme référencé plus tôt dans ce travail, il en va de même chez les humains, où des déséquilibres de la protéostase sont souvent observés dans des maladies neurodégénératives telles que l'Alzheimer, le Parkinson ou le ALS, ainsi que dans le diabète de type 2 et le cancer (Chiti & Dobson, 2017).

Ces modifications peuvent aussi se produire sur d'autres positions du corps de l'ARN de transfert (Figure 1.2). Ces dernières servent de contrôle sur la structure de l'ARNt, elles affectent sa stabilité, flexibilité et capacité à se lier (Lorenz *et al.*, 2017). Les modifications de la pseudouridine stabilisent généralement des motifs structurels spécifiques et les dihydrouridines aident à maintenir la flexibilité de conformation (Väre *et al.*, 2017). Un exemple de cela est la modification du 2'O-ribosyl phosphate à la position 64 de l'ARNtMet de *S. cerevisiae*, qui sert à



la différenciation entre l'ARNtMet des complexes d'initiation ou d'élongation (Hopper, 2013). Dans l'ensemble, il existe une multitude de processus au sein de la cellule contribuant à l'hétérogénéité de la traduction, l'ARNt en fait partie et joue un rôle dans la variation de l'apport de la synthèse protéique en réponse à différents stimuli (Chan et al., 2012).

### **1.2.5 Le repliement protéique co-traductionnel**

Le repliement adéquat des protéines est un mécanisme fondamental à leur bon fonctionnement dans la cellule et est généralement exécuté de façon co-traductionnelle (Zhang & Ignatova, 2011). Un simple changement dans la conformation d'une structure protéique, soit une divergence à son état natif, l'empêcherait d'interagir proprement avec son ou ses substrat(s) correspondants. Pire encore, il les amènerait à s'agréger l'une à l'autre, impliquant ainsi de graves conséquences pour la vitalité cellulaire et celle d'un organisme. Plusieurs maladies, telles que l'Alzheimer et même certaines formes de cancer, sont dues à l'agrégat de produits protéiques (Valastyan & Lindquist, 2014). Maintenu de façon judicieuse malgré la complexité qu'elle représente, cette opération est réalisée par la cellule de manière presque parfaite et extrêmement rapide. Il est estimé que le temps nécessaire pour parcourir tout l'espace de conformations de structures protéiques possibles à partir de séquences d'ARNm prendrait plus que l'âge de l'univers actuel. De plus, plusieurs expériences ont montré que dans les cellules de mammifères, les domaines protéiques acquièrent leur état natif peu de temps après leur émergence du tunnel de sortie du ribosome (Yan Han *et al.*, 2012). Cela dit, le repliement protéique s'exécute en quelques microsecondes, de façon très spécifique et simultanée par millions d'exemplaires dans la cellule (Dill & MacCallum, 2012), il est donc clair que de nombreux mécanismes de contrôle y sont impliqués. Les scientifiques étudient le problème de prédiction de la structure protéique depuis presque 60 ans et ce dernier n'est toujours pas élucidé (Dill & MacCallum, 2012). Plusieurs recherches ont démontré que la structure et la fonction des protéines peuvent être affectées par un changement dans la cinétique de traduction et ne dépendent donc pas exclusivement des forces de la thermodynamique (Sharma & O'Brien, 2018). Il a également été démontré que la majorité des protéines sont repliées en même temps qu'elles sont traduites par le ribosome, soit de façon co-traductionnelle (Rodnina & Wintermeyer, 2016). La formation de complexes protéiques (Shiber *et al.*, 2018) ainsi que la dégradation des protéines (Duttler *et al.*, 2013) peuvent également être effectuées de façon cotraductionnelle. La protéine croissante, soit la chaîne polypeptidique émanant du tunnel de sortie du ribosome, se retrouve exposée au milieu cellulaire et en permet l'interaction

avec plusieurs autres composantes. Des exemples de cela sont les chaperons; enzymes qui catalysent le repliement protéique, les particules de reconnaissance de signaux (SRP); facilitant le déplacement de ses protéines naissantes à des localisations cellulaires spécifiques, ou encore des complexes de dégradation protéique qui en digèrent la chaîne naissante, de façon cotraductionnelle, par souci de régulation ou de maintiens de la protéostase, parmi d'autres (Richter & Collier, 2015).

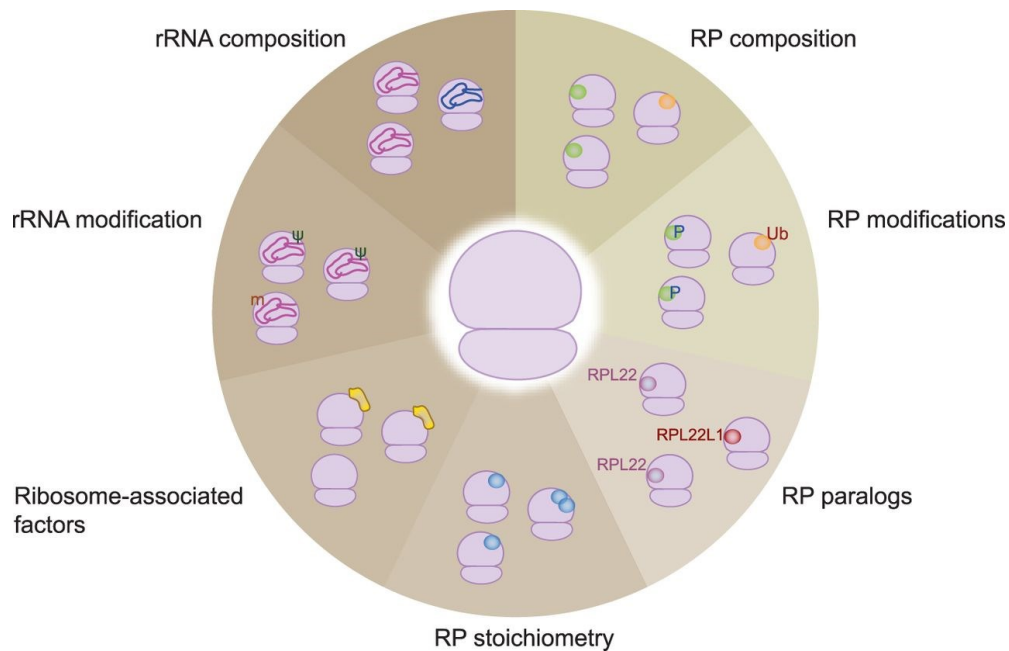
Cela dit, le couplage de la synthèse polypeptidique avec le repliement offre plusieurs avantages. Dans les protéines à plusieurs domaines, le repliement co-traductionnel facilite l'acquisition de l'état natif (Netzer & Hartl, 1997). Le repliement linéaire, avec la pointe N-terminale de la chaîne naissante en tête, facilite des trajectoires de repliement plus ordonnées, en particulier dans les protéines à domaines multiples (Zhang & Ignatova, 2011). Ensuite, le mouvement des ribosomes le long des cadres de lecture ouverts (ORF) n'est pas uniforme et comprend des sites de pause. L'incorporation de ses phases lentes peut permettre plus de temps pour le repliement du polypeptide, pour le ciblage membranaire ainsi que pour le traitement enzymatique. De plus, la zone de sortie du tunnel ribosomal agit également comme plateforme facilitant la liaison de plusieurs protéines interagissant avec la chaîne naissante, un exemple de cela est la RP L23 située à la sortie du tunnel et agissant comme site d'interaction avec le chaperon Trigger Factor (TR), reconnu pour se lier à la grande sous-unité ribosomale (LSU) 60S (Kramer *et al.*, 2002). Finalement, l'acquisition précoce des structures secondaire et tertiaire, parfois à l'intérieur même du tunnel ribosomal, mais principalement par les chaperons et autres mécanismes externes diminue le temps d'exposition de la chaîne naissante à l'environnement cellulaire et prévient ainsi un mauvais repliement de la structure protéique (Tu & Deutsch, 2010).

De plus en plus de sources pointent au ribosome comme chef d'orchestre du repliement protéique co-traductionnel. Tout d'abord, la chaîne polypeptidique naissante peut prendre des conformations de structures secondaires lors de son passage à travers le tunnel de sortie du ribosome. Ce dernier peut accommoder des chaînes d'environ 30-35 acides aminés si complètement étendues (Gloge *et al.*, 2014). De plus, les composantes ribosomales situées à l'embouchure du tunnel interagissent avec la chaîne naissante ainsi qu'avec le milieu cellulaire. Ces « zones de repliement » servent de port d'amarrages à des structures telles que les facteurs ribosomaux de biosynthèse protéique (RPBs) ou les chaperons. Le ribosome influence également

le repliement protéique par sa cinétique de translocation (Gloge *et al.*, 2014; Rodnina & Wintermeyer, 2016).

### 1.2.6 Les ribosomes spécialisés

Les ribosomes ont couramment été considérés des structures statiques et immuables. Toutefois, dans les dernières années, cette vision des choses s'est rapidement transformée et est présentement considérée quasi obsolète (Guo, 2018; Xue & Barna, 2012). Les ribosomes sont des structures complexes constituées de plusieurs ribonucléoprotéines et ARNr, composant deux sous-unités principales (Khatter *et al.*, 2015; Wilson & Cate, 2012). Leur synthèse et leur assemblage sont très coûteux en énergie et entièrement effectués dans le noyau cellulaire, précédant leur exportation vers le cytoplasme. Il est récemment devenu clair que les ribosomes présentent des différences dans leur structure (Guo, 2018). L'hétérogénéité de ces dernières est responsable de leur spécialisation et est causée autant par des modifications et/ou des différences dans les séquences des ARNr et des RPs qui les composent (Figure 1.3).



**Figure 1.3:** Sources de différences structurales chez les ribosomes eucaryotes.  
Provenance: (Guo, 2018)

Le contenu en protéines ribosomales (RP) peut différer de diverses manières. Leurs protéines constitutives peuvent avoir des paralogues ainsi que des types alternatifs. Les protéines du ribosome peuvent aussi subir des modifications telles que l'acétylation, la phosphorylation ou

la méthylation (Guo, 2018). Il existe également des types alternatifs d'une même molécule d'ARN ribosomique, par exemple, le génome de *S. coelicolor* abrite six copies de gènes d'ARNr codants pour cinq différentes LSUs (Kim et al., 2008). Les ARN ribosomiques peuvent également être modifiés comme les RPs, par exemple par 2-hydroxyméthylation ou par la conversion de l'uridine en pseudouridine (Sauert *et al.*, 2015). Ces modifications sont d'ailleurs souvent présentes dans des régions au fonctionnement important, e.g. : le centre de la peptidyltransférase, les sites ribosomiques A, P et E, le tunnel de sortie du ribosome et les sites d'interaction entre sous-unités (Decatur & Fournier, 2002). Cela suggère que de nombreuses fonctions ribosomales, incluant la traduction, sont sujettes à des modifications nucléotidiques, dont les plus courantes sont la méthylation 2'-O (du ribose) et la pseudouridylation (isomérisation de l'uridine en pseudouridine) (Decatur & Fournier, 2002). Ces modifications sont très abondantes et souvent guidées par de petits ARN nucléolaires (snoRNA) (Lowe & Eddy, 1999). Intéressamment, la majorité des modifications chez les ARNr sont répertoriées dans des régions considérées comme conservées, ce qui vient fortement appuyer l'importance de leur rôle dans la fonction structurale ainsi que dans le contrôle de la traduction (Natchiar *et al.*, 2017).

De plus, les éléments régulateurs *cis*, tels que les sites d'entrée internes des ribosomes (IRES) ou les cadres de lecture ouverts en amont (uORF), ont également démontré une influence sur la régulation par leurs interactions avec des éléments ribosomiques hétérogènes, favorisant ou non leur recrutement (Hui & Boer, 1987). En effet, il a été démontré que des ribosomes intrinsèquement distincts sont liés à la traduction de sous-groupes spécifiques d'ARNm, suggérant une régulation de la localisation subcellulaire de la traduction (Shi *et al.*, 2017), rendant par exemple, la traduction de protéines destinées à l'exportation vers les mitochondries plus efficace. Aussi, certaines protéines ciblées pour le réticulum endoplasmique (ER), seraient préférentiellement traduites par des ribosomes spécifiquement compétents pour cette localisation (Chartron *et al.*, 2016).

En outre, tous ces processus régulateurs possèdent en commun une capacité à contrôler la cinétique de la translocation des ribosomes. En effet, ils affectent la vitesse et le rythme de déplacement des ribosomes le long des transcrits d'ARNm, ce qui a pour effet de contrôler directement les niveaux de produits protéiques (Pechmann, 2018). L'avènement de la technique du profilage ribosomique, ou Ribo-Seq, permet désormais aux chercheurs d'observer ces changements

dans la cinétique de la traduction à un niveau de résolution nucléotidique novateur et perpétue ainsi l'espoir de mieux comprendre ces mécanismes (Ingolia *et al.*, 2009, 2012).

### **1.3 Approches expérimentales et computationnelles à l'étude de la régulation de la traduction à haut débit**

#### **1.3.1 Le séquençage de nouvelle génération et ses dérivés (NGS)**

Le séquençage de nouvelle génération (NGS) est une des plus importantes inventions modernes en biotechnologie. Bien qu'il soit performé par différentes plateformes : Roche, Illumina et SOLiD, parmi d'autres, ces dernières possèdent une base méthodologique commune : le séquençage massivement en parallèle de molécules d'ADN amplifiées de façon clonale. Cette conception expérimentale est un changement drastique de celle du séquençage de Sanger, lequel est basé sur la séparation électrophorétique des produits de terminaison de chaînes issus des réactions de séquençage individuelles avec du clonage bactérien. Cette dernière méthode fut utilisée pour le premier séquençage du génome humain, le *Human Genome Project* complété en 2013 a demandé 13 ans de travail et 2.7 milliards de dollars américains. Cela dit, le NGS permet un séquençage de l'ADN beaucoup plus rapide et à des coûts très réduits par rapport à son prédécesseur (Grada, 2013; Voelkerding *et al.*, 2009). De nos jours, le séquençage d'un génome humain prend seulement quelques jours et ne coute que quelques milliers de dollars, il est également bien plus précis (Mardis, 2011), permettant ainsi d'innombrables avancées courantes dans tous les domaines de la biologie. Le plus gros défi actuel réside dans les techniques de manipulation, de stockage et d'analyses de ces énormes quantités de données générées. Des outils bio-informatiques avancés sont essentiels pour une application réussie de la technologie NGS. Ces connaissances sont, encouragement, acquises à des rythmes fulgurants. Le NGS permet le séquençage de l'information génétique complète de plusieurs espèces ainsi qu'une bien meilleure caractérisation de chacun de leurs génomes. Il permet également une meilleure compréhension de la biologie moléculaire dans son ensemble, notamment au travers de plusieurs de ses variantes, dont le RNA-Seq. Il est indéniable que les technologies NGS auront un impact révolutionnaire sur la recherche en génomique ainsi que sur l'ensemble du domaine biologique pour les années à venir (Zhang *et al.*, 2011).

### 1.3.2 Le séquençage du transcriptome (RNA-Seq)

Depuis 1980, la méthode de micropuces à ADN (microarray), impliquant l'hybridation de transcrits d'ARNm à des sondes de cDNA présynthétisées et marquées par biofluorescence (Govindarajan *et al.*, 2012), était utilisée pour étudier le contenu et l'abondance en transcrits d'une cellule. En revanche, cette technique comportait plusieurs désavantages qui ont poussé le milieu scientifique à opter progressivement pour des méthodes basées sur le séquençage. D'abord le séquençage de type Sanger de marqueurs de séquences exprimées, ou *expressed sequence tag* (EST) (Adams *et al.*, 1991). Toutefois, jugé trop bas débit et pas assez apte quantitativement, il fut suivi des méthodologies avec tag, soit l'analyse en série de l'expression des gènes (SAGE) (Velculescu *et al.*, 1995), l'analyse avec cap (CAGE) (Kodzius *et al.*, 2006) et le séquençage par signature massivement en parallèle (MPSS) (Reinartz *et al.*, 2002). Les désavantages du *microarray* sont principalement la forte présence d'artéfacts dus à une importante présence de bruit de fond, une incapacité à détecter des gènes faiblement ou très fortement exprimés en raison d'une limite sur les quantifications et principalement à la nécessité de connaître les séquences d'intérêt *à priori*. Les désavantages des méthodes avec séquençage sont principalement le prix et la complexité dans l'analyse des résultats (Han *et al.*, 2015). Cependant, avec l'arrivée du séquençage du NGS, ces limitations représentent progressivement moins un problème, notamment pour les raisons décrites dans le paragraphe précédent. Les techniques mentionnées plus haut ont graduellement été remplacées, en faveur du RNA-Seq.

Le RNA-Seq fait usage des apports technologies du NGS afin de caractériser l'entièreté des ARNs présents dans un échantillon de cellules représentant une certaine condition. Il en permet l'estimation du niveau d'expression ainsi que la découverte de nouvelles structures telles que les ARN non codants (ncRNAs) et les miARNs. Il assiste également à la caractérisation de composantes du génome, telles que les régions d'épissage alternatif ou encore les UTRs (Wang *et al.*, 2009). Il permet notamment de cataloguer tous les types de transcrits, y compris les ARNm, les ARN non codants et les petits ARNs, de déterminer la structure transcriptionnelle des gènes selon leurs sites de départ, d'explorer les extrémités 5' et 3' de la séquence, de générer des modèles d'épissage alternatif et autres modifications post-transcriptionnelles ainsi que de quantifier les changements dans les niveaux d'expression génétique de chaque transcrit lorsque sous différentes conditions cellulaires.

La méthode générale du séquençage d'ARN consiste en l'isolement et la construction d'une librairie d'ARN d'intérêt pour un groupe de cellules exprimant une condition spécifique. Les transcrits d'ARNm sont alors convertis en cDNA, ces derniers ayant une structure plus stable. Suite à cela, le protocole est sensiblement le même pour toutes les techniques utilisant le NGS; les molécules sont flanquées par des adaptateurs sur les deux extrémités, amplifiées par réaction en chaîne par polymérase (PCR) et hybridées sur une *flow cell* afin d'être séquencées (Wang et al., 2009). Le séquençage peut se faire à partir d'une seule extrémité (*single-end*) ou des deux (*pair-end*). Les fragments séquencés seront alors alignés à un génome de référence ou utilisés pour l'assemblage d'un nouveau transcriptome ou pour la quantification de leur expression. Le *pair-end* permet un meilleur assemblage du transcriptome, surtout lors d'analyses *de-novo* (Corley et al., 2017).

Le RNA-Seq a donné naissance à une panoplie de méthodes, non seulement pour chaque type d'ARN distinct nécessitant son propre protocole (rARN, tRNA, snRNA, snoRNA, scaRNA, etc.), mais aussi pour les différentes composantes interagissant avec les ARNm; (CHIP-Seq, Ribo-Seq, TL-Seq, FAIRE-Seq etc.) (Conesa et al., 2016; Hrdlickova et al., 2017). Les avancées technologiques ont notamment permis le séquençage de fragments plus long, e.g., 250–300 paires de bases chez MiSeq de Illumina, de plus amples profondeurs pour le même coût ainsi qu'une quantité plus basse de matériels requis. Cela dit, d'autres technologies de tierce partie telles que *Pacific Biosciences* (PacBio) ou *Oxford Nanopore Technologies* (ONT), permettent le séquençage de molécules singulières d'ADN ainsi que de fragments beaucoup plus longs, allant jusqu'à la longueur complète d'un transcrit, et ce pour l'entièreté du transcriptome (Besser et al., 2018; B. Wang et al., 2016). Le séquençage de molécules isolées singulièrement offre l'avantage de sauter l'étape d'amplification par PCR, ce qui permet notamment de réduire le biais de duplication causé par cette dernière (Laver et al., 2015; Ross et al., 2013). De plus, la possibilité de séquencer des fragments plus longs permet de réduire l'ambiguïté de l'alignement au génome, diminuer les difficultés d'assemblage, d'identification ainsi que de quantification d'isoformes (203-205). Une autre technologie en pleine émergence est le séquençage *single-cell* (scRNA-seq). Comme son nom l'indique, cette dernière rend possible le séquençage de l'ARN contenu une cellule seulement. Cela permet d'obtenir des résultats plus spécifiques et avec moins de bruits que le RNA-Seq traditionnel. Son principal défi demeure toutefois la quantité très limitée d'ARN disponible dans une cellule singulière (Kolodziejczyk et al., 2015). En somme, la multiplicité de méthodes

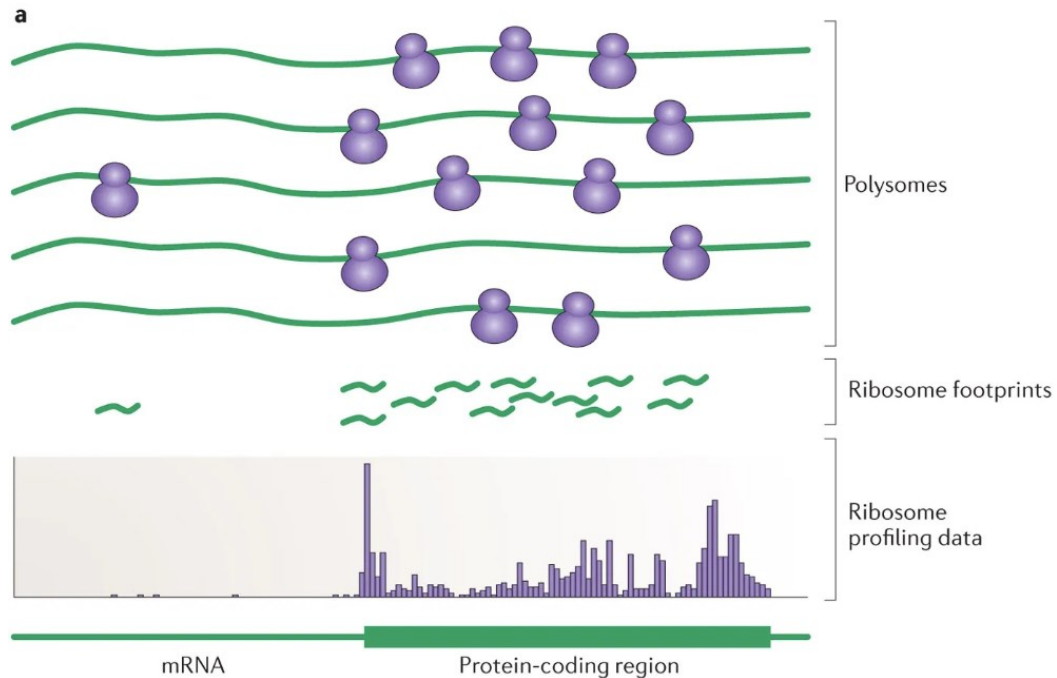
émergentes de RNA-Seq permet à la communauté scientifique d'étudier le transcriptome sous différents angles. Dans le cas de l'élongation de la traduction, la technique du Ribo-Seq y apporte une profondeur de résolution nouvelle et prometteuse.

### 1.3.3 Le profilage de ribosome (Ribo-Seq)

La technique de profilage des ribosomes est une approche expérimentale émergente développée par Jonathan Weissman et Nicholas Ingolia (Ingolia *et al.*, 2009). Une variante du RNA-Seq, le procédé consiste en l'interruption à l'échelle cellulaire de la traduction et du séquençage à haut-débit des fragments d'ARNm recouverts par la structure ribosomale exclusivement (Figure 1.3). Par la suite, l'alignement des fragments résultants au génome permet l'observation et l'analyse subséquente de régions de hautes densités ribosomales (RD) au nucléotide près (MGlincy & Ingolia, 2017). En considérant les régions CDS, les signaux RD représentent ultimement une image de la cinétique de la traduction en temps réel dans la cellule.

En principe, la traduction est interrompue au moyen de divers composés chimiques pouvant inhiber la translocation du ribosome. Plus récemment, ces derniers se sont fait remplacer par la congélation rapide dans l'azote liquide (*flash-freezing*), notamment dû aux distorsions que leurs utilisations engendraient dans les résultats (Ingolia, 2016). Par la suite, les fragments d'ARNm recouverts par les structures ribosomales sont digérés par des ribonucléases (RNase), dissociés des structures RP environnantes et séquencés à haut débit. Les fragments de séquences résultants sont ensuite alignés à la séquence génomique cellulaire annotée (Ingolia *et al.*, 2009). Cela permet d'obtenir des données de RD pour les différentes structures du génome, les séquences codantes et les UTRs environnants (Ingolia, 2016), fournissant ainsi des informations sur la localisation ribosomale en temps réel et sa quantification à l'échelle du transcriptome.





**Figure 1.4:** Illustration du principe du Ribo-Seq et la dérivation des DRs correspondantes. Provenance de (Ingolia, 2014).

Par conséquent, en plus d'une meilleure représentation de la synthèse protéique par la quantification des transcrits étant activement traduits, ces données présentent de l'information unique et précise sur la position du ribosome sur la séquence d'ARNm (Calviello *et al.*, 2016). Le Ribo-Seq a conduit à de nombreuses découvertes de mécanismes fondamentaux qui interagissent avec la cinétique de la traduction (Brar & Weissman, 2015). Des exemples de cela sont l'identification des sites d'initiation non-AUG en amont chez les eucaryotes (Lee *et al.*, 2012), la régulation spécifique de la réponse au stress au niveau de la traduction (Andreev *et al.*, 2015), l'identification des ribosomes impliqués dans des processus cellulaires distincts pour la synthèse localisée des protéines chez la levure (Jan *et al.*, 2014) ainsi que le repliement et l'assemblage de complexes protéiques concomitants à la traduction (Oh *et al.*, 2011; Shiber *et al.*, 2018). Les résultats de Ribo-Seq ont d'ailleurs démontré des corrélations plus élevées avec les résultats de spectrométrie de masse, soit la mesure même du contenu en protéine d'une cellule, que l'ARN-Seq classique (Weinberg *et al.*, 2016), témoignant de la capacité de la technique à fournir de l'information sur cette étape de la traduction.

Malgré cela, de nombreux aspects de la traduction demeurent mal compris (Bartholomäus *et al.*, 2015; Ingolia *et al.*, 2009). Cela est dû en partie à la complexité des systèmes biologiques

ainsi qu'à l'évidente nécessité de découvrir de nouvelles composantes de ce mécanisme. Le raffinement des technologies courantes demeure également un important défi. En effet, les données Ribo-Seq, sont caractérisées par une forte variabilité et une basse reproductibilité des résultats (Diament & Tuller, 2016). Il a été observé que les signaux de RD des codons individuels à partir des données de la technique sont représentés par des distributions amples et évanescentes plutôt qu'étroites et distinguées (Gamble *et al.*, 2016; Ingolia *et al.*, 2012). Cela rend difficile la modélisation précise des mécanismes de la traduction, principalement à des résolutions nucléotidiques (Matsuo *et al.*, 2017). Une part de l'incertitude dans les données peut être attribuée aux difficultés techniques ainsi qu'aux biais expérimentaux du Ribo-Seq. Parmi ces incertitudes figurent la difficulté du protocole expérimental, la technique d'interruption de la traduction, la digestion de l'ARNm, la disjonction de l'ARNr, la construction et l'annotation du génome, les biais dans l'alignement des séquences et le choix des gènes à inclure dans l'analyse (Hanson & Collier, 2018; Jan *et al.*, 2014; Oh *et al.*, 2011; Weinberg *et al.*, 2016). Il est clair que le bruit technique ne peut à lui seul expliquer toute la variabilité observée dans les données résultantes. Il est donc important de mieux saisir ces limitations afin de pouvoir recueillir en bout de compte des signaux illustrant des connaissances biologiques.

### **1.3.3.1 Les défis expérimentaux du Ribo-Seq**

D'énormes quantités de données ont déjà été générées à l'aide de la technique du Ribo-Seq, cependant de multiples sources d'artefacts ont été identifiées avec le temps (Hussmann *et al.*, 2015). Parmi ceux-ci, l'utilisation de composés inhibiteurs de la traduction, provoquant une distorsion des résultats expérimentaux (Santos *et al.*, 2019). La molécule de cycloheximide (CHX) en est la plus étudiée dû à sa préminente présence dans les premières années d'utilisation de la technologie (Schneider-Poetsch *et al.*, 2010). Effectivement, la CHX ne peut inhiber l'initiation ni la terminaison de la traduction, causant une accumulation des ribosomes au codon de départ (Ingolia *et al.*, 2011). De plus, sous son effet, certains ribosomes continuent leur translocation de façon plus lente, sans s'arrêter immédiatement (Gerashchenko & Gladyshev, 2014). Cela provoque des décalages et, conséquemment, une divergence dans les résultats. Une autre source importante d'artefact expérimental est le choix de la ribonucléase (RNase) (Gerashchenko & Gladyshev, 2017). L'objectif de celle-ci est la digestion précise du fragment résident sous les corps actifs du ribosome. Cela dit, il a été démontré que le choix de la ribonucléase comporte deux principaux risques, soit la digestion de la structure ribosomale, ainsi que la digestion incomplète de l'ARNm

flanquant le fragment sous le complexe de traduction (Gerashchenko & Gladyshev, 2017). Gerashchenko et Gladyshev ont astucieusement observé que les effets des RNase diffèrent selon l'organisme. Par exemple, la RNase I est la nucléase de choix lors des expériences employant *S. cerevisiae* comme organisme modèle. Il a été démontré que cela diffère entre espèces, par exemple, les ribosomes de la drosophile sont facilement dégradables par la RNase I (Gerashchenko & Gladyshev, 2017). Les chercheurs ont également observé que certaines nucléases introduisent des biais d'affinité avec des nucléotides particuliers pour les sites de digestion, générant ainsi chez certaines espèces différentes longueurs de fragments, contribuant au bruit des résultats du Ribo-Seq (Gerashchenko & Gladyshev, 2017; MGlinicy & Ingolia, 2017). Le choix de méthode pour la purification des monosomes du lysat digéré, soit par leur densité (*sucrose gradients*, *sucrose cushions*) ou leur taille (chromatographie d'exclusion stérique) (MGlinicy & Ingolia, 2017) ainsi que pour la méthode de déplétion des ARNr précédant le séquençage (Weinberg *et al.*, 2016), peuvent contribuer aux variations dans les données. À cela il faut également ajouter les erreurs par parallaxe, soit la marge d'erreur dans la performance humaine lors de l'exécution d'un protocole expérimental. La biologie se révèle bien sûr être très complexe et d'infimes changements dans les milieux cellulaires peuvent affecter les résultats d'une méthode aussi précise et élaborée que le Ribo-Seq. La promesse de cette dernière est claire et puissante, mais nécessite plus de raffinement, la liste des biais et artefacts potentiels se révèle en effet être assez substantielle (Bartholomäus *et al.*, 2015; Diamant & Tuller, 2016).

### **1.3.3.2 Les défis analytiques du Ribo-Seq**

De nombreux logiciels de bio-informatique ont été publiés, espérant générer une analyse plus intégrée des résultats de Ribo-Seq (Carja *et al.*, 2017; Chung *et al.*, 2015; Clauwaert *et al.*, 2018; Husmann *et al.*, 2015; Legendre *et al.*, 2015; MGlinicy & Ingolia, 2017; Michel *et al.*, 2016). Deux principaux défis affectent le côté analytique des données qui en résultent.

Le premier est l'identification de la position exacte du codon présent au site A sur un fragment de transcrite. Il a été démontré que le site A de la structure du ribosome, étant le site du déroulement de la liaison peptidique, est un des principaux facteurs limitants de la vitesse de traduction (Michel & Baranov, 2013). Déterminer sa localisation exacte sur un transcrite permettrait de réduire le bruit dans les données de Ribo-Seq. Cela aiderait également à identifier si un fragment séquencé fait partie ou non du cadre de lecture canonique (Fang *et al.*, 2018). De nombreuses

études ont tenté de déduire son emplacement. La majorité d'entre elles ont choisi d'assumer que le site A se retrouve à précisément 15 nucléotides en aval de l'extrémité 5' du fragment d'ARNm séquencé, issu du protocole expérimental d'origine du Ribo-Seq (Ingolia *et al.*, 2009). Ce dernier étant basé sur les recherches de (Wolin & Walter, 1988), démontre que le site A serait situé sur le 15<sup>ème</sup> nucléotide sur les fragments de longueur approximative de 28 nucléotides, produits lors de la digestion par l'ARNase I. Des études ont cependant observé que les différentes sous-structures constituant le ribosome oscillent constamment lors de sa translocation, provoquant une variation des longueurs de fragments protégés par le ribosome (Lareau *et al.*, 2014). Comme mentionné plus haut, la taille de l'empreinte peut également varier en fonction de l'organisme étudié, du choix de la nucléase et des conditions de lyse cellulaire (Hussmann *et al.*, 2015). Pour contourner cela, certaines études ont aligné des séquences de différentes longueurs sur des positions reconnues comme étant celles du site P au moment de l'initiation de la traduction, soit le codon ATG du début d'une CDS. Cette approche est encore limitée par la complexité multifactorielle de la longueur de fragments de transcrits (Martens *et al.*, 2015). Ce n'est que très récemment que des outils bio-informatiques, tentant de fournir une solution à cette problématique, ont été publiés. Par exemple, *RiboDeblur29* normalise les données de RD afin d'obtenir des signaux de position plus homogènes afin de mieux pouvoir évaluer les positions du site A (Wang *et al.*, 2018). Alternativement, le logiciel *Scikit-ribo* utilise une approche d'apprentissage supervisé pour former un modèle d'identification du codon de site A dans les données Ribo-Seq à l'aide des fragments contenant la position du codon de méthionine d'initiation de la traduction. Il utilise un modèle de *random forest* pour évaluer huit facteurs différents d'alignement sur le génome, soit la longueur du fragment, la distance entre l'extrémité 5 'ou 3' de la séquence et le codon de départ, et les nucléotides flanquant les extrémités du transcrit (Fang *et al.*, 2018).

Le deuxième défi analytique majeur d'interprétation des données de Ribo-Seq consiste à prendre en compte l'incertitude expérimentale lors des analyses statistiques et informatiques. Autrement dit, d'estimer la variabilité des données de nature technique, à la fois dans une même expérience que par rapport à d'autres observations, et de minimiser son impact sur les potentielles découvertes biologiques pouvant être représentés chez cette incertitude. *RUST* (Ribo-Seq Unit Step Transformation) est un programme ayant examiné cette variabilité intrinsèque (O'Connor *et al.*, 2016). Il utilise une simple transformation des profils de densité des ribosomes en une fonction binaire afin d'en retirer le vrai signal biologique des artefacts techniques. Étonnamment, il est

précisé dans l'article même que l'algorithme ne fonctionne que marginalement mieux qu'une simple normalisation des valeurs de RD des profils par le signal moyen à travers la CDS. Il n'existe jusqu'à présent aucune autre étude ayant présenté une manière moins triviale de normaliser les profils de RD (Bartholomäus *et al.*, 2015). Il a été suggéré de normaliser par les valeurs d'expression correspondantes issues d'un RNA-Seq classique, mais cela n'apporte pas d'avantages assez marquants pour pouvoir en justifier les coûts plus élevés (Artieri & Fraser, 2014). D'autres groupes ont tenté de dissocier le bruit et le signal biologique dans les données en proposant des modèles plus élaborés pouvant être utilisés pour prendre en compte les différents biais qui y sont présent sans toutefois en éliminer de l'information potentiellement pertinente. C'est le cas par exemple de (Gritsenko *et al.*, 2015) et leur modèle TASEP (Totally Asymmetric Exclusion Process), soit un simple modèle dynamique combiné avec une nouvelle stratégie de simulation de Monte-Carlo au sein d'un algorithme d'optimisation évolutive. Il en va de même pour le groupe de (Dana & Tuller, 2014) et leur modèle de normalisation des données de RD. Ce dernier filtre certains biais connus dans les données, tels que l'enrichissement des codons aux deux extrémités des transcrits, mais tente aussi d'identifier et d'inclure dans le modèle la détection de sites de pauses ribosomales rares. Ces deux dernières études ont d'ailleurs présenté des conclusions contradictoires sur la corrélation entre les RD propres aux codons individuels et la disponibilité de leur ARNts respectif dans la cellule (Gritsenko *et al.*, 2015) ne reportant pas de corrélation significative et (Dana & Tuller, 2014), l'inverse. Cela contribue à démontrer l'incertitude liée au choix de méthode pour l'analyse des données de Ribo-Seq et, du fait même, le besoin de développer une technique avec davantage de fiabilité et de reproductibilité. Ce dernier critère fut d'ailleurs remis en question par certaines études. L'équipe de (Diamant & Tuller, 2016) a remis en doute l'habilité du Ribo-Seq de mesurer la RD à une échelle nucléotidique. Dans leur étude, ils analysent 15 différentes expériences de Ribo-Seq avec de multiples répliques et pour 6 différents organismes. En comparant les valeurs moyennes de RD pour chaque gène, ils obtiennent des coefficients de corrélation  $r$  au-dessus de 0,85. Cependant, la comparaison des valeurs moyennes de RD pour chaque position individuelle de nucléotides dans les gènes, retourne des corrélations de moins de 0,24. Les idées de solutions semblent toutefois être unanimes, le protocole analytique des données de Ribo-Seq nécessite un pipeline uniforme prenant en considération ces diverses limitations et s'y adaptant correctement. Il en va de même pour les étapes de contrôle de qualité.

Cela dit, la méthode acquiert davantage de maturité avec le temps, la technique de Ribo-Seq ne fait que débiter sa contribution à la science (Gritsenko *et al.*, 2015; Ingolia *et al.*, 2011).

### 1.3.4 L'analyse de signaux et la détection de pics

Le profilage des ribosomes permet l'étude de la traduction en fournissant des informations sur la densité ribosomale en temps réel et à des résolutions nucléotidiques, pouvant être extrapolées aux codons. La manière logique de représenter et d'étudier un tel signal consiste à attribuer des valeurs de densité ribosomale (RD) aux positions des codons dans les régions codantes du génome (Ingolia, 2016). Cela pose un défi dans l'identification d'un signal pouvant être considéré significatif, en particulier pour les personnes intéressées par l'étude de la cinétique de traduction. La détection de pics importants dans un signal bruyant est un défi depuis longtemps partagé par presque tous les domaines et pratiques scientifiques tels que la médecine, les télécommunications, l'astrophysique, etc. (Biswas *et al.*, 2014). Avec des revues telles que *IEEE Transactions on Signal Processing*, elle détient même son propre champ de recherche. En génomique, en particulier avec l'avènement des données de séquençage profond à haut débit et de nouvelle génération (NGS), elle est devenu un enjeu central à l'interprétation des résultats (Kumar *et al.*, 2013). Dans l'effectif, plusieurs implémentations distinctes de la technique RNA-Seq souffrent de défis similaires dans l'identification et l'interprétation de leurs signaux (ChIP-seq, DNase-seq, FAIRE-seq, etc.). L'interprétation des résultats de techniques légèrement plus anciennes comme ChIP-seq a déjà été plus largement étudiée, avec une pléthore de logiciels et d'algorithmes ayant été publiés pour l'analyse des pics dans les résultats de la méthode (Hocking *et al.*, 2017). Souvent, de nombreuses années séparent les différentes publications, ce qui démontre le manque de protocoles standard et la possibilité constante d'améliorations. La méthode de profilage des ribosomes étant plus récente, un protocole satisfaisant pour l'analyse de ses données reste en suspens (Bartholomäus *et al.*, 2015; Diamant & Tuller, 2016).

La nature de l'analyse des données de Ribo-Seq varie en fonction de son objectif. Si ce dernier consiste à l'évaluation des informations sur l'abondance d'expression de transcrits, comme dans le design classique du RNA-Seq, une méthode d'appel de pics précise n'est pas nécessaire. Des outils comme *Scikit-ribo* font un excellent travail pour fournir ces informations, avec des résultats en TPM par exemple (Fang *et al.*, 2018). De plus, *GWIPS-viz* offre une expérience de visualisation par navigateur de génome, prenant en compte plusieurs résultats d'expériences de

profilage de ribosomes publiés (Michel *et al.*, 2014). L'interprétation précise des résultats de Ribo-Seq se heurte cependant à deux défis majeurs si l'évaluation des signaux de pauses ribosomales dans la traduction avec une résolution au nucléotide prêt en est son objectif. Tout d'abord, l'isolation de valeurs d'amplitude de pics significatives par rapport au bruit de fond. Ensuite, la synchronisation de ses signaux lorsque de multiples expériences sont considérées. Plusieurs groupes ont tenté d'utiliser des algorithmes universels de détection de pics ou des logiciels développés pour d'autres méthodes RNA-Seq pour effectuer cela (Ibrahim *et al.*, 2015).

Cette approche s'avère toutefois complexe puisque les signaux de ces méthodes sont de nature très différente. Le signal Ribo-Seq à une résolution au codon près présente des caractéristiques similaires à celles des séries chronologiques, soit avec des variations d'amplitude élevées et des largeurs de pic très faibles et nettes. En revanche, le signal pour les données ChIP-seq présente des zones de largeurs étendues avec une amplitude accrue (Jiang & Mortazavi, 2018). Un autre type de signal qui a tendance à ressembler à des séries chronologiques est celui de l'électroencéphalogramme (Biswas *et al.*, 2014). Il représente les mêmes pics nets avec une grande variation d'amplitude, mais affiche également un besoin de synchronisation, car il exhibe de légères variations de signal provenant des différents capteurs reliés à la machine. Une solution élégante à ces deux artefacts est présentée par (Biswas *et al.*, 2014). Ces derniers ont publié un article présentant un algorithme de synchronisation ascendante sur des pics nets dans plusieurs signaux. L'algorithme fonctionne en attribuant des valeurs binaires pour les positions des pics dans le profil du signal, un pic est identifié par un modèle de détection de pic de seuil d'amplitude simple, c'est-à-dire que tout signal présentant plus du double de la valeur de l'écart-type pour le profil est considéré comme un pic. Ils ont ensuite attribué des poids aux différents signaux de pics, sur la base d'une fonction de densité de probabilité. Si un pic est présent dans plus d'un profil de signal, un maximum lui est attribué et ses valeurs voisines se voient attribué un poids selon une fonction de probabilité décroissante. Les paramètres de la fonction de probabilité peuvent être modifiés pour lui accorder plus d'amplitude (valeur du poids) ainsi que plus de largeur (étendue de distribution). Le score chaque positions dans tous les signaux est ensuite additionné et normalisé afin de produire un profil de consensus représentant tous les profils de RD individuels réunis.

### 1.3.5 L'annotation fonctionnelle des gènes

Le progrès dans les technologies de séquençage du génome représente une ressource incroyable en recherche biologique, cependant, encore faut-il pouvoir interpréter et classifier l'énorme quantité de données générées. Une séquence génomique révèle très peu sans être associée à un élément spécifique. Cela dit, l'annotation génétique est aussi ancienne que les premiers travaux de séquençages (Mudge & Harrow, 2016). Alors que ces travaux portaient sur des gènes codants pour des protéines individuelles, les données génomiques modernes comportent des structures plus complexes à annoter, par exemple; l'épissage alternatif, la transcription non codante et les pseudogènes (Mudge & Harrow, 2016). L'annotation des gènes vient répondre à cette technicalité et permet de faire le lien direct entre la séquence génomique et les fonctions biologiques (Aken *et al.*, 2016). Avec la quantité énorme de données générées, des dizaines de millions d'entrées dans la base de données PubMed, il est au-delà de la capacité humaine d'assimiler toute cette information. Les scientifiques se sont donc tournés vers la conception de projets d'annotations à haut débit, basés sur le calcul informatique. L'annotation est un problème compliqué, qui se complexifie d'ailleurs davantage avec l'ajout de nouvelles découvertes dans la structure de l'expression génétique de la cellule. Plus la cible d'annotation s'éloigne du code génétique, plus les déficiences dans le système d'annotation se font sentir (Mudge & Harrow, 2016). En fait, la complexité même du transcriptome peut mener à se demander s'il est vraiment possible qu'il puisse être complètement décrit *in silico*. Une question d'annotation fondamentale concerne la proportion du transcriptome qui contribue à la fonction cellulaire, certains chercheurs disent même que l'objectif des projets d'annotation devrait être de décrire uniquement le transcriptome fonctionnel (Mudge & Harrow, 2016). Une meilleure compréhension de la régulation de la traduction peut venir en aide à cette problématique et à une meilleure organisation des fonctions cellulaires globales. Notamment au travers de la traduction sélective de sous-ensembles de gènes qui contribueraient à distinguer davantage le signal d'intérêt du bruit.

L'annotation se base sur les données de l'ontologie génétique, soit l'une des ressources les plus robustes et les plus utilisées pour classer les entités biologiques en groupes fonctionnellement liés. L'objectif des annotations GO est de générer la définition la plus précise possible du rôle d'un gène dans la biologie et de rendre les données compréhensibles autant à l'homme qu'à la machine (Attrill *et al.*, 2019). Elle se compose de trois principales structures de classification en ce qui concerne les produits géniques; les composants cellulaires, la fonction moléculaire et les processus



biologiques (Ashburner *et al.*, 2000). Chaque entrée de gène dans l'ontologie GO s'accompagne d'un identifiant unique, un nom et un nom de groupe qui indique le domaine auquel il appartient. La relation entre ces termes est décrite par un graphe orienté acyclique (DAG) (Eden *et al.*, 2009). Depuis sa création, de nombreux outils ont été développés pour explorer, filtrer et rechercher dans la base de données GO. Le souci de standardisation de ces informations a créé le *Gene Ontology Consortium*, un groupe de quelque dizaines de scientifiques qui s'occupent du maintien de la base de données web *The Gene Ontology Resource* (Ashburner *et al.*, 2000), contenant toutes les informations d'annotation. Un large répertoire d'outils pour l'analyse d'enrichissement a été développé ces dernières années, y compris GOAT (Xu & Shaulsky, 2005), FatiGO (Al-Shahrour *et al.*, 2004), DAVID (Huang *et al.*, 2007) et g:Profiler (Raudvere *et al.*, 2019), parmi d'autres. En général, ces outils acceptent en entrée des listes de gènes d'intérêts et performant des tests statistiques afin de déterminer leurs enrichissements chez les trois ontologies principales mentionnées plus haut. Le test statistique utilisé pour l'analyse d'enrichissement est généralement basé sur un modèle hypergéométrique ou binomial (Eden *et al.*, 2009). En somme, l'annotation fonctionnelle des gènes est complémentaire aux résultats de séquençage et permet non seulement d'aider à les élucider, mais aussi à guider les futures pistes de questionnements.

## 1.4 Hypothèse et objectif

L'élongation de la traduction est une des plus importantes étapes de l'élaboré système de l'expression génétique. La technique du profilage de ribosome en permet l'observation approfondie par le décèlement de régions de ralentissement de la translocation du complexe ribosomale sur les transcrits d'ARNm. Malgré les promesses de cette technique, l'analyse des données qui en résulte présente des incertitudes dans leurs reproductibilités ainsi que leurs variations entre les expériences, particulièrement lors de l'observation des résultats avec une précision au nucléotide près.

Dans le cadre de ce travail, nous avons émis l'hypothèse que l'hétérogénéité de la traduction contribue à la variabilité décrite ci-dessus. C'est-à-dire, qu'une portion des signaux attribués aux bruits techniques dans les résultats de Ribo-Seq contient plutôt de l'information d'origine biologique, jouant un rôle dans la régulation de la traduction.

Pour ce faire nous avons deux principaux objectifs. Premièrement, la conception d'un modèle mathématique permettant la mise en place d'un profil consensus de valeurs de densités ribosomales (RD) du transcriptome de la levure (*S. cerevisiae*) à partir de l'agrégation de données issues de distinctes expériences de Ribo-Seq. Pour continuer, nous avons comme deuxième objectif de comparer les séquences différentiellement traduites (DTs), soit ayant la même composition en termes de résidus, mais une disparité dans leurs signaux RDs consensus, en termes de leur impact sur les systèmes concomitants à la traduction. Le but de ces travaux étant de contribuer à une meilleure compréhension et caractérisation de l'hétérogénéité de la traduction.

## 2. Article

### **Inferring translational heterogeneity from ribosome profiling data**

Pedro do Couto Bordignon and Sebastian Pechmann

*bioRxiv* preprint first posted online Dec. 5, 2019; doi: <http://dx.doi.org/10.1101/866582>

#### **Contribution of authors**

Experiment design and analysis developed and implemented by Pedro do Couto Bordignon, under the supervision of Professor S. Pechmann, who also assisted in the preparation of final figures and manuscript text.

#### **Keywords**

homeostasis, proteostasis, gene expression, mRNA, translation, elongation, regulation, *Saccharomyces*, Ribo-Seq, ribosome, tRNA, modification, sequence, eukaryotic, heterogeneity

## 2.1 Abstract

Translation of messenger RNAs into proteins by the ribosome is the most important step of protein biosynthesis. Accordingly, translation is tightly controlled and heavily regulated to maintain cellular homeostasis. Ribosome profiling (Ribo-Seq) has revolutionized the study of translation by revealing many of its underlying mechanisms. However, equally many aspects of translation remain mysterious, in part also due to persisting challenges in the interpretation of data obtained from Ribo-Seq experiments. Here, we show that some of the variability observed in Ribo-Seq data has biological origins and reflects programmed heterogeneity of translation. To systematically identify sequences that are differentially translated (DT) across mRNAs beyond what can be attributed to experimental variability, we performed a comparative analysis of Ribo-Seq data from *Saccharomyces cerevisiae* and derived a consensus ribosome density profile that reflects consistent signals in individual experiments. Remarkably, the thus identified DT sequences link to mechanisms known to regulate translation elongation and are enriched in genes important for protein and organelle biosynthesis. Our results thus highlight examples of translational heterogeneity that are encoded in the genomic sequences and tuned to optimizing cellular homeostasis. More generally, our work highlights the power of Ribo-Seq to understand the complexities of translation regulation.

## 2.2 Introduction

Translation of messenger RNA (mRNA) into protein sequences by the ribosome is the central step of protein biosynthesis and the energetically most expensive process in the cell (Kafri *et al.*, 2016). Accordingly, translation is tightly controlled and heavily regulated to sustain cellular homeostasis (Sonenberg & Hinnebusch, 2009). In turn, the dysregulation of translation is associated with many severe human diseases including cancer and neurodegeneration (Schafer *et al.*, 2015; Tahmasebi *et al.*, 2018). The kinetics of translation initiation exert the strongest influence on final protein synthesis rates (Sonenberg & Hinnebusch, 2009). However, translation elongation, the speed and dynamics at which ribosomes move along mRNAs, is equally heavily regulated and emerging as an important aspect to defining the fate of the encoded nascent polypeptides (Pechmann *et al.*, 2013; Richter & Collier, 2015).

Ribosome profiling (Ribo-Seq), the experimental high-resolution measurement of translation kinetics through deep-sequencing of ribosome-protected footprints, has revolutionized

the study of translation (Brar & Weissman, 2015; Ingolia, 2014; Ingolia *et al.*, 2009). Next to the accurate identification of translated sequences in genomes (Calviello *et al.*, 2016), many fundamental insights into the determinants of translation elongation kinetics could so far be derived from Ribo-Seq data. For instance, codon translation rates on average correlate with cellular tRNA abundances (Dana & Tuller, 2014; Hanson & Collier, 2018; Hussmann *et al.*, 2015). Positively charged amino acids can slow down translation through interaction with the negatively charged interior of the ribosome exit tunnel (Charneski & Hurst, 2013; Lu & Deutsch, 2008), as can upstream RNA secondary structures that hinder ribosome translocation (Yang *et al.*, 2014). Specific sequence motifs can stall translating ribosomes such as select codon pairings (Gamble *et al.*, 2016), or successive proline residues that require the specialized translation factor EF-P for their translation (Doerfel *et al.*, 2013). Moreover, these determinants of translation elongation kinetics (Riba *et al.*, 2019; Sabi & Tuller, 2017; Zhao *et al.*, 2019) are selectively placed along mRNA and protein sequences to prevent ribosome traffic jams and overall optimize translation.

Equally importantly, the kinetics of translation elongation directly coordinate downstream processes on the ribosome important for determining the fate of newly synthesized proteins (Pechmann *et al.*, 2013). This includes cotranslational protein folding (Park & Subramaniam, 2019; Reuveni *et al.*, 2011; Tuller *et al.*, 2010), interaction of nascent chains with chaperones or ribosome-associated targeting factors (Geller *et al.*, 2018; Pechmann *et al.*, 2014; Stein *et al.*, 2019), protein complex assembly (Shiber *et al.*, 2018), and even cotranslational protein degradation (Duttler *et al.*, 2013). Accordingly, the tuning of translation elongation is coupled to powerful feedback systems of cellular regulation (Chu *et al.*, 2014), for instance in response to stress (Goodarzi *et al.*, 2016; Shalgi *et al.*, 2013; Wu *et al.*, 2019).

Yet, despite tremendous progress made possible through the emergence of Ribo-Seq (Ingolia *et al.*, 2009), many aspects of translation remain mysterious. In part this must be due to the complexity of the underlying biology that awaits to be further uncovered at rapid pace. However, in part this may also stem from persisting challenges in the interpretation of Ribo-Seq data that are often characterized by high intrinsic variability. For instance, the inference of individual codon translation rates from Ribo-Seq data commonly results in wide distributions rather than narrowly defined individual values (Duc & Song, 2018; Nedialkova & Leidel, 2015; Sharma *et al.*, 2019; Zinshteyn & Gilbert, 2013). This renders the mechanistic modelling of

translation, even though strong progress could be made (Duc & Song, 2018; Reuveni *et al.*, 2011; Tunney *et al.*, 2018; von der Haar, 2012; Zarai *et al.*, 2017), challenging. Some uncertainty in Ribo-Seq data could be directly attributed to technical difficulties and experimental biases (Bartholomäus *et al.*, 2015; Diamant & Tuller, 2016; Hussmann *et al.*, 2015; O'Connor *et al.*, 2016). However, Ribo-Seq experiments have been continuously improved (Lecanda *et al.*, 2016; Weinberg *et al.*, 2016) and technical noise alone cannot explain all the variability observed in Ribo-Seq data.

To this end, the process of translation itself is known to exhibit substantial heterogeneity (Sauert *et al.*, 2015). Eukaryotic cells contain up to 108 ribosomes (Milo *et al.*, 2010), and variation in composition, modification, interaction, and localization can specialize their functionality (Guo, 2018; Hui & Boer, 1987; Xue & Barna, 2012). Such specialized ribosomes have been found for instance to prioritize the efficient translation of proteins destined for export to mitochondria (Shi *et al.*, 2017). Similarly, ER-targeted proteins are preferentially translated by a select pool of translocation competent ribosomes (Chartron *et al.*, 2016). Heterogeneity equally exists at the level of translation elongation. The cellular abundances of tRNAs affect the efficiency of diffusion to the ribosome, thus codon-specific translation rates (Reis *et al.*, 2004). Long considered stable in their expression, the principles and functional implications of regulating tRNA abundances are only starting to emerge (Goodarzi *et al.*, 2016). Another mechanism that directly affects the decoding efficiency is achieved through RNA modifications at or near the tRNA stem loop (60, 61). Of note, tRNA modifications are well characterized biochemically but their physiological roles in most cases remain poorly understood (Grosjean *et al.*, 2010; Quax *et al.*, 2015). An example of a universally conserved tRNA modification is the modification of the anticodon wobble uridine (U34) in the tRNA genes tEUUC, tKUUU, and tQUUG that increases the efficiency of translating AAA, CAA, and GAA codons (Nedialkova & Leidel, 2015, 2015). Taken together, a plethora of processes within the cell exists that can render translation heterogeneous, which should be reflected in Ribo-Seq data.

Here, we have performed a comparative analysis of Ribo-Seq data from *S. cerevisiae* to systematically detect principles of translational heterogeneity from data obtained by bulk Ribo-Seq experiments. By systematically identifying subsequences whose differential translation cannot be explained solely by variability between experiments, we present cases of biological

heterogeneity of translation that are encoded in the genomic sequences and likely tuned to optimizing protein and organellar biogenesis. Our results suggest that some variability commonly observed in Ribo-Seq data actually represents fascinating functional biology.

## 2.3 Results

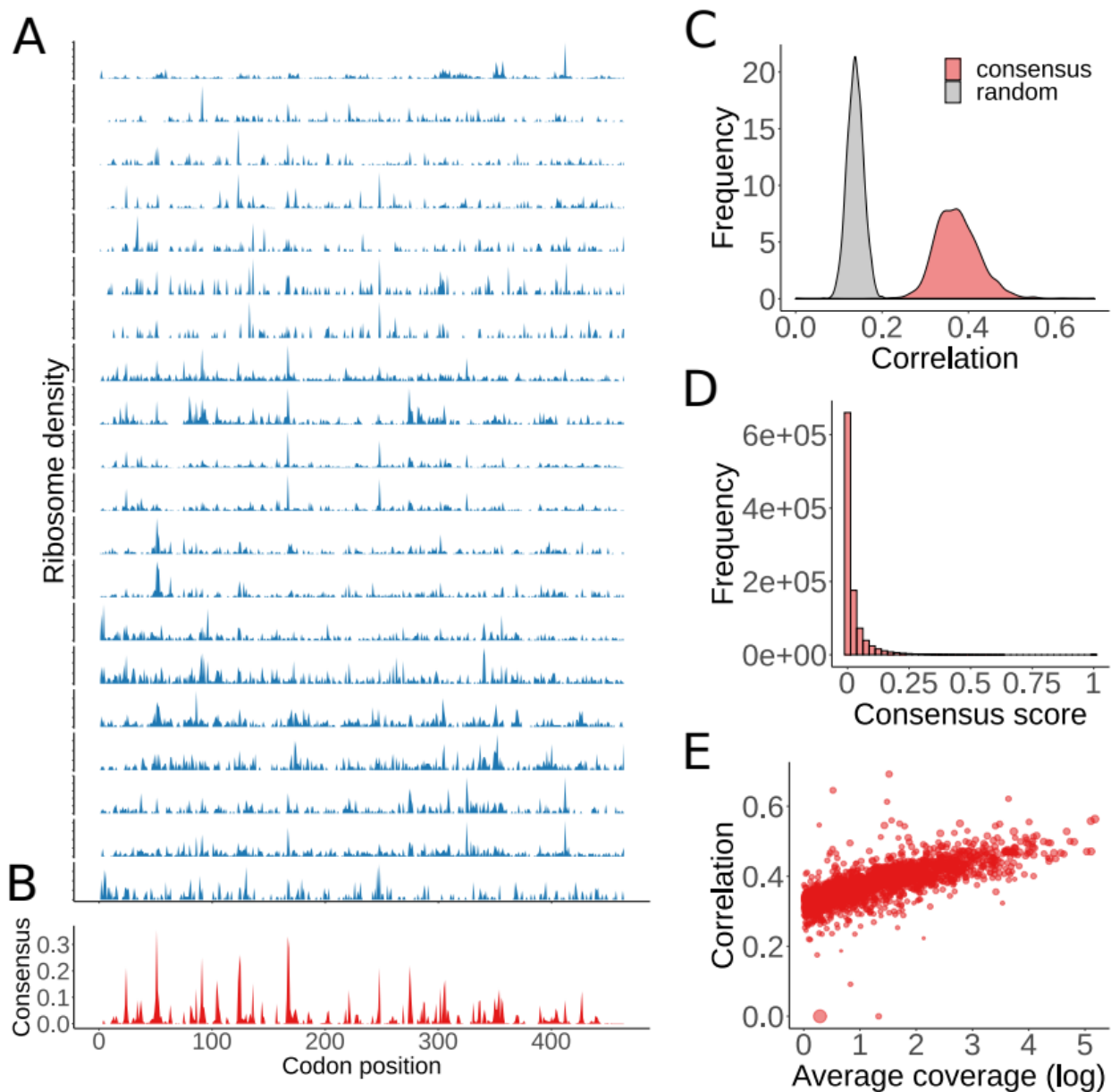
Ribo-Seq experiments afford to probe the dynamics of translation at very high resolution through the selective sequencing of ribosome protected footprints. Nonetheless, due to the complexities of the experiment and underlying biology, Ribo-Seq data published so far often contain a high degree of intrinsic variability (Diament & Tuller, 2016). This was equally apparent from our comparative analysis of data from 20 independent Ribo-Seq experiments comprising both biological replicas from the same laboratories as well as data obtained by different research groups (see Methods; Table 2.1). Illustrated for the exemplary yeast gene *LTV1*, we found at first sight substantial variability between the corresponding ribosome density (RD) profiles of individual experiments (Figure 2.1A). At the same time, many characteristic peaks of high RD that indicate regions of slow translation were strongly preserved across experiments (Figure 2.1A). Thus, while there was variability in these data, strong translational attenuation could clearly be detected by Ribo-Seq systematically.

### 2.3.1 Consensus RD profiles represent individual experiments

To identify a representative RD signal that is supported by independent individual experiments, we sought to derive a consensus RD profile along transcripts. Specifically, regions of high RD along mRNAs that are present in all or most individual profiles should be clearly reflected in a consensus profile as they likely represent strong signatures of translation attenuation. Due to the sensitivity of arresting translation during Ribo-Seq experiments without use of cycloheximide, RD peaks in individual experiments may not be found at exactly the same positions but very slightly shifted. To account for this, we chose to derive a consensus profile upon applying a peak-synchronization algorithm (Biswas *et al.*, 2014) (see Methods). Importantly, the resulting consensus RD profile visibly reflected the main features of the individual profiles for the gene *LTV1* (Figure 2.1B).

To test whether our consensus generally represented the individual experiments, we computed the per-gene average of the correlation coefficients between the individual profiles and the consensus. As a control, we calculated a consensus and average correlation for sets of

individual profiles randomly chosen from different genes (see Methods). We found much higher correlation coefficients for the true data compared to the random control, suggesting that our consensus profiles indeed well reflected the individual experiments (Figure 2.1C).



**Figure 2.1:** Consensus ribosome density profiles. (A) Ribosome density (RD) profiles from independent experiments for the exemplary yeast gene *LTV1*. (B) Consensus ribosome density profile for the gene *LTV1* derived from the individual experiments shown in (A). (C) Consensus profiles are representative of individual experiments. Shown are the distributions for the average correlation coefficient between individual experiments and the corresponding consensus compared to a randomized control (*see Methods*). (D) High consensus values are rare. Distribution of RD consensus scores across all genes analyzed. (E) Sequencing depth limits the resolution of ribosome



profiling data. Higher sequencing coverage on average results in stronger correlations between individual experiments and their consensus.

The obtained consensus scores ranged by definition from [0, 1], but high consensus scores were found to be very rare. Across all genes the vast majority of positions exhibited consensus scores of or just above 0, while only few positions had relatively high consensus scores (Figure 2.1D). As a result, peaks in the consensus profile were generally sparse and very clearly separated from the background signal (Figure 2.1B).

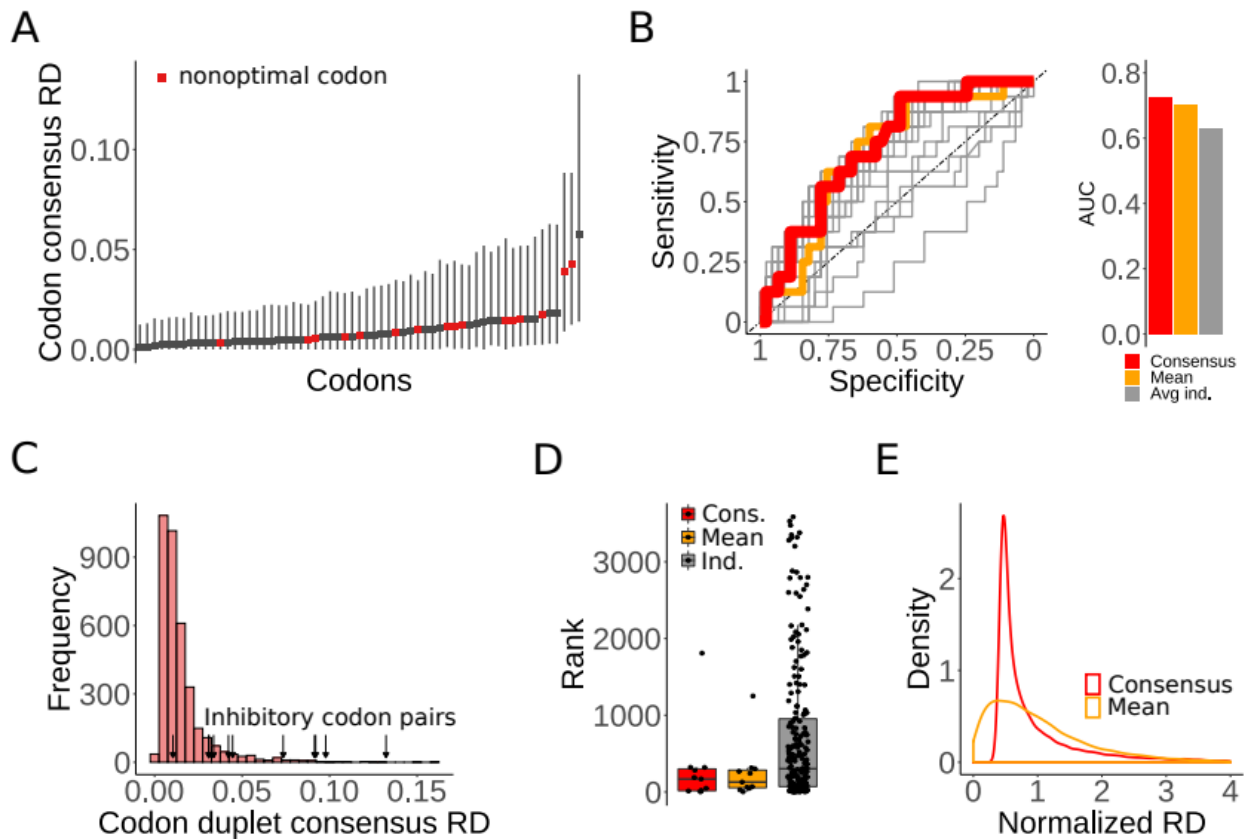
Moreover, we found that how well the consensus represented the individual profiles depended, as expected, strongly on the sequencing coverage. Genes with higher read-depth of ribosome protected footprints were clearly found with higher correlations between individual experiments and the consensus (Figure 2.1E). Our analyses thus suggested that a current limiting factor of Ribo-Seq may be foremost insufficient sequencing depth. Taken together, the derived RD consensus profiles clearly represented individual experiments and highlighted regions whose slow translation was detected systematically by independent experiments with a strong consensus signal that clearly separated from the background.

### **2.3.2 Consensus profiles are biologically meaningful**

We next sought to evaluate whether the derived consensus profiles were also carrying biological meaning. To test this, we compared the different RD profiles to known determinants of translation elongation. Characteristic values for codon-specific ribosome densities are commonly interpreted as dwell times and used to infer codon translation rates (Gobet *et al.*, 2019). Accordingly, the median of the distributions of the RD values for the 61 sense codons was computed for the individual experiments, the consensus, and as a control for a simple average (mean) profile of the individual experiments (see Methods). More slowly translated, or nonoptimal codons usually correspond to those that have a low tRNA adaptation index (tAI) (Reis *et al.*, 2004). To test whether the consensus profile was as good or better than individual profiles at identifying nonoptimal codons, we defined the bottom 25% of the sense codons with the lowest tAI as nonoptimal and evaluated the discriminative power of the inferred codon median RDs to correctly classify them.

As previously reported, individual codon RDs distributed broadly, but nonoptimal codons were found on average with higher codon consensus RDs (Figure 2.2A). Receiver Operating

Characteristic (ROC) curves for the individual experiments, the mean profile and the consensus quantified that the consensus was indeed as good or even slightly better at classifying the nonoptimal codons (Figure 2.2B). Foremost, this observation validated that the consensus profiles did not introduce artificial peaks without biological meaning, but it was nice to see that they performed very well at identifying nonoptimal codons. These results were found qualitatively independent of the definition of the nonoptimal codons. Some individual experiments performed as well or even very slightly better than the consensus, suggesting variability also between the experimental data sets. Notably, the consensus method was found very powerful to condense heterogeneous data to a strong and meaningful signal.



**Figure 2.2:** Consensus profiles identify nonoptimal codons and stalling sequences. (A) Distributions of consensus codon RDs for the 61 sense codons are represented by their median values (squares) and their 25 and 75 percentiles (solid lines). Nonoptimal codons (in red) correspond on average to higher consensus codon RDs. (B) The consensus profiles can identify nonoptimal codons. ROC curves for the classification of nonoptimal codons from codon RDs are shown for the consensus, mean, and individual profiles. The consensus has more discriminative power than the mean or average of the individual profiles. (C) Inhibitory codon pairs have high consensus RD scores. Distribution of the consensus RDs of all codon pairs. Codon pairs that have been reported as strong stalling sequences are highlighted and are predominantly found in the tail

of the distribution. **(D)** Rank order of inhibitory codon pairs. Both consensus and mean profiles identify inhibitory codon pairs as those with especially high RD scores. **(E)** The distribution of consensus RD scores follows a strongly convex upper tail in contrast to an initially concave upper tail for the mean RD scores. As a result, high RD signals in the consensus separate better from the background of low RD scores.

Similarly, experimental work has established that select codon duplets act as very strong stalling sequences (Gamble *et al.*, 2016). Here, the codon duplet consensus RDs of these 'inhibitory' codon pairs (Gamble *et al.*, 2016) were almost all found to be much higher than average (Figure 2.2C). Of note, we found inhibitory codon pairs strongly selected against in the set of analyzed genes, reflected by only very few occurrences in yeast mRNA sequences compared to the numbers of other codon pairs. Codon duplet RDs derived from both the consensus and the mean profiles ranked generally higher than in the individual experiments when sorting all codon duplet scores in ascending order (Figure 2.2D). These observations further underlined the power of a joint analysis of multiple ribosome profiling datasets through a consensus profile to filter out a consistent signal from heterogenous data sources.

Last, while both the consensus and the simple average of the individual profiles gave an improvement with respect to the prediction and classification of nonoptimal codons and inhibitory codon pairs, they clearly differed in how they distributed. The distribution of mean RD values followed a initially concave upper tail while the distribution of consensus RDs showed a strongly convex upper tail (Figure 2.2E). As a result, the high RD peaks in the consensus profile in general better separated from the background signal of low RD values.

#### **2.3.4 Differentially translated (DT) sequences are rare**

Individual codon RDs obtained from the consensus profiles displayed broad distributions similar to what had been reported previously (Duc & Song, 2018; Nedialkova & Leidel, 2015; Sharma *et al.*, 2019; Zinshteyn & Gilbert, 2013). This observation suggested substantial variation of the speeds at which individual codons are translated. Some of these differences may simply be due to intrinsic noise in the complex sequencing experiment. However, if the same codons, or subsequences of codons, could be found differentially translated beyond what can be accredited to experimental variability, then the observed differences may also have biological origins that reflect cases of translational heterogeneity.



our definition, only 159 as non-differentially translated with systematically low RD (nDTlo), and only 17 as non-differentially translated with systematically high RD (nDThi). **(D)** Codon frequencies in DT sequences. The observed enrichment of individual codons in DT sequences directly links to mechanisms known to modulate translation elongation kinetics.

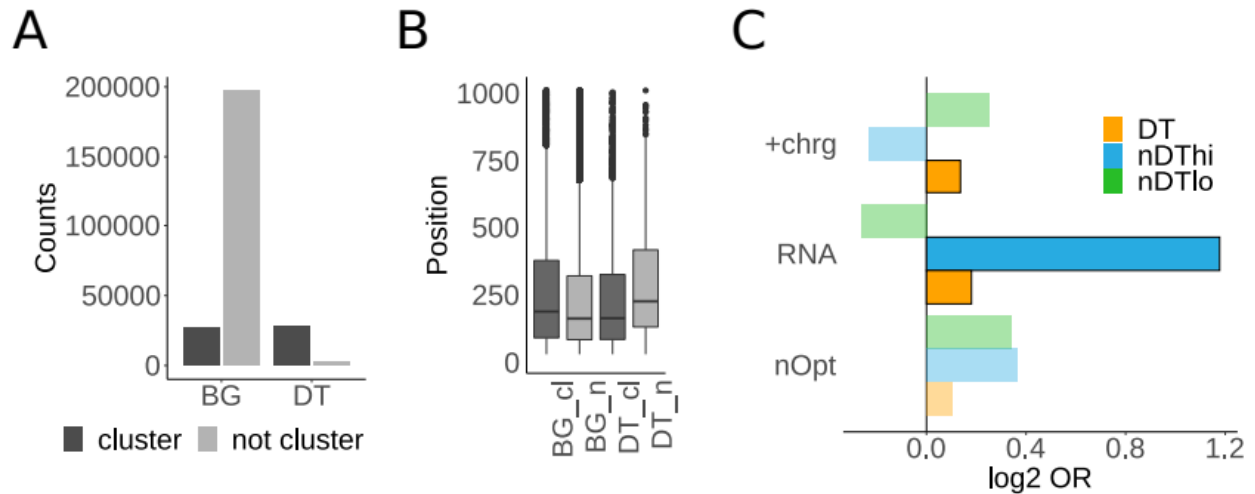
Next, we set out to systematically identify DT sequences in the *S. cerevisiae* genome. Focussing on subsequences of length 3 codons, we defined DT sequences as those that could be found at least 10 times at or below the lower gene-specific 10% threshold and at least 10 times above the gene-specific 90% threshold (Figure 2.3B). Of note, the analyzed 2443 *S. cerevisiae* mRNA sequences contained 5525 subsequences of 3 codons with a general redundancy of more than 20 occurrences. The median number of occurrences of these 5525 sequences was 27, and only the top 10% were found to have a redundancy of over 44 counts. Our definition of DT sequences was therefore sufficiently stringent and selective. Importantly, of these 5525 codon triplets, 451 fulfilled our definition as DT sequences (Figure 2.3C). While overall the number of subsequences outside the chosen significance thresholds correlated with their overall redundancy, the observed number of DT sequences was sparse, leaving the possibility that they serve specific functions. For comparison, only 159 sequences were found as translated consistently fast (nDTlo), and only 17 as translated consistently slowly (nDThi) (see Methods).

We next asked whether the DT sequences were enriched in specific codons. Strikingly, the codons the most over-represented in the DT sequences were those recognized by tRNAs whose known modifications at the wobble-uridine position modulate the speed of translation, primarily sequences containing GAA codons (Figure 2.3D). Equally importantly, absent were codons encoding Prolines that are known to elicit strong stalling and likely leave little room for contextual modulation of translation speeds. In contrast, nDTlo sequence were enriched in CAA codons, another codon linked to the U34 tRNA modification, while nDThi sequence showed a high frequency of the Proline codons CCA and CCT that act as known strong stalling signals. Equally remarkably, when normalizing amino acid usage by transcript abundance, i.e. how often the corresponding codons are translated by the ribosome, Glutamine residues become one of the most heavily used amino acid (Pechmann & Frydman, 2013). As such, choice between CAA and CAG codons encoding Gln residues may strongly impact overall translation efficiency. Accordingly, we observed a clear preference for the optimal codon CAA, whose decoding efficiency is equally further increased by tRNA modification, in the nDTlo sequence, and of the nonoptimal codon CAG in the nDThi sequences. In addition to distinct codon preferences, low RD DT sequences

were on average found later along mRNA sequences and thus may also reflect known biases in decreasing coverage along transcripts. Taken together, our analysis suggested that DT sequences exhibited distinct codon usage patterns that strongly linked to known determinants of modulating translation elongation kinetics.

### **2.3.5 DT sequences may be important for protein biogenesis**

Having identified select sequences that are translated differentially in systematic fashion beyond what can readily be attributed to variability between experiments, we next sought to better understand what characterizes whether the same subsequence is translated slowly or fast. Clusters of several adjacent positions of low, and especially of high RD can have a more pronounced impact on translation elongation compared to individual positions. Regions of low RD can be under selection for fast translation, but can also result from limited sequencing coverage in the Ribo-Seq experiments. In contrast, high RD regions generally exert a stronger influence on translation elongation and coupled processes. We therefore focused on regions of high RD and first asked whether high RD DT sequence preferentially occur in clusters of other high RD positions or not. Across the set of analyzed mRNA sequence positions of high consensus RD generally did not show a preference for being in clusters, here defined as at least 3 sites with RDs above the gene-specific upper threshold in a window of 7 positions (Figure 2.4A). In contrast, the vast majority of DT sequences were found to occur in clusters (Figure 2.4A). This observation suggested that DT sequences may contribute to the selective and potentially functional attenuation of translation.



**Figure 2.4:** Characteristics of DT sequences. **(A)** DT sequences are preferentially found in clusters. While only a small fraction of non-redundant high consensus RD regions (BG for background) cluster, the vast majority of DT sequences with high consensus RD appear in clusters. **(B)** Preference for DT sequences in clusters early in sequence. DT sequences in clusters (DT c) are found on average closer to the protein N-terminus compared to BG, and in contrast to BG earlier in sequence compared to DT sequences not in clusters (DT n). **(C)** Known stalling signals contribute to DT sequences. Illustrated are statistical associations between high consensus RD occurrences of DT sequences and RNA folding strength, nonoptimal codons, and clusters of positively charged amino acids. Significant relationships ( $p < 0.05$ ) are indicated by solid color and black border around the bars indicating the log<sub>2</sub> odds-ratio (OR).

Moreover, we found a bias of DT sequences in clusters to be positioned earlier in sequence than those not in clusters, and in contrast to the background signal of all high consensus RD positions (Figure 2.4B). If translation elongation influenced protein synthesis, folding, and homeostasis, any cues to optimize translation or promote downstream processes would be expected earlier in sequence. Guiding the formation of a folding nucleus, coordinate chaperone binding, or, equally importantly, regulating ribosome spacing to prevent ribosome traffic jams, thus facilitating efficient overall translation, are all decisive events for the fate of a nascent polypeptide that usually occur early on during translation. Our result of a positional bias thus left open the possibility that DT sequences may play a role in regulating protein biogenesis.

To better understand how the sequence context may determine whether instances of the same sequence were sometimes translated fast and sometimes slowly, we analyzed a putative link to three known stalling signals, clusters of positively charged amino acids, high RNA folding strength, and clusters of nonoptimal codons. In comparison to the DT sequences that were defined as sequences with at least 10 occurrences each below and above the significance thresholds, we

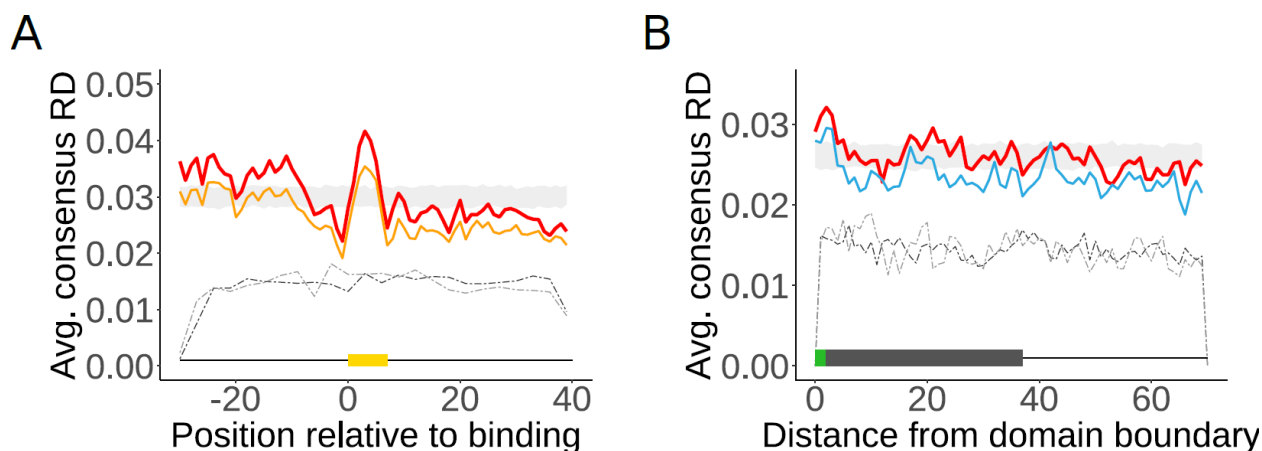
analyzed sequences that were consistently translated fast or slowly, i.e. nondifferentially translated sequences of consistently high RD (nDThi) and low RD (nDTlo) respectively (see Methods). Positively charged amino acids can slow down translation through interaction with the interior of the ribosome exit tunnel. Our analysis suggested that high RD DT sequences were significantly ( $p = 0.007$  by Fisher's Exact test) enriched in downstream clusters of positive charges (Figure 2.4C). Similarly, high RD DT sequences were significantly enriched in high RNA folding strength upstream ( $p = 0.002$  by Fisher's Exact test), which may hinder ribosome translocation along the mRNA, thus slowing down translation (Figure 2.4C). Remarkably, nDThi sequences, i.e. those that are consistently translated more slowly, linked to a very strong enrichment of high RNA folding strength ( $p = 0.03$  by Fisher's Exact test) (Figure 2.4C). In all cases we observed a higher number of clusters of nonoptimal codons just upstream of the translated position of high RD, albeit without significance. Taken together, these results suggested that known stalling signals play a clear role in the observed heterogeneity of translating the same sequences.

### **2.3.6 No systematic role of DT sequences in protein folding**

Based on the findings that the rhythm of translation elongation can coordinate protein folding events at the ribosome (Chaney & Clark, 2015; Pechmann & Frydman, 2013; Sharma & O'Brien, 2018), we next asked whether DT sequences may systematically link to two major events in cotranslational protein folding, the binding of the main cytosolic Hsp70 chaperone SSB, and the folding of protein domains. It has been previously shown that a local slowdown of translation can coordinate the binding of chaperones or targeting factors to nascent polypeptides at the ribosome (Pechmann *et al.*, 2014; Stein *et al.*, 2019). Once bound to an elongating nascent chain, a chaperone can fundamentally change the folding landscape of the protein. Therefore, a coordinated attenuation of translation may increase the likelihood of chaperone binding to promote correct folding. We observed, as previously reported, a distinct and localized higher average consensus RD in the translated mRNAs just when Hsp70 is binding to the elongating nascent polypeptide outside the ribosome exit tunnel (Figure 2.5A). Notably, this signal is much more pronounced when only considering the first binding site per protein rather than all of them. However, we could not observe any localized enrichment of DT sequences within the same distance of Hsp70 binding sites (Figure 2.5A). This result reinforced that local translation kinetics may contribute to coordinating the binding of cotranslationally acting chaperones, but suggested that this is independent of DT sequences.



Similarly, it had been previously observed that a local attenuation of translation is linked to protein domain boundaries, leaving additional time for a completely translated and exposed domain to fold outside the ribosome before translation continues at high speed (G. Zhang *et al.*, 2009). However, this could so far only be observed for individual proteins, not systematically (Saunders & Deane, 2010). Similarly, we could not detect any systematic local slowdown of translation that may link to internal protein domain boundaries (Figure 2.5B). Accordingly, we found no enrichment of DT sequences that may coordinate protein domain folding (Figure 2.5B).



**Figure 2.5:** DT sequences are not systematically coordinating protein folding on the ribosome. **(a)** Meta-gene analysis Hsp70 SSB binding sites. Reported are per-position average consensus RD profiles of mRNA regions that are translated while Hsp70 is binding cotranslationally to the encoded elongating nascent polypeptide outside the ribosome exit tunnel; shown are profiles for the first Hsp70 binding site in each substrate protein (red line) and all Hsp70 binding sites (orange line) relative to a randomized background indicating  $mean \pm SD$  (grey ribbon). On average, translation slows down when Hsp70 is binding (indicated by yellow area), as observed by (Stein *et al.*, 2019). The density of high (light grey dashed line) and low (dark grey dashed line) DT sequences shows no elevated occurrence that could contribute to this local translation attenuation. **(b)** Meta-gene analysis of protein domain boundaries. Reported are per-position average profiles consensus RD downstream of the first domain boundary (green area) per multi-domain protein (red line), and for only domain boundaries that precede any Hsp70 chaperone binding (blue line). No systematic translation attenuation linked to domain folding could be observed for the consensus RD values, although a local slowdown of translation may coordinate the folding of some protein domains that precede Hsp70 binding (blue line) outside the ribosome exit tunnel (grey area). The density of high (light grey dashed line) and low (dark grey dashed line) DT sequences shows no elevated occurrence that could contribute to local translation attenuation.

Thus, DT sequences appeared to be neither coordinating Hsp70 binding or domain folding in a systematic fashion. Because these are pivotal events in *de novo* protein folding, any programmed translation slowdown may not be encoded through DT sequences but rather through

other, more unambiguous stalling signals. Our results leave open the possibility that DT sequences are involved in the dynamically regulated coordination of such protein folding events in individual cases that however will have to be understood at much higher resolution.

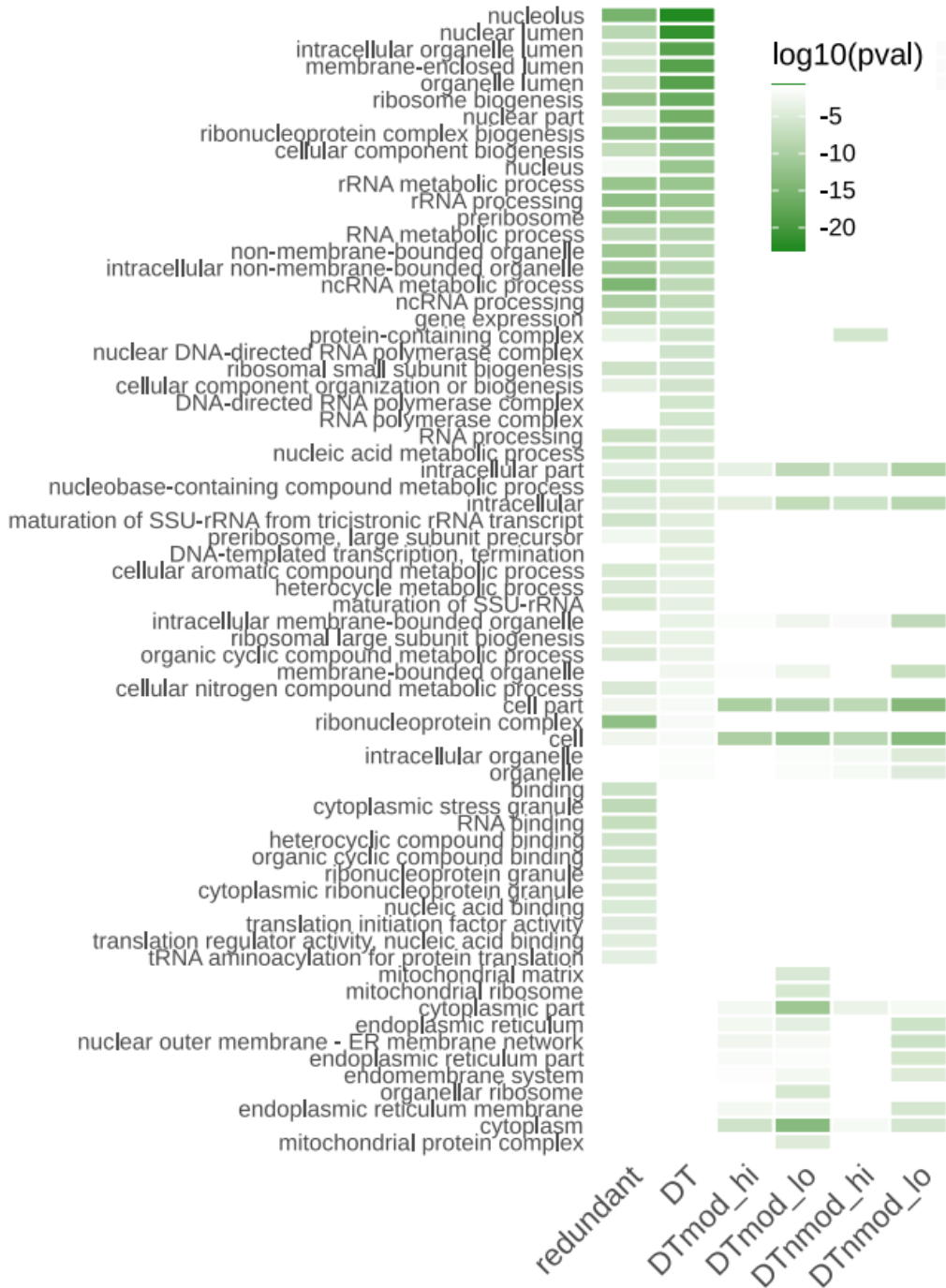
### **2.3.7 DT sequences link to protein and organelle biogenesis**

Our findings that DT sequences preferentially occurred in clusters and early in sequence indicated the potential for a functional role in coordinating early protein folding events at the ribosome. However, we could not find any systematic link between DT sequences and the binding of the main cytosolic Hsp70 SSB or the folding of protein domains. Next to the direct coordination of protein folding, which will have to be understood at higher resolution of the individual folding contexts, clusters of slowly or fast translated sequence regions can also have a strong cumulative impact on the overall translational efficiency, thus final protein levels. In addition to just more efficient overall translation, clusters of slowly translated codons early in sequence can contribute to the careful spacing of ribosomes along mRNA sequences to prevent ribosome traffic jams (Park & Subramaniam, 2019; Reuveni *et al.*, 2011; Tuller *et al.*, 2010).

To test whether genes enriched in DT sequences linked to specific functions within the cell, we performed Gene Ontology (GO) analyses (Figure 2.6). Specifically, we compared the functional annotations of the set of 'redundant' genes, i.e. mRNAs with the overall highest density of redundant subsequences, to five categories of genes enriched in DT sequences. Herein, we considered 'DT' genes as the mRNAs with the highest density of DT sequences outside the gene-specific thresholds (see Methods); in comparison, 'DTmod hi' and 'DTmod lo' were defined as the sets of genes with the largest difference between high and low RD DT sequences that contain GAA codons, i.e. the most overrepresented codon that is recognized by a tRNA subject to wobble-uridine modification; similarly, we considered 'DTnmod hi' and 'DTnmod lo' as the set of genes with the largest difference between high and low RD DT sequences not containing GAA codons (Figure 2.3D). These four additional lists of genes thus reflected mRNAs that contained DT sequence with strong bias towards consistently fast or slowly translated, and stratified by a possible effect of the wobble-uridine tRNA modification.

Several important observations could be made. First, each category corresponded to clearly defined GO categories. This suggested that these sets of genes not only shared distinct sequence characteristics, but also functional roles (Figure 2.6). A general correlation between the

'redundant' and 'DT' categories was expected as a higher level of redundancy on average and by our definition also leads to higher counts of DT sequences outside the significance thresholds. Both categories were strongly enriched in genes involved in nuclear biology, RNA processing, and ribosome biogenesis. The main difference between these two categories was that the set of 'redundant' genes was also strongly enriched in GO categories linked to stress granule formation. Redundant or repetitive sequences are known to link to phase transition and stress granule formation (Franzmann & Alberti, 2019; E. W. Martin & Mittag, 2018). In contrast, the 'DT' genes showed a stronger enrichment in GO categories linked to nucleus, nucleolus, and ribosome biogenesis biology (Figure 2.6), i.e. categories directly linked to protein biosynthesis.



**Figure 2.6:** Gene Ontology (GO) analysis of DT sequences suggests DT sequences link to protein and organelle biosynthesis. Enrichment of GO categories color-coded base on adjusted p-value as provided by the g:Profiler webserver. 'Redundant' denotes the top 5% of genes with the highest density of sequences of 3 codons that have at least 20 occurrences across the analyzed mRNAs; 'DT' are the top 5% of genes with the highest density of DT sequences. 'DTmod hi' and 'DTmod lo' are the top 5% genes with the largest difference between their counts of DT sequences of high and low consensus RD respectively and that contain GAA codons recognized by tRNAs subject

to modification of the anticodon wobble uridine; 'DTnmod hi', and 'DTnmod lo' are similarly defined for sequences that do not contain GAA codons.

Remarkably, the remaining lists of genes representing mRNAs with DT sequences that were almost only fast (low RD) or slowly (high RD) translated associated with an orthogonal set of GO terms, namely mitochondrial and endoplasmic reticulum (ER) biology (Figure 2.6). These included several GO terms describing processes of organelle biosynthesis, from 'endoplasmic reticulum membrane network' to 'mitochondrial matrix' to 'mitochondrial ribosome' (Figure 2.6). In general, more functional categories associated to the low than the high RD categories. The main difference between DT sequences that included codons whose translation can be modulated by wobble-uridine tRNA modification (Nedialkova & Leidel, 2015) was found to be an additional enrichment in the 'mitochondrial ribosome' and 'organellar ribosome' GO categories, as well as a link to 'mitochondrial matrix' and 'mitochondrial protein complex' GO terms. Strikingly, of the genes with consistently high consensus RD, DT sequences that depended on the tRNA modification were found predominantly in mRNAs important for endoplasmic reticulum biogenesis, while genes with consistently low consensus RD of their DT sequences subject to tRNA modification associated with mitochondrial biology. Similar results could be obtained by considering all three codons (AAA, CAA, and GAA) whose translation is modulated by this tRNA modification (Nedialkova & Leidel, 2015). Moreover, while this tRNA modification is normally pervasive and applied to most corresponding cytoplasmic tRNAs (Grosjean *et al.*, 2010), it may be less present in cases of localized translation known to facilitate the synthesis of proteins destined to the major organelles. Of note, mitochondrial proteins often fold in the cytosol and their export to the mitochondria often share components of the cytoplasmic protein quality control network (Jores *et al.*, 2018; Priesnitz & Becker, 2018; J. C. Young *et al.*, 2003), in *S. cerevisiae* for instance the Hsp70 chaperone SSA1 (Eliyahu *et al.*, 2012). In contrast, translocation to the ER is directly coupled to translation as ER-destined proteins usually cannot fold inside the cytoplasm. Moreover, ER-destined proteins are often translated by a specialized pool of translocation-competent ribosomes near the ER membrane (Chartron *et al.*, 2016). Our results thus suggest that the translational heterogeneity observed may be directly coupled to the organization of proteins to the major organelles, and indicate that tRNA modifications could play a role in this.

## 2.4 Discussion

Data from Ribo-Seq experiments have led to many fundamental discoveries of mechanisms that govern the kinetics of translation. However, challenges persist in the analysis and interpretation of Ribo-Seq data that are often characterized by intrinsic variability. Here, we show that some of this variability represents cases of biological heterogeneity of translation rather than experimental biases or noise. Through a comparative analysis of published *S. cerevisiae* Ribo-Seq data sets we have systematically identified short subsequences that exhibit substantial variation in how they are translated. Importantly, whether the same sequence was translated slowly or fast associated with known stalling signatures of translation elongation, including high RNA folding strength and positively charged amino acids, or linked to mechanisms that can dynamically regulate translation kinetics such as tRNA modifications. Moreover, these DT sequences were found selectively enriched in genes central to protein and organelle biosynthesis. Our results thus indicate how translational heterogeneity that is directly encoded in the genomic sequences may serve to optimize overall protein biosynthesis and cellular homeostasis.

Many aspects of translational heterogeneity, especially of localized translation or translation on specialized ribosomes, have been under increasingly intense investigation. A spectacular example illustrating the importance and immediate functional consequences of local protein biosynthesis is given by neurons where efficient protein production in distal axons is required to control synaptic transmission (Hafner *et al.*, 2019; Holt *et al.*, 2019). Similarly, the importance of localized translation in the production of proteins that constitute the major organelles is generally well established (Chartron *et al.*, 2016; Priesnitz & Becker, 2018). Interestingly, while the translation of ER-destined proteins is almost completely separated from the production of cytosolic proteins, substantial overlap exists between the protein biosynthesis and quality control pathways for mitochondrial and cytoplasmic proteins. Accordingly, cytosolic and mitochondrial translation are tightly coordinated (Couvillion *et al.*, 2016). The use of DT sequences, i.e. sequences that are differentially translated in a context-dependent manner, open an intriguing window into additional mechanisms that may contribute to cellular organization through translation regulation.

One such central regulatory mechanism that remains under dynamic cellular control is given by tRNA modifications, which, for instance, have been shown to play an important role in

the global regulation of protein levels in response to stress (C. Chan *et al.*, 2018; C. T. Y. Chan *et al.*, 2012; Deng *et al.*, 2015; Fernández-Vázquez *et al.*, 2013). We found that consistently slowly translated DT sequences containing codons decoded by tRNAs subject to wobble-uridine modification, i.e. suggesting an absence of the modification, are linked to the ER. In contrast, consistently fast translated sequences containing codons recognized by tRNAs amenable by wobble-uridine modification, which is normally present in cytosolic translation, associated with genes involved in mitochondrial biology. Remarkably, this overall link between the biased translation of DT sequences and organellar biology is directly following the established differences in the biosynthesis of mitochondrial and ER proteins. While an intriguing observation, our work cannot currently resolve whether our result is cause or consequence: does localized translation in a different cellular milieu merely differentially predispose to the availability of this tRNA modification? Or does the differential translation regulated through absence or presence of a tRNA modification aid in the targeting of the elongating nascent chains to their correct destination? While these questions warrant further investigation, our results already highlight a fundamental conundrum: biological sequences are initially invariant, but the cellular system is highly dynamic. Robust control could easily be achieved by using unique sequences decoding slow or fast translation. Given the complexity of the cell as a system, it is most likely that the combinatorics achieved through hierarchical layers of regulation are needed to ultimately sustain homeostasis. Herein, the use of redundant sequences can pose several distinct advantages. A synchronization of the dependency on cellular resources, e.g. specific amino acids that need to be attached to the elongating nascent chain, also render these sets of genes susceptible to shared regulatory pathways. Moreover, shared sequences at both the RNA and protein levels can serve as recognition and binding sites. A well characterized example that is supported by our results is given by stress granules which are known to depend on low complexity and repetitive sequences for their formation (Das *et al.*, 2015; Franzmann & Alberti, 2019; E. W. Martin & Mittag, 2018). Similarly, the strong enrichment of DT sequences in genes involved in ribosome biogenesis may be strongly tied to an autocatalytic control of ribosome biogenesis (Reuveni *et al.*, 2011). Taken together, our results suggest a use of DT sequences that may enable selective feedback in regulating and optimizing protein biosynthesis.

Extensively discussed in the context of gene expression (Raj & van Oudenaarden, 2008), our results also highlight a pervasive stochastic character of translation. Many of the steps required

for translation elongation are stochastic, foremost the diffusion of charged tRNA-ternary complexes to the ribosome that can strongly influence decoding efficiencies. Next to the randomness of biomolecular interactions, some stochasticity may actually be built into the system. The cell has to operate efficiently under many different conditions, and this is especially true for translation as the energetically most expensive process. Translation under non-stress conditions may be not completely optimized for overall protein production efficiency, but rather to maintain agility and rapidly adapt to changing and more challenging conditions. This has been directly observed for instance under conditions of amino acid starvation that rapidly affect codon-specific translation rates (Darnell *et al.*, 2018; Raj & van Oudenaarden, 2008). More generally, the suppression of stochasticity in biological systems is very expansive and subject to cost-benefit trade-offs that usually tolerate some level of stochasticity (Lestas *et al.*, 2010). Thus, for modeling translation as well as predicting the effect of translation elongation kinetics on cellular processes, it may become equally important to explicitly consider a stochastic character of translation than to improve on more narrowly defined estimates of codon-specific translation rates.

Finally, our results foremost underline the power of Ribo-Seq to discover fundamental aspects of translation. A main limiting factor in current Ribo-Seq studies may simply be too shallow sequencing depth. While it has been suggested that inference from individual experiments may be limited (Szavits-Nossan & Ciandrini, 2019), our work of deriving and analyzing a consensus signal demonstrates that a comparative meta-analysis of several datasets can yield important additional insights. Many details of programmed translational heterogeneity result from the complex interplay of many different contributions and await to be characterized at much higher resolution. Next to a better understanding of the sequence determinants of translational heterogeneity, this includes better understanding the functional roles of central tRNA modifications (Novoa *et al.*, 2012), rRNA modifications (Decatur & Fournier, 2002), and the interplay between tRNA and epigenetic mRNA modifications (Ranjan & Leidel, 2019). Our results on identifying principles of translational heterogeneity from bulk Ribo-Seq data suggest a shift in the interpretation of variability in these data and open a fascinating window into further understanding the intricacies of translation, one of the most central and energetically the most costly process in the cell. What has been considered unwanted variability in Ribo-Seq data in part actually reflects exciting biology.



## 2.5 Materials and Methods

### 2.5.1 Code and data availability

Computer code and data to reproduce all presented results is available at:

[www.github.com/pechmannlab/riboconsensus](http://www.github.com/pechmannlab/riboconsensus)

### 2.5.2 Data sources and processing

We analyzed the ribosome profiling datasets from the experiments performed without use of the translation inhibitor cycloheximide (Duncan & Mata, 2017; Gerashchenko & Gladyshev, 2014) that are listed in Table 2.1. Sequencing adapters were removed with Cutadapt (M. Martin, 2011). Reads that subsequently did not align to yeast ribosomal RNA or transfer RNA sequences were mapped to the *S.cerevisiae* reference genome R64-1-1 (Engel *et al.*, 2014) with STAR (Dobin *et al.*, 2013). Ribosome density (RD) profiles of unambiguously mapped reads were computed with Scikit-ribo (Fang *et al.*, 2018). Genes with an average coverage of less than 1 read per position across datasets, or with more than 30% multi-mapping, i.e. the fraction of mRNA sequences that yields ambiguous read mapping (Halpin *et al.*, 2020), were omitted. This resulted in a final dataset of 2443 yeast genes. Furthermore, the first and last 20 codons from each gene were excluded from our analyses of translation elongation as they reflect known biases that link to translation initiation and termination (Tunney *et al.*, 2018).

**Table 2.1:** Ribosome profiling datasets analyzed in this study.

<b>Publication</b>	<b>SRA accession</b>	<b>Reference</b>
<b>Gardin (2014)</b>	SRR1002819	(Gardin <i>et al.</i> , 2014)
<b>Gerashchenko (2014)</b>	SRR1520311	(Gerashchenko & Gladyshev, 2014)
<b>Guydosh (2014)</b>	SRR1042855	(Guydosh & Green, 2014)
<b>Pop (2014)</b>	SRR1688545	(Pop <i>et al.</i> , 2014)
<b>Nedialkova (2015)</b>	SRR1944981-3	(Nedialkova & Leidel, 2015)
<b>Young (2015)</b>	SRR2046309,10	(Young <i>et al.</i> , 2015)
<b>Lecanda (2016)</b>	SRR3945926,8	(Lecanda <i>et al.</i> , 2016)
<b>Radhakrishnan (2016)</b>	SRR3493886,7	(Radhakrishnan <i>et al.</i> , 2016)
<b>Beaupere (2017)</b>	SRR4000288,9	(Beaupere <i>et al.</i> , 2017)
<b>Gerashchenko (2017)</b>	SRR363557,8	(Gerashchenko & Gladyshev, 2017)
<b>Schuller (2017)</b>	SRR5008134,5	(Schuller <i>et al.</i> , 2017)
<b>Zou (2017)</b>	SRR5090936	(Zou <i>et al.</i> , 2017)

### 2.5.3 Consensus RD profiles

Consensus ribosome density profiles were computed with a peak synchronization algorithm initially developed for the analysis of EEG data (Biswas *et al.*, 2014). Herein, peaks in individual profiles were detected by segmentation above a threshold of profile  $mean+1SD$  (standard deviation). Peaks across individual profiles were then averaged upon applying a Gaussian weight relative to their distances from the current position; in this way peaks that were systematically present at the same position yielded a maximum consensus signal, while peaks that were shifted made a reduced contribution with their weight reflecting their distance. Gaussian weights were inferred based on a probability density function with threshold  $th = 0.0001$  after which tails were considered insignificant and omitted, a central coefficient  $ca = 0.5$  describing the bin width at which the maximum weight was applied, and  $SD = 1$ . The resulting consensus profiles were found very robust to the choice of these parameters as previously reported (Biswas *et al.*, 2014). Multi-mapping positions in sequences were excluded from computing the consensus profile. The average correlation to consensus was computed as the per-gene average of the pairwise Spearman correlation coefficients between individual experiments and the corresponding

consensus profile. The randomized control was generated by computing a consensus profile of individual experiments randomly drawn from different genes with a minimum length of 100 codons. Similarly, per-gene background percentiles were derived from computing the consensus from shuffled individual profiles of the same gene 100 times and analyzing the distribution of the resulting consensus values. In both cases multi-mapping positions were omitted. For comparison, a simple average profile was computed as the per-position average of the individual profiles normalized to  $mean = 1$ .

#### **2.5.4 Validation of consensus profiles**

To validate that the consensus RD profiles carried biological meaning, we sought to evaluate their predictive power for known stalling signals, notably nonoptimal codons and inhibitory codon pairs. The RDs of individual codons are commonly interpreted as ribosome dwell times and used to infer codon specific translation rates. Because Scikit-ribo accurately maps the translated codon in the ribosome A-site (Fang *et al.*, 2018), we used the Scikit-ribo output to infer characteristic codon RDs as the median of the distribution of RDs of all occurrences for each sense codon or codon pair. The first and last 20 codons from each gene and multi-mapping positions were omitted. Individual profiles were normalized to  $mean = 1$ . Receiver Operating Characteristic (ROC) curves for the classification of nonoptimal codons defined as the bottom 25% of the tRNA adaptation index (tAI) (Reis *et al.*, 2004) were computed with the R package pROC. Inhibitory codon pairs were defined as those in (Gamble *et al.*, 2016). For the comparison of consensus and mean RDs, their distributions were normalized to  $mean = 1$  and  $SD = 1$ .

#### **2.5.5 Identification of differentially translated (DT) sequences**

We focused on subsequences of length 3 codons as longer sequences rapidly drop in redundancy across the transcriptome thus limiting occurrences for further analysis. Differentially translated (DT) sequences were defined as sequences of 3 codons that had both at least 10 occurrences above gene-specific 90% and at least 10 occurrences below gene-specific 10% thresholds in their consensus profiles across all mRNAs. Thresholds were chosen as good trade-off between sufficient selectivity and data set size; subsequent results were found qualitatively very robust to these choices. For comparison, we extracted all sequences of length 3 codons with at least 20 occurrences throughout the yeast mRNAs. Sequences that were characterized by consistently high RD in the consensus (nDThi) were defined as those with at least 16 occurrences

above the 90% threshold and at most 4 occurrences below the 10% threshold. Sequences that were characterized by consistently low RD in the consensus (nDTlo) were defined as those with at most 4 occurrences above the 90% threshold and at least 16 occurrences below the 10% threshold. Codon frequencies were compared to the background frequency of the mRNAs with sufficient coverage that passed our quality control (see above).

### **2.5.6 Analysis of DT sequences**

To test whether known sequence features of translation attenuation contribute to DT sequences, we analyzed a link to nonoptimal codon clusters, RNA folding strength, and clusters of positively charged amino acids. Nonoptimal codon clusters were defined as at least 2 nonoptimal codons in a stretch of length 5 translated within 2 codons, i.e. just before the current subsequence. Clusters of positively charged amino acids were defined as polypeptide stretches of length 8 amino acids containing at least 3 positive charges (Arg, Lys, His) with no negatively charged amino acids in between, and placed up to 35 sequence positions before, i.e. concurrently present in the ribosome exit tunnel. High RNA folding strength that can hamper ribosome translocation was tested by evaluating positive PARS scores (Kertesz *et al.*, 2010) 4 codons downstream (Yang *et al.*, 2014). Fisher's exact test was used to test for statistical associations between the presence of one of these features and the category of fast and slowly translated subsequences. Domain boundaries of all yeast proteins were retrieved from (Malmström *et al.*, 2007), and binding sites of the yeast Hsp70 SSB were downloaded from (Doerfel *et al.*, 2013). Average profiles of the 70 codons upstream regions of SSB binding sites or domain boundaries were computed as per-position averages of the consensus RD profiles, and compared to averages of shuffled profiles of the defined regions. Gene ontology (GO) analyses were performed with the g:Profiler webserver (Raudvere *et al.*, 2019; Reimand *et al.*, 2007) through their Python API. All GO terms with an enrichment at  $p < 0.001$  for at least one input gene list were reported.

## **2.6 Acknowledgments**

The authors are grateful for funding through a Discovery grant from the Natural Sciences and Engineering Research Council of Canada. S.P. holds the Canada Research Chair in Computational Systems Biology.

### 3. Conclusions et perspectives

En guise de conclusion, la traduction est une étape importante de la régulation génétique, principalement car il s'agit de l'étape la plus coûteuse en termes d'énergie de la cellule (Lindqvist *et al.*, 2018). Logiquement, elle se voit donc minutieusement régulée afin d'optimiser et modérer l'utilisation de ces ressources énergétiques. De plus, la présence de multiples mécanismes de régulation, principalement dans sa phase d'élongation, lui confère une position privilégiée à l'adaptation rapide, agile et de façon localisée de l'expression génétique (C. Chan *et al.*, 2018; C. T. Y. Chan *et al.*, 2012; Deng *et al.*, 2015; Fernández-Vázquez *et al.*, 2013). Cela a d'ailleurs été démontré sous des conditions de carence en acides aminés. Sous un stress comme ce dernier, les cellules reposent sur la régulation de la traduction afin de regagner la balance de la production protéique, notamment en affectant la vitesse ribosomale. Par exemple, des conditions de déficience d'arginine causent une pause de la translocation des ribosomes à deux des six codons d'arginines, contrairement à la carence en leucine qui entraîne un changement négligeable dans la cinétique ribosomale (Darnell *et al.*, 2018).

Comme le démontrent certaines études, notamment la nôtre, un des principaux régulateurs de l'efficacité de l'élongation de la traduction semble être le choix de codons sur le transcrit d'ARNm (Hanson & Collier, 2018). Ces derniers présentent différents taux de vitesse de décodage, intimement liée à l'abondance en ARNs de transfert environnants ainsi qu'à leurs interactions avec la machinerie de traduction (Brule & Grayhack, 2017). Ces mécanismes sont également tous sujets à des modifications épitranscriptomiques puisqu'ils impliquent des ARNs (Dominissini & Rechavi, 2018). Nos observations de séquences statiques et identiques du point de vue moléculaire, bien que différenciellement traduites (DT), viennent supporter ce modèle de la régulation hétérogène de la traduction. En effet, ces séquences sont principalement enrichies dans les mécanismes de biosynthèse du ribosome ainsi que dans la traduction localisée chez des organelles spécifiques de la cellule comme le ER et la mitochondrie. Les signaux de RD des séquences DT diffèrent le plus pour une même séquence chez des gènes dont l'annotation ontologique correspond à ces deux organelles. Les séquences DT se composent également d'un taux élevé de codons s'associent fréquemment à des ARNs modifiés. Cela évoque cette couche additionnelle de régulation de l'expression génétique et illustre un fascinant dynamisme et une hétérogénéité de la traduction.

Ce qui permet l'ampleur des récentes avancées dans l'étude de la régulation de la traduction est le développement de nouvelles technologies de pointe, notamment le Ribo-Seq. Cette technique a démontré plusieurs fois sa capacité à faire des découvertes fondamentales de mécanismes qui régissent la cinétique de la traduction (Ingolia, 2016). Néanmoins, plusieurs défis en affectent toujours l'analyse et l'interprétation des données quelle qu'elle soit (Husmann *et al.*, 2015). Nos résultats démontrent qu'une analyse comparative de ces données en agrégat parvient à mieux isoler le signal biologique du bruit technique, soit qu'une fraction de ses bruits serait en évidence causée par des composantes de l'hétérogénéité biologique de la traduction. Cette agglomération d'expériences apporte une meilleure distinction et une plus grande confiance dans la présence des signaux de ralentissement du ribosome. Cela est démontré par l'association de nos résultats avec des facteurs de la traduction connus, tels les forces de repliement des ARNm, la charge protéique et les mécanismes d'interactions avec les ARNts. Ceci suggère qu'un séquençage plus en profondeur permettrait de limiter davantage l'incertitude liée aux bruits techniques du Ribo-Seq. De nouvelles technologies de séquençage profond telles que *10X Sequencing* (Zheng *et al.*, 2017) ou encore le séquençage ciblant des régions plus spécifiques de sous-section du génome (Korkmazhan *et al.*, 2017), se révéleraient d'intéressantes pistes de solutions à ces problèmes. Nos résultats démontrent également l'intérêt d'étudier des mécanismes dynamiques de la régulation de la traduction, tels que sa localisation (Medioni *et al.*, 2012) et sa spécialisation (Filipovska & Rackham, 2013), notamment dans le cas des complexes de traduction propres au réticulum endoplasmique (Jan *et al.*, 2014). Il serait intéressant d'explorer plus en profondeur leurs effets sur le repliement protéique ou encore la formation de complexes de manière cotraductionnelle à la traduction (Shiber *et al.*, 2018). De meilleurs modèles computationnels prenant en compte les connaissances récemment acquises sur les différents facteurs agissant sur la régulation de la traduction permettraient de réduire davantage le bruit et rendre les analyses plus précises. L'application de techniques d'apprentissage profond avec des paramètres correspondants en est un exemple (Zhang *et al.*, 2017). Nos résultats ouvrent également la voie à de futures expériences cherchant à mieux cibler la nature de l'hétérogénéité de la traduction. Par exemple, de démêler les rôles des modifications d'ARN, en étudiant de façon individuelle le rôle de chacune sur la cinétique de la traduction, mais aussi sous un angle plus holistique, afin de mieux tracer les réseaux formés par leurs interactions (Linder & Jaffrey, 2019). Aussi, d'étude comment ces différentes composantes de régulation de la traduction réagiraient à des changements dans les conditions

cellulaires, plus et moins exigeantes. Une autre voie de recherche inspirée de nos résultats serait de confirmer un rôle définitif de l'élongation de la traduction dans la synthèse protéique. Par exemple, par la mutation de codons synonymes dans des gènes normalement corégulés et l'observation du maintien ou non de cette corégulation.

En somme, la traduction est un processus vastement complexe et d'importance fondamentale à la vitalité et au dynamisme nécessaire à la survie cellulaire. Cela en fait une composante importante de l'évolution et du maintien de toutes espèces vivantes. L'avènement de nouvelles technologies ouvre une fenêtre d'opportunités pour mieux la comprendre et ainsi augmenter nos connaissances de la biologie moléculaire. La nature exacte de l'hétérogénéité de la traduction ne fait que commencer à être tracée. Les concepts de spécialisation des ribosomes, les modifications aux bases nucléotidiques des ARNs et le transport cellulaire ne sont qu'à leurs débuts et ont encore beaucoup à dévoiler sur l'intrication et les connexions du système cellulaire. Il est fascinant de pouvoir développer de nouvelles techniques permettant leur étude et d'espérer un jour pouvoir élucider le fonctionnement si complexe de la cellule. Une meilleure compréhension du fonctionnement de la traduction a le potentiel de révéler de nouveaux concepts aidant à mieux comprendre la biologie fondamentale, ainsi qu'à la prévention et l'élucidation des nombreuses maladies causés par ses dérégulations.

## Références bibliographiques

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., & Et, A. (1991). Complementary dna sequencing: Expressed sequence tags and human genome project. *Science*, *252*(5013), 1651–1656. <https://doi.org/10.1126/science.2047873>
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., ... Searle, S. M. J. (2016). The ensembl gene annotation system. *Database*, *2016*. <https://doi.org/10.1093/database/baw093>
- Al-Shahrour, F., Díaz-Uriarte, R., & Dopazo, J. (2004). FatiGO: A web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics*, *20*(4), 578–580. <https://doi.org/10.1093/bioinformatics/btg455>
- Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., Cormican, P., Morris, D. W., Shatsky, I. N., & Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eif2 repression. *ELife*, *4*, e03971. <https://doi.org/10.7554/eLife.03971>
- Artieri, C. G., & Fraser, H. B. (2014). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Research*, *24*(12), 2011–2021. <https://doi.org/10.1101/gr.175893.114>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. <https://doi.org/10.1038/75556>
- Attrill, H., Gaudet, P., Huntley, R. P., Lovering, R. C., Engel, S. R., Poux, S., Van Auken, K. M., Georghiou, G., Chibucos, M. C., Berardini, T. Z., Wood, V., Drabkin, H., Fey, P., Garmiri, P., Harris, M. A., Sawford, T., Reiser, L., Tauber, R., & Toro, S. (2019). Annotation of gene product function from high-throughput studies using the gene ontology. *Database*, *2019*. <https://doi.org/10.1093/database/baz007>
- Balch, W. E., Morimoto, R. I., Dillin, A., & Kelly, J. W. (2008). Adapting proteostasis for disease intervention. *Science*, *319*(5865), 916–919. <https://doi.org/10.1126/science.1141448>
- Barbosa, C., Peixeiro, I., & Romão, L. (2013). Gene expression regulation by upstream open reading frames and human disease. *PLOS Genetics*, *9*(8), e1003529. <https://doi.org/10.1371/journal.pgen.1003529>
- Bartholomäus, A., Del, C. C., & Ignatova, Z. (2015). Mapping the non-standardized biases of ribosome profiling. *Biological Chemistry*, *397*(1), 23–35. <https://doi.org/10.1515/hsz-2015-0197>



- Beaupere, C., Wasko, B. M., Lorusso, J., Kennedy, B. K., Kaeberlein, M., & Labunskyy, V. M. (2017). CAN1 arginine permease deficiency extends yeast replicative lifespan via translational activation of stress response genes. *Cell Reports*, *18*(8), 1884–1892. <https://doi.org/10.1016/j.celrep.2017.01.077>
- Bednářová, A., Hanna, M., Durham, I., VanCleave, T., England, A., Chaudhuri, A., & Krishnan, N. (2017). Lost in translation: Defects in transfer rna modifications and neurological disorders. *Frontiers in Molecular Neuroscience*, *10*. <https://doi.org/10.3389/fnmol.2017.00135>
- Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). Protein structure and function. *Biochemistry*. 5th Edition. <https://www.ncbi.nlm.nih.gov/books/NBK21177/>
- Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, *24*(4), 335–341. <https://doi.org/10.1016/j.cmi.2017.10.013>
- Biswas, R., Khamaru, K., & Majumdar, K. K. (2014). A peak synchronization measure for multiple signals. *IEEE Transactions on Signal Processing*, *62*(17), 4390–4398. <https://doi.org/10.1109/TSP.2014.2333568>
- Björk, G. R., Jacobsson, K., Nilsson, K., Johansson, M. J. O., Byström, A. S., & Persson, O. P. (2001). A primordial trna modification required for the evolution of life? *The EMBO Journal*, *20*(1–2), 231–239. <https://doi.org/10.1093/emboj/20.1.231>
- Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., de Crécy-Lagard, V., Ross, R., Limbach, P. A., Kotter, A., Helm, M., & Bujnicki, J. M. (2018). MODOMICS: A database of rna modification pathways. 2017 update. *Nucleic Acids Research*, *46*(D1), D303–D307. <https://doi.org/10.1093/nar/gkx1030>
- Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology*, *16*(11), 651–664. <https://doi.org/10.1038/nrm4069>
- Brule, C. E., & Grayhack, E. J. (2017). Synonymous codons: Choose wisely for expression. *Trends in Genetics*, *33*(4), 283–297. <https://doi.org/10.1016/j.tig.2017.02.001>
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., & Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nature Methods*, *13*(2), 165–170. <https://doi.org/10.1038/nmeth.3688>
- Carja, O., Xing, T., Wallace, E. W. J., Plotkin, J. B., & Shah, P. (2017). Riboviz: Analysis and visualization of ribosome profiling datasets. *BMC Bioinformatics*, *18*(1), 461. <https://doi.org/10.1186/s12859-017-1873-8>
- Catalanotto, C., Cogoni, C., & Zardo, G. (2016). MicroRNA in control of gene expression: An overview of nuclear functions. *International Journal of Molecular Sciences*, *17*(10), 1712. <https://doi.org/10.3390/ijms17101712>

Chan, C., Pham, P., Dedon, P. C., & Begley, T. J. (2018). Lifestyle modifications: Coordinating the trna epitranscriptome with codon bias to adapt translation during stress responses. *Genome Biology*, *19*(1), 228. <https://doi.org/10.1186/s13059-018-1611-1>

Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., & Dedon, P. C. (2012). Reprogramming of trna modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications*, *3*(1), 1–9. <https://doi.org/10.1038/ncomms1938>

Chaney, J. L., & Clark, P. L. (2015). Roles for synonymous codon usage in protein biogenesis. *Annual Review of Biophysics*, *44*(1), 143–166. <https://doi.org/10.1146/annurev-biophys-060414-034333>

Charneski, C. A., & Hurst, L. D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLOS Biology*, *11*(3), e1001508. <https://doi.org/10.1371/journal.pbio.1001508>

Chartron, J. W., Hunt, K. C. L., & Frydman, J. (2016). Cotranslational signal-independent srp preloading during membrane targeting. *Nature*, *536*(7615), 224–228. <https://doi.org/10.1038/nature19309>

Chekulaeva, M., & Landthaler, M. (2016). Eyes on translation. *Molecular Cell*, *63*(6), 918–925. <https://doi.org/10.1016/j.molcel.2016.08.031>

Chen, P., Jäger, G., & Zheng, B. (2010). Transfer rna modifications and genes for modifying enzymes in arabidopsis thaliana. *BMC Plant Biology*, *10*(1), 201. <https://doi.org/10.1186/1471-2229-10-201>

Cheng, J., North, B. J., Zhang, T., Dai, X., Tao, K., Guo, J., & Wei, W. (2018). The emerging roles of protein homeostasis-governing pathways in alzheimer's disease. *Aging Cell*, *17*(5). <https://doi.org/10.1111/accel.12801>

Chiti, F., & Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annual Review of Biochemistry*, *86*(1), 27–68. <https://doi.org/10.1146/annurev-biochem-061516-045115>

Choi, J., Grosely, R., Prabhakar, A., Lapointe, C. P., Wang, J., & Puglisi, J. D. (2018). How messenger rna and nascent chain sequences regulate translation elongation. *Annual Review of Biochemistry*, *87*(1), 421–449. <https://doi.org/10.1146/annurev-biochem-060815-014818>

Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *The EMBO Journal*, *33*(1), 21–34. <https://doi.org/10.1002/embj.201385651>

Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., & Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for ribo-seq data analysis. *RNA*, *21*(10), 1731–1745. <https://doi.org/10.1261/rna.052548.115>

Clauwaert, J., Menschaert, G., & Waegeman, W. (2018). DeepRibo: Precise gene annotation of prokaryotes using deep learning and ribosome profiling data. *BioRxiv*, 317180. <https://doi.org/10.1101/317180>

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for rna-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>

Corley, S. M., MacKenzie, K. L., Beverdam, A., Roddam, L. F., & Wilkins, M. R. (2017). Differentially expressed genes from rna-seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, 18(1), 399. <https://doi.org/10.1186/s12864-017-3797-0>

Couvillion, M. T., Soto, I. C., Shipkovenska, G., & Churchman, L. S. (2016). Synchronized mitochondrial and cytosolic translation programs. *Nature*, 533(7604), 499–503. <https://doi.org/10.1038/nature18015>

Crain, P. F., & McCloskey, J. A. (1997). The rna modification database. *Nucleic Acids Research*, 25(1), 126–127. <https://doi.org/10.1093/nar/25.1.126>

Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227, 3.

Dana, A., & Tuller, T. (2014). The effect of trna levels on decoding times of mrna codons. *Nucleic Acids Research*, 42(14), 9171–9181. <https://doi.org/10.1093/nar/gku646>

Darnell, A. M., Subramaniam, A. R., & O’Shea, E. K. (2018). Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Molecular Cell*, 71(2), 229–243.e11. <https://doi.org/10.1016/j.molcel.2018.06.041>

Das, R. K., Ruff, K. M., & Pappu, R. V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 32, 102–112. <https://doi.org/10.1016/j.sbi.2015.03.008>

de Klerk, E., & ‘t Hoen, P. A. C. (2015). Alternative mrna transcription, processing, and translation: Insights from rna sequencing. *Trends in Genetics*, 31(3), 128–139. <https://doi.org/10.1016/j.tig.2015.01.001>

Decatur, W. A., & Fournier, M. J. (2002). RRNA modifications and ribosome function. *Trends in Biochemical Sciences*, 27(7), 344–351. [https://doi.org/10.1016/S0968-0004\(02\)02109-6](https://doi.org/10.1016/S0968-0004(02)02109-6)

Deng, W., Babu, I. R., Su, D., Yin, S., Begley, T. J., & Dedon, P. C. (2015). Trm9-catalyzed trna modifications regulate global protein expression by codon-biased translation. *PLOS Genetics*, 11(12), e1005706. <https://doi.org/10.1371/journal.pgen.1005706>

Dever, T. E., Dinman, J. D., & Green, R. (2018). Translation elongation and recoding in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, a032649. <https://doi.org/10.1101/cshperspect.a032649>

- Diament, A., & Tuller, T. (2016). Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biology Direct*, *11*(1), 24. <https://doi.org/10.1186/s13062-016-0127-4>
- Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, *338*(6110), 1042–1046. <https://doi.org/10.1126/science.1219021>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal rna-seq aligner. *Bioinformatics*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Doerfel, L. K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., & Rodnina, M. V. (2013). EF-p is essential for rapid synthesis of proteins containing consecutive proline residues. *Science*, *339*(6115), 85–88. <https://doi.org/10.1126/science.1229017>
- Dominianni, D., & Rechavi, G. (2018, September 28). *Epitranscriptome regulation* [News]. Nature Structural & Molecular Biology. <https://doi.org/10.1038/s41594-018-0140-7>
- Duc, K. D., & Song, Y. S. (2018). Correction: The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLOS Genetics*, *14*(8), e1007620. <https://doi.org/10.1371/journal.pgen.1007620>
- Duncan, C. D. S., & Mata, J. (2017). Effects of cycloheximide on the interpretation of ribosome profiling experiments in *Schizosaccharomyces pombe*. *Scientific Reports*, *7*(1), 10331. <https://doi.org/10.1038/s41598-017-10650-1>
- Duttler, S., Pechmann, S., & Frydman, J. (2013). Principles of cotranslational ubiquitination and quality control at the ribosome. *Molecular Cell*, *50*(3), 379–393. <https://doi.org/10.1016/j.molcel.2013.03.010>
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics*, *10*(1), 48. <https://doi.org/10.1186/1471-2105-10-48>
- El Yacoubi, B., Bailly, M., & de Crécy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer rnas. *Annual Review of Genetics*, *46*(1), 69–95. <https://doi.org/10.1146/annurev-genet-110711-155641>
- Eliyahu, E., Lesnik, C., & Arava, Y. (2012). The protein chaperone ssa1 affects mrna localization to the mitochondria. *FEBS Letters*, *586*(1), 64–69. <https://doi.org/10.1016/j.febslet.2011.11.025>
- Engel, S. R., Dietrich, F. S., Fisk, D. G., Binkley, G., Balakrishnan, R., Costanzo, M. C., Dwight, S. S., Hitz, B. C., Karra, K., Nash, R. S., Weng, S., Wong, E. D., Lloyd, P., Skrzypek, M. S., Miyasato, S. R., Simison, M., & Cherry, J. M. (2014). The reference genome sequence of *Saccharomyces cerevisiae*: Then and now. *G3: Genes, Genomes, Genetics*, *4*(3), 389–398. <https://doi.org/10.1534/g3.113.008995>

- Fang, H., Huang, Y.-F., Radhakrishnan, A., Siepel, A., Lyon, G. J., & Schatz, M. C. (2018). Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Systems*, 6(2), 180-191.e4. <https://doi.org/10.1016/j.cels.2017.12.007>
- Fernández-Vázquez, J., Vargas-Pérez, I., Sansó, M., Buhne, K., Carmona, M., Paulo, E., Hermand, D., Rodríguez-Gabriel, M., Ayté, J., Leidel, S., & Hidalgo, E. (2013). Modification of trnalysuuu by elongator is essential for efficient translation of stress mrnas. *PLOS Genetics*, 9(7), e1003647. <https://doi.org/10.1371/journal.pgen.1003647>
- Filipovska, A., & Rackham, O. (2013). Specialization from synthesis: How ribosome diversity can customize protein function. *FEBS Letters*, 587(8), 1189–1197. <https://doi.org/10.1016/j.febslet.2013.02.032>
- Franzmann, T. M., & Alberti, S. (2019). Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *Journal of Biological Chemistry*, 294(18), 7128–7136. <https://doi.org/10.1074/jbc.TM118.001190>
- Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., & Grayhack, E. J. (2016). Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*, 166(3), 679–690. <https://doi.org/10.1016/j.cell.2016.05.070>
- García-Sánchez, A., & Marqués-García, F. (2016). Review of methods to study gene expression regulation applied to asthma. In M. Isidoro García (Ed.), *Molecular Genetics of Asthma* (pp. 71–89). Springer. [https://doi.org/10.1007/978-1-4939-3652-6\\_6](https://doi.org/10.1007/978-1-4939-3652-6_6)
- Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., & Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *ELife*, 3, e03735. <https://doi.org/10.7554/eLife.03735>
- Geller, R., Pechmann, S., Acevedo, A., Andino, R., & Frydman, J. (2018). Hsp90 shapes protein and rna evolution to balance trade-offs between protein stability and aggregation. *Nature Communications*, 9(1), 1–11. <https://doi.org/10.1038/s41467-018-04203-x>
- Gerashchenko, M. V., & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Research*, 42(17), e134–e134. <https://doi.org/10.1093/nar/gku671>
- Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Research*, 45(2), e6–e6. <https://doi.org/10.1093/nar/gkw822>
- Giorgi, C., & Moore, M. J. (2007). The nuclear nurture and cytoplasmic nature of localized mrnps. *Seminars in Cell & Developmental Biology*, 18(2), 186–193. <https://doi.org/10.1016/j.semcdb.2007.01.002>
- Gloge, F., Becker, A. H., Kramer, G., & Bukau, B. (2014). Co-translational mechanisms of protein maturation. *Current Opinion in Structural Biology*, 24, 24–33. <https://doi.org/10.1016/j.sbi.2013.11.004>

- Gobet, C., Weger, B., Marquis, J., Martin, E., Gachon, F., & Naef, F. (2019). Landscapes of ribosome dwell times and relationship with aminoacyl-trna levels in mammals. *BioRxiv*, 551838. <https://doi.org/10.1101/551838>
- Goodarzi, H., Nguyen, H. C. B., Zhang, S., Dill, B. D., Molina, H., & Tavazoie, S. F. (2016). Modulated expression of specific trnas drives gene expression and cancer progression. *Cell*, 165(6), 1416–1427. <https://doi.org/10.1016/j.cell.2016.05.046>
- Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). Microarray and its applications. *Journal of Pharmacy And Bioallied Sciences*, 4(6), 310. <https://doi.org/10.4103/0975-7406.100283>
- Grada, A. (2013). *Next-generation sequencing: Methodology and application*. 4.
- Granneman, S., & Tollervy, D. (2007). Building ribosomes: Even more expensive than expected? *Current Biology*, 17(11), R415–R417. <https://doi.org/10.1016/j.cub.2007.04.011>
- Gritsenko, A. A., Hulsman, M., Reinders, M. J. T., & Ridder, D. de. (2015). Unbiased quantitative models of protein translation derived from ribosome profiling data. *PLOS Computational Biology*, 11(8), e1004336. <https://doi.org/10.1371/journal.pcbi.1004336>
- Grosjean, H., Crécy-Lagard, V. de, & Marck, C. (2010). Deciphering synonymous codons in the three domains of life: Co-evolution with specific trna modification enzymes. *FEBS Letters*, 584(2), 252–264. <https://doi.org/10.1016/j.febslet.2009.11.052>
- Guo, H. (2018). Specialized ribosomes and the control of translation. *Biochemical Society Transactions*, 46(4), 855–869. <https://doi.org/10.1042/BST20160426>
- Guydosh, N. R., & Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell*, 156(5), 950–962. <https://doi.org/10.1016/j.cell.2014.02.006>
- Hafner, A.-S., Donlin-Asp, P. G., Leitch, B., Herzog, E., & Schuman, E. M. (2019). Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. *Science*, 364(6441). <https://doi.org/10.1126/science.aau3644>
- Halpin, J. C., Jangi, R., & Street, T. O. (2020). Multimapping confounds ribosome profiling analysis: A case-study of the hsp90 molecular chaperone. *Proteins: Structure, Function, and Bioinformatics*, 88(1), 57–68. <https://doi.org/10.1002/prot.25766>
- Han, Yan, David, A., Liu, B., Magadán, J. G., Bennink, J. R., Yewdell, J. W., & Qian, S.-B. (2012). Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proceedings of the National Academy of Sciences*, 109(31), 12467–12472. <https://doi.org/10.1073/pnas.1208138109>
- Han, Yixing, Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced applications of rna sequencing and challenges: *Bioinformatics and Biology Insights*. <https://doi.org/10.4137/BBI.S28991>

- Hanson, G., & Collier, J. (2018). Codon optimality, bias and usage in translation and mrna decay. *Nature Reviews Molecular Cell Biology*, 19(1), 20–30. <https://doi.org/10.1038/nrm.2017.91>
- Hetz, C., & Glimcher, L. H. (2011). Protein homeostasis networks in physiology and disease. *Current Opinion in Cell Biology*, 23(2), 123–125. <https://doi.org/10.1016/j.ceb.2011.01.004>
- Hocking, T. D., Goerner-Potvin, P., Morin, A., Shao, X., Pastinen, T., & Bourque, G. (2017). Optimizing chip-seq peak detectors using visual labels and supervised machine learning. *Bioinformatics*, 33(4), 491–499. <https://doi.org/10.1093/bioinformatics/btw672>
- Holt, C. E., Martin, K. C., & Schuman, E. M. (2019). Local translation in neurons: Visualization and function. *Nature Structural & Molecular Biology*, 26(7), 557–566. <https://doi.org/10.1038/s41594-019-0263-5>
- Hopper, A. K. (2013). Transfer rna post-transcriptional processing, turnover, and subcellular dynamics in the yeast *saccharomyces cerevisiae*. *Genetics*, 194(1), 43–67. <https://doi.org/10.1534/genetics.112.147470>
- Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-seq methods for transcriptome analysis. *WIREs RNA*, 8(1), e1364. <https://doi.org/10.1002/wrna.1364>
- Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H. C., & Lempicki, R. A. (2007). The david gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, 8(9), R183. <https://doi.org/10.1186/gb-2007-8-9-r183>
- Hui, A., & Boer, H. A. de. (1987). Specialized ribosome system: Preferential translation of a single mrna species by a subpopulation of mutated ribosomes in *escherichia coli*. *Proceedings of the National Academy of Sciences*, 84(14), 4762–4766. <https://doi.org/10.1073/pnas.84.14.4762>
- Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S., & Press, W. H. (2015). Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLOS Genetics*, 11(12), e1005732. <https://doi.org/10.1371/journal.pgen.1005732>
- Ibrahim, M. M., Lacadie, S. A., & Ohler, U. (2015). JAMM: A peak finder for joint analysis of ngs replicates. *Bioinformatics*, 31(1), 48–55. <https://doi.org/10.1093/bioinformatics/btu568>
- Ingolia, N. T. (2014). Ribosome profiling: New views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3), 205–213. <https://doi.org/10.1038/nrg3645>
- Ingolia, N. T. (2016). Ribosome footprint profiling of translation throughout the genome. *Cell*, 165(1), 22–33. <https://doi.org/10.1016/j.cell.2016.02.066>
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mrna fragments. *Nature Protocols*, 7(8), 1534–1550. <https://doi.org/10.1038/nprot.2012.086>

- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), 218–223. <https://doi.org/10.1126/science.1168978>
- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789–802. <https://doi.org/10.1016/j.cell.2011.10.002>
- Jakobsson, M. E., Małeckı, J., & Falnes, P. Ø. (2018). Regulation of eukaryotic elongation factor 1 alpha (eef1a) by dynamic lysine methylation. *RNA Biology*, 15(3), 314–319. <https://doi.org/10.1080/15476286.2018.1440875>
- Jan, C. H., Williams, C. C., & Weissman, J. S. (2014). Principles of cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, 346(6210), 1257521. <https://doi.org/10.1126/science.1257521>
- Jiang, S., & Mortazavi, A. (2018). Integrating chip-seq with other functional genomics data. *Briefings in Functional Genomics*, 17(2), 104–115. <https://doi.org/10.1093/bfgp/ely002>
- Jores, T., Lawatscheck, J., Beke, V., Franz-Wachtel, M., Yunoki, K., Fitzgerald, J. C., Macek, B., Endo, T., Kalbacher, H., Buchner, J., & Rapaport, D. (2018). Cytosolic hsp70 and hsp40 chaperones enable the biogenesis of mitochondrial  $\beta$ -barrel proteins. *The Journal of Cell Biology*, 217(9), 3091–3108. <https://doi.org/10.1083/jcb.201712029>
- Kafri, M., Metzl-Raz, E., Jona, G., & Barkai, N. (2016). The cost of protein production. *Cell Reports*, 14(1), 22–31. <https://doi.org/10.1016/j.celrep.2015.12.015>
- Kanai, A. (2014). Welcome to the new trna world! *Frontiers in Genetics*, 5. <https://doi.org/10.3389/fgene.2014.00336>
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010). Genome-wide measurement of rna secondary structure in yeast. *Nature*, 467(7311), 103–107. <https://doi.org/10.1038/nature09322>
- Khatter, H., Myasnikov, A. G., Natchiar, S. K., & Klaholz, B. P. (2015). Structure of the human 80s ribosome. *Nature*, 520(7549), 640–645. <https://doi.org/10.1038/nature14427>
- Kim, H.-L., Song, W.-S., Kim, K., & Lee, K. (2008). Characterization of heterogeneous lsu rna profiles in streptomyces coelicolor under different growth stages and conditions. *Current Microbiology*, 57(6), 537–541. <https://doi.org/10.1007/s00284-008-9238-1>
- Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Ulrich Hartl, F. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry*, 82(1), 323–355. <https://doi.org/10.1146/annurev-biochem-060208-092442>
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., Hayashizaki, Y., & Carninci, P. (2006). CAGE: Cap analysis of gene expression. *Nature Methods*, 3(3), 211–222. <https://doi.org/10.1038/nmeth0306-211>



- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell rna sequencing. *Molecular Cell*, *58*(4), 610–620. <https://doi.org/10.1016/j.molcel.2015.04.005>
- Komili, S., & Silver, P. A. (2008). Coupling and coordination in gene expression processes: A systems biology view. *Nature Reviews Genetics*, *9*(1), 38–48. <https://doi.org/10.1038/nrg2223>
- Korkmazhan, E., Teimouri, H., Peterman, N., & Levine, E. (2017). Dynamics of translation can determine the spatial organization of membrane-bound proteins and their mrna. *Proceedings of the National Academy of Sciences*, *114*(51), 13424–13429. <https://doi.org/10.1073/pnas.1700941114>
- Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., & Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. *Nature*, *419*(6903), 171–174. <https://doi.org/10.1038/nature01047>
- Krichevsky, A. M., & Kosik, K. S. (2001). Neuronal rna granules: A link between rna localization and stimulation-dependent translation. *Neuron*, *32*(4), 683–696. [https://doi.org/10.1016/S0896-6273\(01\)00508-6](https://doi.org/10.1016/S0896-6273(01)00508-6)
- Kumar, V., Muratani, M., Rayan, N. A., Kraus, P., Lufkin, T., Ng, H. H., & Prabhakar, S. (2013). Uniform, optimal signal processing of mapped deep-sequencing data. *Nature Biotechnology*, *31*(7), 615–622. <https://doi.org/10.1038/nbt.2596>
- Lareau, L. F., Hite, D. H., Hogan, G. J., & Brown, P. O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mrna fragments. *ELife*, *3*, e01257. <https://doi.org/10.7554/eLife.01257>
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the oxford nanopore technologies minion. *Biomolecular Detection and Quantification*, *3*, 1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>
- Lecanda, A., Nilges, B. S., Sharma, P., Nedialkova, D. D., Schwarz, J., Vaquerizas, J. M., & Leidel, S. A. (2016). Dual randomization of oligonucleotides to reduce the bias in ribosome-profiling libraries. *Methods*, *107*, 89–97. <https://doi.org/10.1016/j.ymeth.2016.07.011>
- Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., & Qian, S.-B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, *109*(37), E2424–E2432. <https://doi.org/10.1073/pnas.1207846109>
- Legendre, R., Baudin-Baillieu, A., Hatin, I., & Namy, O. (2015). RiboTools: A galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics*, *31*(15), 2586–2588. <https://doi.org/10.1093/bioinformatics/btv174>
- Lempiäinen, H., & Shore, D. (2009). Growth control and ribosome biogenesis. *Current Opinion in Cell Biology*, *21*(6), 855–863. <https://doi.org/10.1016/j.ceb.2009.09.002>
- Lestas, I., Vinnicombe, G., & Paulsson, J. (2010). Fundamental limits on the suppression of molecular fluctuations. *Nature*, *467*(7312), 174–178. <https://doi.org/10.1038/nature09333>

- Linder, B., & Jaffrey, S. R. (2019). Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mrna. *Cold Spring Harbor Perspectives in Biology*, 11(6), a032201. <https://doi.org/10.1101/cshperspect.a032201>
- Lindqvist, L. M., Tandoc, K., Topisirovic, I., & Furic, L. (2018). Cross-talk between protein synthesis, energy metabolism and autophagy in cancer. *Current Opinion in Genetics & Development*, 48, 104–111. <https://doi.org/10.1016/j.gde.2017.11.003>
- Lorenz, C., Lünse, C. E., & Mörl, M. (2017). tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules*, 7(2), 35. <https://doi.org/10.3390/biom7020035>
- Lowe, T. M., & Eddy, S. R. (1999). A computational screen for methylation guide snoras in yeast. *Science*, 283(5405), 1168–1171. <https://doi.org/10.1126/science.283.5405.1168>
- Lu, J., & Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. *Journal of Molecular Biology*, 384(1), 73–86. <https://doi.org/10.1016/j.jmb.2008.08.089>
- Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mrna and protein in complex biological samples. *FEBS Letters*, 583(24), 3966–3973. <https://doi.org/10.1016/j.febslet.2009.10.036>
- Malmström, L., Riffle, M., Strauss, C. E. M., Chivian, D., Davis, T. N., Bonneau, R., & Baker, D. (2007). Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLOS Biology*, 5(4), e76. <https://doi.org/10.1371/journal.pbio.0050076>
- Mardis, E. R. (2011). A decade's perspective on dna sequencing technology. *Nature*, 470(7333), 198–203. <https://doi.org/10.1038/nature09796>
- Martens, A. T., Taylor, J., & Hilser, V. J. (2015). Ribosome a and p sites revealed by length analysis of ribosome profiling data. *Nucleic Acids Research*, 43(7), 3680–3687. <https://doi.org/10.1093/nar/gkv200>
- Martin, E. W., & Mittag, T. (2018). Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry*, 57(17), 2478–2487. <https://doi.org/10.1021/acs.biochem.8b00008>
- Martin, K. C., & Ephrussi, A. (2009). mRNA localization: Gene expression in the spatial dimension. *Cell*, 136(4), 719. <https://doi.org/10.1016/j.cell.2009.01.044>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Matsuo, Y., Ikeuchi, K., Saeki, Y., Iwasaki, S., Schmidt, C., Udagawa, T., Sato, F., Tsuchiya, H., Becker, T., Tanaka, K., Ingolia, N. T., Beckmann, R., & Inada, T. (2017). Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nature Communications*, 8(1), 159. <https://doi.org/10.1038/s41467-017-00188-1>
- Medioni, C., Mowry, K., & Besse, F. (2012). Principles and roles of mrna localization in animal development. *Development*, 139(18), 3263–3276. <https://doi.org/10.1242/dev.078626>

- MGlincy, N. J., & Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods*, *126*, 112–129. <https://doi.org/10.1016/j.ymeth.2017.05.028>
- Michel, A. M., & Baranov, P. V. (2013). Ribosome profiling: A hi-def monitor for protein synthesis at the genome-wide scale. *Wiley Interdisciplinary Reviews: RNA*, *4*(5), 473–490. <https://doi.org/10.1002/wrna.1172>
- Michel, A. M., Fox, G., M. Kiran, A., De Bo, C., O'Connor, P. B. F., Heaphy, S. M., Mullan, J. P. A., Donohue, C. A., Higgins, D. G., & Baranov, P. V. (2014). GWIPS-viz: Development of a ribo-seq genome browser. *Nucleic Acids Research*, *42*(D1), D859–D864. <https://doi.org/10.1093/nar/gkt1035>
- Michel, A. M., Mullan, J. P. A., Velayudhan, V., O'Connor, P. B. F., Donohue, C. A., & Baranov, P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biology*, *13*(3), 316–319. <https://doi.org/10.1080/15476286.2016.1141862>
- Milo, R., Jorgensen, P., Moran, U., Weber, G., & Springer, M. (2010). BioNumbers—The database of key numbers in molecular and cell biology. *Nucleic Acids Research*, *38*(suppl\_1), D750–D753. <https://doi.org/10.1093/nar/gkp889>
- Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory rna. *Nature Reviews Genetics*, *15*(6), 423–437. <https://doi.org/10.1038/nrg3722>
- Mudge, J. M., & Harrow, J. (2016). The state of play in higher eukaryote gene annotation. *Nature Reviews Genetics*, *17*(12), 758–772. <https://doi.org/10.1038/nrg.2016.119>
- Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I., & Klaholz, B. P. (2017). Visualization of chemical modifications in the human 80s ribosome structure. *Nature*, *551*(7681), 472–477. <https://doi.org/10.1038/nature24482>
- Nedialkova, D. D., & Leidel, S. A. (2015). Optimization of codon translation rates via trna modifications maintains proteome integrity. *Cell*, *161*(7), 1606–1618. <https://doi.org/10.1016/j.cell.2015.05.022>
- Negrutskii, B., Vlasenko, D., Mirande, M., Futernyk, P., & El'skaya, A. (2018). MRNA-independent way to regulate translation elongation rate in eukaryotic cells. *IUBMB Life*, *70*(3), 192–196. <https://doi.org/10.1002/iub.1724>
- Netzer, W. J., & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*, *388*(6640), 343–349. <https://doi.org/10.1038/41024>
- Neuman, N. (2017). An updated translation. *Trends in Biochemical Sciences*, *42*(8), 583–584. <https://doi.org/10.1016/j.tibs.2017.06.003>
- Novoa, E. M., Pavon-Eternod, M., Pan, T., & Ribas de Pouplana, L. (2012). A role for trna modifications in genome structure and codon usage. *Cell*, *149*(1), 202–213. <https://doi.org/10.1016/j.cell.2012.01.050>

- O'Connor, P. B. F., Andreev, D. E., & Baranov, P. V. (2016). Comparative survey of the relative impact of mrna features on local ribosome profiling read density. *Nature Communications*, 7, 12915. <https://doi.org/10.1038/ncomms12915>
- Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R. J., Typas, A., Gross, C. A., Kramer, G., Weissman, J. S., & Bukau, B. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, 147(6), 1295–1308. <https://doi.org/10.1016/j.cell.2011.10.044>
- Pan, T. (2018). Modifications and functional genomics of human transfer rna. *Cell Research*, 28(4), 395–404. <https://doi.org/10.1038/s41422-018-0013-y>
- Paquin, N., & Chartrand, P. (2008). Local regulation of mrna translation: New insights from the bud. *Trends in Cell Biology*, 18(3), 105–111. <https://doi.org/10.1016/j.tcb.2007.12.004>
- Park, H., & Subramaniam, A. R. (2019). Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLOS Biology*, 17(9), e3000396. <https://doi.org/10.1371/journal.pbio.3000396>
- Pechmann, S. (2018). Coping with stress by regulating trnas. *Sci. Signal.*, 11(546), eaau1098. <https://doi.org/10.1126/scisignal.aau1098>
- Pechmann, S., Chartron, J. W., & Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by srp in vivo. *Nature Structural & Molecular Biology*, 21(12), 1100–1105. <https://doi.org/10.1038/nsmb.2919>
- Pechmann, S., & Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural & Molecular Biology*, 20(2), 237–243. <https://doi.org/10.1038/nsmb.2466>
- Pechmann, S., Willmund, F., & Frydman, J. (2013). The ribosome as a hub for protein quality control. *Molecular Cell*, 49(3), 411–421. <https://doi.org/10.1016/j.molcel.2013.01.020>
- Pereira, M., Francisco, S., Varanda, A. S., Santos, M., Santos, M. A. S., & Soares, A. R. (2018). Impact of trna modifications and trna-modifying enzymes on proteostasis and human disease. *International Journal of Molecular Sciences*, 19(12), 3738. <https://doi.org/10.3390/ijms19123738>
- Phizicky, E. M., & Hopper, A. K. (2010). TRNA biology charges to the front. *Genes & Development*, 24(17), 1832–1860. <https://doi.org/10.1101/gad.1956510>
- Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014). Causal signals between codon bias, mrna structure, and the efficiency of translation and elongation. *Molecular Systems Biology*, 10(12), 770. <https://doi.org/10.15252/msb.20145524>
- Poulin, F., & Sonenberg, N. (2013). *Mechanism of translation initiation in eukaryotes*. Landes Bioscience. <https://www.ncbi.nlm.nih.gov/books/NBK6597/>

- Priesnitz, C., & Becker, T. (2018). Pathways to balance mitochondrial translation and protein import. *Genes & Development*, 32(19–20), 1285–1296. <https://doi.org/10.1101/gad.316547.118>
- Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. *Molecular Cell*, 59(2), 149–161. <https://doi.org/10.1016/j.molcel.2015.05.035>
- Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., & Collier, J. (2016). The dead-box protein dhh1p couples mrna decay and translation by monitoring codon optimality. *Cell*, 167(1), 122–132.e9. <https://doi.org/10.1016/j.cell.2016.08.053>
- Raghavan, A., Ogilvie, R. L., Reilly, C., Abelson, M. L., Raghavan, S., Vasdewani, J., Krathwohl, M., & Bohjanen, P. R. (2002). Genome-wide analysis of mrna decay in resting and activated primary human t lymphocytes. *Nucleic Acids Research*, 30(24), 5529–5538. <https://doi.org/10.1093/nar/gkf682>
- Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135(2), 216–226. <https://doi.org/10.1016/j.cell.2008.09.050>
- Ranjan, N., & Leidel, S. A. (2019). The epitranscriptome in translation regulation: mRNA and trna modifications as the two sides of the same coin? *FEBS Letters*, 593(13), 1483–1493. <https://doi.org/10.1002/1873-3468.13491>
- Ranjan, N., & Rodnina, M. V. (2016). tRNA wobble modifications and protein homeostasis. *Translation*, 4(1), e1143076. <https://doi.org/10.1080/21690731.2016.1143076>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198. <https://doi.org/10.1093/nar/gkz369>
- Reck-Peterson, S. L., Tyska, M. J., Novick, P. J., & Mooseker, M. S. (2001). The yeast class v myosins, myo2p and myo4p, are nonprocessive actin-based motors. *The Journal of Cell Biology*, 153(5), 1121–1126. <https://doi.org/10.1083/jcb.153.5.1121>
- Reimand, J., Kull, M., Peterson, H., Hansen, J., & Vilo, J. (2007). G:profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, 35(suppl\_2), W193–W200. <https://doi.org/10.1093/nar/gkm226>
- Reinartz, J., Bruyns, E., Lin, J.-Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M., & Woychik, R. (2002). Massively parallel signature sequencing (mpss) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics*, 1(1), 95–104. <https://doi.org/10.1093/bfgp/1.1.95>
- Reis, M. dos, Savva, R., & Wernisch, L. (2004). Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Research*, 32(17), 5036–5044. <https://doi.org/10.1093/nar/gkh834>

- Reuveni, S., Meilijson, I., Kupiec, M., Ruppim, E., & Tuller, T. (2011). Genome-scale analysis of translation elongation with a ribosome flow model. *PLOS Computational Biology*, 7(9), e1002127. <https://doi.org/10.1371/journal.pcbi.1002127>
- Riba, A., Nanni, N. D., Mittal, N., Arhné, E., Schmidt, A., & Zavolan, M. (2019). Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proceedings of the National Academy of Sciences*, 116(30), 15023–15032. <https://doi.org/10.1073/pnas.1817299116>
- Richter, J. D., & Collier, J. (2015). Pausing on polyribosomes: Make way for elongation in translational control. *Cell*, 163(2), 292–300. <https://doi.org/10.1016/j.cell.2015.09.041>
- Rodnina, M. V., Korniy, N., Klimova, M., Karki, P., Peng, B.-Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E., & Peske, F. (n.d.). Translational recoding: Canonical translation mechanisms reinterpreted. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz783>
- Rodnina, M. V., & Wintermeyer, W. (2016). Protein elongation, co-translational folding and targeting. *Journal of Molecular Biology*, 428(10, Part B), 2165–2185. <https://doi.org/10.1016/j.jmb.2016.03.022>
- Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., & Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biology*, 14(5), R51. <https://doi.org/10.1186/gb-2013-14-5-r51>
- Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic rna modifications in gene expression regulation. *Cell*, 169(7), 1187–1200. <https://doi.org/10.1016/j.cell.2017.05.045>
- Russell, J. B., & Cook, G. M. (1995). Energetics of bacterial growth: Balance of anabolic and catabolic reactions. *Microbiological Reviews*, 59(1), 48–62.
- Sabi, R., & Tuller, T. (2015). A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics*, 16(10), S5. <https://doi.org/10.1186/1471-2164-16-S10-S5>
- Sabi, R., & Tuller, T. (2017). Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in *saccharomyces cerevisiae*. *RNA*, 23(7), 983–994. <https://doi.org/10.1261/rna.059188.116>
- Saletore, Y., Meyer, K., Korlach, J., Vilfan, I. D., Jaffrey, S., & Mason, C. E. (2012). The birth of the epitranscriptome: Deciphering the function of rna modifications. *Genome Biology*, 13(10), 175. <https://doi.org/10.1186/gb-2012-13-10-175>
- Santos, D. A., Shi, L., Tu, B. P., & Weissman, J. S. (2019). Cycloheximide can distort measurements of mrna levels and translation efficiency. *Nucleic Acids Research*, 47(10), 4974–4985. <https://doi.org/10.1093/nar/gkz205>

- Sauert, M., Temmel, H., & Moll, I. (2015). Heterogeneity of the translational machinery: Variations on a common theme. *Biochimie*, *114*, 39–47. <https://doi.org/10.1016/j.biochi.2014.12.011>
- Saunders, R., & Deane, C. M. (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Research*, *38*(19), 6719–6728. <https://doi.org/10.1093/nar/gkq495>
- Schafer, S., Adami, E., Heinig, M., Rodrigues, K. E. C., Kreuchwig, F., Silhavy, J., Heesch, S. van, Simaite, D., Rajewsky, N., Cuppen, E., Pravenec, M., Vingron, M., Cook, S. A., & Hubner, N. (2015). Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nature Communications*, *6*(1), 1–9. <https://doi.org/10.1038/ncomms8200>
- Schmid, M., Jaedicke, A., Du, T.-G., & Jansen, R.-P. (2006). Coordination of endoplasmic reticulum and mrna localization to the yeast bud. *Current Biology*, *16*(15), 1538–1543. <https://doi.org/10.1016/j.cub.2006.06.025>
- Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., Green, R., Shen, B., & Liu, J. O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature Chemical Biology*, *6*(3), 209–217. <https://doi.org/10.1038/nchembio.304>
- Schuller, A. P., & Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*, *19*(8), 526. <https://doi.org/10.1038/s41580-018-0011-4>
- Schuller, A. P., Wu, C. C.-C., Dever, T. E., Buskirk, A. R., & Green, R. (2017). EIF5A functions globally in translation elongation and termination. *Molecular Cell*, *66*(2), 194–205.e5. <https://doi.org/10.1016/j.molcel.2017.03.003>
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342. <https://doi.org/10.1038/nature10098>
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-limiting steps in yeast protein translation. *Cell*, *153*(7), 1589–1601. <https://doi.org/10.1016/j.cell.2013.05.049>
- Shalgi, R., Hurt, J. A., Krykbaeva, I., Taipale, M., Lindquist, S., & Burge, C. B. (2013). Widespread regulation of translation by elongation pausing in heat shock. *Molecular Cell*, *49*(3), 439–452. <https://doi.org/10.1016/j.molcel.2012.11.028>
- Sharma, A. K., & O'Brien, E. P. (2018). Non-equilibrium coupling of protein structure and function to translation–elongation kinetics. *Current Opinion in Structural Biology*, *49*, 94–103. <https://doi.org/10.1016/j.sbi.2018.01.005>
- Sharma, A. K., Sormanni, P., Ahmed, N., Ciryam, P., Friedrich, U. A., Kramer, G., & O'Brien, E. P. (2019). A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLOS Computational Biology*, *15*(5), e1007070. <https://doi.org/10.1371/journal.pcbi.1007070>

- Shi, Z., Fujii, K., Kovary, K. M., Genuth, N. R., Röst, H. L., Teruel, M. N., & Barna, M. (2017). Heterogeneous ribosomes preferentially translate distinct subpools of mrnas genome-wide. *Molecular Cell*, 67(1), 71-83.e7. <https://doi.org/10.1016/j.molcel.2017.05.021>
- Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., & Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, 561(7722), 268. <https://doi.org/10.1038/s41586-018-0462-y>
- Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: Mechanisms and biological targets. *Cell*, 136(4), 731–745. <https://doi.org/10.1016/j.cell.2009.01.042>
- Stein, K. C., & Frydman, J. (2019). The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *Journal of Biological Chemistry*, 294(6), 2076–2084. <https://doi.org/10.1074/jbc.REV118.002814>
- Stein, K. C., Kriel, A., & Frydman, J. (2019). Nascent polypeptide domain topology and elongation rate direct the cotranslational hierarchy of hsp70 and tric/cct. *Molecular Cell*, 75(6), 1117-1130.e5. <https://doi.org/10.1016/j.molcel.2019.06.036>
- Susor, A., Jansova, D., Cerna, R., Danylevska, A., Anger, M., Toralova, T., Malik, R., Supolikova, J., Cook, M. S., Oh, J. S., & Kubelka, M. (2015). Temporal and spatial regulation of translation in the mammalian oocyte via the mtor-eif4f pathway. *Nature Communications*, 6, 6078. <https://doi.org/10.1038/ncomms7078>
- Szavits-Nossan, J., & Ciandrini, L. (2019). Accurate measures of translation efficiency and traffic using ribosome profiling. *BioRxiv*, 719302. <https://doi.org/10.1101/719302>
- Tahmasebi, S., Khoutorsky, A., Mathews, M. B., & Sonenberg, N. (2018). Translation deregulation in human disease. *Nature Reviews Molecular Cell Biology*, 1. <https://doi.org/10.1038/s41580-018-0034-x>
- Taylor, R. C., Berendzen, K. M., & Dillin, A. (2014). Systemic stress signalling: Understanding the cell non-autonomous control of proteostasis. *Nature Reviews Molecular Cell Biology*, 15(3), 211–217. <https://doi.org/10.1038/nrm3752>
- Torrent, M., Chalancon, G., Groot, N. S. de, Wuster, A., & Babu, M. M. (2018). Cells alter their trna abundance to selectively regulate protein synthesis during stress conditions. *Sci. Signal.*, 11(546), eaat6409. <https://doi.org/10.1126/scisignal.aat6409>
- Torres, A. G., Batlle, E., & Ribas de Pouplana, L. (2014). Role of trna modifications in human diseases. *Trends in Molecular Medicine*, 20(6), 306–314. <https://doi.org/10.1016/j.molmed.2014.01.008>
- Tu, L. W., & Deutsch, C. (2010). A folding zone in the ribosomal exit tunnel for kv1.3 helix formation. *Journal of Molecular Biology*, 396(5), 1346–1360. <https://doi.org/10.1016/j.jmb.2009.12.059>



- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, *141*(2), 344–354. <https://doi.org/10.1016/j.cell.2010.03.031>
- Tunney, R., McGlinchy, N. J., Graham, M. E., Naddaf, N., Pachter, L., & Lareau, L. F. (2018). Accurate design of translational output by a neural network model of ribosome distribution. *Nature Structural & Molecular Biology*, *25*(7), 577–582. <https://doi.org/10.1038/s41594-018-0080-2>
- Tuorto, F., & Lyko, F. (2016). Genome recoding by trna modifications. *Open Biology*, *6*(12), 160287. <https://doi.org/10.1098/rsob.160287>
- Valastyan, J. S., & Lindquist, S. (2014). Mechanisms of protein-folding diseases at a glance. *Disease Models & Mechanisms*, *7*(1), 9–14. <https://doi.org/10.1242/dmm.013474>
- Väre, V. Y. P., Eruysal, E. R., Narendran, A., Sarachan, K. L., & Agris, P. F. (2017). Chemical and conformational diversity of modified nucleosides affects trna structure and function. *Biomolecules*, *7*(1), 29. <https://doi.org/10.3390/biom7010029>
- Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270*(5235), 484–487. <https://doi.org/10.1126/science.270.5235.484>
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: From basic research to diagnostics. *Clinical Chemistry*, *55*(4), 641–658. <https://doi.org/10.1373/clinchem.2008.112789>
- Voet, Daniel, & Voet, Judith. (2005). *Biochimie* (2e édition).
- von der Haar, T. (2012). MATHEMATICAL and computational modelling of ribosomal movement and protein synthesis: AN overview. *Computational and Structural Biotechnology Journal*, *1*(1), e201204002. <https://doi.org/10.5936/csbj.201204002>
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J. C., & Ware, D. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications*, *7*(1), 1–13. <https://doi.org/10.1038/ncomms11708>
- Wang, H., Kingsford, C., & McManus, C. J. (2018). Using the ribodeblur pipeline to recover a-sites from yeast ribosome profiling data. *Methods*, *137*, 67–70. <https://doi.org/10.1016/j.ymeth.2018.01.002>
- Wang, H., McManus, J., & Kingsford, C. (2016). Accurate recovery of ribosome positions reveals slow translation of wobble-pairing codons in yeast. *Journal of Computational Biology*, *24*(6), 486–500. <https://doi.org/10.1089/cmb.2016.0147>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, *24*(11), 437–440. [https://doi.org/10.1016/S0968-0004\(99\)01460-7](https://doi.org/10.1016/S0968-0004(99)01460-7)

- Weil, T. T., Parton, R., Davis, I., & Gavis, E. R. (2008). Changes in bicoid mRNA anchoring highlight conserved mechanisms during the oocyte-to-embryo transition. *Current Biology*, *18*(14), 1055–1061. <https://doi.org/10.1016/j.cub.2008.06.046>
- Weinberg, D. E., Shah, P., Eichhorn, S. W., Hussmann, J. A., Plotkin, J. B., & Bartel, D. P. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Reports*, *14*(7), 1787–1799. <https://doi.org/10.1016/j.celrep.2016.01.043>
- Wilson, D. N., & Cate, J. H. D. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harbor Perspectives in Biology*, *4*(5), a011536. <https://doi.org/10.1101/cshperspect.a011536>
- Wolin, S. L., & Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal*, *7*(11), 3559–3569.
- Wu, C. C.-C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019). High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Molecular Cell*, *73*(5), 959–970.e5. <https://doi.org/10.1016/j.molcel.2018.12.009>
- Xu, Q., & Shaulsky, G. (2005). GOAT. *Applied Bioinformatics*, *4*(4), 281–283. <https://doi.org/10.2165/00822942-200504040-00008>
- Xue, S., & Barna, M. (2012). Specialized ribosomes: A new frontier in gene regulation and organismal biology. *Nature Reviews Molecular Cell Biology*, *13*(6), 355–369. <https://doi.org/10.1038/nrm3359>
- Yang, J.-R., Chen, X., & Zhang, J. (2014). Codon-by-codon modulation of translational speed and accuracy via mRNA folding. *PLOS Biology*, *12*(7), e1001910. <https://doi.org/10.1371/journal.pbio.1001910>
- Young, D. J., Guydosh, N. R., Zhang, F., Hinnebusch, A. G., & Green, R. (2015). Rli1/abce1 recycles terminating ribosomes and controls translation reinitiation in 3'UTRs in vivo. *Cell*, *162*(4), 872–884. <https://doi.org/10.1016/j.cell.2015.07.041>
- Young, J. C., Hoogenraad, N. J., & Hartl, F. U. (2003). Molecular chaperones hsp90 and hsp70 deliver preproteins to the mitochondrial import receptor tom70. *Cell*, *112*(1), 41–50. [https://doi.org/10.1016/S0092-8674\(02\)01250-3](https://doi.org/10.1016/S0092-8674(02)01250-3)
- Zarai, Y., Margaliot, M., & Tuller, T. (2017). Ribosome flow model with extended objects. *Journal of The Royal Society Interface*, *14*(135), 20170128. <https://doi.org/10.1098/rsif.2017.0128>
- Zhang, G., Hubalewska, M., & Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology*, *16*(3), 274–280. <https://doi.org/10.1038/nsmb.1554>

- Zhang, G., & Ignatova, Z. (2011). Folding at the birth of the nascent chain: Coordinating translation with co-translational folding. *Current Opinion in Structural Biology*, 21(1), 25–31. <https://doi.org/10.1016/j.sbi.2010.10.008>
- Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3), 95–109. <https://doi.org/10.1016/j.jgg.2011.02.003>
- Zhang, S., Hu, H., Zhou, J., He, X., Jiang, T., & Zeng, J. (2017). Analysis of ribosome stalling and translation elongation dynamics by deep learning. *Cell Systems*, 5(3), 212–220.e6. <https://doi.org/10.1016/j.cels.2017.08.004>
- Zhao, D., Baez, W. D., Fredrick, K., & Bundschuh, R. (2019). RiboProP: A probabilistic ribosome positioning algorithm for ribosome profiling. *Bioinformatics*, 35(9), 1486–1493. <https://doi.org/10.1093/bioinformatics/bty854>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1), 1–12. <https://doi.org/10.1038/ncomms14049>
- Zinshteyn, B., & Gilbert, W. V. (2013). Loss of a conserved trna anticodon modification perturbs cellular signaling. *PLOS Genetics*, 9(8), e1003675. <https://doi.org/10.1371/journal.pgen.1003675>
- Zou, K., Ouyang, Q., Li, H., & Zheng, J. (2017). A global characterization of the translational and transcriptional programs induced by methionine restriction through ribosome profiling and rna-seq. *BMC Genomics*, 18(1), 189. <https://doi.org/10.1186/s12864-017-3483-2>