



**HAL**  
open science

# Regulation of mRNA translation during the inflammatory response in murine macrophages

Juliana Blin-Gonthier

► **To cite this version:**

Juliana Blin-Gonthier. Regulation of mRNA translation during the inflammatory response in murine macrophages. Molecular biology. Université de Lyon, 2020. English. NNT : 2020LYSEN032 . tel-03122969

**HAL Id: tel-03122969**

**<https://theses.hal.science/tel-03122969>**

Submitted on 27 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Numéro National de Thèse : 2020LYSEN032

**THÈSE de DOCTORAT DE L'UNIVERSITE DE LYON**  
opérée par  
**l'Ecole Normale Supérieure de Lyon**

**Ecole Doctorale N°340**  
**Biologie Moléculaire, Intégrative et Cellulaire (BMIC)**

**Discipline** : Sciences de la vie et de la santé

Soutenue publiquement le 30 octobre 2020, par :

**Juliana BLIN-GONTHIER**

---

**Regulation of mRNA translation during  
the inflammatory response in murine  
macrophages**

Étude de la régulation de la traduction dans les  
macrophages au cours de la réponse  
inflammatoire

---

Devant le jury composé de :

WEIL Dominique, Directrice de recherche IBPS, Paris	Rapporteuse
SAVEANU Cosmin, Directeur de recherche Institut Pasteur, Paris	Rapporteur
GOUJON Caroline, Chargée de recherche IRIM, Montpellier	Examinatrice
JALINOT Pierre, Directeur de recherche LBMC, ENS de Lyon	Examineur
RICCI Emiliano, Chargé de recherche LBMC, ENS de Lyon	Directeur de thèse





# Foreword

Since my arrival in the laboratory, my objective was to set up tools to study the role of post-transcriptional control in the regulation of gene expression in immune cells.

My first project consisted in the validation of a new tool for the genetic engineering of difficult to transfect cells, such as primary bone-marrow derived macrophages, using viral particles mediated delivery of the components of the CRISPR/Cas9 system (Mangeot et al., 2019). This work was part of a publication annexed to this manuscript (Annex 1).

After this, I focused on the study of translation regulation using newly developed high-throughput sequencing approaches such as ribosome profiling. While learning the basics of this technique, I wrote a review on its utility to decipher the impact of viral infection on the cell biology (Blin and Ricci, 2016) that is annexed to this manuscript (Annex 2).

Following this literature review, I performed parallel RNA-seq and ribosome profiling on humanized mouse T lymphocytes infected with HTLV-1 to characterize the role of the PDZ-domain of the viral oncoprotein Tax in the immortalization of infected cells (Pérès et al., 2018). This work was also part of a publication annexed to this manuscript (Annex 3).

The main part of my PhD was consecrated to the adaptation of an innovative approach to study the translation from different ribosomal populations, monosomes and polysomes, in very great detail during the inflammatory response in murine macrophages. I will thus focus on this work during the rest of this manuscript.



# Abstract

The dynamic regulation of the protein synthesis process participates in the cell adaptation to a constantly evolving environment. Despite its critical role in gene expression regulation, the understanding of translational control in fundamental biological processes, such as immune responses, is still incomplete. The implementation of new approaches based on deep sequencing can be used to fill the gap in the knowledge of protein synthesis regulation. Notably, monosome vs polysome footprinting is an innovative approach derived from ribosome profiling that allows the characterization of 80S footprints derived either from monosomes or polysomes associated ribosomes. In this work, I identified the key parameters required to obtain a robust picture of ribosomal densities across cellular mRNAs using monosome vs polysome footprinting in murine primary bone-marrow derived macrophages (pBMDM). These immune cells are particularly interesting to study protein synthesis regulation in evolving conditions as they display a high sensitivity towards their environment and have the ability to trigger different gene expression programs depending on external cues. Their high phenotypic plasticity is in fact essential to ensure their protective functions in the organism such as the triggering and the resolution of the inflammatory response. As monosome vs polysome footprinting was initially developed in yeast, the adaptation of this method to study murine immune cells required extensive optimizations. The resulting protocol developed in this work was used to confirm that, contrary to a long lasting belief in the scientific community, murine pBMDM monosomes are actively involved in the translation process. Interestingly, we were able to recapitulate similar observations to what was previously observed in yeast regarding the features of mRNAs preferentially bound to monosomes or polysomes in murine pBMDM. This could suggest that the differential trafficking of ribosomes depending on specific features of the cellular mRNAs is a conserved mechanism of translational control. Importantly, the distribution of ribosomes across the different mRNAs is not random and the proper ribosome allocation pattern could be critical to adapt protein synthesis levels to the cellular needs. Here we developed a robust strategy to study this overlooked transcript-specific mechanism of translational control. Moreover, our optimized protocol can now be used to study the impact of translation through monosomes or polysomes at different stages of the inflammatory response in murine macrophages.

Key-words : Inflammation, Macrophage, Translation, Ribosome profiling, Monosome vs Polysome footprinting



# Résumé de la thèse

La régulation de la traduction permet d'adapter les niveaux de synthèse des protéines en fonction des besoins de la cellule. Ce type de régulation joue notamment un rôle particulier lorsque la cellule est confrontée à des modifications de son environnement ou lors d'un stress. Ces dernières années, le développement de nouvelles techniques basées sur le séquençage à haut débit, comme le *ribosome profiling*, a permis de mettre en lumière l'importance de la régulation de la traduction au cours du processus d'expression des gènes. Cette régulation peut se faire à différents niveaux : à l'échelle globale, par le ciblage des différents acteurs impliqués dans le processus de traduction, ou de façon plus spécifique, pour chaque protéine individuellement en ciblant l'ARN messager (ARNm) correspondant. Les mécanismes permettant de déterminer quels ARNm sont traduits dans la cellule à un moment donné sont particulièrement complexes et ne sont pas encore tous complètement caractérisés. Le *monosome vs polysome footprinting* est une nouvelle méthode d'étude de la traduction dérivée du *ribosome profiling*. Cette approche a permis de montrer que certains ARNm dont l'expression est très régulée au cours du temps sont majoritairement traduits par un seul ribosome, ou monosome, alors que la synthèse des protéines fait le plus souvent appel au recrutement de plusieurs ribosomes, ou polysomes. Les ARNm préférentiellement traduits par des monosomes permettent notamment la synthèse de protéines jouant un rôle de régulation dans la cellule dont l'expression doit être très contrôlée. L'association préférentielle à des monosomes pourrait ainsi permettre de réguler l'expression des protéines ayant un impact très fort sur la vie de la cellule de façon très dynamique en fonction de l'environnement.

Au cours de ma thèse, j'ai utilisé des macrophages primaires dérivés de la moelle osseuse de souris pour étudier la régulation de la synthèse des protéines par *ribosome profiling* et *monosome vs polysome profiling*. Ces cellules sont particulièrement intéressantes car elles présentent une grande capacité d'adaptation à leur environnement et peuvent moduler de façon très dynamique leurs taux de synthèse protéique suite à la détection d'un signal de danger. Cette importante plasticité est cruciale pour assurer leurs fonctions de cellules immunitaires protectrices dans l'organisme. La synthèse des protéines dans les macrophages est ainsi particulièrement contrôlée afin d'assurer la mise en place de réponses immunitaires efficaces et adaptées. En effet, en cas de perturbations, la réponse inflammatoire déclenchée dans les macrophages peut avoir des effets délétères à l'échelle de l'organisme pouvant aller jusqu'au déclenchement d'une inflammation chronique ou de pathologies auto-immunitaires.

L'approche de *monosome vs polysome* footprinting ayant été initialement développée chez la levure, son application à des macrophages primaires de souris a nécessité de nombreuses optimisations techniques. Après cette étape de mise au point, j'ai pu valider la qualité des résultats obtenus dans les macrophages murins en les comparant à ceux obtenus chez la levure. Notamment, j'ai pu confirmer que les monosomes étaient impliqués dans toutes les étapes du processus de traduction. Cette observation est particulièrement importante car historiquement, les monosomes étaient considérés comme inactifs ou comme une étape de transition avant de devenir des polysomes. J'ai pu ainsi identifier des ARNm qui sont préférentiellement traduits par des monosomes plutôt que par des polysomes dans les macrophages primaires de souris. Comme chez la levure, une partie de ces ARNm correspond à des protéines membranaires pour lesquelles une association avec le réticulum endoplasmique en cours de synthèse est nécessaire ou à des protéines régulatrices dont l'expression est très contrôlée. J'ai ensuite utilisé une approche de *machine learning* pour identifier les caractéristiques des ARNm qui pourraient expliquer leur traduction préférentiellement par les monosomes plutôt que par les polysomes. L'utilisation du protocole optimisé de *monosome vs polysome footprinting* ouvre ainsi la possibilité d'étudier les modifications de synthèse protéique induites suite à un changement de l'environnement de façon très détaillée. Cette approche pourra notamment être utilisée pour caractériser l'effet de la régulation de la traduction sur la mise en place et la résolution de la réponse inflammatoire dans les macrophages primaires.

# Contents

<b>Foreword</b> .....	1
<b>Abstract</b> .....	3
<b>Résumé de la thèse</b> .....	5
<b>Contents</b> .....	7
<b>Introduction</b> .....	9
<b>Chapter I. The RNA-centered regulation of protein synthesis</b> .....	9
<b>I.1 Back to the 20<sup>th</sup> century, from proteins to genes : the unexpected role of RNA</b> 9	
Historic view on the central dogma of molecular biology .....	9
The central role of messenger RNA.....	11
<b>I.2 Global gene expression regulation</b> .....	15
RNA sequencing to study gene expression regulation.....	15
Quantification of protein synthesis rates using polysome profiling .....	16
Quantitative proteomics bring a new view of the cell proteome regulation .....	18
Ribosome profiling provides a snapshot of translation rates .....	20
<b>Chapter II. Dynamic regulation of protein synthesis rates</b> .....	24
<b>II. 1. General principles of mRNA translation</b> .....	24
Universal features of protein synthesis .....	24
The mechanism of translation in eukaryotes.....	25
<b>II. 2. Mechanisms of translational control</b> .....	32
Cellular resources limit global protein production levels.....	32
Regulation of global protein synthesis by targeting of translation factors .....	33
Transcript-specific regulation dependent on specific features of the mRNA.....	34
Trans-acting factors modulate specific mRNA translation efficiency .....	37
<b>II. 3. Monosomes : overlooked players in translational control</b> .....	38
Protein synthesis occurs mainly in polysomes .....	38
Monosomes participate in the translation of highly regulated mRNAs.....	40
Alternative cellular functions of monosomes .....	44
<b>Chapter III. Gene expression regulation during the inflammatory response in macrophages</b> .....	47
<b>III.1 Macrophages plasticity is critical for the inflammatory response efficiency</b> .	47
<b>III.2 Regulation of inflammation related genes in macrophages</b> .....	49



<b>Results</b> .....	54
Setting-up the harvesting of cytoplasmic lysates from macrophages .....	55
Depletion of highly abundant ribosomal RNA contaminants.....	59
The RNase treatment greatly influences the ribosomal footprints quality .....	70
RNase H mediated depletion of rRNA contaminants .....	73
Endogenous RNases disturb the purification of monosomes and polysomes from pBMDMs .....	77
Improving ribosomal footprints selection using high salt conditions .....	83
Monosome vs polysome footprinting using pBMDMs.....	92
Characterization of monosome vs polysome enriched transcripts.....	106
In vitro translation to decipher monosomes translational status .....	115
<b>Discussion</b> .....	121
Adaptation of the monosome vs polysome footprinting protocol from yeast to macrophages.....	121
Deciphering monosomes translational status .....	126
Characterization of the features explaining preferential association with monosomes or polysomes .....	130
How monosomes or polysomes binding shape the inflammatory response in macrophages ?.....	135
How monosomes and polysomes participate in the modulation of the cellular proteome depending on the conditions ? .....	137
<b>Material and Methods</b> .....	139
<b>References</b> .....	145

# Introduction

## Chapter I. The RNA-centered regulation of protein synthesis

### I.1 Back to the 20<sup>th</sup> century, from proteins to genes : the unexpected role of RNA

#### *Historic view on the central dogma of molecular biology*

Protein synthesis is at the center of every biological system. Scrutinized for decades, this fundamental process is still not fully characterized. How protein synthesis can be re-shaped by the cell in order to express all the components that are required for its survival in a constantly evolving environment is a particularly puzzling question. The study of regulatory mechanisms involved in gene expression control has already provided many answers to this question. In the beginning of this exploration, the nature of the different factors involved in protein synthesis was not even known. The idea that genes could direct the protein expression process was suggested for the first time in 1901 by Archibald Garrod (Piro et al., 2009). In 1941, George Beadle and Edward Tatum validated this hypothesis by developing the first reverse genetics experiment : they irradiated *Neurospora* fungus spores using X-rays to introduce mutations in their genes and studied the impact on cell phenotype (Beadle and Tatum, 1941). They observed that some irradiated strains had lost the ability to carry out biochemical reactions essential for their metabolism and survival on a minimal medium. Each mutant lost the ability to use one specific metabolite required for *Neurospora* survival. The role of enzymes in the catalysis of biochemical reactions in the cell was already well characterized. The authors could thus conclude that each mutant probably corresponded to one specific inactivated enzyme. It was the birth of the first gene expression model : “One gene produces one protein”. It must be noted that the notion of gene at that time was quite elusive. Despite the establishment of a link between genes and protein synthesis, the chemical nature of the cellular components involved was unclear. Beadle and Tatum’s experiment nevertheless launched a race for the characterization of the protein synthesis process. As a consequence, most of the basic knowledge regarding this major cellular mechanism was established during the time period spanning the 1940-1960s.

By the 1940s, the importance of proteins in cell life was well established. In fact, their diversity of structures and versatility of functions inspired many contemporary scientists to believe that they would be the support of genetic information (Strauss, 2016). The different tools available were not sensitive enough to dissociate the role of DNA and proteins in nucleoprotein complexes. The mechanism by which proteins would be able to synthesize more proteins revealed to be a baffling problem. It was the development of novel approaches to study the chemical nature of cellular components that brought new hints for the understanding of protein synthesis. In 1944, Avery and colleagues identified DNA as the transforming factor responsible for the acquisition of specific characteristics in bacteria (Avery et al., 1944). Moreover, the development of chromatography around 1945 was critical to link the structure of nucleic acids to that of proteins. The data obtained in 1950 by Chargaff, studying the variations in base composition of DNA from different organisms, were essential for the proper interpretation of DNA's role in the cell life. In 1953, Watson and Crick used X-ray crystallography data to uncover DNA structure and concluded that this molecule is the master regulator of protein synthesis (Watson and Crick, 1953a, 1953b). The mechanism explaining how information from DNA in the nucleus could be translated into proteins in the cytoplasm remained however unclear. The role of RNA in protein synthesis was revealed by a series of experiments conducted in 1956 on Tobacco Mosaic Virus (TMV) by Gierer and colleagues. They observed that the viral RNA was sufficient to produce infective particles and that different RNA could produce different nucleoproteins in infected plants (Gierer and Schramm, 1956). Moreover, the treatment of normal TMV RNA with nitrous acid induced mutations in the viral RNA and an alteration of the amino acid sequence of the proteins expressed in the corresponding viral particles (Gierer and Mundry, 1958). The same year, Francis Crick published an essay recapitulating the different views "On protein synthesis", revealing the mindset of the scientific community around that time (Crick, 1958):

" [...] the main function of the genetic material is to control (not necessarily directly) the synthesis of proteins. [...] Once the central and unique role of proteins is admitted there seems little point in genes doing anything else. Although proteins can act in so many different ways, the way in which they are synthesized is probably uniform and rather simple [...]. Biologists should not deceive themselves with the thought that some new class of biological molecules, of comparable importance to the proteins, remains to be discovered. This seems highly unlikely. In the protein molecule Nature has devised a unique instrument in which an underlying simplicity is used to express great subtlety and versatility."

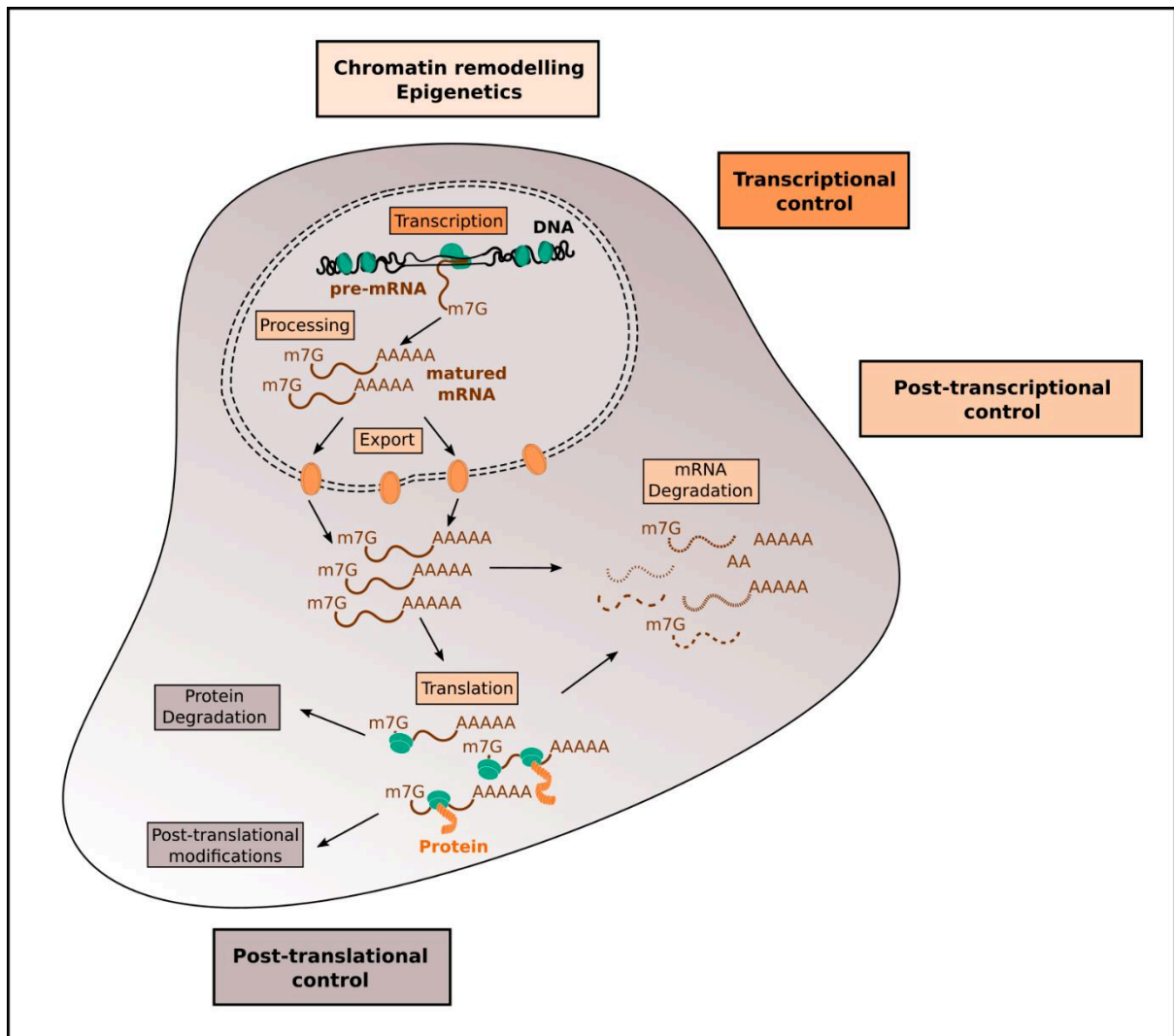
Historically RNA was thus not expected to have a big impact on the global gene expression workflow where DNA and proteins already played the biggest part. Its main purpose was to be a faithful representation of genetic information encoded in DNA and to be a template directing protein synthesis. Francis Crick's view on the gene expression process became the widely accepted central dogma of molecular biology that prevailed for decades.

### ***The central role of messenger RNA***

In the beginning of the 1960s, the scientific community had accumulated an extensive knowledge about the ribosome composition and its role in protein synthesis. Virtually all the components involved in the reaction were described but the mechanism by which their combination would actually ensure protein synthesis was not understood (Warner and Knopf, 2002; Warner et al., 1963). One question in particular represented a big challenge : what kind of RNA would be the template that guides the sequential addition of amino acids in the growing polypeptide chain? Transfer RNA (tRNA) role as an adaptor molecule bridging the nucleotide codon to its corresponding amino-acid was already described. Some scientists suggested that the template might be the RNA composing the ribosome itself. However, it was known that ribosomal RNA levels were quite stable through time (Ts'o, 1962). If there was a template translating the genetic information contained in DNA sequence into a protein sequence, then it would be expected that this molecule would be quite unstable in order to maintain a constant flow of information. Compiling the results of the experiments performed by Gierer on TMV and their data acquired studying the expression of beta-galactosidase in bacteria, Jacob and Monod elaborated the theory of a messenger RNA (mRNA) in 1961 (Jacob and Monod, 1961). This hypothesis was further validated by the experiment performed by Nirenberg and Matthaei published the same year. Using a cell-free translation system treated with DNase, they demonstrated that the addition of polyuridylic acid induces the synthesis of polypeptides composed exclusively of phenylalanines (Nirenberg and Matthaei, 1961). In 1963, three independent studies revealed that most mRNAs are translated by several ribosomes at a time forming structures called polysomes (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). By the end of the 1960s, all the pieces of the protein synthesis puzzle were thus finally reassembled. The new objective of the field was then to further characterize the different components involved in protein synthesis leading to a great improvement of the methods to study their molecular structures including electron microscopy and X-ray crystallography during the following years.

The central dogma of molecular biology postulated in 1958 was challenged for the first time in 1977 with the discovery of introns by two independent laboratories (Berget et al., 1977; Berk and Sharp, 1977; Chow et al., 1977). To localize the position of genes within the genome, they performed DNA-RNA hybridization experiments and observed that the genetic information was discontinuously organized within the DNA molecule. As a consequence, mRNA is not a faithful copy of DNA but contains intronic sequences that must be removed and exonic sequences brought back together through splicing to serve as template for protein synthesis. This phenomenon that is mostly restricted to eukaryotes could not be uncovered from previous studies on protein synthesis mainly focused on bacteria.

The vision of mRNA restricted to its messenger function in protein synthesis remained the standard until the development of sequencing technologies and launching of the whole genome sequencing projects in the 1990s. The unprecedented amount of information obtained from these experiments revealed that the actual number of genes in an eukaryotic genome is quite low compared to the diversity of proteins in a cell. This discrepancy could be explained by the fact that one gene could direct the synthesis of many proteins and not just only one (Siomi and Dreyfuss, 1997). One mechanism in particular, alternative splicing, is essential to ensure the expression of multiple peptides from single stretches of DNA through the use of multiple start sites and different patterns of exon use. This discovery therefore revealed that the genetic message carried by mRNA can be modified even after its transcription. More than just a messenger, mRNA plays an important part in the gene expression process. In fact, several regulatory mechanisms act at different levels of the mRNA life to modulate protein synthesis in eukaryotes (Figure 1). Notably, the role of mechanisms controlling mRNA processing, localization, stability and translation into proteins in the whole gene expression process was greatly undervalued previously. Importantly, the nature of the mRNA can also impact protein synthesis levels, thus increasing the complexity of this mechanism initially described as rather simple. This RNA-centered regulation of protein synthesis is of particular interest for the cell adaptability in fluctuating conditions.



**Figure 1: Gene expression can be regulated at different levels in eukaryotes.** The genetic information stored in DNA is copied in the pre-mRNA molecule through transcription. Gene expression regulation involves chromatin modifications or remodeling (epigenetics) that modulate the accessibility of the DNA for the transcription machinery. Transcriptional control implies the selection of the transcribed gene through the binding of specific transcription factors. Post-transcriptional control comprises several steps of the mRNA life after its synthesis in the nucleus : Processing, Export to the cytoplasm, Translation or Degradation. The processing of the pre-mRNA begins co-transcriptionally with the addition of a m7G cap at the 5' end, splicing to remove introns, and 3' end cleavage/polyadenylation. Only the matured mRNAs are competent for export through the nuclear pores. In the cytoplasm, the mRNAs can be targeted for translation or degradation depending on the associated RNA-binding proteins. Newly synthesized proteins are co-translationally folded and can undergo additional post-translational modifications upon release from the ribosome. Controlled protein degradation occurs constantly to regulate protein levels.

To sum up, the regulation at both transcriptional and post-transcriptional levels is required to allow the expression of all the proteins needed in a cell at the right time. This combination is important because both mechanisms are not effective on the same scale. While transcription control determines the pool of expressed genes, post-transcriptional mechanisms fine tune the timing and levels of protein synthesis (Vogel and Marcotte, 2012). This property is essential for the adaptability to the environment as it ensures the rapid and versatile modulation of protein levels. Numerous post-transcriptional events also participate in the expression of different genetic programs depending on the cell function in a multicellular organism. Furthermore, defects in post-transcriptional control mechanisms can be linked to various human pathologies (Corbett, 2018). The study of the link between a specific mRNA and its corresponding protein abundance is thus particularly important. Despite its critical role in gene expression regulation, the understanding of post-transcriptional control in fundamental biological processes is still lagging behind compared to transcriptional control for both historical and technical reasons (Mata et al., 2005). Indeed, until the last decade, the tools to study post-transcriptional control impact genome-wide were quite limited. The implementation of new approaches was thus pivotal to fill the gap in the knowledge of protein synthesis regulation (Hershey et al., 2012).

## **I.2 Global gene expression regulation**

Back in the 1990s, the importance of post-transcriptional control was widely accepted in the scientific community (Siomi and Dreyfuss, 1997). Cells adapt to their environment by modulating their protein synthesis levels through regulatory mechanisms involved at different stages of the gene expression flow. However, to what extent the diverse post-transcriptional mechanisms contribute to the regulation of global protein synthesis rates was not fully understood. The regulation of mRNA stability, localization and translation during several developmental stages in eukaryotes was well described (Siomi and Dreyfuss, 1997). Early studies on the protein synthesis process had revealed that translation efficiency can vary depending on intrinsic features shared by different subsets of mRNAs (Kozak, 1991a). It was becoming clear that as opposed to Francis Crick's conclusion, the way proteins are synthesized is not uniform and definitely not simple. To get a clearer picture, the next challenge was thus to study the correlation between the protein levels in a cell and the expression of a specific genetic program. The spectacular development of methods to quantify mRNA and protein levels during the following decades prompted the multiplication of genome-wide approaches to characterize gene expression patterns.

### ***RNA sequencing to study gene expression regulation***

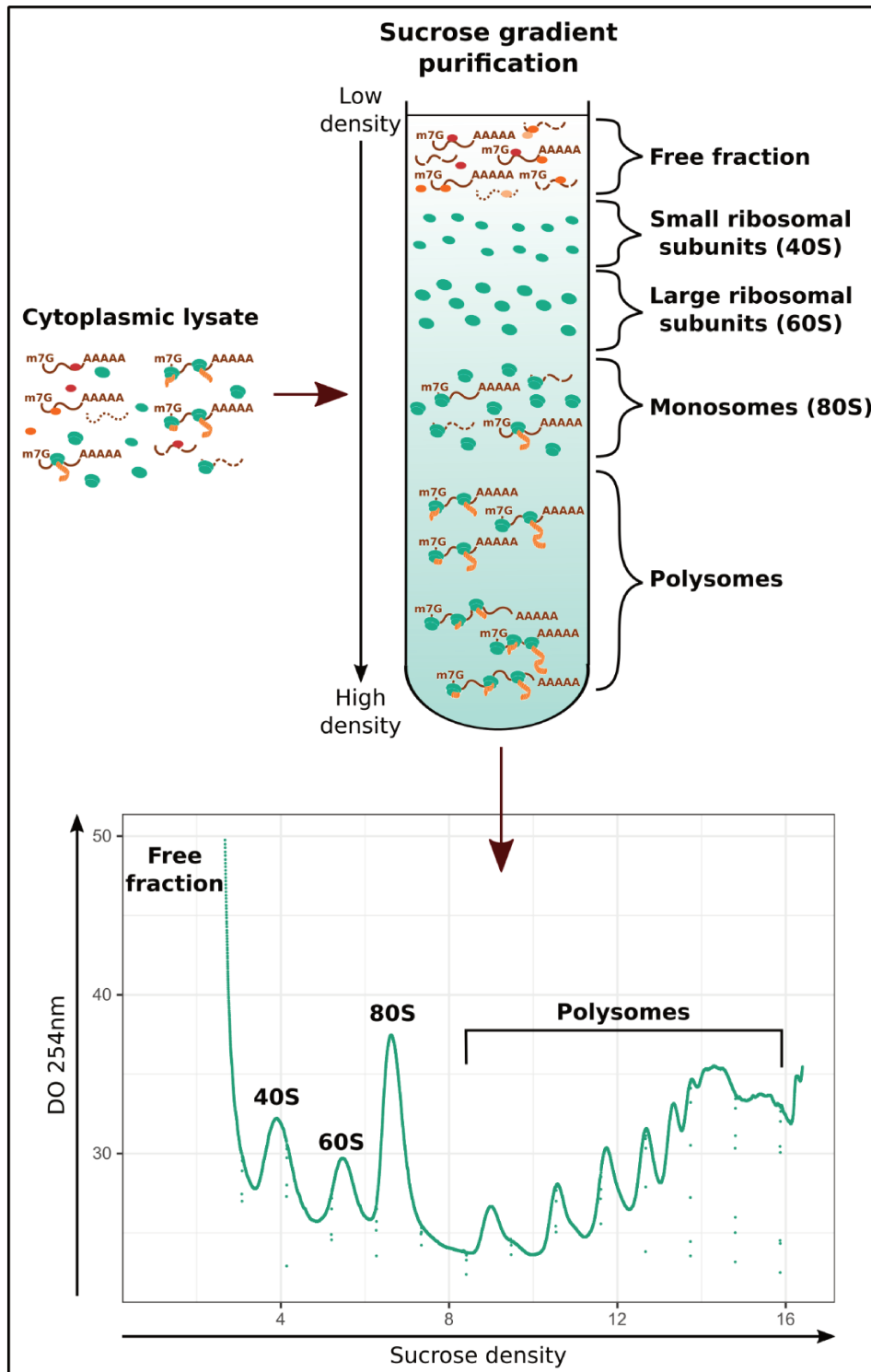
Amongst the first organisms for which whole genome sequencing was completed, yeast was also used for the first analysis comparing mRNA expression to protein levels in eukaryotes (Gygi et al., 1999). This study revealed that mRNA quantification is not sufficient to predict protein expression rates : only 40% of global protein abundance could be explained by mRNA levels. Yet, their results were questionable because of the restricted number of mRNAs analyzed and the technical biases imputed to the methods used for both mRNA and protein quantification. Notably, the approaches to identify and measure individual protein levels were very limited at that time. Using 2D polyacrylamide gel electrophoresis coupled with radioactive labelling and mass spectrometry, they were able to quantify unambiguously 156 proteins while they had access to the levels of 4665 mRNAs using Serial Analysis of Gene Expression (SAGE). Moreover, another study published the same year using virtually the same experimental conditions supported on the contrary that mRNA and protein levels were well correlated (Futcher et al., 1999). The main output of this first analysis was thus that better methods for protein quantification would be required before going further in the investigation of post-transcriptional impact on protein synthesis.



Few years later, Greenbaum and colleagues re-analyzed both the mRNA and protein quantification data collected in the two studies published in 1999. They also included the results obtained in two other large-scale proteomics analysis in yeast using MudPit approach that combines chromatography and mass spectrometry. By doing so, they managed to generate a dataset containing protein abundance information for approximately 2000 mRNAs. The analysis of this wider dataset revealed that in yeast, 66% of protein levels could be explained by those of mRNAs (Greenbaum et al., 2003). To interpret the low correlation, the authors explain that experimental errors, differences of protein half-lives and post-transcriptional control can introduce variations between mRNA and protein levels. Furthermore, when looking at different subsets of proteins, depending either on their subcellular localization or functions, the correlation coefficient can be decreased or increased. This suggested that the regulatory mechanisms involved could be different depending on the target mRNA.

### ***Quantification of protein synthesis rates using polysome profiling***

Considering the difficulties to characterize the cell proteome using the available methods, another approach that consists in the quantification of translatable mRNAs was promoted around the same time (Pradet-Balade et al., 2001; Zong et al., 1999). As measuring total mRNA levels was not reliable to predict the corresponding protein abundances, the selection of actively translated mRNAs could be a better indicator. This approach, named polysome profiling, implies the separation of cellular mRNAs according to their degree of ribosome loading through ultracentrifugation on a sucrose gradient (Figure 2). The number of ribosomes loaded on a specific mRNA depends both on the rate of translation initiation and the speed at which the ribosomes elongate the newly produced polypeptide chain (Ruan et al., 1997). As it was generally accepted that the initiation step is rate-limiting and more tightly controlled than the rest of the protein synthesis process, the ribosomal loading would be controlled mainly by the recruitment of new ribosomes onto the mRNA. Moreover, mRNAs associated with a single ribosome, or monosomes, are less translationally active than those bound to several ribosomes, or polysomes (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). Therefore, the number of ribosomes bound to one mRNA was expected to be a robust indicator of the protein synthesis rate.



**Figure 2: Polysome profiling separates cellular mRNA depending on the number of associated ribosomes.** Hypotonic cell lysis is performed to separate the cytoplasmic fraction from the nucleus. The resulting cytoplasmic lysate contains a mixture of free proteins, large (60S) and small (40S) ribosomal subunits, full ribosomes (80S) bound to mRNAs or not, intact or degraded non-translated mRNAs associated with RNA binding proteins forming ribonucleic particles or mRNPs complexes. All these molecules are then separated depending on their respective densities by ultracentrifugation through a sucrose gradient. The heaviest molecules tend to migrate faster through the gradient leaving the lightest molecules (free fraction) on top of the tube. After ultracentrifugation, the gradient is splitted in several fractions while the absorbance at 254 nm is recorded to keep track of the positions of all molecules containing nucleic acids. Each RNA containing species thus can be assigned to the different fractions of the sucrose gradient and purified independently.

Once purified from the sucrose gradient fractions, the mRNAs can be used directly in Northern Blot assays or reverse transcribed into cDNA for quantification by various methods including qRT-PCR, microarrays or high-throughput sequencing (Del Prete et al., 2007; Ruan et al., 1997). Beside the quantification of ribosome-bound mRNAs, polysome profiling is also useful to identify transcripts submitted to translational control (Beilharz and Preiss, 2004). Using polysome profiling combined with microarrays, Arava and his colleagues were the first to confirm the importance of translational control for all mRNAs expressed in yeast. Notably, they observed that the ribosome loading for most mRNAs was well below what would be expected if no regulation was occurring (Arava et al., 2003). They also discovered that for some genes, most of the corresponding mRNAs are associated only with a single ribosome or not engaged in translation at all. This observation thus revealed that translational control could have a stronger impact for specific subsets of mRNAs expressed in a cell.

### ***Quantitative proteomics bring a new view of the cell proteome regulation***

In order to characterize the impact of the different post-transcriptional control mechanisms on global gene expression, the best approach is to compare total vs translatable mRNA levels combined with proteomics analysis (Mata et al., 2005). As a matter of fact, the number of mRNA molecules existing in a cell at a specific time is controlled by both mRNA synthesis and degradation rates. Consequently, the number of translated mRNAs could reflect directly the protein synthesis rates. Additionally, the comparison between the translational activity and protein abundance depicts the impact of protein stability on the gene expression pattern. The first integrative study that analyzed these different parameters genome-wide was performed by Beyer et al., in 2004. For this, they combined data obtained by several groups working on gene expression regulation in yeast. To get robust mRNA quantification, they combined 36 microarray datasets with SAGE results described previously (Futcher et al., 1999; Greenbaum et al., 2003; Gygi et al., 1999). For protein quantification, the dataset obtained was less robust due to a reduced number of whole proteome studies and the low reliability of the methods used. They managed however to characterize the impact of translation regulation and protein stability on the abundance of 1669 proteins. They observed that in spite of a general tendency for homodirectional changes of mRNA and protein synthesis levels, protein abundance is weakly correlated to the number of mRNAs engaged in translation determined by polysome profiling. Therefore, they concluded that protein stability plays a large part in the modulation of yeast proteome composition. This conclusion was rejected a few years later, following the development of quantitative proteomics, in a study revealing that most proteins expressed in yeast are quite stable through time (Christiano et al., 2014).

Another explanation is that polysome profiling is not appropriate to assess the global translational activity as it underestimates the effect of translation elongation regulation on the protein synthesis process. They also confirmed previous observations that some mRNAs are subjected to suppressed translation under normal conditions. For these particular mRNAs, they suggest a mechanism of “translation on demand” where protein synthesis rates could be enhanced in response to environmental cues in order to give the cell more adaptability. Furthermore, they also discuss the link between differential translation regulation and protein subcellular localization and functions. Interestingly, mRNAs encoding regulatory proteins tend to be translated at very low rates. This conclusion unveils translational control as a powerful regulatory mechanism which could alter the cell phenotype by modifying the expression of proteins that impact the whole gene expression pattern.

To confirm the results obtained previously in yeast, but using mammalian cells this time, Schwanhäusser et al. took advantage of novel approaches to quantify simultaneously mRNA and protein synthesis rates and stability (Schwanhäusser et al., 2011). For this, they performed a parallel metabolic pulse labelling using Stable Isotope Labelling by Amino acids in Cell culture (pSILAC) to discriminate newly synthesized proteins from the pre-existing ones and the nucleoside analogue 4-thiouridine (4SU) to tag newly transcribed mRNAs. By measuring both protein and mRNA turnover, they expected to obtain a better picture of the impact of the different layers of regulation on the global gene expression levels. Their analysis indicates that 40% of the variations in protein abundances could be explained by the modification of mRNA levels, mostly through transcriptional regulation. Regarding protein stability effect, they suggest that its impact is rather small in mouse fibroblasts contrary to what was initially observed in yeast (Beyer et al., 2004). Moreover, they conclude that protein abundances are mainly controlled by regulation at the translational level. They also describe groups of genes with similar combinations of mRNA and protein stability that share common functions. Notably, genes encoding unstable mRNAs and proteins are strongly enriched in transcription factors, signaling proteins and chromatin modifying enzymes. Additionally, they predict that the effect of the different regulatory mechanisms would be different depending on the mRNA and protein turnover rates. Hence, translational control would have a bigger impact on genes encoding unstable proteins independently of their corresponding mRNA stability.

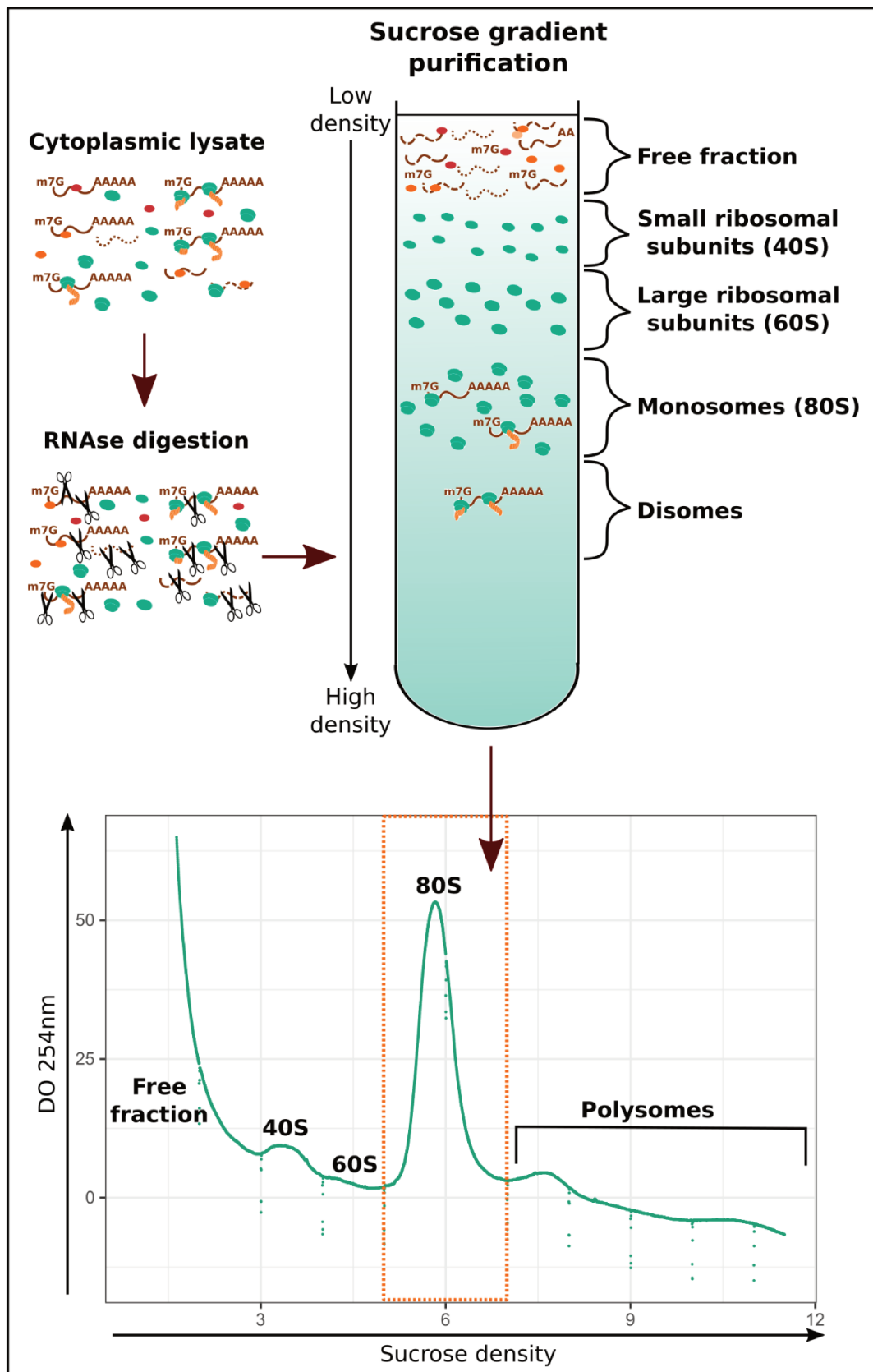
This study also demonstrates how the advances in large-scale quantitative proteomics contributed to the understanding of gene expression regulation. Indeed, by the 2010s the recent technical improvements allowed the systematic quantification of absolute abundances for thousands of proteins in a single experiment (Vogel and Marcotte, 2012). The proper analytical treatment of such datasets was nevertheless a complicated task (Liu et al., 2016; Wang et al., 2019). The results described by Schwanhäusser et al. were actually biased by an error made during the calculations of protein abundance estimates. Despite the publication of a corrected version in 2013, the re-analysis of the datasets by another group revealed that the calibration of the protein levels was inaccurate leading to an underestimation of the less abundant protein levels (Jingyi et al., 2020; Schwanhäusser et al., 2013). After rescaling the protein dataset and taking into account experimental errors, the mRNA levels explained at least 56% of the protein levels. Moreover, translational control could explain 30% of the variations of protein concentrations while transcription accounted for 38% of the differences. Transcriptional control is thus the primary determinant of gene expression patterns. Translational control is yet the main post-transcriptional regulatory mechanism involved as mRNA stability represents only 18% of the variations.

### ***Ribosome profiling provides a snapshot of translation rates***

To better estimate protein synthesis rates, a new method relying on high-throughput sequencing to quantify the levels of mRNAs actively translated, ribosome profiling, was developed in 2009. The short term goal of this approach was to circumvent the lack of reliability of contemporary quantitative proteomics that was particularly strong for the less abundant proteins (Ingolia et al., 2009). Its impact was greater as it opened the possibility of uncovering new regions of the genome that participate in the whole gene expression process such as upstream Open Reading Frames or uORFs (McGeachy and Ingolia, 2016). It also can be used to study global and transcript-specific translational control mechanisms and to localize all ribosomes bound to mRNAs at a specific time in a cell. This approach takes advantage of the ability of the ribosome complex to protect a portion of the mRNA being translated from RNase mediated degradation. After digestion, the remaining portions of mRNAs effectively protected by a ribosome are selected through sucrose gradient ultracentrifugation and sequenced at high rates to give a detailed picture of the ribosomes positions on all cellular mRNAs (Figure 3). These ribosome footprints (RPFs) are the direct reflection of the translational status of each mRNA as they are generated by ribosomes involved in any stage of the protein synthesis process (Initiation, Elongation, Termination or Stalling). As a consequence, an increased ribosome density in a specific region can point to a slowly translated or pausing sequence within a mRNA. Ribosome profiling thus provides a more consistent strategy to measure

translation efficiency than polysome profiling. This was further confirmed through the parallel measurements of protein synthesis levels using ribosome profiling and pSILAC (Liu et al., 2017). The correlation between RPFs abundance and protein synthesis rates measured by quantitative mass spectrometry was really good in steady-state conditions ( $R=0.8$ ). The results were less convincing upon the modification of the cell environment demonstrating that this approach alone is still not sufficient to predict dynamic remodeling of protein concentrations.

The use of ribosome profiling in different model organisms revealed that elongation is also critical to modulate protein synthesis rates as RPFs density along the coding sequence (CDS) can change depending on the transcript translated (Riba et al., 2019). As ribosome footprints on a mRNA can be produced both by elongating or stalled ribosomes, RPFs density cannot be taken as a direct measure of protein synthesis rates. Although combining mRNA deep sequencing and ribosome profiling is a good approximation of protein synthesis rates, proteomics approaches should not be neglected to obtain a complete picture of the gene expression process.



**Figure 3: Overview of the ribosome profiling protocol.** The cytoplasmic lysate is submitted to controlled RNase digestion to generate ribosomal footprints. After RNase treatment, the properly digested ribosomes are separated from the other RNA containing molecules by ultracentrifugation through a sucrose gradient. Only the fractions corresponding to the monosomes are selected for further processing.

To conclude, looking at mRNA levels can be informative to predict if a protein is likely expressed or not in a cell at a particular time. However, focusing only on transcription regulation gives a partial view of the gene expression process and is not sufficient to predict protein levels (Liu et al., 2016; Vogel and Marcotte, 2012; Wegler et al., 2020). To fully characterize gene expression patterns involved in different stages of cell life, it is important to integrate as well post-transcriptional control and protein turnover regulation. This is particularly critical when studying gene expression modifications induced by a change in the cell environment (Bludau and Aebersold, 2020; McManus et al., 2015). For example, when cells are exposed to stress conditions, their protein synthesis levels can be re-shaped in order to respond properly to their changing environment. In such conditions, the correlation between mRNA and protein levels can be further reduced. Moreover, translational control could also buffer fluctuations in mRNA abundances to maintain the expression of essential proteins at constant levels upon a change of state (Kozlovski and Agami, 2019; Lorent et al., 2019). All these observations reveal how the regulation of protein synthesis is complex and requires precise mechanisms to adapt the cell proteome depending on the conditions.



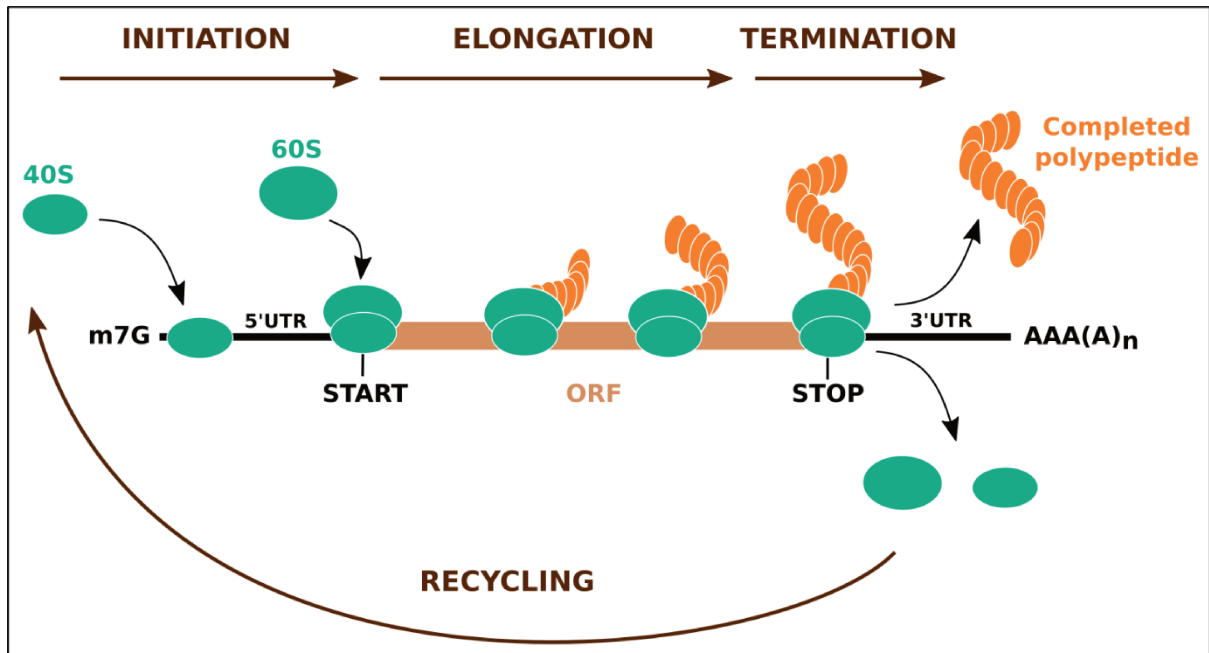
## Chapter II. Dynamic regulation of protein synthesis rates

The regulation of synthesis rates for each individual protein is quite complex as it is the result of the entire gene expression flow that can be modulated at different levels. Additionally, the process of protein synthesis itself can be modulated depending on environmental cues, providing a strategy to adapt protein abundances at the global scale. This chapter will focus more specifically on mRNA translation and the various mechanisms involved in its control.

### II. 1. General principles of mRNA translation

#### *Universal features of protein synthesis*

Early studies aiming to decipher the protein synthesis mechanism were performed using bacteria and mainly revealed universal mechanisms that are similar between prokaryotes and eukaryotes. The basic principles of translation are in fact conserved through all kingdoms of life. Notably, protein synthesis is among the most energy consuming processes in the cell (Buttgereit and Brand, 1995). As a consequence, the synthesis rates are heavily affected depending on nutrient availability in the environment. Every organism has several regulatory pathways to adapt the levels of translation according to the energy and amino acids supplies. Particularly, protein synthesis rates can be adapted by tuning off mRNA levels and their ribosome loading. The protein synthesis process itself is a conserved sequence composed of four main steps (Figure 4). The first step, initiation, consists in the recruitment of the ribosome to the mRNA start site which is composed of a set of three specific consecutive nucleotides or codon. During the next step, elongation, the new polypeptide chain is synthesized as the ribosome decodes the information contained in the open reading frame (ORF) of the mRNA. Finally, when the ribosome encounters a stop codon (UAG, UGA or UAA), the polypeptide chain is released by the help of termination factors and the ribosome subunits can be recycled to perform new rounds of translation (Termination and Recycling). Alternatively, the ribosome can remain attached to the translated mRNA to perform a new round of protein synthesis (Reinitiation) (Skabkin et al., 2013). As this entire process has a high energy cost, it was reasoned that initiation should be the most rate-limiting step to prevent useless energy expense if the ribosome was unable to fully complete protein synthesis. The development of ribosome profiling revealed that elongation and termination steps could in fact have a significant impact despite the primary dependency on the initiation step (Riba et al., 2019). This is particularly true in eukaryotes where several layers of gene expression control can interact to shape the cell proteome.



**Figure 4. Global view of the translation process.** During initiation, the two ribosomal subunits (40S and 60S) are sequentially recruited to the mRNA with the help of initiation factors. The 40S subunit scans the 5'UTR region of the mRNA until it reaches a start codon localized in a good nucleotidic context. The 60S subunit is recruited following 40S binding to the start codon. During elongation, the ribosome translates the open reading frame (ORF) into a polypeptide chain. Termination occurs when the ribosome reaches a stop codon, leading to the release of the polypeptide chain. Finally, the dissociated ribosomal subunits can be recycled to perform another round of translation.

### ***The mechanism of translation in eukaryotes***

The increased complexity of the translation process in eukaryotes compared to prokaryotes is transposed by the requirement of more proteins to coordinate the different steps. In addition to core proteins that are essential for proper translation, other factors can play regulatory or organizational roles (Fraser, 2015). For instance, biochemical studies performed in yeast revealed the existence of numerous eukaryotic-specific initiation factors (Aylett and Ban, 2017). Besides core initiation proteins, the ATP-dependent helicases eIF4a, DDX3 and DHX29, the GTPase activating protein eIF5 and the multi-protein scaffolds eIF3 and eIF4 participate in the recruitment and positioning of a ribosome at the proper start site on a mRNA. In fact, translation initiation is the most complex step of protein synthesis in eukaryotes as it involves the coordinated recruitment of many factors in a sequential pattern (Figure 5).

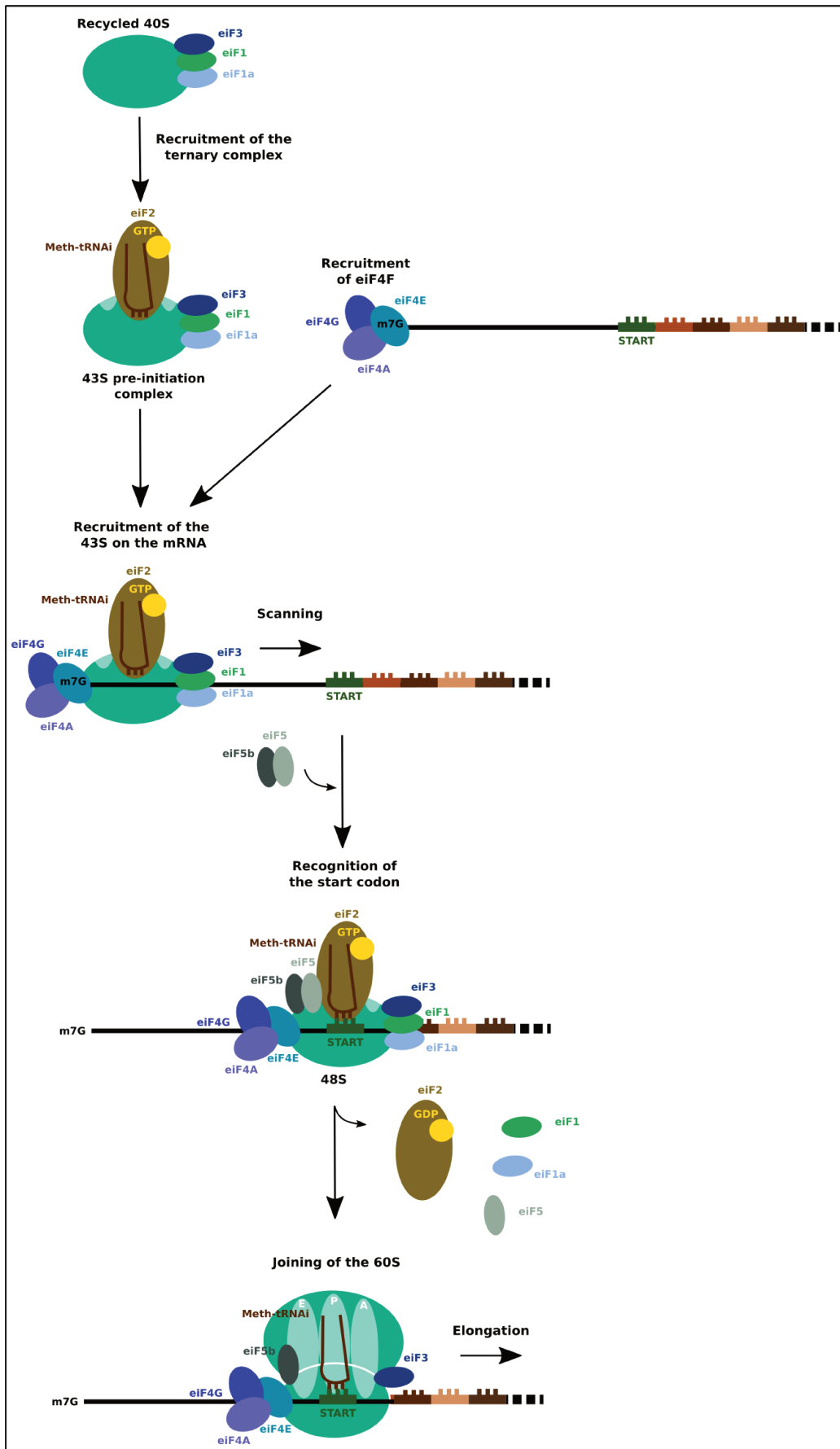
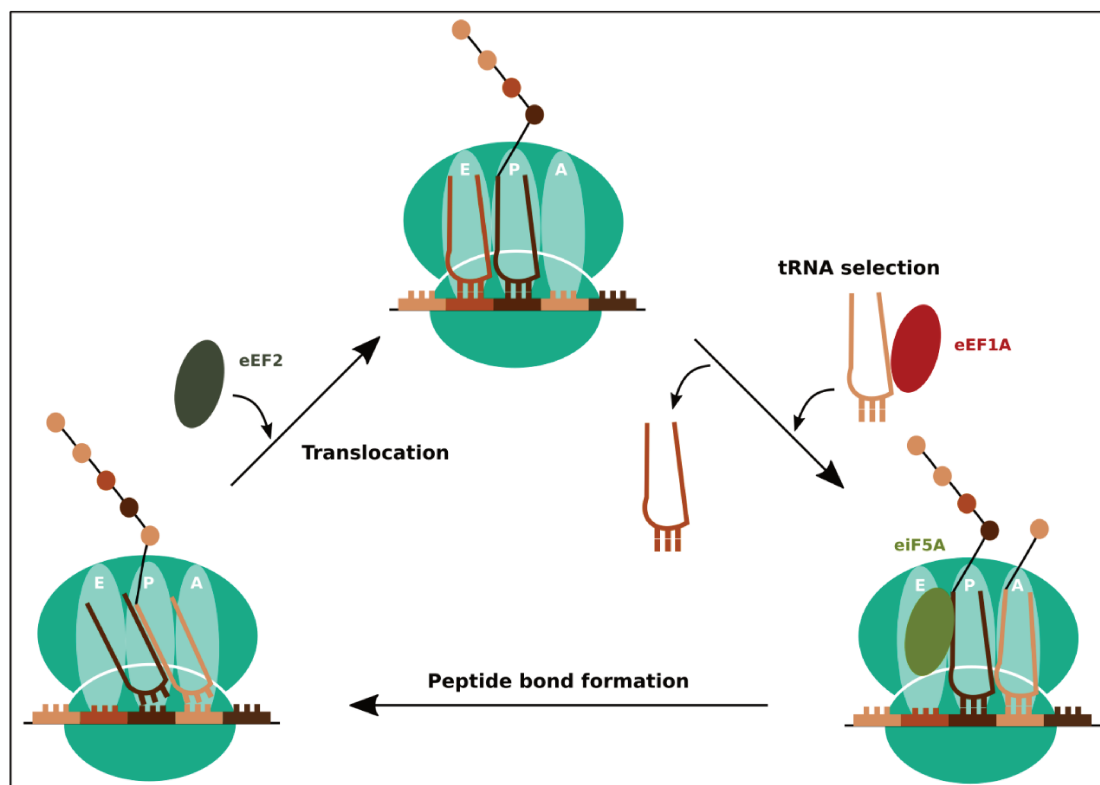


Figure 5 : The canonical mechanism of translation initiation in eukaryotes.

**Figure 5 : The canonical mechanism of translation initiation in eukaryotes.** The canonical cap-dependent initiation of translation requires the recognition of the m7G cap of the mRNA by eIF4E, a component of the multiprotein complex eIF4F. In parallel, the small ribosomal subunit bound to eIF3, eIF1 and eIF1a is associated to the ternary complex, composed of eIF2-GTP bound to the Methionine-tRNA initiator (Meth-tRNA<sub>i</sub>), to form the 43S pre-initiation complex (PIC). The 43S PIC is recruited to the mRNA through interaction with eIF4F and starts scanning the 5'UTR until it reaches a start codon. Upon start codon recognition, eIF2 bound GTP is hydrolyzed to GDP leading to a conformational change that stabilizes the interactions between the initiation complex and the start codon. Consequently, eIF2-GDP, eIF1, eIF1a and eIF5 are released while eIF5b promotes the joining of the 60S subunit. Once the 80S initiating complex is completed, the ribosome enters the elongation cycle and remaining initiation factors no longer required are progressively detached.

First, the small subunit of the ribosome must be rendered competent for mRNA binding. For this, the 40S subunit must be bound by eIF1, eIF1A and eIF3 (Jackson et al., 2010). The subsequent recruitment of the ternary complex, composed of eIF2-GTP and a methionine tRNA initiator (Meth tRNA<sub>i</sub>) leads to the formation of a 43S complex. The ribosome binding step depends both on specific features of the mRNA and the initiation factors involved. In most cases, the ribosome is recruited to the mRNA through binding to its m7G cap with the help of eIF4F scaffold and must scan the 5' untranslated region (UTR) before committing to the first start codon in a good nucleotidic context (Kozak, 1991a, 2002). The recruitment of the mRNA to be translated is mainly ensured by the cap-binding protein eIF4E, a component of eIF4F complex (Sokabe and Fraser, 2019). The selection of the start codon is next controlled by the combined action of eIF1, eIF2 and eIF5. Alternatively, the ribosome can bind to the mRNA independently of the m7G cap through highly structured sequences in the 5'UTR called internal ribosome entry sites (IRES). The presence of modifications on the position N6 of adenosines (m6A) in the 5' end of the mRNA could also promote ribosomal binding independently of the cap (Zhou et al., 2018). Once positioned at the proper initiation codon, the 60S ribosome subunit is recruited with the help of eIF5b to form a 80S complex ready for translation elongation. The initiation factors that are no longer required, including eIF1, eIF1a, eIF2 and eIF3, are progressively detached from the ribosome during the early elongation phase (Jackson et al., 2010).

Elongation is the most evolutionarily constrained step of translation (Burroughs and Aravind, 2019). Importantly, the fidelity of addition of new amino acids in the growing peptide chain according to the information carried by the mRNAs must be constantly evaluated. As a consequence, this process must be highly specific to avoid any mistakes that could lead to protein malfunction. Protein chain elongation is therefore achieved through a controlled sequence of three steps repeated until the ribosome reaches a stop codon: tRNA selection, peptide-bond formation, and translocation of the mRNA and associated tRNAs (Figure 6). Each step involves several interactions between the different components of the translation reaction that act as conformational checkpoints to control its accuracy (Dever et al., 2018).

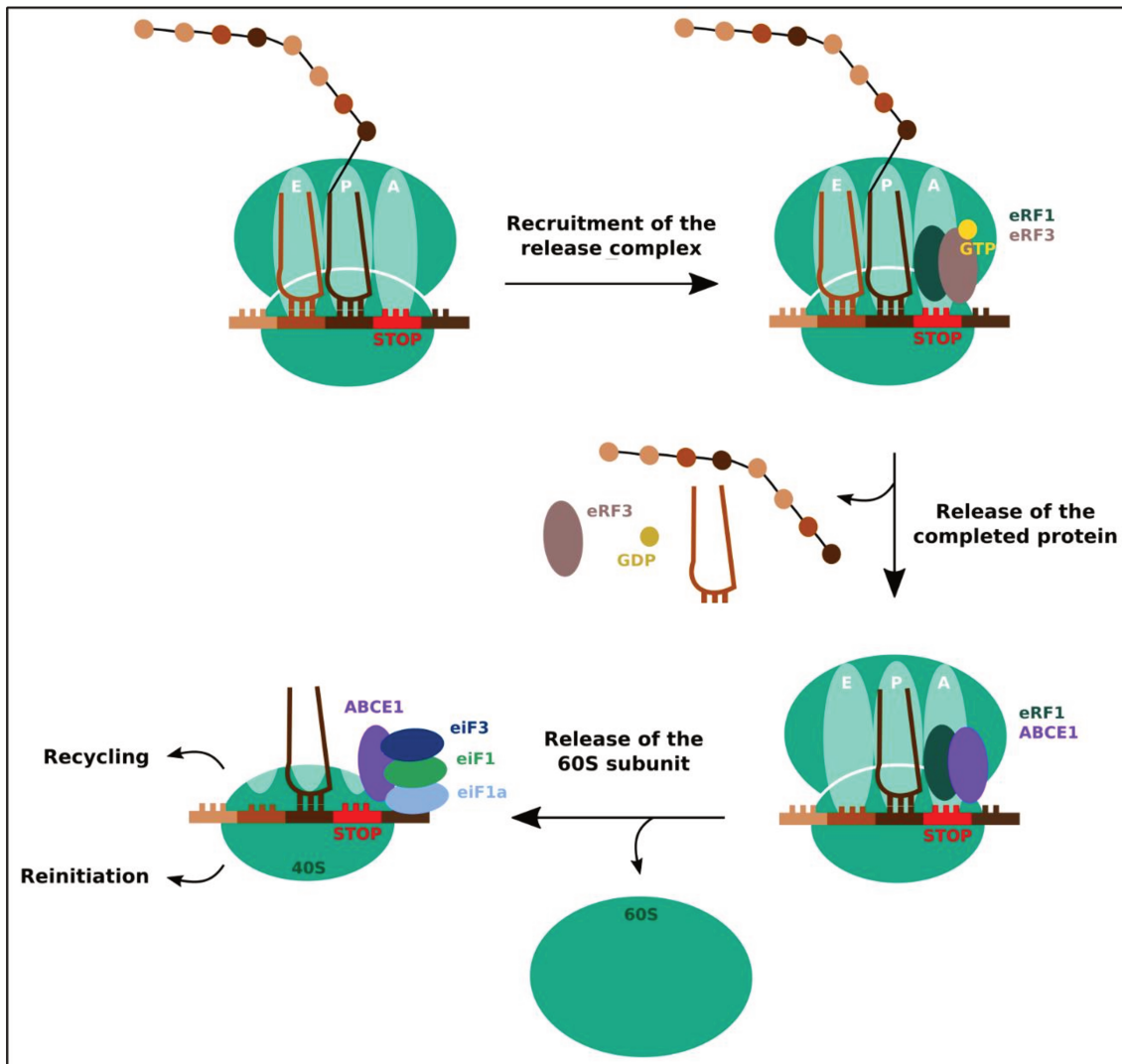


**Figure 6 : Schematic representation of the translation elongation cycle.** During the first step of elongation, an aminoacyl tRNA corresponding to the mRNA codon localized in the ribosome A site is recruited with the help of eEF1A. Meanwhile eiF5A enters the E site to ensure the favorable positioning of the nascent peptide chain in order to allow the subsequent addition of a new amino acid. The proper binding of the aminoacyl tRNA to the codon in the A site triggers a conformational change leading to the peptide bond formation. During this reaction, the elongated peptide chain is transferred from in the tRNA in the P site to the acceptor stem of the aminoacyl tRNA newly recruited. At the same time, the ribosomal subunits rotate, leaving the attached tRNAs in hybrid states between the different ribosomal sites (P/E and A/P). During the last step, the translocation of the next mRNA codon in the A site is catalyzed by eEF2 ensuring the complete relocalization of the attached tRNAs in the E and P sites. At the end of the cycle, the A site is thus empty again and ready for the recruitment of a new aminoacyl tRNA. A : Aminoacyl site ; P : Peptidyl site; E : Exit site.

Ribosomes contain three tRNA-binding regions: the aminoacyl or A site, the peptidyl or P site, and the exit or E site. At the end of initiation, the methionine tRNA initiator interacting with the start codon is localized in the P site of the ribosome. To add a new amino acid corresponding to the next mRNA codon, the cognate aminoacyl tRNA is recruited to the A site with the help of the elongation factor eEF1A (Proud, 1994). When the aminoacyl tRNA is properly bound to the mRNA codon, its accommodation in the ribosome allows the formation of the peptide bond catalyzed by the peptidyl transferase center (PTC). Meanwhile, the favorable positioning of the peptidyl tRNA is ensured by the binding of eiF5A in the E site (Dever et al., 2018). During the peptide bond formation, the nascent peptide is transferred from the tRNA in the P site to the acceptor stem of the aminoacyl tRNA. At the same time, the ribosomal subunits rotate, leaving the attached tRNAs in hybrid states between the different ribosomal sites (P/E and A/P) (Joseph, 2003). Translocation of the next mRNA codon in the A site is then catalyzed by eEF2 leading to the complete relocation of the tRNAs in the E and P sites (Noller et al., 2017). At the end of this process, the tRNA carrying the growing chain is thus in the P site while the A site is ready to interact with a new aminoacyl tRNA.

The termination step begins when a stop codon enters the A site (Figure 7). The release complex eRF1/eRF3-GTP is recruited to this empty site and catalyzes the hydrolysis of the bond between the elongated peptide chain and the tRNA located in the P site (Jackson et al., 2010). Upon peptide release, eRF1 remains bound while eRF3 is detached. The recycling factor ABCE1 in mammals, or Rli1 in yeast, then interacts with eRF1 to promote the dissociation of the 60S subunit (Skabkin et al., 2013). The remaining 40S is thus still bound to the mRNA, tRNA and ABCE1 recycling protein that can interact with the initiation factors eiF1, eiF1A and eiF3 (Heuer et al., 2017). This interaction is probably critical to prepare the small subunit for a new round of translation initiation on the same or a different mRNA. The selection of reinitiation instead of recycling of the ribosomal subunits depends on the CDS length and the kinetic binding of initiation factors (Sokabe and Fraser, 2019). Interestingly, reinitiation events are more frequent for short ORFs or uORFs where eiF3 is not properly detached from the elongating 80S before it reaches a stop codon (Kozak, 1987; Mohammad et al., 2017). Consequently, reinitiation rates decrease quite abruptly with increasing length of the uORFs (Kozak, 2001). Additionally, the binding of the eiF4F complex could also promote translation reinitiation. Indeed, eiF4F can interact with both mRNA ends through binding to its m7G cap in the 5' end and to the polyA binding protein (PABP) associated to the 3' end of cellular mRNAs. This circularized conformation could ensure the rapid recruitment of 40S subunits ready for translation initiation to the close by 5' end (Marshall et al., 2014; Sokabe and Fraser, 2019). Moreover, eEF2 binding could also promote the translocation of the terminating 80S

ribosomes before ABCE1 recruitment allowing them to reinitiate without dissociation (Skabkin et al., 2013).



**Figure 7: Schematic representation of the translation termination process.** Translation termination is triggered when a STOP codon enters the A site of the ribosome. The STOP codon is recognized by the release complex formed by eRF1 and eRF3 bound to GTP. The proper binding of the release complex in the A site catalyzes the hydrolysis of the GTP and the ester bond linking the polypeptide chain to the peptidyl tRNA. The release of the newly produced protein also induces the dissociation of eRF3 and the GDP from the ribosome. The recycling factor ABCE1 then binds to eRF1 and induces the release of the large (60S) ribosomal subunit from the 80S complex. The initiation factors eIF1, eIF1A and eIF3 recruited by ABCE1 facilitate the subsequent separation and recycling of the deacylated tRNA, mRNA, and small (40S) ribosomal subunit. Alternatively, the 40S can be recruited for reinitiation on the same mRNA.

In some cases, a non-cognate aminoacyl tRNA can enter the empty A site when the ribosome reaches a stop codon allowing the elongation step to continue. In fact, specific sequence features could also impact the strength of a stop codon to promote translation termination. These readthrough events could be favored by a particular nucleotide sequence context around the termination codon (Cassan and Rousset, 2001). The sequence of the termination codon itself along with the presence of secondary structures or modifications in the mRNA could also alter the recognition of the stop codon (Sokabe and Fraser, 2019). A fascinating example is the effect of highly structured sequences in the 3'UTR called SECIS, for selenocysteine insertion sequences, that induce the addition of a selenocysteine amino acid, instead of arrest of the ribosome on the stop codon (Vindry et al., 2018). This recoding event, conserved in both prokaryotes and eukaryotes, is essential for the expression of selenoproteins triggered when selenium is incorporated in the cell.



## II. 2. Mechanisms of translational control

### ***Cellular resources limit global protein production levels***

Protein neo-synthesis relies primarily on the availability of ribosomes that can be engaged into the translation process in both basal and dynamic conditions. When describing ribosomes for the first time from electron microscopy pictures of animal cells, Palade already observed that ribosome abundances were quite different between quiescent and highly proliferating populations (Palade, 1955). Several studies in bacteria and yeast have revealed that the cell capacity to tolerate increased demands for protein synthesis implies the production of more ribosomal particles (Kafri et al., 2015; Vind et al., 1993). The correlation between ribosome biogenesis and cell growth was also confirmed in metazoan using ribosome profiling (Ingolia et al., 2019). The number of ribosomal subunits that can be recruited for *de novo* translation initiation is thus an important parameter for the regulation of protein synthesis rates (Chu and von der Haar, 2012). Importantly, the formation of inactive 80S complexes in absence of mRNA can decrease global cellular protein synthesis capacities by sequestering ribosomal subunits. The formation of such complexes can however be protective during stress conditions as it reduces protein synthesis rates and limits the degradation of the ribosomal subunits (Brina et al., 2011).

The fact that most mRNAs are translated simultaneously by several ribosomes has opened the question of how ribosomes are distributed across mRNAs (MacDonald and Gibbs, 1969). Without any tool to answer this question in biological conditions, mathematical models were used, instead, to understand the principles of ribosome allocation. The most popular model to study translation dynamics was the totally asymmetric simple exclusion process (TASEP). Most of TASEP-based studies assumed a constant supply of free ribosomes and tRNAs in the cell (Shah et al., 2013). In such conditions, ribosome loading and speed along the mRNA are the major determinants of protein synthesis rates. Consequently, several patterns could be observed depending mainly on translation initiation rates. In most cases, ribosome loading would directly reflect protein synthesis levels providing that initiation and elongation rates are correlated. When initiation is not frequent, ribosome density and protein output would be low. Alternatively, when initiation rates are too high, ribosome density would increase possibly leading to ribosomes collisions that ultimately decrease the levels of protein produced. When confronted with experimental data obtained with modern genome-wide approaches, combined ribosome profiling and pSILAC, TASEP-based model of translation dynamics revealed to be quite robust (Riba et al., 2019).

The importance of translation initiation regulation to control protein production and reduce the energy expense due to improper ribosome recruitment was validated. However, no reduced protein production caused by a ribosome overload was detected in yeast. Another model taking into account known measurements of ribosome and tRNA concentrations alternatively points ribosome availability as the major limiting factor in the control of protein synthesis (Shah et al., 2013). The subsequent regulation of the initiation and elongation levels could thus participate in the adaptation of the ribosome allocation pattern depending on the conditions. Particularly, the amount of small ribosomal subunits and initiation factors can become limiting when the majority of them are engaged in translation (Dykeman, 2020). Under such conditions, the rates of recycling and reinitiation could play a significant role in controlling protein synthesis levels (Sokabe and Fraser, 2019).

The availability of tRNAs competent for translation elongation was also described as a major determinant of protein synthesis efficiency (Sharp and Li, 1987). Several tRNA species can carry the same amino acid despite recognizing a different codon on the mRNA. These synonymous codons are not represented at the same frequency across mRNAs. For sub-optimal codons, the frequency of the corresponding tRNA is low leading to a reduced translation efficiency. On the contrary, there are significantly more codons corresponding to abundant tRNAs in the most highly expressed transcripts in yeast (Tuller et al., 2010). This codon bias can thus impact transcript-specific elongation speed and is a highly conserved regulatory mechanism (Duret and Mouchiroud, 1999). Notably, it could contribute to the control of translation fidelity, protein folding and mRNA stability in many organisms (Hanson and Collier, 2018). The importance of this mechanism in mammals is however less clear as other layers of regulation may exert a stronger effect on translation (Ingolia et al., 2011; Plotkin and Kudla, 2011).

### ***Regulation of global protein synthesis by targeting of translation factors***

Protein synthesis control is quite complex, particularly in eukaryotes, and relies on regulatory mechanisms that can act at different levels. Global modifications of protein synthesis rates through the targeting of translation factors mainly involves the regulation of the initiation process (Gebauer and Hentze, 2004; Jackson et al., 2010) but also of elongation. In fact, many global regulatory mechanisms controlling translation factors have been described in eukaryotes but only a few examples will be described here to reduce the complexity of this demonstration.

An efficient strategy to rapidly and drastically reduce global protein synthesis levels is to block the activity of translation initiation factors. The regulation of eIF2 by reversible protein phosphorylation is among the most well described examples. This phosphorylation event is achieved by stress-induced kinases such as PKR, activated during the antiviral response, and PERK, activated upon accumulation of misfolded proteins (Jackson et al., 2010). Consequently, the number of competent 43S complexes and translation initiation rates are critically reduced. Translation elongation factors can also be phosphorylated to regulate their activity in mammals (Browne and Proud, 2002). Notably, eEF2 activity can be modulated through the mTOR (mechanistic Target Of Rapamycin) pathway to connect translation elongation rates to nutrient availability (Dever et al., 2018). The control of eEF2 by the inhibitory kinase eEF2K has been recently linked to a reduction of translational errors (Xie et al., 2019). The regulation of elongation speed is thus critical to maintain the efficiency of protein production without altering its accuracy. In addition to global regulations, translational control is not uniform and can also have transcript-specific impact depending on specific features of the mRNA (cis-regulation) or the binding of external factors (trans-regulation).

### ***Transcript-specific regulation dependent on specific features of the mRNA***

The fact that all mRNAs are not equal regarding ribosome loading efficiency has been a long standing assumption (Kozak, 1991a; Lodish, 1974). Using *in vitro* translation of genetically engineered mRNAs, Marilyn Kozak identified five features that can regulate translation initiation rates on a specific start codon : the presence of a m7G cap, the nucleotide context around the start codon, the position of the initiation site in the 5'UTR, the 5'UTR length and the presence of secondary structures (Kozak, 1991b). The importance of these features can be directly linked to the mechanism of translation initiation. The recruitment of ribosomal subunits on a mRNA depends on the recognition of the m7G cap by eIF4E and the presence of highly structured sequences in the beginning of the 5'UTR can alter the binding efficiency (Jackson et al., 2010; Sokabe and Fraser, 2019). These sequences can also slow down the ribosomal complexes scanning for the initiation site and modulate its detection. For instance, the insertion of a stem-loop structure between the cap and the first AUG codon can interrupt the scanning of the 5'UTR (Kozak, 2002). The nature of the nucleotides surrounding the start codon, or Kozak context, is also of great importance to promote the recognition of the initiation site. The optimal context was determined from the study of 699 vertebrate mRNAs as following : a purine must be placed three nucleotides (nt) upstream of the AUG codon and the first nucleotide after the start codon must be a G (gccA/GccAUGG) (Kozak, 1991a). Notably, the different start codons are not all as efficient to promote initiation : the use of non-AUG codons is less common even if the nucleotide context is optimal (Kozak, 2002). Moreover, a poor

Kozak context around the first AUG codon promotes the leaky scanning of the 43S complex and initiation on a downstream codon with a better context. This mechanism can thus lead to the expression of two proteins of different size from the same mRNA. The presence of upstream ORFs (uORFs) in the 5'UTR region can also impact the translation of the canonical ORF by subtracting 43S complexes for example. The effect of uORFs on initiation rates depends on their size and mechanically on the 5'UTR length (Kozak, 1987). A particularly well described example of such regulation corresponds to the regulation of GCN4 (in yeast) or ATF4 (in mammals) expression, restricted to stress conditions involving nutrient deprivation (Mueller and Hinnebusch, 1986; Vatter and Wek, 2004). In basal conditions, ribosomes are recruited to GCN4 or ATF4 mRNAs in a positive-acting uORF placed in the 5'UTR region. Inhibitory uORFs placed directly downstream the first uORF trap the recruited ribosomes and limit reinitiation events in the main ORF. GCN4 and ATF4 protein synthesis is thus inhibited. Upon nutrient deprivation, eIF2-GTP levels are decreased and the time required for reinitiation is increased thus allowing more ribosomes to reinitiate in the main ORF and the synthesis of GCN4 and ATF4 proteins. All these features reveal the importance of the 5'UTR region for efficient initiation on a specific mRNA (Li et al., 2019). Furthermore, an increased 3'UTR length could promote reinitiation by the formation of a closed loop or circularized conformation through eIF4G and PABP binding (Amrani et al., 2008; Costello et al., 2015). As a consequence, alternative splicing events that modify the UTR length or sequence can produce differentially translated transcripts originating from the same gene and encoding similar proteins. This notion was notably validated in human cells using high-throughput sequencing to identify the different mRNA isoforms expressed in a cell and deduce their ribosomal loading from a sucrose gradient (Floor and Doudna, 2016).

Most of the regulatory parameters postulated by Kozak were recently validated using the high throughput profiling of the small 40S subunits on cellular mRNAs (Giess et al., 2020). Nonetheless, this new approach, ribosome complex profiling or RCP-seq revealed that the Kozak context is not as optimal as expected for all mRNAs in zebrafish embryos. Indeed by purifying footprints produced by both 43S and 80S complexes after crosslinking on the mRNA and RNase digestion, it is now possible to identify precisely the sequence recognized for translation initiation on every cellular mRNA. As the context optimality was initially calculated from a subset of 699 vertebrate mRNAs, this discrepancy demonstrates how global predictions of translational control cannot recapitulate the complex regulation of translation initiation occurring *in vivo*. Particularly, transcript specific features could have a stronger impact on translational control than previously expected (Sharma et al., 2019).

Translational control during the elongation step can also heavily affect protein synthesis rates. This type of regulation may be more important for mRNAs for which initiation is not highly constrained (Sokabe and Fraser, 2019). In yeast, elongation rates are directly correlated to initiation levels for most transcripts (Riba et al., 2019). However, for some transcripts, the dynamics of translation elongation can be modified by cis-acting elements (Li et al., 2019). A well described effect is the presence of codons that correspond to low abundance tRNAs in the translated mRNA. These non-optimal codons reduce translation elongation speed and thus contribute to fidelity of translation (Hanson and Collier, 2018). The nature of the nascent protein is also of importance as cofolding of the polypeptide chain and the presence of positively charged amino acids have been linked to variation in elongation rates (Hanson and Collier, 2018; Tuller et al., 2010). The presence of local secondary structures along the mRNA sequence can also greatly modulate decoding speed to either increase or decrease translation efficiency (Mao et al., 2014; Mauger et al., 2019).

The cis-elements that slow down translation elongation rates can also cause ribosomal pausing. Short ribosome stalling can be resolved quickly by the conserved translation factor eIF5A and promote recoding events (Dever et al., 2018). For instance, stalling on slippery sequences can provoke a frameshifting meaning the recruitment of a codon that is not consecutive to the three previously decoded nucleotides in the A site. Another example is the readthrough events induced by topological features such as stem-loops or pseudoknots downstream of the stop codon. In addition to this, long pausing of the ribosome on aberrant mRNA sequences induces the termination of translation and triggers the activation of ribosome rescue pathways (Schuller and Green, 2018). When stalled for too long, the ribosome is disassembled through the recruitment of recycling factors such as Dom34/Hbs1 or Pelota/HBS1L and both the mRNA and the nascent protein are targeted for degradation. Notably, the recruitment of factors involved in the ribosome-associated quality control pathway (RQC) ensures the ubiquitination of the newly synthesized protein and its targeting to the proteasome (Joazeiro, 2019; Schuller and Green, 2018). For the defective mRNA, several decay pathways have been described depending on its features (Karamyshev and Karamysheva, 2018; Stein and Frydman, 2019). Non-sense mediated decay (NMD) is triggered when the ribosome encounters a premature stop codon. Alternatively, non-stop decay occurs when elongation is not ended at the stop codon and the ribosome is stalled within the poly-A tail. Finally, long pauses in the ORF due to mRNA truncation, secondary structures or rare codons activate the No-go decay pathway (Chandrasekaran et al., 2019; D'Orazio et al., 2019; Harigaya and Parker, 2010; Shao et al., 2015). As a consequence, translation rates can directly impact mRNA stability (Hanson and Collier, 2018; Presnyak et al., 2015). In fact, ribosomal flux across the translated mRNA is tightly regulated and stalling

actively participates in the control of gene expression (Stein and Frydman, 2019). For example in yeast, the helicase Dhh1 was described as binding to slow-moving ribosomes to trigger mRNA decay (Hanson and Collier, 2018; Radhakrishnan et al., 2016). All these mechanisms participate in the transcript specific regulation of elongation rates and thus contribute to the modulation of protein synthesis rates.

In addition to the different cis-acting mechanisms previously described, the length of the translated ORF has been highlighted repetitively as a critical parameter to regulate both ribosomal loading and protein synthesis rates (Jingyi et al., 2020; Riba et al., 2019). The mechanism involved is however not clear. Despite the established fact that the number of ribosomes recruited per mRNA is correlated to the length of the ORF (Tuller et al., 2010), it is not well understood how the size of the translated region can impact protein levels. The effect of mRNA length was described as important for the regulation of both initiation and elongation rates (Jingyi et al., 2020; Wang et al., 2013b). Another explanation could thus be that the ORF length regulates reinitiation rates and consequently ribosome recruitment on a particular mRNA (Fernandes et al., 2017; Rogers et al., 2017). To conclude, the differential trafficking of ribosomes depending on specific features of the cellular mRNAs could also be essential to ensure the dynamic regulation of protein synthesis levels.

### ***Trans-acting factors modulate specific mRNA translation efficiency***

Transcript-specific translational control also relies on the expression of trans-acting factors to allow a dynamic regulation of protein levels depending on the cellular environment. The binding of trans-regulatory factors to cis-regulating elements found in mRNAs can substantially modulate the translation rates depending on external cues. These trans-factors can be themselves targets of signaling pathways allowing an integration of external signals at different levels of the gene expression process (Sokabe and Fraser, 2019). This suggests a strong co-evolution between specific features of mRNAs and the factors that can recognize them to ultimately achieve a coordinated regulation of protein synthesis (Li et al., 2019).

Among the different trans-acting factors, the role of RNA binding proteins (RBPs) was assessed very early on in the study of protein synthesis control (Siomi and Dreyfuss, 1997). Many RBPs involved in the regulation of the different steps of a mRNA life were already described early in the 1990s. Interestingly, the hypothesis that RBPs activity could be modulated by post-translational modifications in response to a stimulus to modulate protein synthesis was also well discussed. More recently, the differential expression of RBPs between various cell types was connected to their ability to specify the proteome depending on the cell

functions (Corbett, 2018). Importantly, RBPs can recognize specific sequences or modifications such as m6A methylation on mRNAs to target them either for translation or degradation (Chen et al., 2019; Harvey et al., 2018).

Non coding RNAs, such as microRNAs (miRNAs) or long non coding RNAs (lncRNAs), are also well known trans-acting factors that can modulate the translation efficiency of specific mRNAs. These non-coding RNAs can interact with the mRNA associated with ribosomal subunits and translation factors to modulate their fate at different levels (Fabian et al., 2010; Noh et al., 2018). Notably, they can induce their degradation or remove them from the pool of translatable mRNAs. For example, the binding of miRNAs on specific sequences in the 3'UTR can considerably reduce translation initiation efficiency for a particular mRNA (Humphreys et al., 2005).

To conclude, the regulation of the translation process is very complex as it can be targeted at the global and transcript-specific levels using various mechanisms, through cis-regulating elements and/or the binding of trans-acting factors. All of these mechanisms are however required to ensure the fine tuning of protein synthesis rates depending on the cellular needs in a constantly evolving environment.

## **II. 3. Monosomes : overlooked players in translational control**

### ***Protein synthesis occurs mainly in polysomes***

Historically, monosomes were not expected to actively participate in the gene expression process. Indeed, for several decades polysomes were considered as the main and only relevant effectors for protein synthesis. This widely supported view came from the first studies performed in the 1960s to characterize the site of protein synthesis. By 1962, ribosomes were identified as a major component of the translation reaction but the mechanistic details were not well understood (Ts'o, 1962). The development of ultracentrifugation through a sucrose gradient had allowed the definition of different cytoplasmic fractions depending on the density and the shape of the macromolecular complexes involved. Using this approach, several groups revealed the existence of light and heavy ribonucleic particles in the cytoplasm of both prokaryotic and eukaryotic cells (Siekevitz and Zamecnik, 1981). Radioactive labelling of newly synthesized proteins was routinely performed using amino acids containing alternative isotopes such as C<sup>14</sup>. The first translation events characterized from bacterial lysates were associated with the light fraction or monosomes. However, it was not clear how a relatively small complex (230Å diameter) could polymerize large proteins corresponding to

mRNAs of more than 1000Å (Warner et al., 1963). In 1963, three different teams combined radioactive labelling and ultracentrifugation to identify the main site of *de novo* protein synthesis using mammalian *in vitro* translation systems (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). Few cell types could be used for such experiments : most of these assays were performed using rat liver extracts and rabbit reticulocytes lysate (RRL) as the protocols were quite well described. Both rat liver and RRL systems involved cells functionally optimized for high levels of protein synthesis. The results obtained by the three teams revealed that most of protein synthesis occurred in heavy polysomes structures where a single mRNA was translated by several ribosomes at the same time. Notably, the RRL is derived from specialized cells that have lost their nuclei and only translate a few mRNAs to produce high levels of two proteins : hemoglobin and peroxidase. Consequently in RRL, most protein neo-synthesis occurs in structures composed of five ribosomes bound to the hemoglobin mRNAs (Warner et al., 1963). As previous studies were made using less validated systems, it was concluded that the observations of monosomes actively involved in protein synthesis was most likely due to the degradation of polysomes by RNases (Raacke and Fiala, 1964). Moreover, as the translation mechanism was expected to be highly uniform, it was admitted that all proteins shall be synthesized in polysomes.

In an attempt to decipher how exogenous mRNA can be translated, Gierer and colleagues also added polyuridylic acid (poly-U) RNA to their *in vitro* translation reactions. They observed that protein synthesis rates in the polysomal fractions were not dramatically modified in presence of poly-U. On the contrary, poly-U addition greatly stimulated the incorporation of radioactive phenylalanine amino acids in the monosome fraction. They concluded that monosomes could be easily recruited for a pioneer round of translation initiation upon addition of an exogenous mRNA probably because most of them were not bound to mRNAs in the first place (Gierer, 1963). Once the translation of the mRNA is properly initiated, additional ribosomes could be recruited leading to a relocalization to the polysomal fractions. This vision of the monosome fraction as a transition state between inactive particles and initiating ribosomes while actively translating ribosomes are restricted to polysomes prevailed for many years (Gebauer and Hentze, 2004; Heyer and Moore, 2016; Liu and Qian, 2016).

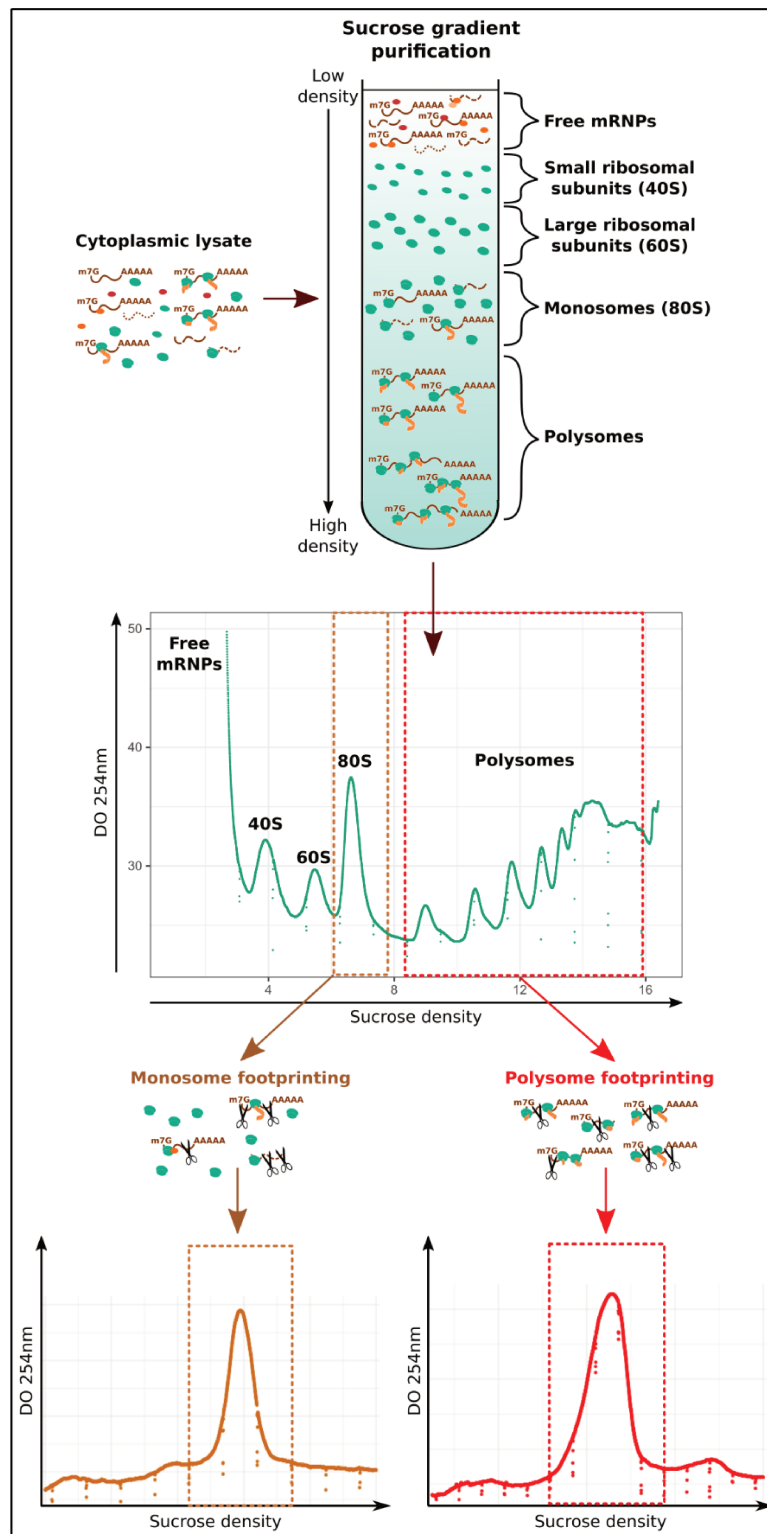


### ***Monosomes participate in the translation of highly regulated mRNAs***

The preponderant role of polysomes in the translation of most cellular transcripts was further validated in 2003 using the state of the art of the available techniques to study mRNAs : polysome profiling combined with microarrays (Arava et al., 2003). While the majority of cytoplasmic mRNAs were associated with several ribosomes, the authors also identified two subsets of transcripts with an unexpected behavior. Few mRNAs were not interacting at all with ribosomal subunits suggesting that they were stored in the cytoplasm while non translationally engaged. Furthermore, some mRNAs were associated with a single ribosome despite the excess of free ribosomal subunits. Indeed, the results showed that at least 85% of ribosomes are bound to mRNAs in actively growing yeast. The number of ribosomal subunits was therefore not likely limiting in these cells. These observations led to the conclusion that some transcripts are subjected to particular translational control mechanisms that reduce their ribosomal binding levels. To support this hypothesis, they further described three monosome-bound mRNAs corresponding to proteins for which translational control was already well described (GCN4, CPA1 and ICY2).

The existence of translationally repressed mRNAs was already discussed previously. Notably, the relative inefficiency of some cellular mRNAs to recruit new ribosomal subunits was well described in vertebrates (Kozak, 1991a). Interestingly, mRNAs with unfavorable features that reduce their ribosome loading capacities mainly encode regulatory proteins whose expression must be tightly controlled depending on the conditions such as growth factors, kinases, transcription factors and cytokines. Upon the discovery of RNA-binding proteins roles in the mRNA life, it was additionally suggested that their binding could prevent specific mRNAs from the recruitment to the translational apparatus (Siomi and Dreyfuss, 1997). In the response to a stimulus, the masking proteins could be removed to allow the rapid translation of mRNAs previously stored in the cytoplasm. To what extent these rather exceptional and transcript-specific events could have an impact on the global cell phenotype was however completely unclear.

To better characterize how this type of regulation can impact global protein synthesis rates, it is critical to determine the features that affect both ribosome recruitment and elongation speed depending on the transcript. The development of ribosome profiling was pivotal for this as it provides a direct and accurate measure of the ribosomal density along all cellular mRNAs. However, the ribosomal footprints are produced independently of the number of ribosomes initially bound to the mRNA. As a consequence, it is not possible to infer if a specific mRNA is more associated with monosomes or polysomes using this approach. To keep track of the ribosomal loading efficiency and quantify the ribosome occupancy, the best strategy is to combine polysome profiling with ribosome profiling. A new approach based on this principle, monosome vs polysome footprinting, was developed in 2016 to decipher the translational status of monosomes in yeast (Heyer and Moore, 2016). For this, the monosome or polysome bound mRNAs are first separated by ultracentrifugation through a sucrose gradient as for polysome profiling. The two pools are then subjected to RNase digestion before purification on a second sucrose gradient to select specifically the mRNA regions protected by the ribosomes (Figure 8).



**Figure 8: Overview of the monosome vs polysome footprinting protocol.** Monosomes and polysomes are selected after sucrose gradient ultracentrifugation of the cytoplasmic lysate. The different fractions corresponding to each ribosomal population are pooled separately, diluted to reduce sucrose concentration, concentrated on a Amicon-Ultra filtration column and submitted to RNase digestion in similar conditions to ribosome profiling. After RNase treatment, only the properly digested ribosomes (monosomes) are selected through a second sucrose gradient purification before further processing.

The results revealed that monosomes are not exclusively inactive or initiating ribosomes (Heyer and Moore, 2016). On the contrary, most of them generate RPFs across the entire ORF of the associated mRNAs demonstrating that they are in fact elongating ribosomes. A similar observation was recently made in neuronal cells derived from rodents (Biever et al., 2020). Overall, the number of associated ribosomes is increased proportionally to the ORF length. The same relationship was previously observed using polysome profiling (Arava et al., 2003). Hence, monosomes ensure the translation of small mRNAs (<100nt) that could not accommodate more than one ribosome and other short ORFs including uORFs. Additionally, some mRNAs are preferentially bound to monosomes despite being long enough to accept several ribosomes. These particular mRNAs encode for low-abundance regulatory proteins such as kinases and transcription factors whose expression must be highly controlled. The physiological relevance of monosomes mediated translation in the gene expression process is thus more important than expected previously : they could play a particular role for the translation of highly regulated mRNAs.

Moreover, monosome-bound mRNAs are less stable than those associated with polysomes in yeast (Heyer and Moore, 2016). While many of them are targeted by the NMD degradation pathway, it is not the case for all monosome-bound mRNAs. Other features such as cis-elements or the association to particular RBPs could thus explain the reduced stability of these mRNAs. To go further, the authors suggest that the differential association to monosomes or polysomes depends on the ratio between initiation and total elongation time. Consequently, if the initiation step is not highly regulated and faster than elongation, then several ribosomes could be loaded on the same mRNA. Inversely, when initiation time is increased and elongation not so controlled, the mRNAs are preferentially monosome-bound. This could explain how protein synthesis levels can be modified depending on differential ribosome occupancy. Consistent with this hypothesis, most of the abundant proteins are preferentially synthesized by polysomes. Alternatively, proteins that must be expressed at low levels and for a relatively short time period are preferentially produced by monosomes. To conclude, mRNA translation through monosomes is also relevant in the gene expression process as they could ensure the translation of specific transcripts encoding for regulatory proteins that can modulate the global cell phenotype.

## ***Alternative cellular functions of monosomes***

As demonstrated in yeast, only a small fraction of cellular mRNAs are preferentially translated by monosomes in basal conditions (Arava et al., 2003; Heyer and Moore, 2016). It should also be noted that not all ribosomes composing the monosomal fraction are actually active in protein synthesis. In addition to ribosomes that initiate the pioneer round of translation, monosomes can be bound to mRNAs targeted for degradation. Consistent with this assumption, many NMD targets or improperly spliced mRNAs are associated with monosomes in yeast (Heyer and Moore, 2016). In fact, the monosome compartment could be enriched in mRNAs that provoke aberrant translation events requiring the recruitment of the ribosome quality control (RQC) pathway and associated mRNA decay pathways. Indeed, mRNA stability can be directly impacted by translation levels (Hanson and Collier, 2018). Consequently, when translation initiation or elongation rates are dramatically reduced as it is possibly the case for monosome bound transcripts, the probability to observe ribosomal pausing leading to the mRNA degradation is increased. Supporting this view, several recent studies confirmed the widespread coupling between cytoplasmic mRNA decay and the protein synthesis process (Collart and Weiss, 2020; Pelechano et al., 2015). Notably, cotranslational and ribosome-phased endonucleolytic cuts could occur widely across translated mRNAs through ribothrypsis (Ibrahim et al., 2018). Furthermore, mRNA decay factors such as SKIV2L and XRN1 can directly bind ribosomes to regulate translated mRNA stability in mammalian cells (Tuck et al., 2020). The co-translational degradation of monosome bound mRNAs could thus participate in the elimination of aberrant transcripts as well as in the control of protein synthesis for highly regulated transcripts.

In addition to the previously depicted monosomes populations, a substantial fraction of the monosomes are inactive 80S complexes that are not bound to mRNA. In fact, the number of ribosomes available for *de novo* translation is not limiting in basal conditions (Arava et al., 2003). On the contrary, the stock of ribosomal particles exceeds the cell needs for protein synthesis (Metzl-Raz et al., 2017). While non engaged in translation, the two main ribosomal subunits, 40S and 60S, are associated with initiation factors, eIF3 and eIF6 respectively, that prevent their reassociation in absence of mRNA (Brina et al., 2011). The accumulation of inactive 80S complexes is thus a controlled phenomenon. Notably, empty monosomes are a transitory state to assess and maintain ribosomal subunits integrity depending on the conditions.

First, during the maturation of pre-40S subunits in the cytoplasm, the formation of 80S-like complexes is part of a quality control check during which the ribosomal subunits undergo a translation-like cycle (Strunk et al., 2012). During this maturation step, the association to tRNA and mRNA is prevented by repressor proteins. Once the proper binding of the 60S to the pre-40S subunit is validated, the complexes are disassembled through the recruitment of recycling factors and the matured subunits enter the translating pool.

Secondly, during stress conditions that induce a global translation inhibition, free ribosomal subunits reassociate to form a large pool of non-translating 80S ribosomes stabilized by the clamping factor Stm1 in yeast (den Elzen et al., 2014). The formation of such complexes may protect the ribosomal subunits from degradation and promote the resumption of protein synthesis upon stress relief (Brina et al., 2011). For example, the accumulation of inactive monosomes in yeast during nutrient deprivation has been well characterized. In such case, eEF2 stably binds to ribosomes, acting like a stalling factor that inhibits translation elongation (Leprivier et al., 2013). The monosome fraction isolated by sucrose sedimentation of nutrient deprived yeast lysate thus contains a large quantity of inactive ribosomes that do not engage on mRNA (Liu and Qian, 2016). After stress relief, when the conditions are more favorable, the inactive complexes are dissociated by the classical recycling factors (Dom34-Hbs1 or Pelota/HBS1L) and the subunits can be recruited to resume protein synthesis without requirement for ribosome biogenesis (den Elzen et al., 2014). Similarly, the excess of ribosomal particles observed even in basal conditions could be a stock that cells preserve to be able to increase protein synthesis rates quickly after a change in the environment (Metzl-Raz et al., 2017). Indeed, ribosome biogenesis has a high energy cost and translation of ribosomal proteins competes with the production of other proteins (Chu and von der Haar, 2012). As the cellular resources limit translation rates, it is a viable strategy to store ribosomal particles to allow a better adaptation to less favorable conditions. To conclude, monosomes encompass an heterogeneous population of ribosomes that were previously overlooked but that could actually participate in the adaptation of protein synthesis levels in a fluctuating environment.

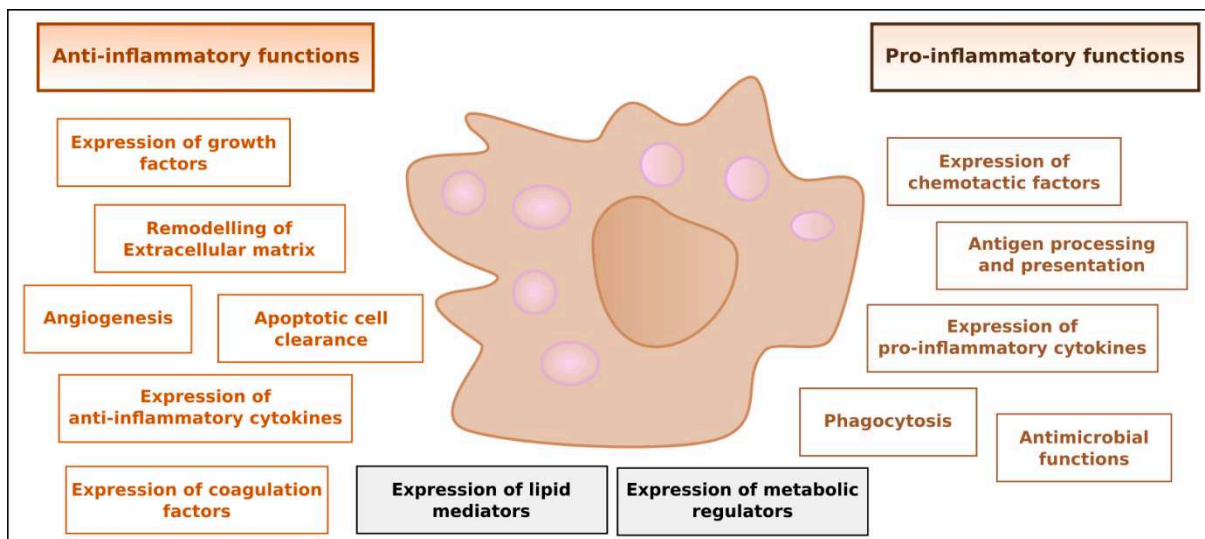
Translation dynamics were mostly studied at steady-state which only provides a limited view on how the process is regulated depending on the cellular environment. Particularly, how the proteome can be reshaped to give cells more adaptability in changing conditions is not well characterized. To understand protein synthesis kinetics, it is critical to take in account that the different translation steps can be achieved at variable speed depending on transcript features or binding of regulatory factors (Sokabe and Fraser, 2019). It is also important to integrate information on the availability of each component required for translation. Notably, the amount of ribosomal subunits available for new rounds of translation can become limiting in conditions where protein synthesis levels must be increased rapidly (Marshall et al., 2014; Sokabe and Fraser, 2019, Dykeman, 2020). Under such conditions, the distribution of ribosomes across the pool of cellular mRNAs could be a decisive parameter to shape protein synthesis depending on the cell behavior.

# Chapter III. Gene expression regulation during the inflammatory response in macrophages

Dynamic protein synthesis regulation facilitates the cell adaptation to environmental changes. In this work, we selected murine macrophages as a model to study the mechanisms involved in this process as these cells display a high phenotypic and functional plasticity. In this last introductory chapter, I will describe the importance of gene expression control in macrophage's biology.

## III.1 Macrophages plasticity is critical for the inflammatory response efficiency

In the last decade, great efforts have been made to study the impact of post-transcriptional control in dynamic systems such as immune cells (Carpenter et al., 2014; Jovanovic et al., 2015). These cells are particularly interesting to decipher the regulation of gene expression patterns in a changing environment as they undergo a complete switch of protein expression after challenging with an activating signal. The rapid remodeling of their proteome is essential to ensure the efficiency of their protective functions in the organism.



**Figure 9 : The diverse functions of macrophages.** Macrophages have a high functional plasticity as they can display pro-inflammatory, or degradative, and anti-inflammatory, or reparative, functions depending on their environment. The expression of lipid mediators and metabolic regulators participate in both pro and anti-inflammatory functions of the macrophages depending on the factors expressed.



Macrophages are very particular immune cells that display a wide range of phenotypes and functions depending on their environment (Gordon and Mantovani, 2011; Gordon and Plüddemann, 2019; Wang et al., 2013a). Indeed, the term “macrophages” encompasses an heterogeneous population with distinct origins, pathways of differentiation and behavior upon activation (Sica and Mantovani, 2012). Their differences can be introduced by distinct developmental origins with tissue-resident macrophages generated from embryonic progenitors or monocyte-derived macrophages produced from circulating cells in reaction to inflammation (Gordon and Plüddemann, 2019; Molawi and Sieweke, 2013; Murray and Wynn, 2011). They can also acquire various phenotypic features, from pro-inflammatory to protective functions, depending on the signals detected in their environment (Figure 9). Their heterogeneity accounts for their participation in numerous physiological processes in the organism : clearance of dying cells during the development and throughout adult life, tissue repair following injury, immune surveillance, antimicrobial defense, antigen presentation to adaptive immune cells, metabolism regulation (Gordon and Plüddemann, 2019; Watanabe et al., 2019). Interestingly, macrophages can express a wide range of receptors that allow them to scan their environment and detect any alterations of tissue homeostasis or infection (Gordon and Plüddemann, 2017). The great sensitivity of macrophages towards external cues contributes to their high functional plasticity and adaptability. Indeed, many studies have documented their ability to switch from one functional phenotype to another in response to new microenvironmental signals (Galli et al., 2011; Murray and Wynn, 2011).

Amongst the most important roles of macrophages in the organism is their participation in both the triggering and the resolution of the inflammatory response (Hamidzadeh et al., 2017; Medzhitov and Horng, 2009). Inflammation is a fundamental biological process essential to ensure the organism's integrity under basal and stress conditions, following an injury or microbial infection (Medzhitov, 2008, 2010). This physiological process can be divided in several stages : the onset is triggered by local immune cells, such as tissue-resident macrophages, that produce pro-inflammatory factors upon detection of pathogen or danger-associated molecules (PAMPs or DAMPs). These factors include a large range of molecules, such as proteinases, chemokines, cytokines, growth and differentiation factors, as well as metabolites derived from oxygen, nitrogen, arachidonate and other lipids (Gordon and Plüddemann, 2017). Their secretion in the extracellular medium provokes an increased permeability of the nearby vascular endothelium and the recruitment of other immune cells, such as neutrophils and monocytes derived macrophages, to the site of injury. This phenomenon is at the origin of the known symptoms of inflammation : redness, swelling, heat and pain (Molawi and Sieweke, 2013). The immune cells newly recruited participate in the

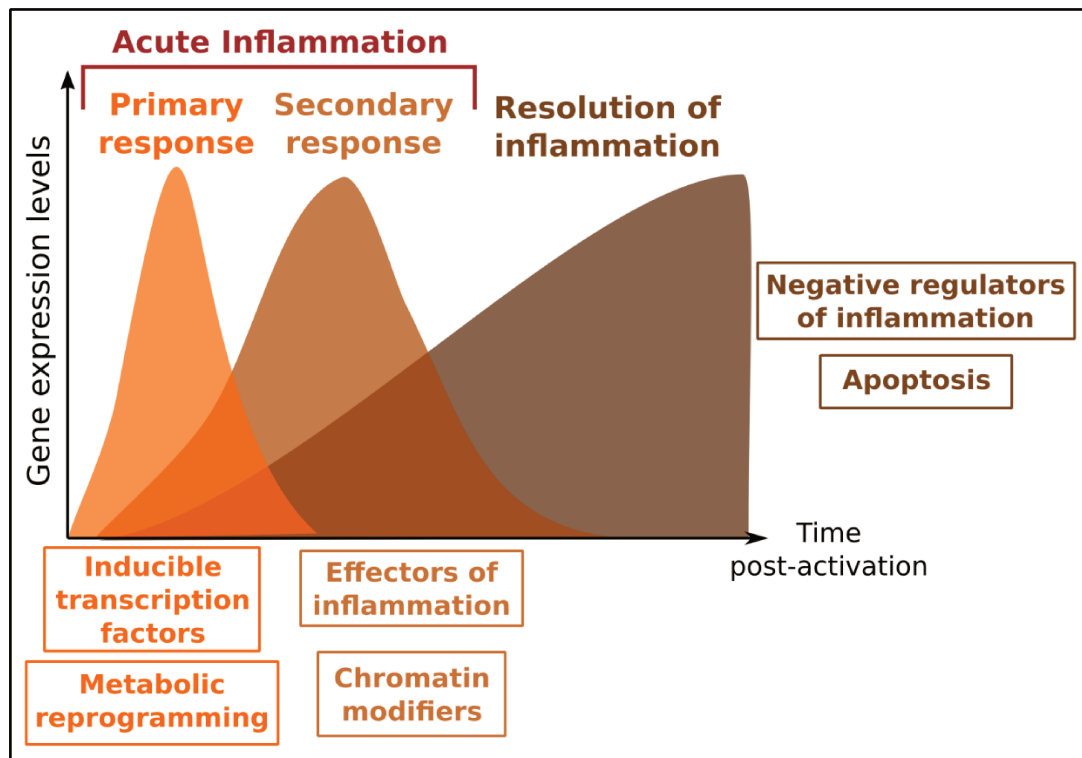
elimination of the threat through phagocytosis and antimicrobial activity. The last stage consists in the restoration of tissue homeostasis after the danger elimination. Consequently, physiological, acute inflammatory response is normally followed by a recovery phase during which macrophages actively participate in the healing process (Hamidzadeh et al., 2017; Oishi and Manabe, 2018; Watanabe et al., 2019). If this recovery phase is not properly completed, either because of a failure to remove the threat or inappropriately sustained inflammation, tissue damages can be increased on a long term scale ultimately leading to chronic inflammatory disorders or autoimmune pathologies (Feehan and Gilroy, 2019; Oishi and Manabe, 2018; Sica and Mantovani, 2012; Takeuchi and Akira, 2010).

Upon triggering of inflammation, macrophages undergo through a complete metabolic rewiring in order to acquire increased migratory, phagocytic and digestive capacities (Bossche et al., 2017; Kelly and O'Neill, 2015). Notably, they express higher levels of proteases, RNases, nitric oxide and reactive oxygen species (NOS and ROS) upon activation (Liu et al., 2016). Their lipid metabolism is also highly increased in order to synthesize lipid mediators that promote or reduce inflammation and to ensure the membrane expansion of all subcellular compartments (Everts et al., 2014; Martinez et al., 2006). This metabolic reprogramming additionally confers them an increased resistance to all bioactive antimicrobial molecules that could damage their basic cell components such as NOS and ROS (Virág et al., 2019). After inflammation resolution, the majority of neutrophils and monocyte-derived macrophages are cleared by programmed apoptosis.

### **III.2 Regulation of inflammation related genes in macrophages**

To achieve a balanced inflammatory response, macrophage immune gene expression must be tightly regulated to limit the production of pro-inflammatory molecules while promoting the expression of recovery functions (Hamidzadeh et al., 2017; Medzhitov and Horng, 2009). Gene expression regulation in macrophages involves various mechanisms that act at both transcriptional and post-transcriptional levels (Anderson, 2008; Carpenter et al., 2014; Mino and Takeuchi, 2018; Molawi and Sieweke, 2013). The combination of these different mechanisms is required to obtain a well-orchestrated response with time dependent expression of specific proteins according to the different stages of inflammation (Figure 10). Additionally, as the expression of pro-inflammatory molecules must be adapted in a context-specific manner, the different layers of regulation are also important to fine tune the protein synthesis levels depending on the cellular needs.

Gene expression regulation in immune cells has been studied at the transcriptional level in very great details (Smale and Natoli, 2014). Obviously, many changes occur through transcriptional modifications during the inflammatory response in macrophages (Medzhitov and Horng, 2009). Several functional modules or clusters of genes, specifying the response magnitude and intensity, are activated by different transcription factors in a constrained temporal pattern (Figure 10). Following the detection of a triggering factor, the first wave of newly transcribed gene expression begins very rapidly with the recruitment of constitutively expressed transcription factors activated by post-translational modifications (Smale and Natoli, 2014). Among these early response genes are found other transcription factors that are produced in a stimulus dependent manner and control the expression of a secondary wave of pro-inflammatory genes. Moreover, additional lineage specifying transcription factors are synthesized during the second wave of gene expression (Glass and Natoli, 2015; Medzhitov and Horng, 2009). These factors notably target the expression of chromatin remodeling factors that ultimately cause durable epigenetic modifications in the activated macrophages (Lauterbach et al., 2019; Molawi and Sieweke, 2013; Saeed et al., 2014). As their expression depends on the activity of signal-specific transcription factors, the subsequent remodeling of gene expression thus depends on the triggering stimulus nature and intensity (Smale et al., 2014). This mechanism is particularly important as it can induce an hypersensitive, or on the contrary an hyporesponsive, phenotype following the detection of a new threat by previously stimulated macrophages (Feehan and Gilroy, 2019; Hamidzadeh et al., 2017; Molawi and Sieweke, 2013). Importantly, the expression of negative regulators of the inflammatory response and factors involved in the resolution phase is triggered very early after inflammation onset to modulate the levels of activation and promote a rapid recovery after the threat clearance (Hamidzadeh et al., 2017; Serhan and Savill, 2005).



**Figure 10: Different waves of gene expression shape the inflammatory response in macrophages.** The first wave of gene expression is triggered very rapidly after the inflammation onset allowing the expression of primary response genes. These primary response genes correspond to inducible transcription factors that will ensure the expression of other inflammation-related genes and the metabolic reprogramming of the macrophages. During the second wave of gene expression, chromatin modifiers are expressed along effectors of inflammation to ensure a long-term remodeling of macrophage gene expression. The expression of negative regulators of inflammation and inducers of apoptosis is triggered early after the inflammation onset and increases following the removal of the threat.

Post-transcriptional regulation in immune cells, and particularly macrophages, has gained interest over the last decades (Carpenter et al., 2014). Indeed, transcriptional control alone is insufficient to explain the great macrophages plasticity and their ability to transiently modify their protein synthesis levels very rapidly after stimulation to trigger a properly orchestrated inflammatory response (Anderson, 2008). Notably, an extensive knowledge of specific mRNA features recognized by trans-acting regulatory factors was accumulated. The most well characterized trans-acting factors are RNA binding proteins that can act at different stages of the mRNA life (Anderson, 2008; Mino and Takeuchi, 2018; Turner and Díaz-Muñoz, 2018). A common example is the binding of RBPs to A-U-Rich Elements (ARE) in the 3'UTR of mRNAs to induce to modulate their localization, stability and translation (Otsuka et al., 2019; Turner and Díaz-Muñoz, 2018). Additionally, many sequences targeted by miRNAs were

identified in the 3'UTR of pro-inflammatory transcripts (Hamidzadeh et al., 2017). The regulation through the binding of non-coding RNAs, such as long non coding RNAs (lncRNAs), have also drawn a lot of attention since several years (Carpenter et al., 2014; Jackson et al., 2018). Importantly, the regulation of mRNA stability, particularly during the late stages of inflammation, is critical to ensure that the expression of pro-inflammatory factors remains transient (Carpenter et al., 2014; Zhang et al., 2017).

Despite the growing knowledge about post-transcriptional control in macrophages, still little is known about translational control in these cells. Large scale RBPs screening in macrophages revealed that many of them interact with ribosomal proteins, opening a possibility for translational control (Turner and Díaz-Muñoz, 2018). Moreover, RBPs can target translation initiation factors to inhibit global protein synthesis in response to infection or to adapt immune cell metabolism depending on the type of response induced (Carpenter et al., 2014; Piccirillo et al., 2014). A recent ribosome profiling study in murine bone-marrow derived macrophages confirmed the widespread regulation of pro-inflammatory mRNAs at the translational level through the recruitment of specific RBPs (Zhang et al., 2017). Notably, they described a mechanism by which the ARE-binding protein Zfp36 directly binds to the PolyA Binding Protein (PABP) to inhibit translation and subsequently recruits mRNA decay factors. Another ribosome profiling study in activated macrophages revealed that translational regulation selectively affects pathways important for cytokine expression, protein synthesis and cell metabolism (Su et al., 2015). Furthermore, short-lived negative regulators of inflammation are particularly sensitive to translation blockade (Lemaitre and Girardin, 2013). Their suppression following translation inhibition could be a signal recognized by the innate immune system to respond to particular pathogens (Barry et al., 2017). Altogether, these studies highlight the existence of a coordinated network of regulation acting at different stages of the mRNA life, including translation, to modulate protein synthesis during the inflammatory response.

To conclude, macrophages are a very fascinating model to study protein synthesis adaptation depending on fluctuating conditions due to their high sensitivity towards their environment and their ability to trigger different gene expression programs accordingly (Pope and Medzhitov, 2018). The purpose of my work was to study how modifications of the translation process could participate in the shaping of the inflammatory response in macrophages. Particularly, I was interested in how variations of the ribosomal binding pattern could affect the expression of inflammation related genes. Indeed, the competition for ribosomes, tRNAs and translation factors could be even more important upon triggering of the inflammatory response as transcription rates are rapidly increased. Consequently, the cellular resources must be used efficiently to produce only the proteins that are biologically relevant depending on the nature of the danger. Particularly, the over-expression of pro-inflammatory factors can have deleterious side effects on the organism and several layers of regulation are combined to limit their production. Importantly, the distribution of ribosomes across the different mRNAs is not random and the proper ribosome allocation pattern could be critical to regulate protein synthesis levels during the inflammatory response. Additionally, translational control could also participate in the adaptation of protein synthesis kinetics depending on the different inflammation phases (Koppenol-Raab et al., 2017). Interestingly, a lag was previously observed between transcriptional induction and protein synthesis for a subset of transcripts in macrophages (Eichelbaum and Krijgsveld, 2014). To characterize how the ribosomal binding pattern could be modified during inflammation, I thus sought to perform monosome vs polysome footprinting using activated macrophages at different stages post-stimulation.

# Results

Mouse Bone-Marrow Derived Macrophages (BMDMs) are broadly used to study gene expression regulation during innate immune responses and particularly inflammation (Medzhitov and Horng, 2009; Wang et al., 2013). They are well-characterized and easier to cultivate than human macrophages. We thus decided to use these cells as a model for our study. Given the fact that macrophages represent an inherently heterogeneous population, the production and use of these cells imply to follow a rigorous protocol to achieve reproducible results. For example, when cultivating these cells *in vitro*, any variations of the cell density can ultimately affect their functional capacities (Lee and Hu, 2013).

To obtain BMDMs, myeloid progenitors extracted from mouse bone-marrow are cultivated in the presence of a lineage-specific growth factor, Macrophage Colony-Stimulating Factor or M-CSF, until complete differentiation (Weischenfeldt and Porse, 2008). After seven days of *in vitro* culture, the progenitors are fully differentiated into mature macrophages that can be used in various immunological studies. The exposition to diverse microbial components activates the macrophages and serves as a proxy to study the behavior of innate immune cells upon modification of their environment. The activation of BMDMs by lipopolysaccharide (LPS), a glycolipid found in the outer membrane of Gram negative bacteria, is the most commonly used protocol to study the inflammatory response (Medzhitov and Horng, 2009).

## ***Setting-up the harvesting of cytoplasmic lysates from macrophages***

Despite using a well-established model in the immunology field, the adaptation of the monosome vs polysome footprinting protocol from yeast to mouse macrophages was not straightforward. On the contrary, several technical aspects required optimization for the proper execution of this assay using this particular cell type. Amongst the first issues was the relatively limited proliferative capacities of BMDMs. As a consequence, obtaining enough material for some specific experiments was quite challenging. To circumvent this problem, early stage optimizations were performed using both primary and immortalized cells. BMDMs can be immortalized through the infection with an oncogenic virus (Gandino and Varesio, 1990). Immortalized BMDMs (iBMDMs) present the advantages to proliferate faster than primary bone-marrow derived macrophages (pBMDMs) and are also easier to cultivate. However, after immortalization, these cells display less physiological features and are functionally different from pBMDMs (Troupin et al., 2013). As a matter of fact, while iBMDMs can be useful to set up an experiment, pBMDMs should be preferred for assays aiming at understanding the normal cell biology.

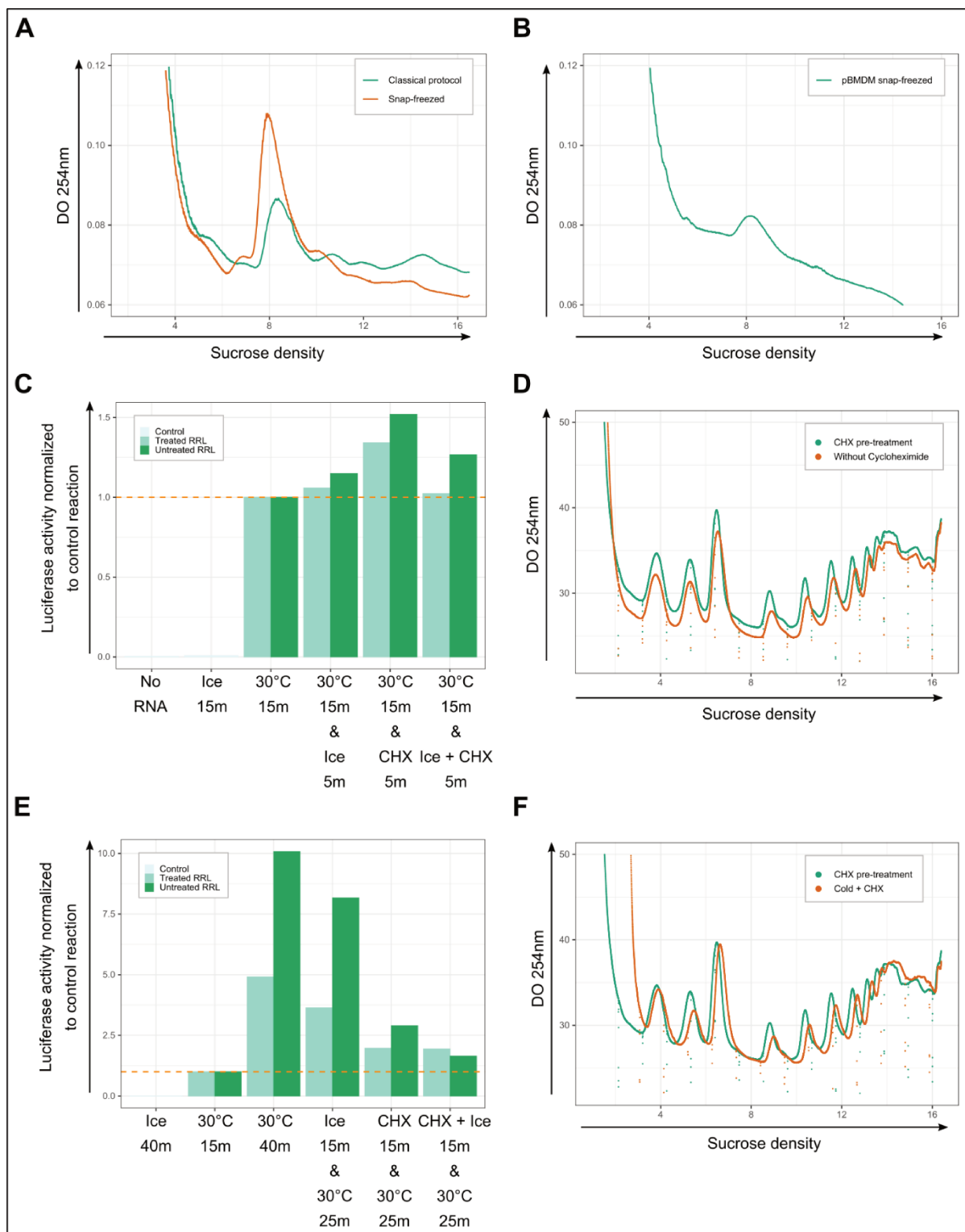
In addition to the great sensitivity of macrophages towards their environment, protein synthesis is a highly dynamic process that can be altered very rapidly following any modification of the cell medium. Consequently, any approach aiming at capturing the physiological changes of translation in macrophages must avoid the introduction of technical distortions that could skew normal translation dynamics. This is particularly important for ribosome profiling based techniques as freezing the ribosomes at their exact positions on the mRNAs is essential to obtain an accurate picture of ribosomal densities. The proper arrest of ribosomes on their associated mRNAs is even more critical for monosome vs polysome footprinting as the aim is to compare the binding pattern from different ribosome populations.

In the beginning of my thesis, I sought to find the best approach to lyse the macrophages without disturbing the ribosomal binding pattern across mRNAs. In fact, if translation is not properly blocked during sample collection, the ribosomes continue to translocate until they fall off of the mRNA leading to an artifactual increase of light polysomes and monosomes. This phenomenon, called ribosomal run-off, can be efficiently inhibited by the addition of translation inhibitors such as cycloheximide (CHX). In the initial ribosome profiling protocol (Ingolia et al., 2009), cells were pre-treated with CHX at 37°C for 10 minutes before lysis. However, several subsequent studies revealed that this treatment could disturb ribosomal densities at different levels in yeast. Notably, as CHX inhibits translation elongation but not initiation, new ribosomes can be recruited on the mRNAs and halted only in the



beginning of the CDS causing an accumulation of RPFs near the start codon (Santos et al., 2019; Weinberg et al., 2016). Moreover, CHX mediated inhibition is not immediate as the molecule has to diffuse into the cells and its effect depends on the ribosome conformation provoking codon-specific alterations of the ribosomal binding pattern (Gerashchenko and Gladyshev, 2014; Hussmann et al., 2015; Weinberg et al., 2016). A recent study revealed that the impact of CHX induced biases could be less significant in mammalian cells compared to yeast (Sharma et al., 2019b). Despite this, the pre-treatment with CHX at 37°C is not recommended for studies aiming at studying translation at the codon resolution (MGlincy and Ingolia, 2017; Weinberg et al., 2016).

Several studies performed using yeast or bacteria revealed that the incubation with translation inhibitors before cell lysis was not necessary, providing that the cells are lysed rapidly in very cold conditions. For this, the cells are snap-frozen using liquid nitrogen and then grinded in presence of lysis buffer while still frozen (MGlincy and Ingolia, 2017; Mohammad and Buskirk, 2019). This strategy is easy to set up for non-adherent cells that can be rapidly harvested by simply collecting the culture medium. The collection timing is however increased for adherent cells such as macrophages that must be scraped off the culture plate before resuspension in the lysis buffer. Consequently, the increased collection timing could reduce the efficiency of the translational arrest mediated by snap-freezing. To assess if this strategy could be used for monosome vs polysome fractionation, iBMDMs and pBMDMs containing plates were directly placed in a liquid nitrogen bath after a quick wash with ice cold PBS. The cells were then scraped in 1 mL of ice cold lysis buffer. After homogenization by several pipetting and clarification, the resulting cytoplasmic lysates were loaded on sucrose gradients to check the integrity of the polysomes. The results revealed that the polysome fractions were dramatically reduced following snap-freezing (Figure 11.A and B). This observation could be explained by a run-off of the translating ribosomes or a mechanical break-down of the polysomes into monosomes. Hence, the snap-freezing approach could not be used to perform monosome vs polysome fractionation from macrophages.



**Figure 11 : Ice-cooling and cycloheximide addition directly in the lysis buffer is the best strategy to maintain ribosomal binding pattern during sample preparation. A.** Sucrose gradients profiles of iBMDMs cytoplasmic lysates prepared with cycloheximide (CHX) pre-treatment or using snap-freezing in liquid nitrogen. **B.** Sucrose gradient profile of pBMDMs cytoplasmic lysates prepared using snap-freezing. **C.** Quantification of luciferase activity after *in vitro* translation of the renilla luciferase mRNA with treated or untreated Rabbit Reticulocyte Lysate (RRL) using ice or CHX, or their combination to block the reaction. **D.** Sucrose gradients profiles of pBMDMs cytoplasmic lysates prepared with

cycloheximide (CHX) pre-treatment or using ice to block translation. **E.** Quantification of luciferase activity after resuming of *in vitro* translation of the renilla luciferase mRNA with treated or untreated Rabbit Reticulocyte Lysate (RRL) following the use of ice or CHX, or their combination to block the reaction. **F.** Sucrose gradient profiles of pBMDMs cytoplasmic lysates prepared with cycloheximide (CHX) pre-treatment or using ice and CHX to block translation.

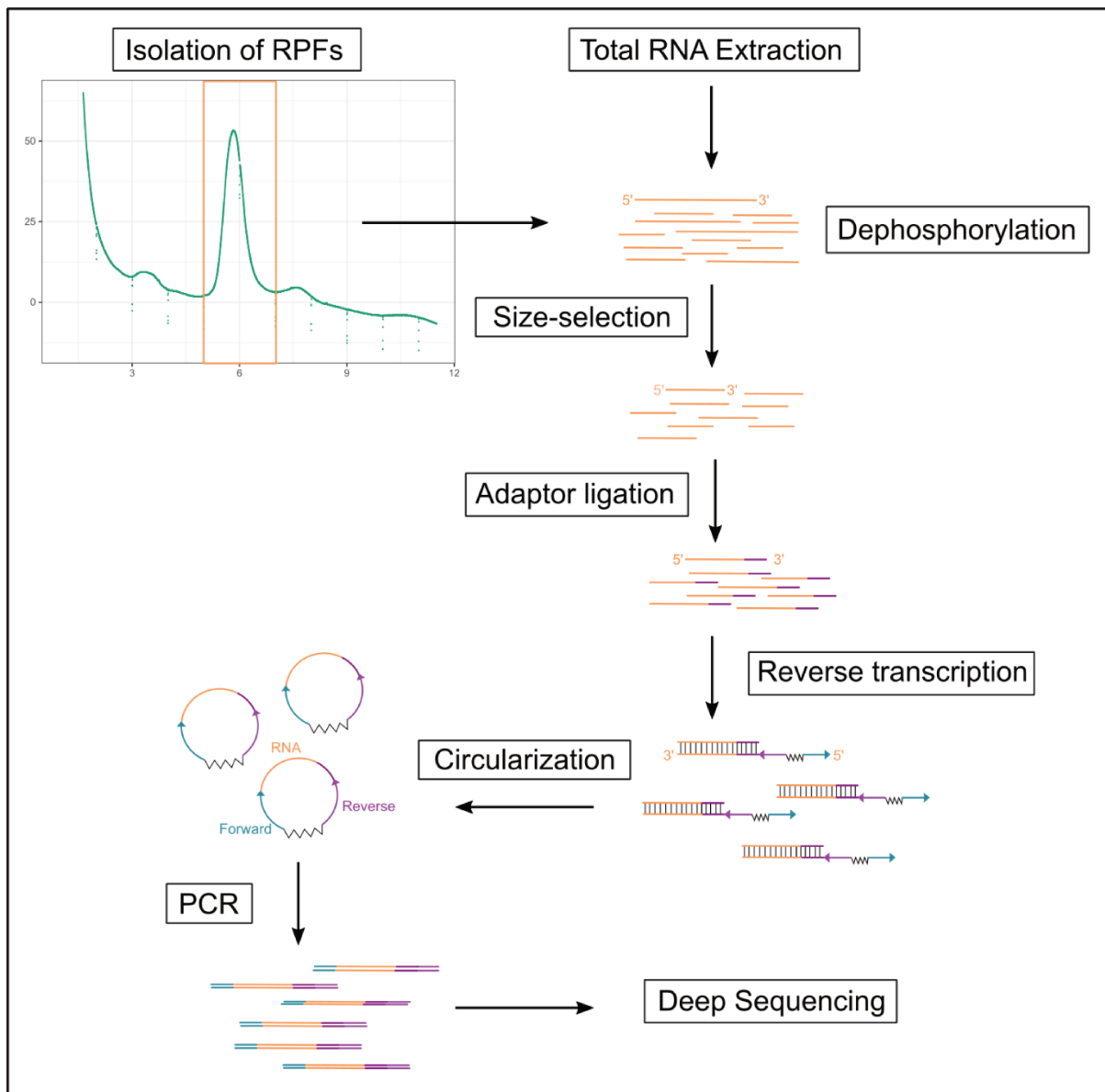
To rapidly decrease the temperature while preserving the polysomes integrity, another approach consists in placing the culture plate on ice during the wash with ice cold PBS. The subsequent reduction of temperature is however slower than snap-freezing. To assess if ice-cooling the plate could efficiently block translation, I performed an *in vitro* translation assay using a mRNA encoding for the renilla luciferase (Figure 11.C). The luciferase activity measured after stopping the reaction by the addition of lysis buffer reflects the amounts of renilla protein produced and is thus a direct proxy of translation initiation and elongation efficiency. This assay was performed using Rabbit Reticulocyte Lysate (RRL) either still containing or depleted of its endogenous mRNAs (untreated or treated). This difference of composition influences the efficiency of translation for exogenous mRNA with untreated RRL being closer to physiological *in cellulo* conditions (Rifo et al., 2007). When the translation mixture is kept on ice for 15 min, the luciferase activity measured is comparable to what is observed in the negative control that does not contain the luciferase mRNA. Thus, ice-cooling is a good strategy to inhibit translation. To capture the most accurate picture of ribosomal densities in translating cells, lysis must be performed quickly after taking the cells out of the incubator at 37°C. To reproduce this drop of temperature using the *in vitro* translation system, the reaction mixtures were incubated for 15 min at 30°C before stopping translation using ice cooling with or without CHX. As expected, with CHX the levels of renilla luciferase synthesized continue to increase compared to the condition where the reaction is stopped after 15 min at 30°C (Figure 11.C). When the translation mixture is placed on ice, the synthesis of renilla luciferase is not immediately blocked as well but the amount produced is reduced. The translation blockade is thus more efficient using ice-cooling compared to the incubation with CHX at higher temperatures. Moreover, when both cycloheximide and ice-cooling are combined to block translation, the translation blockade is more rapid and effective for treated but not for untreated RRL. To assess if similar results could be obtained using cells, sucrose gradients prepared using iBMDMs pre-treated for 10 min at 37°C with CHX or collected with the ice-cooling protocol were compared (Figure 11.D). The sucrose gradient profiles did not reveal an increased run-off when only ice-cooling is used to block translation, confirming that the CHX pre-treatment is not mandatory and that good quality polysomes can be obtained in absence of translation inhibitors if cold conditions are strictly maintained during cell lysis.

To track back the position of ribosomes on the translated mRNAs using ribosome profiling derived methods, the cytoplasmic lysates must be subjected to a RNase digestion step. For most of the current protocols, this step is performed at room temperature rather than on ice. As a consequence, the rise of temperature during this digestion step could relieve the translation blockade if no translation inhibitors are added in the lysis buffer. To confirm this, I compared the efficiency of the translation blockade mediated by CHX, ice-cooling or both when the synthesis reaction is resumed due to a rise of temperature using a similar *in vitro* translation assay than previously (Figure 11.E). Unsurprisingly, translation rates were rapidly increased when the reaction was placed at 30°C after ice-cooling only. The pre-treatment with CHX for 15 min at room temperature efficiently prevented the resuming of the translation reaction at 30°C. When CHX is added on ice instead, the translation blockade is more efficient probably because the protein production is inhibited with less delay compared to a room temperature incubation. The sucrose gradient profiles obtained using iBMDMs pre-treated with CHX at 37°C or ice-cooled and lysed in presence of CHX confirmed this observation (Figure 11.F). Indeed, the heavy polysomes peaks are higher when the lysate is prepared using both ice-cooling and CHX. Conversely, in the sample pre-treated at a higher temperature, the peaks are higher in the light polysome fractions revealing some ribosomal run-off. To conclude, collecting samples on ice and adding CHX directly in the lysis buffer is the best strategy to maintain ribosomal binding pattern during the preparation of cytoplasmic lysates and RNase digestion.

### ***Depletion of highly abundant ribosomal RNA contaminants***

rRNA contamination is a recurrent problem in all RNA sequencing based methods including ribosome profiling (Chung et al., 2015; Zinshteyn et al., 2020). To sequence a pool of RNAs with confidence, the average number of times a given RNA sequence is read, or coverage, must be increased proportionally to the levels of the least abundant RNA species of interest in the library. If many sequencing reads correspond to rRNA sequences, the amount of information that can be collected regarding mRNA sequences is inevitably reduced. To gain more sensitivity, one solution is to increase the total number of reads sequenced or sequencing depth. As this can be quite expensive, other strategies were developed to instead decrease the amount of rRNA fragments in the sequencing library. Most of these approaches were designed for RNA-seq samples that display different characteristics compared to ribosome footprinting samples. Notably, RPFs are smaller than classical RNA-seq fragments, reducing the efficiency of most classical commercial kits available for rRNA depletion from sequencing libraries. Consequently, the depletion of rRNA fragments from ribosome profiling libraries is not straightforward and a standard strategy is still lacking in the community.

To obtain a snapshot of the position of translating ribosomes on their associated mRNAs, the RPFs are generated by RNase digestion. During this step, mRNA portions left unprotected are degraded but the rRNA composing the ribosomes can also be targeted. After purification of the digested ribosomes from a sucrose gradient, the mRNA fragments represent only a small fraction of the total RNA obtained. The vast majority of the material (80-95%) is in fact composed of rRNA with another small fraction corresponding to tRNAs stably associated to the ribosomes. To reduce the contamination levels, only RNA fragments corresponding to the expected RPFs size (around 30nt) are selected on a highly resolutive denaturing polyacrylamide gel. However, rRNA and tRNA fragments of a similar size can also be co-purified during size-selection. As a consequence, RPFs derived from mRNAs are not the only RNA species sequenced in ribosome profiling based approaches. To prepare the RNAs for deep sequencing, they must be ligated to adaptors that will be necessary to initialize the sequencing reaction and reverse transcribed into cDNA (Figure 12). For this, the purified RNA fragments are first dephosphorylated to promote the ligation of the 3' end adaptor. After 3' adaptor ligation, the RNAs are reverse transcribed using a barcoded reverse-transcription (RT) primer that anneals to this adaptor. The RT primer also contains another adaptor sequence that will be used to circularize the resulting cDNA using CirLigase I (Lucigen). A long flexible linker (18-atom hexaethylene glycol spacer) is placed between the two adaptor sequences in the RT primer to minimize structural constraints during circularization. The final sequencing library is produced by PCR amplification using primers targeting both adaptor sequences surrounding the cDNA insert. An additional barcode sequence is added in the reverse PCR primer allowing the mixing, or multiplexing, of different samples in the same sequencing reaction.



**Figure 12: Overview of the library preparation protocol using ribosomal footprints.** Following isolation of the monosomes fractions through sucrose gradient ultracentrifugation, the RNAs are extracted using acid phenol chloroform purification. The purified RNAs are dephosphorylated to have 3'OH ends and size selected to enrich the sample in true Ribosomes Produced Footprints (RPFs). After gel purification, RNAs are ligated to a pre-adenylated DNA adaptor and then reverse transcribed. The reverse transcription (RT) primer contains both the reverse and forward priming sequences for Illumina sequencing, as well as a barcode to uniquely identify the sample. The RT product is gel purified, removing unligated adaptors and unextended RT primers. The gel purified RT product is circularized, forming a template for PCR. After PCR amplification, the library is gel purified, quantified and pooled to other libraries to form a multi-sample mix used for deep sequencing.

As the RNA fragments generated during ribosome footprinting correspond to the material left after the selection of ribosome associated RNAs from a sucrose gradient and size-selection, the amount of contaminating rRNAs could be reduced compared to classical RNA-seq. With this in mind, no depletion strategy was used for the first ribosome profiling study (Ingolia et al., 2009). Despite the rRNA contamination, roughly 16% of reads were mapped to mRNA sequences allowing the estimation of translation rates and the characterization of RPFs position on mRNAs at the sub-codon level. As similar results were obtained by other groups using different organisms in the following years, we decided not to use any depletion method for the first ribosome profiling and monosome vs polysome footprinting experiments that we carried out using murine pBMDMs. While some rRNA contamination was expected, the results revealed that the amount of rRNA sequences in our libraries was actually massive. Unlike what was observed using other models, such as yeast or human HEK293T cells, the rRNA contamination levels were largely superior to 90% thus severely decreasing the amount of information that could be retrieved regarding mRNAs (Figure 13.A).

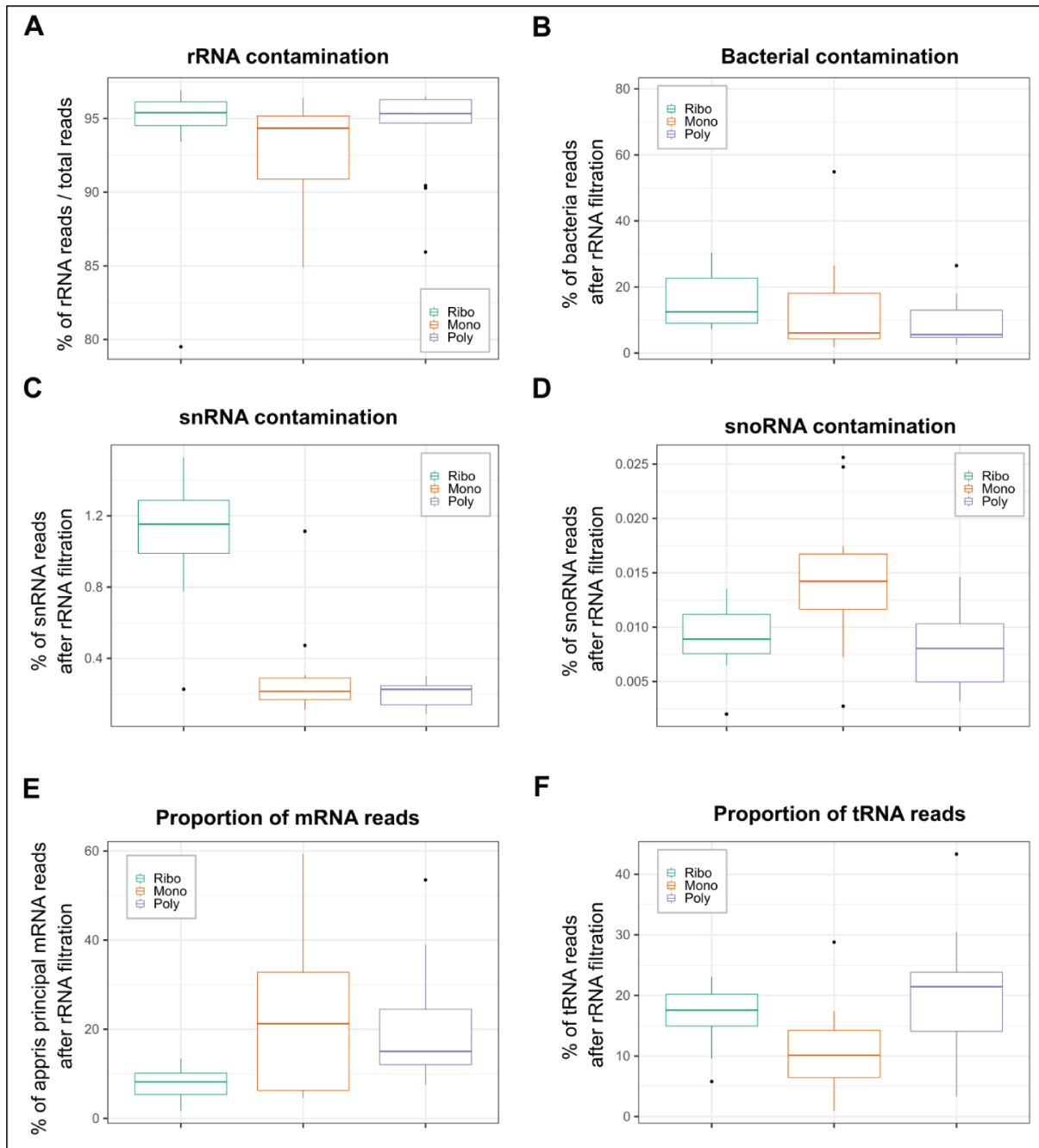
In addition to the heavy rRNA contamination, bacteria derived sequences were also found in our samples (Figure 13.B). As macrophages are particularly sensitive, many precautions were taken to avoid and also to monitor the apparition of any type of contamination during all cell culture steps. Hence, the contamination was more likely occurring during the sample collection or library preparation steps. This was confirmed by the identity of the bacterial sequences found in our libraries : a significant part of them corresponded to PhiX reads that were improperly assigned to our samples after demultiplexing. These *Enterobacteria* phage derived sequences are commonly added as spike-in in Illumina sequencing to improve base calling accuracy. The rest of the contamination was dominated by *Acinetobacter* or *Pseudomonas* derived sequences corresponding to bacterial species that frequently contaminate the laboratory environment (Park et al., 2019; Strong et al., 2014). To avoid this type of contamination, extra care must be taken for the cleaning of the material used during library preparation and only very high quality RNase-free water should be used.

To assess if the ribosome associated RNAs were properly purified using sucrose gradient separation, the levels of small nuclear or nucleolar RNAs (snRNAs or snoRNAs) were also quantified in our libraries (Figure 13.C and D). The proportion of reads mapping to these sequences was really low (less than 1.5% of the reads after filtering of rRNA sequences) demonstrating that ribosome associated RNAs were highly enriched in our libraries. Moreover, snRNA contamination was well reduced in monosome and polysome footprinting libraries compared to global ribosome profiling libraries (Figure 13.C). This could be explained by the

double sucrose gradient purification steps that limit the purification of RNA species not directly bound to mRNA-ribosomes complexes. This observation was not reproduced when looking at snoRNA levels in monosome vs polysome libraries compared to global ribosome profiling. The proportion of snoRNA reads found in the monosome libraries was actually higher than for polysome or global footprinting libraries (Figure 13.D). As snoRNAs participate in the maturation of ribosomal complexes, this could be explained by the presence of co-purified immature 60S particles in the monosome fraction. However, it was not possible to draw solid conclusions on the levels of small RNAs associated with the ribosomes as their coverage was insufficient.

As a consequence of the massive rRNA contamination, the proportion of reads mapping to mRNA sequences was very low in our libraries (less than 1% of total sequenced reads). After computer filtering of rRNA reads, 10-20% of the remaining sequences mapped to the principal mouse mRNA isoforms determined by the APPRIS database (Rodriguez et al., 2018). Importantly, despite the reduced coverage due to high rRNA contamination levels, many reads corresponding to mRNA sequences were found in the monosome libraries (Figure 13.E). This result confirmed that monosomes contain a high proportion of mRNA bound ribosomes as previously observed in other model organisms. Nevertheless, it was not possible to conduct more detailed analyses in order to confidently conclude about their translational status using these results as the coverage was very low. The quantification of tRNA levels was also informative despite a low coverage. Indeed, the proportion of tRNA reads was reduced in monosome libraries compared to polysome or global footprinting (Figure 13.F).

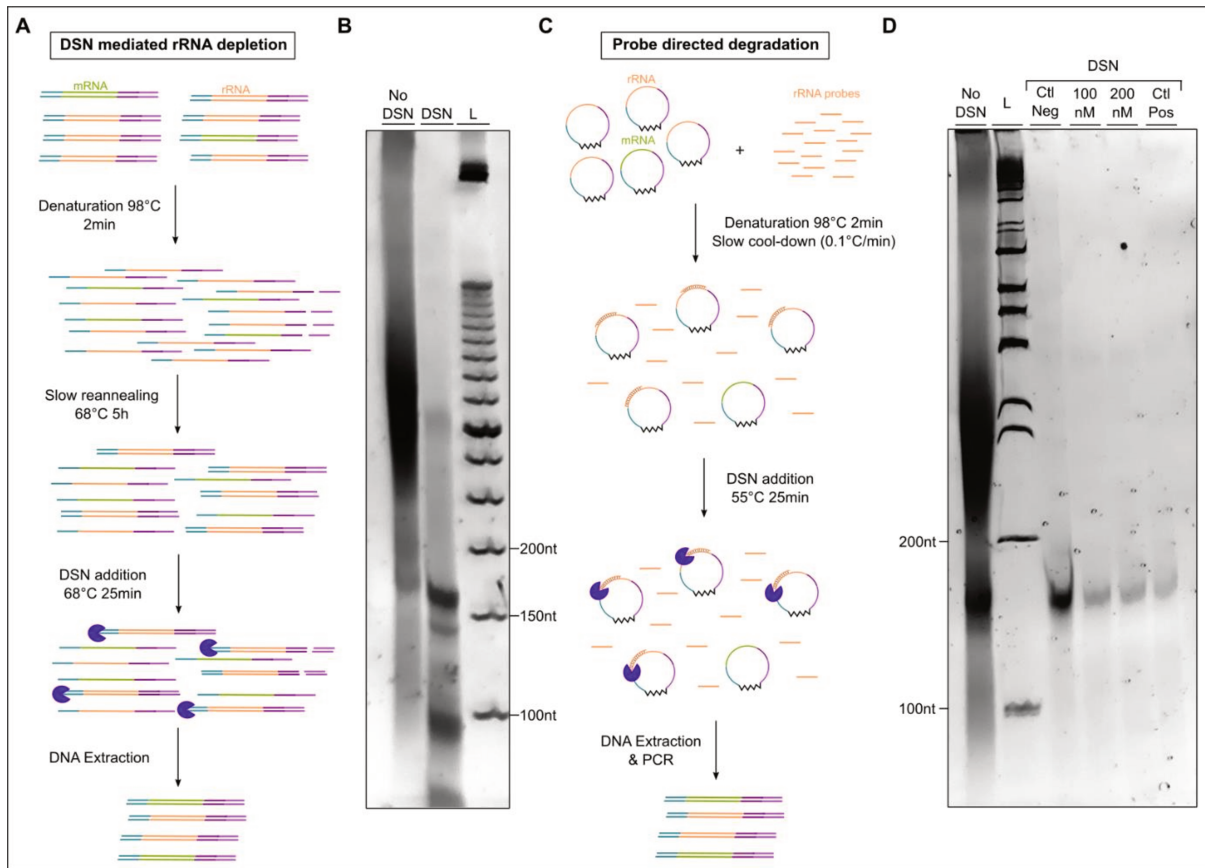




**Figure 13: Quantitative analysis of the different RNA species identified in our first global, monosome and polysome footprinting libraries.** **A.** Proportion of rRNA reads relative to the total reads sequenced for each type of library. Proportions of **B.** bacteria, **C.** small nuclear or snRNA, **D.** small nucleolar or snoRNA, **E.** messenger or mRNA, **F.** Transfer or tRNA reads relative the number of reads left after rRNA filtration for each type of library. The results obtained for 12 global (Ribo), 15 monosome (Mono) and 15 polysome (Poly) footprinting libraries are represented.

While the sequencing depth for the most interesting RNA species was low, the results obtained for this first attempt were quite interesting as we successfully retrieved mRNA sequences preferentially associated to monosomes or polysomes. For this reason, I sought to find a strategy to deplete rRNA sequences from the already constructed libraries. A strategy to do this is to use the Duplex-Specific Nuclease (DSN) that targets specifically DNA-DNA or DNA-RNA hybrids (Chung et al., 2015). This approach, widely tested for rRNA depletion from RNA-seq libraries, is based on the fact that DNA re-annealing rate is correlated to its concentration (Bogdanova et al., 2009; Christodoulou et al., 2011). Consequently, in a pool of DNAs, the most abundant molecules re-hybridize more rapidly than the others after heat-denaturation. Similarly, in PCR amplified sequencing libraries, the rRNA containing molecules should re-hybridize more frequently than the mRNA ones after denaturation. The addition of DSN after slow re-annealing can thus be used to deplete sequencing libraries from abundant contaminants (Figure 14.A). As this enzyme is resistant to relatively high temperatures, the reaction can be performed at 68°C to limit the re-hybridization of the least abundant species. After inactivation of the enzyme, the remaining DNA molecules can be amplified again by PCR to obtain enough material for sequencing.

To assess the efficiency of this approach, I compared the amount of material obtained for untreated or DSN depleted libraries generated from the same circularization product after migration on a non-denaturing 10% polyacrylamide gel (Figure 14.B). As the number of PCR cycles were optimized to obtain a good amplification for the DSN depleted sample, a large smear is observed in the non-depleted sample corresponding to overamplification products higher than the expected library size (170nt). This demonstrates that the amount of material was well decreased in the DSN treated sample. Additionally, the presence of lower bands in this sample confirms that DSN mediated digestion was effective. To assess the quality of the libraries before sending them to deep sequencing, they are usually cloned into a bacterial vector. After bacterial transformation, several clones are then sequenced using the classical Sanger method. This low throughput strategy can be used to predict the ratio between the different RNA species in sequencing libraries. I used this approach to check if the rRNA contamination was reduced in the DSN treated libraries and observed no differences (data not shown). Even if the results obtained after the migration of the untreated or depleted libraries were promising, the efficiency of this method was thus not sufficient to significantly modify the levels of rRNA contamination. Moreover, increasing the number of PCR rounds can introduce bias in the abundance of each mRNA molecule as the least abundant sequences tend to be less amplified than the most represented ones (Aird et al., 2011; Head et al., 2014).

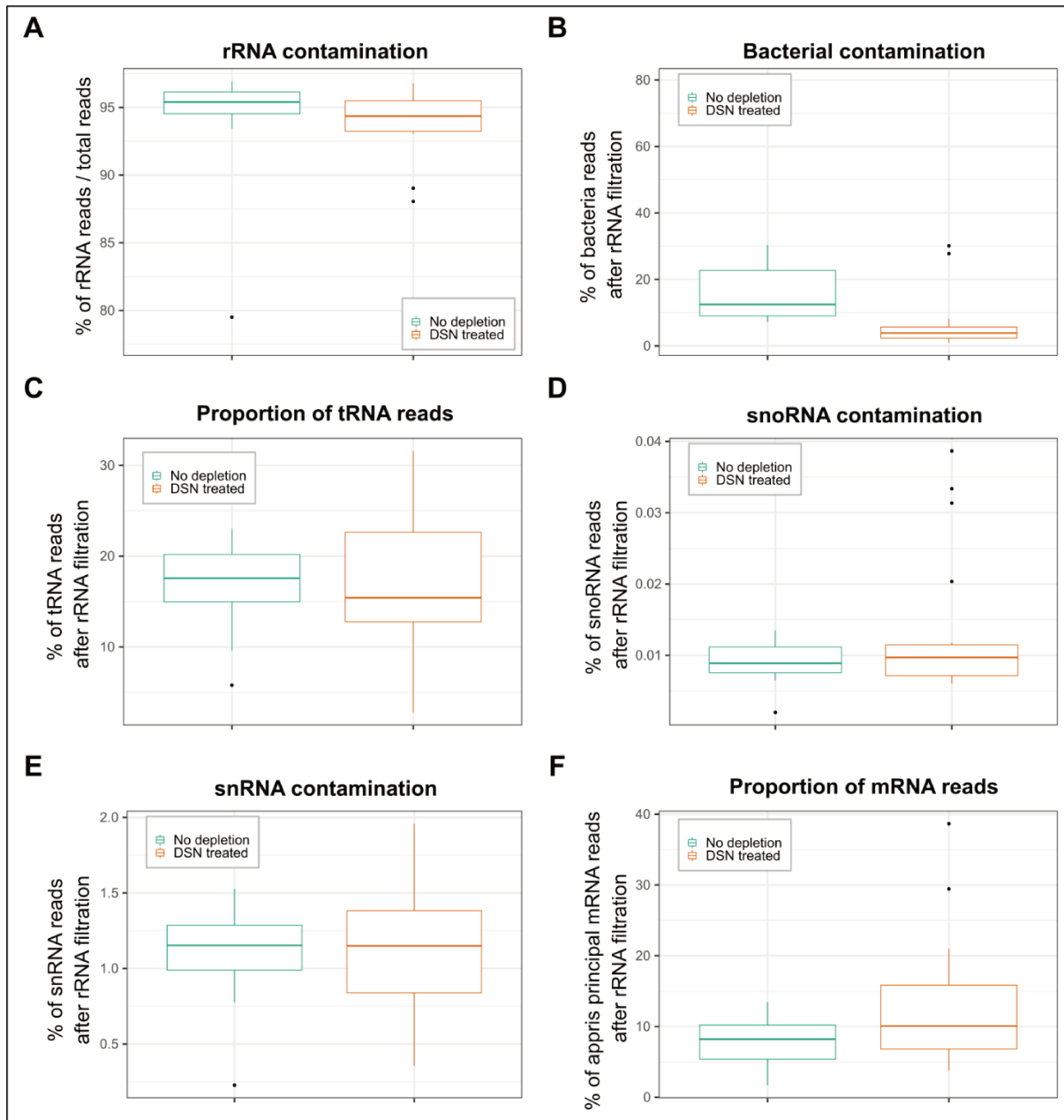


**Figure 14: Duplex Specific Nuclease (DSN) mediated depletion of highly abundant rRNA contaminants from ribosome footprinting samples. A.** Schematic representation of the DSN strategy to remove rRNA contaminants in already prepared libraries after PCR amplification. **B.** Gel electrophoresis of libraries treated or not with DSN after PCR amplification. **C.** Schematic representation of the DSN strategy to remove rRNA contaminants in already prepared libraries by targeting the circularization products. **D.** Gel electrophoresis of libraries prepared using circularization products treated or not with DSN in presence of different oligonucleotides. The negative control (Ctl Neg) sample was prepared using an oligonucleotide antisense to the renilla luciferase mRNA. The positive control (Ctl Pos) sample was prepared using an oligonucleotide antisense to the forward priming sequence for Illumina sequencing found in all circularization products. Two different concentrations (100mM and 200mM) were tested for our pool of oligonucleotides targeting frequent rRNA sequences.

Another strategy based on DSN mediated digestion consists in the addition of complementary oligonucleotides in the depletion reaction to improve the targeting of contaminant sequences (Archer et al., 2014). To implement this probe directed degradation (PDD) strategy, I took advantage of the high coverage of rRNA sequences in our previously sequenced libraries. As observed in other studies, most of the contaminating rRNA sequences were derived from a few specific sites on the ribosomal RNAs (Chung et al., 2015; Ingolia et al., 2009). I thus designed 32 probes targeting the most frequent contaminant sequences found in all of the 42 global, monosome or polysome footprinting libraries prepared previously. Moreover, the depletion was performed directly on the circularization products to avoid the introduction of PCR overamplification induced bias. Similarly to the DSN only protocol, the circularization products mixed with the antisense rRNA oligonucleotides were first heat-denatured before slow re-annealing (Figure 14.C). The depletion reaction was performed at 55°C to increase the probe hybridization efficiency. The results revealed that probe addition further reduced the amount of material left after digestion compared to the negative control treated only with DSN (Figure 14.D). The depletion efficiency was also increased by adding a higher quantity of antisense oligonucleotides. Furthermore, the addition of an oligonucleotide targeting an adaptor sequence found in all sequencing libraries further decreased the amount of material retrieved after digestion. This implies that the DSN directed digestion can specifically target the circularization products containing rRNA sequences annealed with the probes. The results obtained after cloning in a bacterial vector and Sanger sequencing were also promising as one clone out of 19 sequenced contained a mRNA derived RPF. All the libraries sequenced during the first attempt were thus depleted using this strategy and sent to deep sequencing.

Despite the promising results obtained during the optimization of the PDD protocol, the deep sequencing results revealed that the DSN depletion did not dramatically reduce the amount of rRNA contaminating the libraries (Figure 15.A). Indeed, the proportion of rRNA reads was decreased by less than 5% for most libraries. This could be explained by the fact that only a short list of the most abundant contaminants was selected. Interestingly, the level of bacterial contamination was well reduced despite the fact that no probes targeting these sequences were added in the digestion reaction (Figure 15.B). This suggests that the DSN could also target hybridized circularization products. Notably, the DSN requires as little as ten perfectly complementary base pairs to cut so even a partial hybridization could be sufficient to induce degradation (Archer et al., 2014). A similar effect was observed for tRNA levels that were slightly reduced as well (Figure 15.C). The depletion of circularization products non targeted by the probes could in fact depend on their abundance. Indeed, the contamination with less abundant species such as snoRNA and snRNA contamination was very similar with

or without DSN depletion (Figure 15.D and E). Finally, the number of reads mapping to the mouse transcriptome after rRNA filtering was slightly increased (Figure 15.F). Hence, the depletion of contaminants such as rRNA and bacterial sequences from the sequencing libraries really improves the coverage on mRNA sequences. To conclude, these results showed how the depletion of rRNA contaminants in ribosome footprinting libraries can be challenging. Moreover, the different strategies existing to decrease rRNA contamination from already constructed libraries are overall ineffective.



**Figure 15: Quantitative analysis of the different RNA species identified in DSN-treated or untreated ribosome profiling libraries. A.** Proportion of rRNA reads relative to the total reads sequenced for each type of library. Proportions of **B.** bacteria, **C.** snRNA, **D.** snoRNA, **E.** mRNA, **F.** tRNA reads relative to the number of reads left after rRNA filtration for each type of library. The results obtained for 20 DSN treated and 12 untreated ribosome profiling libraries are represented.

### ***The RNase treatment greatly influences the ribosomal footprints quality***

Considering the inefficiency of the strategies tested to improve mRNA coverage using the already constructed libraries, I next focused on methods that would decrease rRNA contamination from the RNA sample before library construction. Interestingly, most of the rRNA fragments that dominated our sequencing libraries were produced by the cleavage of surface exposed regions of the ribosomes targeted during the RNase digestion step. Similar observations were made previously in other studies and revealed that the level of contamination is highly affected by the experimental procedure used to generate the RPFs (Chung et al., 2015; Ingolia et al., 2009). Additionally, organism and cell type dependent variations of the contamination rate were highlighted even when the same protocol was followed for sample preparation (Miettinen and Björklund, 2015). As a matter of fact, ribosomes from different origins have very different resistance to RNase digestion (Gerashchenko and Gladyshev, 2017). While yeast ribosomes are very resilient to RNase digestion, mouse ribosomes can lose their structural integrity when the RNase treatment is too aggressive. This can be quite problematic as RPFs from unstable ribosomes can be lost during the sucrose gradient purification step provoking a significant skewing of gene expression estimates. Apart from this, the amount of rRNA fragments with a similar size to true RPFs can be highly increased.

To estimate the impact of this phenomenon using pBMDM lysates, I compared the amount of material retrieved on a polyacrylamide gel in the RPFs size range after treatment with different RNases (Figure 16.A). When using the same combination of RNase A and T1 as for the previously prepared libraries (3 $\mu$ L of RNase A and 300U RNase of T1), the RNA fragments obtained were highly degraded. Notably, a large smear was observed in the region spanning the RPFs size range and not so many bands were left at higher molecular sizes. This high degradation of rRNA could thus partly explain the massive rRNA contamination in our sequencing libraries. As RNase S7 and T1 were identified as the least aggressive towards mouse ribosomes, I next tested these enzymes to perform the ribosomal footprinting step (Gerashchenko and Gladyshev, 2017). It should be noted that RNase T1 has a strongly biased cutting pattern as it only cuts after guanosine while the RNase S7 can target every nucleotide (delCardayré and Raines, 1995). For this reason, I did not consider preparing the footprinting samples using only RNase T1. The pBMDM lysate treated with the RNase S7 alone contained more RNA fragments of a higher molecular size and less material in the RPFs size range. This observation thus confirmed that S7 nuclease degrades to a lesser extent the mouse ribosome and generates fewer small rRNA fragments that could be co-purified with the RPFs. When combined with RNase T1, more material was obtained in the RPFs range but higher size RNA

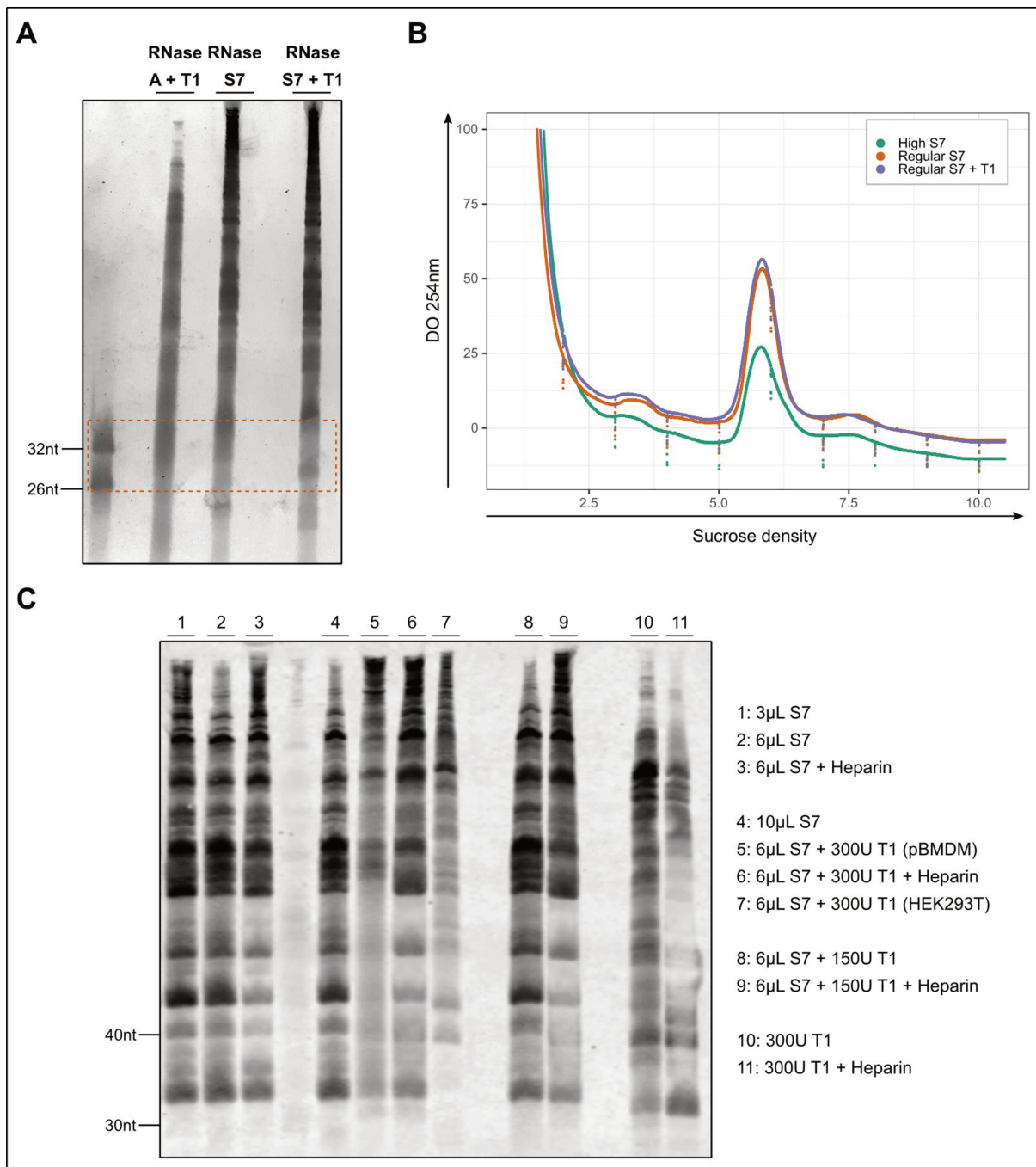
fragments were also preserved. This result is consistent with the observation that mouse ribosomes are particularly sensitive to RNase A treatment (Gerashchenko and Gladyshev, 2017). The important degradation of rRNA observed using our initial digestion conditions could thus be explained by the murine pBMDM ribosome sensitivity to RNase A digestion. Consequently, the use of other RNases could significantly reduce the amount of rRNA contamination in our footprinting libraries.

To control the ability of S7 nuclease to collapse all the polysomes into properly digested monosomes, pBMDM lysates were treated with this enzyme, alone or in combination with RNase T1, and then loaded on sucrose gradients (Figure 16.B). When the lysates were digested with 3 $\mu$ L of RNase S7 (regular amount), the polysomes were properly collapsed into monosomes. A small disome peak was left, corresponding to a fraction of collided ribosomes that cannot be separated by a mild RNase treatment (Han et al., 2020; Tuck et al., 2020). When combined with RNase T1, the sucrose gradient profile was very similar confirming that both RNases do not induce a loss of structural integrity on the digested ribosomes. On the contrary, when S7 nuclease amount was increased to 10 $\mu$ L (high S7), the 80S peak recovered was decreased showing that the RNase also partially degraded the ribosomal complexes at higher concentration. The amount of RNase used during the digestion step is therefore equally important to the RNase identity to retrieve RPFs of high quality.

To optimize the RNase digestion step for murine pBMDMs, several concentrations and combinations of both RNase T1 and S7 were tested on the same cytoplasmic lysate. The RNA fragments obtained were analyzed by Northern Blot using a probe targeting a highly frequent rRNA sequence to select the best condition to reduce rRNA contamination (Figure 16.C). As the smallest RNA fragments tend to migrate faster, they were not properly retained on the nitrocellulose membrane after transfer. For this reason, fragments lower than 30 nucleotides are not visible on our membrane. The results obtained were consistent with what was previously observed with the polyacrylamide gel analysis. RNase S7 treatment alone produced less small rRNA fragments compared to RNase T1 alone or their combination. Increasing RNase amount in the digestion reaction produced more small rRNA fragments. The production of small rRNA fragments was well correlated to the RNase digestion efficiency as the addition of a RNase inhibitor, heparin, induced a reduced fragmentation of high molecular size rRNAs. Additionally, I confirmed that cytoplasmic lysate of murine pBMDMs are more sensitive to RNase digestion than those of other classically used cell lines, such as the human HEK 293T cells. Indeed, the rRNA fragments obtained after digestion were higher for the human sample compared to any condition using pBMDM lysate even in presence of RNase inhibitor. Finally, the incubation with 3 $\mu$ L of S7 nuclease was selected as the best condition to



generate RPFs as the conversion of polysomes into monosomes was optimal. Meanwhile, the production of rRNA fragments that could be co-purified with the RPFs was also well reduced at this concentration.

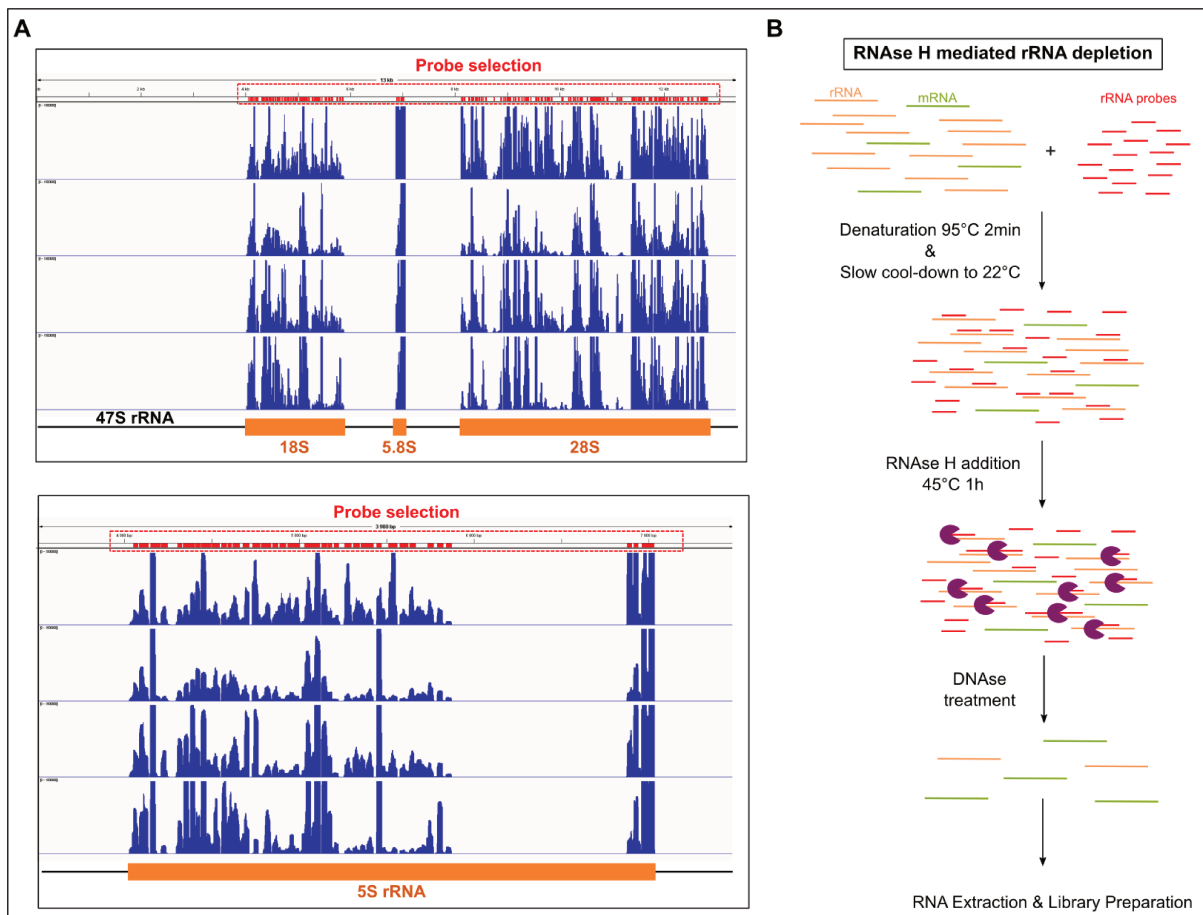


**Figure 16. Optimization of the RNase digestion step decreases the fragmentation of rRNA during footprinting sample preparation.** **A.** Gel electrophoresis of RPFs prepared using different combinations of RNase A, T1 and S7. The orange dotted rectangle highlights the range of RPFs typically selected for library preparation. **B.** Sucrose gradient profiles of pBMDMs lysates treated with RNase S7, T1 or their combination. Regular amount : 3 $\mu$ L of S7 RNase or 300U of RNase T1. High amount : 10 $\mu$ L of S7 RNase. **C.** Northern Blot analysis of samples prepared from one cytoplasmic lysate submitted to different RNase conditions as indicated. The specific detection of rRNA was ensured using a fluorescent probe targeting a highly abundant 18S rRNA fragment.

### ***RNase H mediated depletion of rRNA contaminants***

In addition to the optimization of the RNase digestion step, I also tested the depletion of rRNA fragments directly at the RNA level before library preparation. Most of the strategies to perform this consists in the use of specific oligonucleotides probes that bind to the targeted contaminant sequences. After hybridization, the unwanted fragments are separated from the RNA sample using magnetic beads or degraded by enzymatic digestion. All these methods thus require the characterization of the contaminant sequences before their implementation. Alternatively, commercial kits promising the depletion of a wide range of rRNA sequences from RNA-seq samples could also be efficient to some extent for ribosome footprinting samples. The most efficient commercial depletion kits, Ribo-Zero and Ribo-Gold distributed by Illumina, were recently discontinued leaving few effective commercial alternatives. Notably, I tested the RiboMinus technology distributed by Thermo Fisher and observed a limited effect on the rRNA contamination levels in ribosome footprinting samples (data not shown). The Ribo-Cop kit from Lexogen was also tested in the lab and did not induce a clear reduction of rRNA contamination.

As the commercial kits did not perform well, I next focused on setting-up a cost effective depletion strategy using oligonucleotides to target the contaminants. For this, all the data obtained for the previously sequenced libraries were combined to identify the regions that generate most of the rRNA fragments co-purified with the RPFs. The regions of the 47S pre rRNA and the 5S rRNA that produced most of the rRNA reads were visualized using the Integrative Genomics Viewer (IGV) visualization tool and 129 probes corresponding to the highest peaks were designed (Figure 17.A). In parallel, a new protocol relying on the degradation of targeted RNAs using RNase H was tested in our lab (Adiconis et al., 2013). This protocol was initially developed to target rRNA contaminants using a set of DNA probes that cover entirely all the human rRNA sequences. Interestingly, mouse and human rRNA sequences are highly similar so a large part of the 129 mouse probes designed previously were redundant with the human probes. In addition to the human rRNA probes, I ordered 63 mouse specific probes to increase the depletion efficiency from mouse samples using the RNase H protocol.

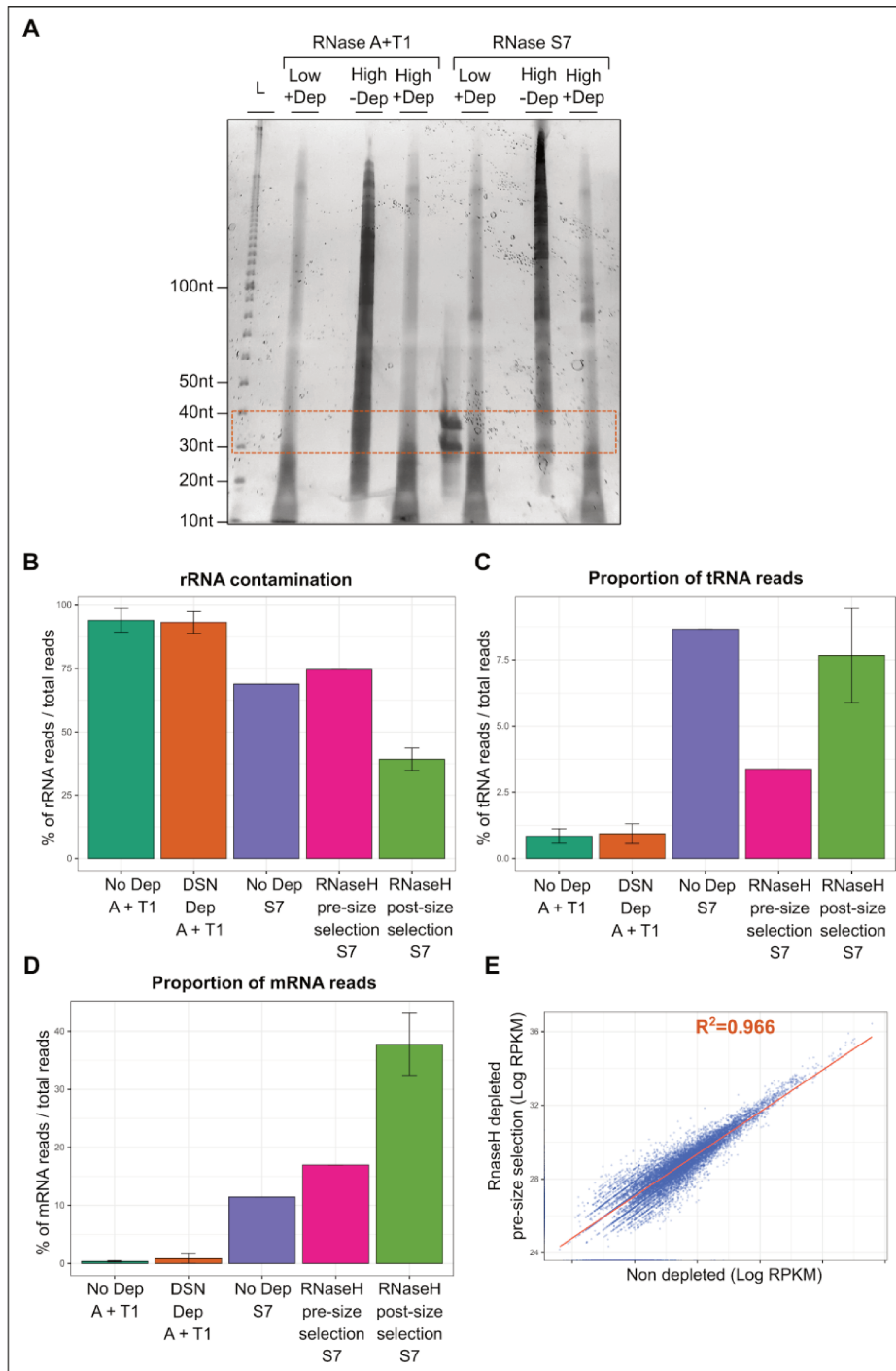


**Figure 17: The RNase H mediated depletion of rRNA contaminants.** **A.** The most frequently contaminating rRNA fragments found in all our non-depleted or DSN-depleted libraries were selected using the Integrative Genomics Viewer (IGV) visualization tool. In total, 129 probes (in red) corresponding to the highest peaks identified in the 47S pre-rRNA and 5S rRNA sequences were designed. **B.** Schematic representation of the RNase H depletion protocol. Purified RNA fragments are mixed with antisense DNA probes targeting the rRNA contaminants. After heat-denaturation and slow reannealing, RNase H is added to degrade all the DNA:RNA hybrids. After depletion, the antisense probes are removed by DNase digestion and the remaining RNAs are purified before proceeding to library preparation.

To perform RNase H mediated depletion of rRNA fragments, the RNA sample is mixed with antisense DNA probes and then heat-denatured (Figure 17.B). After slow re-annealing, RNase H is added to degrade the RNA fragments hybridized with the DNA probes. As this enzyme specifically targets DNA-RNA hybrids, the depletion is restricted to the sequences targeted by probes. After digestion, the DNA probes are degraded to keep only the remaining RNA fragments for library preparation.

To assess the efficiency of this strategy, the amounts of material retrieved before or after RNase H depletion on the same pBMDM digested lysate were compared on a denaturing polyacrylamide gel (Figure 18.A). After RNase H treatment of low or high amounts of input RNA, most of the high molecular size fragments are degraded and many small fragments lower than 30 nucleotides are obtained showing the high digestion efficiency. Importantly, the material obtained in the RPFs size range is well reduced in the depleted samples compared to the non-depleted. Additionally, the depletion is efficient for the two different RNase conditions tested indicating that RNase H treatment could perform well independently of the sample preparation protocol.

To check if the results obtained after deep sequencing were also improved, libraries were prepared using the two depleted or non-depleted samples treated with the S7 nuclease that were gel purified previously. Interestingly, the use of S7 nuclease instead of the combination of RNase A and T1 was already sufficient to decrease the rRNA contamination levels (Figure 18.B). Additionally, the coverage of tRNA and mRNA sequences were also improved (Figure 18.C and D). Indeed, the proportion of rRNA reads was reduced by 25% and mRNA sequences represented nearly 12% of total reads sequenced. This confirmed that RNase S7 mediated footprinting generates less contaminating rRNA fragments. The results were less convincing for the depleted sample as rRNA levels were higher than in the non-depleted sample (75% vs 69% respectively ; Figure 18.B). This could be explained by the degradation of high molecular size rRNAs mediated by RNase H that generates smaller contaminating fragments in the range of the RPFs. Additionally, the tRNA coverage was reduced by half compared to the non-depleted sample (Figure 18.C). This was intriguing as tRNAs were not targeted by DNA probes. On the other hand, the results for mRNA reads were encouraging as they represented 17% of total reads sequenced (Figure 18.D). Moreover, no significant skewing of gene expression estimates was observed using this strategy as the coverage of most transcripts was highly correlated before and after RNase H depletion (correlation coefficient of 0.996 ; Figure 18.E). To avoid a potential contamination due to the cleaving of larger rRNA fragments, we performed the RNase H treatment after gel size-selection. This greatly improved the results as rRNA reads then represented less than 50% of total reads (Figure 18.B). Meanwhile, the proportion of tRNA reads was nearly similar to what was observed in the non-depleted sample and the mRNA coverage was well increased (Figure 18.C and D). Indeed, the proportion of reads corresponding to the principal mouse isoforms represented 35% of all sequenced reads.

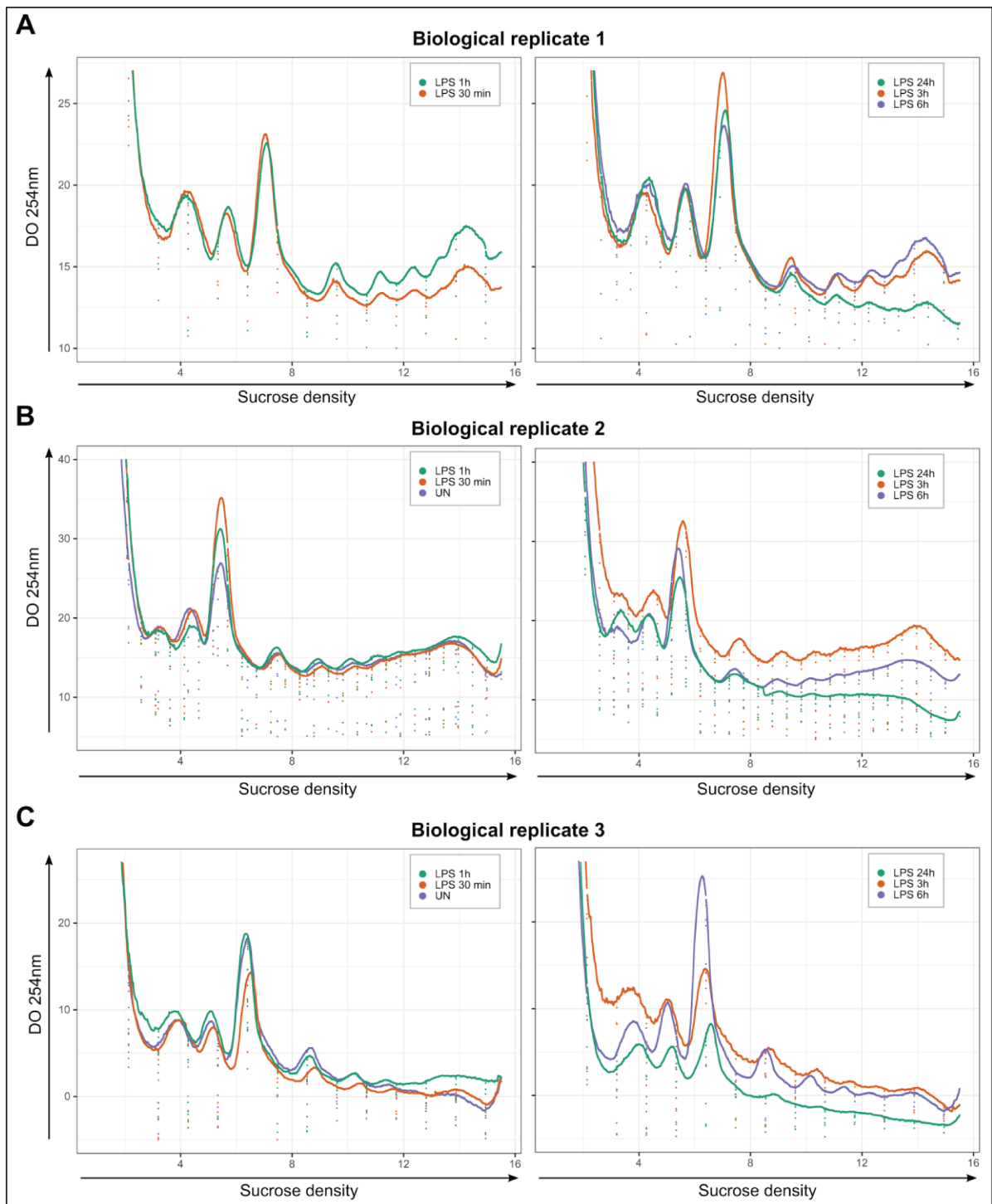


**Figure 18: The RNase H mediated depletion efficiently removes abundant contaminating rRNA fragments from footprinting samples. A.** Gel electrophoresis of RNase A and T1 or S7 treated RPFs submitted (+Dep) or not (-Dep) to RNase H depletion. Two lysates were used as input with High or Low RNA concentrations. The orange dotted rectangle highlights the range of RPFs selected for subsequent library preparation. Proportions of **B.** rRNA, **C.** tRNA, and **D.** mRNA reads relative the total reads sequenced for each type of library : RNase A and T1 and DSN treated (DSN Dep A+T1 ; 18 libraries) or non-depleted (No Dep A+T1 ; 12 libraries) or RNase S7 treated non depleted (No Dep S7 ; 1 library) or depleted with RNase H before RPFs size selection (1 library) or after (17 libraries). The interval bars correspond to the standard deviation. **E.** Scatterplot comparison of the normalized number of reads mapped per gene (Reads Per Kilobase per Million mapped or RPKM) in the non-depleted sample treated with RNase S7 and its corresponding RNase H depleted sample (depletion before size-selection).

To sum up, both the optimization of RNase digestion and the implementation of an efficient depletion strategy were necessary to obtain a sufficient coverage of mRNA sequences from pBMDM ribosome footprinting samples. The depletion of rRNA contaminants using RNase H is particularly effective and greatly improves the coverage of mRNA sequences. A reduced efficiency is however observed if RNase H treatment is performed before size selection due to the cleavage of higher rRNA fragments that are efficiently removed during gel purification. The high depletion efficiency can notably be explained by the fact that many probes were used to target the contaminating rRNA sequences. Importantly, the high coverage of rRNA sequences in the previously sequenced libraries allowed the success of this new depletion strategy as the probes were designed according to the most frequent contaminants characterized from these libraries.

### ***Endogenous RNases disturb the purification of monosomes and polysomes from pBMDMs***

Upon stimulation, macrophages undergo a rapid change of phenotype to ensure the triggering of an efficient inflammatory response. Consequently, their transcriptional and translational activities are very rapidly modified following a change in the environment. While polysome profiling is insufficient to fully characterize the variations of translation across all cellular transcripts, the amount of polysomes in a cell can serve as an indirect measure of the global protein synthesis levels. To assess the impact of LPS stimulation on protein synthesis in pBMDMs, cytoplasmic lysates were collected at different time post-activation (0, 30 min, 1h, 3h, 6h and 24h) and loaded on sucrose gradients (Figure 19). After stimulation, the amount of polysomes obtained was increased regularly through time until 6 hours post-treatment. This observation confirmed that global translation rates were impacted following LPS treatment. Interestingly, the amount of polysomes was well reduced at 24 h post-stimulation possibly revealing a translation blockade in the late phase of the inflammatory response. Even if these observations were made several times using biological replicates, the proportions of monosomes and polysomes were not consistent across experiments. Despite rigorous optimization of the sample collection procedure using iBMDMs, sucrose gradients prepared with pBMDMs lysates were still lacking reproducibility. This was a critical issue as our objective was to use monosome vs polysome footprinting to study the regulation of translation during the inflammatory response in macrophages.

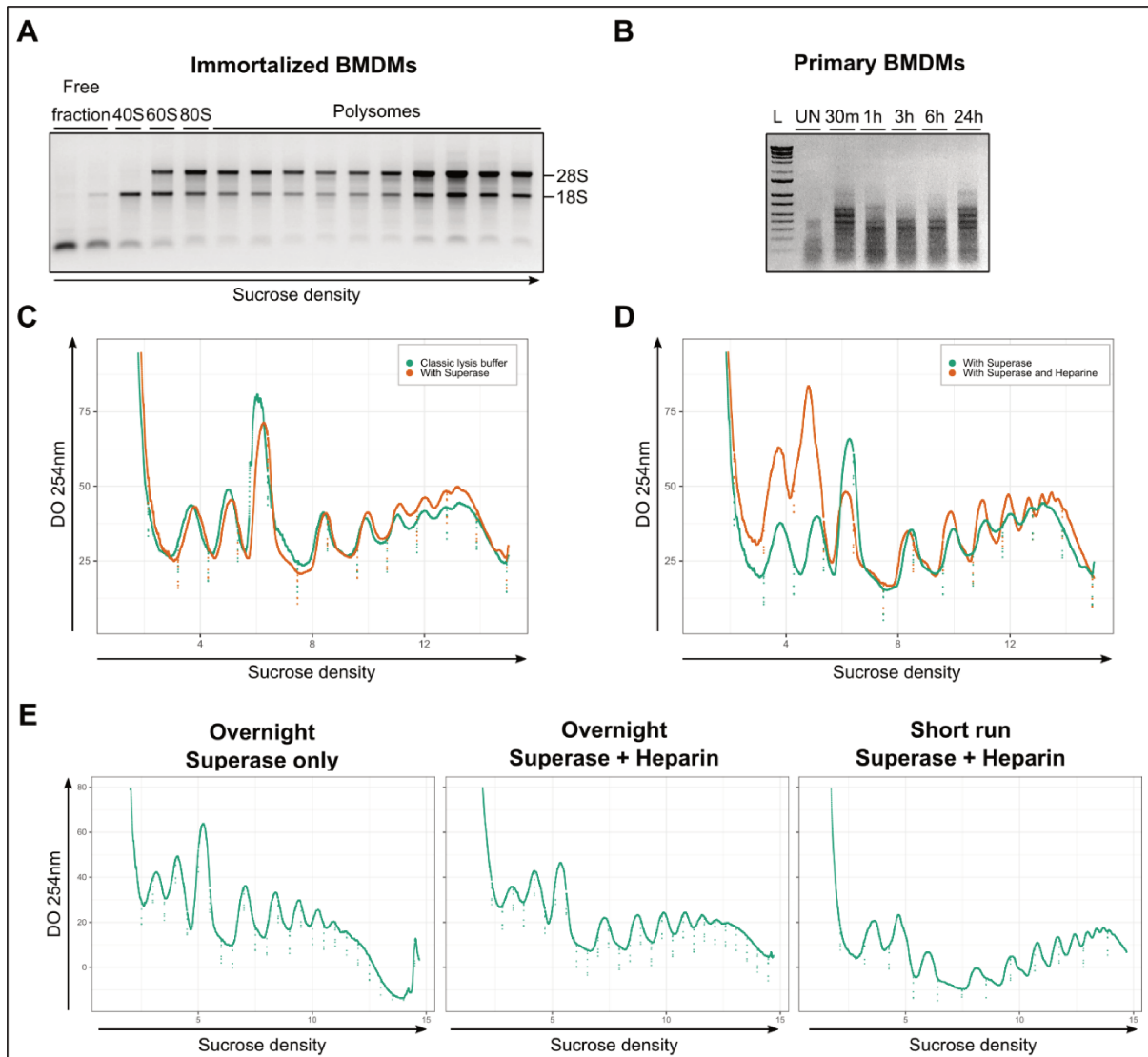


**Figure 19: Sucrose gradient profiles of pBMDMs lysates collected at different times following macrophage activation reveal a lack of reproducibility. A., B., C.** Sucrose gradient profiles obtained for pBMDMs lysates from three independent biological replicates collected at different times post-LPS stimulation (untreated, 30m, 1h, 3h, 6h, 24h).

Apart from ribosomal run-off, RNase contamination during sample collection can also disturb the ribosomal binding pattern across mRNAs and introduce variations in the monosomes vs polysomes ratio. During the initial optimization phase, the absence of RNase contamination during sample collection was tested using immortalized BMDMs. For this, a non-digested cytoplasmic lysate was loaded on a sucrose gradient and the RNAs extracted from the different fractions obtained were visualized by gel electrophoresis (Figure 20.A). Sharp 28S and 18S rRNA bands, that are characteristics of intact RNA, were observed in all monosomes and polysomes fractions. As a matter of fact, the sucrose gradient profiles prepared using iBMDMs were highly reproducible and all the RNAs obtained after purification were of really high quality. Hence, it was concluded that no RNase contamination was occurring during the sample preparation step using our optimized protocol.

After noticing the lack of reproducibility in the sucrose gradient profiles obtained for pBMDMs, I decided to check the integrity of the RNAs in the cytoplasmic lysates prepared using these cells (Figure 20.B). Unexpectedly, no 28S and 18S rRNA bands were visible in these samples and a large smear was observed at the lowest molecular sizes. The RNAs contained in the pBMDM cytoplasmic lysates were thus highly degraded even in absence of added RNases. Additionally, the degradation rates were increased when the cytoplasmic lysates were stored before loading on a sucrose gradient even if they were snap-frozen just after cell lysis (data not shown). This suggested that endogenous RNases were actively degrading the RNAs since the very beginning of the sample collection procedure. Moreover, the degradation rates were different depending on the timing post-LPS stimulation (Figure 20.B). The expression levels of the endogenous RNases could thus vary depending on the stage of the inflammatory response. These results were replicated several times using different pBMDM batches proving that the endogenous RNase contamination was a recurrent problem.



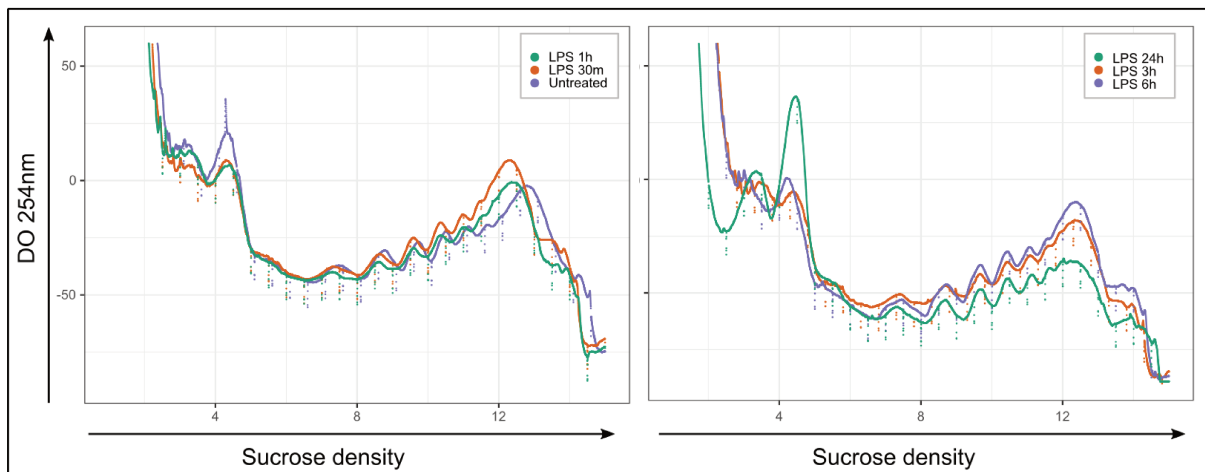


**Figure 20: Potent RNase inhibitors must be added in the lysis buffer to limit polysomes degradation due to endogenous RNases expressed in pBMDMs.** **A.** Gel electrophoresis of the RNAs purified from the different fractions of a sucrose gradient prepared with iBMDM lysate. **B.** Gel electrophoresis of the RNA purified from different pBMDM lysates collected at different time post LPS-stimulation. **C.** Sucrose gradient profiles of pBMDM lysates prepared without RNases inhibitor or with Superase (10U/mL of lysis buffer). **D.** Sucrose gradient profiles of pBMDM lysates prepared with Superase (10U/mL of lysis buffer) alone or in combination with heparin (800 $\mu$ g/mL of lysis buffer). **E.** Sucrose gradient profiles of pBMDM lysates prepared with Superase alone or in combination with heparin after overnight (15h) or short (2h40) ultracentrifugation runs.

To counteract the degradation mediated by endogenous RNases during the preparation of pBMDM cytoplasmic lysates, several RNase inhibitors were tested. Among the commercial protein-based RNase inhibitors, only the Suprase\*In from Thermo Fisher efficiently protected high density polysomes from degradation (Figure 20.C). Despite Suprase\*In ability to robustly bind to a wide range of RNases, the inhibition was only partial as polysomes degradation was still occurring if the lysate was not loaded directly on a sucrose gradient after lysis (data not shown). The addition of another broad range RNase inhibitor was thus necessary to efficiently prevent RNA degradation in the pBMDM lysates.

The anionic polymer heparin is a non-specific competitive RNase inhibitor that was previously used to limit polysomes degradation in sensitive samples (Del Prete et al., 2007; Gauthier and Ven Murthy, 1987; Gerashchenko and Gladyshev, 2017). The combination of this molecule with Suprase\*In in the lysis buffer clearly improved the recovery of high density polysomes from pBMDM lysates (Figure 20.D). In parallel, the amounts of 40S and 60S ribosomal subunits were also increased in the heparin treated sample. This could be explained by the fact that heparin induced a partial nuclear membrane disruption and a release of immature ribosomal subunits. Because of its detergent activity, heparin cannot be used to inhibit RNase degradation from any cell type. While pBMDMs were quite resistant, human HEK 293T cells that are more fragile were completely disrupted preventing the preparation of cytoplasmic lysates that could be loaded on a sucrose gradient. Despite this ribosomal subunits contamination, the quality of the sucrose gradient profiles obtained for pBMDMs was greatly increased even when the lysate was not loaded directly after lysis. Moreover, when long or overnight (15h) ultracentrifugation runs were performed to separate monosomes and polysomes, Suprase\*In only did not manage to prevent polysomal degradation (Figure 20.E). The combination of heparin with Suprase\*In was the most effective strategy to inhibit RNase degradation. Additionally, the amounts of high density polysomes were systematically higher when monosomes and polysomes were separated during short ultracentrifugation runs (2h40min). It was thus necessary to improve our protocol in order to reduce the time between cell lysis and sucrose gradient separation to obtain reliable monosomes and polysomes samples. As a consequence, for all subsequent monosome vs polysome footprinting experiments, the lysates were loaded on a gradient very rapidly after lysis and only short ultracentrifugation runs were performed.

In conclusion, endogenous RNases released during pBMDM lysis dramatically reduce RNA samples quality and disturb the separation of monosomes and polysomes. To achieve reliable ribosomal purification, monosomes must be separated from polysomes as quickly as possible. Moreover, the use of potent RNase inhibitors such as Suprase\*In and heparin is required. By following all these precautions, the reproducibility of the sucrose gradient profiles obtained for pBMDMs was greatly improved. The observations made regarding the modifications of global protein synthesis rates in pBMDMs following LPS stimulation at different time post-activation (0, 30 min, 1h, 3h, 6h and 24h) were also replicated (Figure 21).

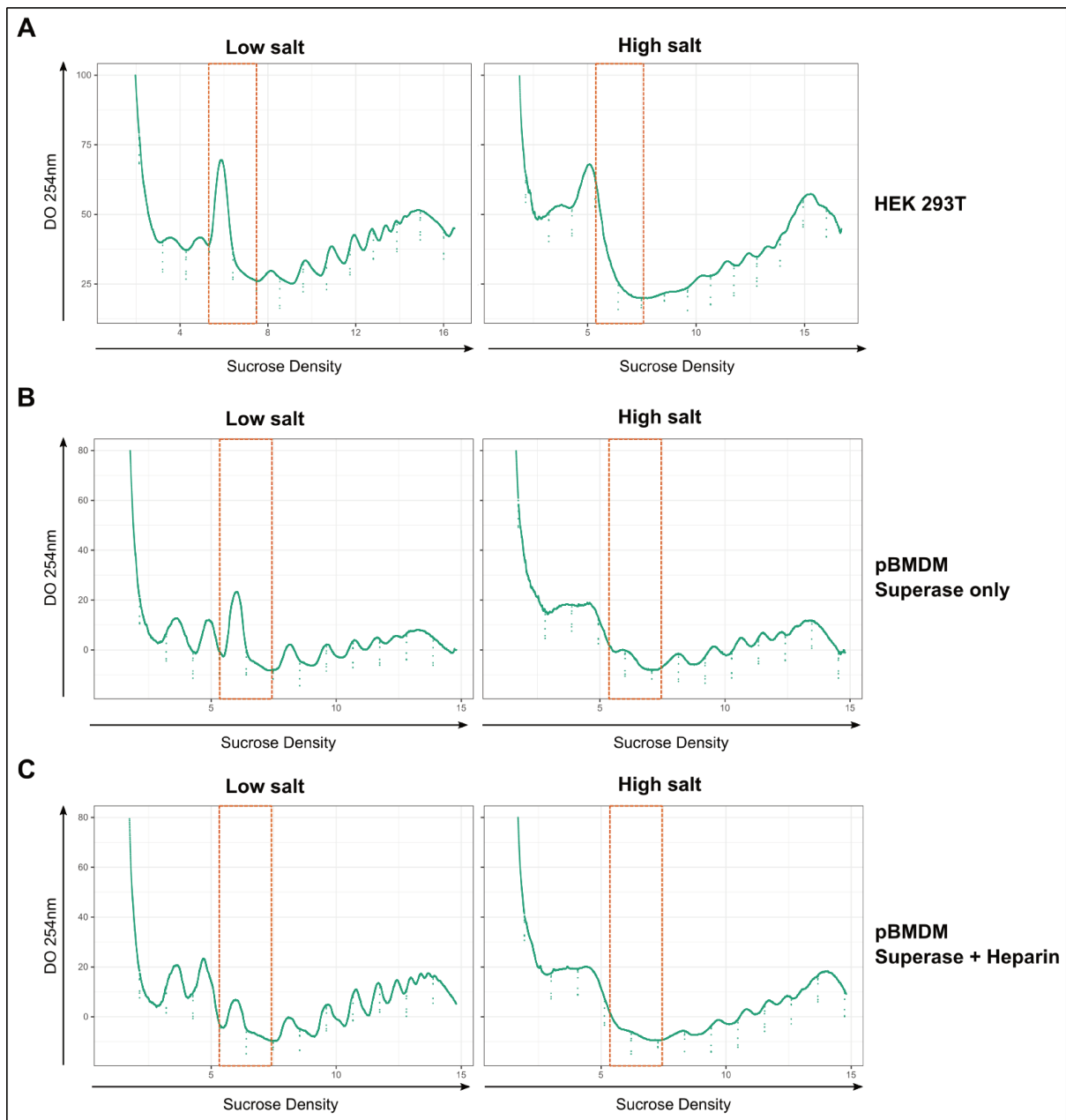


**Figure 21: Optimized sample preparation to inhibit endogenous RNases activity improves the reproducibility of sucrose gradient profiles obtained for pBMDMs.** Sucrose gradient profiles obtained for pBMDMs lysates from collected at different times post-LPS stimulation (untreated, 30m, 1h, 3h, 6h, 24h). These gradients are representative of eight independent biological replicates.

### ***Improving ribosomal footprints selection using high salt conditions***

After improving the sample collection and the library preparation steps, I sought to find a strategy to select the RPFs most likely produced by translating ribosomes and increase the reliability of our monosome vs polysome footprinting results. Indeed, according to previous studies, not all the ribosomes composing the monosomal fraction are actually bound to mRNAs *in vivo* (Arava et al., 2003; Heyer and Moore, 2016). Ribosomal subunits released upon cell lysis could also re-associate randomly with the mRNAs producing artifactual RPFs. Moreover, RBP-mRNA complexes of similar size than the 80S complexes can produce footprints that are then co-purified with RPFs. High salt treatment can limit weak ionic protein-RNA interactions and thus remove false footprints derived from RBPs or nonspecific 80S binding (Miettinen and Björklund, 2015). Interestingly, 80S complexes can be stabilized by interacting with other molecules such as mRNA, tRNAs or proteins and resist high salt conditions. The addition of 1M KCl in the cytoplasmic lysates before separation of monosomes and polysomes could thus be used to select ribosomes associated to translated mRNAs (Blobel and Sabatini, 1971; Martin and Hartwell, 1970; Mills et al., 2016; Zylber and Penman, 1970). Additionally, high ionic conditions prevent the interactions between the ribosomes and translation factors that were not tightly bound thus limiting post-lysis ribosomal movements across mRNAs (Mohammad et al., 2019).

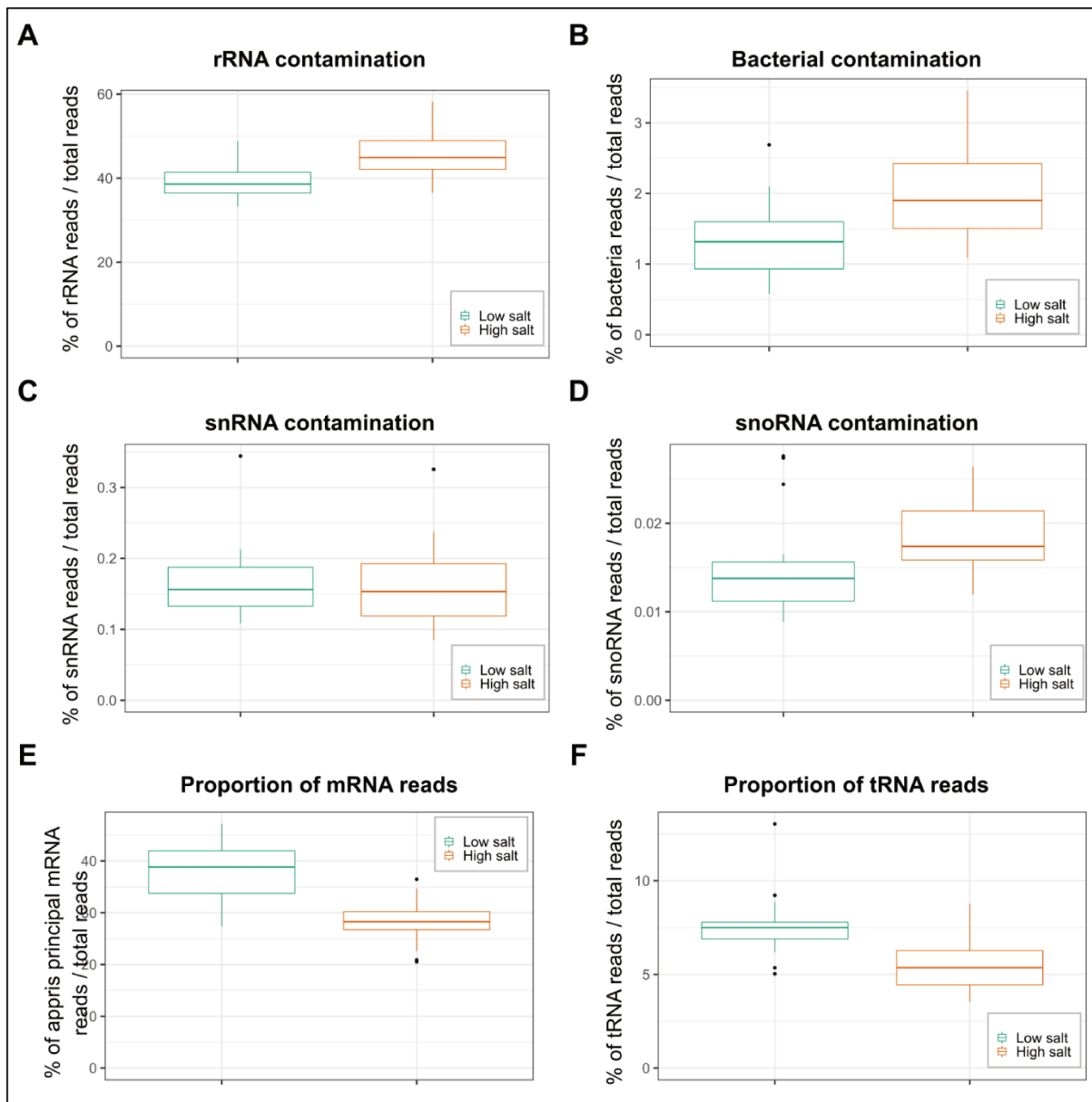
To assess the effect of high salt treatment on monosomal and polysomal populations, cytoplasmic lysates were supplemented with KCl to reach the final concentration of 1M and incubated on ice for 20 min before loading on a sucrose gradient. When lysates derived from HEK 293T cells were exposed to such high salt treatment, the monosomes were completely dissociated leading to a high increase of 40S and 60S subunits peaks compared to the low salt or classical condition (Figure 22.A). Meanwhile, the total amount of polysomes was not significantly impacted even if an increase of high density polysomes was observed. This confirmed that most translation events occur in polysomes in highly proliferating cells such as HEK 293T.



**Figure 22: High salt treatment facilitates the purification of ribosomes likely involved in translation. A.** Sucrose gradient profiles of HEK 293T lysates incubated on ice for 20 min in low salt (100mM KCl) or high salt (1M KCl) conditions prior loading on the gradient. **B.** Sucrose gradient profiles of pBMDM lysates prepared with Superase and incubated on ice for 20 min in low salt (100mM KCl) or high salt (1M KCl) conditions prior loading on the gradient. **C.** Sucrose gradient profiles of pBMDM lysates prepared with both Superase and heparin before incubation on ice for 20 min in low salt (100mM KCl) or high salt (1M KCl) conditions prior loading on the gradient.

The dissociation of monosomes into separate small ribosomal subunits was also visible in pBMDMs lysates treated with 1M KCl but to a lesser extent (Figure 22.B). Interestingly, the amount of polysomes was higher in the high salt treated sample compared to the low salt condition. This observation suggested that in addition to the depletion of monosomes non engaged in translation, the incubation in high salt conditions could also slow down polysome fragmentation induced by endogenous RNases. Notably, the amount of monosomes was well reduced after addition of 1M KCl in pBMDM lysates prepared with heparin (Figure 22.C). As RNase digestion was still occurring in the pBMDM lysate treated only with Suprase\*In, the monosomal peak obtained previously was in fact a mixture of monosomes produced by polysome degradation and truly translating monosomes (Figure 22.B). By using the combination of high RNase inhibitors and 1M KCl treatment, we were thus able to visualize the very small monosomal fraction likely to be translationally active in pBMDMs (Figure 22.C). Moreover, the amount of high density polysomes was well increased in the high salt treated pBMDM lysate prepared with heparin. The incubation with 1M KCl thus further inhibited the endogenous RNases activity. Altogether, these results demonstrated that the addition of the high salt treatment could really improve the reliability and the resolution of ribosomal footprinting studies targeting translationally active ribosomes.

The impact of the 1M KCl treatment was first assessed on classical ribosome profiling samples. For this, each pBMDM lysate used was splitted in two samples incubated in low or high salt conditions before further processing. As the exposure to high ionic conditions increased the ribosome susceptibility to exogenous RNase digestion, the samples were desalted using Zeba Spin columns from Thermo Fisher before addition of the S7 nuclease. It should be noted that this step was not necessary for the preparation of monosome vs polysome footprinting samples as the KCl excess is diluted during the first sucrose gradient separation. After digestion and monosomes purification, the samples were subjected to RNase H treatment to deplete rRNA fragments and used for library preparation.



**Figure 23: Quantitative analysis of the different RNA species identified in classical or high salt treated ribosome profiling libraries.** Proportion of **A.** rRNA, **B.** bacteria, **C.** snRNA, **D.** snoRNA, **E.** mRNA, **F.** tRNA reads relative to the total reads sequenced for each type of library. The low salt samples correspond to libraries prepared with the classical KCl concentration (100mM). For the high salt samples, lysates were incubated with 1M KCl on ice for 20min prior loading on a sucrose gradient. The results obtained for 17 low salt and 18 high salt treated ribosome profiling libraries are represented.

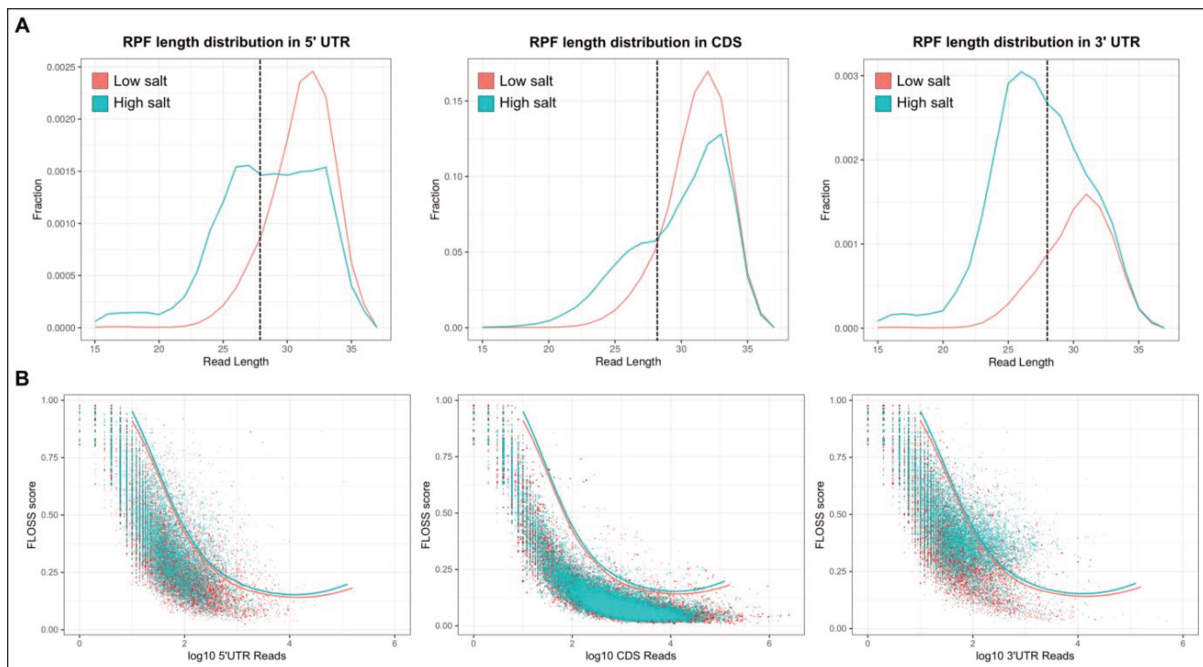
The analysis of the most frequent contaminants found in these new ribosome profiling libraries confirmed the efficiency of our optimized protocol. Indeed, the proportion of rRNA reads was below 50% and the bacterial contamination was reduced from 10% in our first attempt to less than 2% in most samples (Figure 23.A and B). Moreover, the amounts of snRNA and snoRNA contaminating sequences were still very low (Figure 23.C and D). Concomitantly, the proportions of mRNA and tRNA sequences amongst total reads were well increased compared to our first ribosome profiling samples (Figure 23.E and F). Importantly, the amounts of mRNA and tRNA reads were decreased in the high salt treated samples compared to their corresponding low salt samples. This could be partially explained by the more stringent selection of RPFs as footprints generated by RBPs or non-specific ribosomal binding are depleted in this condition. Additionally, ribosomes treated with 1M KCl could be more sensitive to RNase degradation even after desalting leading to a partial loss of RPFs (decrease of 25% in the high samples compared to the low salt ones). This hypothesis could notably explain the increased amounts of rRNA reads observed in the high salt treated samples compared to the low salt conditions (Figure 23.A). This increased rRNA contamination is however not sufficient to explain entirely the reduction of mRNA coverage.

To check the accuracy of the ribosomal footprinting, mRNA reads were then analyzed using the RiboFlow pipeline (Ozadam et al., 2020). The great advantage of this bioinformatic tool is that it allows the analyses of ribosomal densities depending on the RPFs size. Indeed, RPF lengths are variable and carry information on the translational stage of the ribosomes (Mohammad et al., 2019). Due to the different structural rearrangements occurring during protein synthesis, the portion of mRNA protected by the ribosome can change depending on the translation step (Lareau et al., 2014; Wu et al., 2019). Notably, during elongation, the ribosomes oscillate between two different conformations before and after peptide bond formation (Lareau et al., 2014). Before peptide bond formation, the ribosomal A site is occupied by an aminoacyl tRNA and the mRNA portion protected is around 28-30 nucleotides long. Upon peptide bond formation, the ribosomal subunits rotate and the previously aminoacylated tRNA is relocalized to the P site leaving a higher portion of the mRNA vulnerable to RNase digestion. Post peptide bond or translocating ribosomes thus produce shorter RPFs (18-22 nucleotides long). These smaller footprints cannot be visualized if the samples are pre-treated with CHX as this drug mostly stabilizes pre-peptide bond ribosomes. Moreover, they can be lost during gel size selection if the RPFs size range selected is too high. To obtain a good picture of the ribosomes at various stages of the translational cycle, RPFs between 19 and 38 nucleotides were selected to prepare our footprinting libraries. As S7 nuclease has a tendency to leave one or more nucleotides undigested at the 3' boundary of the ribosomes, the RPFs produced by S7 digestion usually display a broader length



distribution compared to other RNases (Gerashchenko and Gladyshev, 2017; Hwang and Buskirk, 2017). Consequently, the RPFs produced by elongating ribosomes are expected to be slightly larger in our samples compared to the stereotyped 18-22nt and 28-30nt RPFs obtained in yeast ribosome profiling studies.

When looking at the RPFs length distribution in our ribosome profiling libraries, the two populations corresponding to distinct elongation stages are only observed in the 1M KCl treated samples (Figure 24.A). This result thus confirmed that the high salt treatment could really improve the footprinting resolution. Moreover, the RPFs produced by pre or post peptide bond ribosomes are indeed a few nucleotides longer than what was obtained in previous yeast studies that did not use S7 nuclease (around 29-35nt and 22-28nt respectively). The RPF length distribution was also very different in the UTR regions compared to the CDS region in the 1M KCl treated sample with more small RPFs observed in 5' and 3' UTRs. These small RPFs could be produced by ribosomes with an empty A-site such ribosomes reinitiating after uORF translation in the 5'UTR or post-termination non recycled ribosomes in the 3'UTR (Skabkin et al., 2013). Interestingly, the association with an mRNA and a tRNA could be sufficient to stabilize the ribosome and confer resistance to high salt treatment independently of the presence of a nascent peptide. The smaller size RPFs observed in the UTRs are thus produced by a different ribosomal population than the one actively elongating restricted to the CDS. These different ribosome populations were not visible in the classical or low salt samples as only large size RPFs were observed in all the mRNA regions. This could be explained by a reduced RNase treatment efficiency compared to the high salt condition leading to an incomplete digestion at the ribosome boundaries in the CDS and the UTRs. Alternatively, the high salt treatment could have removed translation factors or RBPs that were loosely bound to the 5'UTR and 3'UTR ribosomes leaving larger unprotected mRNA portions.

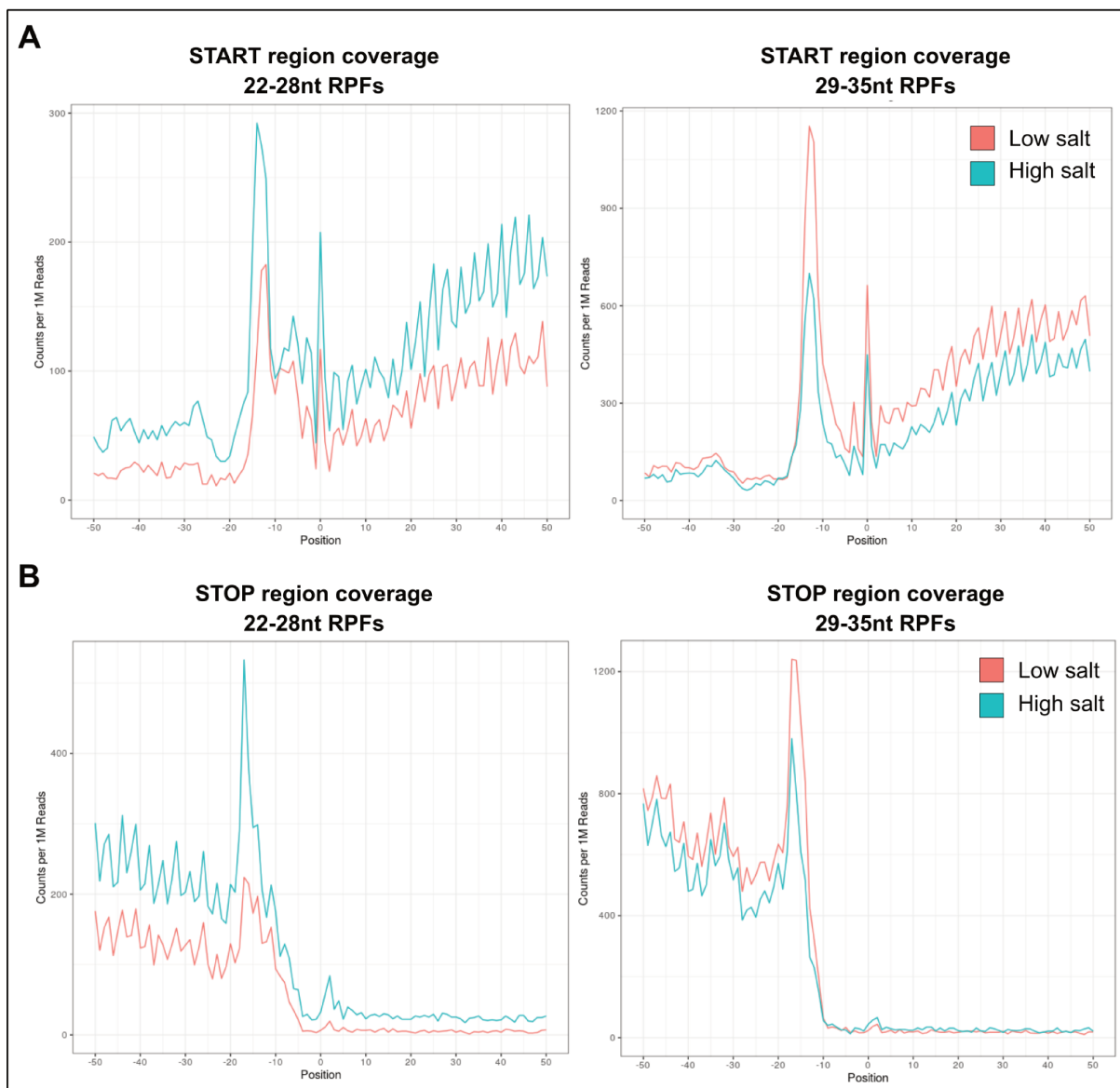


**Figure 24: High salt treatment improves the ribosome footprinting quality and the detection of different ribosomal conformations. A.** Comparative analysis of the length distribution of RPFs mapping to the 5'UTR, CDS or 3'UTR regions in the mouse pBMDM mRNAs using the RiboFlow pipeline. The frequency of each RPF length is normalized to the number of reads mapping in each mRNA region (5'UTR, CDS or 3'UTR) separately. **B.** Fragment Length Organizing Similarity Score (FLOSS) analysis of the RPFs mapping in each mRNA region (5'UTR, CDS or 3'UTR) in low or high salt ribosome profiling libraries.

To further validate our observations, the RPFs size distribution in the different regions of the mRNAs were compared using the Fragment Length Organizing Similarity Score (FLOSS) algorithm for high or low salt treated samples (Figure 24.B). This bioinformatic tool measures the degree of difference between the RPFs size distribution in a region of interest, like the 5' or the 3' UTRs, and a reference set, such as the CDS region (Ingolia et al., 2014). Higher FLOSS score values are thus representative of an increased difference between the reference and the region of interest. As the variability is increased when the sequencing coverage is low, the FLOSS values are less informative if the number of mapped reads is reduced. Interestingly, the FLOSS score is calculated for each transcript individually so it is possible to identify those that are not fitting to the global distribution. For this, the FLOSS score values were plotted depending on the number of reads mapping to each mRNA region. Each individual transcript was represented as a dot and all dots below a statistical cut-off line correspond to transcripts that generate RPFs of similar size to the reference. As the CDS region was selected as the reference for this analysis, most of the transcripts displayed a footprinting pattern that is similar to the global distribution in this region. In the 5'UTR, more differences were visible even in the low salt samples as more individual FLOSS values were above the global distribution. The differences were however higher in the 1M KCl treated

samples as observed when looking directly at the RPFs size distribution. The biggest differences were observed in the 3'UTR region in high salt treated samples as many transcripts displayed a different RPFs distribution in this region compared to the CDS. Altogether these results demonstrate that RPFs obtained in the UTRs display a different pattern than CDS derived RPFs revealing their engagement in another step of the translation cycle. Importantly, high salt treatment can be used to discriminate more easily ribosomes that are actively elongating from those that are not.

Finally, to control the translational status of the different RPF populations, the 5' end read densities were aggregated across all transcripts to assess the global coverage rates around the translation start and stop sites using RiboFlow pipeline (Ozadam et al., 2020). The results revealed that both small and large RPFs display a clear 3 nucleotide periodicity in the CDS region but not in the UTRs (Figure 25.A and B). Thus this further demonstrated that most ribosomes bound in the 5' and 3' UTR are not actively elongating. Intriguingly, a higher small RPFs signal is observed in the 5'UTR region close to the start codon in the high salt samples. These 22-28nt RPFs could be produced by ribosomes reinitiating after uORF translation with an empty A-site that are better detected in this condition. Moreover, the periodicity peaks are sharper and better defined in the 1M KCl treated samples compared to the low salt conditions for both pre and post peptide bond ribosomes. In conclusion, the high salt treatment really improved the quality and the resolution of the footprinting results obtained.

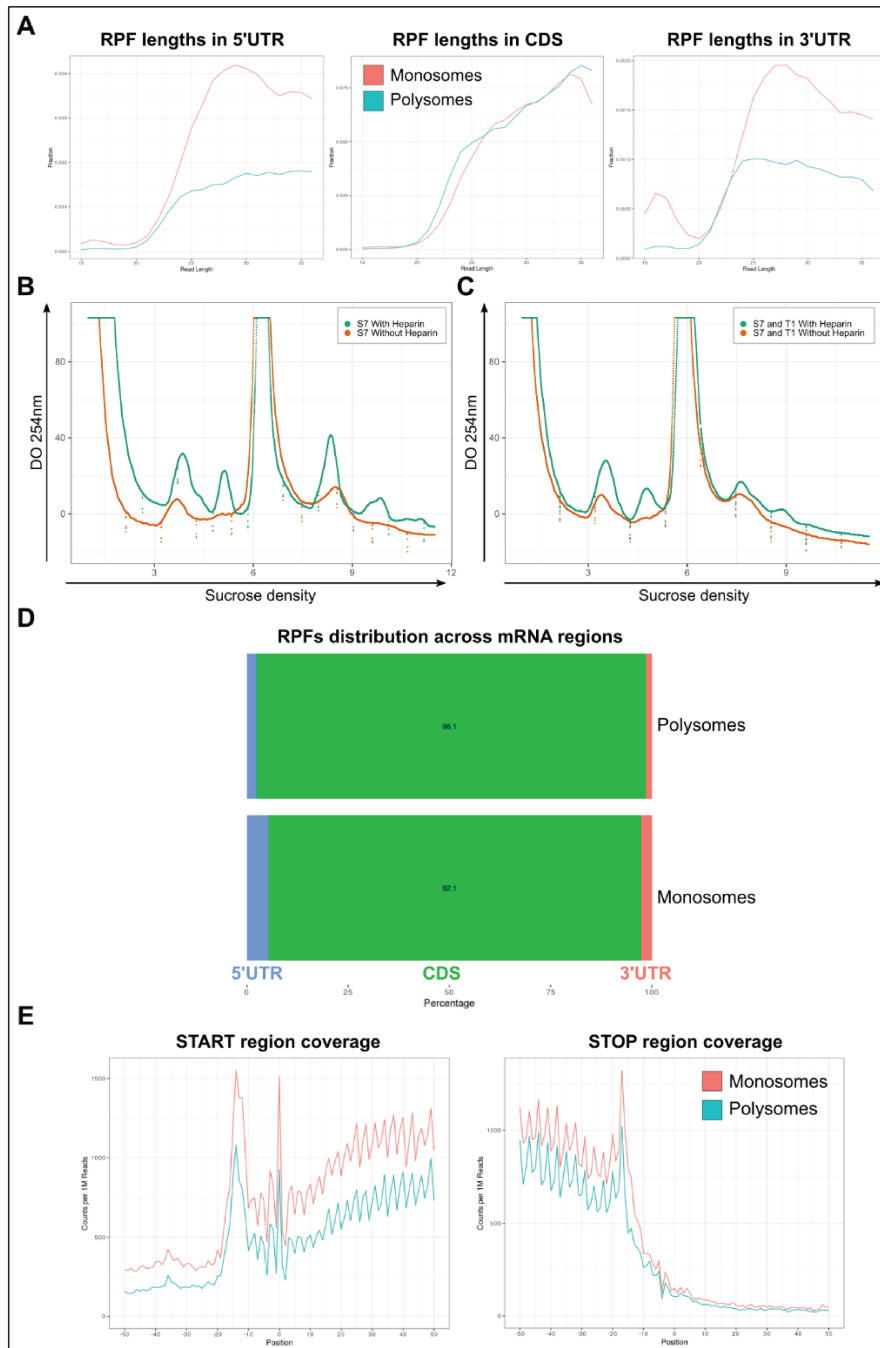


**Figure 25: High salt treatment improves the three nucleotides periodicity in ribosome profiling libraries. A.** Coverage of the start codon region (50nt range before and after the start codon) across mRNAs in the low or high salt treated ribosome profiling libraries. The analysis was performed separately for the two RPFs size ranges corresponding to pre (29-35nt) or post (22-28nt) peptide bond conformations of the ribosomes. **B.** Coverage of the stop codon region (50nt range before and after the stop codon) across mRNAs in the low or high salt treated ribosome profiling libraries. The analysis was performed separately for the two RPFs size ranges corresponding to pre (29-35nt) or post (22-28nt) peptide bond conformations of the ribosomes.

## ***Monosome vs polysome footprinting using pBMDMs***

Following the extensive optimization of our protocol to obtain robust and reliable results, I prepared monosome vs polysome libraries from LPS stimulated pBMDMs lysed in presence of Suprase\*In and heparin, submitted to high salt treatment before loading on the first gradient, digested with 3 $\mu$ L of S7 nuclease and RNase H, and depleted before library preparation. The results obtained were surprisingly bad : on average only 6% and 7.5% of total sequenced reads corresponded to mRNA fragments in the monosome and polysome footprinting libraries. The rRNA contamination levels were higher than what was obtained for global ribosomal profiling but could not explain entirely why the coverage of mRNA sequences was so low (average contamination rates of 53% in monosome libraries and 80% in polysome libraries).

The explanation came after analyzing the mRNA derived RPFs length distributions (Figure 26.A). The RPFs size range was in fact higher than expected in all mRNA regions in both monosomes and polysomes footprinting libraries. Particularly, the distribution of RPFs generated by CDS associated ribosomes was very broad with some footprints being larger than 36 nucleotides long. As the RPFs were gel size-selected between 19 and 38 nucleotides prior library preparation, larger fragments were depleted thus preventing the sequencing of all mRNA portions actually protected by the ribosomes. This reduced efficiency of the RNase treatment could be explained by the presence of heparin traces remaining in the monosomal and the polysomal fractions even after the first sucrose gradient purification. Indeed, we previously tested the impact of heparin addition on the RNase treatment when performed directly in the cytoplasmic lysate and observed that it could reduce its efficiency (Figure 26.B). Notably, the collapse of polysomes into monosomes was incomplete in the presence of heparin. For this reason, heparin was omitted during the preparation of global ribosome profiling samples. On the other hand, we did not expect that heparin traces left after sucrose gradient purification could effectively inhibit the RNase digestion step for monosome and polysome footprinting samples. As RNase T1 was described as less sensitive to heparin inhibition (Gerashchenko and Gladyshev, 2017), I next tested if its combination with S7 nuclease could improve polysomes collapse despite the presence of high RNases inhibitors amounts in the cytoplasmic lysates. The results obtained using this combination of RNases were encouraging as less polysomes were remaining after treatment even in heparin containing samples (Figure 26.C). With this in mind, I decided to set up another monosome vs polysome footprinting experiment specially designed to find the best RNase digestion conditions for our samples.



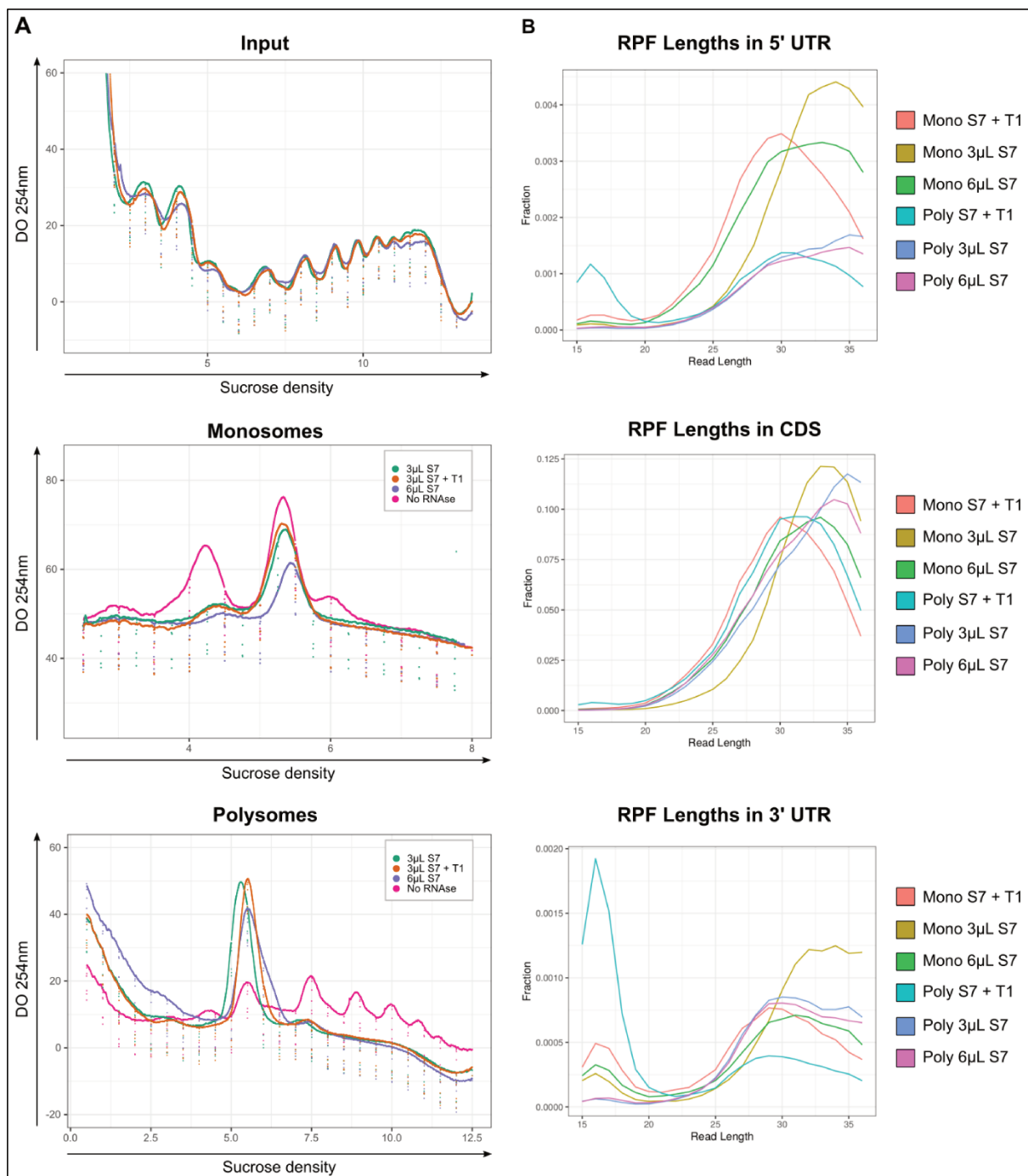
**Figure 26: Heparin traces in purified monosomes and polysomes reduce their footprinting quality.** **A.** Comparative analysis of the length distribution of RPFs mapping to the 5'UTR, CDS or 3'UTR regions in the mouse pBMDM mRNAs using the RiboFlow pipeline. The frequency of each RPF length is normalized to the number of reads mapping in each mRNA region (5'UTR, CDS or 3'UTR) separately. **B.** Sucrose gradient profiles of pBMDMs lysates treated with 3 $\mu$ L of RNase S7 in presence or in absence of heparin. **C.** Sucrose gradient profiles of pBMDMs lysates treated with 3 $\mu$ L of RNase S7 and 150U of RNase T1 in presence or in absence of heparin. **D.** Comparison of the proportions of RPFs mapping to the 5'UTR, CDS or 3'UTR mRNA regions in monosome or polysome footprinting libraries using the RiboFlow pipeline. **E.** Coverage of the start and stop codon regions (50nt range before and after the start/stop codons) across mRNAs in the monosome vs polysome footprinting libraries.

Even though the footprinting pattern was probably biased in our incompletely digested monosome vs polysome libraries, I decided to analyze the few mRNA reads obtained using the RiboFlow pipeline. Importantly, most of the RPFs mapping to mRNA sequences were derived from the CDS region in both monosome and polysome footprinting samples (92.1% and 96.1% respectively ; Figure 26.D). This result clearly suggests that translationally active monosomes are not restricted to the pioneer round of translation initiation or to terminating ribosomes. On the contrary, as observed in yeast, most mRNA bound monosomes are likely elongating or stalled all along the CDS region. Another proof of this was the clear 3nt periodicity observed within the CDS but not in the UTR regions bordering the start and stop codons in our monosome footprinting libraries (Figure 26.D). In conclusion, despite the incomplete digestion and the loss of many RPFs during sample preparation, we were already able to confirm that monosomes are also involved in translation in murine pBMDMs in basal conditions and after LPS activation.

In order to select the best RNase conditions to obtain reliable monosome and polysome footprinting results, monosomal and polysomal fractions prepared after sucrose gradient ultracentrifugation of several untreated pBMDM lysates were pooled. The different fractions were then digested either with the same amount of S7 nuclease as previously (3 $\mu$ L S7), with a doubled amount of S7 alone (6 $\mu$ L S7) or in combination with the RNase T1 (3 $\mu$ L S7 + 150U T1). Moreover, a non-digested sample was also prepared to be able to check the RNase digestion efficiency directly during the second sucrose gradient purification. For all RNase conditions tested, the digestion was efficient despite the heparin traces as all treated polysomes were properly collapsed into monosomes (Figure 27.A). In monosome samples, the 60S peaks were also well reduced after RNase treatment showing that no inhibition due to heparin was occurring. The 80S peaks obtained post-digestion were well reduced in the monosome and polysome samples treated with a higher amount of S7 nuclease (6 $\mu$ L) revealing a partial loss of ribosomal integrity. Conversely, the combination of RNase T1 with S7 nuclease yielded approximately the same amount of 80S complexes as when the fractions were digested with a regular amount of S7 alone. The combination of both RNases at lower concentrations could thus be less aggressive towards the ribosomes.

Clear differences of the RNase digestion efficiency were observed when looking at the RPF lengths distribution in the monosome or polysome footprinting libraries prepared from the different digestion conditions (Figure 27.B). As expected, the RPFs size range was very broad in the monosome and polysome samples prepared with 3 $\mu$ L of S7 nuclease. Many large fragments representative of an incomplete digestion were observed in these samples. These results in accordance with our previous experiment thus indicated that only a small adjustment of the RNase conditions could be sufficient to improve the footprinting quality. Furthermore, doubling the S7 nuclease amount shifted the RPF lengths distribution toward lower sizes. Yet, too many large fragments were still remaining in both monosome and polysome footprinting samples. The RPF lengths distributions were greatly improved when a regular amount of S7 nuclease was combined with a small amount of RNase T1. This condition could thus be the best to perform monosome vs polysome footprinting using pBMDM samples. The length distributions obtained were however not as narrowed as what was observed previously for global ribosome profiling. This could notably explain why it was not possible to discriminate between pre and post peptide bond formation derived RPFs in our monosome and polysome libraries. Further optimization of the RNase digestion conditions could be required if a better discrimination of these two ribosomal populations is desired.

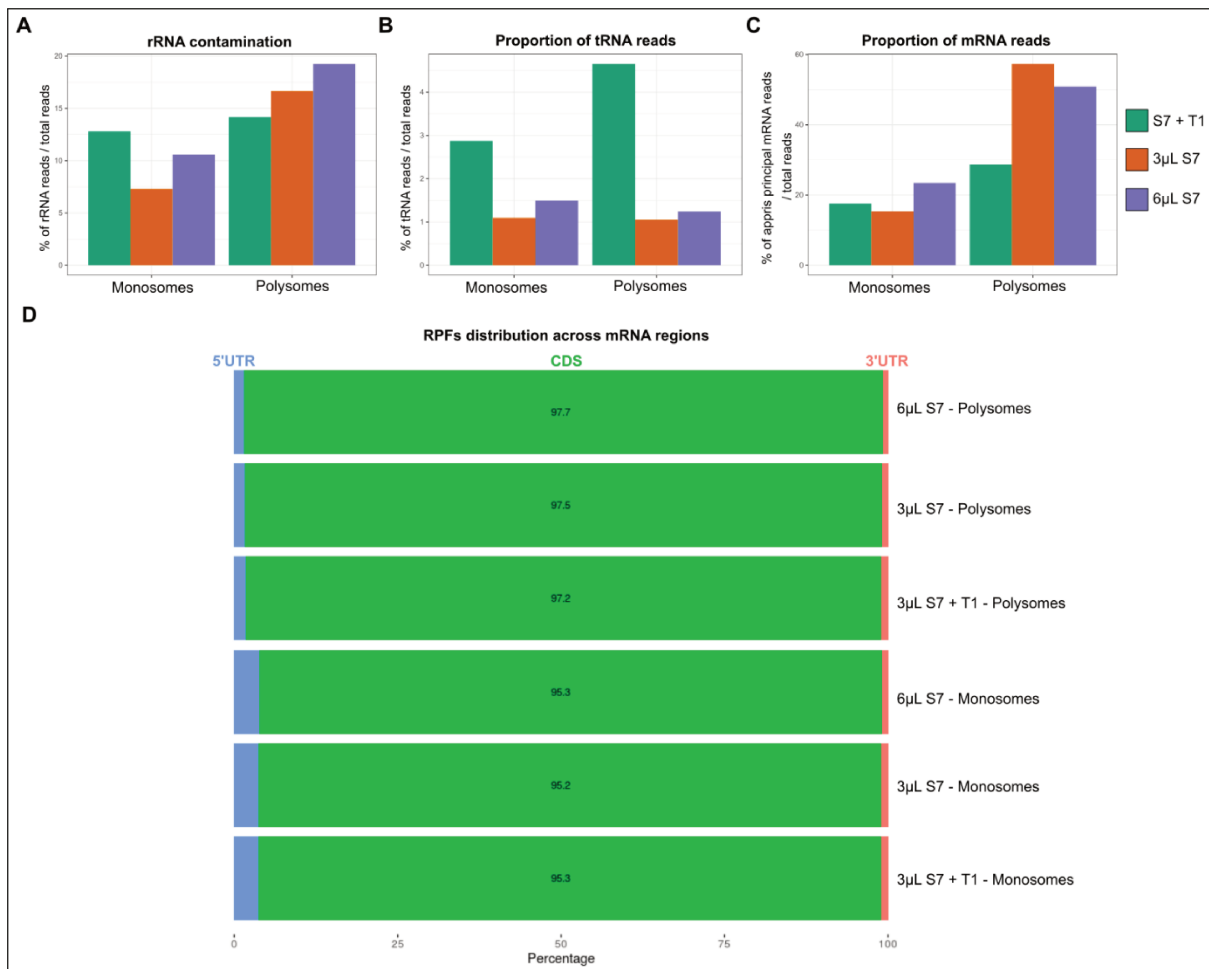




**Figure 27: Optimization of the RNase digestion conditions to improve the monosome vs polysome footprinting protocol. A.** Sucrose gradient profiles of the input pBMDM lysates used for the setting up of the monosome vs polysome footprinting digestion step. Monosomes and polysomes fractions from several sucrose gradients were pooled separately and then submitted to different RNase digestion conditions as indicated. The sucrose gradient profiles obtained during the selection of properly digested monosomes and polysomes are also shown. **B.** Comparative analysis of the length distribution of RPFs mapping to the 5'UTR, CDS or 3'UTR regions in the mouse pBMDM mRNAs for each sample. The frequency of each RPF length is normalized to the number of reads mapping in each mRNA region (5'UTR, CDS or 3'UTR) separately.

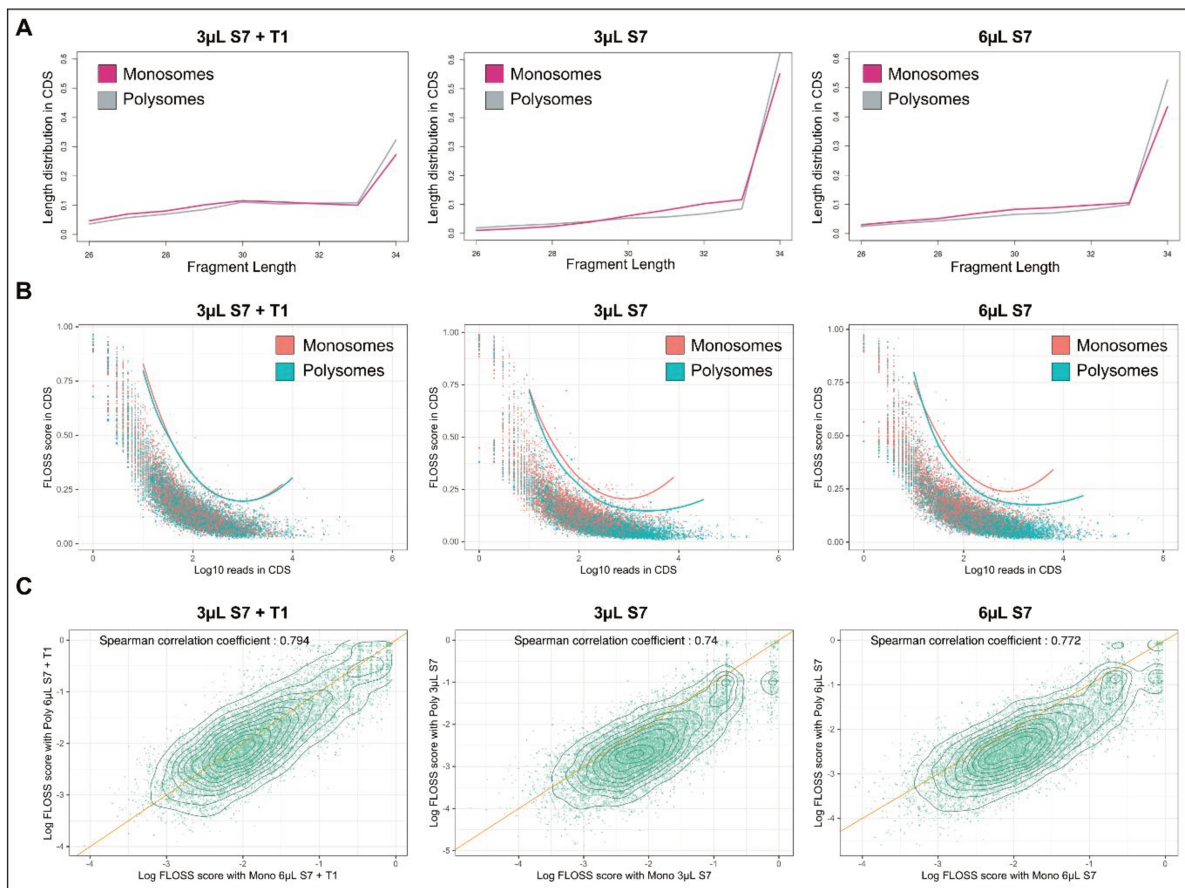
Interestingly, a higher peak corresponding to small 16nt RPFs was observed in the 5' and 3' UTR regions in the polysome sample treated with RNase S7 and T1 (Figure 27.B). This RPF size has been previously associated with ribosomes stalled on 3' end truncated mRNAs that require the recruitment of recycling factors to be removed (Guydosh and Green, 2014; Young et al., 2015). This observation thus suggested that ribosome recycling after termination in the 3' UTR could be less efficient on some polysome bound mRNAs. Consequently, the 80S complexes could remain associated with the mRNA even after peptide release allowing their movement and/or stalling in the 3'UTR region. This previously described mechanism could promote reinitiation on the same mRNA as scanning could be resumed directly after the stop codon (Guydosh and Green, 2014; Skabkin et al., 2013). A similar phenomenon could be occurring in the 5'UTR region with ribosomes not efficiently recycled at the end of uORFs. If the ribosomes cannot resume scanning, they could trigger the No Go decay pathway hence leading to the production of 16nt RPFs. Additionally, it was recently shown that uORF translation could trigger NMD through the recruitment of UPF1 (Jia et al., 2020). Alternatively, these very small RPFs could be generated by the over digestion of pre peptide bond ribosomes by the RNase T1. Indeed, this RNase is more processive than the S7 nuclease and could efficiently degrade the mRNA fragment at each side of the ribosomal boundaries. The comparison of the RPF lengths distribution in monosome and polysome footprinting libraries prepared with a higher amount of RNase S7 alone would be necessary to discriminate between these two scenarios.

When looking at the proportion of the different RNA species in each of these monosome and polysome footprinting libraries, we observed that the amounts of contaminating rRNA reads were globally lower than for all the previous attempts (less than 20% of total reads ; Figure 28.A). In addition to the RNase H mediated depletion effect, the ribosomes integrity was thus better preserved for all RNase conditions tested. The combination of RNases S7 and T1 seemed less aggressive towards polysomes than the other RNase conditions as less rRNA contaminants were produced. On the contrary, for monosomes, more rRNA fragments were obtained using this combination. This observation thus revealed that the best RNase condition could be different for monosomes and polysomes footprinting. The amounts of tRNA fragments retrieved were also higher when RNase S7 and T1 were combined, consistent with a stronger RNase digestion efficiency (Figure 28.B). Finally, the amounts of mRNA reads obtained were higher when only RNase S7 was used during sample preparation (Figure 28.C). Despite improving the RPF lengths distribution, the addition of RNase T1 could induce a loss of RPFs thus decreasing the amount of information retrieved from our footprinting libraries. Interestingly, the amounts of mRNA derived RPFs were quite similar in the samples treated with 3 $\mu$ L or 6 $\mu$ L of RNase S7. Although the 80S peaks observed on the sucrose gradients were reduced when high S7 nuclease amounts were used (Figure 27.A), it did not induce a significant loss of RPFs. Due to its reduced digestion efficiency, the RNase S7 is less aggressive towards the ribosomes and consequently could provide the most intact picture of the ribosomal occupancies in pBMDMs. Finally, independently of the coverage differences between RNase conditions, most of the RPFs mapping to mRNA sequences were derived from the CDS region in all our monosome and polysome footprinting libraries. Indeed, after analysis with the RiboFlow pipeline, we observed that on average 95% of mRNA reads corresponded to CDS in the monosomes libraries and 97% in the polysomes ones (Figure 28.D). This further confirmed that monosomes are not restricted to initiating or terminating ribosomes and can also elongate all across the CDS. Interestingly, the amounts of 5'UTR derived RPFs were higher in the monosome footprinting libraries in accordance with a higher involvement of this ribosomal population in the translation of short uORFs.



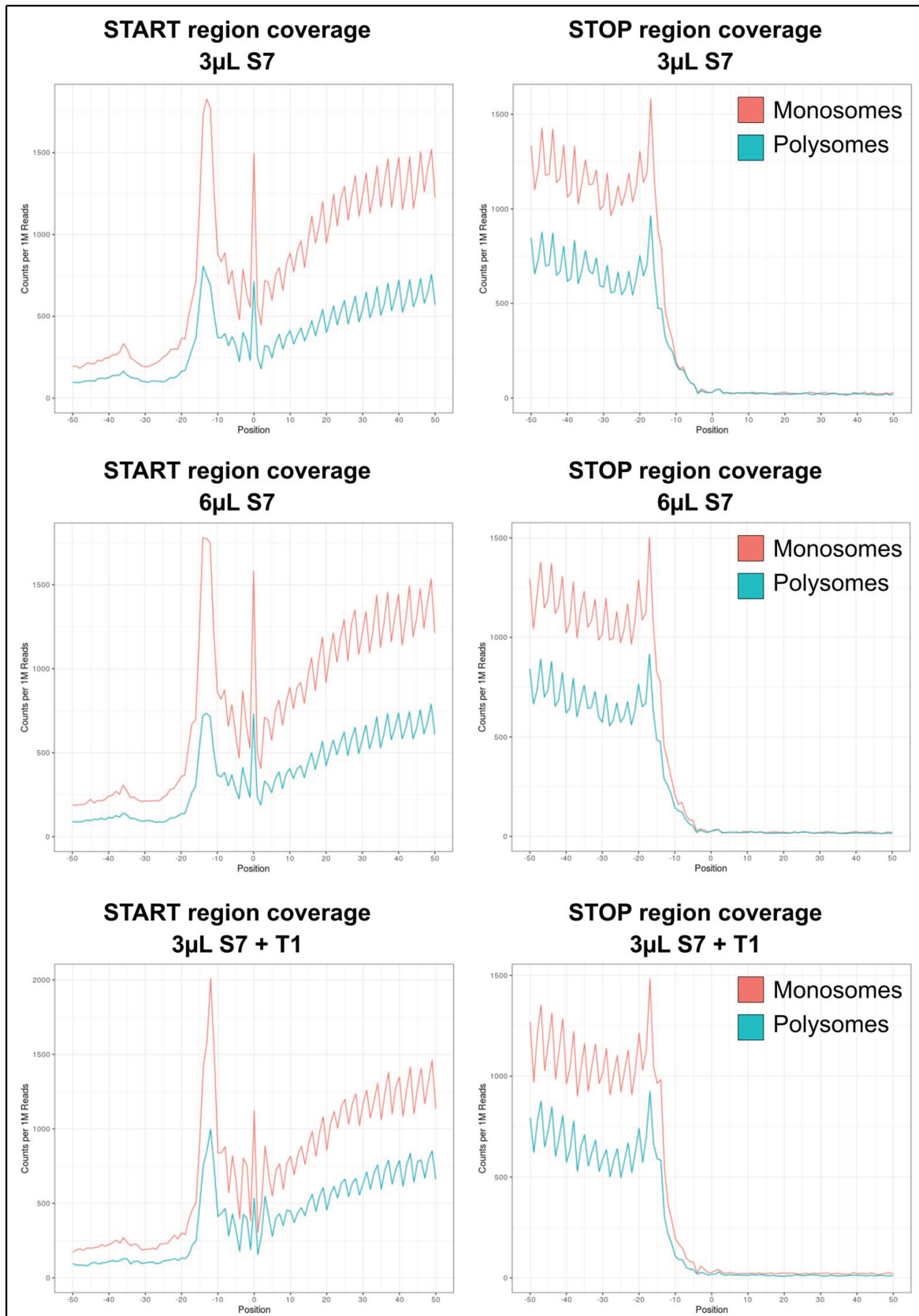
**Figure 28: Characterization of the different RNA sequences represented in the monosome and polysome footprinting libraries prepared using distinct RNase conditions.** Proportions of **A.** rRNA, **B.** tRNA, and **C.** mRNA reads relative to the total reads sequenced for each type of library. **D.** Comparison of the proportions of RPFs mapping to the 5'UTR, CDS or 3'UTR mRNA regions in the different monosome or polysome footprinting libraries using the RiboFlow pipeline.

To confirm that the RPFs mapped to the CDS region in the monosome footprinting libraries are really generated from actively elongating ribosomes, the RPFs size distribution in these samples was compared to the pattern obtained for polysomes samples. For all RNase conditions tested, the length distributions of RPFs mapped to the CDS between monosome and polysome libraries were highly similar (Figure 29.A). This result clearly proves that monosomes can also be actively engaged in translation and particularly in the elongation of newly synthesized polypeptide chains. Moreover, variations were visible at the transcript-level when individual FLOSS values were plotted depending on the number of CDS mapped reads. Notably, the FLOSS values obtained for monosome samples treated with S7 only were increased compared to those obtained for polysome samples (Figure 29.B). This can be explained by the reduced coverage of mRNA sequences in the S7 treated monosome libraries compared to the polysome ones. Consequently, the correlation between the monosome and polysome FLOSS scores was increased for the conditions where the mRNA coverages were more similar (Figure 29.C).



**Figure 29: RPFs derived from monosomes display similar features to polysomal RPFs. A.** Comparison of the frequency of each RPFs length in the CDS region of mRNAs associated to monosomes or polysomes. **B.** Fragment Length Organizing Similarity Score (FLOSS) analysis of the RPFs mapping in each mRNA region (5'UTR, CDS or 3'UTR) in monosome vs polysome libraries. **C.** Scatterplot comparison of the FLOSS scores calculated for monosome and polysome footprinting libraries. All the analyses were performed individually for each RNase condition tested.

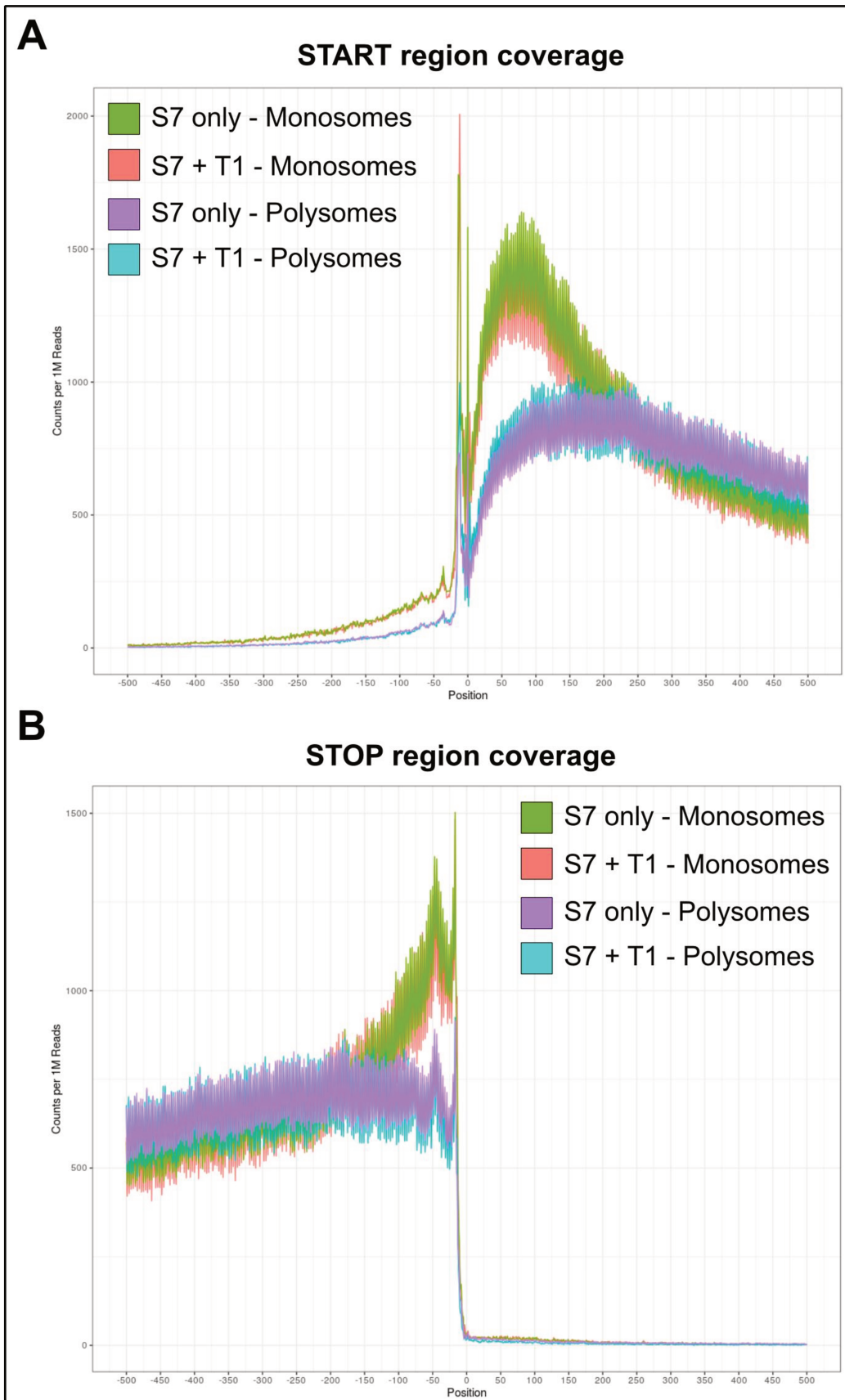
Finally, the 5' end read densities were aggregated across all transcripts to assess the global coverage rates around the translation start and stop sites using RiboFlow pipeline. A clear 3 nucleotide periodicity was observed in the CDS region but not in the UTRs in all our monosome and polysome footprinting libraries independently of the RNase condition used (Figure 30). This result thus supported the conclusion that most mRNA bound monosomes are actively elongating. Accordingly, a high peak, characteristics of AUG codon positioning the ribosomal P-site, is observed in the 5'UTR region around 13nt before the start site in both ribosomal populations. Moreover, a brutal drop of RPFs density after the stop codon was detected in both monosome and polysome footprinting libraries. Altogether, these observations confirmed that monosomes behave similarly to polysomes and also participate in the global cellular protein synthesis.



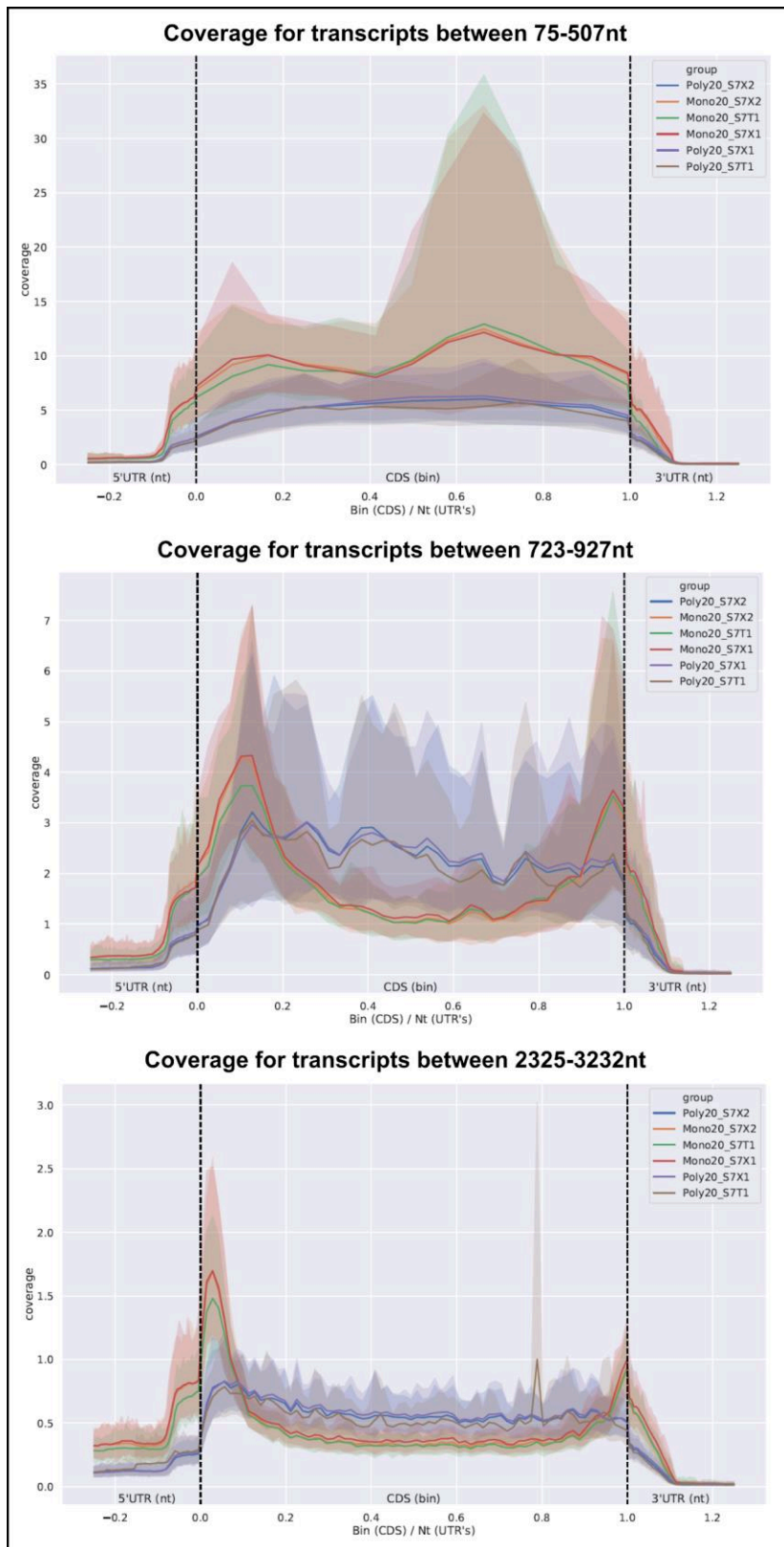
**Figure 30: Monosomes display a clear three nucleotides periodicity in the CDS region similarly to polysomes.** Coverage of the start and stop codon regions (50nt range before and after the start codon) across mRNAs in our different monosome vs polysome footprinting libraries prepared to compare the efficiency of distinct RNase treatments.

Intriguingly, the RPFs densities in the CDS region close to the start and stop codons were higher in the monosome libraries despite the higher CDS coverage observed earlier in polysome samples. As the coverage analysis was performed across the same transcripts for monosome and polysome footprinting libraries, it is unlikely that this pattern could be explained only by a differential association depending on the mRNAs size. In addition to this, transcripts of less than 500nt represented only 2% of all the transcripts used for this analysis (380 out of 21668 transcripts) suggesting that the pattern obtained mainly reflects ribosomal densities across transcripts higher than 500nt long. To assess if these reduced RPFs density in polysomes was also observed at a higher scale along the CDS, the coverage rates were plotted for a larger window around the translation start and stop sites (500nt instead of 50nt). The results revealed that monosomes display a higher ribosomal occupancy than polysomes within the first 200nt of the CDS (Figure 31.A). This can be explained by the fact that monosomes are involved in the pioneer round of translation and also tend to have a slower transit speed during early elongation rounds (Heyer and Moore, 2016). A higher ribosomal density was also observed in the monosome samples for the last 100-150 nucleotides before the stop codon (Figure 31.B). This increased RPFs density could be explained by ribosomes completing the last round of translation on otherwise polysome-associated mRNAs. We also generated metagene plots by aggregating the coverage data for all nucleotides across the CDS depending on the transcripts size (Figure 32). For transcripts lower than 500nt, monosomal densities are higher than polysomal densities all along the CDS, consistent with the preferential association of short mRNAs to monosomes. For transcripts of longer than 500nt, monosome occupancy is higher near the start and the stop codon regions similarly to what was observed previously. Moreover, as expected, polysome occupancy is increased along the CDS confirming that the monosome peaks observed in 5' and 3' most likely correspond to initiating and terminating ribosomes. The selective analysis of the transcripts displaying this increased peak near the stop codon in monosome libraries could give more information on the origin of this phenomenon. To conclude, our results were consistent with previous conclusions made on the different contributions of monosomes and polysomes to the global ribosomal footprint pattern in yeast (Heyer and Moore, 2016).





**Figure 31: Monosomes display a different coverage pattern in the 5' and 3' end of the CDS compared to polysomes. Coverage of the A. start and B. stop codon regions (500nt range before and after the start codon) across mRNAs in our different monosome vs polysome footprinting libraries prepared to compare the efficiency of distinct RNase treatments.**

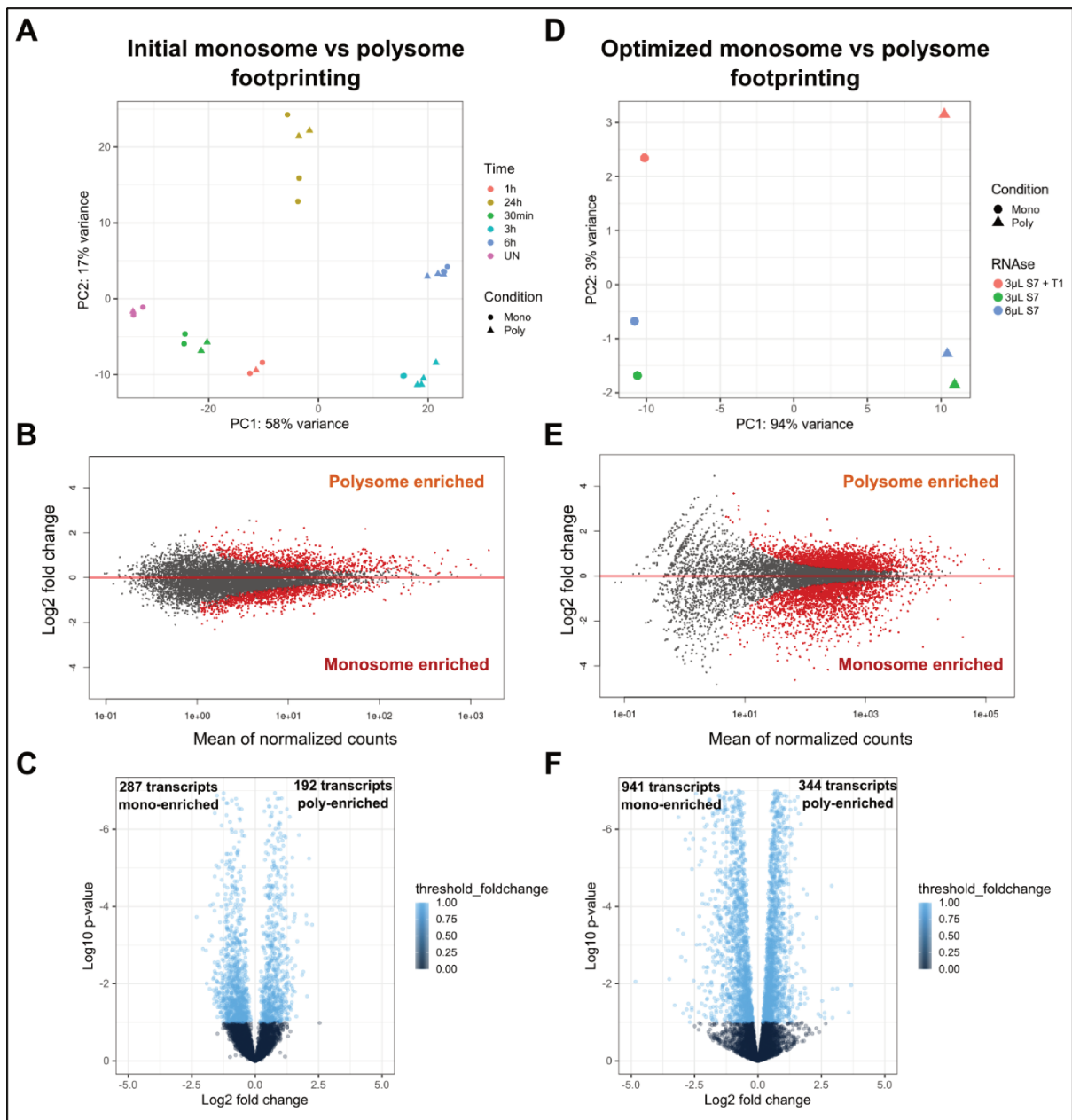


**Figure 32: Monosomes and polysomes display a different coverage pattern across the CDS depending on the transcript size range.** Normalized ribosomal densities across mRNAs in the CDS region for our different monosome vs polysome footprinting libraries prepared to compare the efficiency of distinct RNase treatments. The coverage data were calculated for several groups of transcripts divided in bins depending on their size. The last 50nt of the 5'UTRs and the first 50nt of the 3'UTRs were also included in the analysis.

### ***Characterization of monosome vs polysome enriched transcripts***

To identify transcripts that are preferentially associated to monosomes or polysomes, the number of reads mapping to each individual CDS was measured using the python package HTSeq counts in our different libraries (Anders et al., 2015). The reads counts obtained were then analyzed using the differential expression package DESeq2 (Love et al., 2014). This tool relies on a statistical approach optimized for the analysis of high-throughput RNA sequencing results. It can notably be used to visualize the dispersion between different sequencing samples by generating principal component analysis (PCA) plots. It can also calculate an enrichment score between two sets of samples, such as monosome or polysome libraries, while integrating differences due to other factors such as LPS treatment or variations between biological replicates. Consequently, the monosome versus polysome enrichment scores for each transcript were defined as the log<sub>2</sub> fold changes (Log<sub>2</sub>FC) calculated by DESeq2. The results obtained for our optimized monosome vs polysome footprinting libraries were analyzed alongside with the data obtained for our first attempt (Figure 33).

For our initial experiment, samples were collected at different times post LPS-stimulation. Interestingly, most of the variance (58%) between samples observed using the PCA analysis could be explained by a different stage of macrophage activation (Figure 33.A). Intriguingly, 17% of the variance could be explained by another factor that affected only samples from the late stages of the inflammatory response (6h and 24h). This could notably reflect variations of gene expression due to alterations of the epigenetic modifications pattern. The impact of the differential association between monosomes and polysomes was only visible if the PCA analysis was performed for each time point individually. This can be explained by the massive transcriptional reaction induced following LPS stimulation of the macrophages. After taking in account the transcription induced and experimental variations, 287 transcripts were significantly identified as monosome-enriched and 192 transcripts were polysome-enriched (absolute Log<sub>2</sub>FC value superior to 1 and adjusted p-value of 0.1 ; Figure 33.B and C).

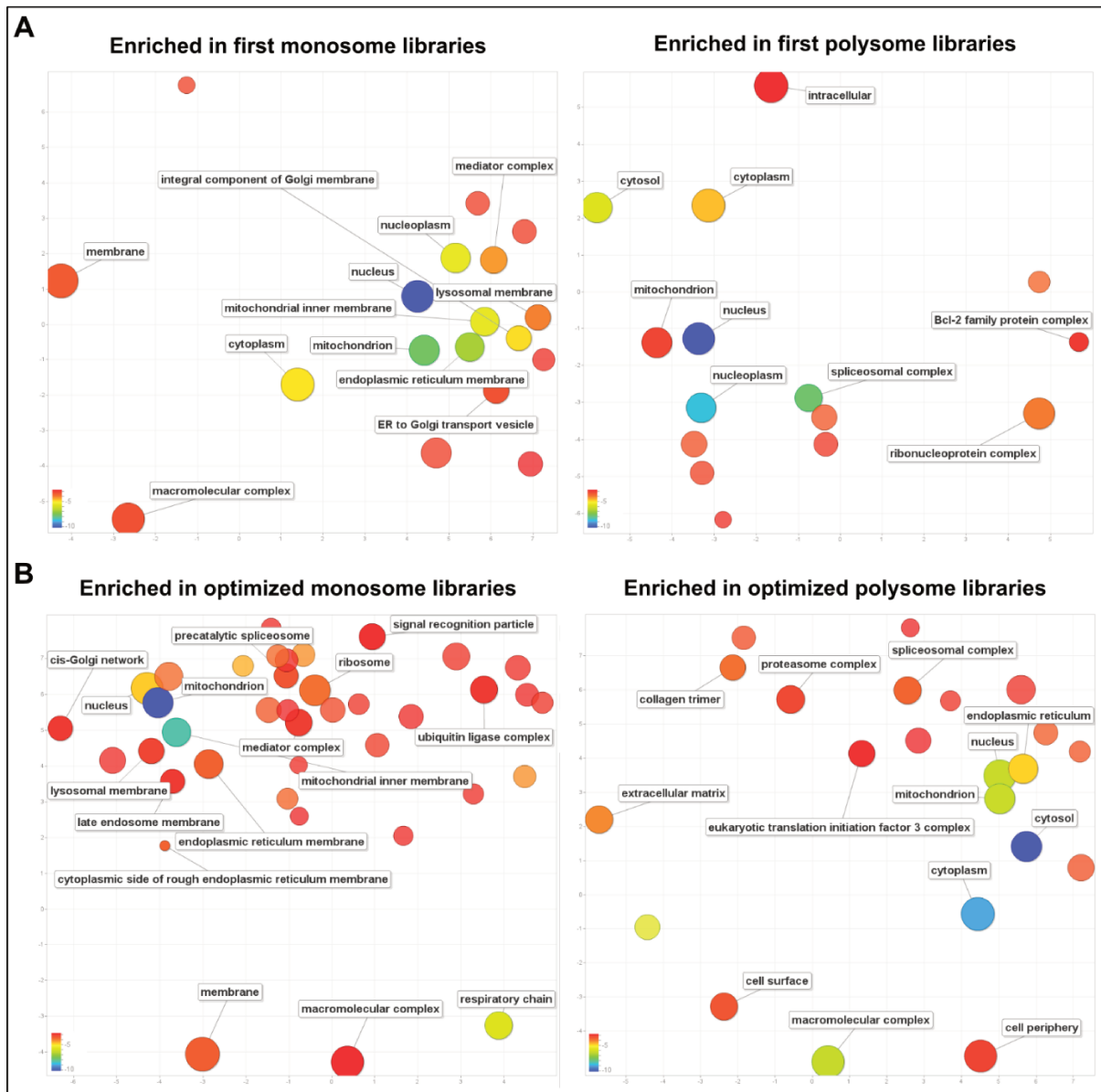


**Figure 33: Differential expression analysis to identify transcripts preferentially associated to monosomes or polysomes.** **A.** and **D.** PCA analysis displaying the two main components that explain the variance between samples in our initial or optimized libraries respectively. **B.** and **E.** MA plots computing the log<sub>2</sub> fold change enrichment score calculated using DESeq2 depending on the mean of normalized counts obtained for each transcript in our initial or optimized libraries respectively. Each transcript is represented as a dot. Grey dots correspond to transcripts not significantly enriched in monosomes or polysomes. Red dots correspond to transcripts preferentially associated to monosomes or polysomes. **C.** and **F.** Volcano plots computing the log<sub>2</sub> fold change enrichment score depending on the log<sub>10</sub> of the adjusted p-value calculated for each transcript in our initial or optimized libraries respectively. Only transcripts with an absolute value of log<sub>2</sub>FC > 1 and an adjusted p-value of 0.1 (threshold fold change in blue) are considered significantly enriched in the monosomes or polysomes subsets. The threshold fold change values are calculated by combining both log<sub>2</sub>FC and adjusted p-values. Each transcript is represented as a dot.

A similar analysis was performed using the samples prepared to test different RNases conditions from the same cytoplasmic lysate for monosome vs polysome footprinting. Despite the distinct sample treatments, most of the variance (94%) observed between samples could be explained by the preferential association to monosomes or polysomes (Figure 33.D). Given that the differences between all the RNase conditions tested were minimal compared to the effect of monosome or polysome purification, the data obtained were pooled together to calculate the monosome vs polysome ratios using DESeq2. The enrichment analysis revealed that 941 transcripts were significantly monosome-enriched and 344 transcripts were polysome-enriched (absolute Log<sub>2</sub>FC value superior to 1 and adjusted p-value of 0.1 ; Figure 33.E and F). The increased number of significantly enriched transcripts in both conditions compared to our first attempt can be explained by the higher mRNA coverage as more reads were assigned to each transcript individually (Figure 33.E).

To see if monosome or polysome enriched transcripts code for proteins displaying different localizations in the cell, a gene ontology (GO) analysis focusing on cellular components was performed using GeneCodis 4.0 (Tabas-Madrid et al., 2012). This analysis informs us on the localization of proteins in distinct cellular compartments, such as nucleus or mitochondria, or in stable macromolecular complexes, such as the ribosomes. As several GO subsets can be redundant and lists of GO can be difficult to visualize, the clustering algorithm REVIGO was used to summarize the results on semantic similarity-based scatterplots (Supek et al., 2011).

As the number of significantly differentially enriched transcripts was low in our initial monosome vs polysome footprinting libraries, fewer different GO subsets were identified in these samples compared to the optimized libraries (Figure 34). Moreover, the differences in the localizations of preferentially monosome or polysome translated proteins were not clear as the top GO clusters found in both conditions were “nucleus” and “cytoplasm” (Figure 34.A). Interestingly, monosomes translated proteins were highly associated with “membrane” as previously observed in yeast (Heyer and Moore, 2016). This result was further confirmed in our optimized libraries as 385 out of 941 monosome-enriched transcripts corresponded to membrane associated proteins. Consequently, many GO clusters corresponding to the membranes of diverse subcellular compartments were identified in our set of monosome-enriched transcripts (Figure 34.B). This can be explained by the model of membrane-associated protein import through the ER membrane during which the signal sequence is first translated by a single cytoplasmic ribosome prior to signal recognition particle (SRP) recruitment and membrane engagement. The two other main GO clusters identified in our monosome-enriched set were “nucleus” and “mitochondria”.

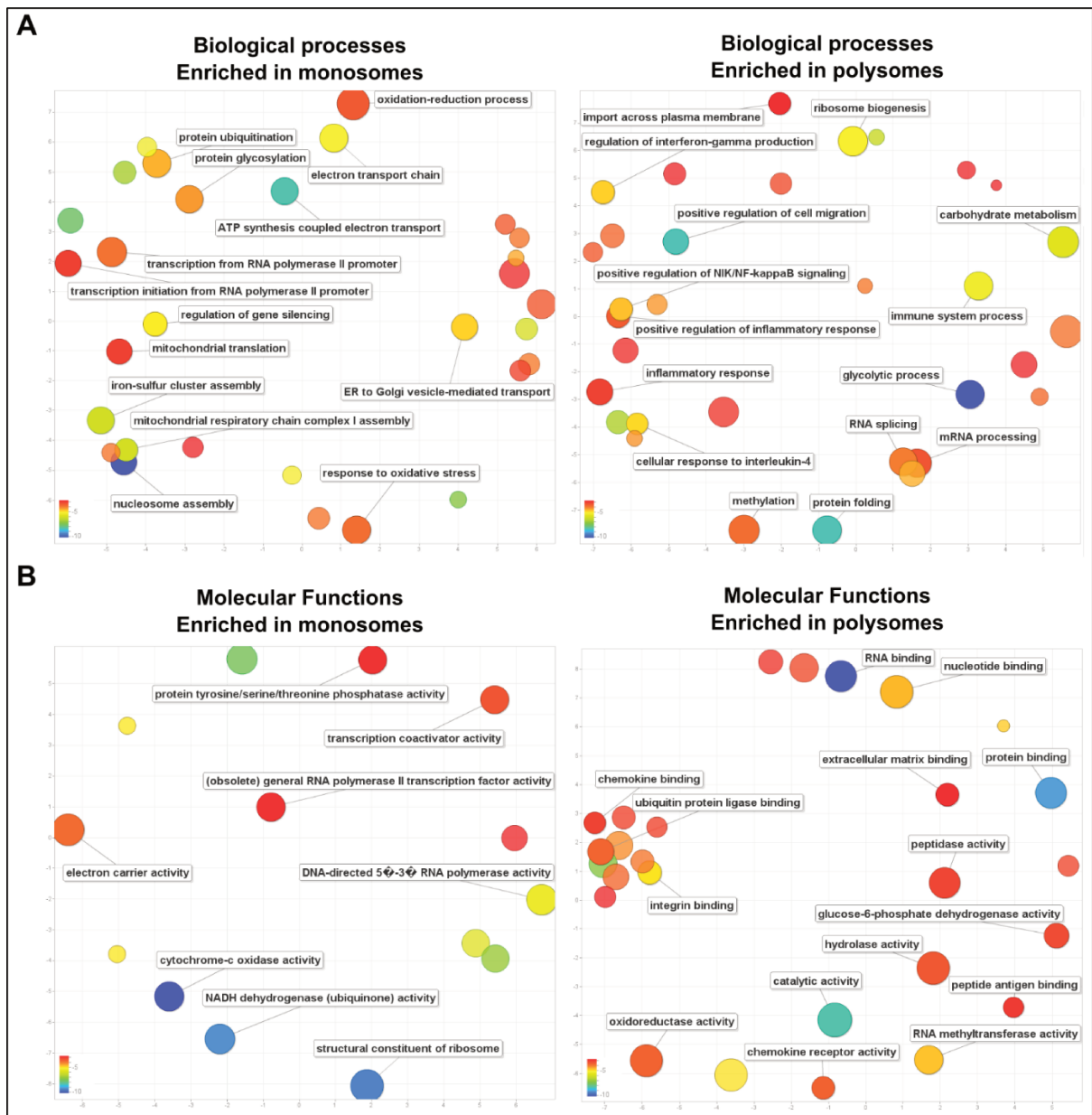


**Figure 34: Cellular component GO analysis reveals differences in the subcellular localization of proteins encoded by monosome or polysome enriched transcripts. A.** GO analysis for monosome or polysome enriched transcripts identified in our initial monosome vs polysome libraries. **B.** GO analysis for monosome or polysome enriched transcripts identified in our optimized monosome vs polysome libraries. The most significant GO clusters (adjusted p-values <0.05) are represented in scatterplots with x and y axes that optimize the closeness between the most similar GO clusters based on semantic similarity. The size of the circles represents the number of transcripts identified per GO cluster. The color of the circles corresponds to a gradient scale of the log<sub>10</sub> adjusted p-value represented in the left bottom corner of each graph.

Contrary to what was observed using our initial monosome vs polysome footprinting data, a clear difference was visible in the top cellular component GO clusters identified in the polysome-enriched set compared to the monosome one. Most polysome-enriched transcripts corresponded to proteins localized in the “cytoplasm” or “cytosol” and in the “nucleus” to a lesser extent. Undoubtedly, the polysomes fragmentation provoked by endogenous RNases seriously perturbed the quality of the data obtained for our first monosome vs polysome footprinting experiment. While some trends were visible using these results, the real physiological differences between monosomes and polysomes associated transcripts were masked in these datasets. For this reason, only the data obtained using our optimized protocol were used for further functional analysis of the proteins preferentially translated by monosomes or polysomes.

Following the analysis of the cellular localizations corresponding to proteins preferentially produced by monosomes and polysomes, I focused on the biological programs, such as immune response or signal transduction, in which they could be involved. Clear differences in GO were again visible between the two subsets (Figure 35.A). Interestingly, monosome-associated transcripts were enriched in basal biological functions, such as mitochondria metabolism or protein modifications by ubiquitination or glycosylation, while immunity related functions were mainly associated with polysomes. Moreover, proteins involved in the processing and splicing of mRNAs were enriched in the polysomes subsets suggesting that this ribosomal population plays the biggest role during the inflammatory response. Importantly, transcripts corresponding to transcriptional regulators were preferentially associated to monosomes in basal conditions. This could suggest that translation through monosomes participates in the regulation of the expression of proteins involved in the triggering of the inflammatory response. Once triggered, factors involved in the next stages of inflammation could then be mostly synthesized in polysomes. A differential association of transcripts encoding transcription factors between monosomes and polysomes was also detected during the different time points of the inflammatory response in our initial libraries. However, as the ribosomal binding patterns were biased due to the endogenous RNases, this experiment should be repeated using the optimized protocol.



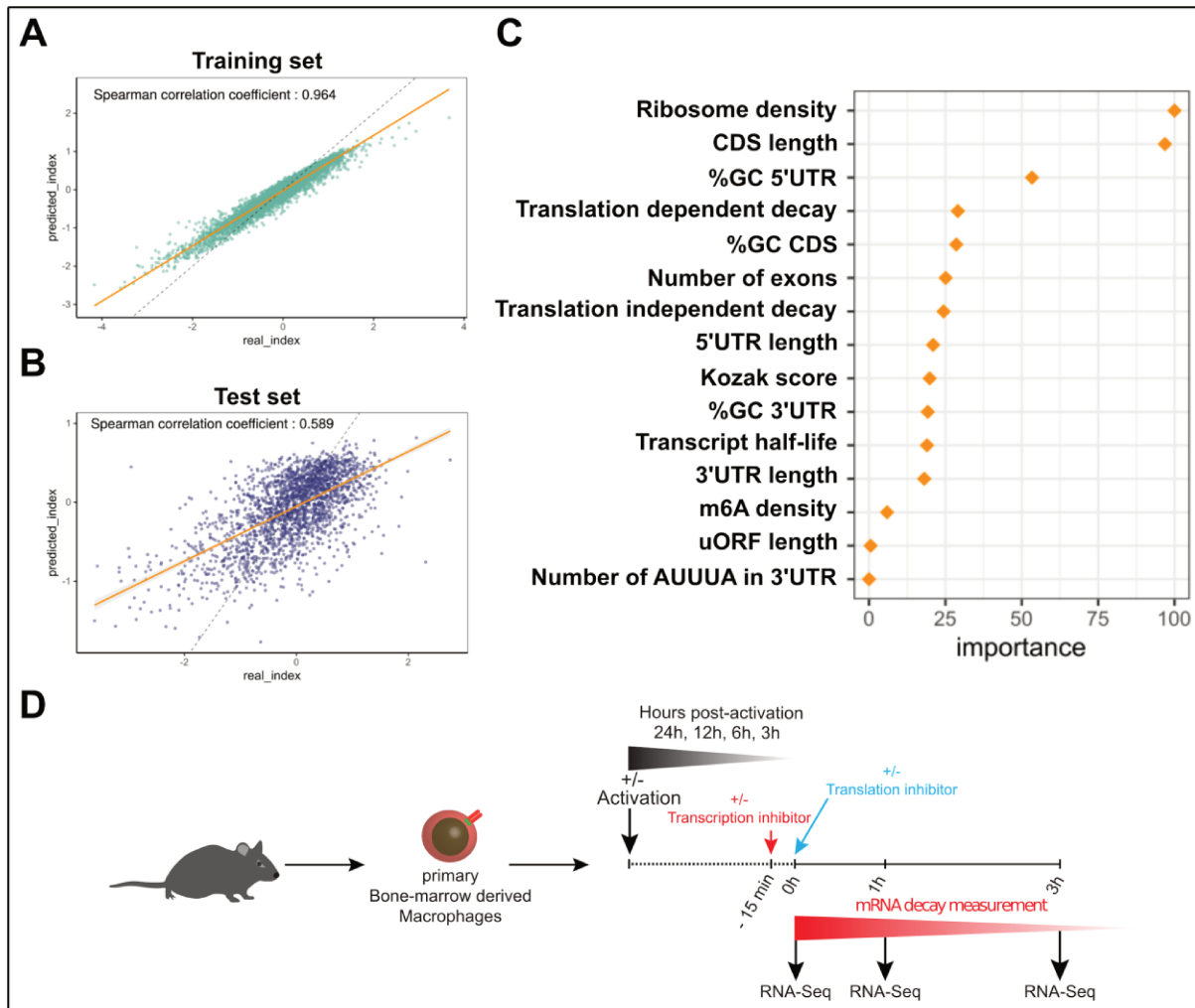


**Figure 35: GO analysis reveals functional differences between proteins encoded by monosome or polysome enriched transcripts.** **A.** Biological process GO analysis for monosome or polysome enriched transcripts identified in our optimized monosome vs polysome libraries. **B.** Molecular functions GO analysis for monosome or polysome enriched transcripts identified in our optimized monosome vs polysome libraries. The most significant GO clusters (adjusted p-values <0.05) are represented in scatterplots with x and y axes that optimize the closeness between the most similar GO clusters based on semantic similarity. The size of the circles represents the number of transcripts identified per GO cluster. The color of the circles corresponds to a gradient scale of the log<sub>10</sub> adjusted p-value represented in the left bottom corner of each graph.



Finally, I performed another GO analysis to characterize the molecular functions associated with monosomes or polysomes enriched transcripts (Figure 35.B). Concordantly with what was previously observed in yeast, monosomes translated proteins are enriched in “transcription coactivator activity” and polysomes are enriched in functions corresponding to highly expressed factors such as “chemokine binding” or “peptide antigen binding” and “extracellular matrix binding”. Many differences were thus visible in our dataset validating our optimized approach for the study of monosome vs polysome association during the inflammatory response in pBMDMs.

To identify specific mRNA features that could explain their preferential association to monosomes or polysomes, a random forest regression approach was implemented using selected parameters that could predict the monosome vs polysome enrichment score previously calculated using DESeq2. The parameters used to train the model were selected based on the literature regarding the most important mRNA encoded features that could impact translation efficiency. To construct the final regression model, multiple regressions are performed by a machine learning algorithm using 30% of the input transcripts randomly selected. A prediction score is then calculated based on the results of the different regressions and used to predict informatically the monosome vs polysome enrichment score. The regression model is then validated by comparing the predicted enrichment scores to the real values. When the predicted enrichment scores are compared to the real values used to train the model (30% of the input transcripts), the correlation coefficient is high as expected (correlation coefficient of 0.964 ; Figure 36.A). When the predicted enrichment scores are compared to the real values corresponding to the other 70% of the input transcripts, the correlation coefficient is 0.589 (Figure 36.B). This suggests that the model constructed using the selected parameters can partly explain the real monosome vs polysome enrichment scores but additional parameters would be required to improve the prediction.



**Figure 36: Random forest regression reveals specific mRNA features that could explain preferential association with monosomes or polysomes.** **A.** Scatterplot comparing predicted monosome vs polysome enrichment scores to real scores in the training set (30% of the transcripts randomly selected). **B.** Scatterplot comparing predicted monosome vs polysome enrichment scores to real scores in the test set (70% of the transcripts not used in the training set). **C.** Hierarchical clustering of the different parameters used to construct the random forest regression model based on their relative importance for the prediction of monosome vs polysome enrichment scores. **D.** Monitoring of translation-dependent and independent mRNA decay in pBMDMs. Activated or untreated pBMDMs are incubated or not with transcription inhibitors (Triptolide or DRB). 15 minutes following addition of transcription inhibitors, translation inhibitors (cycloheximide or harringtonine) are added to the cells. Cells are collected at 0, 1 and 3 hours following transcription inhibition to monitor transcript expression by RNA-sequencing.

The different parameters used to construct the regression model can be then hierarchized depending on their relative importance in the final prediction model (Figure 36.C). It should be noted that here, the percentages of relative importance reported for each parameter are not correlated to their absolute importance to explain the real enrichment scores. Among the various parameters tested, the ribosome density is unsurprisingly the top parameter to explain the monosome vs polysome enrichment scores. This observation thus validates our strategy to perform monosome vs polysome footprinting and calculate an enrichment score. As previously observed in yeast, the CDS length is a key parameter to explain the preferential association to monosomes or polysomes (Arava et al., 2003; Heyer and Moore, 2016). Interestingly, several parameters involved in translation initiation efficiency, including the GC content and length of the 5'UTR or the Kozak score, are also important to explain the monosome vs polysome ratios. On the other hand, the uORF length did not have a big effect on the enrichment score. Moreover, parameters concerning mRNA stability, like the transcript half-life or the translation dependent or independent decay rates, are also associated with monosome vs polysome differences. The translation dependent and independent decay rates were calculated in our lab using RNA-seq in pBMDMs treated with transcription and/or translation inhibitors at different times to identify unstable transcripts whose stability was affected or not by the association to the translation machinery (Figure 36.D). The results obtained are thus particularly interesting because they support the previous observations made in yeast regarding the different parameters that could explain monosome vs polysome enrichments (Heyer and Moore, 2016).

### ***In vitro translation to decipher monosomes translational status***

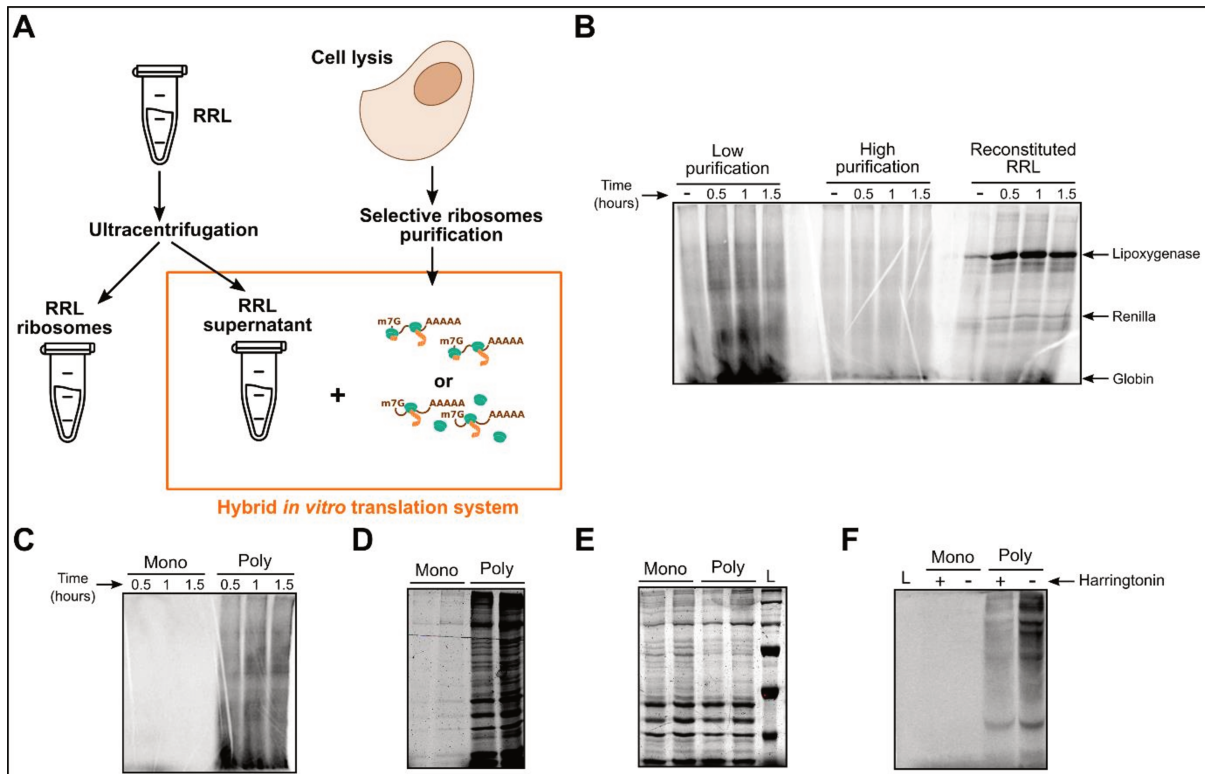
Using the monosome vs polysome footprinting technique, we were able to visualize the ribosomes movement across transcripts at a high resolution and conclude about monosomes translational activity. Notably, the 3nt periodicity visible only in the CDS region and the RPF size distribution very similar to the polysomes one are strong evidences for the existence of actively elongating monosomes. Despite of this, we cannot definitively conclude about their ability to synthesize full proteins. Indeed, the 3nt periodicity was also associated with the co-translational degradation of aberrant mRNA in yeast and other organisms (Ibrahim et al., 2018; Pelechano et al., 2015; Tuck et al., 2020). In the last results chapter of this work, I will present experiments that were designed to detect the level of protein neo-synthesis associated to monosomes using radioactive labelling. Indeed, this highly sensitive technique was historically used to characterize *de novo* protein synthesis from polysomes (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). To date, this is still the best approach to detect very low amount of newly synthesized proteins.

To assess if monosomes translation can produce proteins, I used an hybrid *in vitro* translation system to measure the levels of protein synthesis from endogenous cellular mRNAs (Figure 37.A). In this system, ribosomes purified from the cell type of interest are mixed with rabbit reticulocyte lysate (RRL) supernatant, previously depleted of rabbit ribosomes by ultracentrifugation, to provide all the components required for *in vitro* translation in presence of radioactive amino acids such as S<sup>35</sup> labelled methionine (Panthu et al., 2015). The purified ribosomes can be freezed on their associated mRNAs and subjected to sucrose gradient ultracentrifugation to separate monosomes and polysomes. After isolation of monosomes and polysomes, ribosomes are concentrated by pelleting through a sucrose cushion. The ribosomal populations obtained can be further purified by increasing salt concentrations in the sucrose cushion (Penzo et al., 2015). However, highly purified ribosomes are depleted of all non-tightly bound ribosomal proteins and could thus be less translationally active.

To select the best protocol to measure endogenous mRNAs translation from murine macrophages derived ribosomes, I first prepared global ribosomes samples from iBMDMs using high or low amounts of salts in the sucrose cushion (low or high purification grade). The translational activities of these two ribosomal preparations were compared to a reconstituted sample where the RRL supernatant was mixed with the rabbit ribosomes previously removed (Figure 37.B). Several translation reactions were prepared and incubated for different times at 30°C to check if the amounts of radioactive proteins produced were well increased through

time (incubation for 30m, 1h and 1h30). Moreover, a very small amount (1 fmol) of renilla luciferase mRNA was added to see if an exogenous mRNA could be translated as well. The negative control reactions were incubated for 30 min at room temperature instead of 30°C. The results obtained for the reconstituted sample demonstrated that the hybrid *in vitro* translation reactions were properly set up as both endogenous (lipoxygenase and globin) and exogenous (renilla luciferase) mRNAs were translated. In the highly purified ribosomes samples, almost no protein synthesis occurred as the signals obtained for the reactions incubated at 30°C were very similar to the negative control. This confirmed that high salt conditions can remove translation factors and block protein synthesis. On the contrary, a clear increase of radioactivity was observed in all reactions using low salt purified ribosomes incubated at 30°C compared to the negative control. After 1h of incubation, no further increase of the newly synthesized proteins was observed probably due to the exhaustion of the energy supplies (Panthu et al., 2018). As I was not aware at that time that high salt conditions could better preserve ribosomal integrity by inhibiting endogenous RNases digestion, I concluded that low salt concentrations was the best option to prepare macrophages ribosomes for *in vitro* translation.

I next prepared monosomes and polysomes purified for iBMDMs using a low salt cushion and set up several reactions incubated for different times at 30°C with RRL supernatant (Figure 37.C). The results obtained for the polysomes were similar to what was observed previously for global ribosomes samples. Conversely, no signal was detected in all the reactions prepared using purified monosomes. As I noticed that the monosomes amounts were lower than polysomes in iBMDMs sucrose gradient profiles, I compared the protein levels in two monosomes and polysomes samples prepared similarly as for *in vitro* translation using SDS-PAGE gel migration. The gel was prepared with 2,2,2-trichloroethanol (TCE) to detect all the proteins containing tryptophans very easily after UV exposure (Ladner et al., 2004). The results revealed that the monosomes amounts were indeed very low compared to the polysomes (Figure 37.D). I thus optimized the purification protocol to prepare monosomes and polysomes ready for *in vitro* translation in comparable amounts (Figure 37.E). Despite this, no signal was detected in all translation reactions prepared using purified iBMDMs monosomes (Figure 37.F). Interestingly, the amounts of newly synthesized proteins produced by iBMDMs polysomes were decreased if harringtonin was added to the translation reaction. As all new initiation events were inhibited in this condition, we were able to specifically visualize the protein synthesis ongoing from endogenous mRNAs in the cell at the time of lysis. Moreover, the increased radioactivity labelling observed in absence of harringtonin could be explained by high reinitiation or recycling rates in the polysomes.



**Figure 37. *In vitro* translation assays to assess monosomes translational activity.** **A.** Schematic representation of the hybrid *in vitro* translation system. **B.** *In vitro* translation products obtained following the incubation at 30°C of low salt (low purification) or high salt (high purification) purified iBMDM ribosomes with RRL supernatant were run on a SDS-PAGE gel. The corresponding levels of radioactivity were then measured by autoradiography (overnight exposition). A control reaction was performed by mixing RRL ribosomes with RRL supernatant. Samples were collected at different times of the translation reaction and a negative control reaction was left at room temperature for each combination. **C.** *In vitro* translation products obtained following the incubation at 30°C of purified iBMDM monosomes or polysomes with RRL supernatant were run on a SDS-PAGE gel. The corresponding levels of radioactivity were then measured by autoradiography (overnight exposition). **D.** TCE labelling of all the proteins present in the iBMDMs monosomes and polysomes samples used in the translation reaction C after SDS-PAGE migration. Each ribosomal preparation was loaded twice. **E.** TCE labelling of all the proteins present in the iBMDMs monosomes and polysomes preparations used in the translation reaction F after SDS-PAGE migration. Each ribosomal preparation was loaded twice. **F.** Detection of the *in vitro* translation products obtained following the incubation of equal amounts of purified iBMDM monosomes or polysomes with RRL supernatant by autoradiography after SDS-PAGE migration.

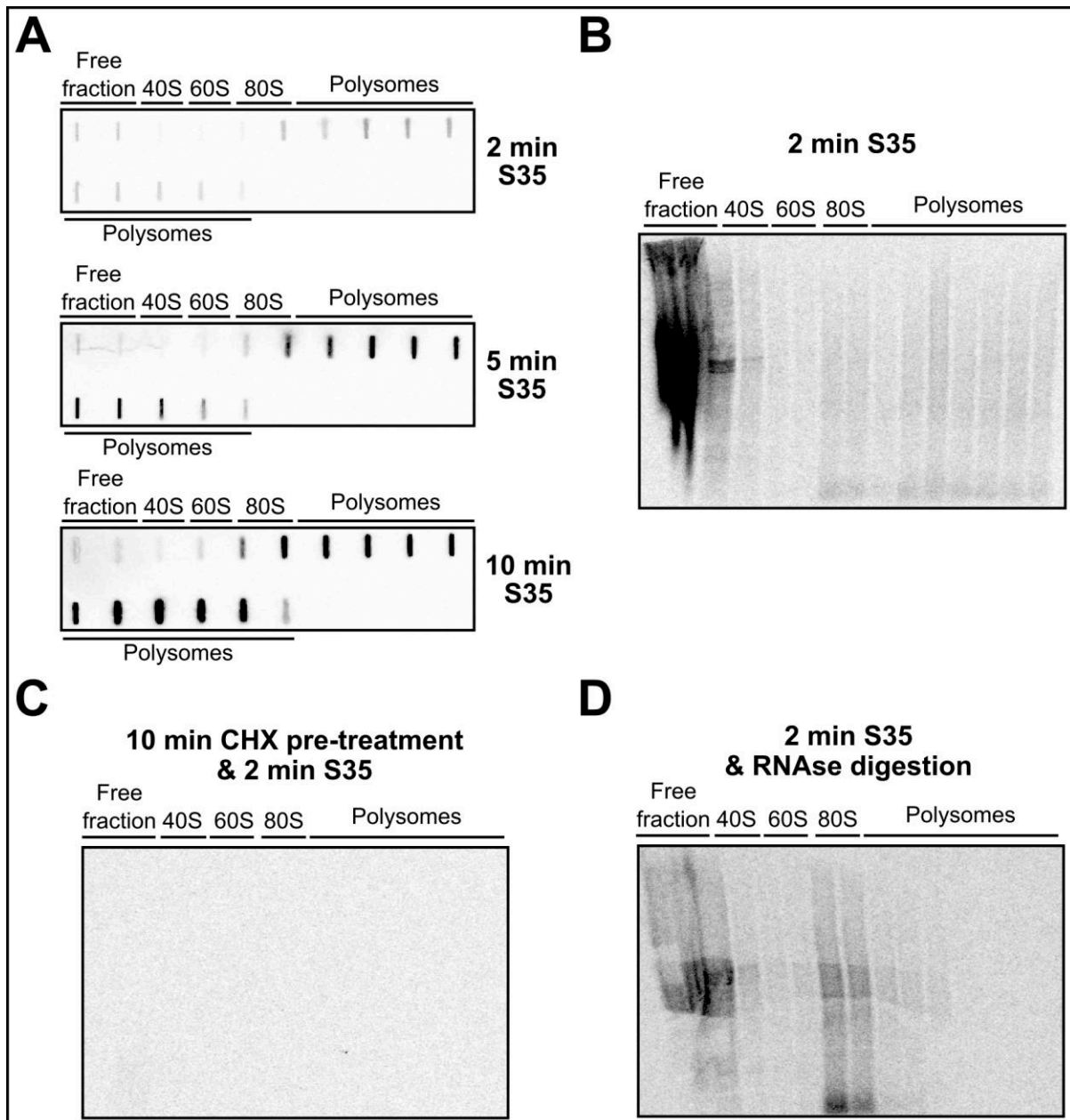
As we could not detect monosomes translation products using the hybrid *in vitro* translation system with iBMDM lysates, I wondered if this failure could be partially explained by the very long sample preparation procedure. Indeed, as monosomes were previously described as preferentially associated with short mRNAs, it could be assumed that their translation speed would be faster than for polysomes. As the sample preparation required purification from the fractions of a sucrose gradient and then overnight pelleting through a sucrose cushion, I was concerned that most translation events occurring in monosomes were completed by the time that the ribosomes were used in the translation reaction. To circumvent this issue, I tested another strategy consisting in a short labelling pulse with radioactive amino acids right before cell lysis. Radioactive methionine and cysteine can diffuse very rapidly through the cellular membrane and are incorporated in newly synthesized peptides in less than two minutes after their addition directly in the culture medium. After lysis, the cytoplasmic lysates were rapidly loaded a sucrose gradient and the fractions collected precipitated with trichloroacetic acid (TCA) to remove sucrose and free radioactive amino acids remained unbound. The resulting proteins can be spotted on a nitrocellulose membrane or loaded on a SDS-PAGE and the levels of radioactivity incorporated can be measured by autoradiography.

Before using this strategy to characterize the translational status of the different ribosomal populations in pBMDMs, several incubation timings were tested. Indeed, if the exposition timing is too long, radioactive proteins can be released in the cytoplasm as regular proteins and relocalized to their natural cellular compartment. This can become problematic for ribosomal proteins as they can be recruited to the ribosomes and induce the detection of radioactivity in the monosomes and polysomes fractions independently of *de novo* protein synthesis events. For this reason, when pBMDMs were incubated for ten minutes with S35 labelled amino acids, all the fractions obtained after sucrose gradient ultracentrifugation, including the 40S and 60S, displayed a good radioactivity signal (Figure 38.A). The signal was higher in all polysomes fractions and could also be detected in the monosomes fractions. When incubated for five minutes, the global labelling efficiency was reduced and less signal was observed in the fractions corresponding to the 40S and 60S subunits. A clear signal was still detected in the monosomes and the polysomes fractions. The results were even cleaner when pBMDMs were labelled only for two minutes. Consequently, the incubation timings with radioactive amino acids were then limited to two minutes in order to limit the production of radioactive ribosomal proteins that could be recruited to the ribosomes. When proteins precipitated from the sucrose gradient fractions prepared accordingly were loaded on a SDS-PAGE gel, a slight increase of the radioactivity signal associated to the monosomes fractions was detected (Figure 38.B). I confirmed that this signal was really due to S35 amino acids incorporation into newly synthesized proteins as no radioactivity was detected when the

pBMDMs were pre-incubated with cycloheximide for 10 min before labelling (Figure 38.C). Moreover, upon RNase digestion, most of the radioactivity signal was shifted from the polysomes fractions to the monosomes fractions (Figure 38.D).

While these results were encouraging, they are insufficient to ascertain the monosomes capacity to synthesize proteins as all these experiments were performed before optimizations of the protocol to limit endogenous RNases activity. Consequently, the signal associated with the monosomes fractions could very likely be caused by the degradation of polysomes into monosomes. A similar assay should be performed using heparin and 1M KCl treatment after cell lysis to be able to conclude definitively about monosomes potential to perform *de novo* protein synthesis.





**Figure 38: Pulse labelling with radioactive amino acids to detect monosomes translational activity.** **A.** Cytoplasmic lysates were prepared from pBMDMs incubated with S<sup>35</sup> methionine and cysteine for short times (2, 5 and 10min). After sucrose gradient separation, each fraction was spotted on a nitrocellulose membrane and autoradiography levels were measured after drying of the membrane (one week exposition). **B.** Cytoplasmic lysate prepared from pBMDMs pulse labelled with S35 methionine and cysteine for 2min was loaded on a sucrose gradient. After ultracentrifugation, the proteins from each fraction were TCA precipitated to remove the sucrose and non-incorporated amino acids. The resulting proteins were then separated through SDS-PAGE migration and autoradiography levels were measured (three days exposition). **C.** Same as in B. except the pBMDMs were pre-treated with 100mg/mL of cycloheximide before pulse labelling. **D.** Same as in B. except the cytoplasmic lysate was treated with RNase A and T1 before loading on the sucrose gradient.

# Discussion

## ***Adaptation of the monosome vs polysome footprinting protocol from yeast to macrophages***

Deep sequencing has revolutionized the study of gene expression regulation by allowing the detection and quantification of any species of RNA at the whole transcriptome level. Since the 2000s, many high-throughput sequencing based methods have been developed to characterize in detail fundamental cellular processes including mRNA translation through the ribosome profiling protocol. A variation of the classical ribosome profiling protocol was recently used to demonstrate that, contrary to what was historically thought, monosomes are translationally active in yeast (Heyer and Moore, 2016). A follow-up study performed using rodent neurons also confirmed this observation in mammalian cells (Biever et al., 2020). Before implementing monosome vs polysome footprinting to pBMDMs, I had to optimize the protocol to capture the ribosomal binding pattern from these different ribosomal populations as faithfully as possible using a particularly sensitive cell type. Indeed, the initial protocol was developed using yeast, the same model organism as for the first ribosome profiling experiments. For this reason, many technical optimizations were already included notably regarding the cell lysis and the RNase digestion steps. These steps required thorough optimizations to be adapted to mouse macrophages.

In the beginning of my work, the procedure to properly freeze the ribosomes at their exact positions on mRNAs was debated (Hussmann et al., 2015; Weinberg et al., 2016). Notably, the common pre-treatment with the elongation inhibitor cycloheximide was reported to induce ribosomal footprinting distortions as the translation blockade occurred at various speed for different decoded codons (Gerashchenko and Gladyshev, 2014; MGlinicy and Ingolia, 2017). This can be explained by the fact that cycloheximide needs to interact with the ribosome directly within the E site to effectively block translation (Sharma et al., 2019b). As this site is not accessible during all stages of translation elongation, cycloheximide mostly freezes ribosomes at the pre-translocation stage (Wu et al., 2019). Hence, for the ribosomes involved in the other steps of elongation, ribosomal movement across the mRNA is not prevented until the next elongation cycle (Hussmann et al., 2015). During my PhD, I was able to demonstrate that a rapid drop temperature by placing the cells directly on ice combined with the addition of cycloheximide directly in the lysis buffer was an efficient method to freeze pBMDMs ribosomes rapidly and durably during the footprinting experiment. Consistent with the fact that translation blockade was not biased by cycloheximide in our ribosome profiling

samples, we were able to visualize RPFs produced by different conformations of the ribosomes before and after peptide bond formation (Lareau et al., 2014; Wu et al., 2019)

Another big challenge introduced by the use of pBMDMs was the presence of endogenous RNases that strongly corrupt the pattern of monosomes vs polysomes within a few minutes after cell lysis. The presence of these RNases was not suspected at first as no RNase contamination leading to the polysomes degradation was detected using iBMDMs lysates. Consequently, we did not expect a different behavior in pBMDMs lysates. This experience highlights the importance of testing the sample preparation protocol specifically in the system used afterwards. Particularly, it was known that the immortalization process could modify the expression of some proteins and macrophage phenotype compared to primary cells (Troupin et al., 2013). Besides, the presence of endogenous RNases is not so surprising when taking into account regular physiological functions of macrophages in the clearance of microbes or dying cells through phagocytosis (Gordon and Plüddemann, 2019; Murray and Wynn, 2011). Interestingly, the degradation rates were not similar in activated pBMDMs collected at different times post LPS-stimulation despite following the similar sample preparation procedure. This could indicate that the expression or the activity of the endogenous RNases involved can vary depending on the macrophage state of activation. Indeed, it was shown that the expression of proteases and other degradation factors were increased following macrophage activation (Na et al., 2018; Virág et al., 2019).

The origin of these endogenous RNases is also an interesting question as they could be derived from several sources. The release of these endogenous RNases could originate from the concomitant lysis of the phagolysosomes during cytoplasmic lysate preparation. Alternatively, these RNases could be naturally found free in the cytoplasm or ribosomes bound. This hypothesis is supported by the fact that I observed RNA degradation occurring when ribosomes were partially separated on a sucrose gradient then left for 1h at 4°C before further ultracentrifugation in low salt samples but not in high salt samples. Moreover, co-translational degradation of pro-inflammatory transcripts in pBMDMs was previously described in the literature through Zfp36 binding and recruitment of mRNA decay pathways (Zhang et al., 2017). To characterize the origin of these RNases, it could be interesting to perform a subcellular fractionation to separate the lysosomes and phagolysosomes from cytosol and endoplasmic reticulum (ER) associated ribosomes and check the level of RNase induced degradation. It could also be interesting to selectively purify the ribosomes by immunoprecipitation and then check their association with known RNases by mass spectrometry.

In this work, I further confirmed the importance of the RNase treatment to calibrate RPFs quality. Importantly, the optimization of the RNase digestion step is a prerequisite to obtain clean ribosomal footprinting results (Gerashchenko and Gladyshev, 2017; Liu et al., 2019). Studies using pre-established RNases concentrations selected using a different experimental model and a sucrose cushion instead of a sucrose gradient purification could suffer from heavy technical biases. The thorough optimization of the sample preparation procedure is not optional and should be integrated in any new ribosome profiling experiment to improve the reproducibility in the field. Clearly, RNase digestion efficiency varies depending on the species, the amount of input material, the concentration and the origin of the RNase (Gerashchenko and Gladyshev, 2017; Liu et al., 2019). In our study, we introduced an additional parameter to take in account for the optimization of the RNase digestion by adding heparin in the lysis buffer. Consequently, while our ribosome profiling protocol is well validated for pBMDMs, the results of monosome vs polysome footprinting could benefit from further optimization of the RNases conditions. Notably, the addition of RNase T1 greatly improved the digestion efficiency but concomitantly increased the amounts of rRNA contamination and potentially caused a loss of mRNAs derived RPFs. Using only the S7 nuclease, the amount of RPFs mapping to mRNAs sequences was greater but the RPFs size distribution was not as good as what was observed for ribosome profiling using similar S7 nuclease amounts. A solution could thus be to use a slightly increased concentration of RNase S7 alone to obtain a better resolution in monosome vs polysome footprinting samples prepared from pBMDMs. Additionally, further testing would be required before implementing this approach to another cell type such as human macrophages.

Ribosome profiling and monosome vs polysome footprinting produce high-resolution maps of ribosome positions on mRNAs. A higher peak of RPFs on a codon can be interpreted as a site of pause for the ribosomes or could just be an artifact. In fact, the experimental reproducibility of the RPFs footprinting profiles for individual transcripts is low (Valleriani and Chiarugi, 2020). This is due to a low signal-to-noise ratio between true peaks signalling an increase of RPFs and global footprinting pattern across one mRNA. To reduce this low signal to noise ratio, one strategy could be to considerably increase the sequencing coverage of each mRNA sequence. However, the cost of this approach can be quite prohibitive, reducing its application. Alternatively, the use of a high salt treatment prior to footprinting could partially reduce signal to noise ratio by removing false positive RPFs produced by RBPs complexes or random 80S post-lysis reassociation. Indeed, the results obtained for low and high salt treated samples prepared from the same cytoplasmic lysate revealed that the footprinting quality, and notably the three nucleotides periodicity, was improved upon incubation with 1M KCl. Further

bioinformatic analysis would be required to check if RPFs profiles are better defined for individual transcripts in the high salt libraries compared to the low salt ones.

Using the high salt treatment, we were also able to visualize different states of translating ribosomes due to distinct structural conformations that were not visible in classical low salt samples. This improved visibility after treatment could be explained by an increased digestion efficiency of ribosomes with an empty A site after the removal of non-tightly bound ribosome associated factors following the 1M KCl incubation. Consequently, the RPFs produced by ribosomes not directly involved in the polypeptide elongation process generate smaller RPFs following high salt treatment compared to classical samples. The size of these smaller RPFs could depend on the activity of the RNases used for footprinting. Indeed, the RPFs are systematically longer in samples treated with S7 nuclease alone while many shorter fragments accumulate when aggressive RNase, such as RNase T1 is added. Apart from this, the high salt treatment also gives a better approximation of the percentage of ribosomes likely involved in translation as all non mRNA bound ribosomes are removed. Interestingly, while the amount of monosomes actually bound to mRNA was very low in the pBMDM lysates, most of them were involved in the translation process as more than 90% of mRNA derived reads corresponded to the CDS region in these samples.

An additional technical challenge was the depletion of contaminating rRNA fragments in our footprinting libraries. The optimization of the sample preparation protocol was already sufficient to greatly improve the coverage of mRNA sequences in our samples by inhibiting the activity of endogenous RNases. I also tried several depletion protocols to further reduce the rRNA contamination in our libraries and all the commercial kits tested revealed to be quite inefficient. Similar results were recently published by another lab regarding the inefficiency of the Qiaseq FastSelect and RiboCop kits for rRNA depletion from ribosome profiling samples (Zinshteyn et al., 2020). In our hands, the RNase H mediated depletion using custom oligonucleotides was the most effective strategy. However, a recent study suggests that nuclease based depletion should be avoided as it could induce a bias in the footprinting profiles through off-target trimming of the RPFs (Zinshteyn et al., 2020). Consequently, the nuclease induced degradation of the RPFs could perturb their size distribution and blur positioning information preventing the detection of a clear three nucleotide periodicity. This bias was not clearly detected in our ribosome profiling and monosome vs polysome samples. The wrong RPFs size distributions obtained for some of our samples were mainly explained by an inappropriate RNase digestion condition. It will be therefore interesting to prepare properly depleted or non-depleted samples using our RNase H protocol to make a side to side comparison of the RPFs length distributions and periodicity. Indeed, during my PhD, I did

perform a side to side comparison of untreated or depleted libraries prepared from the same cytoplasmic lysate but the RNase H treatment was performed before gel size selection leading to a higher rRNA contamination in the depleted sample. These samples thus cannot be used for a robust comparison of RPFs quality. Moreover, the authors suggest that it would be better to use custom biotinylated probes to perform rRNA depletion (Zinshteyn et al., 2020). However, the number of probes required for such depletion is higher for mouse macrophages lysates than for yeast leading to a great increase of the library preparation cost. Additionally, the depletion efficiency using the biotinylated oligonucleotides strategy was very poor compared to the results that we obtained using our RNase H protocol (rRNA proportion was reduced by 3% using biotinylated probes). Hence, we cannot be sure that this protocol would be suitable for our highly degraded samples. To date, as no clear bias of the RPFs quality preventing detailed bioinformatic analyses were detected in our footprinting data, I would suggest that RNase H is still the best approach for rRNA depletion in ribosome profiling samples.

After a long optimization process, our monosome vs polysome footprinting protocol using pBMDMs is now quite robust. Even if some adjustments could be made on the RNase digestion step, the footprinting quality is sufficient to study the differences between monosomal and polysomal populations. The results obtained so far open a wide range of questions that still need to be answered.

## ***Deciphering monosomes translational status***

Historical studies to decipher the protein synthesis mechanism used radioactive labelling of newly synthesized protein in highly translating systems such as rat liver extracts and RRL (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). Using this approach, the authors were able to detect newly synthesized proteins from polysomes but not from monosomes. It was thus concluded that 80S complexes found in the cell are not likely involved in translation. This vision of inactive monosomes prevailed for decades and was probably at the origin of restrictive conclusions particularly from polysome profiling experiments (Heyer and Moore, 2016). Indeed, it was considered that monosomes could be involved in the pioneer round of translation initiation or in the last round of termination but not in elongation. Thus only the polysomes fractions were collected to identify translated mRNAs in different conditions. This could notably explain a part of the differences between the amount of translated mRNAs identified by polysome profiling and concomitant measure of protein levels (Beyer et al., 2004).

Intriguingly, other historic experiments had revealed that ribosomal distributions vary depending on the cell type. When describing ribosomes for the first time in animal cells using electron microscopy, Palade already noticed different ribosomal association patterns depending on the cellular growth rates (Palade, 1955). Notably, packed ribosomal structures close to the endoplasmic reticulum were observed in highly proliferating cells. Conversely, in less proliferating cells, ribosomal structures were more scattered in the cytoplasm. During the set-up of the monosome vs polysome footprinting approach, I observed that the amount of monosomes retrieved in iBMDMs were reduced compared to primary macrophages with decreased proliferative capacities. This could be explained by the fact that the amount of ribosomal proteins produced can be dynamically adapted depending on the proliferation rates (Ingolia et al., 2019; Kafri et al., 2015; Vind et al., 1993). Hence, in highly proliferative cells, the competition for mRNA binding to ribosomal particles and translation factors could be reduced leading to a decreased amount of monosomes. On the contrary, in cells that do not proliferate, or more slowly, the regulation of ribosomal binding could be tighter due to an increased competition for ribosomal components and translation factors. Consistent with this hypothesis, I also observed a well reduced amount of monosomes in the human HEK293T cell line that has high proliferative capacities. Moreover, no newly synthesized proteins were detected from monosomes while protein synthesis clearly occurred in polysomes using HEK293T and iBMDMs samples to perform hybrid *in vitro* translation. Consequently, there could be an impact of the immortalization process on the ribosomes production and distribution across mRNAs. Notably, perturbations of the ribosomal binding pattern could be observed concomitantly to the modifications of protein synthesis rates in cancerous cells.

To characterize monosomes translational status using primary macrophages, the sample preparation for *in vitro* translation requires optimization because of the presence of endogenous RNases that perturb the separation of monosomes from polysomes. Particularly, a high amount of potent RNases inhibitors such as heparin must be added directly in the lysis buffer to protect the polysomes. Heparin is a non-specific competitive RNases inhibitor that could mimic RNAs (Gauthier and Ven Murthy, 1987). Interestingly, it has also been used to purify translation initiation factors, ribosomes and even aminoacyl-tRNA synthetases from cytoplasmic lysates (Hradec and Dusek, 1980; Hradec and Kríz, 1978). Heparin contamination can dramatically reduce the efficiency of reactions involving RNAs such as reverse transcription (Bai et al., 2000; Del Prete et al., 2007). Consequently, the presence of heparin in the translation mixture could drastically inhibit the reaction. After selection on a sucrose gradient and purification on a sucrose cushion, the heparin amounts should be well reduced and thorough washing of the ribosomal pellets should be performed to remove any traces left. Furthermore, it could be desirable to select only monosomes bound to mRNA and possibly involved in translation using a high salt treatment. This would notably help to properly calibrate the *in vitro* translation reactions in order to add the same amount of potentially active monosomes and polysomes. However, high salt treatment removes translation factors not tightly bound to the ribosomes thus leading to a translation blockade (Mohammad et al., 2019). To relieve this blockade, translation factors purified from pBMDMs could be added in the reaction. For this, one approach could be to use pBMDMs supernatant, obtained after the pelleting of the ribosomes from cytoplasmic lysates, similarly to what is done for RRL. However, this could also lead to the introduction of endogenous RNases released upon cell lysis and thus inhibit the reaction. Another solution could be to add purified translation elongation factors.

An alternative strategy to study monosomes translation activity is to perform the radioactive labelling of newly synthesized proteins directly in the cell culture and then sucrose gradient purification. Using this method, the addition of heparin and high salt treatment cannot negatively affect our capacity to detect new protein synthesis events. Additionally, as the pulse labelling is very short (2 min) and the cytoplasmic lysate directly loaded on a sucrose gradient, the timing between cell lysis and monosomes vs polysomes separation is reduced preventing the deleterious action of endogenous RNases. During my PhD, I implemented this approach before optimizing the sample collection procedure to retrieve intact monosomes and polysomes from pBMDMs. Consequently, I was able to detect newly synthesized proteins in the monosomes fractions but could not conclude definitively about the biological relevance of this observation. This experiment should thus be performed again using the optimized sample preparation protocol to obtain a robust validation of monosomes ability to produce new proteins.



Apart from this, the use of deep sequencing technologies is an effective strategy to circumvent the technical difficulty to detect monosomes translation. High-throughput sequencing was indeed necessary to provide the proof that some cellular mRNAs are mainly translated by monosomes in yeast and mammalian cells (Biever et al., 2020; Heyer and Moore, 2016). In accordance with the results obtained in these previous studies, transcripts preferentially translated by monosomes in pBMDMs were successfully identified in this work. Particularly, we were able to detect a clear three nucleotides periodicity specifically in the CDS region of monosomes associated mRNAs. Moreover, the comparison of RPFs size distribution between monosomes and polysomes samples revealed that the pattern obtained from the different ribosomal populations are very similar. Altogether, these results are a great evidence of monosomes implication in every step of the translation process including elongation. Our results are further strengthened by the fact that the RPFs were generated specifically from mRNA and tRNA bound ribosomes after a high salt treatment that decreased the probability of false positive RPFs.

To definitively conclude about monosomes translational status, a ribosomal run-off experiment using translation initiation inhibitors could rule out the possibility that CDS mapping RPFs are produced from ribosomes stalled all along the mRNAs. Similar approaches based on the labelling of newly synthesized proteins using puromycin were recently described (Argüello et al., 2018; Gao et al., 2015). For this, harringtonin or lactidomycin are added to the samples to block the early elongation step. Consequently, there is an accumulation of initiating ribosomes in the 5'UTR and around the start codon region while elongating ribosomes continue their movement until they fall off the mRNAs. To perform the run-off assay, non-initiating ribosomes must be actively elongating so the use of elongation inhibitors such as cycloheximide should be avoided. To avoid an increase of the monosomes induced by polysomes run-off, the initiation inhibitor must be added directly in the sample buffer and not pre-incubated with the cells. Cytoplasmic lysates should be incubated at 37°C with energy supplies to resume translation and in presence of puromycin to label newly synthesized proteins. Sucrose gradient purification of samples collected at several short time points after incubation at 37°C could be used to measure the amounts of puromycin labelled proteins from monosomes and polysomes fractions. This strategy could however give confounding results as puromycin is also an elongation inhibitor that induces the release of the nascent peptides from the ribosomes (Azzam and Algranati, 1973). This molecule could alternatively be used for *in vitro* run-off assays as the subsequent peptide release would not be a big issue after monosomes vs polysomes separation. Furthermore, the use of radioactive labelling in these run-off assays could be a more sensitive approach with decreased probability of newly synthesized peptide release. The same issues regarding the sample preparation procedure

that were described for the regular *in vitro* translation assay would still be true for these run-off assays using pBMDMs derived monosomes and polysomes.

Finally, to circumvent the detection issue, another strategy could be to use deep sequencing after run-off from purified monosomes and polysomes and then footprinting. The comparison of the patterns obtained on the 5' and 3' end of the CDS region following run-off in monosomes and polysomes samples could give the definitive proof that both ribosomal populations are actively elongating. If monosomes are indeed actively elongating, a decreased signal should be observed in the 5' end of the CDS while the signal in the 3' end should be increased providing that the run-off is short enough. Interestingly, for long run-off timing, only stalled ribosomes would still be bound to the CDS region. This experiment could thus be also interesting to characterize specifically the positions of stalled ribosomes on cellular mRNAs. In addition to this, as ribosomes recruitment to the 5'UTR is not blocked by translation initiation inhibitors, we could be able to detect if reinitiation efficiency is effectively higher in polysomes compared to monosomes. If this is the case, then the ribosomal peak near the start codon should display a higher fold change after a long run-off compared to sample without run-off in the polysomes samples.

To sum up, monosomes translational activity was confirmed in our study using deep sequencing of monosomes or polysomes derived footprints. The demonstration of their activity using a biochemical assay such as *in vitro* translation or radioactive pulse labelling directly in the cells and a run-off experiment could reinforce our conclusion that monosomes associated translation is significant in the global gene expression process.

## ***Characterization of the features explaining preferential association with monosomes or polysomes***

After the validation of our protocol to identify transcripts preferentially translated through monosomes or polysomes, a machine learning algorithm was used to identify features encoded by the mRNAs that could explain the preferential association to these distinct ribosomal populations. The random forest regression analysis can only be used to highlight correlations between a variable of interest, here the monosome vs polysome enrichment score, and a small set of pre-defined parameters. The parameters selected for our analysis corresponded mostly to mRNA features that were previously identified in the literature as important to regulate translation efficiency, both at the initiation or elongation levels. The final regression model constructed was able to recapitulate roughly 59% of the variance explaining differential monosome vs polysome association. Consequently, the predictions of the monosome vs polysome enrichment scores based on the mRNA features selected were quite good but incomplete. Notably, other cis-elements contained in the mRNA sequences could also be involved. Particularly, we did not thoroughly test the impact of the presence of miRNA binding sites or ARE-elements in our model. Moreover, trans-acting factors such as RBPs could also play a role in the differential association to monosomes or polysomes.

Using our random forest based regression model, we were able to confirm that monosomes and polysomes associated transcripts were properly segregated using our optimized protocol. Indeed, the ribosome density measured from pBMDMs in a different ribosome profiling experiment was the most important factor explaining our monosome vs polysome enrichment scores. Among the other parameters tested, the CDS length was the most important to explain monosome vs polysome association. This result is in accordance with other observations previously made in yeast (Arava et al., 2003; Heyer and Moore, 2016). While the mechanism explaining why ribosomal loading is so well correlated to the CDS length *in vivo* is unclear, other associated parameters leading to differences in the translation initiation efficiency depending on the mRNA size could be involved. Indeed, parameters that influence translation initiation rates, and particularly the GC content in the 5'UTR, had a great impact on predicted monosome vs polysome association levels. It will be thus very interesting to assess the impact of the other mRNA features on monosome vs polysome enrichment scores depending on the CDS size. For this, the enrichment scores should be plotted against the values of the different parameters for individual transcripts using bins of transcripts with similar CDS length.

Interestingly, the fact that all mRNAs are not equal regarding ribosome loading efficiency has been a long standing assumption (Kozak, 1991a; Lodish, 1974). Studying cis-encoded mRNA features that influence translation initiation efficiency, Marilyn Kozak identified a set of transcripts that “seem designed to be translated poorly” (Kozak, 1991a). Notably, some of them had a high GC content in their 5'UTRs, implying that these regions were highly structured leading to reduced translation initiation rates. This observation was recently validated using a high-throughput reporter assay with a synthetic mRNA library (Jia et al., 2020). Namely, the authors identified high GC structures in 5'UTRs that impair ribosome scanning and relocate the mRNAs to Processing or P-bodies to be degraded. As a reduced translation initiation efficiency was previously suggested as a mechanism to explain the preferential association to monosomes (Heyer and Moore, 2016), the comparison of the GC proportions in the 5'UTR regions of monosomes or polysomes enriched transcripts could be quite instructive. Besides the high GC content in the 5'UTRs, other mRNA features influencing translation initiation efficiency could also explain the preferential association to monosomes or polysomes. In Kozak's study, the presence of uORFs in the 5'UTR was described as a factor reducing initiation efficiency. However, the uORF length did not appear as a critical parameter in our regression model. Furthermore, the nucleotide context surrounding the start codon can also alter translation initiation rates (Kozak, 1991b, 2002). The Kozak context was indeed identified as a parameter that explains a part of the monosome vs polysome enrichment score in our random forest model. Additionally, it was recently shown that transcripts with a weak Kozak context display different behavior in response to the alterations of global translation initiation and elongation rates (Acevedo et al., 2018). Particularly, they are highly sensitive to drop in initiation rates and not really affected by global modifications of elongation rates. Consequently, the Kozak context impacts both initiation and elongation rates and could be a good parameter to promote translation through monosomes. It should be noted that the poorly translated transcripts identified by Marilyn Kozak encoded for oncoproteins, growth factors, transcription factors, signal transduction components and housekeeping genes known to be expressed at low levels (Kozak, 1991a). It is thus quite conceivable to imagine that such proteins would be mainly translated through monosomes regarding the results that we have obtained in our GO analyses.

Altogether, these observations open the fascinating question of whether cis-encoded mRNA features are sufficient to specify the preferential translation through monosomes. To answer it, mRNAs associated with monosomes and polysomes should be specifically purified from a sucrose gradient and then used for *in vitro* translation in presence of radioactive amino acids. Samples should be collected at several times during the translation reaction to see if the protein synthesis kinetics is different for mRNAs purified from monosomes or polysomes.

If cis-elements present in the mRNA sequences are indeed sufficient to reduce their translation efficiency, monosomes purified mRNAs would produce less proteins at a given time compared to those associated to polysomes. If it is not the case, then it would mean that the role of trans-acting factors such as RBPs is equally important.

Importantly, RBPs were previously identified as critical regulators of mRNAs fate in macrophages and many other cell types (Anderson, 2008; Mino and Takeuchi, 2018; Siomi and Dreyfuss, 1997). Consequently, the association of specific RBPs could also participate in the regulation of monosomes or polysomes binding. To identify the RBPs potentially involved in such regulation, mass spectrometry could be used to see if some of them are preferentially associated to monosomes or polysomes fractions at different times during the inflammatory response. As many proteins can interact with the ribosomes, directly or indirectly, mass spectrometry analysis from directly from the different sucrose gradient fractions might not be sufficiently resolute. To improve the detection of RBPs specifically associated to the different ribosomal populations, an immunoprecipitation targeting a core ribosomal protein should be performed after the sucrose gradient purification of monosomes and polysomes.

Monosomes enriched mRNAs were previously described as more unstable than the polysomes enriched in yeast (Heyer and Moore, 2016). For this reason, parameters accounting for mRNA stability were included in the random forest regression analysis. Using this approach, we observed that mRNA decay, occurring co-translationally or not, was also an important parameter to explain preferential monosome vs polysome enrichment. Notably, monosome enriched mRNAs could be intermediates targeted by cellular mRNA decay pathways. This could be particularly relevant for all co-translational decays mechanisms such as NMD, No-Go or Non-stop decays that implies ribosomal binding. The decay pathways independent of translation could also promote monosomes enrichment through the non-specific fragmentation of polysomes associated mRNAs. It would be particularly interesting to perform a differential degradome sequencing analysis (German et al., 2009) from monosomes and polysomes to characterize the degraded mRNAs population associated to each ribosomal subset. Furthermore, the role of specific RBPs binding could be very important to create a link between translational control through differential ribosomal loading efficiency and mRNA stability. As both mechanisms are critical to control the pool of mRNAs translated at a given time (Chan et al., 2018), interactions with RBPs could thus ensure the equilibrium between these two layers of regulation.

Another intriguing question is to what extent monosomes and polysomes composition is similar. Recent studies of the ribosomal particles composition revealed that, contrary to what was thought for many years, ribosomes association with ribosomal proteins can be quite variable depending on the cell type and the conditions (Emmott et al., 2019; Z et al., 2017). Ribosomal proteins composing the ribosomes can also be modified leading to different functional subsets of ribosomes (Genuth and Barna, 2018; Shi and Barna, 2015). It would be quite interesting if monosomes and polysomes were associated with distinct ribosomal proteins or ribosomal modifications providing them the ability to be recruited on different subsets of mRNAs. Moreover, ribosome heterogeneity, and hence activity, was described as a critical regulator of cell growth and metabolism (Calamita et al., 2018). Notably, basal cellular functions such as energy supply or mitochondrial function can be greatly affected following alterations of ribosomal availability in pathological situations. This could be connected to the regulation of the ribosomal binding pattern depending on the cellular needs. The characterization of the proteins specifically bound to monosomes vs polysomes by mass spectrometry after sucrose gradient purification and immunoprecipitation could be used to reveal an heterogeneity in ribosomal proteins association.

Alternatively, the differential ribosomal loading could be controlled through the regulation of the interactions with different translation factors. Notably, our monosome vs polysome footprinting results suggested a reduced translation termination efficiency in polysomes compared to monosomes. Consequently, the recycling rates following termination could be reduced in polysomes potentially promoting reinitiation events as the ribosomes continue to migrate in the 3'UTR region (Skabkin et al., 2013). This could be explained by a difference in the association of translation factors bound to the 3'UTR that would favor a circularized conformation promoting translation reinitiation (Alekhina et al., 2020; Archer et al., 2015). This conformation could notably be the consequence of interactions between the cap-binding initiation factor eIF4E and the adaptor protein eIF4G in 5'UTR and the poly-A binding protein (PABP) in 3'UTR that hold both mRNA ends in close proximity (Gallie, 1991; Wells et al., 1998). An interaction between the initiation factors in 5' and the termination complex eRF1/eRF3 in the 3' end was also demonstrated in yeast cell-free extracts (Amrani et al., 2008). Interestingly, electron microscopy studies performed to characterize ribosomal structures revealed an interplay between reinitiation rates and polysomal structures. Particularly, polysomes conformation is very dense and could regroup several ribosomes in arrays of tetramers working together to translate the same mRNAs (Karpova and Gillet, 2018). It was also previously suggested that translation initiation in circularized polysomes could occur mainly through reinitiation independently of the scanning of the 5'UTR region (Kopeina et al., 2008). Different levels of polysomal compaction were notably linked to modifications of

translation efficiencies (Viero et al., 2015). Consequently, increased reinitiation rates in polysomes could explain why protein synthesis rates are higher in this ribosomal compartment compared to monosomes (Heyer and Moore, 2016). Moreover, high reinitiation capacities could protect polysomes from global alterations of ribosomal subunits availability. Conversely, as ribosomal subunits are more recycled following monosomes translation termination, mRNAs mostly translated through monosomes could be subjected to a higher competition for the recruitment of the translation machinery in their 5' end compared to polysomes enriched transcripts. In conclusion, the rates of recycling vs reinitiation could play a significant role in the preferential association to monosomes or polysomes.

## ***How monosomes or polysomes binding shape the inflammatory response in macrophages ?***

Despite the increased technical difficulties due to the use of pBMDMs in our work, this cell type was still a particularly interesting model to study the adaptation of the ribosomal binding pattern to fluctuating conditions.

At the basal level, we were already able to detect differences in the subcellular localization and functions of the proteins encoded by mRNAs preferentially translated by monosomes or polysomes. These results are very interesting as they are revealing of the processes submitted to the highest gene expression regulation in macrophages. As most monosome-enriched mRNAs are highly regulated, we could expect an easier identification of the master regulators controlling the expression of functional gene clusters leading to the diverse macrophages phenotypes. Contrary to what was expected, not so many transcripts encoding for proteins involved in immunity were enriched in monosomes in basal conditions. Most of them were in fact mainly polysome associated while monosome bound transcripts were enriched in metabolic functions and transcription regulators. This could suggest that some proinflammatory transcripts are synthesized and associated to multiple ribosomes before inflammation triggering but not well translated. Upon stimulation, the translation process would be activated leading to a rapid “on demand” protein synthesis (Mata et al., 2005). Polysomes bound to translationally inhibited mRNAs expressed only in specific conditions were previously described (Baat et al., 2004; Rügsegger et al., 2001). These polysomes bound mRNAs could also be constitutively degraded and stabilized after macrophage activation. For instance, mRNA silencing through miRNAs binding was previously observed in the polysomes fractions (Nottrott et al., 2006; Petersen et al., 2006). Polysomes could also be targeted by specific RBPs to inhibit translation and recruit the mRNA decay pathways (Zhang et al., 2017).

Our results obtained following macrophages activation at different times post-LPS stimulation were biased due to the presence of endogenous RNases. Despite this, we were still able to see clear differences in the functions of monosomes and polysomes associated transcripts that were distinct at all time points during the inflammatory response. Undoubtedly, combining the monosome vs polysome footprinting approach to RNaseq to dissect gene expression regulation during the inflammatory response in pBMDMs will give a clearer picture of the different functional modules recruited upon activation and their interactions. Interestingly, as several negative regulators of the inflammatory response are sensitive to translational control (Lemaitre and Girardin, 2013), perturbations of the ribosomal binding



pattern could particularly affect the resolution of inflammation. The subsequent distortions of these negative regulators synthesis rates could participate in the triggering of chronic inflammatory pathologies.

Several layers of regulation are entangled to achieve controlled and well-orchestrated protein synthesis depending on the various phases of the inflammatory response. Interestingly, the control of gene expression could be adapted depending on evolutionary constraints linking the regulation of the expression of each protein to its functions (Schwanhäusser et al., 2011). Consequently, modifications of the ribosomal binding pattern during the inflammatory response could also be linked to other layers of regulation to adapt properly protein synthesis levels to the cellular needs. The relative instability of transcripts encoding for proinflammatory cytokines was well described previously (Kozak, 1991a; Schwanhäusser et al., 2011) and co-translational decay could also participate in the shaping of the inflammatory response (Zhang et al., 2017). Upon triggering the inflammatory response, many alternative splicing events have been described to allow the expression of specific proinflammatory factors (Carpenter et al., 2014). Notably, different mRNA isoforms can be translated at variable efficiencies (Floor and Doudna, 2016; Weatheritt et al., 2016). Hence, modifications of the 5' and 3' UTR regions through alternative splicing observed upon macrophage activation could regulate their relative association with monosomes or polysomes. Consequently, it would be particularly interesting to assess if monosomes or polysomes enriched transcripts correspond to the same mRNA isoforms during the different stages of the inflammatory response. For this, the different mRNA species found in the different fractions of a sucrose gradient should be characterized by RNaseq or Transcript Isoforms in Polysome sequencing , TriP-seq (Floor and Doudna, 2016).

To conclude, the results obtained using activated pBMDMs confirmed that previous studies performed in basal conditions only provide a partial vision of protein synthesis regulation. Importantly, the monosome vs polysome footprinting strategy can be adapted to various models from the response of other immune cells to bacterial or viral infections to the study of cancerous cell lines or in cells involved in the development to better characterize the impact of translational control in response to environmental changes.

## ***How monosomes and polysomes participate in the modulation of the cellular proteome depending on the conditions ?***

In order to characterize the impact of the differential monosomes or polysomes association on global gene expression rates, an integrative study combining monosome vs polysome footprinting, RNA-seq and proteomics analysis should be performed. Indeed, while our approach could give a good approximation of protein synthesis rates, a combined proteomics analysis would still be necessary to obtain a global view of protein levels regulation notably by taking in account variations of protein stability.

Regarding the modulation of protein synthesis rates depending on the cellular environment, models trained using ribosome profiling data revealed that ribosome allocation is a critical parameter to properly adapt translation to the cellular needs (Riba et al., 2019). Currently, the role of differential ribosome allocation across cellular mRNAs is poorly defined. Many regulatory mechanisms that participate in the control of ribosome loading were previously described (Kozak, 1991b, 2002) but their distinct impact on protein synthesis levels is not clear. The regulation of recycling vs reinitiation rates could play a significant role in the shaping of cellular proteome through the control of the association to monosomes or polysomes. This mechanism could be particularly important during the cell adaptation from one condition to another as the amount of ribosomal subunits and initiation factors can become limiting when most of them are engaged in translation (Dykeman, 2020). Indeed, ribosome availability is a critical factor in the control of protein synthesis (Shah et al., 2013) and increases of transcription induced following a change in the cellular environment could create a competition between mRNAs for ribosomal binding. Poorly translated mRNAs display features that reduce their ribosomal loading suggesting that they are encoded to produce few proteins. As stated in a recent review on translational control: “A goal of future work will be to precisely determine how the translation pathway can be reprogrammed to control what mRNA is selected for translation [...] and how much protein is synthesized from individual mRNAs” (Sokabe and Fraser, 2019). The monosome vs polysome footprinting approach could be a great tool to answer these questions.

Moreover, as the efficiency of ribosome recruitment is mostly impacted by transcript-specific features, studies focusing on global trends of translational control are not sufficient to fully understand how protein synthesis rates are regulated. To understand how translational control participates in the shaping of the cellular proteome depending on the conditions, it is necessary to isolate subsets of transcripts with similar features and/or functions. This can notably explain the difficulty to study this type of regulation using historical approaches. Particularly, initial studies performed to identify the active site of protein synthesis were focused on global protein synthesis rates in highly translating cell lysates, such as liver extracts, or on a small subset of highly translated mRNAs in the case of RRL (Gierer, 1963; Warner et al., 1963; Wettstein et al., 1963). The development of new methods, such as the high-throughput sequencing of ribosome protected RNA fragments, opened the new path to a more comprehensive understanding of protein synthesis regulation.

To sum up, in this work, we were able to identify the critical parameters to study protein synthesis regulation in pBMDMs in great detail using the monosome vs polysome footprinting approach. Notably, we confirmed that ribosome allocation pattern is controlled through cis-regulating features of the mRNAs. These features could be specifically recognized by trans-acting factors such as RBPs to coordinate protein synthesis rates to the other layers of gene expression for each transcript individually. This complicated network could help in the dynamic shaping of the cellular proteome depending on the environment. Particularly, differential association to monosomes or polysomes could participate in the regulation of gene expression following macrophage activation and partly explain their high functional plasticity. This phenomenon could be especially important during the inflammatory response as different macrophages functions are required depending on the stages of inflammation. Finally, the use of monosome vs polysome footprinting in various conditions can bring our understanding of protein synthesis regulation in a fluctuating environment to a different dimension with a more complete view on transcript-specific translational control.

# Material and Methods

## ***Bone-marrow derived macrophages culture***

Bone-marrow derived macrophages were prepared using 8 weeks old wild-type female C57Bl/6J mice from Charles River. Bone marrow cells were flushed out of the mice bones and seeded at  $30 \cdot 10^6$  cells/dish in 15 cm dishes. They were cultured for seven days at 37°C, 5% CO<sub>2</sub> with DMEM medium supplemented with penicillin, streptomycin, 10% heat-inactivated fetal calf serum (FCS) and 20% of L929 conditioned medium as a source of macrophage stimulating factor (M-CSF). Immortalized macrophage cell line was generated from bone-marrow cells infected with an oncogenic virus (J2) as previously described (Blasi et al., 1989). For activation assays, macrophages were stimulated with LPS at 100 ng/mL for indicated times.

## ***Cytoplasmic lysate preparation***

The cells were plated the day before to reach 80% confluency at the time of collection ( $21 \cdot 10^6$  cells in a 15-cm dish for primary bone-marrow derived macrophages). For lysis, cells were placed on ice and quickly washed with 10 mL of ice-cold PBS containing 100 µg/mL of cycloheximide (CHX). After this, they were scraped off the dish in 1 ml of ice-cold PBS with CHX and transferred in a 2 mL tube. The cells were then pelleted at 500g for 5 min at 4°C. The pellet was gently resuspended in 1 mL of lysis buffer (10 mM Tris-HCl pH7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 1% Triton X-100, 100 µg/ml cycloheximide, 2 mM DTT) and incubated on ice for 10 min. Finally, the lysate was clarified by centrifugation at 1300g for 10 min at 4°C. The cleared lysate obtained was directly used either for RNA-seq, ribosome profiling or monosome vs polysome footprinting. Alternatively, the lysate was snap-frozen in liquid nitrogen and stored for several weeks at -80°C before use.

## ***Polysome and Ribosome Profiling***

The cytoplasmic lysate was quantified by measuring its absorbance at 260 nm by Nanodrop. For ribosome profiling, 5 absorbance units (AU) were incubated in presence of RNase for 30 min at 25°C. Different RNases (RNase A from Ambion, RNase T1 from Thermo Fisher and RNase S7 from Sigma) were tested at variable concentrations (detailed in the Results section). Sucrose gradient solutions were prepared weight/volume in gradient buffer (20 mM HEPES-KOH pH 7.4, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 100 µg/mL cycloheximide, 2 mM DTT). Gradients were poured using a Gradient Master (Biocomp). Digested lysate was loaded onto a 11 ml 10-50% gradient and spun for 2h40 at 35,000 rpm at 4°C. For polysome profiling, 5 AU of non-digested lysate was loaded on the sucrose gradient. For high salt treatment, 4X high KCl buffer

(20 mM HEPES-KOH pH 7.4, 4M KCl, 5 mM MgCl<sub>2</sub>, 100 µg/mL cycloheximide, 2 mM DTT) was added directly in the cytoplasmic lysate reaching a 1X final concentration. The lysates were then incubated on ice for 20 min before loading on a sucrose gradient. If a RNase digestion was performed after the high salt treatment, the lysates were desalted using ZebaSpin filtration columns from ThermoFischer accordingly to the manufacturer's guidelines before RNase addition. The absorbance at 254 nm was recorded and the gradient fractions were collected using a Density Gradient Fractionation System (Brandel #BR-188). Fractions corresponding to the 80S monosome peak were collected and pooled.

### ***Monosome vs Polysome Footprinting***

9 AU of clarified lysate was loaded onto a 11 ml 10-50% gradient and spun for 2h40 at 35,000 rpm at 4°C. Fractions corresponding to either the monosome peak or polysome peaks were pooled, resulting in ~5 ml of monosomes and ~15 ml of polysomes. To dilute the sucrose, an equal volume of gradient buffer was added to each pool. Samples were then concentrated on Amicon-Ultra 100K columns (Millipore #UFC910024 and #UFC810024) by spinning at 4,000g for either 8 min (monosome fractions) or 15 min (polysome fractions). Concentrated monosome or polysome fractions (volume between 500-1000µl) were digested with RNase at 25°C for 30 min. As for ribosome profiling, different RNases were tested at variable concentrations (detailed in the Results section). Digested fractions were loaded onto a second 10-50% sucrose gradient and centrifuged at 35,000rpm for 2hr40 at 4°C. Gradient fractions were collected as above, and the monosome fractions were pooled.

### ***Ribosome Footprint Isolation***

The collected monosome fractions were supplemented with EDTA 15 mM final to promote ribosomal subunits dissociation. The sample was next treated with proteinase K in presence of 1% SDS for 45 min at 42°C. After this, RNA extraction was performed using acid phenol-chloroform, followed by ethanol precipitation and resuspension in 25µl of RNase free water. The RNA fragments of 19-38 nt size were selected from a denaturing 8M Urea 10% polyacrylamide gel. RNA was eluted from gel fragments in RNA Elution Buffer (300 mM NaCl and 1 mM EDTA). After an overnight incubation with constant rotation at 4°C, the eluate was isopropanol precipitated. The RNA pellet was resuspended in 25µL of RNase free water and used for library construction.

### ***Library Construction***

Deep sequencing libraries were prepared using the optimized kit-free Omniprep protocol (Heyer et al., 2015). Briefly, purified RNA 3' ends were dephosphorylated at 37°C for 4h using 12.5U of the T4 polynucleotide kinase (PNK) supplied by NEB and the manufacturer's buffer. RNA fragments were then ligated to a pre-adenylated adaptor (5'-rAppAGATCGGAAGAGCACACGTCTGAACTCCAddC-3') using the T4 RNL2 Tr.K227Q (NEB). The ligation reaction was carried out at 30°C for 4h and then heat-inactivated at 65°C for 20mn. The ligated RNAs were reverse transcribed for 45mn at 55°C using Superscript III (Invitrogen) with the first-strand buffer without MgCl<sub>2</sub>. After heat inactivation (15mn at 70°C), the RT products were selected on a 8M Urea 10% polyacrylamide gel. Gel-purified cDNA were circularized with CircLigase I (Lucigen) and PCR-amplified using Illumina's primers 1.0 and 2.0. The number of amplification cycles was optimized depending on the RNA input amount : 12 cycles for RNA-Seq, 7 cycles for ribosome profiling and 9 cycles for monosome and polysome footprinting. The PCR products were purified on a non-denaturing 10% polyacrylamide gel. After this, the libraries were quantified using the TapeStation system (Agilent) and pooled before sequencing in the GenomEast Platform, IGBMC, Illkirch, France.

### ***Library Sequencing and Genome Alignment***

Libraries were sequenced on an Illumina HiSeq 4000 along with PhiX genome derived fragments to increase base calling accuracy (single-end, 50 bp run). Data demultiplexing was performed using the Python library Flexi-splitter 1.0.2. The 3' adaptor sequence (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') was removed and reads smaller than 15 nts were filtered out using Cutadapt 2.1 (Martin, 2011). Reads mapping to PhiX genome or mouse rRNAs, tRNAs, snRNAs and snoRNAs sequences were removed using Bowtie 1 with the parameters "-v 2 -k 1" (Langmead, 2010). Remaining reads were mapped to the mouse genome (GRCm38.p6 primary assembly from Gencode) using HISAT2 v2.1.0 with the arguments "-k 20 --non-deterministic --rna-strandness 'F' --no-unal" and providing a defined set of known splice sites extracted from the Gencode vM23 comprehensive gene annotation primary gtf file. Only primary alignments were used for the following analysis, secondary alignments were filtered using Samtools 1.6 with the parameter "-F 256" (Li et al., 2009).

### ***Genome Counts and Monosome vs Polysome Score***

Counts per gene were calculated from genome-mapping reads using HTSeq (Anders et al., 2014) with parameters "-f bam -s yes -a 10 -t CDS -i gene\_id -m union". Only a single transcript isoform, tagged APPRIS principal, was considered per gene (Rodriguez et al., 2018). Resulting monosome and polysome counts were fed into DESeq2 (Love et al., 2014) for quantification of enrichment in either library. The assigned monosome:polysome score was the log<sub>2</sub> fold change (log<sub>2</sub>FC) calculated by DESeq2.

### ***Contaminant Analysis and Transcriptome Alignment***

Fasta files corresponding to the mouse rRNAs, tRNAs, snRNAs, and snoRNAs sequences were downloaded from the Ensembl database. These files were used to create independent genome reference files prior to mapping with Bowtie 1 with the parameters "-v 2 -k 1". To calculate the percentage of reads originating from mRNAs, all reads left after non coding RNAs and bacterial reads filtering were mapped to the mouse APPRIS principal transcript sequences also using Bowtie1.

### ***Complementary Sequencing Data Analysis***

The FLOSS scores were calculated using the scripts provided in the corresponding study (Ingolia et al., 2014). The RPF length distribution, measurements of the number of reads mapping to the different mRNA regions and RPF length dependent periodicity plots were generated using the RiboFlow pipeline (Ozadam et al., 2020). GO enrichment of monosomes or polysomes associated transcripts was performed using GeneCodis 4.0 (Tabas-Madrid et al., 2012). The clustering algorithm REVIGO was used to summarize GO analyses results on semantic similarity-based scatterplots (Supek et al., 2011). All the plots were generated using the R package ggplot2 (<https://ggplot2.tidyverse.org/>).

### ***Duplex specific nuclease mediated depletion of rRNA***

rRNAs were first depleted from the footprinting libraries after the PCR amplification step. For this, 12 µL of libraries were mixed with 4 µL of 4X hybridization buffer (200 mM HEPES pH 7.5 and 2 M NaCl) and denatured at 98°C for 3 min. The mix was slowly cool-down to 68°C (drop of 3°C/sec) and further incubated at 68°C for 45 min to allow re-annealing. After this, 2 µL of the commercial 10X DSN buffer and 4U of DSN (Evrogen) were added. Digestion was allowed to proceed for 45 min at 68°C. The reaction was stopped by the addition of 20 µL of 10 mM EDTA and incubation for a further 5 min at 68°C. DNA was recovered using Ampure XP beads mediated purification and PCR amplified.

For the probe-directed degradation, 5 $\mu$ L of circularized libraries were ethanol precipitated to remove the MnCl<sub>2</sub> present in the circularization reaction that could decrease DSN activity. The precipitated product was used to set-up the depletion reaction using the commercial 10X DSN buffer and antisense rRNA oligonucleotides at the final concentration of 300 nM each. The mix was denatured on a thermocycler at 95°C for 5 min, brought to 75°C and then slowly cooled (0.1°C/sec) to 55°C. After incubation for 5 min, 5  $\mu$ l of pre-warmed DSN master mix containing 0.4 U DSN (Evrogen) in 1 $\times$  DSN buffer was added. The depletion mix was further incubated for 30 mins at 55°C. The reaction was stopped by the addition of 20  $\mu$ L of 10 mM EDTA and incubated for another 10 minutes. The depleted circularization products were purified using basic phenol:chloroform extraction. The purified product was then PCR amplified using 12 cycles to complete the library construction.

### ***RNase H mediated depletion of rRNA***

To prepare rRNA depletion probes, 195 50-nt long DNA oligonucleotides covering the reverse complement of the entire length of each human rRNA were designed. Mouse specific probes were also added as described in the results section. Equal molar amounts of each oligonucleotide was used in the depletion reaction.

To deplete rRNA, the purified RNA sample was resuspended in 5  $\mu$ L of 5X hybridization buffer (200 mM NaCl, 100 mM Tris-HCl pH 7.4) and then mixed with 20 $\mu$ L of rRNA depletion probes at the final concentration of 0.5 $\mu$ M each. Heat denaturation was performed at 95 °C for 2 min and then the temperature was slowly reduced to 45°C (-0.1 °C/s). We next added 3  $\mu$ L of 10X RNase H digestion buffer (1M NaCl, 200mM MgCl<sub>2</sub>, 500mM Tris-HCl pH 7.4), 5 U of Hybridase Thermostable RNase H (Lucigen) and RNase free water qsp 30 $\mu$ L to the RNA and DNA oligo mix. The mixture was incubated at 45 °C for 1 hour. The DNA probes were then degraded using TURBO DNase according to manufacturer's guidelines and the DNase-treated RNA were purified by acid phenol-chloroform extraction before using them for library preparation.



### ***In vitro translation***

For *in vitro* translation, untreated or treated RRL lysate provided in the Flexi® Rabbit Reticulocyte Lysate System from Promega were used. To pellet the rabbit ribosomes and collect the RRL supernatant, RRL was ultracentrifuged for 2h 15 min at 240,000g at 4°C. Hybrid translation reactions were performed in a final volume of 20µL consisting of 8µL of RRL supernatant, 6µL of purified ribosomes from different sources, KCl 75mM final, MgCl<sub>2</sub> 0.5 mM final, DTT 100mM final, 20µM amino acids mixture without methionine, 3µL of S<sup>35</sup> labelled methionine. The translation mixture was incubated at 30°C for indicated times. The reaction was then stopped by the addition of luciferase lysis buffer. Renilla activity was measured in a Mithras luminometer, using the Renilla Luciferase Assay System (Promega). For detection of radioactive proteins, samples were resolved on a SDS-PAGE (10% gel), dried and subjected to autoradiography for indicated times. The signal was quantified using the Typhoon PhosphorImager System.

# References

- Acevedo, J.M., Hoermann, B., Schlimbach, T., and Teleman, A.A. (2018). Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci Rep* 8, 1–12.
- Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysocker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods* 10, 623–629.
- Alekhina, O.M., Terenin, I.M., Dmitriev, S.E., and Vassilenko, K.S. (2020). Functional Cyclization of Eukaryotic mRNAs. *Int J Mol Sci* 21.
- Amrani, N., Ghosh, S., Mangus, D.A., and Jacobson, A. (2008). Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* 453, 1276–1280.
- Anderson, P. (2008). Post-transcriptional control of cytokine production. *Nature Immunology* 9, 353–359.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 100, 3889–3894.
- Archer, S.K., Shirokikh, N.E., Hallwirth, C.V., Beilharz, T.H., and Preiss, T. (2015). Probing the closed-loop model of mRNA translation in living cells. *RNA Biol* 12, 248–254.
- Argüello, R.J., Reverendo, M., Mendes, A., Camosseto, V., Torres, A.G., Pouplana, L.R. de, Pavert, S.A. van de, Gatti, E., and Pierre, P. (2018). SunRiSE – measuring translation elongation at single-cell resolution by means of flow cytometry. *J Cell Sci* 131.
- Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *J Exp Med* 79, 137–158.
- Aylett, C.H.S., and Ban, N. (2017). Eukaryotic aspects of translation initiation brought into focus. *Philos Trans R Soc Lond B Biol Sci* 372.
- Azzam, M.E., and Algranati, I.D. (1973). Mechanism of puromycin action: fate of ribosomes after release of nascent protein chains from polysomes. *Proc. Natl. Acad. Sci. U.S.A.* 70, 3866–3869.
- Bai, X., Fischer, S., Keshavjee, S., and Liu, M. (2000). Heparin interference with reverse transcriptase polymerase chain reaction of RNA extracted from lungs after ischemia-reperfusion. *Transpl. Int.* 13, 146–150.
- Barry, K.C., Ingolia, N.T., and Vance, R.E. (2017). Global analysis of gene expression reveals mRNA superinduction is required for the inducible immune response to a bacterial pathogen. *ELife* 6, e22707.
- Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in *Neurospora*. *Proc Natl Acad Sci U S A* 27, 499–506.

- Beilharz, T.H., and Preiss, T. (2004). Translational profiling: the genome-wide measure of the nascent proteome. *Brief Funct Genomic Proteomic* 3, 103–111.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74, 3171–3175.
- Berk, A.J., and Sharp, P.A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12, 721–732.
- Beyer, A., Hollunder, J., Nasheuer, H.-P., and Wilhelm, T. (2004). Post-transcriptional Expression Regulation in the Yeast *Saccharomyces cerevisiae* on a Genomic Scale. *Molecular & Cellular Proteomics* 3, 1083–1092.
- Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J.D., and Schuman, E.M. (2020). Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* 367.
- Blin, J., and Ricci, E.P. (2016). [An intimate look at the viral replication cycle through ribosome profiling]. *Med Sci (Paris)* 32, 849–860.
- Blobel, G., and Sabatini, D. (1971). Dissociation of Mammalian Polyribosomes into Subunits by Puromycin. *Proc Natl Acad Sci U S A* 68, 390–394.
- Bludau, I., and Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nat Rev Mol Cell Biol* 1–14.
- Bossche, J.V. den, O'Neill, L.A., and Menon, D. (2017). Macrophage Immunometabolism: Where Are We (Going)? *Trends in Immunology* 38, 395–406.
- Braat, A.K., Yan, N., Arn, E., Harrison, D., and Macdonald, P.M. (2004). Localization-Dependent Oskar Protein Accumulation: Control after the Initiation of Translation. *Developmental Cell* 7, 125–131.
- Brina, D., Grosso, S., Miluzio, A., and Biffo, S. (2011). Translational control by 80S formation and 60S availability: The central role of eIF6, a rate limiting factor in cell cycle progression and tumorigenesis. *Cell Cycle* 10, 3441–3446.
- Browne, G.J., and Proud, C.G. (2002). Regulation of peptide-chain elongation in mammalian cells. *European Journal of Biochemistry* 269, 5360–5368.
- Burroughs, A.M., and Aravind, L. (2019). The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways. *Int J Mol Sci* 20.
- Buttgereit, F., and Brand, M.D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem J* 312, 163–167.
- Calamita, P., Gatti, G., Miluzio, A., Scagliola, A., and Biffo, S. (2018). Translating the Game: Ribosomes as Active Players. *Front Genet* 9.
- Carpenter, S., Ricci, E.P., Mercier, B.C., Moore, M.J., and Fitzgerald, K.A. (2014). Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* 14, 361–376.
- Cassan, M., and Rousset, J.P. (2001). UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol. Biol.* 2, 3.

- Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *ELife* 7, e32536.
- Chandrasekaran, V., Juskiewicz, S., Choi, J., Puglisi, J.D., Brown, A., Shao, S., Ramakrishnan, V., and Hegde, R.S. (2019). Mechanism of ribosome stalling during translation of a poly(A) tail. *Nature Structural & Molecular Biology*.
- Chen, J., Fang, X., Zhong, P., Song, Z., and Hu, X. (2019). N6-methyladenosine modifications: interactions with novel RNA-binding proteins and roles in signal transduction. *RNA Biology* 16, 991–1000.
- Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell* 11, 819–836.
- Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T.C. (2014). Global Proteome Turnover Analyses of the Yeasts *S. cerevisiae* and *S. pombe*. *Cell Rep* 9, 1959–1965.
- Chu, D., and von der Haar, T. (2012). The architecture of eukaryotic translation. *Nucleic Acids Res* 40, 10098–10106.
- Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C., and 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, 1731–1745.
- Collart, M.A., and Weiss, B. (2020). Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res* 48, 1043–1055.
- Corbett, A.H. (2018). Post-transcriptional Regulation of Gene Expression and Human Disease. *Curr Opin Cell Biol* 52, 96–104.
- Costello, J., Castelli, L.M., Rowe, W., Kershaw, C.J., Talavera, D., Mohammad-Qureshi, S.S., Sims, P.F.G., Grant, C.M., Pavitt, G.D., Hubbard, S.J., et al. (2015). Global mRNA selection mechanisms for translation initiation. *Genome Biol* 16, 1–21.
- Crick, F.H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.
- Del Prete, M.J., Vernal, R., Dolznig, H., Müllner, E.W., and Garcia-Sanz, J.A. (2007). Isolation of polysome-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. *RNA* 13, 414–421.
- Dever, T.E., Dinman, J.D., and Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harb Perspect Biol* 10.
- D’Orazio, K.N., Wu, C.C.-C., Sinha, N., Loll-Krippléber, R., Brown, G.W., and Green, R. (2019). The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay (eLife Sciences Publications Limited).
- Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4482.
- Dykeman, E.C. (2020). A stochastic model for simulating ribosome kinetics in vivo. *PLoS Comput Biol* 16.

- Eichelbaum, K., and Krijgsveld, J. (2014). Rapid Temporal Dynamics of Transcription, Protein Synthesis, and Secretion during Macrophage Activation. *Mol Cell Proteomics* *13*, 792–810.
- den Elzen, A.M.G., Schuller, A., Green, R., and Séraphin, B. (2014). Dom34-Hbs1 mediated dissociation of inactive 80S ribosomes promotes restart of translation after stress. *EMBO J* *33*, 265–276.
- Emmott, E., Jovanovic, M., and Slavov, N. (2019). Approaches for Studying Ribosome Specialization. *Trends in Biochemical Sciences* *44*, 478–479.
- Everts, B., Amiel, E., Huang, S.C.-C., Smith, A.M., Chang, C.-H., Lam, W.Y., Redmann, V., Freitas, T.C., Blagih, J., van der Windt, G.J.W., et al. (2014). TLR-driven early glycolytic reprogramming via the kinases TBK1- $\text{IKK}\epsilon$  supports the anabolic demands of dendritic cell activation. *Nat Immunol* *15*, 323–332.
- Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* *79*, 351–379.
- Feehan, K.T., and Gilroy, D.W. (2019). Is Resolution the End of Inflammation? *Trends in Molecular Medicine* *25*, 198–214.
- Fernandes, L.D., Moura, A.P.S. de, and Ciandrini, L. (2017). Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Scientific Reports* *7*.
- Floor, S.N., and Doudna, J.A. (2016). Tunable protein synthesis by transcript isoforms in human cells. *ELife* *5*, e10921.
- Fraser, C.S. (2015). Quantitative studies of mRNA recruitment to the eukaryotic ribosome. *Biochimie* *114*, 58–71.
- Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. (1999). A sampling of the yeast proteome. *Mol. Cell. Biol.* *19*, 7357–7368.
- Galli, S.J., Borregaard, N., and Wynn, T.A. (2011). Phenotypic and functional plasticity of cells of innate immunity: macrophages, mast cells and neutrophils. *Nat Immunol* *12*, 1035–1044.
- Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.* *5*, 2108–2116.
- Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S.-B. (2015). Quantitative profiling of initiating ribosomes in vivo. *Nature Methods* *12*, 147–153.
- Gauthier, D., and Ven Murthy, M.R. (1987). Efficacy of RNase inhibitors during brain polysome isolation. *Neurochem Res* *12*, 335–339.
- Gebauer, F., and Hentze, M.W. (2004). Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* *5*, 827–835.
- Genuth, N.R., and Barna, M. (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell* *71*, 364–374.
- Gerashchenko, M.V., and Gladyshev, V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res* *42*, e134.

- Gerashchenko, M.V., and Gladyshev, V.N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Res* 45, e6.
- German, M.A., Luo, S., Schroth, G., Meyers, B.C., and Green, P.J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nature Protocols* 4, 356–362.
- Gierer, A. (1963). Function of aggregated reticulocyte ribosomes in protein synthesis. *J. Mol. Biol.* 6, 148–157.
- Gierer, A., and Mundry, K.W. (1958). Production of mutants of tobacco mosaic virus by chemical alteration of its ribonucleic acid in vitro. *Nature* 182, 1457–1458.
- Gierer, A., and Schramm, G. (1956). Infectivity of ribonucleic acid from tobacco mosaic virus. *Nature* 177, 702–703.
- Giess, A., Torres Cleuren, Y.N., Tjeldnes, H., Krause, M., Bizuayehu, T.T., Hiensch, S., Okon, A., Wagner, C.R., and Valen, E. (2020). Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation. *Cell Reports* 31, 107534.
- Glass, C.K., and Natoli, G. (2015). Molecular control of activation and priming in macrophages. *Nature Immunology* 17, 26.
- Gordon, S., and Mantovani, A. (2011). Diversity and plasticity of mononuclear phagocytes. *European Journal of Immunology* 41, 2470–2472.
- Gordon, S., and Plüddemann, A. (2019). The Mononuclear Phagocytic System. Generation of Diversity. *Front Immunol* 10.
- Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 4, 117.
- Guydosh, N.R., and Green, R. (2014). Dom34 Rescues Ribosomes in 3' Untranslated Regions. *Cell* 156, 950–962.
- Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology* 19, 1720–1730.
- Hamidzadeh, K., Christensen, S.M., Dalby, E., Chandrasekaran, P., and Mosser, D.M. (2017). Macrophages and the Recovery from Acute and Chronic Inflammation. *Annu Rev Physiol* 79, 567–592.
- Hanson, G., and Collier, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* 19, 20–30.
- Harigaya, Y., and Parker, R. (2010). No-go decay: a quality control mechanism for RNA in translation. *WIREs RNA* 1, 132–141.
- Harvey, R.F., Smith, T.S., Mulroney, T., Queiroz, R.M.L., Pizzinga, M., Dezi, V., Villeneuve, E., Ramakrishna, M., Lilley, K.S., and Willis, A.E. (2018). Trans-acting translational regulatory RNA binding proteins. *Wiley Interdiscip Rev RNA* 9.
- Hershey, J.W.B., Sonenberg, N., and Mathews, M.B. (2012). Principles of Translational Control: An Overview. *Cold Spring Harb Perspect Biol* 4, a011528.

Heuer, A., Genova, M., Schmidt, C., Trowitzsch, S., Preis, A., Kötter, P., Berninghausen, O., Becker, T., Beckmann, R., and Tampé, R. (2017). Structure of the 40S–ABCE1 post-splitting complex in ribosome recycling and translation initiation. *Nat Struct Mol Biol* 24, 453–460.

Heyer, E.E., and Moore, M.J. (2016). Redefining the Translational Status of 80S Monosomes. *Cell* 164, 757–769.

Hradec, J., and Dusek, Z. (1980). Particulate aminoacyl-tRNA synthetases are retained on heparin bound to Sepharose. *Mol. Biol. Rep.* 6, 245–248.

Hradec, J., and Kríz, O. (1978). Heparin-sepharose 4B at low temperatures retains ribosomes. *Biochem J* 173, 349–352.

Humphreys, D.T., Westman, B.J., Martin, D.I.K., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *PNAS* 102, 16961–16966.

Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S., and Press, W.H. (2015). Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLOS Genetics* 11, e1005732.

Ibrahim, F., Maragkakis, M., Alexiou, P., and Mourelatos, Z. (2018). Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat Struct Mol Biol* 25, 302–310.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 324, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* 147, 789–802.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep* 8, 1365–1379.

Ingolia, N.T., Hussmann, J.A., and Weissman, J.S. (2019). Ribosome Profiling: Global Views of Translation. *Cold Spring Harb Perspect Biol* 11, a032698.

Jackson, R., Kroehling, L., Khitun, A., Bailis, W., Jarret, A., York, A.G., Khan, O.M., Brewer, J.R., Skadow, M.H., Duizer, C., et al. (2018). The Translation of Non-Canonical Open Reading Frames Controls Mucosal Immunity. *Nature* 564, 434–438.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). THE MECHANISM OF EUKARYOTIC TRANSLATION INITIATION AND PRINCIPLES OF ITS REGULATION. *Nat Rev Mol Cell Biol* 11, 113–127.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3, 318–356.

Jia, L., Mao, Y., Ji, Q., Dersh, D., Yewdell, J.W., and Qian, S.-B. (2020). Decoding mRNA translatability and stability from the 5' UTR. *Nature Structural & Molecular Biology* 1–8.

Jingyi, J.L., Bickel, P.J., and Biggin, M.D. (2020). System wide analyses have underestimated protein abundances and the importance of transcription in mammals [PeerJ].

- Joazeiro, C.A.P. (2019). Mechanisms and functions of ribosome-associated protein quality control. *Nat Rev Mol Cell Biol* 20, 368–383.
- JOSEPH, S. (2003). After the ribosome structure: How does translocation work? *RNA* 9, 160–164.
- Jovanovic, M., Rooney, M.S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E.H., Fields, A.P., Schwartz, S., Raychowdhury, R., et al. (2015). Dynamic profiling of the protein life cycle in response to pathogens. *Science* 347, 1259038.
- Kafri, M., Metzli-Raz, E., Jona, G., and Barkai, N. (2015). The Cost of Protein Production. *Cell Rep* 14, 22–31.
- Karamyshev, A.L., and Karamysheva, Z.N. (2018). Lost in Translation: Ribosome-Associated mRNA and Protein Quality Controls. *Front Genet* 9, 431.
- Karpova, E.A., and Gillet, R. (2018). The Structural and Functional Organization of Ribosomal Compartment in the Cell: A Mystery or a Reality? *Trends in Biochemical Sciences*.
- Kelly, B., and O'Neill, L.A. (2015). Metabolic reprogramming in macrophages and dendritic cells in innate immunity. *Cell Res* 25, 771–784.
- Kopeina, G.S., Afonina, Z.A., Gromova, K.V., Shirokov, V.A., Vasiliev, V.D., and Spirin, A.S. (2008). Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. *Nucleic Acids Res* 36, 2476–2488.
- Koppenol-Raab, M., Sjoelund, V., Manes, N.P., Gottschalk, R.A., Dutta, B., Benet, Z.L., Fraser, I.D.C., and Nita-Lazar, A. (2017). Proteome and Secretome Analysis Reveals Differential Post-transcriptional Regulation of Toll-like Receptor Responses. *Mol Cell Proteomics* 16, S172–S186.
- Kozak, M. (1987). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol. Cell. Biol.* 7, 3438–3445.
- Kozak, M. (1991a). An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115, 887–903.
- Kozak, M. (1991b). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* 266, 19867–19870.
- Kozak, M. (2001). Constraints on reinitiation of translation in mammals. *Nucleic Acids Res.* 29, 5226–5232.
- Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299, 1–34.
- Kozlovski, I., and Agami, R. (2019). More or less – the same? mRNA fluctuations are balanced during translation. *The EMBO Journal* 38, e103651.
- Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *ELife* 3.



Lauterbach, M.A., Hanke, J.E., Serefidou, M., Mangan, M.S.J., Kolbe, C.-C., Hess, T., Rothe, M., Kaiser, R., Hoss, F., Gehlen, J., et al. (2019). Toll-like Receptor Signaling Rewires Macrophage Metabolism and Promotes Histone Acetylation via ATP-Citrate Lyase. *Immunity* *51*, 997-1011.e7.

Lemaitre, B., and Girardin, S.E. (2013). Translation inhibition and metabolic stress pathways in the host response to bacterial pathogens. *Nat Rev Microbiol* *11*, 365–369.

Leprivier, G., Remke, M., Rotblat, B., Dubuc, A., Mateo, A.-R.F., Kool, M., Agnihotri, S., El-Naggar, A., Yu, B., Somasekharan, S.P., et al. (2013). The eEF2 Kinase Confers Resistance to Nutrient Deprivation by Blocking Translation Elongation. *Cell* *153*, 1064–1079.

Li, J.J., Chew, G.-L., and Biggin, M.D. (2019). Quantitative principles of cis-translational control by general mRNA sequence features in eukaryotes. *Genome Biol* *20*, 1–24.

Liu, B., and Qian, S.-B. (2016). Characterizing inactive ribosomes in translational profiling. *Translation (Austin)* *4*, e1138018.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* *165*, 535–550.

Lodish, H.F. (1974). Model for the regulation of mRNA translation applied to haemoglobin synthesis. *Nature* *251*, 385–388.

Lorent, J., Kusnadi, E.P., van Hoef, V., Rebello, R.J., Leibovitch, M., Ristau, J., Chen, S., Lawrence, M.G., Szkop, K.J., Samreen, B., et al. (2019). Translational offsetting as a mode of estrogen receptor  $\alpha$ -dependent regulation of gene expression. *The EMBO Journal* *38*, e101323.

MacDonald, C.T., and Gibbs, J.H. (1969). Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers* *7*, 707–725.

Mangeot, P.E., Risson, V., Fusil, F., Marnef, A., Laurent, E., Blin, J., Mournetas, V., Massouridès, E., Sohier, T.J.M., Corbin, A., et al. (2019). Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat Commun* *10*, 45.

Mao, Y., Liu, H., Liu, Y., and Tao, S. (2014). Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. *Nucleic Acids Res* *42*, 4813–4822.

Marshall, E., Stansfield, I., and Romano, M.C. (2014). Ribosome recycling induces optimal translation rate at low ribosomal availability. *J R Soc Interface* *11*.

Martin, T.E., and Hartwell, L.H. (1970). Resistance of Active Yeast Ribosomes to Dissociation by KCl. *J. Biol. Chem.* *245*, 1504–1506.

Martinez, F.O., Gordon, S., Locati, M., and Mantovani, A. (2006). Transcriptional Profiling of the Human Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns of Gene Expression. *The Journal of Immunology* *177*, 7303–7311.

Mata, J., Marguerat, S., and Bähler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends in Biochemical Sciences* *30*, 506–514.

- Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J., et al. (2019). mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences* *116*, 24075–24083.
- McGeachy, A.M., and Ingolia, N.T. (2016). Starting too soon: upstream reading frames repress downstream translation. *EMBO J* *35*, 699–700.
- McManus, J., Cheng, Z., and Vogel, C. (2015). Next-generation analysis of gene expression regulation – comparing the roles of synthesis and degradation. *Mol Biosyst* *11*, 2680–2689.
- Medzhitov, R., and Horng, T. (2009). Transcriptional control of the inflammatory response. *Nat Rev Immunol* *9*, 692–703.
- Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., and Barkai, N. (2017). Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *ELife* *6*.
- MGlincy, N.J., and Ingolia, N.T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods* *126*, 112–129.
- Mills, E.W., Wangen, J., Green, R., and Ingolia, N.T. (2016). Dynamic Regulation of a Ribosome Rescue Pathway in Erythroid Cells and Platelets. *Cell Reports* *17*, 1–10.
- Mino, T., and Takeuchi, O. (2018). Post-transcriptional regulation of immune responses by RNA binding proteins. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* *94*, 248–258.
- Mohammad, F., Green, R., and Buskirk, A.R. (2019). A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *ELife* *8*, e42591.
- Mohammad, M.P., Munzarová Pondelícková, V., Zeman, J., Gunišová, S., and Valášek, L.S. (2017). In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. *Nucleic Acids Res.* *45*, 2658–2674.
- Molawi, K., and Sieweke, M.H. (2013). Chapter Ten - Transcriptional Control of Macrophage Identity, Self-Renewal, and Function. In *Advances in Immunology*, K.M. Murphy, and M. Merad, eds. (Academic Press), pp. 269–300.
- Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. *Cell* *45*, 201–207.
- Murray, P.J., and Wynn, T.A. (2011). Obstacles and opportunities for understanding macrophage polarization. *J Leukoc Biol* *89*, 557–563.
- Na, Y.R., Je, S., and Seok, S.H. (2018). Metabolic features of macrophages in inflammatory diseases and cancer. *Cancer Letters* *413*, 46–58.
- Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* *47*, 1588–1602.
- Noh, J.H., Kim, K.M., McClusky, W., Abdelmohsen, K., and Gorospe, M. (2018). Cytoplasmic functions of lncRNAs. *Wiley Interdiscip Rev RNA* *9*, e1471.
- Noller, H.F., Lancaster, L., Zhou, J., and Mohan, S. (2017). The Ribosome Moves: RNA Mechanics and Translocation. *Nat Struct Mol Biol* *24*, 1021–1027.

- Nottrott, S., Simard, M.J., and Richter, J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat. Struct. Mol. Biol.* *13*, 1108–1114.
- Oishi, Y., and Manabe, I. (2018). Macrophages in inflammation, repair and regeneration. *Int Immunol* *30*, 511–528.
- Otsuka, H., Fukao, A., Funakami, Y., Duncan, K.E., and Fujiwara, T. (2019). Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. *Front Genet* *10*.
- Ozadam, H., Geng, M., and Cenik, C. (2020). RiboFlow, RiboR and RiboPy: an ecosystem for analyzing ribosome profiling data at read length resolution. *Bioinformatics* *36*, 2929–2931.
- Palade, G.E. (1955). A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM. *J Biophys Biochem Cytol* *1*, 59–68.
- Park, S.-J., Onizuka, S., Seki, M., Suzuki, Y., Iwata, T., and Nakai, K. (2019). A systematic sequencing-based approach for microbial contaminant detection and functional inference. *BMC Biology* *17*, 72.
- Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* *161*, 1400–1412.
- Pérès, E., Blin, J., Ricci, E.P., Artesi, M., Hahaut, V., Van den Broeke, A., Corbin, A., Gazzolo, L., Ratner, L., Jalinet, P., et al. (2018). PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice. *PLoS Pathog.* *14*, e1006933.
- Petersen, C.P., Bordeleau, M.-E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* *21*, 533–542.
- Piccirillo, C.A., Bjur, E., Topisirovic, I., Sonenberg, N., and Larsson, O. (2014). Translational control of immune responses: from transcripts to translomes. *Nat Immunol* *15*, 503–511.
- Piro, A., Tagarelli, A., Tagarelli, G., Lagonia, P., and Quattrone, A. (2009). Archibald Edward Garrod: the physician father of biochemistry. *Metabolism - Clinical and Experimental* *58*, 427–437.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* *12*, 32–42.
- Pope, S.D., and Medzhitov, R. (2018). Emerging Principles of Gene Expression Programs and Their Regulation. *Molecular Cell* *71*, 389–397.
- Pradet-Balade, B., Boulmé, F., Beug, H., Müllner, E.W., and Garcia-Sanz, J.A. (2001). Translation control: bridging the gap between genomics and proteomics? *Trends in Biochemical Sciences* *26*, 225–229.
- Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* *160*, 1111–1124.
- Proud, C.G. (1994). Peptide-chain elongation in eukaryotes. *Mol Biol Rep* *19*, 161–170.
- Raacke, I.D., and Fiala, J. (1964). POLYRIBOSOME-BOUND NUCLEOSIDE TRIPHOSPHATASES IN ESCHERICHIA COLI\*. *Proc Natl Acad Sci U S A* *51*, 323–329.

- Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., and Collier, J. (2016). The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. *Cell* *167*, 122-132.e9.
- Riba, A., Di Nanni, N., Mittal, N., Arhné, E., Schmidt, A., and Zavolan, M. (2019). Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc Natl Acad Sci U S A* *116*, 15023–15032.
- Rodriguez, J.M., Rodriguez-Rivas, J., Di Domenico, T., Vázquez, J., Valencia, A., and Tress, M.L. (2018). APPRIS 2017: principal isoforms for multiple gene sets. *Nucleic Acids Res* *46*, D213–D217.
- Rogers, D.W., Böttcher, M.A., Traulsen, A., and Greig, D. (2017). Ribosome reinitiation can explain length-dependent translation of messenger RNA. *PLoS Comput Biol* *13*.
- Ruan, H., Brown, C.Y., and Morris, D.R. (1997). Chapter 16 - Analysis of Ribosome Loading onto mRNA Species: Implications for Translational Control. In *MRNA Formation and Function*, J.D. Richter, ed. (New York: Academic Press), pp. 305–321.
- Rüegsegger, U., Leber, J.H., and Walter, P. (2001). Block of HAC1 mRNA Translation by Long-Range Base Pairing Is Released by Cytoplasmic Splicing upon Induction of the Unfolded Protein Response. *Cell* *107*, 103–114.
- Saeed, S., Quintin, J., Kerstens, H.H.D., Rao, N.A., Aghajani-refah, A., Matarese, F., Cheng, S.-C., Ratter, J., Berentsen, K., van der Ent, M.A., et al. (2014). Epigenetic programming during monocyte to macrophage differentiation and trained innate immunity. *Science* *345*, 1251086.
- Santos, D.A., Shi, L., Tu, B.P., and Weissman, J.S. (2019). Cycloheximide can distort measurements of mRNA levels and translation efficiency. *Nucleic Acids Res* *47*, 4974–4985.
- Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nat Rev Mol Cell Biol* *19*, 526–541.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* *473*, 337–342.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2013). Correction: Corrigendum: Global quantification of mammalian gene expression control. *Nature* *495*, 126–127.
- Serhan, C.N., and Savill, J. (2005). Resolution of inflammation: the beginning programs the end. *Nature Immunology* *6*, 1191–1197.
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-Limiting Steps in Yeast Protein Translation. *Cell* *153*, 1589–1601.
- Shao, S., Brown, A., Santhanam, B., and Hegde, R.S. (2015). Structure and Assembly Pathway of the Ribosome Quality Control Complex. *Mol Cell* *57*, 433–444.
- Sharma, A.K., Sormanni, P., Ahmed, N., Ciryam, P., Friedrich, U.A., Kramer, G., and O'Brien, E.P. (2019). A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLOS Computational Biology* *15*, e1007070.

- Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* *15*, 1281.
- Shi, Z., and Barna, M. (2015). Translating the Genome in Time and Space: Specialized Ribosomes, RNA Regulons, and RNA-Binding Proteins. *Annual Review of Cell and Developmental Biology* *31*, 31–54.
- Sica, A., and Mantovani, A. (2012). Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest* *122*, 787–795.
- Siekevitz, P., and Zamecnik, P.C. (1981). Ribosomes and protein synthesis. *J Cell Biol* *91*, 53s–65s.
- Siomi, H., and Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. *Current Opinion in Genetics & Development* *7*, 345–353.
- Skabkin, M.A., Skabkina, O.V., Hellen, C.U.T., and Pestova, T.V. (2013). Reinitiation and Other Unconventional Posttermination Events during Eukaryotic Translation. *Molecular Cell* *51*, 249–264.
- Smale, S.T., and Natoli, G. (2014). Transcriptional Control of Inflammatory Responses. *Cold Spring Harbor Perspectives in Biology* *6*.
- Smale, S.T., Tarakhovsky, A., and Natoli, G. (2014). Chromatin Contributions to the Regulation of Innate Immunity. *Annu. Rev. Immunol.* *32*, 489–511.
- Sokabe, M., and Fraser, C.S. (2019). Toward a Kinetic Understanding of Eukaryotic Translation. *Cold Spring Harb Perspect Biol* *11*.
- Stein, K.C., and Frydman, J. (2019). The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *J Biol Chem* *294*, 2076–2084.
- Strauss, B.S. (2016). Beadle and Tatum and the origins of molecular biology. *Nat Rev Mol Cell Biol* *17*, 266–266.
- Strong, M.J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., Fewell, C., Taylor, C.M., and Flemington, E.K. (2014). Microbial Contamination in Next Generation Sequencing: Implications for Sequence-Based Analysis of Clinical Samples. *PLoS Pathog* *10*.
- Strunk, B.S., Novak, M.N., Young, C.L., and Karbstein, K. (2012). Joining of 60S subunits and a translation-like cycle in 40S ribosome maturation. *Cell* *150*, 111–121.
- Su, X., Yu, Y., Zhong, Y., Giannopoulou, E.G., Hu, X., Liu, H., Cross, J.R., Rättsch, G., Rice, C.M., and Ivashkiv, L.B. (2015). Interferon- $\gamma$  regulates cellular metabolism and mRNA translation to potentiate macrophage activation. *Nat Immunol* *16*, 838–849.
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* *6*, e21800.
- Takeuchi, O., and Akira, S. (2010). Pattern Recognition Receptors and Inflammation. *Cell* *140*, 805–820.

- Trouplin, V., Boucherit, N., Gorvel, L., Conti, F., Mottola, G., and Ghigo, E. (2013). Bone Marrow-derived Macrophage Production. *J Vis Exp*.
- Ts'o, P.O.P. (1962). The Ribosomes-Ribonucleoprotein Particles. *Annual Review of Plant Physiology* *13*, 45–80.
- Tuck, A.C., Rankova, A., Arpat, A.B., Liechti, L.A., Hess, D., Iesmantavicius, V., Castelo-Szekely, V., Gatfield, D., and Bühler, M. (2020). Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Mol Cell* *77*, 1222-1236.e13.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* *141*, 344–354.
- Turner, M., and Díaz-Muñoz, M.D. (2018). RNA-binding proteins control gene expression and cell fate in the immune system. *Nature Immunology* *1*.
- Valleriani, A., and Chiarugi, D. (2020). A workbench for the translational control of gene expression. *BioRxiv* 2020.01.28.923219.
- Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *PNAS* *101*, 11269–11274.
- Viero, G., Lunelli, L., Passerini, A., Bianchini, P., Gilbert, R.J., Bernabò, P., Tebaldi, T., Diaspro, A., Pederzoli, C., and Quattrone, A. (2015). Three distinct ribosome assemblies modulated by translation are the building blocks of polysomes. *J Cell Biol* *208*, 581–596.
- Vind, J., Sørensen, M.A., Rasmussen, M.D., and Pedersen, S. (1993). Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.* *231*, 678–688.
- Vindry, C., Ohlmann, T., and Chavatte, L. (2018). Translation regulation of mammalian selenoproteins. *Biochimica et Biophysica Acta (BBA) - General Subjects* *1862*, 2480–2492.
- Virág, L., Jaén, R.I., Regdon, Z., Boscá, L., and Prieto, P. (2019). Self-defense of macrophages against oxidative injury: Fighting for their own survival. *Redox Biol* *26*.
- Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* *13*, 227–232.
- Wang, C., Yu, X., Cao, Q., Wang, Y., Zheng, G., Tan, T.K., Zhao, H., Zhao, Y., Wang, Y., and Harris, D.C. (2013a). Characterization of murine macrophages from bone marrow, spleen and peritoneum. *BMC Immunology* *14*, 6.
- Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology* *15*, e8503.
- Wang, T., Cui, Y., Jin, J., Guo, J., Wang, G., Yin, X., He, Q.-Y., and Zhang, G. (2013b). Translating mRNAs strongly correlates to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res* *41*, 4743–4754.
- Warner, J.R., and Knopf, P.M. (2002). The discovery of polyribosomes. *Trends in Biochemical Sciences* *27*, 376–380.

Warner, J.R., Knopf, P.M., and Rich, A. (1963). A Multiple Ribosomal Structure in Protein Synthesis. *PNAS* *49*, 122–129.

Watanabe, S., Alexander, M., Misharin, A.V., and Budinger, G.R.S. (2019). The role of macrophages in the resolution of inflammation. *J Clin Invest* *129*, 2619–2628.

Watson, J.D., and Crick, F.H.C. (1953a). Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* *171*, 964–967.

Watson, J.D., and Crick, F.H.C. (1953b). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* *171*, 737–738.

Weatheritt, R.J., Sterne-Weiler, T., and Blencowe, B.J. (2016). The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol* *23*, 1117–1123.

Wegler, C., Ölander, M., Wiśniewski, J.R., Lundquist, P., Zettl, K., Åsberg, A., Hjølmesæth, J., Andersson, T.B., and Artursson, P. (2020). Global variability analysis of mRNA and protein concentrations across and within human tissues. *NAR Genom Bioinform* *2*.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Reports* *14*, 1787–1799.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* *2*, 135–140.

Wettstein, F.O., Staehelin, T., and Noll, H. (1963). Ribosomal Aggregate Engaged in Protein Synthesis: Characterization of the Ergosome. *Nature* *197*, 430–435.

Wu, C.C.-C., Zinshteyn, B., Wehner, K.A., and Green, R. (2019). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Molecular Cell* *73*, 959-970.e5.

Xie, J., Alves, V. de S., Haar, T. von der, O’Keefe, L., Lenchine, R.V., Jensen, K.B., Liu, R., Coldwell, M.J., Wang, X., and Proud, C.G. (2019). Regulation of the Elongation Phase of Protein Synthesis Enhances Translation Accuracy and Modulates Lifespan. *Current Biology* *29*, 737-749.e5.

Young, D.J., Guydosh, N.R., Zhang, F., Hinnebusch, A.G., and Green, R. (2015). Rli1/ABCE1 recycles terminating ribosomes and controls translation reinitiation in 3’UTRs in vivo. *Cell* *162*, 872–884.

Z, S., K, F., Km, K., Nr, G., Hi, R., Mn, T., and M, B. (2017). Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol Cell* *67*, 71-83.e7.

Zhang, X., Chen, X., Liu, Q., Zhang, S., and Hu, W. (2017). Translation repression via modulation of the cytoplasmic poly(A)-binding protein in the inflammatory response. *ELife* *6*, e27786.

Zhou, J., Wan, J., Shu, X.E., Mao, Y., Liu, X.-M., Yuan, X., Zhang, X., Hess, M.E., Brüning, J.C., and Qian, S.-B. (2018). N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. *Mol. Cell* *69*, 636-647.e7.

Zinshteyn, B., Wangen, J.R., Hua, B., and Green, R. (2020). Nuclease-mediated depletion biases in ribosome footprint profiling libraries. *BioRxiv* 2020.03.30.017061.

Zong, Q., Schummer, M., Hood, L., and Morris, D.R. (1999). Messenger RNA translation state: The second dimension of high-throughput expression screening. *Proc Natl Acad Sci U S A* *96*, 10632–10636.

Zylber, E.A., and Penman, S. (1970). The effect of high ionic strength on monomers, polyribosomes, and puromycin-treated polyribosomes. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* *204*, 221–229.



# Annex 1

## **Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins.**

CRISPR-Cas9 system enables rapid gene-editing in a wide range of target cells by the combined action of a single-end guide RNA (sgRNA) that directs the Cas9 endonuclease to specific DNA sequences, which are complementary to the sgRNA, to induce double-strand breaks. These breaks are mainly repaired through the error-prone Non-Homologous End Joining (NHEJ) cellular pathway leading to small mutations or indels in the targeted sequences that can inactivate gene expression when occurring within the reading frame of a gene. Classical procedures consist of transfecting cells with the different components of CRISPR machinery. However, this approach is difficult to apply to cells that are refractory to transfection such as primary macrophages. One way to overcome this issue is the use of lentiviral vectors to deliver a transgene coding for Cas9 and gRNAs into the cells. However, this strategy has adverse effects as the transgene can integrate within cellular genes potentially introducing a bias in gene expression. Therefore, new approaches were needed to deliver CRISPR components in sensitive cells in an efficient and non-toxic manner.

This work describes the development of an innovative CRISPR strategy based on the delivery of Cas9 protein and gRNA by noninfectious virus-like particles (Nanoblades). This method is based on the observation that HEK-293T cells over-expressing the glycoprotein of the Vesicular Stomatitis Virus (VSV-G) and the Murine Leukemia Virus (MLV) retroviral protein GAG produce fusogenic vesicles that can incorporate proteins co-expressed by the producer cell and deliver them into target cells. Nanoblades are thus less toxic than the other classical approaches for CRISPR-Cas9 delivery as they do not contain integrative genetic material and mediate transient protein delivery. Additionally, Nanoblades are pseudotyped with an additional viral envelope, the baboon retroviral envelope glycoprotein (BaEV), to improve the efficiency of their fusion with primary cells.





My contribution to this work was to assess the efficiency of this tool for CRISPR mediated gene-editing in mouse primary macrophages. For this, I optimized the protocol to transduce bone-marrow derived precursors from GFP transgenic mice using Nanoblades loaded with a gRNA targeting the GFP coding sequence. The efficiency of the knock-out was validated by fluorescence microscopy, flow cytometry and T7 endonuclease assay. I also checked that the treatment with Nanoblades early during the differentiation process did not disturb the ability of the precursors to generate functional macrophages. Finally, I compared the gene editing efficiency when primary mouse bone-marrow precursors were treated using Nanoblades or by electroporation of Cas9-sgRNA ribonucleoparticles.

ARTICLE

<https://doi.org/10.1038/s41467-018-07845-z>

OPEN

# Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins

Philippe E. Mangeot<sup>1</sup>, Valérie Risson<sup>2</sup>, Floriane Fusil<sup>1</sup>, Aline Marnef<sup>3</sup>, Emilie Laurent<sup>1</sup>, Juliana Blin <sup>1</sup>, Virginie Mournetas<sup>4</sup>, Emmanuelle Massouridès <sup>4</sup>, Thibault J.M. Sohier <sup>1</sup>, Antoine Corbin<sup>1</sup>, Fabien Aubé<sup>5</sup>, Marie Teixeira<sup>6</sup>, Christian Pinset<sup>4</sup>, Laurent Schaeffer<sup>2</sup>, Gaëlle Legube<sup>3</sup>, François-Loïc Cosset<sup>1</sup>, Els Verhoeyen<sup>1,7</sup>, Théophile Ohlmann<sup>1</sup> & Emiliano P. Ricci <sup>1,5</sup>

Programmable nucleases have enabled rapid and accessible genome engineering in eukaryotic cells and living organisms. However, their delivery into target cells can be technically challenging when working with primary cells or in vivo. Here, we use engineered murine leukemia virus-like particles loaded with Cas9-sgRNA ribonucleoproteins (Nanoblades) to induce efficient genome-editing in cell lines and primary cells including human induced pluripotent stem cells, human hematopoietic stem cells and mouse bone-marrow cells. Transgene-free Nanoblades are also capable of in vivo genome-editing in mouse embryos and in the liver of injected mice. Nanoblades can be complexed with donor DNA for “all-in-one” homology-directed repair or programmed with modified Cas9 variants to mediate transcriptional up-regulation of target genes. Nanoblades preparation process is simple, relatively inexpensive and can be easily implemented in any laboratory equipped for cellular biology.

<sup>1</sup>CIRI, Centre International de Recherche en Infectiologie Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007 Lyon, France. <sup>2</sup>Institut NeuroMyoGène, CNRS 5310, INSERM U121, Université Lyon 1, Faculté de Médecine Lyon Est, Lyon 69008, France. <sup>3</sup>LBCMCP, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, UT3, 118 Route de Narbonne, 31062 Toulouse, France. <sup>4</sup>I-STEM/CECS, Inserm, UMR861 28 rue Henri Desbruères, 91100 Corbeil Essonnes, France. <sup>5</sup>LBMC, Laboratoire de Biologie et Modélisation de la Cellule Univ Lyon, ENS de Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5239, INSERM, U1210, Lyon 69007, France. <sup>6</sup>SFR BioSciences, Plateau de Biologie Expérimentale de la Souris (AniRA-PBES), Ecole Normale Supérieure de Lyon, Université Lyon1, CNRS UMS3444 INSERM US8, 69007 Lyon, France. <sup>7</sup>Present address: CIRI, Université Côte d’Azur, INSERM, C3M, 06204 Nice, France. These authors contributed equally: Valérie Risson, Floriane Fusil, Aline Marnef. Correspondence and requests for materials should be addressed to P.E.M. (email: [philippe.mangeot@inserm.fr](mailto:philippe.mangeot@inserm.fr)) or to E.P.R. (email: [emiliano.ricci@ens-lyon.org](mailto:emiliano.ricci@ens-lyon.org))

Targeted genome editing tools, such as meganucleases (MGN), zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs) and more recently the clustered regularly interspaced short palindromic repeats (CRISPR) have revolutionized most biomedical research fields. Such tools allow to precisely edit the genome of eukaryotic cells by inducing double-stranded DNA (dsDNA) breaks at specific loci. Relying on the cell endogenous repair pathways, dsDNA breaks can then be repaired by non-homologous end-joining (NHEJ) or homology-directed repair (HDR) allowing the removal or insertion of new genetic information at a desired locus.

Among the above-mentioned tools, CRISPR-Cas9 is currently the most simple and versatile method for genome engineering. Indeed, in the two-component system, the bacterial-derived nuclease Cas9 (for CRISPR-associated protein 9) associates with a single-guide RNA (sgRNA) to target a complementary DNA sequence and induce a dsDNA break<sup>1</sup>. Therefore, by the simple modification of the sgRNA sequence, users can specify the genomic locus to be targeted. Consistent with the great promises of CRISPR-Cas9 for genome engineering and gene therapy, considerable efforts have been made in developing efficient tools to deliver the Cas9 and the sgRNA into target cells *ex vivo* either by transfection of plasmids coding for the nucleases, transduction with viral-derived vectors coding for the nucleases or by direct injection or electroporation of Cas9-sgRNA complexes into cells.

Here, we have designed Nanoblades, a protein-delivery vector based on friend murine leukemia virus (MLV) that allows the transfer of Cas9-sgRNA ribonucleoproteins (RNPs) to cell lines and primary cells *in vitro* and *in vivo*. Nanoblades deliver the ribonucleoprotein cargo in a transient and rapid manner without delivering a transgene and can mediate knock-in in cell lines when complexed with a repair template. Nanoblades can also be programmed with modified Cas9 proteins to mediate transient transcriptional activation of targeted genes.

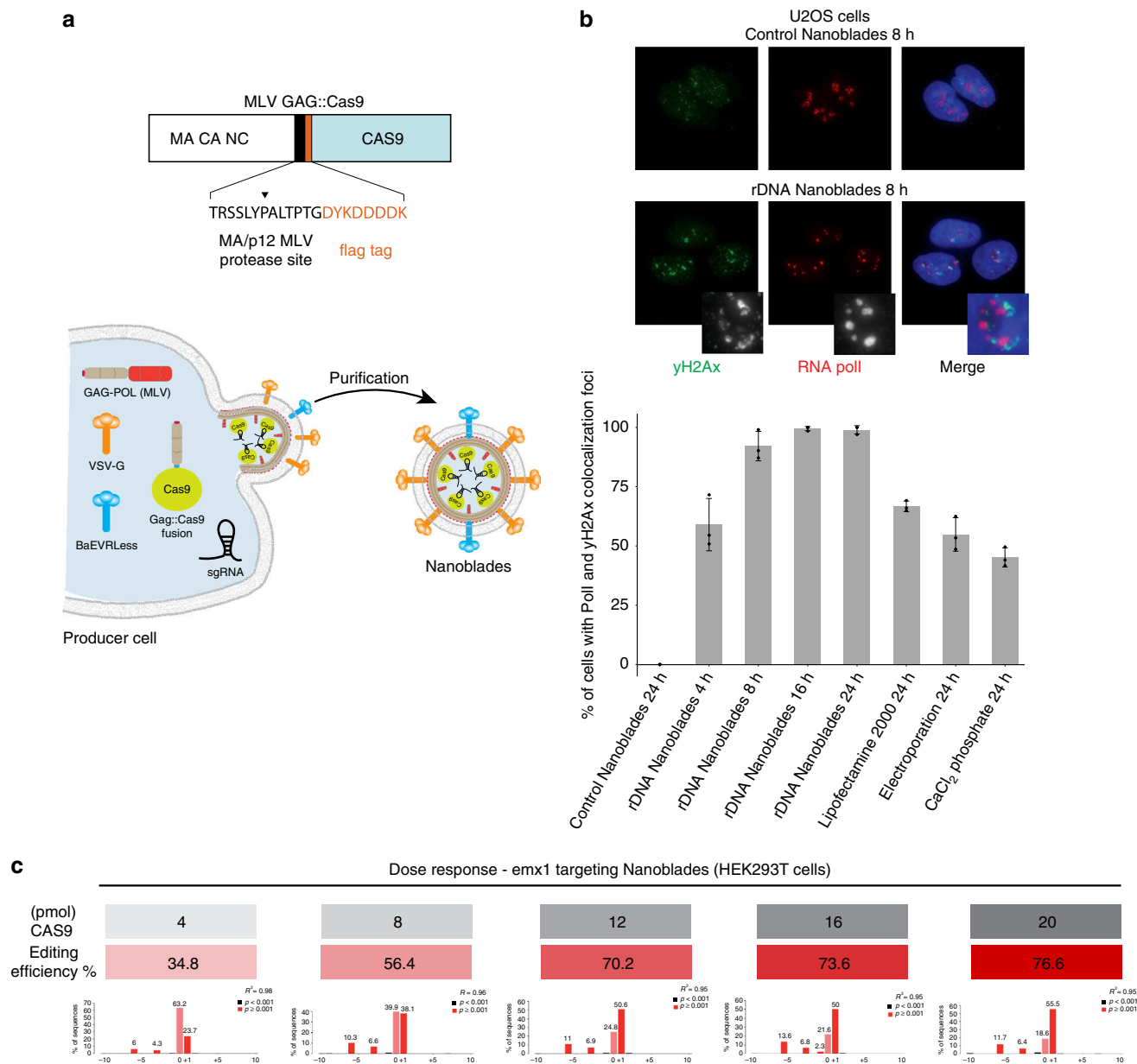
## Results

**Cas9-sgRNA RNP delivery through MLV virus-like particles (VLPs).** Assembly of retroviral particles relies on the viral structural Gag polyprotein, which multimerizes at the cell membrane and is sufficient, when expressed in cultured cells, to induce release of VLPs into the cell supernatant<sup>2</sup>. When Gag is coexpressed together with a fusogenic viral envelope, pseudotyped VLPs are produced that lack a viral genome but still retain their capacity to fuse with target cells and deliver the Gag protein into their cytoplasm. As previously investigated<sup>3,4</sup>, we took advantage of the structural role of Gag and designed an expression vector coding for the MLV Gag polyprotein fused, at its C-terminal end, to a flag-tagged version of *Streptococcus pyogenes* Cas9 protein (Gag::Cas9, Fig. 1a). The two fused proteins are separated by a proteolytic site which can be cleaved by the MLV protease to release the Flag-tagged Cas9 (Fig. 1a). By cotransfecting HEK-293T cells with plasmids coding for Gag::Cas9, Gag-Pro-Pol, a sgRNA, and viral envelopes, fusogenic VLPs are produced and released in the culture medium (herein described as Nanoblades). Biochemical and imaging analysis of purified particles (Supplementary Figure 1a, 1b, 1c and 1d) indicates that Nanoblades (150 nm) are slightly larger than wild-type MLV (Supplementary Figure 1b) but sediment at a density of 1.17 g/ml (Supplementary Figure 1c) as described for MLV VLPs<sup>5</sup>. As detected by western blot, Northern blot, mass-spectrometry, and deep-sequencing, Nanoblades contain the Cas9 protein and sgRNA (Supplementary Figure 1 and 2 and Supplementary Data 1). In addition to Gag, Cas9 and envelope proteins,

mass-spectrometry analysis of Nanoblades identified several cellular proteins, mostly membrane-associated proteins (Supplementary Figure 2a and Supplementary Data 1). Interestingly, the packaging of sgRNA depends on the presence of the Gag::Cas9 fusion protein, since Nanoblades produced from cells that only express the Gag protein fail to incorporate detectable amounts of sgRNA (Supplementary Figure 1d). Furthermore, Cas9-dependent loading of the sgRNA within Nanoblades is not limited by the efficiency of the interaction between the Cas9 and the sgRNA, since expressing an optimized version of the sgRNA that improves binding to Cas9<sup>6</sup> does not appear to increase sgRNA levels within purified VLPs (Supplementary Figure 1d see sgRNA(F+E)).

To assess for Cas9-sgRNA RNP delivery efficiency in target cells and induction of genomic dsDNA breaks, we designed Nanoblades with a sgRNA targeting the 45S rDNA loci. Human 45S rDNA genes are present in hundreds of tandem repeats across five autosomes, locate in the nucleolus and are transcribed exclusively by RNA polymerase (Pol) I<sup>7</sup>. Using immunofluorescence microscopy, it is therefore possible to follow the occurrence of dsDNA breaks at rDNA loci with single-cell resolution by monitoring the nucleolus using the nucleolar marker RNA Pol I and the well-established dsDNA break-marker, histone variant  $\gamma$ -H2AX<sup>8</sup>, that localizes at the nucleolar periphery after dsDNA break induction within rDNA<sup>9</sup>. U2OS (osteosarcoma cell line) cells transduced for 24 h with Nanoblades programmed with a sgRNA targeting rDNA display the typical  $\gamma$ -H2AX distribution at the nucleolar periphery with RNA Pol I, indicative of rDNA breaks, whilst cells transduced with Nanoblades with control sgRNAs do not (Fig. 1b, top panel). Interestingly, this distribution of  $\gamma$ -H2AX at the nucleolar periphery can be observed as early as 4 h after transduction in 60% of cells with a maximum effect observed at 16 h after transduction, where almost 100% of observed cells display this  $\gamma$ -H2AX distribution (Fig. 1b, bottom panel and quantification below). In comparison, only 60% of cells transfected with a plasmid coding for Cas9 and the sgRNA display the perinucleolar  $\gamma$ -H2AX/RNA Pol I localization 24 h after transfection. Similar results were obtained in human primary fibroblasts with more than 85% cells displaying this distribution after 16 h (Supplementary Figure 1e). These results suggest that Nanoblade-mediated delivery of the Cas9-sgRNA RNP is both efficient and rapid in cell lines and primary human cells. To further confirm these results, we designed and dosed Nanoblades (by ELISA assay using anti-Cas9 antibodies) programmed with a sgRNA widely used in the literature<sup>10</sup> that targets the human *EMX1* gene to induce dsDNA cleavage at a single locus. HEK-293T cells were then transduced with increasing amounts of Nanoblades and gene editing was measured from the bulk population 48 h after transduction (Fig. 1c). Under these conditions, we observed a dose-dependent effect of Nanoblades ranging from 35% of *EMX1* (at 4 pmol of Cas9) editing to 77% of editing at the highest dose (20 pmol) of Cas9 (Fig. 1c).

Because Nanoblades carry cellular proteins from producer cells in addition to Cas9 (Supplementary Data 1), we tested whether these proteins could also be delivered to recipient cells. For this, we over-expressed the firefly luciferase in producer cells and collected Nanoblades targeting *EMX1* from the supernatant. Luciferase-loaded Nanoblades were then used to transduce HEK293T cells for 24 h. Cells were then washed twice in PBS and incubated in fresh medium for 4, 8, 24, and 48 h. Luciferase activity was measured at each time point, as well as in input Nanoblades (Supplementary Figure 2c). As observed, we could detect a mild luciferase signal (4–6% of input) at 4 and 8 h upon transduction. However, the signal rapidly faded at 24 h (2% of input) and

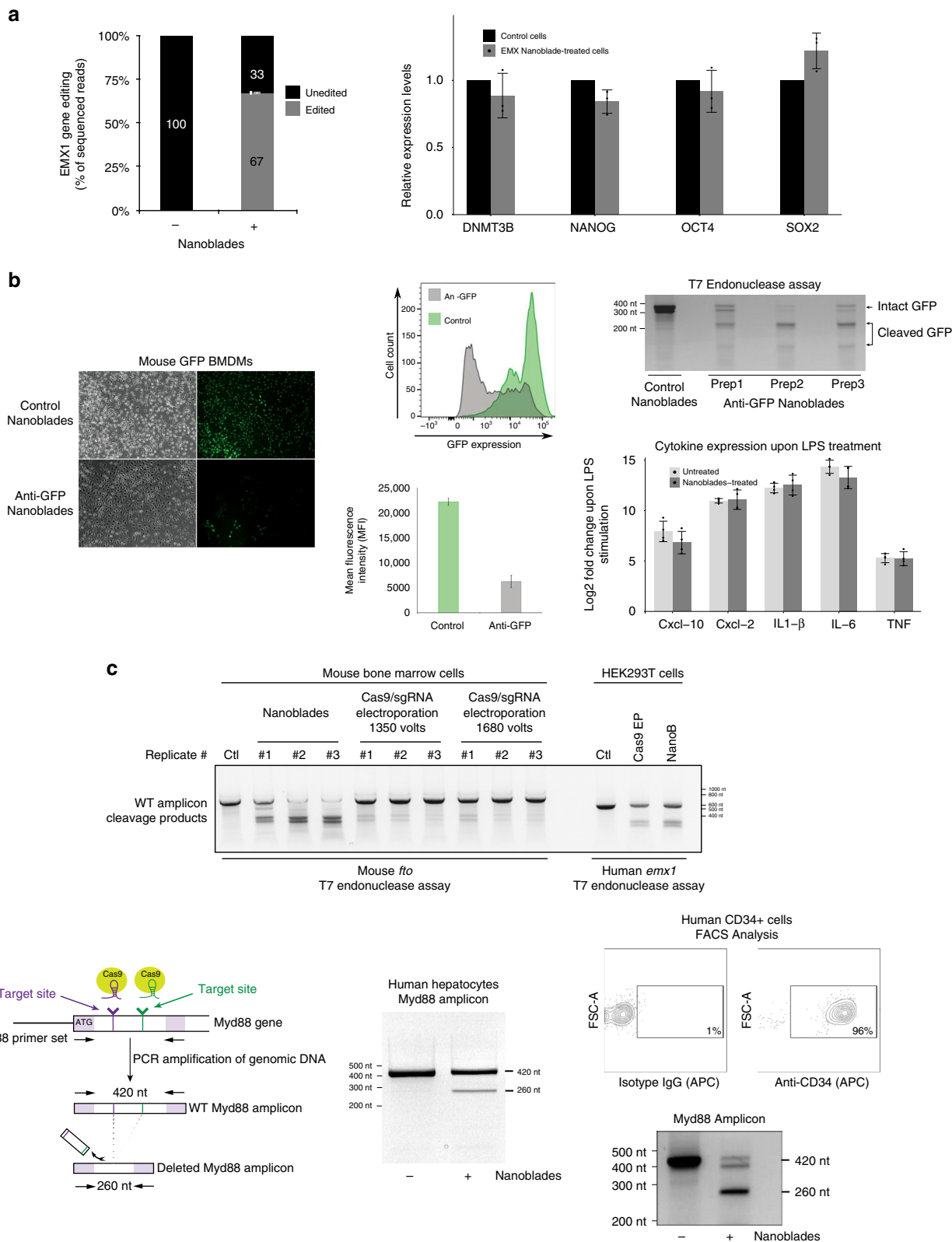


**Fig. 1** Nanoblade-mediated genome editing. **a** Scheme describing the MLV Gag::Cas9 fusion and the Nanoblade production protocol based on the transfection of HEK-293T cells by plasmids coding for Gag-Pol, Gag::Cas9, VSV-G, BaEVRLess, and the sgRNA. **b** Top panel, immunofluorescence analysis of  $\gamma$ -H2AX (green), RNA poll (red) in U2OS cells 8 h after being transduced with control Nanoblades or with Nanoblades targeting ribosomal DNA genes. Bottom panel, quantification of  $\gamma$ -H2AX and RNA poll colocalization foci in U2OS cells at different times after Nanoblades transduction or after classical DNA transfection methods ( $n = 3$ , error bars correspond to standard deviation). **c** Dose response of Nanoblades. HEK-293T cells were transduced with increasing amounts of Nanoblades targeting human *EMX1* ( $n = 1$  displayed). The exact amount of Cas9 used for transduction was measured by dot blot (in gray). Genome editing was assessed by Sanger sequencing and Tide analysis (in red)

was almost undetectable at 48 h (Supplementary Figure 2c). In addition to the ectopically expressed firefly luciferase, we also investigated transmission of the CD81 cell-surface protein, which is highly expressed in HEK293T producer cells and is present in Nanoblades as revealed by mass spectrometry (Supplementary Data 1). HepG2 cells, a hepatic cell line that lacks CD81 expression<sup>11</sup>, were transduced for 24 h with Nanoblades targeting *EMX1* and then washed twice with PBS before monitoring CD81 residual signal immediately after the washes or 8 and 48 h after incubation with fresh medium (Supplementary Figure 2d). As observed, even though CD81 was very abundant at the cell surface of producer cells

and completely absent in recipient cells (Supplementary Figure 2d, left and middle panels), we could only detect a mild CD81 signal immediately after transduction (see Supplementary Figure 2d, right panel). Later time points (8 and 48 h) did not show any specific CD81 labeling in recipient HepG2 cells. The impact of cellular proteins delivered by Nanoblades into recipient cells appears therefore limited and restricted to a short time frame.

Taken together, our results indicate that Nanoblades can be efficiently used to mediate genome editing in a rapid and dose-dependent manner with limited impact on the proteome of target cells.



**Nanoblades-mediated genome editing in primary cells.** Genome editing in primary cells and patient-derived pluripotent cells represents a major interest both for basic science and therapeutical applications. However, primary cells are often refractory to DNA transfection and other gene delivery methods. Because Nanoblades were capable of efficient delivery of functional Cas9-sgRNA RNPs into primary fibroblasts, we tested whether they

were effective in other primary cells for genome editing. To this aim, Nanoblades targeting *EMX1* were used to transduce human-induced pluripotent stem cells (hiPSCs). Genome editing at the *EMX1* locus was assessed in the bulk cellular population 48 h after transduction by deep-sequencing of the *EMX1* locus (Fig. 2a, left panel). As observed, Nanoblades were capable of mediating 67% genome editing at the *EMX1* locus in hiPSCs. Notably, hiPSCs



**Fig. 2** Genome editing in primary cells transduced with Nanoblades. **a** Left panel, editing efficiency at the *EMX1* locus (measured by high-throughput sequencing on the Illumina Miseq platform) of human-induced pluripotent stem cells (hiPSCs) transduced with Nanoblades targeting human *EMX1* ( $n = 3$ ). Right panel, expression of pluripotency markers measured by qPCR in control cells and cells transduced with Nanoblades targeting *EMX1* ( $n = 3$ ). **b** Left and middle panels, fluorescence microscopy and FACS analysis of GFP expressing BMDMs transduced at the bone marrow stage (day 0 after bone marrow collection) with control Nanoblades or Nanoblades targeting the *GFP*-coding sequence ( $n = 3$ ). Right top panel, T7 endonuclease assay against the GFP sequence from Nanoblades-treated BMDMs. Right bottom panel, cytokine expression levels (measured by qPCR) in untreated or Nanoblade-treated cells upon LPS stimulation ( $n = 4$ ). **c** T7 endonuclease assay against mouse *Fto* or human *EMX1* genomic sequences amplified by PCR from primary mouse bone marrow cells transduced with Nanoblades or electroporated with recombinant Cas9-sgRNA RNPs. For bone marrow cells, two electroporation settings were tested. Lanes numbered #1–#3 correspond to biological replicates. Editing efficiencies were calculated by TIDE<sup>13</sup> analysis of the Sanger sequencing electropherograms for each PCR amplicon **d** Left panel, excision of a 160 bp DNA fragment of *MYD88* using Nanoblades. Middle panel PCR results obtained in human primary hepatocytes transduced with Nanoblades. Right-panel (top), FACS analysis of CD34+ cells purified from human cord-blood. Bottom, genome editing at the *MYD88* locus assessed by PCR in untreated and Nanoblades-treated CD34+ cells. Error bars in all figures correspond to standard deviation

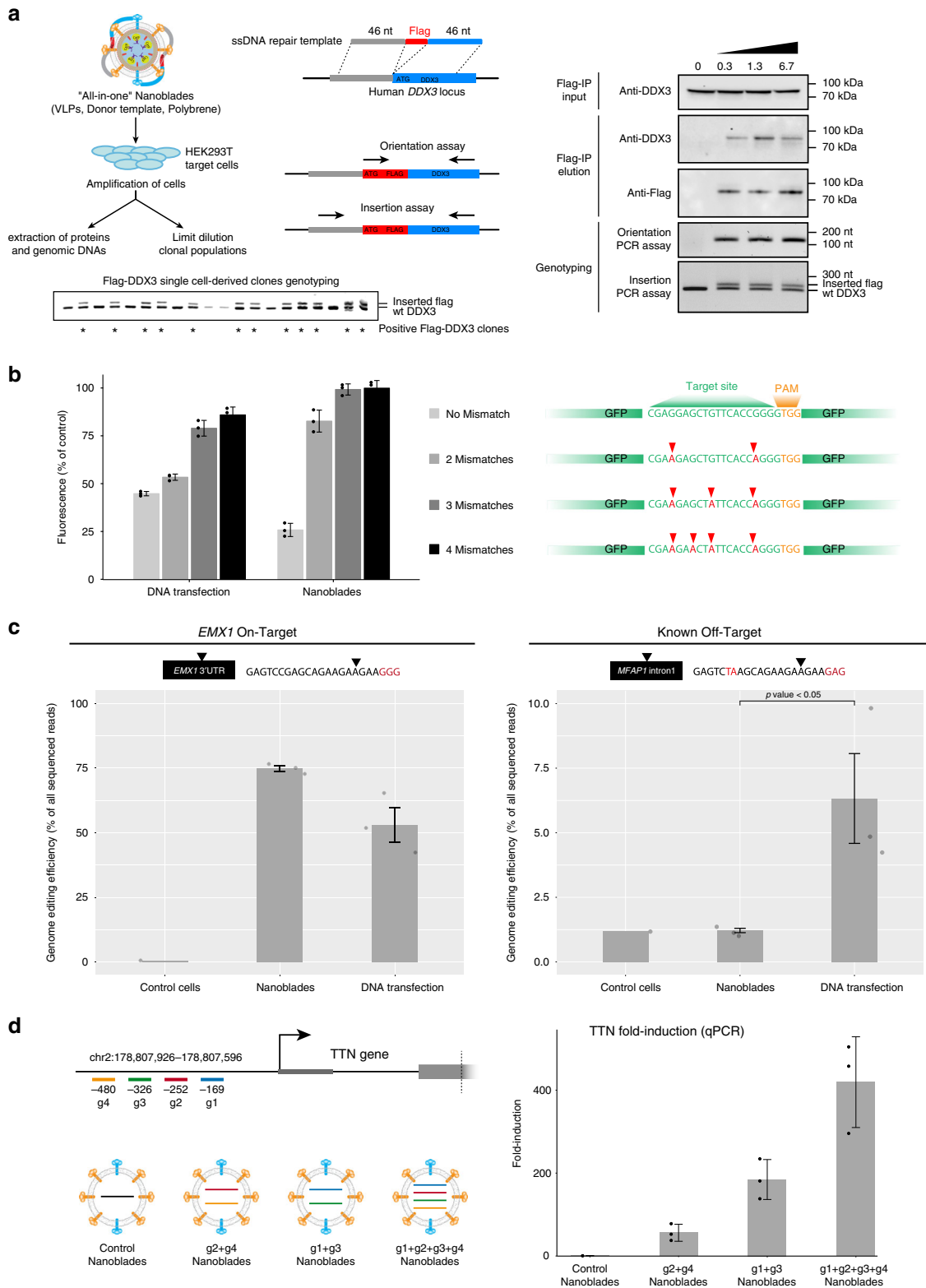
treated with *EMX1* Nanoblades maintained constant levels of pluripotency markers compared to control cells (Fig. 2a, right panel) thus indicating that their multipotent status did not appear to be affected.

Similarly to hiPSCs, mouse bone marrow (BM) cells can be collected and differentiated in vitro into various hematopoietic cell types, such as macrophages (bone marrow-derived macrophages or BMDMs) and dendritic cells. Efficient genome editing of specific genes in BM cells would therefore allow for the corresponding pre-existing protein to be degraded during differentiation and obtain a functional knockout. To test this hypothesis, BM cells obtained from GFP transgenic mice<sup>12</sup> were transduced with Nanoblades programmed with a sgRNA targeting the *GFP* coding sequence. 6 h after transduction, cells were washed and incubated in presence of macrophage colony-stimulating factor (M-CSF) for 1 week. After this, cells were collected to monitor GFP levels by fluorescence microscopy, FACS and genome editing by T7 endonuclease assay (Fig. 2b). We consistently obtained close to 75% reduction of GFP expression as measured by FACS analysis and around 60–65% genome editing at the *GFP* locus as measured by T7 endonuclease assays (Fig. 2b). Importantly, genome editing through Nanoblades did not affect the capacity of BMDMs to respond to LPS as their cytokine expression remains identical to that of untreated control cells (Fig. 2b bottom right panel). Nanoblades can therefore be used to inactivate genes in BM cells and study their function in differentiated cells. To further complement these results, we compared the efficiency of Nanoblades to that of recombinant Cas9-sgRNA RNP electroporation in targeting an endogenous gene in primary mouse BM cells. For this, Nanoblades or Cas9-sgRNA RNPs programmed to target the *Fto* gene were used, respectively, to transduce or electroporate primary BM cells freshly extracted from mice. As a control, Nanoblades or Cas9-sgRNA RNPs programmed to target human *EMX1* were also tested in HEK293T cells. In both cases, the efficiency of genome editing was assessed 24 h after transduction or electroporation. As observed (Fig. 2c), both Nanoblades and Cas9-sgRNA electroporation mediate efficient genome editing in HEK293T at 71% (Nanoblades) and 44% (Electroporation) of editing efficiency at the *EMX1* locus. Interestingly, in primary BM cells, while Nanoblades achieve highly efficient genome editing of the *Fto* locus (up to 76% as measured by TIDE<sup>13</sup> analysis), Cas9 electroporation was much less efficient at both conditions that we tested (1350 and 1680 V) yielding a mild but visible signal in the T7 endonuclease assay which was nevertheless below the detection limit for TIDE analysis. Interestingly both protocols (Nanoblades and protein electroporation) did not have an important impact on cell viability 24 h after Cas9 delivery (Supplementary Figure 2e).

Nanoblades efficiency was also investigated in human cells that represent a major interest in research and gene therapy like human primary hepatocytes and human hematopoietic stem cells (HSCs) that both have the capacity to colonize and regenerate fully functional tissues. For both these cell types, Nanoblades programmed with two sgRNAs targeting the human *Myd88* gene were prepared and achieved significant cleavage efficiencies, as revealed by flanking PCR assays (Fig. 2d). Interestingly, HSCs are difficult to transduce with classic VSV-G pseudotyped lentiviral vectors (LVs) because they lack the LDL receptor<sup>14</sup>, a limitation that can be alleviated by the use of the baboon retroviral envelope glycoprotein (BaEV)<sup>15</sup>. This prompted us to equip Nanoblades with both BaEV and VSV G-envelopes for these cells and finally in all our study as the combination of both envelopes improved Cas9 delivery in most cells (Supplementary Figure 6a and b). As observed, Nanoblades were also able to induce genome editing in these cells (50% genome editing based on T7 endonuclease assay, Fig. 2d) thus expanding the catalog of primary cells that can be edited using Nanoblades.

Taken together, our results indicate that Nanoblades are an efficient delivery system to induce rapid and effective genome editing in murine and human primary cells of high therapeutic value that are notoriously difficult to transfect.

**“All-in-one” Nanoblades for homology directed repair.** Precise insertion of genetic material (also known as Knock-in) using CRISPR-Cas9 can be achieved through HDR. This occurs when a donor DNA template with sequence homology to the region surrounding the targeted genomic locus is provided to cells together with the Cas9-sgRNA RNP. Based on a previous finding showing that retroviral-particles can be complexed with DNA in the presence of polybrene to allow for virus-dependent DNA transfection<sup>16</sup>, we tested whether Nanoblades could be directly complexed with a DNA template to mediate HDR in target cells. To test this approach, Nanoblades programmed to target a locus close to the AUG start codon of the human *DDX3* gene were complexed to a single-stranded DNA oligomer bearing the FLAG-tag sequence flanked with 46 nucleotide (nt) homology arms corresponding to the region surrounding the start-codon of *DDX3* (Fig. 3a, left panel). HEK293T were transduced with these “All-in-one” Nanoblades and passed 6 times before assessing HDR efficiency in the bulk cellular population both by PCR and by Flag-immunoprecipitation followed by western-blotting (using a *DDX3* and FLAG-antibody). As observed (Fig. 3a, right panel), cells transduced with “All-in-one” Nanoblades showed incorporation of the FLAG-tag at the *DDX3* locus both genetically and at the level of protein expression (Fig. 3a right panel, see Flag-IP elution and Genotyping panels). In parallel, single-cell



clones were derived from the Flag-DDX3 bulk population and tested for Flag incorporation by PCR. As shown (Fig. 3a left bottom panel), 12 out of 20 isolated clones displayed incorporation of the Flag-sequence at the DDX3 locus thus suggesting a knock-in efficiency of more than 50% of cells using “all-in-one” Nanoblades.

Knock-in assisted by “All-in-one” Nanoblades was also obtained at the AAVS1 locus which has been described as a safe

harbor for transgene insertion<sup>17</sup>. For this we designed a dsDNA template of 4 kb bearing the puromycin resistance gene with homology arms to the AAVS1 locus. After transduction of HEK-293T cells with Nanoblades complexed with this template using polybrene, single-cell-derived clones were selected with puromycin. Out of  $1 \times 10^5$  transduced cells, we obtained 47 puromycin-resistant clones (Supplementary Figure 3b, c and d). A PCR-assay revealed that 42 out of 47 puromycin-resistant clones tested had

**Fig. 3** “All-in-one” Nanoblades for knock-in experiments and assessment of Nanoblades off-target activity. **a** Left panel, Nanoblades targeting human *DDX3* close to its start codon were complexed with a donor ssDNA bearing homology arms to the targeted locus and a Flag-tag sequence in the presence of polybrene. HEK293T cells were then transduced with these “All-in-one” Nanoblades. After cell amplification, a fraction of cells were collected to extract genomic DNA and total proteins while the remaining cells were cultured to obtain single-cell clonal populations. Right panel, insertion of the Flag-tag in HEK-293T cells transduced with “all-in-one” Nanoblades complexed with increasing amounts of donor ssDNA was assessed by Flag-immunoprecipitation followed by western-blot using anti-flag or anti-DDX3 antibodies in the input and Flag-immunoprecipitation elution fractions. Flag insertion was also assessed by PCR using a forward primer in the flag-sequence and a reverse primer in the *DDX3* locus (Orientation PCR assay) or using primers flanking the Flag sequence (Insertion PCR assay). Bottom panel, Flag-insertion in 20 different single-cell-derived clones was assessed by PCR using primers flanking the Flag-sequence. **b** Left panel, off-target monitoring in immortalized mouse macrophages stably expressing *GFP* transgenes bearing silent mutations in the region targeted by the sgRNA. Right panel, cells were transfected with plasmids coding for Cas9 and the sgRNA or transduced with Nanoblades. GFP expression was measured by FACS 72 h after transfection/transduction ( $n = 3$ ). **c** Left and right panels, gene-editing at the *EMX1* on-target site and the *MFAP1* intronic off-target site measured by high-throughput sequencing in untreated cells (control cells) and cells transduced with *EMX1* Nanoblades (Nanoblades) or transfected with plasmids coding for Cas9 and the *EMX1* sgRNA (DNA transfection) ( $n = 3$ ). Statistical significance of the Nanoblades and DNA transfection comparison at the on-target site was computed using a two-tail Student test. **d** Left panel, position of sgRNAs targeting the promoter of *TTN* and VLPs with different combination of sgRNAs produced for the experiment. Right-panel, *TTN* mRNA expression levels (normalized to Control) as measured by qPCR in MCF7 transduced with VLPs ( $n = 3$ ). Error bars in all figures correspond to standard deviation

the puromycin cassette inserted at the *AAVS1* locus (Supplementary Figure 3d).

Taken together, our results show that Nanoblades can be used for the precise insertion of genetic material through HDR both with ssDNA and dsDNA donor DNA template and no requirement for any transfection reagent.

**Nanoblades confer low off-target genome-editing.** A major concern related to the use of CRISPR/Cas9-mediated gene editing are the potential off-target effects that can occur at genomic loci that are similar in sequence to the original target. Interestingly, several reports have shown that transient delivery of the Cas9-sgRNA complex by injection or RNP transfection generally leads to reduced off-target effects as compared to constitutive expression of Cas9 and sgRNA from DNA transfection experiments<sup>18</sup>. Since Nanoblades deliver the Cas9-sgRNA complex in a dose-dependent and transient fashion, we tested whether they could also lead to reduced off-target effects when compared to classical DNA transfection. For this, we developed an approach similar to that described by Fu and colleagues<sup>19</sup> by creating a series of HEK-293T reporter cell lines transduced with different versions of a *GFP* transgene bearing silent point mutations located in the sgRNA target site (Fig. 3b, right panel). These cells were either transfected with plasmids coding for Cas9 and the sgRNA targeting the *GFP* or transduced with Nanoblades programmed with the same sgRNA. 96 h after transfection/transduction, cells were collected and GFP expression was monitored by FACS (Fig. 3b, left panel). As expected, GFP expression from cells bearing the wild-type *GFP* sequence (No Mismatch) was efficiently repressed both after Nanoblades transduction (close to 80% repression) and DNA transfection (close to 60% repression) (Fig. 3b, left panel “No Mismatch”). When two mismatches were introduced in the target site, Nanoblades were no longer able to efficiently repress GFP expression (20% compared to control) while GFP expression from transfected cells was still reduced to levels similar to that of the *GFP* bearing a perfect match with the sgRNA. Interestingly, the presence of three or four mismatches completely abolished *GFP* editing in Nanoblades-treated cells while cells transfected with the Cas9 and sgRNA plasmids still displayed a mild inhibition of GFP expression (Fig. 3b see 3 and 4 Mismatches).

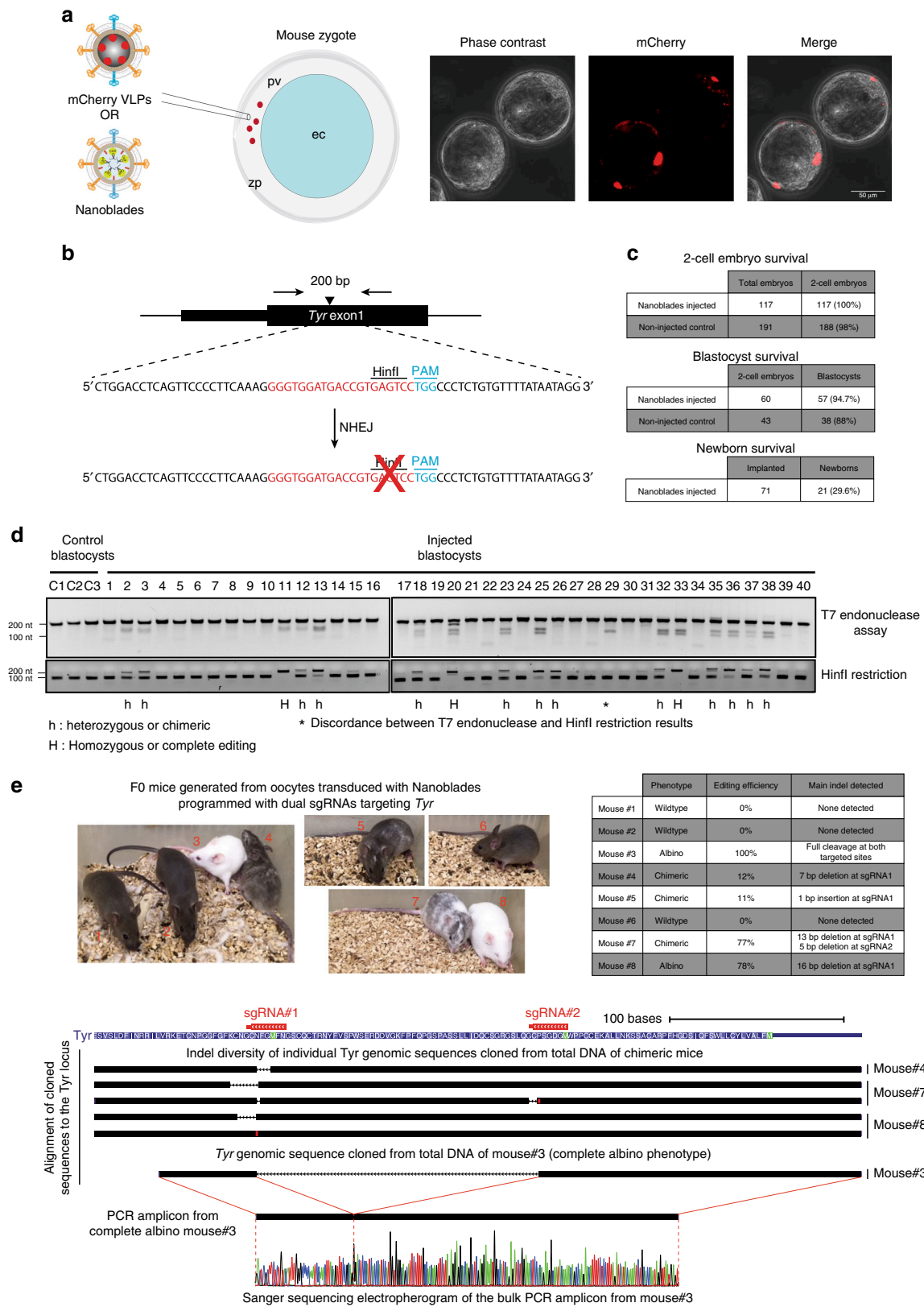
To complement these results, we further tested for genomic off-target effects using the well-characterized sgRNA targeting human *EMX1*. Off-targets for this sgRNA have been extensively studied using T7 endonuclease assays and high-throughput sequencing approaches<sup>10</sup>. We PCR-amplified the *EMX1* locus and one of the previously described *EMX1* genomic off-target loci occurring at the intron of *MFAP1*<sup>10</sup> in cells treated for 72 h with

Nanoblades programmed with the *EMX1* sgRNA or transfected with a DNA construct coding for Cas9 and the *EMX1* sgRNA. We then assessed genome-editing on each sample by high-throughput sequencing (Fig. 3c)<sup>13</sup>. Editing at the on-target site was efficient in Nanoblade-treated cells (75% in average) and to a less extent in cells transfected with the DNA coding for Cas9 and the sgRNA (53% in average) (Fig. 3c, left panel). As expected, small INDELS (insertions and deletions) occurred close to the expected Cas9 cleavage site located 3nt upstream the PAM sequence both in Nanoblades treated and in DNA-transfected cells (Supplementary Figure 4). Surprisingly, in spite of the higher editing efficiency at the on-target site, we could not detect any significant editing at the *MFAP1* off-target site in Nanoblades-treated cells (Fig. 3c, right panel). In contrast, cells transfected with the DNA coding for Cas9 and the sgRNA displayed significant editing (close 6%) at the off-target site (Fig. 3c, right panel) and had INDELS at the expected cut site (Supplementary Figure 4).

Taken together, our results indicate that similarly to other protocols that lead to transient delivery of the Cas9-sgRNA RNP, Nanoblades display low off-target effects.

**Targeted transcriptional activation through Nanoblades.** Having shown efficient genome editing using Nanoblades loaded with the catalytically active Cas9, we tested whether Nanoblades could also deliver Cas9 variant proteins for applications, such as targeted transcriptional activation. To this aim, we fused the Cas9-derived transcriptional activator (SP-dCas9-VPR)<sup>20</sup> to Gag from MLV and expressed the fusion protein in producer cells together with a control sgRNA or different combinations of sgRNAs targeting the promoter region of human Titin (*TTN*) as previously described<sup>20</sup> (Fig. 3d, left panel). Nanoblades loaded with SP-dCas9-VPR were then incubated with MCF-7 cells and induction of *TTN* measured by quantitative RT-PCR (normalized to GAPDH expression). As observed (Fig. 3d, right panel), when two different sgRNAs were used in combination, *TTN* transcription was stimulated from 50 to 200 fold compared to the control situation. Interestingly, when combining the four different sgRNAs in a single VLP, we obtained up to 400-fold transcription stimulation of *TTN* after 4 h of transduction. Our results therefore suggest that in spite of the large molecular size of the SP-dCas9-VPR (predicted at 224 kDa alone and 286 kDa when fused to MLV Gag), neither its encapsidation within VLPs nor its delivery and function within target cells are impaired. The use of Cas9 variants could therefore expand the toolbox of potential applications of Nanoblades in immortalized and primary cells.





**Nanoblades-mediated transduction of mouse zygotes.** CRISPR-Cas9 has been extensively used to generate transgenic animals through microinjection of zygotes with DNA coding for Cas9 and the sgRNA or with the synthetic sgRNA and a Cas9 coding mRNA or directly with the preassembled Cas9-sgRNA RNP<sup>21</sup>. However, some of these options usually require injection into the

pronucleus or the cytoplasm of zygotes, which can significantly impact their viability. Moreover, in some species, pronucleus and even cytoplasmic microinjection can be technically challenging.

Because Nanoblades are programmed to fuse with their target cells, we reasoned that they could also transduce murine zygotes without requiring intracellular microinjection. To test this

**Fig. 4** Generation of transgenic mice using Nanoblades. **a** Left panel, scheme describing injection of mCherry VLPs or Nanoblades in the perivitelline space of mouse 1-cell embryos. Right panel, fluorescence microscopy of mouse blastocysts injected with mCherry VLPs at the single-cell stage. **b** Scheme of the design strategy to target the mouse *Tyr* locus (adapted from ref. 22). Upon editing and NHEJ repair, the *Hin*I restriction site becomes inactive. **c** Survival rates of injected embryos at two-cell, blastocyst, and newborn stage (the latter obtained from experiments presented in Supplementary figure 5). **d** T7 endonuclease (top panel) and *Hin*I restrictions (bottom panel) assays on PCR fragments amplified from the *Tyr* locus of Control or Nanoblades-injected embryos. **e** Top left panel, photographs of F0 mice generated from embryos injected with Nanoblades programmed with two sgRNAs targeting the *Tyr* locus. Top-right panel, phenotype, editing efficiency (as measured by TIDE analysis of the Sanger-sequencing electropherograms) and the main INDEL type as detected by Sanger sequencing of individual PCR clones. Bottom-panel, alignment of individual PCR clones obtained from the *Tyr* locus of F0 mice against the mouse mm10 genome indicating the main observed INDELS in chimeric mice (mouse #4, #7, and #8) and total excision of the *Tyr* sequence between the sgRNA1 and sgRNA2 targeting loci for the complete albino mouse (mouse #3). The Sanger sequencing electropherogram from the bulk PCR amplicon obtained from mouse #3 indicates complete editing at both targeted sites

hypothesis, VLPs loaded with the mCherry protein (instead of Cas9) were produced and injected in the perivitelline space of mouse zygotes (Fig. 4a, top panel). Embryos were harvested 80 h after injection (blastocyst stage) and visualized by fluorescence microscopy, showing mCherry protein delivery within embryo cells (Fig. 4a, right panel).

Nanoblades programmed with a sgRNA targeting the first exon of the tyrosinase (*Tyr*) gene previously described in ref. 22 were produced and injected in the perivitelline space of mouse zygotes. This particular sgRNA was specifically designed to target a *Hin*I restriction site in the *Tyr* gene that should be disrupted upon dsDNA cleavage and NHEJ repair<sup>22</sup> (Fig. 4b). 80 h after injection, blastocysts were harvested and genomic DNA extracted to monitor genome-editing by PCR amplification followed by T7 endonuclease assay or *Hin*I restriction. As observed (Fig. 4d), 16 out of 40 blastocysts were positive for genome-editing at the *Tyr* gene both for the T7 endonuclease and the *Hin*I restriction assays. Interestingly, three blastocysts (#11, #20, and #33) appeared to bear complete *Tyr* editing as we could not detect any residual *Hin*I restriction products (Fig. 4d). In the remaining 13 blastocysts that were positive for genome editing at the *Tyr* locus, we observed different editing efficiencies thus arguing for variable levels of mosaicism between individuals (Fig. 4d). Interestingly, injection of Nanoblades in the perivitelline was not associated with embryo mortality as we did not obtain any significant difference in survival rates between injected and non-inject embryos (Fig. 4c). To further validate these results, we produced Nanoblades programmed with two sgRNAs targeting the *Tyr* locus (see Fig. 4e bottom scheme) that were injected in the perivitelline space of single-cell embryos, which were then implanted into pseudopregnant females and carried to term. In this case, five out of eight F0 mice obtained carried detectable *Tyr* editing both at the phenotype and genotype level as assayed by PCR amplification of the *Tyr* locus from genomic DNA extracted from the fingers of each animal (Fig. 4e). Interestingly, one of the two fully albino mice carried a complete deletion of the DNA segment between the two sgRNA-targeted loci in all tested cells (as assayed by Sanger sequencing of the bulk PCR product and Sanger sequencing of single clone PCR fragments (Fig. 4e bottom panels)). The remaining F0 mice that displayed a partial *Tyr* disruption phenotype had an editing efficiency ranging from 11% up to 78% of all *Tyr* alleles (Fig. 4e see table). Sanger sequencing of individual PCR clones amplified from these mice indicated that one of the two sgRNAs (sgRNA1) was more efficient in inducing INDELS (Fig. 4e bottom scheme). Moreover, we also detected some degree of mosaicism within each individual mouse (with the exception of mouse #3 which had complete bi-allelic excision of the *Tyr* sequence between the two target loci) with at least two types of INDELS detected in mice 7 and 8 (Fig. 4e, see genomic alignment scheme). This, however, is very similar to the degree of mosaicism found in other approaches<sup>22,23</sup>. Taken together, these results validate the use of

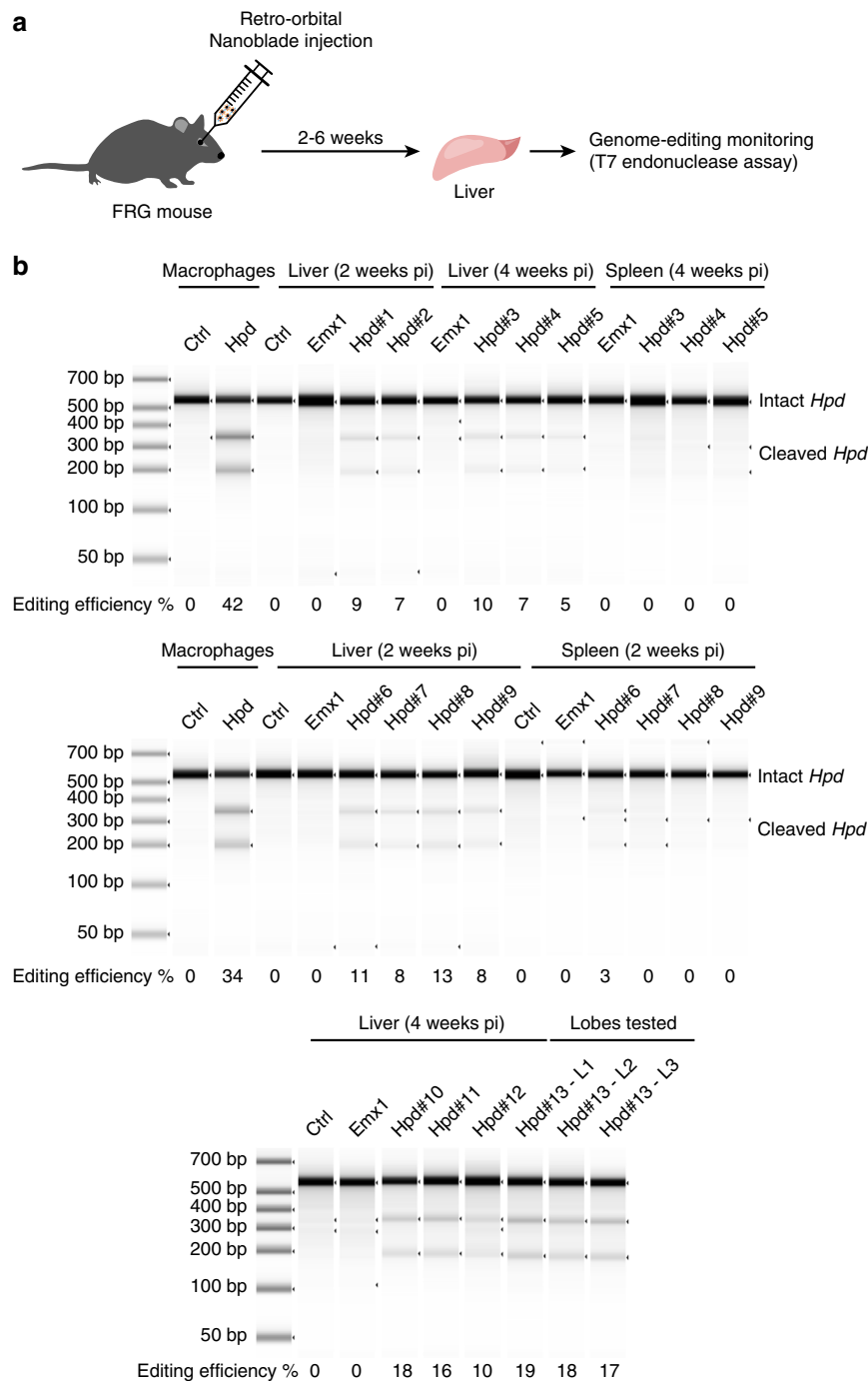
Nanoblades to generate transgenic mice upon perivitelline injection of single-cell embryos.

To further confirm the ability of Nanoblades to mediate genome-editing in mouse embryos and transmission of the edited locus to the offspring, we designed a sgRNA targeting the loxP sequence that could mimic the action of the Cre recombinase by removing a loxP flanked cassette (Supplementary Figure 5, left panel). These Nanoblades were first tested in primary BM cells derived from R26R-EYFP transgenic mice bearing a single-copy of the YFP transgene under control of a “lox-stop-lox” cassette<sup>24</sup> (Supplementary Figure 5, top right panel). Nanoblades were then injected in the perivitelline space of heterozygous R26R-EYFP 1-cell embryos which were then implanted into pseudopregnant females and carried to term. In this case, 1 out of 14 founder animals was YFP positive under ultraviolet (UV) light and displayed efficient excision of the “lox-stop-lox” cassette as confirmed by PCR<sup>25</sup> (Supplementary Figure 5, bottom left panel). Consistent with our previous results, the F1 progeny obtained after mating the loxed F0 mouse with a wild-type mouse contained the “loxed” version of the YFP allele and displayed YFP expression in tails and muscle fibers (Supplementary Figure 5, bottom right panel), indicating efficient transmission of the loxed allele from the F0 founder to its progeny.

Taken together, Nanoblades can represent a viable alternative to classical microinjection experiments for the generation of transgenic animals, in particular for species with fragile embryos or with poorly visible pronuclei.

#### In vivo editing of *Hpd* in the liver of tyrosinaemic FRG mice.

Hereditary tyrosinemia type I (HT1) is a metabolic disease caused by disruption of fumarylacetoacetate hydrolase (*Fah*), which is an enzyme required in the tyrosine catabolic pathway. *Fah*<sup>-/-</sup> mice recapitulate many phenotypic characteristics of HT1 in humans, such as hypertyrosinemia and liver failure and have to be treated with nitisinone for their survival. Disruption of hydroxyphenylpyruvate dioxygenase (HPD), the enzyme targeted by nitisinone) through hydrodynamic tail vein injection in *Fah*<sup>-/-</sup> mice was recently shown to restore their survival in the absence of nitisinone thanks to the selective advantage of *Hpd* negative hepatocytes<sup>26</sup>. We therefore reasoned that Nanoblades could represent a non-invasive method to inactivate the *Hpd* gene in NRG (NOD<sup>Fah</sup><sup>-/-</sup>/Rag2<sup>-/-</sup>/Il2rg<sup>-/-</sup>) mice<sup>27</sup>. To this aim, we designed a sgRNA directed against the fourth exon of *Hpd*, which should disrupt the reading frame through the INDELS caused by NHEJ (see Methods section for the sequence). Nanoblades directed against *Hpd* or against human *EMX1* (control) were introduced in NRG mice through retro-orbital injection (Fig. 5a). Upon injection, mice were weaned off nitisinone until they reached a 20% loss of their body weight, in which case nitisinone was subsequently administered punctually. Two weeks after injection, all mice injected with Nanoblades targeting *Hpd* displayed detectable editing in the liver (between 7% and 13%



**Fig. 5** Inactivation of *Hpd* in the liver of tyrosinaemic FRG mice. **a** Scheme of the experimental approach to target the liver of FRG mice. **b** T7 endonuclease assay to monitor genome editing at the *Hpd* gene in immortalized mouse macrophages and in the liver or spleen of injected mice. Samples were quantified using a TapeStation chip

efficiency, Fig. 5b). On the contrary, no editing was detected in control (uninjected) mice or in mice injected with Nanoblades targeting human *EMX1* (Fig. 5b). Similar results were obtained 4 weeks post-injection where all mice injected with Nanoblades targeting *Hpd* displayed genome editing in the liver (Fig. 5b). Furthermore, genome-editing occurred in a homogenous fashion across the liver as shown by T7 endonuclease assay from biopsies recovered from three different lobes of a single animal (Fig. 5b, bottom panel). In contrast, editing in other organs, such as spleen was weak or not detectable (Fig. 5b).

Interestingly, we observed a small overall increase in editing levels at 4 weeks post-injection compared to 2 weeks post-

injection suggesting that cells with *Hpd* editing could have a selective advantage over non-edited cells (Fig. 5b compare middle and bottom panel). Because we did not monitor genome editing earlier than 2 weeks post injection, we cannot rule out that a similar selective advantage of edited cells might have occurred during this incubation time. Nevertheless, based on the weak increase of the editing efficiency observed between 2 and 4 weeks after injection, we do not expect this selective advantage to significantly improve the observed editing efficiency during the first 2 weeks after injection. Importantly, Nanoblades injection was not associated with any signs of morbidity.

## Discussion

Genome editing should ideally be achieved in a fast and precise fashion to limit toxicity and possible off-target effects due to a sustained expression of effectors. In this regard, extensive efforts have been recently described to vehicle Cas9-sgRNA RNPs in cultured cells and in vivo by non-coding material including Nanocarriers<sup>28</sup>, optimized transfection reagents<sup>18</sup>, or lentivirus-derived particles<sup>29</sup>.

This work describes and characterizes VLPs to efficiently vectorize the CRISPR-Cas9 system into primary cells, embryos, and animals. These non-coding agents—we called herein Nanoblades—incorporate the Cas9 endonuclease into their internal structure. The molecular basis of this technology is the fusion of Cas9 from *Streptococcus pyogenes* to Gag from MLV. Expressed with other components of viral assembly and construct encoding gRNA(s), this molecule can bind sgRNAs into producer cells, forms RNP complexes and cohabit with Gag and Gag-Pol within particles. We indeed show that robust packaging of sgRNAs into Nanoblades depends on their interaction with Gag::Cas9 (Supplementary Figure 1d).

When compared to other methods of delivery such as lipofection or electroporation, Nanoblades were more efficient and rapid in inducing dsDNA breaks both in immortalized U2OS cells, primary fibroblasts (Fig. 1b, Supplementary Figure 1e). Nanoblades are also functional in primary cells that are known to be difficult to transfect and transduce using classical delivery methods, such as human iPS cells, human CD34+ and primary mouse bone-marrow cells (Fig. 2) reaching efficiencies comparable or even superior to other recent methods<sup>30,31</sup>, such as Cas9-sgRNA ribonucleoprotein electroporation (Fig. 2c), together with low off-target effects (Fig. 3b and c). Furthermore, Nanoblades achieve genome editing in a dose-dependent manner (Fig. 1c). Beyond delivery of Cas9-sgRNA complexes, we also show that Nanoblades can be complexed with DNA repair templates to mediate homologous recombination-based knock-in cultured cells in the absence of any transfection reagent. Our results also validate the use of Nanoblades in vivo for generating transgenic mice upon embryo injection in the perivitelline space (Fig. 4 and Supplementary Figure 5) or in the liver of injected animals (Fig. 5). Although, other recent methods for in vivo genome editing of zygotes and animals have reached higher editing rates<sup>22,23,32–34</sup>, Nanoblades represent a viable, inexpensive, and accessible alternative that can still benefit from further improvements.

Similarly to other cell-derived particles (including most viral vectors), Nanoblades incorporate RNAs and proteins from producer cells that could be responsible for the transmission of undesired effects. Mass spectrometry analysis of the content of Nanoblades revealed that plasma membrane terms were particularly enriched, which is consistent with the vesicular nature of Nanoblades (Supplementary Figure 2a and Supplementary Data 1). As previously described for retroviral-VLPs<sup>35</sup>, characterization of the RNA content revealed that Nanoblades contain thousands of individual cellular mRNA species, most of these being encapsidated stochastically, in proportion to their abundance in the producer cell. We found that transcripts overexpressed for production purposes (GAG, VSV-G, etc.) represent <0.4% of Nanoblades RNAs (Supplementary Figure 2b) supporting the notion that their delivery to recipient cells is marginal. Confirming this observation, transfer of cellular proteins loaded in Nanoblades from producer cells to recipient cells appears to be minimal and restricted to a short time window between 8 and 24 h after transduction (Supplementary Figure 2c and d). While we cannot exclude the fact that VLPs may be responsible for some cellular responses, depending on the nature of recipient cells, efficient doses of Nanoblades were globally harmless for most

primary cells we tested and in injected animals. In our effort to exploit the retroviral nature of Nanoblades, we explored diverse pseudotyping options (Supplementary Figure 6) and finally focused on the use of an original mixture of two envelopes (VSV-G plus BRL), a recipe that we have optimized (Supplementary Figure 6) and which systematically displayed the best cleavage results in most recipient cells. Depending on the cellular target, it may be possible to pseudotype Nanoblades with envelopes from Measles virus<sup>36</sup>, influenza virus<sup>37</sup>, or other targeting systems<sup>38,39</sup> to restrict or improve Cas9 delivery to certain cell types (Supplementary Figure 6a).

Next generation Nanoblades may also benefit from the continual evolutions of Cas9-derivatives that can support fusion with Gag from MLV (Fig. 3) and could be adapted to other gene-editing targetable nucleases like Cpf1 nucleases<sup>40</sup> or even the latest generation of programmable base editors<sup>41</sup>. We also noted that Nanoblades can be engineered to accommodate other proteins/RNAs in addition to Cas9-RNPs and serve as multifunctional agents. Nanoblades capable of delivering both Cas9-RNPs and a reverse-transcribed template that can serve for reparation by homologous-recombination could therefore be envisioned. Furthermore, multiple sgRNAs can be incorporated within Nanoblades thus allowing gene excisions or multiple genes to be targeted. Multiplexing of sgRNAs may also allow the introduction of an additional sgRNA targeting a specific gene that will allow selection of cells efficiently edited by Nanoblade-mediated CRISPR<sup>42</sup>.

This versatility allows any laboratory equipped with BSL2 facilities to generate its own batches of particles. Beyond cell lines, our VLP-based technique provides a powerful tool to mediate gene editing in hiPSCs and primary cells including macrophages, human hematopoietic progenitors and primary hepatocytes. We have shown that Nanoblades injection into the perivitelline space of mouse-zygotes was particularly harmless for the recipient cells, since none of the injected zygotes were affected in their development after treatment. Generation of transgenic animals upon perivitelline space injection of VLPs could be adapted to other species, including larger animals for which the number of zygotes is limited. Finally, we achieved significant gene-editing in the liver of injected adult mice with no consequences on their viability. Nanoblades, could therefore represent an interesting route for the delivery of Cas9 in vivo to inactivate gene expression but also used in combination with other viral delivery tools carrying a donor DNA template (such as Adeno-associated virus (AAV)) to perform in vivo HDR experiments as recently shown<sup>32</sup>.

Considering the examples provided in our work, we believe that the Nanoblade technology will facilitate gene editing in academic laboratories working with primary cells and could represent a viable alternative for therapeutical purposes and the rapid generation of primary cell-types harboring genetic diseases, humanized-liver mouse models and transgenic animal models.

## Methods

**Plasmids.** SP-dCas9-VPR was a gift from George Church (Addgene plasmid #63798). Lenti CRISPR was a gift from F. Zhang (Addgene plasmid #49535). The GagMLV-CAS9 fusion was constructed by sequential insertions of PCR-amplified fragments in an eukaryotic expression plasmid harboring the human cytomegalovirus early promoter (CMV), the rabbit Beta-globin intron and polyadenylation signals. The MA-CA-NC sequence from Friend MLV (Accession Number: M93134) was fused to the MA/p12 protease-cleavage site (9 aa) and the Flag-nls-spCas9 amplified from pLenti CRISPR.

**Cell culture.** Gesicle Producer 293T (Clontech 632617), U2OS cells, and primary human fibroblasts (Coriell Institute, GM00312) were grown in DMEM supplemented with 10% fetal calf serum (FCS).

hiPSCs were obtained and cultured as described in ref. <sup>43</sup>.

Bone marrow-derived macrophages (BMDMs) were differentiated from BM cells obtained from wild-type C57BL/6 mice. Cells were grown in DMEM



supplemented with 10% FCS and 20% L929 supernatant containing MCSF as described in ref. 44. Macrophages were stimulated for the indicated times with LPS (Invivogen) at a final concentration of 100 ng/ml.

**CD34+ cell sample collection, isolation, and transduction.** Cord blood (CB) samples were collected in sterile tubes containing the anti-coagulant, citrate-dextrose (ACD, Sigma, France) after informed consent and approval was obtained by the institutional review board (Centre international d'inféctiologie (CIRI), Lyon, France) according to the Helsinki declaration. Low-density cells were separated over, Ficoll-Hypaque. CD34+ isolation was performed by means of positive selection using magnetic cell separation (Miltenyi MACs) columns according to the manufacturer's instructions (Miltenyi Biotec, Bergisch Gladbach, Germany). Purity of the selected CD34+ fraction was assessed by FACS analysis with a phycoerythrin (PE)-conjugated anti-CD34 antibody (Miltenyi Biotec, Bergisch Gladbach, Germany) and exceeded 95% for all experiments. Human CD34+ cells were incubated for 18–24 h in 24-well plates in serum-free medium (CellGro, CellGenix, Germany) supplemented with human recombinant: SCF (100 ng/ml), TPO (20 ng/ml), Flt3-L (100 ng/ml) (Miltenyi, France).  $5 \times 10^4$  prestimulated CD34+ cells were then incubated with nanoblades in 48-well plates in serum-free medium.

**sgRNA design and sequences (+PAM).** sgRNAs targeting *MYD88*, *DDX3*, *GFP*, *Hpd*, *Fto*, *Tyr*, and the *LoxP* sequence were designed using CRISPRseek<sup>45</sup>.

Human AAVS1: 5' ACCCCACAGTGGGGCCACTAggg 3'  
 Human DDX3: 5' AGGGATGAGTCATGTGGCAGTgg 3'  
 Human EMX1: 5' GAGTCCGAGCAGAAGAAGAagg 3'  
 Human MYD88 #1: 5' GAGACCTCAAGGGTAGAGGTggg 3'  
 Human MYD88 #2: 5' GCAGCCATGGCGGGCGGTCCtgg 3'  
 Human rDNA: 5' CCTTCTTAGCGATCTGAGagg 3'  
 Human TTN -169: 5' CCTTGGTGAAGTCTCCTTTgagg 3'  
 Human TTN -252: 5' ATGTTAAATCCGAAAATGcagg 3'  
 Human TTN -326: 5' GGGCACAGTCTCAGGTTTgggg 3'  
 Human TTN -480: 5' ATGAGCTCTCTCAACGTTAagg 3'  
 Mouse Fto: 5' CATGAAGCGCGTCCAGACCgagg 3'  
 Mouse Hpd: 5' GAGTTTCTATAGGTGGTGGTGGTgggg 3'  
 Mouse Tyr: 5' GGGTGGATACCGTGAAGTCCtgg 3' obtained from Chen et al. 22  
 Mouse Tyr: 5' AACTTCATGGGTTTCAACTGcgg 3' obtained from Yoon et al. 23  
 Mouse Tyr: 5' ATGGGTGATGGGAGTCCCTGcgg 3' this study  
 LoxP: 5' CATTATACGAAGTTATATTAagg 3'  
 GFP: 5' CGAGGAGCTGTTACCAGGGGtgg 3'

**Production of Nanoblades.** Nanoblades were produced from transfected gestic producer 293T cells plated at  $5 \times 10^6$  cells/10 cm plate 24 h before transfection with the JetPrime reagent (Polyplus). Plasmids encoding the GagMLV-CAS9 fusion (1.7 µg), Gag-POLMLV (2.8 µg), gRNA expressing plasmid(s) (4.4 µg), VSV-G (0.4 µg), the Baboon Endogenous retrovirus Rless glycoprotein (BaEVRless)<sup>15</sup> (0.7 µg) were cotransfected and supernatants were collected from producer cells after 40 h. For production of serum-free particles, medium was replaced 24 h after transfection by 10 ml of Optimum (Gibco) supplemented with penicillin–streptomycin. Nanoblade-containing medium was clarified by a short centrifugation (500 × g 5 min) and filtered through a 0.8 µm pore-size filter before ultracentrifugation (1h30 at 96,000 × g). Pellet was resuspended by gentle agitation in 100 µl of cold 1X PBS. Nanoblades were classically concentrated 100-fold. X-Nanoblades referred as Nanoblades loaded with gRNA(s) targeting the x-gene.

To dose Cas9 packaged into particles, Nanoblades or recombinant Cas9 (New England Biolabs) were diluted in 1X PBS and serial dilutions were spotted onto a Nitrocellulose membrane. After incubation with a blocking buffer (nonfat Milk 5% w/v in TBST), membrane was stained with a Cas9 antibody (7A9-3A3 clone, Cell signaling) and revealed by a secondary anti-mouse antibody coupled to horseradish peroxidase. Cas9 spots were quantified by Chemidoc touch imaging system (Biorad).

**Transduction procedure.** Transductions with Nanoblades were performed in a minimal volume to optimize cell/particles interactions for at least 2 h before supplementing with fresh medium. When specified, polybrene was used at a final concentration of 4 µg/ml in the transduction medium. After dosing Cas9 amount in each Nanoblades preparation, we typically used 10 pmol of encapsidated Cas9 for  $1 \times 10^5$  adherent cells.

**sgRNA in vitro transcriptions.** sgRNAs were in vitro transcribed using the EnGen sgRNA Synthesis kit, *S. pyogenes* (New England Biolabs; E3322S) following the manufacturer's protocol with the following oligonucleotides:

Human EMX1: 5' TTCTAATACGACTCACTATAgatccgagcagaagaagaGTTTTAGAGCTAGA 3'  
 Mouse Fto: 5' TTCTAATACGACTCACTATAgatcaagcgcgtcagaccgGTTTTAGAGCTAGA 3'

After transcription, sgRNAs were purified by acidic phenol/chloroform extraction and precipitated using 2.5 volumes of 100% ethanol. sgRNA integrity was then assessed by denaturing urea polyacrylamide gel electrophoresis.

**Cas9-sgRNA RNP electroporation procedure.** Cas9-sgRNA RNP electroporation was performed as described in the manufacturer's protocol. Briefly, 12 pmol of recombinant Cas9 (EnGen Cas9 NLS, *S. pyogenes*; New England Biolabs; M0646T) were incubated with 12 pmol of in vitro transcribed sgRNAs in the presence of Resuspension Buffer R (Neon Transfection System; ThermoFisher Scientific; MPK1025) for 20 min at room temperature. After this,  $1 \times 10^5$  cells resuspended in 5 µl of resuspension buffer R (for HEK293T cells) or resuspension buffer T (for primary mouse BM cells) are added to the Cas9-sgRNA mix and the whole mixture electroporated with the following settings:

-1700 V, 20 ms, 1 pulse (HEK293T cells)  
 -1350 V, 10 ms, 4 pulses (mouse BM cells)  
 -1680 V, 20 ms, 1 pulse (mouse BM cells)

Upon electroporation, cells were incubated in their corresponding medium (DMEM complemented with 10% FCS for HEK293T cells and DMEM complemented with 10% FCS and 20% L929 supernatant containing MCSF for 24 h before extracting their genomic DNA to assess genome editing.

**Combination of Nanoblades with ssDNA and dsDNA.** Nanoblades programmed to target the AUG codon of *DDX3* were resuspended in PBS 2% FBS and combined with ssDNA donor repair template (see the sequence of "Flag-DDX3 primer" below) at a final concentration of 0.3, 1.3 or 6.7 µM in 30 µl of PBS supplemented with polybrene (Sigma) at 4 µg/ml. Complexes were let 15 min on ice before addition to  $7 \times 10^4$  HEK293T cells plated 6 h before in 400 µl of complete medium supplemented with polybrene (4 µg/ml). 24 h later, transduction medium was supplemented with 1 ml of fresh medium (10% FCS) and cells were passed the day after into six-well plates for amplification. Cells were amplified in 10 cm dishes and passed six times during 3 weeks before extraction of proteins and genomic DNAs.

Sequence of the Flag-DDX3 primer (HPLC-purified):

5'-ACTCGCTTAGCAGCGGAAGACTCCGagTTCTCGGTA  
 CTCTTCAGGGATGGA  
 CTACAAGGACGACGATGACAAgagTCATGTGGCAGTG  
 GAAAATGCGCTCGGGCTGGACCAGCAGGTGA-3'

DDX3 amplification was performed using the following primers: DDX3-Forward 5'-CTTCGCGGTGGAACAAACAC-3' and DDX3-Reverse1 5'-CGCCATTAGCCAGGTTAGGT-3' for the "Insertion PCR assay" and Flag-Forward 5'-GACTACAAGG  
 ACGACGATGACAAG-3' and DDX3-Reverse2 5'-CGCCATTA  
 GCCAGGTTAGGT-3' for the "Orientation PCR assay". PCR conditions were performed as follows: 94 °C 5 min, followed by three cycles (94 °C 30 s, 64 °C 30 s, 72 °C 30 s), followed by 25 cycles (94 °C 30 s, 57 °C 30 s, 72 °C 30 s), followed by 5 min at 72 °C.

dsDNA (AAVS1): 10 µl of concentrated Nanoblades were complexed with 650 ng of dsDNA in a total volume of 30 µl of PBS with polybrene at a final concentration of 4 µg/ml. After 15 min of incubation on ice, complexes were used to transduce  $1 \times 10^5$  HEK293T cells in a 24-well plate containing medium supplemented with polybrene (4 µg/ml). Two days later cells were reseeded in a 10 cm dish before puromycin selection (0.5 µg/ml). Single-cell-derived clones were next isolated and cultivated in a 12-well plates before PCR analysis performed on genomic DNAs (500 ng).

Primers used to assess the presence of the puromycin cassette are:

Puromycin-forward 1: 5'-GGCAGTCTGCTGTTCTCTGAC-3'  
 Puromycin-reverse 1: 5'-GATCCAGATCTGGTGTGGCGCG  
 TGGCGGGGTAG-3'

Followed by a nested-PCR using the following primers:

Puromycin-forward 2: 5'-GATATACGCGTCCCAGGGCCGG  
 TTAATGTGGCTC-3'  
 Puromycin-reverse 1: 5'-GATCCAGATCTGGTGTGGCGCG  
 TGGCGGGGTAG-3'

Primers used to assess correct integration of the cassette at the AAVS1 locus are:

AAVS1-forward: 5'-CGAACTCTGCCCTCTAACGCTG-3'  
 Puromycin reverse 2: 5'-GATCCAGATCTGGTGTGGCGCG  
 TGGCGGGGTAG-3'

Followed by a nested-PCR using the following primers:

AAVS1-forward: 5'-GGCAGGTCCTGCTTCTCTGAC-3'  
 Puromycin reverse 3: 5'-CACCGTGGCTGTACTCGGT  
 CAT-3'

**Flag-immunoprecipitation and western-blotting.** For Flag-immunoprecipitation,  $5 \times 10^6$  cells were lysed in 500 µl of lysis buffer (NaCl 300 mM, MgCl<sub>2</sub> 6 mM, Tris-HCl 15 mM, 0.5% NP40). 250 µl of the cell lysate (1 mg of total proteins) was incubated with 40 µl of M2-antiFlag magnetic beads (Sigma M8823) equilibrated in TBS. After incubation for 2 h at 4 °C, beads were washed four times in lysis buffer and proteins eluted in 60 µl of TBS supplemented with Flag-peptide (120 µg/ml final) for 2 h at 4 °C. The supernatant (without beads) was then collected and used for western-blot analyses.

Western-blotting against Flag-DDX3 and endogenous DDX3 was performed using the following antibodies: anti-DDX3 (rabbit, Sigma 19B4, 1/1000 dilution), Flag-M2 Antibody (mouse, Sigma F3165, 1/2000 dilution), and actin antibody (mouse, Sigma A1978, 1/10,000 dilution). The uncropped images for

Supplementary Figs. 1a, 2d, 3d and 2b–d, 3a, 4d are provided in Supplementary Fig. 7.

**T7 endonuclease assay.** Genomic DNA was extracted from VLP-treated cells using the Nucleospin gDNA extraction kit (Macherey-Nagel). 150 ng of genomic DNA was then used for PCR amplification. PCR products were diluted by a factor 2 and complemented with Buffer 2 (New England Biolabs) to a final concentration of 1×. Diluted PCR amplicons were then heat denatured at 95 °C and cooled down to 20 °C with a 0.1 °C/s ramp. Heteroduplexes were incubated for 30 min at 37 °C in presence of 10 units of T7 Endonuclease I (NEB). Samples were finally run on a 2.5% agarose gel or on a BioAnalyzer chip (Agilent) to assess editing efficiency.

**Reverse-transcription and quantitative PCR.** Total RNAs were extracted using TriPure Isolation Reagent (Roche, 11667165001) following the manufacturer's instructions. 1.5 µg of total RNA was treated with DNase and reverse-transcribed using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific, K1672) following the manufacturer's instructions. qPCR experiments were performed on a LightCycler 480 (ROCHE) in technical triplicates in 10 µl reaction volume as follows: 5 µl of 2X SYBR qPCR Premix Ex Taq (Thi RNaseH Plus) (TAKARA, TAKRR420W); forward and reverse primers (0.5 µM each final); 7.5 µg of cDNA.

**Immunofluorescence and imaging.** Cells were fixed in 1X PBS supplemented with 4% of paraformaldehyde (PFA) for 20 min, washed three times with 1X PBS and permeabilized with 0.5% Triton X-100 for 4.5 min. Cells were incubated with primary antibodies overnight at 4 °C. Primary antibodies used are: rabbit yH2AX (1:1000; Abcam 81299) and mouse RNA pol I RPA194 (1:500; Santacruz sc48385). Cells were washed three times in 1X PBS, followed by incubation of the secondary antibodies conjugated to Alexa 488 or 594 used at a 1:1000 dilution (Life Technologies) for 1 h at room temperature. After three 1X PBS washes, nucleus were stained with Hoechst 33342 at 1 µg/ml for 5 min. The coverslips were mounted in Citifluor medium (AF1, Citifluor, London, UK). Cells were observed under a Leica DM6000. At least 100 cells were counted in each indicated experiment. Averages and standard deviation values were obtained from three independent biological replicates.

**Flow cytometry analysis of CD81 expression.**  $1 \times 10^6$  HepG2 or HEK293T cells were detached from the cell culture plate using Accutase (Stemcell technologies #07920) and washed twice in PBS + 2%BSA. Cells were then incubated in 100 µl of PBS + 2%BSA + Anti-CD81 (BD Biosciences #555675, clone JS-81, 1/200 dilution) for 30 min at 4 °C. Cells were then washed three times in PBS + 2% BSA and incubated in 100 µl of PBS + 2 %BSA + anti-mouse FITC (Biollegend # 406001, 1/2000 dilution) for 30 min at 4 °C in the dark. Cells were then washed three times in PBS + 2%BSA and fixed with 4% of paraformaldehyde (PFA) for 15 min and washed in PBS + 2%BSA before flow cytometry analysis on a BD FACSCanto II.

**Northern-blot of sgRNAs.** 2 µg of total RNA extracted from Nanoblades or Nanoblade-producing cells were run on a 10% acrylamide, 8 M Urea, 0.5X TBE gel for 1 h at 35 W. RNAs were then transferred onto a Nitrocellulose membrane (Hybond Amersham) by semi-dry transfert for 1 h at 300 mA in 0.5X TBE. The membrane was UV-irradiated for 1 min using a stratalinker 1800 and then baked at 80 °C for 30 min. The membrane was then incubated in 50 ml of Church buffer (125 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.085% phosphoric acid, 1 mM EDTA, 7% SDS, 1% BSA) and washed twice in 10 ml of Church buffer. The 5' P32-labeled ( $1 \times 10^7$  cpm total) and heat-denatured ssDNA probe directed against the constant sequence of the guideRNA (sequence of the sgRNA antisense probe: 5' GCACCGACTCGGTGCCA CTTTTCAGTTGATAACGGACTAGCCTTATTTAACCTTGCTATTCTA GCTCTA3') was diluted in 10 ml of Church buffer and incubated with the membrane overnight at 37 °C. The membrane was washed four times in 50 ml of wash buffer (1X SSC + 0.1% SDS) before proceeding to phosphorimaging.

**Transmission electron microscopy (TEM) and mass spectrometry (MS).** Nanoblades programmed to target the YFP were prepared and processed for TEM and MS as previously described<sup>46</sup>. Briefly, Nanoblades were produced from transfected Gesicles Producer 293T cells plated at  $5 \times 10^6$  cells/10 cm plate 24 h before transfection with the JetPrime reagent (Polyplus) and supernatants were collected from producer cells after 40 h, passed through a 0.45 µm filter and concentrated 100-fold by overnight centrifugation at  $3800 \times g$ . This preparation was next laid overlaid on a continuous optiprep gradient and ultracentrifuged to obtain density fractions. Fractions containing Nanoblades were next pooled and centrifuged overnight at  $3800 \times g$  before PBS resuspension to obtain a 6000×-concentrated sample.

For electron microscopy, after a flash-fixation in glutaraldehyde, staining was amplified using the R-Gent Kit (Biovalley, Marne-la-Vallée, France) before the negative coloration (phosphotungstic acid 2%). Specimen were observed under a JEM-1400 microscope (Jeol, Tokyo, Japan) coupled with the Orius-600 camera (Gatan, Pleasanton, CA).

**High-throughput sequencing of RNAs extracted from Nanoblades.** Total RNA was extracted from purified Nanoblades programmed to target the YFP using Trizol. RNAs were then fragmented to 100nt and used as input for the preparation of cDNA libraries following the protocol described in ref. <sup>47</sup>. Briefly, RNA fragments with a 3'-OH were ligated to a preadenylated DNA adaptor. Following this, ligated RNAs were reverse transcribed with Superscript III (Invitrogen) with a barcoded reverse-transcription primer that anneals to the preadenylated adaptor. After reverse transcription, cDNAs were resolved in a denaturing gel (10% acrylamide and 8 M urea) for 1 h and 45 min at 35 W. Gel-purified cDNAs were then circularized with CircLigase I (Epicentre) and PCR-amplified with Illumina's paired-end primers 1.0 and 2.0.

Analysis of high-throughput sequencing data was performed as previously described<sup>48</sup>. Briefly, reads were split with respect to their 5'-barcode sequence. After this, 5'-barcode and 3'-adaptor sequences were removed from reads. Reads were mapped to a custom set of sequences including 18S, 28S, 45S, 5S, and 5.8S rRNA, tRNAs, the sgRNA directed against the GFP sequence and all transcripts coding for Nanoblades components (Envelopes, Gag and Pol, Cas9) using Bowtie<sup>49</sup>. Reads that failed to map to this custom set of sequences were next aligned to University of California, Santa Cruz (UCSC) human hg18 assembly using TopHat2<sup>50</sup>. Read counts on all transcripts of interest were obtained using the HTSeq count package<sup>51</sup>.

**High-throughput sequencing of Emx1 On-target and Off-target loci.** Genomic DNA was extracted from Nanoblades-treated cells using the Nucleospin gDNA extraction kit (Macherey-Nagel). 150 ng of genomic DNA was then used for PCR amplification using primers specific for the *EMX1* On-target locus (*EMX1*-Forward 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCAGAACCGG AGGACAAAGTAC-3' and *EMX1*-Reverse 5'-GTGACTGGAGTCCCTCTCTAT GGGCAGTCGGTGAAGCCCATTGCTTGTCCCTCTGTCAATG-3') and the previously described Off-target locus in the intron of *MFAP1* (*MFAP1*-Forward 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCATCACGGCCTTTG CAAATAGAGCCC-3' and *MFAP1*-Reverse 5'-GTGACTGGAGTCCCTCTCTA TGGCAGTCGGTGACAGAGGGAACACTACAAGATCGCTGAGC-3') bearing adapters sequencing for Illumina's Miseq platform. Obtained PCR products were purified and PCR amplified with a second set of primers bearing specific barcodes for multiplex sequencing. Final PCR products were sequenced on the Miseq platform using a custom sequencing primer (Miseq-Custom 1: 5' ATCACCGACTGCCCATAGAGAGGACTCCAGTCAC 3') and a custom index sequencing primer (Miseq-Custom 2: 5' GTGACTGGAGTCCCTCTCTATGGGC AGTCGGTGAT 3').

**Animal experimentation.** All animal experiments were approved by a local ethics committee of the Université de Lyon (CECCAPP, registered as CEEA015 by the French ministry of research) and subsequently authorized by the French ministry of research (APAFIS#8154-2016112814462837 v2 for the generation of transgenic animals and C 69 123 0303 for the usage of Nanoblades in vivo). All procedures were in accordance with the European Community Council Directives of September 22, 2010 (2010/63/EU) regarding the protection of animals used for scientific purposes.

**Mouse oocyte injection.** Four or five weeks old FVB/NRj female mice (Janvier Labs, France) were superovulated by intraperitoneal (i.p.) administration of 5 IU of pregnant mare serum gonadotropin (PMSG, Alcyon, France), followed by an additional i.p. injection of 5 IU human chorion gonadotropin 48 h later (hCG, Alcyon, France). Superovulated females were mated with B6D2F1 adult males (1 male/2 females) and euthanized at 0.5 day post coitum (usually between 10 and 11 a.m.). Oviduct were dissected, and the ampulla nicked to release zygotes associated with surrounding cumulus cells into a 200 µl droplet of hyaluronidase (Sigma) in M2 solution (300 µg/ml, Sigma) under a stereomicroscope (Olympus SZX9). Zygotes were incubated for 1 min at room temperature and passed with a mouth pipette through three washes of M2 medium to remove cumulus cells. Zygotes were kept in M16 medium (Sigma) in a water jacketed CO<sub>2</sub> incubator (5% CO<sub>2</sub>, 37 °C) until microinjection with Nanoblades. Micro-injection were carried-out under a stereomicroscope (Olympus SZX9) using a FemtoJet 4i (Eppendorf) microinjector. Briefly, 1 pl of Nanoblades were injected in the perivitelline space of oocytes. Zygotes were then transferred into M16 medium and kept overnight in incubator. The embryos that reached the two-cell stage were transferred into the oviduct of B6CBAF1 (Charles River, France) pseudopregnant females (15–20 embryos per female).

**Retro-orbital injection of Nanoblades.** All experiments were performed in accordance with the European Union guidelines for approval of the protocols by the local ethics committee (Authorization Agreement C2EA 15, "Comité Rhône-Alpes d'Ethique pour l'Expérimentation Animale", Lyon, France). The highly immunosuppressed NOD FRG mice (Fah<sup>-/-</sup>/Rag2<sup>-/-</sup>/Il2rg<sup>-/-</sup>) (Yecuris coaration), deficient for T-cell, B-cell, and NK-cell are maintained in pathogen-free facility. Retro-orbital injection (SRO) were performed under isoflurane anesthesia.

Genomic DNA from each mouse (treated either by control or *Hpd* targeting Nanoblades) was extracted from three distinct liver lobes and pooled together.

Following this, a two-step PCR was performed on 300 ng of gDNA template, the first PCR using primers Hpd-Forward 1: 5'-CTTAGGAGGTTCAGCAAGATG GGAG-3' and Hpd-Reverse 1: 5'-TCTAGTCTCTATCCAGGGTCCAGCC-3' to amplify the *Hpd* gene (94 °C 5 min, 3 cycles 94 °C, 64 °C, 72 °C, and 20 cycles 94 °C, 58 °C, 72 °C, 5 min 72 °C). The second nested-PCR used primers Hpd-Forward 2: 5'-GAACTGGGATTGGCTAGTGCG-3' and Hpd-Reverse 2: 5'-CACCCAG CACCACCTATAGAAACTC-3' (94 °C 5 min, 3 cycles 94 °C, 64 °C, 72 °C and 30 cycles 94 °C, 57 °C, 72 °C, 5 min 72 °C). Amplicons were next analyzed by T7-endonuclease assay as described.

**Raw data files.** Uncropped scans of ethidium bromide gels and western-blotting figures are displayed in Supplementary Figure 7.

### Data availability

Gene Expression Omnibus: [GSE107035](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107035). The following plasmids will be available from Addgene: Gag::Cas9 fusion (BIC-Gag-CAS9, Plasmid ID: 119942), the Gag::Cas9-VPR fusion (BICstim-Gag-dCAS9-VPR, Plasmid ID: 120922) and the Gag::Cre fusion (GAG-CRErec, Plasmid ID: 119971).

Received: 11 December 2017 Accepted: 30 November 2018

Published online: 03 January 2019

### References

- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Sci. N. Y. NY* **337**, 816–821 (2012).
- Gheysens, D., Jacobs, E., de Foresta, F. & Thiriart, C. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* **59**, 103–112 (1989).
- Kaczmarczyk, S. J., Sitaraman, K., Young, H. A., Hughes, S. H. & Chatterjee, D. K. Protein delivery using engineered virus-like particles. *Proc. Natl Acad. Sci. USA* **108**, 16998–17003 (2011).
- Voelkel, C. et al. Protein transduction from retroviral Gag precursors. *Proc. Natl Acad. Sci. USA* **107**, 7805–7810 (2010).
- O'Connor, T. E., Rauscher, F. J. & Zeigel, R. F. Density gradient centrifugation of a murine leukemia virus. *Sci. N. Y. NY* **144**, 1144–1147 (1964).
- Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
- Gibbons, J. G., Branco, A. T., Yu, S. & Lemos, B. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. *Nat. Commun.* **5**, 4850 (2014).
- Kinner, A., Wu, W., Staudt, C. & Iliakis, G. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res.* **36**, 5678–5694 (2008).
- van Sluis, M. & McStay, B. A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev.* **29**, 1151–1163 (2015).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Zhang, J. et al. CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J. Virol.* **78**, 1448–1455 (2004).
- Okabe, M., Ikawa, M., Kominami, K. & Nakanishi, T. Green mice as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
- Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
- Amirache, F. et al. Mystery solved: VSV-G-LVs do not allow efficient gene transfer into unstimulated T cells, B cells, and HSCs because they lack the LDL receptor. *Blood* **123**, 1422–1424 (2014).
- Girard-Gagnepain, A. et al. Baboon envelope pseudotyped LVs outperform VSV-G-LVs for gene transfer into early-cytokine-stimulated and resting HSCs. *Blood* **124**, 1221–1231 (2014).
- Okimoto, T., Friedmann, T. & Miyano, A. VSV-G envelope glycoprotein forms complexes with plasmid DNA and MLV retrovirus-like particles in cell-free conditions and enhances DNA transfection. *Mol. Ther. J. Am. Soc. Gene Ther.* **4**, 232–238 (2001).
- Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51 (2012).
- Zuris, J. A., Thompson, D. B., Shu, Y. & Guilinger, J. P. Efficient delivery of genome-editing proteins in vitro and in vivo. *Nature* **33**, 73–80 (2015).
- Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
- Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
- Hori, T. et al. Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Sci. Rep.* **4**, 4513 (2014).
- Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J. & He, L. Highly efficient mouse genome editing by CRISPR ribonucleoprotein electroporation of zygotes. *J. Biol. Chem.* **291**, 14457–14467 (2016).
- Yoon, Y. et al. Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses. *Nat. Commun.* **9**, 412 (2018).
- Srinivas, S., Watanabe, T. & Lin, C. S. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Ldts* **1**, 4 (2001).
- Zhang, D. J. et al. Selective expression of the Cre Recombinase in late-stage thymocytes using the distal promoter of the *Lck* gene. *J. Immunol.* **174**, 6725–6731 (2005).
- Pankowicz, F. P. et al. Reprogramming metabolic pathways in vivo with CRISPR-Cas9 genome editing to treat hereditary tyrosinaemia. *Nat. Commun.* **7**, 12642 (2016).
- Azuma, H. et al. Robust expansion of human hepatocytes in *Fah<sup>-/-</sup>/Rag2<sup>-/-</sup>/Il2rg<sup>-/-</sup>* mice. *Nat. Biotechnol.* **25**, 903–910 (2007).
- Qazi, S. et al. Programmed self-assembly of an active P22-Cas9 nanocarrier system. *Mol. Pharm.* **13**, 1191–1196 (2016).
- Choi, J. G. et al. Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Ther.* **23**, 627–634 (2016).
- Wang, G. et al. Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. *Nat. Protoc.* **12**, 88–103 (2017).
- Modarai, S. R. et al. Efficient delivery and nuclear uptake is not sufficient to detect gene editing in CD34+ cells directed by a ribonucleoprotein complex. *Mol. Ther.—Nucleic Acids* **11**, 116–129 (2018).
- Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
- Lau, C.-H. & Suh, Y. In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. *F1000Res.* **6**, 2153 (2017).
- Yin, H. et al. Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).
- Rulli, S. J. et al. Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* **81**, 6623–6631 (2007).
- Frecha, C. et al. Stable transduction of quiescent T cells without induction of cycle progression by a novel lentiviral vector pseudotyped with measles virus glycoproteins. *Blood* **112**, 4843–4852 (2008).
- Szécsi, J. et al. Targeted retroviral vectors displaying a cleavage site-engineered hemagglutinin (HA) through HA-protease interactions. *Mol. Ther. J. Am. Soc. Gene Ther.* **14**, 735–744 (2006).
- Morizono, K. et al. Lentiviral vector retargeting to P-glycoprotein on metastatic melanoma through intravenous injection. *Nat. Med.* **11**, 346–352 (2005).
- Morizono, K. et al. Redirecting lentiviral vectors pseudotyped with Sindbis virus-derived envelope proteins to DC-SIGN by modification of N-linked glycans of envelope proteins. *J. Virol.* **84**, 6923–6934 (2010).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
- Agudelo, D. et al. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods* **14**, 615–620 (2017).
- Massouridès, E. et al. Dp412e: a novel human embryonic dystrophin isoform induced by BMP4 in early differentiated cells. *Skelet. Muscle* **5**, 40 (2015).
- Carpenter, S. et al. A long noncoding RNA mediates both activation and repression of immune response genes. *Sci. N. Y. NY* **341**, 789–792 (2013).
- Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS ONE* **9**, e108424 (2014).
- Mangeot, P.-E. et al. Protein transfer into human cells by VSV-G-induced nanovesicles. *Mol. Ther. J. Am. Soc. Gene Ther.* **19**, 1656–1666 (2011).
- Heyer, E. E., Ozadam, H., Ricci, E. P., Cenik, C. & Moore, M. J. An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res.* **43**, e2 (2015).
- Ricci, E. P. et al. Staufeni senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol.* **21**, 26–35 (2014).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).



50. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
51. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

### Acknowledgements

Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the ‘France Génomique’ consortium (ANR-10-INBS-0009). We acknowledge the contribution of SFR Biosciences (UMS3444/CNRS, US8/Inserm, ENS de Lyon, UCBL) facilities: Platim and PBES (Celphedia, AniRA). We also thank J.F. Henry, N. Aguilera, and J.L. Thoumas from the animal facility (PBES, Plateau de Biologie Expérimentale de la Souris, ENS de Lyon), as well as A. Ollivier for their technical help in handling mice. We thank Claire Lionnet from Platim for technical assistance in taking confocal fluorescence images. We thank Elisabeth Errazuriz-Cerda and the CeCIL-facility (Lyon, France) for the preparation and the observation of samples by TEM and Yohann Couté and the edyp-service (Grenoble, France) for the proteomic analysis. We thank Gérard Benoît for his help in preparing final figures. This work was funded by Labex Ecofect (ANR-11-LABX-0048) of the Université de Lyon, within the program Investissements d’Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR), Fondation FINOVI and Agence Nationale des Recherches sur le SIDA et les Hépatites Virales (ANRS—ECTZ3306) to E.P.R. Open access fees were funded by the European Research Council (ERC-StG-LS6-805500 to E.P.R.) under the European Union’s Horizon 2020 research and innovation programs.

### Author contributions

P.E.M. and E.P.R. conceived the study and designed most experiments. P.E.M., E.P.R., V.R., A.M., E.L., F.F., E.V., F.L.C., T.S. and F.A. designed experiments. P.E.M., E.P.R., E.L., V.R., A.M., F.F., T.S., F.A., J.B., E.V., V.M., M.T., and E.M. performed experiments and analyzed data. P.E.M. and E.P.R. wrote the paper with contributions from all authors.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-07845-z>.

**Competing interests:** P.E.M., T.O., and E.P.R. are named as inventors on a patent relating to the Nanoblades technology (patent applicants: Institut National de la Santé et de la Recherche Médicale (INSERM), Centre National de la Recherche Scientifique (CNRS), Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, Villeurbanne Cedex; name of inventors: Theophile Ohlmann, Mathieu Misery, Philippe Mangeot, Emiliano Ricci; application number: WO 2017/068077 A1; patent status: published, 27th April 2017; all aspects of the manuscript are covered by the patent application. The remaining authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019



## **Annex 2**

### **An intimate look at the viral replication cycle through ribosome profiling.**

**BLIN J**, RICCI EP. Med Sci (Paris). 2016 Oct;32(10):849-860. Epub 2016 Oct 19. (Review in French)

► L'explosion du nombre de techniques basées sur le séquençage massif parallèle est actuellement en train de révolutionner l'étude des systèmes biologiques en permettant à l'expérimentateur d'avoir une vision globale des processus se déroulant à l'échelle moléculaire. Parmi ces nouvelles approches, le profilage ribosomique est un outil particulièrement puissant pour l'étude de la traduction à un niveau de détail jamais égalé auparavant. Cette technique permet notamment de cartographier très précisément la position des ribosomes sur l'ensemble des ARN messagers en cours de traduction dans la cellule à un moment donné. Dans le cas d'une infection virale, il est ainsi possible d'étudier les mécanismes souvent très complexes et encore mal compris qui sont mis en place par les virus pour assurer la production des protéines nécessaires à leur multiplication. Cette synthèse a pour but de discuter la manière dont le profilage ribosomique peut nous permettre de mieux comprendre le cycle de réplication virale, mais aussi de montrer les biais liés à la technique à prendre en compte lors de l'analyse des résultats. ◀

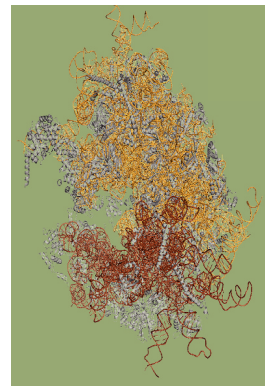
Les virus sont des « parasites » intracellulaires obligatoires qui dépendent de la cellule infectée pour se multiplier. Cette dépendance se traduit à différents niveaux au cours d'une infection virale et notamment lors de la synthèse des protéines du virus. En effet, les virus sont dépourvus de ribosomes qui sont essentiels pour la synthèse de protéines à partir des ARN messagers (ARNm). La traduction est donc une étape critique du cycle de réplication virale pendant laquelle la cellule est trompée afin de traduire les ARNm viraux en dépit de certaines différences structurelles qu'ils présentent avec les ARNm cellulaires. En effet, les contraintes évolutives, induites par le volume limité des particules virales, ont conduit les

Vignette (Photo © Lionel Tafforeau).

## Le profilage ribosomique

### Une technique nouvelle génération pour l'étude de la traduction au cours d'une infection virale

Juliana Blin<sup>1-5</sup>, Emiliano P. Ricci<sup>1-5</sup>



<sup>1</sup> CIRI, international center for infectiology research, université de Lyon, 46, allée d'Italie, Lyon, France ;  
<sup>2</sup> Inserm, U1111, Lyon, France ;  
<sup>3</sup> École normale supérieure de Lyon, 46, allée d'Italie, 69007 Lyon, France ;  
<sup>4</sup> Université Claude Bernard Lyon 1, centre international de recherche en infectiologie, Lyon, France ;  
<sup>5</sup> CNRS, UMR5308, Lyon, France.  
[emiliano.ricci@ens-lyon.org](mailto:emiliano.ricci@ens-lyon.org)  
[emiliano.ricci@inserm.fr](mailto:emiliano.ricci@inserm.fr)

acides nucléiques viraux à être organisés de façon extrêmement compacte afin de contenir un maximum d'informations génétiques dans un espace réduit [1]. L'expression des gènes viraux se fait selon des mécanismes complexes qui ne sont pas tous encore compris. La compréhension de ces mécanismes ainsi que de ceux impliqués dans la subversion de la cellule est particulièrement importante pour la mise en place de thérapies antivirales plus ciblées et efficaces. Le profilage ribosomique, ou *ribosome profiling* en anglais, est une technique innovante permettant de suivre à la trace la position de l'ensemble des ribosomes présents sur les ARNm à un moment donné. Elle a récemment permis de faire des avancées considérables dans l'étude de la traduction au niveau cellulaire [2], notamment au cours d'une infection virale [3]. Cette méthode s'inspire des techniques d'empreintes, ou *footprinting*, où l'acide nucléique, soumis à une digestion enzymatique, est protégé localement par les protéines qui lui sont associées. En comparant la séquence de départ à la séquence digérée, il est ainsi possible de retrouver très précisément la position d'un complexe protéique sur n'importe quelle séquence nucléotidique. En couplant cette approche aux méthodes de séquençage à haut-débit de nouvelle génération, on peut désormais étudier en détail la traduction de l'ensemble des ARNm exprimés dans la cellule [4]. Dans cette synthèse, nous discuterons de la manière dont le profilage ribosomique a révolutionné l'étude de la traduction à l'échelle cellulaire et de la façon dont cette technique peut nous permettre de mieux comprendre le cycle de réplication virale. Nous nous intéresserons également aux limitations et aux biais liés à cette méthode qui doivent être pris en compte afin d'obtenir des résultats de qualité.

## Le profilage ribosomique dans l'étude de la traduction cellulaire

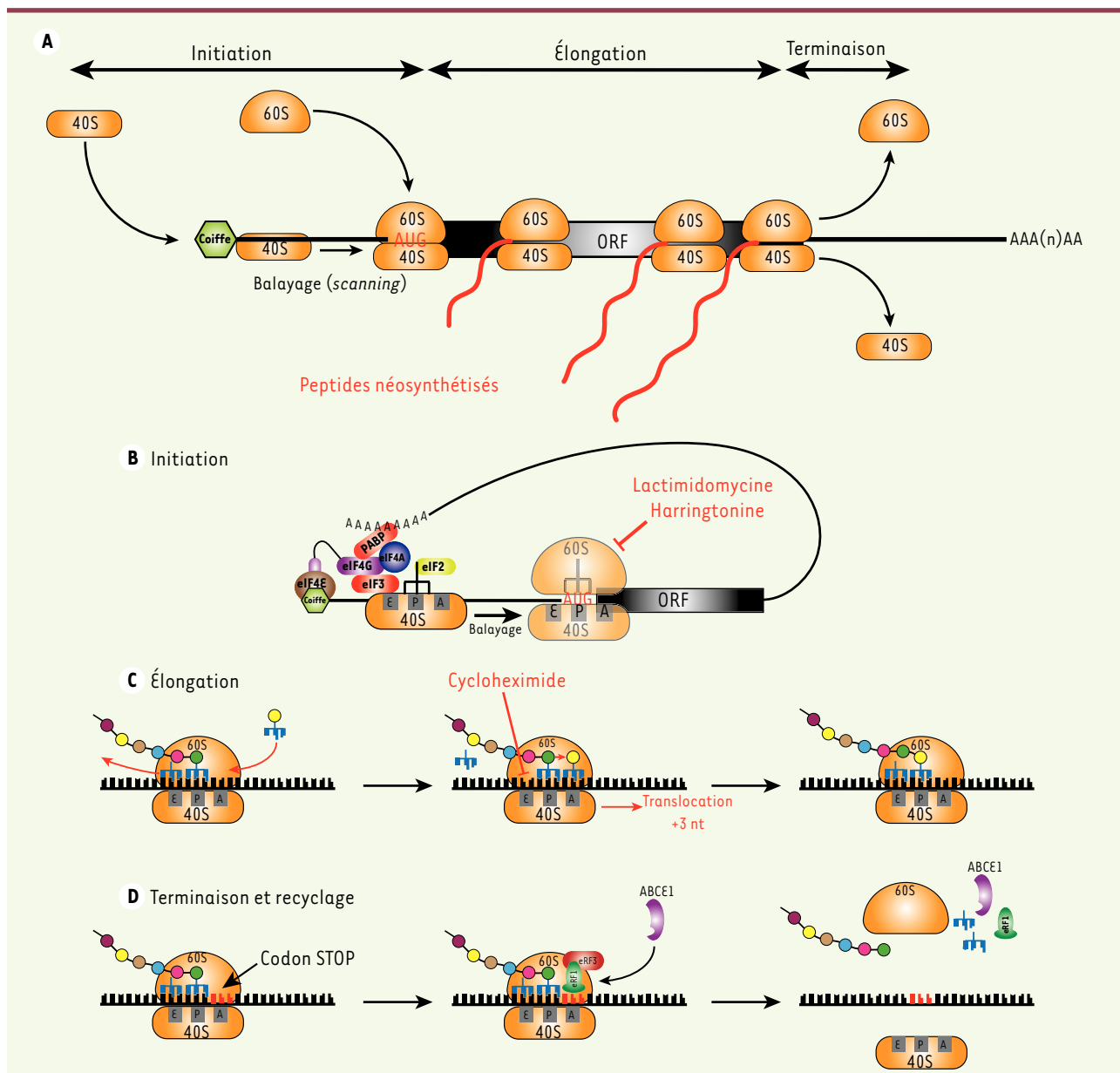
La synthèse des protéines est assurée par les ribosomes [43] (→), des complexes macromoléculaires composés à la fois de protéines et d'ARN ribosomaux (ARNr), capables de reconnaître les ARNm et de les traduire. Cette reconnaissance est rendue possible grâce à la présence de structures conservées telles que la coiffe méthylée ( $m^7GppN-$ ,  $N^7$ -méthylguanosine-triphosphate) et la queue de poly-adénosines, ou poly(A), sur les ARNm cellulaires. Le processus de traduction se divise en quatre étapes majeures que sont l'initiation, l'élongation, la terminaison et le recyclage des ribosomes et des facteurs de traduction (Figure 1A). La première étape, l'initiation, consiste à recruter un ribosome fonctionnel sur le site d'initiation de la traduction situé le plus souvent au niveau du premier codon AUG de l'extrémité 5' de l'ARNm [5]. Pour cela, la petite sous-unité du ribosome (40S), associée à un ARN de transfert (ARNt) initiateur, et guidée par différents facteurs d'initiation (regroupés sous le nom d'eIF ou *eukaryotic initiation factors*), se fixe au niveau de la coiffe pour former un complexe qui va balayer l'ARNm jusqu'à atteindre un codon AUG (Figure 1B). Lorsqu'il l'atteint, le complexe s'arrête et recrute la grosse sous-unité ribosomale (60S) formant ainsi un ribosome (80S) prêt à démarrer la synthèse protéique [6]. L'étape suivante, l'élongation, consiste en l'ajout successif d'acides aminés dans la chaîne polypeptidique en cours de synthèse par le ribosome (Figure 1C). Les acides aminés sont apportés un à un par des ARNt portant chacun un acide aminé correspondant au codon en cours de lecture par le ribosome. L'élongation va ainsi se poursuivre jusqu'à ce que le ribosome atteigne un codon stop (UAG, UGA ou UAA). Lors de l'étape de terminaison, le codon stop est reconnu par le facteur de terminaison eRF1 (*eukaryote release factor 1*) qui, avec l'aide du facteur eRF3, catalyse l'hydrolyse de la liaison entre le dernier acide aminé incorporé et l'ARNt qui l'apporte [7] (Figure 1D). La protéine nouvellement formée est ainsi libérée du ribosome qui est ensuite dissocié grâce à l'action conjointe des facteurs ABCE1 (*ATP binding cassette E1*) et eRF1 (Figure 1D). Les différents composants du ribosome sont enfin recyclés pour permettre la traduction de nouveaux ARNm. La plupart des ARNm peuvent être traduits par plusieurs ribosomes en même temps, formant des structures appelées polysomes, ce qui augmente la quantité de protéines pouvant être produite à partir d'un même message.

Chez les eucaryotes, différents mécanismes de contrôle de la traduction existent pour réguler la synthèse protéique en fonction des stimulus perçus par la cellule, comme lors de la différenciation cellulaire ou d'un stress [8]. L'expression individuelle d'un gène, ou de toute une classe de gènes, peut ainsi être modulée afin de permettre à la cellule de s'adapter à son environnement, en conditions physiologiques comme pathologiques. Le profilage ribosomique est une technique particulièrement utile pour étudier les mécanismes de contrôle traductionnel mis en place par la cellule pour s'adapter dans différents contextes, y compris au cours d'une infection virale [3]. Cette

(→) Voir la Nouvelle de V. Marcel et al., m/s n° 1, janvier 2014, page 21

approche permet, notamment, d'étudier les processus moléculaires mis en jeu pour contrôler la synthèse protéique à un niveau de détail jamais égalé auparavant en s'appuyant sur le séquençage à haut-débit des ARNm en cours de traduction. L'avènement du séquençage massif parallèle, ou *next generation sequencing* (NGS), a en effet révolutionné de nombreux domaines de la biologie en permettant aux expérimentateurs d'avoir une vision globale des processus cellulaires faisant intervenir des acides nucléiques [9]. Le principe de ce type d'approche est de séquencer un grand nombre de petits fragments nucléotidiques, généralement plusieurs millions de fragments de quelques dizaines à quelques centaines de nucléotides de long. Cette technologie a été rapidement adoptée pour l'étude de l'expression des gènes à différents niveaux. En effet, un des avantages majeurs du NGS, comparé aux méthodes préexistantes, comme les *microarrays*, est qu'il permet la détection de l'ensemble des transcrits présents dans un échantillon de manière très sensible et sans aucun *a priori*. Ces dernières années, la grande versatilité des plates-formes de NGS et la baisse continue de leur coût d'utilisation a favorisé le développement de nombreux protocoles dédiés à l'étude des ARN dans différents contextes. Parmi ces nouvelles approches, le *profilage ribosomique*, qui permet donc de suivre très précisément la traduction de l'ensemble des ARNm cellulaires à un instant donné, consiste à cartographier les sites d'interaction de la sous-unité 80S des ribosomes sur les ARNm en cours de traduction (Figure 2). Cela est rendu possible par le fait que le ribosome protège l'ARNm auquel il est lié sur une zone d'environ 30 nucléotides (Figure 2B) [10]. De ce fait, les régions nues de l'ARNm sont plus sensibles à la dégradation que celles qui sont associées aux ribosomes, notamment lors d'un traitement avec des nucléases. Le séquençage à haut-débit des fragments obtenus après digestion enzymatique de polysomes permet ainsi de déterminer la position exacte de chaque ribosome associé à un ARNm et pour un ARNm donné, la quantité de ribosomes qui sont en train de le traduire [4, 11].

D'un point de vue pratique, la première étape du profilage ribosomique consiste à bloquer la progression de la traduction dans la cellule grâce à l'ajout d'inhibiteurs de la traduction ou par incubation à froid des échantillons (Figure 2A). Ces inhibiteurs agissent soit en bloquant l'avancée des ribosomes sur les ARNm et en empêchant l'élongation, comme c'est le cas de la cycloheximide (Figure 1C), soit en ciblant l'initiation de la traduction et en figeant les ribosomes sur le codon d'initiation, comme c'est le cas de la lactimidomycine ou de l'haringtonine [12] (Figure 1B). Lorsque l'élongation est



**Figure 1. Les différentes étapes de la traduction chez les eucaryotes.** **A.** Vision globale des étapes principales de la traduction coiffe-dépendante : initiation, élongation et terminaison. **B.** Au cours de l'initiation, la petite sous-unité ribosomale (40S), associée au ARNt-initiateur méthionine et à eIF2 (*eukaryotic initiation factors*2), est recrutée au niveau de la coiffe par une multitude de facteurs d'initiation. Parmi ceux-ci, eIF4E lie directement la coiffe située à l'extrémité 5' des ARNm cellulaires et interagit avec eIF4G qui joue un rôle d'échafaudage en interagissant avec de multiples autres facteurs comme eIF3 (qui permet de faire le lien avec le ribosome 40S) et PABP (*poly(A)-binding protein*) qui lie la queue poly(A) située à l'extrémité 3' de l'ARNm et qui permet de circulariser l'ARNm. Une fois le ribosome recruté sur l'ARNm, il balaye la région 5' non traduite dans le sens 5' vers 3' aidée par l'activité hélicase d'eIF4A qui permet de dérouler les éventuelles structures secondaires de l'ARN qui pourraient bloquer sa progression. Le balayage s'arrête lorsque le ribosome 40S arrive au niveau d'un codon AUG situé dans un bon contexte nucléotidique (aussi appelé contexte de Kozak). Une fois placé sur le codon AUG, les facteurs d'initiation sont relâchés et la grande sous-unité ribosomale (60S) est recrutée pour former un ribosome 80S. **C.** Lors de l'étape d'élongation, un ARNt chargé entre dans le site A du ribosome. Le ribosome catalyse la liaison peptidique entre l'acide aminé porté par l'ARNt du site A et le peptide situé sur l'ARNt du site P. Une fois la liaison effectuée, le ribosome réalise une translocation de 3 nucléotides en direction 3' permettant le déplacement de l'ARNt qui était auparavant sur le site P vers le site E et l'arrivée de l'ARNt couplé à la chaîne peptidique dans le site P. Cette translocation libère le site A qui peut donc accueillir un nouvel ARNt chargé. **D.** Lorsque le ribosome arrive au niveau d'un codon stop, le facteur de terminaison eRF1 (*eukaryote release factor 1*) associé à eRF3 s'insère dans le site A du ribosome et induit le relargage de la protéine néosynthétisée. La dissociation des deux sous-unités ribosomales est ensuite catalysée par l'action conjointe de ABCE1 (*ATP binding cassette E1*) et d'eRF1. ORF : *open reading frame* ; nt : nucléotide.

bloquée, la mesure de la densité de ribosomes associés à une région codante donne une mesure indirecte de l'efficacité de traduction du transcrite (paramètre qui dépend à la fois du nombre de ribosomes associés à l'ARNm et de la vitesse d'élongation) [13]. De plus, en étudiant l'empreinte obtenue pour chaque ribosome, il est possible de retrouver la phase de lecture de ces ribosomes sur une région codante. En couplant cette approche au séquençage des ARNm, il est ainsi possible de caractériser l'ensemble du transcriptome cellulaire et de cartographier les régions du génome qui sont codantes [14] (Figure 2C). L'utilisation d'inhibiteurs de l'initiation a plutôt pour application la localisation exacte des sites d'initiation de la traduction mais aussi la caractérisation de la vitesse d'élongation des ribosomes [11]. En effet, lors de l'ajout de ces drogues, les ribosomes en phase d'élongation vont poursuivre la traduction jusqu'à atteindre le codon stop selon un phénomène connu sous le nom de *ribosome run-off*. En bloquant l'élongation à différents temps après l'ajout de l'inhibiteur d'initiation, il est possible de mesurer la distance parcourue par les ribosomes depuis le site d'initiation et ainsi de déterminer la vitesse d'élongation et les sites de pause du ribosome.

Depuis le développement de cette technique, en 2009 [4], l'étude de la traduction par profilage ribosomique a permis de mettre en évidence un grand nombre de mécanismes de régulation de la traduction au niveau cellulaire. Ainsi, il a été montré que les cellules faisaient régulièrement appel à des mécanismes de contrôle traductionnel originaux tels que l'utilisation de codons d'initiation non canoniques (différents du codon AUG), de petits cadres de lecture ouverts dans les régions non traduites en 5' des ARNm, connus sous le nom d'*upstream open reading frames* ou uORF, ou encore de cadres de lecture alternatifs dans les ARNm [2]. Le profilage ribosomique a aussi permis de caractériser finement les changements globaux de traduction qui se produisent en cas de stress cellulaire, comme au cours d'un choc thermique [15], d'un stress oxydatif [16] ou d'une infection virale [17].

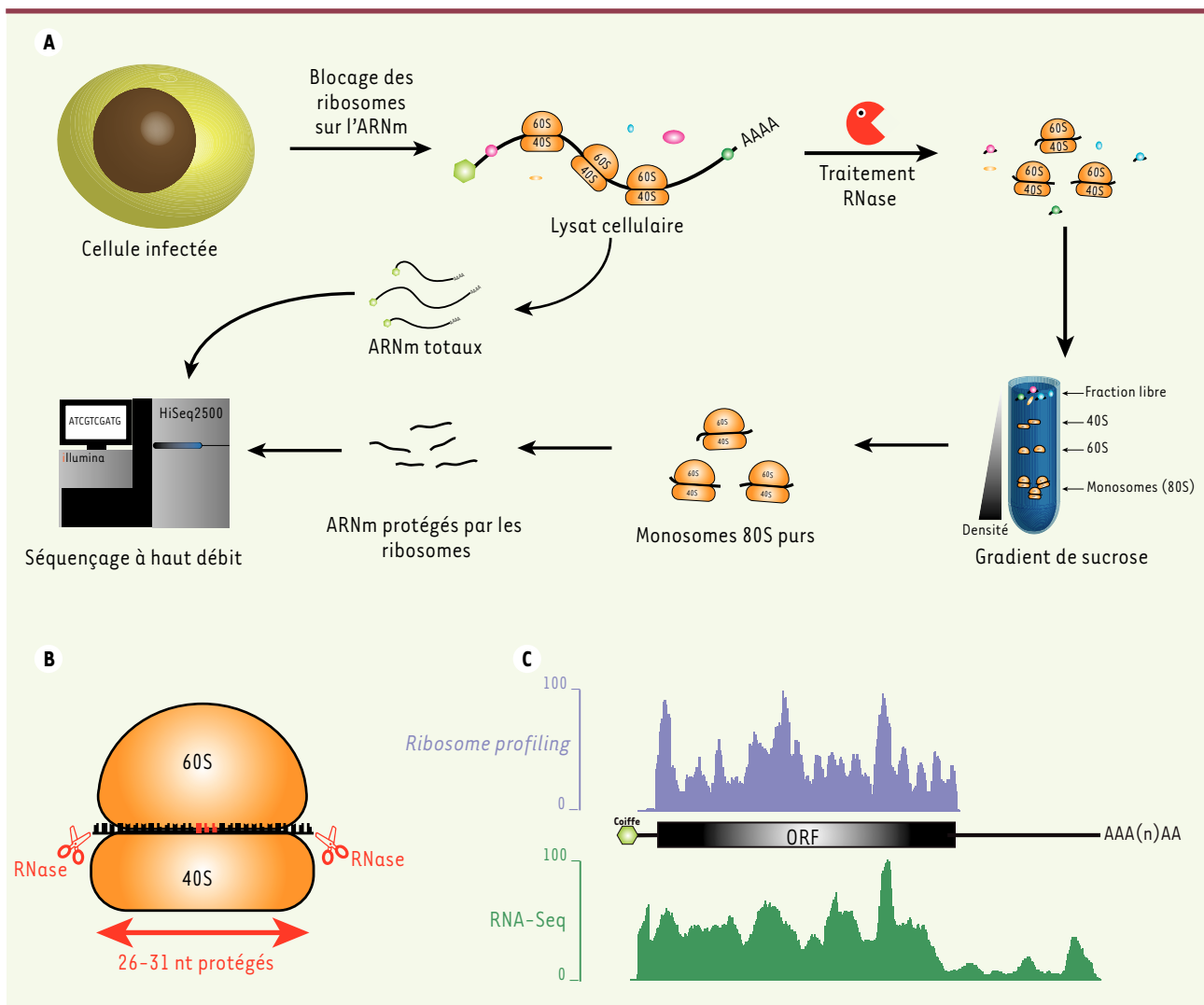
### L'étude de la traduction des ARNm viraux par profilage ribosomique

La traduction est une étape critique au cours d'une infection virale durant laquelle les ARNm viraux sont en compétition directe avec les ARNm cellulaires pour s'associer aux ribosomes. De nombreux ARNm viraux présentent des caractéristiques qui réduisent leur capacité à être traduits par la machinerie cellulaire, comme l'absence de coiffe ou de queue poly(A), ou encore la présence de structures secondaires essentielles pour la réplication des ARN viraux mais qui peuvent bloquer la progression des ribosomes. Cela est particulièrement vrai pour les virus à ARN. En conséquence, les virus ont évolué de manière à développer des mécanismes alternatifs pour assurer l'expression efficace de leurs protéines. Leur principale stratégie consiste à cibler les différents facteurs de traduction cellulaires. De nombreux mécanismes de subversion interviennent notamment au niveau de l'initiation de la traduction, car c'est à cette étape que la régulation est la plus importante [18]. Ainsi, la capacité à recruter les ribosomes sur les ARN viraux est capitale pour le succès de l'infection [19]. Si de

nombreux mécanismes de détournement de la cellule-hôte ont déjà été mis en évidence par le passé, leurs conséquences sur l'expression des ARNm cellulaires restent aujourd'hui encore assez mal comprises. Le profilage ribosomique est une technique particulièrement prometteuse pour l'analyse approfondie des mécanismes de détournement de la cellule au cours d'une infection virale [3, 20]. Cette méthode peut notamment s'appliquer à l'étude des infections virales *in vitro*, sur des cultures de cellules infectées, mais aussi *ex vivo*, en travaillant directement à partir de tissus animaux infectés. Dans la suite de cette revue, nous allons détailler plusieurs mécanismes d'expression des gènes viraux et de détournement de la cellule-hôte qui peuvent être étudiés par le profilage ribosomique en décrivant les avantages par rapport aux techniques classiques d'étude de la traduction.

### Initiation de la traduction coiffe-indépendante par les IRES

Les IRES (*internal ribosome entry site*) correspondent à des séquences portées par les ARNm capables de recruter la sous-unité 40S du ribosome indépendamment de la présence d'une coiffe à l'extrémité 5' du transcrite (Figure 3A). En dépit de leur grande variabilité au niveau de leur séquence primaire, la plupart des IRES possèdent des régions fortement structurées permettant une interaction directe avec le ribosome ou certains facteurs d'initiation de la traduction. La variabilité des séquences IRES se traduit aussi par une forte diversité de mécanismes moléculaires utilisés pour l'initiation de la traduction. Ainsi, toutes les IRES ne vont pas recruter les mêmes facteurs d'initiation de la traduction et certaines sont même capables de s'en affranchir totalement. Dans la cellule eucaryote, les séquences IRES sont présentes dans un nombre limité de transcrits et permettent d'assurer le maintien de la production de protéines importantes en conditions de stress. Ces séquences sont essentielles pour la synthèse des protéines viraux à partir d'ARNm naturellement dépourvus de coiffe, comme c'est le cas pour de nombreux virus à ARN. Elles sont également retrouvées dans des transcrits viraux portant une coiffe et une queue poly(A). Dans ce cas, elles permettent la production de protéines viraux même lorsque la traduction coiffe-dépendante est perturbée, comme par exemple lors du phénomène de *host shut-off* que nous décrirons dans la suite de cette revue [1]. Par ailleurs, les séquences IRES peuvent permettre d'accroître le nombre de protéines différentes codées par un même messager, notamment grâce à la traduction de l'ARNm viral à la fois par la voie canonique et *via* une IRES. Cette stratégie est retrouvée



**Figure 2. Étude globale de la traduction cellulaire par profilage ribosomique.** **A.** Les cellules ou tissus d'intérêt sont traités avec des inhibiteurs de la traduction comme le cycloheximide dans le but de figer les ribosomes sur les ARNm qu'ils traduisent. Les cellules ou tissus d'intérêt sont ensuite lysés et la fraction cytoplasmique récupérée (compartiment où a lieu la traduction des ARNm). Les lysats obtenus sont ensuite séparés en deux fractions. La première est utilisée pour extraire les ARNm et mesurer leur niveau d'expression par séquençage massif parallèle (RNA-seq, RNA-Seq). La deuxième fraction est incubée avec des ribonucléases (généralement avec la RNase I qui clive l'ARN simple brin quelle que soit la séquence nucléotidique) qui vont dégrader toutes les régions d'ARN accessibles à l'exception de celles qui sont associées aux protéines liant l'ARN et celles associées aux ribosomes 80S. Les ribosomes étant plus denses et volumineux que les protéines libres et autres complexes ribonucléoprotéiques, ils peuvent être purifiés sur un gradient de sucrose, sur coussin de sucrose ou par chromatographie d'exclusion. Une fois les ribosomes 80S isolés, les fragments d'ARNm protégés sont récupérés puis utilisés pour préparer des banques d'ADN complémentaire (ADNc) qui seront finalement séquencées. Une fois les résultats du séquençage obtenus, les séquences protégées par le ribosome ainsi que les séquences correspondant au RNA-Seq (ARNm en entier) sont alignées contre le génome de référence correspondant à l'espèce du matériel de départ. Une fois les alignements obtenus, l'expérimentateur peut mesurer la densité de ribosomes par molécule d'ARNm en calculant le ratio entre le nombre de séquences protégées par les ribosomes pour un ARNm donné et le nombre de séquences obtenues pour le même ARNm à partir du RNA-Seq. **B.** Les fragments protégés font généralement entre 26 et 31 nt (nucléotides) de long (la majorité des fragments font 28 nt de long). En conséquence, il est possible de retrouver la phase de lecture en prenant comme information la position du premier nucléotide de chaque séquence. **C.** Exemple de résultat obtenu par profilage ribosomique et RNA-Seq sur le gène humain codant la glyceraldéhyde-3-phosphate déhydrogénase (GAPDH). Sur le panel du haut (ribosome profiling), le nombre de fragments protégés par le ribosome est présenté en ordonnée par rapport à chaque position du transcrite codant la GAPDH présentée en abscisse. Sur le panel du bas (RNA-Seq), est présenté le nombre de fragments d'ARNm séquencés pour chaque position du transcrite codant la GAPDH. On peut observer que le signal correspondant au profilage ribosomique est limité à la région codante du transcrite alors que le signal correspondant au RNA-Seq s'étend sur les régions 5' et 3' non traduites. ORF : *open reading frame*.

par exemple chez les virus de l'immunodéficience humaine (VIH-1 et VIH-2) qui sont capables d'exprimer plusieurs isoformes de la protéine structurale Gag en utilisant à la fois les mécanismes de traduction coiffe-dépendant et IRES-dépendant [21].

Historiquement, l'étude de la traduction IRES-dépendante chez les virus a été réalisée en utilisant des gènes rapporteurs monocistroniques (portant la séquence IRES d'intérêt au niveau de la région 5'UTR [5' *untranslated region*]), ou bicistroniques (portant la séquence IRES entre deux gènes rapporteurs). Ces constructions étaient ensuite utilisées dans des systèmes de traduction *in vitro*, ou transfectées dans des cellules en culture. Ces méthodes ont permis de caractériser l'activité IRES dans de nombreux virus en rendant possible l'étude détaillée de leurs besoins en facteurs d'initiation de la traduction et la localisation précise du site d'initiation. Cependant, lorsque l'on utilise ce type d'approche, de nombreux contrôles sont nécessaires afin de valider l'activité IRES d'une séquence donnée et d'exclure les artefacts dus à la présence de promoteurs ou de sites d'épissage cryptiques dans la région supposée contenir une IRES. Le profilage ribosomique pourrait faciliter la détection et l'étude des séquences IRES directement au cours de l'infection virale et donc dans un contexte plus physiologique. Notamment, dans le cas de virus induisant un blocage de la traduction coiffe-dépendante, le profilage ribosomique permettrait d'identifier les ARNm cellulaires et viraux dont la traduction est résistante à ce blocage, potentiellement grâce à une activité IRES. En couplant cette approche avec des techniques complémentaires permettant une analyse poussée de l'expression des transcrits telles que le séquençage ARN ou le CAGE-Seq (*cap analysis gene expression*, permettant la cartographie de l'extrémité 5' des transcrits), il est même possible d'aller encore plus loin dans la caractérisation des séquences contenant des IRES en excluant celles qui correspondent à des artefacts liés à la présence de promoteurs ou de sites d'épissage cryptiques.

### **Host shut-off ou modulation de la traduction des ARNm de la cellule-hôte**

Comme décrit précédemment, la grande majorité des ARNm cellulaires sont traduits selon le mécanisme conventionnel coiffe-dépendant (Figure 1B). En ciblant ce mode de traduction, le virus peut donc perturber la production globale des protéines cellulaires. Cette stratégie, nommée *host shut-off*, permet aux virus qui sont capables d'utiliser une voie alternative d'initiation de la traduction, comme les IRES, de se débarrasser de toute compétition pour le recrutement des ribosomes sur leurs ARNm. Certains virus vont par exemple cibler les zones conservées des ARNm cellulaires pour empêcher leur traduction. C'est le cas des poxvirus qui expriment des enzymes de dégradation de la coiffe permettant de réduire le *pool* d'ARNm traduits selon la voie coiffe-dépendante (Figure 3B) [22]. D'autres virus, comme les picornavirus et les rétrovirus, vont quant à eux cibler des facteurs essentiels pour l'initiation de la traduction coiffe-dépendante, tels que eIF4G et PABP (*poly(A)-binding protein*), et induire leur dégradation (Figure 3B) [18]. Ce phénomène d'extinction de la synthèse protéique cellulaire joue aussi un rôle dans l'échappement face aux défenses immunitaires. En effet, en réduisant le taux de synthèse protéique

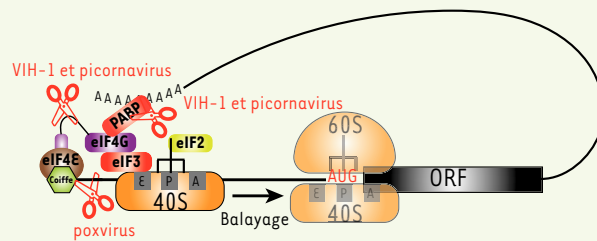
global, l'expression de protéines nécessaires à la mise en place d'une réponse immunitaire antivirale efficace est également limitée [20]. Le profilage ribosomique est particulièrement utile pour étudier ce phénomène, car il permet de quantifier simultanément et précisément le taux de traduction des ARNm cellulaires et viraux au cours d'une infection. Il est ainsi possible de caractériser les mécanismes moléculaires mis en place pour induire un *host shut-off* et de mesurer leur impact sur la traduction cellulaire. En couplant cette approche avec la mesure des niveaux d'expression des transcrits cellulaires et viraux en temps réel par séquençage à haut débit, il a été récemment montré que le virus HSV-1 (*herpes simplex virus 1*) était capable de perturber spécifiquement l'étape de maturation des ARNm cellulaires pour conduire au *shut-off* des protéines cellulaires et favoriser l'expression de certains de ses ARNm dont la maturation n'est pas sensible au blocage qu'il induit [23]. Une étude de profilage ribosomique récemment réalisée chez un coronavirus murin a, au contraire, montré un mécanisme alternatif pour augmenter la production de protéines virales lors du cycle de réplication [24]. Dans ce cas, la transcription des ARN viraux est tellement importante qu'ils représentent jusqu'à 90 % de l'ensemble des ARN codants dans la cellule hôte. Cette stratégie permet ainsi au virus de produire de grandes quantités de protéines virales sans que l'efficacité de traduction de ses ARNm ne soit plus importante que celle des ARNm cellulaires.

### **Leaky scanning et initiation de la traduction à partir de codons non-AUG**

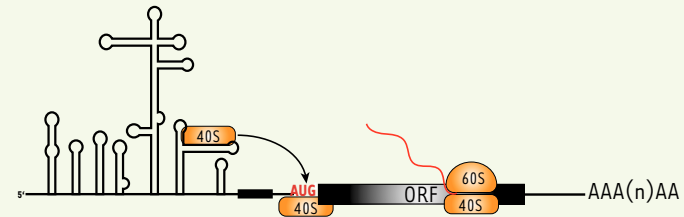
Un autre mécanisme de régulation de l'initiation de la traduction qui peut être étudié par profilage ribosomique est le *leaky scanning* (Figure 3C). Dans la voie conventionnelle, le complexe d'initiation balaye l'extrémité 5' de l'ARNm et s'arrête au niveau du premier codon d'initiation AUG qu'il rencontre avant de former un ribosome 80S et de débiter l'élongation (Figure 1B) [6]. Dans le cadre du *leaky scanning*, une partie des sous-unités 40S ne s'arrête pas au niveau du premier codon d'initiation de la traduction mais continue de balayer l'ARNm jusqu'à atteindre le codon AUG suivant (Figure 3C). Le *leaky scanning* est utilisé régulièrement par les cellules eucaryotes comme mécanisme de contrôle traductionnel notamment pour réguler l'expression des protéines de réponse au stress. Ce mécanisme est aussi fréquemment utilisé par les virus, et particulièrement par les virus à ARN, afin d'assurer la traduction d'ARNm viraux polycistroniques, qui codent pour plusieurs protéines à la fois [1]. Ainsi, la présence de codons AUG dans un cadre de lecture différent de



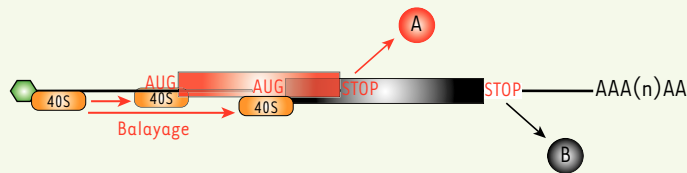
### A Blocage de la traduction cellulaire (*shut-off*)



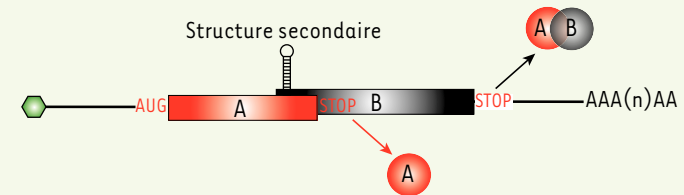
### B Traduction IRES-dépendante



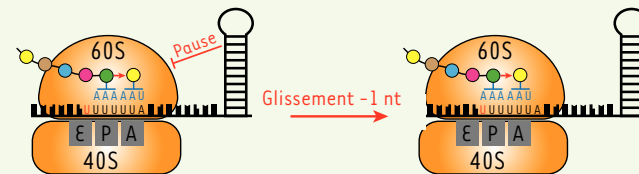
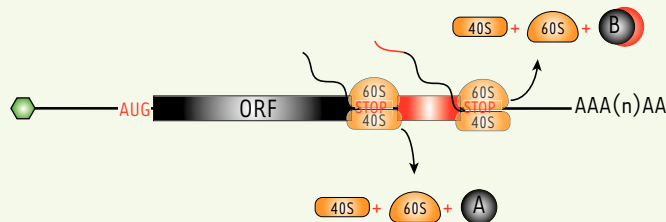
### C Leaky scanning



### D Glissement de phase de lecture (*frameshift*)



### E Readthrough du codon STOP



**Figure 3. Divers mécanismes de régulation de la traduction chez les eucaryotes.** **A.** Les séquences IRES (*internal ribosome entry site*) permettent de recruter la petite sous-unité du ribosome (40S) indépendamment de la coiffe et de certains facteurs d'initiation de la traduction et sont insensibles aux mécanismes de blocage de la traduction mentionnés ci-dessus. **B.** Blocage de la traduction coiffe-dépendante cellulaire par les virus. Certains virus, comme les picornavirus, les VIH (virus de l'immunodéficience humaine) ou les poxvirus, sont capables d'induire la protéolyse de certains facteurs d'initiation de la traduction essentiels à la traduction coiffe-dépendante ou d'induire le clivage de la coiffe à l'extrémité 5' des ARNm et d'inhiber la traduction de la plupart des transcrits cellulaires. **C.** Lors du *leaky scanning*, certains ribosomes 40S ne reconnaissent pas le premier codon AUG situé en 5' de l'ARNm et continuent de balayer la 5'UTR (*untranslated region*) pour initier la traduction au niveau d'un codon AUG situé en aval. De cette façon, plusieurs protéines ou isoformes protéiques peuvent être synthétisées à partir du même ARNm. **D.** Le glissement de phase de lecture se produit généralement lorsque le ribosome 80S est ralenti par une structure secondaire d'ARN, se décale d'un ou deux nucléotides (nt) puis reprend l'élongation dans une autre phase de lecture. **E.** Lors du *readthrough*, les ribosomes 80S échouent à reconnaître le codon stop en incorporant un acide aminé à la place et continuent l'élongation en aval permettant la synthèse d'une isoforme de la protéine avec une extension C-terminale. ORF : *open reading frame* ; eIF : *eukaryotic initiation factors* ; PABP : *poly(A)-binding protein*.



celui du codon initiateur canonique, qui peuvent être reconnus par *leaky scanning*, permet la production simultanée de différentes protéines à partir d'un même ARNm (Figure 3C). Un exemple connu est le cas des papillomavirus humains (HPV) qui utilisent le *leaky scanning* pour assurer la production des oncoprotéines virales E6 et E7 à partir d'un même messenger [25]. Grâce à l'utilisation d'inhibiteurs de l'initiation de la traduction comme l'harringtonine, il est possible de figer les ribosomes au niveau des codons d'initiation auxquels ils sont associés et donc de déterminer leur position sur chaque ARNm. Ainsi, les transcrits ayant recours au *leaky scanning* peuvent être facilement détectés grâce à la présence d'empreintes de la sous-unité 80S multiples au niveau de chaque codon d'initiation utilisé [11]. Cette approche a récemment permis de mettre en évidence l'existence de nombreux sites d'initiation de la traduction non canoniques chez les ARNm cellulaires [26]. En effet, en raison de la dégénérescence du code génétique, l'ARNt initiateur qui s'associe préférentiellement au codon AUG peut également initier la traduction à partir de codons de séquence proche comme le CUG ou le UUG. Chez certains virus, l'initiation peut également se faire sur des codons non-AUG dans le cadre du *leaky scanning*. C'est le cas notamment pour le virus leucémogène murin (MuLV) qui utilise un codon CUG situé en amont du codon canonique pour produire une isoforme de sa polyprotéine Gag qui n'est pas incorporée dans les particules virales, mais qui joue un rôle important dans la dissémination des virions [27, 28]. Le profilage ribosomique appliqué à l'étude de la traduction au cours d'une infection virale a également permis récemment d'identifier de nombreux sites d'initiation non-canoniques chez deux virus à ADN et un virus à ARN [17, 24, 29]. La plupart des sites d'initiation non-AUG découverts à ce jour chez les virus sont utilisés pour la synthèse de petits peptides dont le rôle biologique n'est pas encore connu. Des études supplémentaires seront donc nécessaires afin de vérifier si ces peptides participent activement au cycle de réplication, s'ils jouent plutôt un rôle régulateur sur la traduction d'autres gènes viraux en *cis* ou encore s'ils correspondent à un bruit de fond.

### Frameshifting ou glissement de phase de lecture

Le *frameshifting* (glissement de phase de lecture) est un mécanisme conservé chez les eucaryotes comme chez les virus qui permet l'expression de plusieurs protéines à partir d'un même ARNm (Figure 3D). De manière générale au cours de l'élongation, le ribosome se déplace le long de l'ARNm par translocation de codon en codon, soit de 3 nucléotides à chaque fois. Dans le cas d'un glissement de phase de lecture, le ribosome peut se décaler d'un ou de deux nucléotides supplémentaires ce qui induit un décalage de la phase de lecture (Figure 3D). Ce mécanisme permet notamment l'expression de protéines différentes en fonction du contexte cellulaire. Chez les virus, le *frameshifting* est très répandu, car il augmente sensiblement la quantité d'information génétique que peut contenir un acide nucléique [1]. Chez certains rétrovirus, en particulier les VIH, ce processus est essentiel pour la production des protéines virales. Il représente donc une cible thérapeutique de choix [30]. L'efficacité du glissement de phase de lecture étant relativement faible, il permet en outre au virus de restreindre les niveaux

d'expression de certaines de ses protéines qui pourraient être toxiques pour la cellule à forte dose [31]. Un exemple particulièrement étudié est l'utilisation du *frameshifting* par le VIH-1 afin d'exprimer de manière alternée les précurseurs protéiques Gag et la protéine de fusion Gag/Pol, toutes deux essentielles pour la production de particules virales infectieuses [32]. La mise en évidence d'un décalage de phase de lecture peut être réalisée par des approches de prédiction *in silico* ou en utilisant des méthodes plus classiques de génétique. Le profilage ribosomique est particulièrement intéressant pour l'étude de ce mécanisme car il permet à la fois de détecter un décalage dans la phase de lecture pour un transcrit donné, et de quantifier le taux de traduction au niveau de chaque cadre de lecture en conditions physiologiques. En effet, en suivant la position des empreintes ribosomales laissées sur son extrémité 5' après la digestion par les nucléases, il est possible de déterminer la phase de lecture d'un ARNm en cours de traduction [16]. En cas de *frameshifting*, la position des empreintes obtenues ne suivra pas la périodicité de trois nucléotides au niveau de la région où les deux cadres de lecture se chevauchent. Elle sera décalée de un ou deux nucléotides si le deuxième cadre de lecture s'étend en aval du cadre de lecture canonique. Récemment, cette méthode a permis d'étudier le mécanisme de *frameshift* utilisé par un coronavirus murin appartenant au même genre que le virus SARS-CoV à l'origine du SRAS (syndrome respiratoire aigu sévère) [24]. Les résultats obtenus ont permis de montrer que les ribosomes ne font pas de pause au niveau de la structure en pseudo-nœud responsable du glissement de phase dans les cellules infectées, contrairement à ce qui avait été démontré précédemment *in vitro* [33]. Ces résultats posent ainsi de nouvelles questions quant au rôle des structures secondaires de l'ARNm dans le mécanisme moléculaire impliqué dans le glissement de phase de lecture. Il sera intéressant d'étendre les résultats obtenus chez le coronavirus à d'autres virus ayant recours au *frameshift* comme le VIH-1 et de voir si l'absence de pause des ribosomes au cours de l'élongation observée *in vivo* est une caractéristique générale remettant en cause les résultats obtenus *in vitro*.

### Réinitiation de la traduction et translecture du codon stop

La modulation de la traduction cellulaire peut également avoir lieu au cours des étapes tardives de la synthèse protéique comme lors de la terminaison. Notamment, dans certains cas où le codon stop est situé à proximité de séquences particulières, qui peuvent contenir ou non des structures secondaires, le ribosome

peut l'ignorer et poursuivre l'élongation de la chaîne polypeptidique (Figure 3E). Ce phénomène est appelé translecture ou *readthrough* du codon stop et permet de synthétiser une isoforme d'une protéine avec une extension C-terminale (Figure 3E). À l'inverse, dans le cas de la réinitiation de la traduction, le ribosome reconnaît correctement le codon stop et l'étape de terminaison se fait de manière conventionnelle. Cependant, la petite sous-unité du ribosome (40S) reste accrochée à l'ARNm après l'arrêt de la synthèse protéique et reprend le balayage de l'ARNm en aval du codon stop jusqu'à atteindre un autre codon d'initiation et traduire un nouveau cadre de lecture. Ces deux mécanismes de contrôle traductionnel peuvent être détournés au cours d'une infection virale pour assurer la production de plusieurs protéines différentes à partir d'un même ARNm [1]. Ainsi, de nombreux génomes viraux contiennent des séquences favorisant la translecture et la réinitiation de la traduction afin d'augmenter la quantité d'information génétique que peut contenir un même ARNm viral.

Le profilage ribosomique a déjà été utilisé pour étudier la translecture du codon stop chez l'homme et la drosophile [34, 35]. Les résultats obtenus ont montré que de nombreux gènes utilisent ce mécanisme de régulation pour produire des isoformes de protéines possédant des caractéristiques différentes comme, par exemple, pour leur localisation intracellulaire. Son application sur des cellules infectées pourrait permettre d'étudier l'importance de ce type de phénomène dans le cycle de réplication virale grâce à la localisation précise de l'ensemble des ribosomes en cours de traduction à un moment donné.

### Découverte de nouvelles protéines viraux par profilage ribosomique

Une des applications les plus importantes du profilage ribosomique dans l'étude du cycle de réplication virale consiste en la découverte de nouvelles régions codantes et la confirmation de l'utilisation de cadres de lecture ouverts (*open reading frame* ou ORF) prédits à partir de l'analyse bio-informatique de la séquence primaire des génomes viraux [14]. Les ORF correspondent à des portions des ARNm qui peuvent potentiellement être traduites en protéines et sont définies comme une région comprise entre deux codons stop séparés par une série de triplets. Le profilage ribosomique est particulièrement utile pour l'étude des virus complexes qui possèdent un grand génome codant pour de nombreux transcrits dont le rôle n'est pas encore caractérisé : il permet de mettre en évidence les séquences viraux qui sont effectivement traduites au cours d'une infection. Les résultats obtenus avec cette méthode doivent cependant être validés par des approches complémentaires afin de suivre, en parallèle, la quantité de protéines viraux produites dans la cellule telles que la spectrométrie de masse ou le *western blot*. Récemment, une telle approche a été appliquée à l'analyse de la synthèse des protéines lors de l'infection par le cytomégalovirus humain (HCMV). D'après des prédictions obtenues après analyses bio-informatiques, ce virus possède un génome complexe d'environ 240 kilobases contenant plus de 200 ORF d'au moins 50 acides aminés de long [36]. Pour décrypter le processus de production des protéines de ce virus, l'ensemble des transcrits

produits au cours de l'infection a été séquencé et leur traduction étudiée par profilage ribosomique [17]. De plus, la position de l'extrémité 5' de tous les transcrits viraux a été caractérisée et l'expression des protéines viraux et cellulaires analysée par spectrométrie de masse. Les résultats obtenus ont conduit à la caractérisation de plus de 700 régions codantes dont plus de la moitié codent des petits peptides de moins de 80 acides aminés. Les auteurs ont par ailleurs trouvé que de nombreuses régions codantes se chevauchent soit dans la même phase de lecture, produisant des isoformes tronquées de la même protéine, soit dans des phases de lectures différentes et donc produisant des protéines différentes. Ils ont également découvert l'existence d'ARN viraux antisens (codés à partir des brins opposés du génome) dont les régions codantes se chevauchent au niveau génomique. Ce résultat est particulièrement intéressant en regard des contraintes évolutives nécessaires pour coder deux protéines fonctionnelles à partir des deux brins d'une même séquence nucléotidique. Enfin, les auteurs ont montré que de nombreux ARN auparavant prédits comme non codants par des approches bio-informatiques car ne possédant pas d'ORF suffisamment longs, sont en réalité traduits et correspondent à des ARN polycistroniques codant plusieurs peptides de moins de 90 acides aminés chacun. Même si la plupart des ORF identifiés correspondent effectivement à des protéines viraux, leur rôle au cours du cycle de réplication virale reste cependant encore à être caractérisé. Il est possible que ces régions codantes correspondent à du bruit de fond traductionnel ou qu'elles participent à la régulation de la traduction ou de la stabilité de transcrits en *cis* sans que la protéine produite ne joue un rôle biologique (comme c'est déjà le cas de certains uORF [*upstream open reading frame*]). Ces résultats démontrent à quel point les génomes viraux peuvent être complexes et font appel à des mécanismes d'expression particulièrement originaux. Le développement d'approches innovantes telles que le profilage ribosomique ouvre donc de nouvelles perspectives pour l'étude de la traduction chez de tels virus.

### Limitations et biais liés au profilage ribosomique

Le profilage ribosomique représente donc une avancée technique majeure pour l'étude de la traduction des ARNm. Cependant, cette approche présente un certain nombre de biais et de limitations qu'il ne faut pas négliger lors de l'analyse des résultats.

Une des limitations majeures réside dans le fait que le profilage ribosomique ne permet pas d'obtenir une

mesure directe de l'efficacité de traduction d'un ARNm mais uniquement de mesurer la densité de ribosomes sur un transcrit donné. En effet, dans le cas où les ribosomes seraient bloqués ou ralentis en phase d'élongation, leur densité sur la région codante de l'ARNm traduit pourrait croître sans pour autant conduire à une augmentation du nombre de protéines produites par transcrit. De plus, la présence de ribosomes sur un ARNm n'est pas nécessairement synonyme de synthèse protéique active. Il est donc préférable de valider les résultats obtenus, au moins sur une partie des transcrits d'intérêt, en utilisant des méthodes alternatives permettant de suivre les modifications de l'expression protéique telles que l'utilisation de gènes rapporteurs, le *western blot* ou encore la spectrométrie de masse. Par ailleurs, de nouveaux outils analytiques, comme le « *ribosome release score* » ou RSS, ont été développés pour discriminer, parmi les ARNm associés à des ribosomes, ceux qui codent des protéines de ceux qui sont non-codants [37]. Le RSS se base sur l'hypothèse que sur les transcrits réellement codants, le nombre de ribosomes associés en aval du codon stop décroît drastiquement, tandis que pour les ARN non-codants, l'association avec les ribosomes reste homogène le long de toute la séquence.

Le profilage ribosomique est également sensible à la présence de contaminants qui peuvent créer des faux-positifs. En effet, malgré la purification des ribosomes par ultracentrifugation sur gradient ou coussin de sucrose avant de récupérer les fragments d'ARNm qui leurs sont associés, des complexes ribonucléiques de haute masse moléculaire peuvent cosédimer et introduire des fragments d'ARN contaminants dans les fractions récoltées. Afin d'éviter ce type de contamination, il est possible de purifier les ribosomes par chromatographie d'affinité en utilisant des anticorps dirigés contre une protéine de la sous-unité ribosomale 60S [14]. Il est aussi possible d'introduire des étiquettes moléculaires, telles que les séquences tag (étiquettes) HA (*hemagglutinin*), FLAG (peptide de séquence AspTyrLysAspAspAspAspLys) ou GFP (*green fluorescent protein*), dans une protéine de la sous-unité 60S [38]. Cette approche est très utile *in vivo* puisqu'elle permet d'étudier la traduction spécifiquement dans un tissu donné en restreignant l'expression de la protéine ribosomale portant l'étiquette à ce tissu. De plus, des approches de bio-informatique comme le filtrage des séquences en fonction de la taille attendue pour les fragments protégés par le ribosome, connu sous le nom de « *fragment length organization similarity score* » ou FLOSS, améliorent significativement la qualité des données et minimisent la présence de contaminants [14]. Enfin, il est possible de préparer des échantillons contrôles en utilisant des inhibiteurs des étapes précoces de l'initiation de la traduction comme la patéamine A [17] ou des inhibiteurs non spécifiques de la traduction comme l'EDTA (acide éthylène diamine tétra-acétique). Ces drogues permettent de bloquer complètement l'association des ribosomes aux ARNm afin de vérifier si le signal observé dans les échantillons d'intérêt provient réellement d'empreintes ribosomales ou correspond à des artefacts.

Une limitation supplémentaire est introduite par les inhibiteurs de la traduction généralement utilisés pour figer les ribosomes sur les ARNm qui peuvent modifier leur distribution sur les régions codantes

en introduisant un biais non négligeable [39]. Ainsi, l'utilisation de la cycloheximide aurait tendance à enrichir les empreintes de ribosomes situées à proximité du site d'initiation de la traduction ce qui ne serait pas forcément représentatif de la situation réelle dans la cellule d'intérêt avant l'ajout de la drogue. Pour éviter ce problème, des variantes du protocole n'utilisant pas d'inhibiteurs de la traduction ou au contraire utilisant un large excès d'inhibiteurs ont été développées afin de minimiser le biais de position [40].

Une étude a récemment montré que le choix de la nucléase utilisée pour obtenir les empreintes ribosomales pouvait introduire des biais significatifs dans la distribution des séquences protégées [35]. Ainsi, la RNase I a tendance à enrichir les fragments protégés au niveau du site d'initiation de la traduction, alors que la nucléase micrococcalle révèle la présence de sites protégés dans la région 3'UTR des transcrits. Ces résultats peuvent être expliqués par le fait que la RNase I induit une dégradation non spécifique des ribosomes, particulièrement dans les échantillons provenant de mammifères [41]. La plupart des nucléases utilisées dans les protocoles de profilage ribosomique, telles que la RNase I, la nucléase micrococcalle et la RNase T1, agissent préférentiellement sur des substrats ARN simple brin. Certaines de ces nucléases (comme la RNase A, la RNase T1 et la nucléase micrococcalle) ont une préférence pour digérer l'ARN en 3' ou en 5' de certains nucléotides : par exemple, la RNase A ne clive l'ARN qu'en 3' des nucléotides C et U. Il est donc très probable que le choix de la nucléase impose des biais significatifs dans le profil des séquences protégées par les ribosomes. Ces biais, associés à d'autres comme la présence de structures secondaires sur l'ARNm ou les biais dus à la préparation des banques d'ADNc (ADN complémentaire) conduisent à une distribution hétérogène des séquences obtenues par profilage ribosomique le long des régions codantes (Figure 2C). Cette hétérogénéité de distribution rend difficile certaines analyses comme la détection de sites de pause des ribosomes lors de l'élongation [24].

Un paramètre supplémentaire à prendre en compte lors de l'analyse des résultats de profilage ribosomique est la profondeur de séquençage. En effet, l'analyse fine de la phase de lecture des ribosomes sur une région codante nécessite l'alignement d'un grand nombre de séquences sur le transcrit d'intérêt. Cependant, les échantillons obtenus après la purification des ribosomes 80S sont extrêmement riches en ARN ribosomiques qui, même après déplétion, peuvent représenter une fraction importante des séquences obtenues par séquençage. Cela introduit un biais non négligeable,

notamment dans le cas où les ARNm viraux étudiés sont faiblement exprimés ou traduits. En effet, contrairement aux autres techniques d'analyse globale comme les *microarrays* où la mesure de chaque transcrite est indépendante de celle des autres, lors du séquençage à haut débit tous les transcrits sont en compétition pour être séquencés. Ainsi, un transcrite abondant va être surreprésenté en termes de fragments séquencés comparé à un transcrite peu abondant avec une couverture de séquençage beaucoup plus faible. Il faut donc adapter le nombre de fragments à séquencer au niveau d'expression des transcrits d'intérêt et aux questions biologiques posées. Notamment, l'étude de la phase de lecture nécessite une profondeur de séquençage plus importante qu'une analyse d'expression différentielle. Les protocoles de séquençage actuels étant très performants, il est possible d'obtenir des résultats de qualité en préparant au minimum trois réplicats biologiques pour chaque condition testée.

En dépit de toutes ces limitations, le profilage ribosomique est un nouvel outil qui ouvre de nombreuses perspectives pour l'étude de la traduction en conditions physiologiques comme pathologiques.

## Conclusion

Le profilage ribosomique est une technique innovante dont l'intérêt majeur repose sur la possibilité de cartographier la position des ribosomes sur l'ensemble des ARNm présents dans une cellule et de quantifier précisément leur densité sur chaque transcrite [4]. Cette approche est particulièrement utile pour suivre les mécanismes permettant la synthèse des différentes protéines virales mis en place lors d'une infection [3, 20]. Il est aussi possible d'étudier les mécanismes de traduction des ARNm viraux, déjà décrits par le passé mais dont les détails moléculaires restaient mal définis, en suivant la localisation des ribosomes. Le profilage ribosomique ouvre également de nouvelles perspectives pour l'étude du cycle de réplication de virus particulièrement complexes, tels que les poxvirus [29] ou encore le cytomégalovirus humain (HCMV) [17], grâce à la découverte de nouvelles protéines virales et potentiellement de mécanismes de traduction des ARNm viraux encore jamais décrits. Cette approche est d'autre part très prometteuse pour l'étude des interactions entre le virus et la cellule-hôte en permettant de suivre l'impact d'une infection virale sur la synthèse protéique cellulaire [42]. Cela est particulièrement intéressant dans le cadre des phénomènes d'échappement immunitaire afin de comprendre comment le virus parvient à empêcher la mise en place d'une réponse immunitaire antivirale efficace. ♦

## SUMMARY

### An intimate look at the viral replication cycle through ribosome profiling

Next Generation Sequencing (NGS) techniques have revolutionized most biomedical research fields over the past decade by allowing a broader vision on biological processes that occur at the molecular level. Among these, ribosome profiling or footprinting is a powerful tool to study mRNA translation in a transcriptome-wide manner. Ribosome profiling has been used to study the impact of translational control of

gene expression under many different cellular conditions including viral infections. Indeed, translation is a critical step during the viral replication cycle in which the infected cell is embezzled to produce viral proteins. Ribosome profiling tools can provide new insights on viral translation by monitoring ribosome binding to viral and cellular RNAs with a high definition during the time course of an infection. Here, we describe the potential uses of ribosome profiling for the understanding of viral translational control and the impact of viral infection on host gene expression. We also discuss the main limitations and biases related to the technique that need to be taken into account for its use. ♦

## REMERCIEMENTS

Nous tenons à remercier tous les membres du laboratoire de traduction eucaryote et virale pour la relecture du manuscrit. Nous remercions la fondation FINOVI, l'ANRS (France REcherche Nord & sud Sida-hiv Hépatites), la Ligue nationale contre le cancer et l'Inserm pour leur soutien financier.

## LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

## RÉFÉRENCES

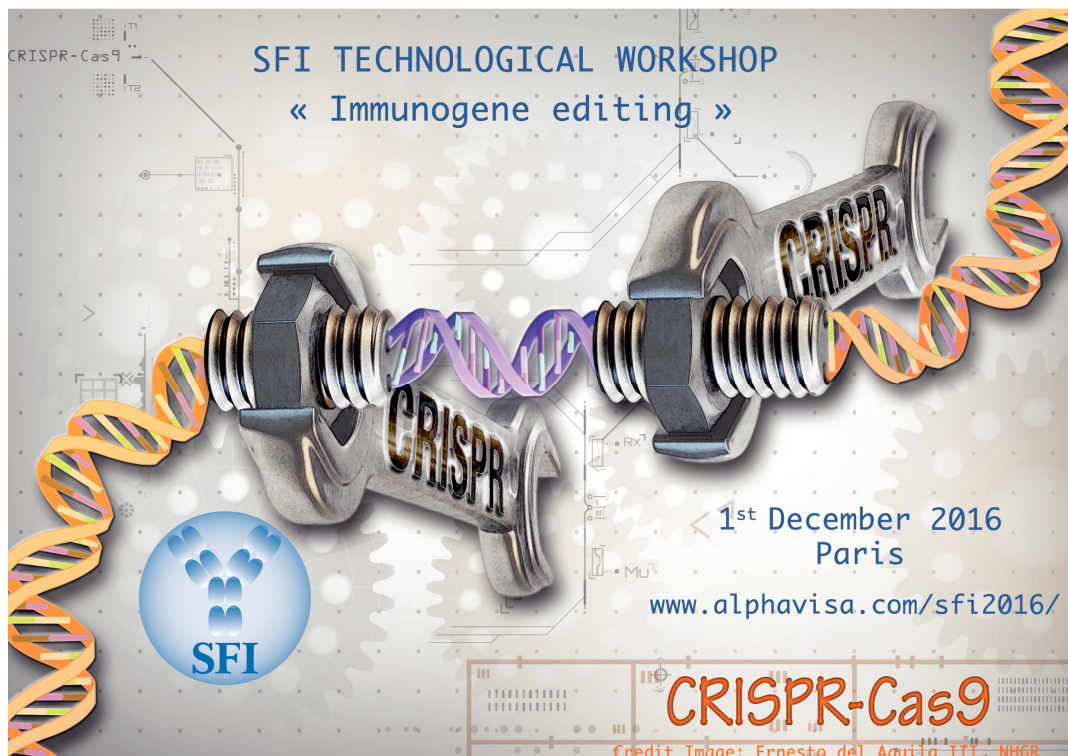
1. Firth AE, Brierley I. Non-canonical translation in RNA viruses. *J Gen Virol* 2012 ; 93 : 1385-409.
2. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 2015 ; 16 : 651-64.
3. Stern-Ginossar N. Decoding viral infection by ribosome profiling. *J Virol* 2015 ; 89 : 6164-6.
4. Ingolia NT, Ghaemmaghami S, Newman JRS, et al. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 2009 ; 324 : 218-23.
5. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in Eukaryotes: Mechanisms and biological targets. *Cell* 2009 ; 136 : 731-45.
6. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 2002 ; 299 : 1-34.
7. Jackson RJ, Hellen CUT, Pestova TV. Termination and post-termination events in eukaryotic translation. *Adv Protein Chem Struct Biol* 2012 ; 86 : 45-93.
8. Gebauer F, Hentze MW. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 2004 ; 5 : 827-35.
9. Morris DR. Ribosomal footprints on a transcriptome landscape. *Genome Biol* 2009 ; 10 : 215.
10. Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J* 1988 ; 7 : 3559-69.
11. Ingolia NT, Brar GA, Rouskin S, et al. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 2012 ; 7 : 1534-50.
12. Michel AM, Baranov PV. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA* 2013 ; 4 : 473-90.
13. Ingolia NT, Brar GA, Rouskin S, et al. Genome-wide annotation and quantitation of translation by ribosome profiling. *Curr Protoc Mol Biol* 2013 ; chapter 4 : Unit-4.18.
14. Ingolia NT, Brar GA, Stern-Ginossar N, et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep* 2014 ; 8 : 1365-79.
15. Shalgi R, Hurt JA, Krykbaeva I, et al. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell* 2013 ; 49 : 439-52.



## RÉFÉRENCES

16. Gerashchenko MV, Lobanov AV, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci USA* 2012 ; 109 : 17394-9.
17. Stern-Ginossar N, Weisburd B, Michalski A, et al. Decoding human cytomegalovirus. *Science* 2012 ; 338 : 1088-93.
18. Walsh D, Mohr I. Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol* 2011 ; 9 : 860-75.
19. Roberts LO, Jopling CL, Jackson RJ, et al. Viral strategies to subvert the mammalian translation machinery. *Prog Mol Biol Transl Sci* 2009 ; 90 : 313-67.
20. Walsh D, Mathews MB, Mohr I. Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb Perspect Biol* 2013 ; 5 : a012351.
21. Ricci EP, Rifo RS, Herbreteau CH, et al. Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem Soc Trans* 2008 ; 36 : 690-3.
22. Lloyd RE. Translational control by viral proteinases. *Virus Res* 2006 ; 119 : 76-88.
23. Rutkowski AJ, Erhard F, L'Hernault A, et al. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun* 2015 ; 6 : 7126.
24. Irigoyen N, Firth AE, Jones JD, et al. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog* 2016 ; 12 : e1005473.
25. Stacey SN, Jordan D, Williamson AJK, et al. Leaky scanning is the predominant mechanism for translation of Human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J Virol* 2000 ; 74 : 7284-97.
26. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity of mammalian proteomes. *Cell* 2011 ; 147 : 789-802.
27. Corbin A, Prats AC, Darlix JL, et al. A nonstructural gag-encoded glycoprotein precursor is necessary for efficient spreading and pathogenesis of murine leukemia viruses. *J Virol* 1994 ; 68 : 3857-67.
28. Prats AC, De Billy G, Wang P, et al. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J Mol Biol* 1989 ; 205 : 363-72.
29. Yang Z, Cao S, Martens CA, et al. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. *J Virol* 2015 ; 89 : 6874-86.
30. Bolinger C, Boris-Lawrie K. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology* 2009 ; 6 : 8.
31. Finch LK, Ling R, Naphine S, et al. Characterization of ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *J Virol* 2015 ; 89 : 8580-9.
32. Guerrero S, Batisse J, Libre C, et al. HIV-1 replication and the cellular eukaryotic translation apparatus. *Viruses* 2015 ; 7 : 199-218.
33. Somogyi P, Jenner AJ, Brierley I, et al. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* 1993 ; 13 : 6931-40.
34. Dunn JG, Foo CK, Belletier NG, et al. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* 2013 ; 2 : e01179.
35. Miettinen TP, Björklund M. Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res* 2015 ; 43 : 1019-34.
36. Murphy E, Rigoutsos I, Shibuya T, et al. Reevaluation of human cytomegalovirus coding potential. *Proc Natl Acad Sci USA* 2003 ; 100 : 13585-90.
37. Guttman M, Russell P, Ingolia NT, et al. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013 ; 154 : 240-51.
38. Sanz E, Yang L, Su T, et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc Natl Acad Sci USA* 2009 ; 106 : 13939-44.
39. Hussmann JA, Patchett S, Johnson A, et al. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet* 2015 ; 11 : e1005732.
40. Gerashchenko MV, Gladyshev VN. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res* 2014 ; 42 : e134.
41. Cenik C, Cenik ES, Byeon GW, et al. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res* 2015 ; 25 : 1610-21.
42. Piccirillo CA, Bjur E, Topisirovic I, et al. Translational control of immune responses: from transcripts to translomes. *Nat Immunol* 2014 ; 15 : 503-11.
43. Marcel V, Catez F, Mertani HC, Diaz JJ. Le ribosome. *Med Sci (Paris)* 2014 ; 30 : 21-4.

**TIRÉS À PART**  
É.P. Ricci



## Annex 3

# PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice.

Pérès E, **Blin J**, RICCI EP, Artesi M, Hahaut V, Van den Broeke A, Corbin A, Gazzolo L, Ratner L, Jalinet P, Duc Dodon M. PLoS Pathog. 2018 Mar 22;14(3):e1006933.

Doi: 10.1371/journal.ppat.1006933. Ecollection 2018 Mar.

Human T-cell leukemia virus type 1 (HTLV-1) is the etiological agent of Adult T-cell Leukemia/Lymphoma (ATLL), an aggressive malignant proliferation of activated CD4<sup>+</sup> T lymphocytes. The viral Tax oncoprotein is critically involved in both HTLV-1-replication and T-cell proliferation, a prerequisite to the development of ATLL. Tax contains a PDZ domain-binding motif (PBM) that can interact with several cellular PDZ proteins. In this study, the contribution of the Tax PDZ domain-Binding Motif (PBM) to the lymphoproliferative process was investigated *in vivo*. For this, T-cell proliferative capacities were assessed in humanized mice (hu-mice) carrying a human hemato-lymphoid system infected with either a wild type (WT) or a Tax PBM-deleted ( $\Delta$ PBM) provirus. The frequency of CD4<sup>+</sup> activated T-cells in the peripheral blood and in the spleen was significantly higher in WT than in  $\Delta$ PBM hu-mice. Likewise, human T-cells collected from WT hu-mice and cultivated *in vitro* in presence of interleukin-2 were proliferating at a higher level than those from  $\Delta$ PBM animals. The association of Tax with the Scribble PDZ protein, a prominent regulator of T-cell polarity, was also analysed in human T-cells either directly after *ex vivo* isolation or later after *in vitro* culture. The binding of the Tax PBM to the PDZ Scribble protein correlated with perturbations of cytoskeletal organization and cell polarity. Finally, a comparative genome-wide transcriptomic analysis was performed to assess the effect of the interactions between Tax PBM and cellular PDZ proteins at a global scale. For this, I performed RNA-seq and ribosome profiling on T cells isolated from WT and  $\Delta$ PBM hu-mice after several passages *in vitro*. The results suggested that the Tax PBM-PDZ proteins association can modulate the expression of genes regulating proliferation, apoptosis and cytoskeletal organization. Tax PBM is thus an auxiliary motif that contributes to the sustained growth of HTLV-1 infected T-cells *in vivo* and *in vitro* and is essential to T-cell immortalization.

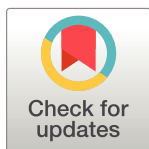
RESEARCH ARTICLE

# PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice

Eléonore Pérès<sup>1</sup>, Juliana Blin<sup>2</sup>, Emiliano P. Ricci<sup>2</sup>, Maria Artesi<sup>3,4</sup>, Vincent Hahaut<sup>3,4</sup>, Anne Van den Broeke<sup>3,4</sup>, Antoine Corbin<sup>2</sup>, Louis Gazzolo<sup>1</sup>, Lee Ratner<sup>5</sup>, Pierre Jalinot<sup>1</sup>, Madeleine Duc Dodon<sup>1\*</sup>

**1** Laboratory of Biology and Modeling of the Cell, Ecole Normale Supérieure (ENS) de Lyon, INSERM U1210, CNRS UMR5239, 46 allée d'Italie, Lyon, France, **2** International Center for Infectiology Research, ENS de Lyon, Université Claude Bernard Lyon 1, INSERM U1111, CNRS UMR 5308, 46 allée d'Italie, Lyon, France, **3** Laboratory of Experimental Hematology, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium, **4** Unit of Animal Genomics, Groupe Interdisciplinaire Génomprotéomique Appliquée (GIGA), Université de Liège, Liège, Belgium, **5** Division of Oncology, Washington University School of Medicine, St Louis, MO, United States of America

\* [mducdodo@ens-lyon.fr](mailto:mducdodo@ens-lyon.fr)



**OPEN ACCESS**

**Citation:** Pérès E, Blin J, Ricci EP, Artesi M, Hahaut V, Van den Broeke A, et al. (2018) PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice. *PLoS Pathog* 14 (3): e1006933. <https://doi.org/10.1371/journal.ppat.1006933>

**Editor:** Susan R. Ross, University of Illinois at Chicago College of Medicine, UNITED STATES

**Received:** December 1, 2017

**Accepted:** February 12, 2018

**Published:** March 22, 2018

**Copyright:** © 2018 Pérès et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** RNA-seq data can be viewed in Fig 7 and S6 Fig. Additionally, the RNA-seq data has been deposited in both the GEO submission, NCBI tracking system (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102220>) and in European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/PRJEB22059>).

**Funding:** This work was supported by grants from Ligue Nationale Contre le Cancer (url: <https://www.ligue-cancer.net>), comité Auvergne-Rhône-Alpes, from Fondation FINOVI (url: <https://www.finovi>).

## Abstract

Human T-cell leukemia virus type 1 (HTLV-1) is the etiological agent of adult T-cell leukemia/lymphoma (ATLL), an aggressive malignant proliferation of activated CD4+ T lymphocytes. The viral Tax oncoprotein is critically involved in both HTLV-1-replication and T-cell proliferation, a prerequisite to the development of ATLL. In this study, we investigated the *in vivo* contribution of the Tax PDZ domain-binding motif (PBM) to the lymphoproliferative process. To that aim, we examined T-cell proliferation in humanized mice (hu-mice) carrying a human hemato-lymphoid system infected with either a wild type (WT) or a Tax PBM-deleted ( $\Delta$ PBM) provirus. We observed that the frequency of CD4+ activated T-cells in the peripheral blood and in the spleen was significantly higher in WT than in  $\Delta$ PBM hu-mice. Likewise, human T-cells collected from WT hu-mice and cultivated *in vitro* in presence of interleukin-2 were proliferating at a higher level than those from  $\Delta$ PBM animals. We next examined the association of Tax with the Scribble PDZ protein, a prominent regulator of T-cell polarity, in human T-cells analyzed either after *ex vivo* isolation or after *in vitro* culture. We confirmed the interaction of Tax with Scribble only in T-cells from the WT hu-mice. This association correlated with the presence of both proteins in aggregates at the leading edge of the cells and with the formation of long actin filopods. Finally, data from a comparative genome-wide transcriptomic analysis suggested that the PBM-PDZ association is implicated in the expression of genes regulating proliferation, apoptosis and cytoskeletal organization. Collectively, our findings suggest that the Tax PBM is an auxiliary motif that contributes to the sustained growth of HTLV-1 infected T-cells *in vivo* and *in vitro* and is essential to T-cell immortalization.

org), from Public Health Service grants CA100730, and CA063417 (LR), from les Amis de l'Institut Bordet (Brussels, Belgium), from Télévie (FRS-FNRS, Belgium) and the International Brachet Stiftung (IBS). EP and JB were funded by Ministère de l'Enseignement Supérieur, de la Recherche et de l'innovation. EP was a recipient of a PhD grant from the Ligue Nationale Contre le Cancer. MA holds a Postdoctoral Researcher fellowship of the FRS-FNRS. VH is funded by a grant from FRS-FNRS-Télévie. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

The viral Tax oncoprotein is a critical contributor to the development of adult T-cell leukemia/lymphoma, an aggressive malignant proliferation of T lymphocytes. Tax contains a PDZ domain-binding motif (PBM) that favors the interaction with several cellular PDZ proteins. Here, we compare the *in vivo* involvement of the Tax PBM in humanized mice infected with either a full-length provirus or a Tax PBM-deleted provirus. We observe that the establishment of the sustained lymphoproliferation in the peripheral blood of infected mice is dependent on the Tax PBM. Furthermore, binding of the Tax PBM to the PDZ Scribble protein correlated with perturbations of cytoskeletal organization and cell polarity. In addition, genome-wide transcriptomic analyses strongly suggest that the association of Tax PBM with cellular PDZ proteins results in the expression of several genes involved in proliferation, apoptosis and cytoskeletal organization. Collectively, these results indicate that the Tax PBM is an auxiliary motif that contributes to the growth of HTLV-1 infected T-cells. As a consequence, targeting the PBM/PDZ nodes using small peptides may have the potential to antagonize the Tax-induced lymphoproliferation, offering a novel strategy for the treatment of this disease.

## Introduction

HTLV-1 (Human T-cell leukemia virus, type 1) is the etiological agent of adult T-cell leukemia/lymphoma (ATLL), an aggressive and fatal form of leukemia characterized by the malignant expansion of activated CD4+ T-cells [1]. Among several non-structural regulatory proteins encoded by HTLV-1, Tax, a crucial transcriptional activator of the viral life cycle, exerts pleiotropic effects during the initial stages of the multistep leukemic process [2]. This viral protein modulates the expression of cellular genes leading to the deregulation of T-cell proliferation, perturbing the integrity of cell cycle checkpoints, the DNA damage response and apoptosis pathways [3–6].

Like other viral oncoproteins such as human adenovirus E4-ORF1 and human papillomavirus (HPV) E6, Tax encodes a carboxyl-terminal (ETEV amino acids 350–353) PDZ domain-Binding Motif (PBM) that mediates interactions with a particular group of cellular proteins containing one or several PDZ (PSD95/DLG/ZO-1) domain(s) [7–9]. Many of these PDZ proteins are involved in processes that control cell attachment, cell proliferation, cell polarity and cell signaling [10, 11]. Previous studies have indicated that the interaction of viral oncoproteins with PDZ proteins may play a critical role in the development of malignancies by perturbing the function of these cellular proteins [12, 13]. The HTLV-1 Tax PBM has been shown to associate with several PDZ cellular proteins such as DLG1 (Discs large 1), Scribble, Erbin, TIP-1 (Tax-interacting protein-1) or MAGI-3 (Membrane-associated guanylate kinase-3) in *in vitro* studies [14–16]. One of them, Scribble that acts as a tumor suppressor and a regulator of cell polarity, is highly expressed in activated T-lymphocytes [17, 18].

Interestingly, the absence of this motif in the Tax of HTLV-2, a non-leukemic strain of HTLV, has led to the assumption that the HTLV-1 Tax PBM fulfills an essential function in the leukemic process [19, 20]. Previous studies have shown that the deletion or mutation of the Tax PBM decreases IL2-independent growth of CTLL-2 cells and the Tax transforming activity in a rat fibroblast cell line [19]. More interestingly, Xie et al have reported that PBM is required for virus-mediated T-cell proliferation and genetic instability *in vitro* and for viral persistence in a rabbit infection model [21]. These observations strongly support the hypothesis that the Tax PBM is critically involved in supporting the infectious process, prompting us to *in vivo*



evaluate the implication of the PBM in T-cell proliferation. To that aim, we used immunodeficient mice, which display a human hemato-lymphoid system, therein referred to as hu-mice. These hu-mice provide a powerful model for investigating the pathogenesis associated with infection by human lymphotropic viruses [22, 23]. Several studies have previously demonstrated that infection of hu-mice with HTLV-1 recapitulates certain features of ATLL [24–26]. More specifically, our group has demonstrated that HTLV-1 is able to perturb early  $\alpha\beta$ T-cell development in humanized BALB/c Rag2<sup>-/-</sup>γc<sup>-/-</sup> (BRG) mice [27]. We showed that HTLV-1 infection propelled thymic human T-cell development towards the mature stages and that this effect was dependent on Tax expression.

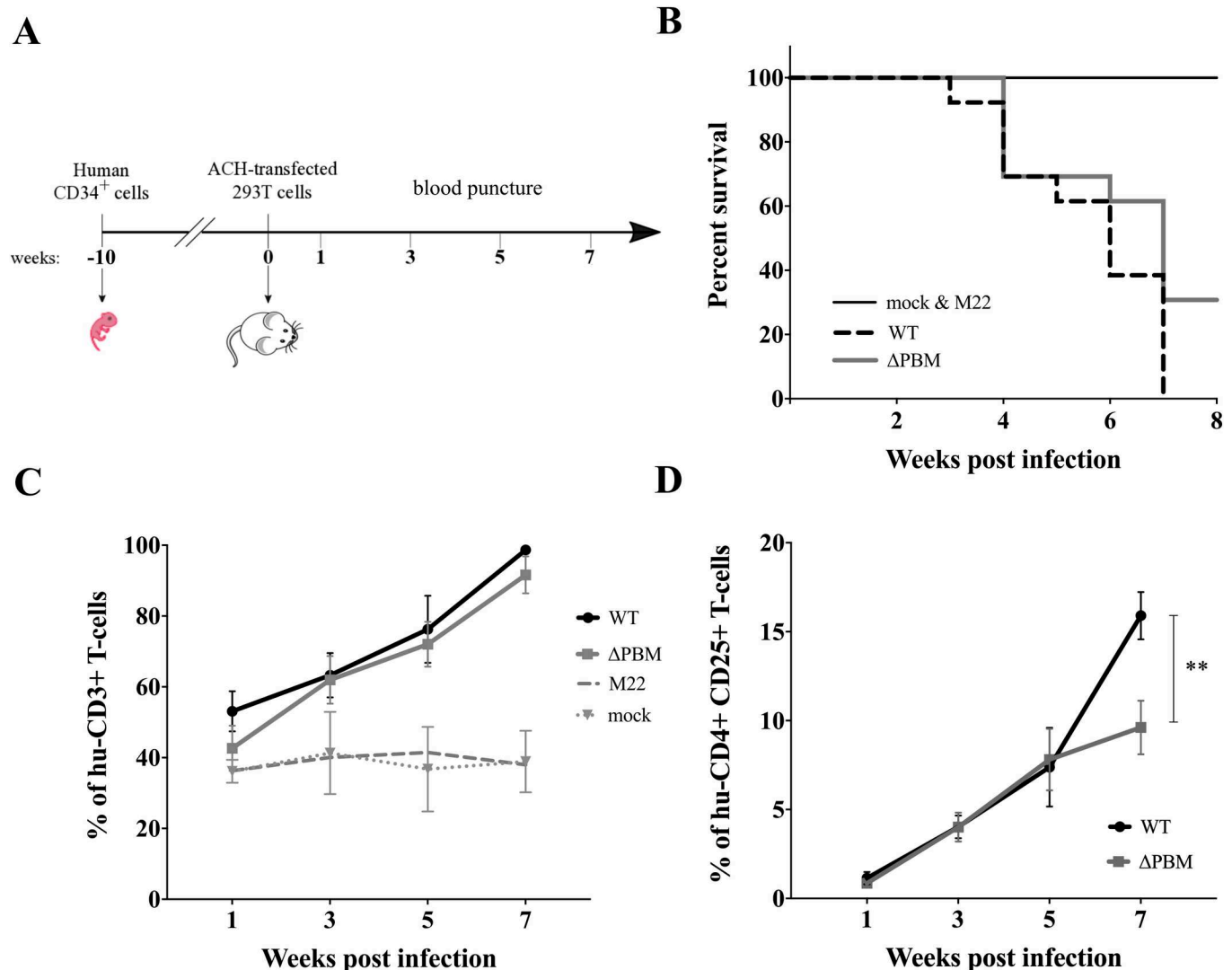
In this study, we addressed the role of the Tax PBM in hu-mice infected with irradiated cells producing either a wild-type virus (HTLV-1 WT) or a virus characterized by a Tax PBM-deleted (HTLV-1 ΔPBM). In the peripheral blood of WT hu-mice, the proliferation of activated CD4+CD25+ T-cells was significantly higher than in the peripheral blood of ΔPBM hu-mice. Likewise, human T-cells collected from WT hu-mice and cultivated *in vitro* in presence of interleukin-2 were proliferating at a higher level than those from ΔPBM animals. We then showed that the PDZ Scribble protein interacts with Tax in *ex vivo* or *in vitro* T-cells from WT hu-mice, but not in cells from ΔPBM hu-mice. These results underline that the PBM-PDZ association is critical for sustaining HTLV-1-induced T-cell proliferation. Finally, a genome-wide transcriptomic analysis of T-cells from infected hu-mice suggests that this association is involved in the regulation of host genes implicated in cell proliferation, inhibition of apoptosis, cell polarity and cytoskeletal changes.

## Results

### Tax PBM enhances HTLV-1-induced T-cell proliferation in infected hu-mice

To evaluate the role of the Tax PBM *in vivo*, a total of 32 hu-mice were used for this study. Thirteen hu-mice were inoculated with X-irradiated 293T cells previously transfected with either ACH-WT or ACH-ΔPBM molecular clones. Three hu-mice were inoculated with either X-irradiated 293T cells transfected with ACH-M22 (that displays a Tax PBM, but is unable to activate the NF-κB pathway) or untransfected (mock infected). These hu-mice were daily monitored for apparent suffering signs, such as weight loss, back arches and prostrated behavior. Furthermore, a small volume of peripheral blood was eye-harvested from each infected hu-mouse every two weeks starting from one week post-infection (Fig 1A) and cytometry analysis was immediately performed to follow the presence of activated human CD25+ T-cells. Accordingly, 8 WT and 5 ΔPBM hu-mice with suffering signs between 3 and 5 weeks after infection were sacrificed. At 7 weeks, the 5 remaining WT hu-mice and 4 ΔPBM hu-mice that exhibited more than 10% of circulating CD25+ T-cells were sacrificed. The 4 surviving ΔPBM hu-mice that did not show any suffering signs were sacrificed two days later together with the 3 ACH M22 and the 3 mock infected animals (Fig 1B).

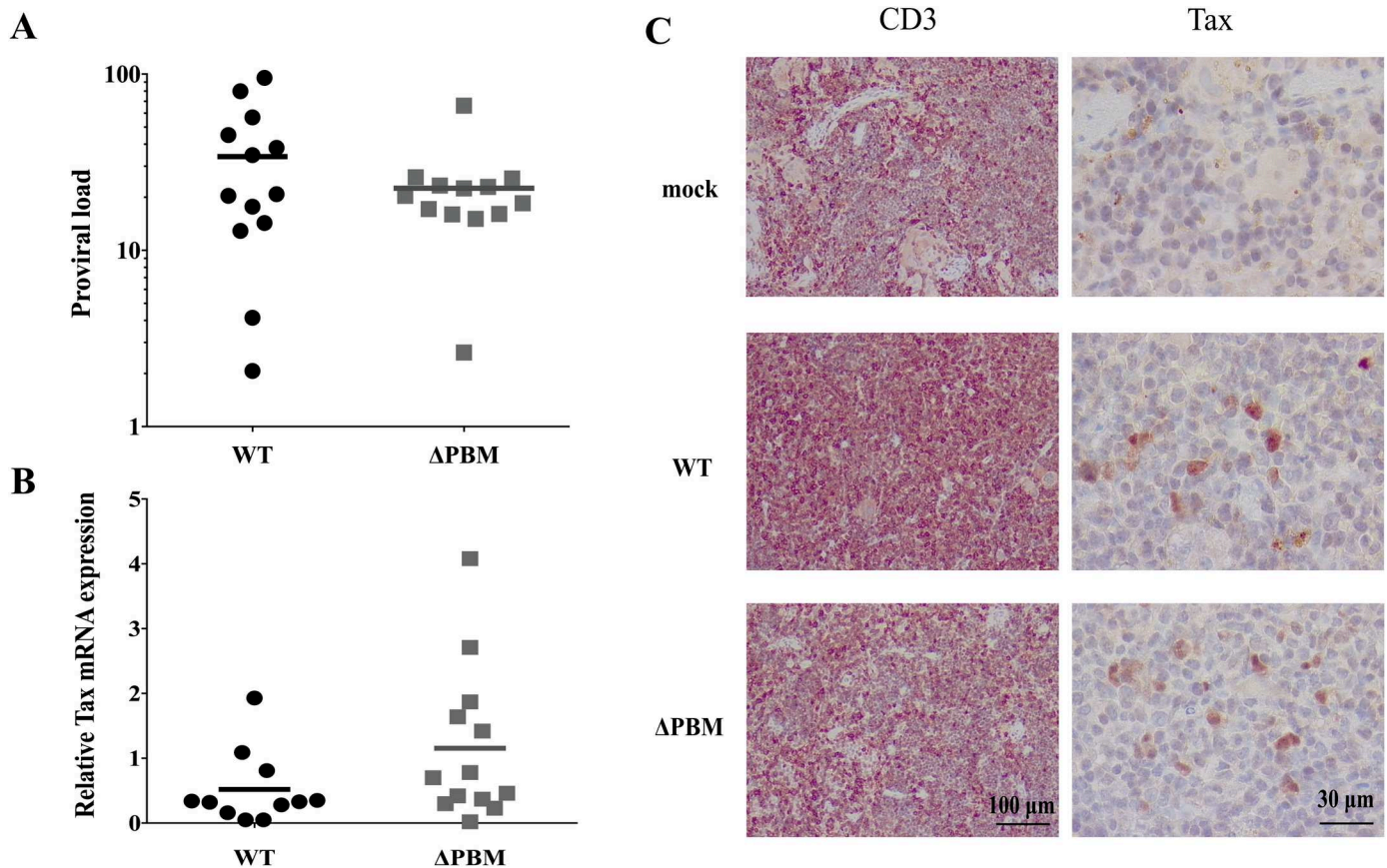
First, contrary to mock and M22 infected hu-mice, the percentage of circulating hu-CD3+ T-cells of WT and ΔPBM hu-mice increased gradually up to 7 weeks (Fig 1C). This correlated with a significant increase of the frequency of hu-CD45 cells in both WT and ΔPBM infected hu-mice compared to the M22 and mock infected mice (S3 Table). Among the hu-CD3+ T-cells, the frequency of activated CD4+CD25+ T-cells increased gradually up to 5 weeks after infection in the peripheral blood of both WT and ΔPBM hu-mice, while at 7 weeks this percentage was significantly higher in WT than in ΔPBM hu-mice (Fig 1D). A low frequency of CD4+CD25+ T-cells was observed in the peripheral blood of ACH-M22 hu-mice as well as in that of mock infected hu-mice (S3 Table). We did not observe a significant proliferation of the



**Fig 1. Tax PBM increases HTLV-1-induced proliferation of human CD4+CD25+ T-cells.** (A) Schematic representation of the procedure for the generation of infected hu-mice: newborn immuno-deficient NSG mice were sub-lethally X-irradiated and intra-hepatically injected with purified huCD34+ stem cells. Ten weeks later, at a time when the human hemato-lymphoid system is established, hu-mice were infected with HTLV-1 by intra-peritoneal inoculation of 293T cells transfected with various ACH plasmids and then X-irradiated. Peripheral blood was collected every two weeks until the sacrifice. (B) Representative Kaplan-Meier analysis of survival of hu-mice infected with ACH-WT (13 animals, dashed line), ACH-ΔPBM (13 animals, grey line), ACH-M22 (3 animals) and 3 mock infected animals (black line). (C) Kinetics analysis of the frequency of human CD3+ T-cells among human cells in peripheral blood of WT-(black line), ΔPBM-(grey line); M22 and mock (dashed lines) infected hu-mice. Data are presented as mean ± SEM. (D) Kinetics analysis of the frequency of human CD4+ CD25+ T-cells among human cells in peripheral blood of 5 WT (black) and 8 ΔPBM (grey) infected hu-mice. To evaluate the frequency, we first gated the hu-CD45+ cells, then the CD3+ cells of hu CD45+ cells; then the CD4+/CD8+ / CD25+ of hu-CD3+ cells. Statistical difference was calculated with Mann-Whitney *U* test with \*\*, *P* = 0.0093.

<https://doi.org/10.1371/journal.ppat.1006933.g001>

CD8+ T-cells in the peripheral blood of either group of mice (S3 Table). It is important to note that the Tax transcriptional activity mediated by the WT and ΔPBM proviruses through both CREB/ATF and NF-κB signalling pathways is independent of the Tax PBM (S1 Fig). In addition, it is evident that the PBM is operational only when the NF-κB pathway is functional. Collectively, these data suggest that the PBM is endowed with a sustaining activity of T-cell proliferation.



**Fig 2. Proviral load and Tax expression in the spleen of WT or  $\Delta$ PBM infected hu-mice.** (A) The proviral load in splenocytes from the 2 groups of 13 hu-mice infected with the respective HTLV-1 variants was determined by quantitative PCR and reported as the number of pX copies per 100 human cells. Bar represents mean. The Mann-Whitney *U* test indicates no statistical difference between the two conditions,  $P = 0.2939$ . (B) Tax mRNA expression in splenocytes isolated from HTLV-1-infected hu-mice with HTLV-1 variants. Levels of Tax mRNA were measured by RT-qPCR; bar represents mean. Mann-Whitney *U* test,  $P = 0.0579$ . (C) Immunohistochemistry of representative sections of spleen of WT and  $\Delta$ PBM infected hu-mice; staining with CD3 and Tax revealed an infiltration of T-lymphocytes with a nuclear and cytoplasmic localization of Tax.

<https://doi.org/10.1371/journal.ppat.1006933.g002>

Examination of sacrificed mice revealed that enlargement of the spleen was the most frequently observed pathological symptom. Splenomegaly was observed in all, but five WT and six  $\Delta$ PBM hu-mice respectively (S4 Table). Spleen was collected as well as bone marrow and when possible mesenteric lymph nodes. There was no significant difference in the spleen weight between groups (mean of  $0.255 \pm 0.199$  g for WT vs  $0.232 \pm 0.126$  g for  $\Delta$ PBM hu-mice compared to  $0.103 \pm 0.037$  g for mock and M22 infected hu-mice) (S4 Table). Sequence analysis of genomic DNA prepared from splenocytes of infected hu-mice confirmed the original sequence of the ACH-WT or ACH- $\Delta$ PBM molecular clones used for infection (S5 Table). PVL of both WT and  $\Delta$ PBM hu-mice splenocytes was between 0.1 to 1 copy/cell with no significant difference between groups (Fig 2A; S4 Table). Using high throughput sequencing (HTS)-based mapping of HTLV-1 integration sites, we did not observe significant differences in the number of unique insertion sites (UIS) corresponding to the number of independent clones between WT and  $\Delta$ PBM hu-mice (S2 Fig). Similar levels of Tax mRNA and protein were detected in CD3+ T-cells from splenocytes of both WT and  $\Delta$ PBM infected hu-mice (Fig 2B and 2C and S3 Fig). There was no correlation between Tax mRNA levels and the weight of the spleens.

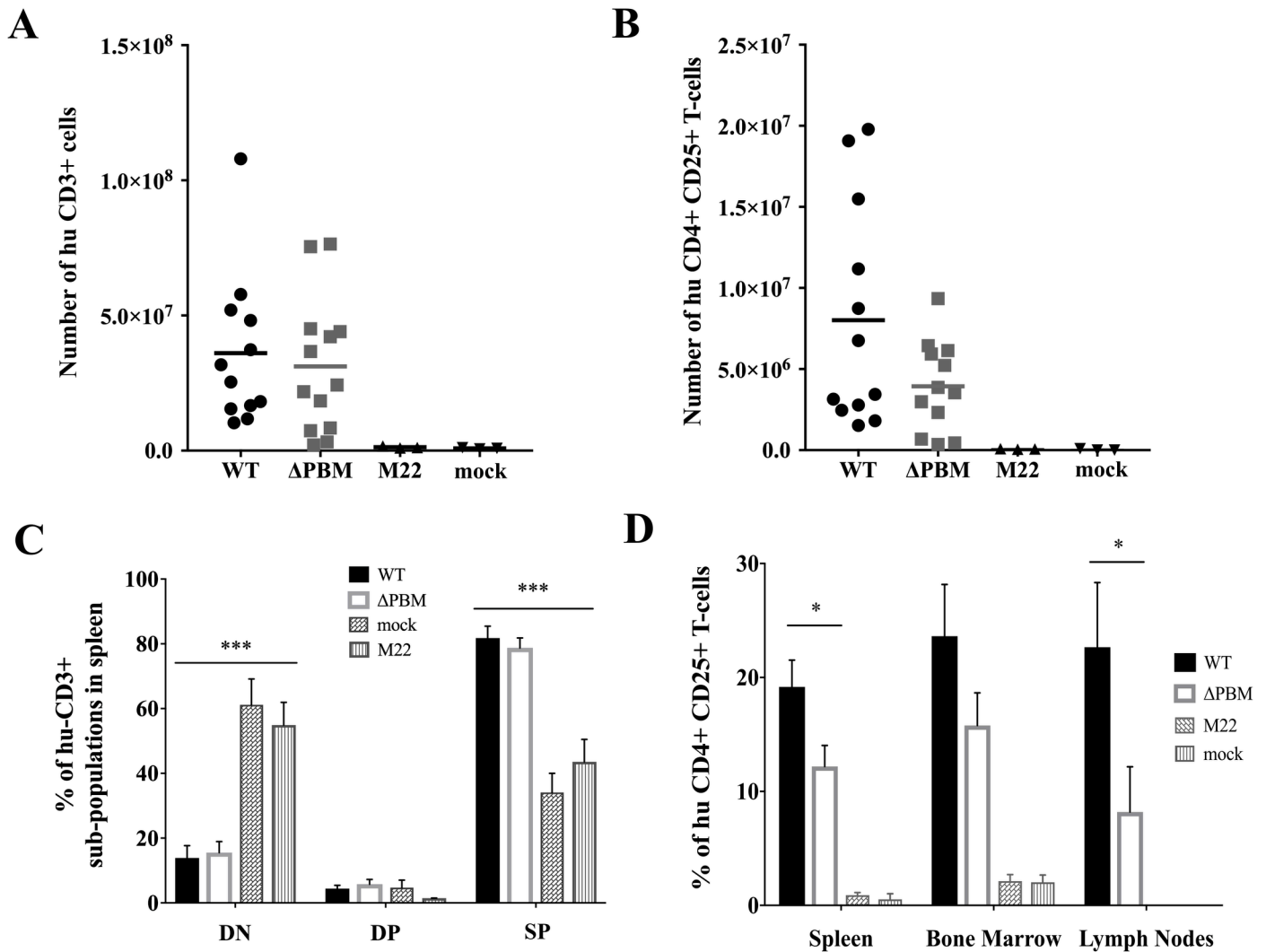
As shown in Fig 3 and S6 Table, the number and the percentage of CD3+ T-cells collected from the spleen of WT and  $\Delta$ PBM infected hu-mice were similar and higher than those

observed in ACH M22 and mock infected hu-mice. Remarkably, a comparative analysis of the T-cell subpopulations (DN: CD4-CD8-; DP: CD4+CD8+, and SP: CD4+ or CD8+) among these CD3+ T cells in both WT and  $\Delta$ PBM infected hu-mice revealed a low percentage of the DN T-cells and a high percentage of the SP T-cell population (among which CD4+ T-cells dominated over CD8+ T-cells), in sharp contrast with the distribution profile of these subpopulations in ACH M22 and mock infected hu-mice (Fig 3C and S4 Fig). Interestingly, even if similar degrees of splenomegaly (S4 Table) that correlated with similar numbers of CD3+ T-cells in the spleen (Fig 3A) were detected, it remains that the number as well as the percentage of CD4+CD25+ T-cells among human splenocytes were higher in WT than in  $\Delta$ PBM hu-mice (Fig 3B–3D). Likewise, lymph nodes and bone marrow collected from WT hu-mice showed a higher frequency of CD4+ CD25+ T-cells than those from  $\Delta$ PBM-hu mice. As indicated above, such a difference also observed in the peripheral blood of animals sacrificed at 7 weeks suggests that the seeding of the periphery by CD4+CD25+ T-cells homing from lymphoid organs is more efficient for WT hu-mice than for  $\Delta$ PBM hu-mice. Altogether, these results indicate that hu-mice are providing an appropriate environment for the proliferation of human T-cells infected with either HTLV-1 WT or  $\Delta$ PBM proviruses. Overall, they underline that the HTLV-1 Tax PBM is acting as an *in vivo* auxiliary motif in the HTLV-1-induced proliferation of infected human T-cells.

### Tax PBM mislocalizes Scribble in T-cells from WT, but not from $\Delta$ PBM hu-mice and sustains proliferation of WT T-cells

As introduced above, several studies have documented that the Tax PBM mediates interactions with a select group of PDZ-containing proteins [8, 19, 21, 28]. In the present study, we focused our attention on one of them, the Scribble protein, that under physiological conditions is differentially localized throughout polarized T-cells and acts as a tumor suppressor [18]. Indeed, Scribble has been shown to undergo mislocalization in cultured HTLV-1 infected T-cells [14, 15]. We first examined the interaction of Tax with Scribble and the localization of the two proteins in *ex vivo* splenocytes collected immediately from WT and  $\Delta$ PBM infected hu-mice after their sacrifice, by using the *in situ* Proximity Ligation Assay (PLA) technology. PLA is a reliable readout of the molecular proximity of two endogenous proteins, thereby facilitating the direct observation of individual protein complexes *in situ* [29]. Thus, the presence of at least 4 dots per cell in about 60% of splenocytes from WT hu-mice clearly revealed a direct contact between Tax and Scribble (Fig 4A, panel 1 and 4). In contrast, we did not observe similar interactions in  $\Delta$ PBM hu-mice (Fig 4A, panel 2). With regards to the intracellular localization of Scribble and Tax in these *ex vivo* splenocytes, immunofluorescent-staining (IF) assays clearly indicated that Scribble was preferentially detected in large polarized aggregates in the cytoplasm of cells from WT mice. In contrast, it was diffusely localized in the cytoplasm and at the plasma membrane of cells from  $\Delta$ PBM hu-mice (Fig 4B). Concomitantly, Tax was found to be mostly localized in the cytoplasm and also visible in condensed aggregates at the plasma membrane of cells from WT hu-mice (Fig 4C). In contrast, Tax was detected in the nucleus and in the cytoplasm of cells from  $\Delta$ PBM hu-mice. These data strongly suggest that PBM is associated with the sequestration of Scribble into polarized aggregates of T-cells from WT hu-mice. As mislocalization of Scribble might interfere with its tumor suppressor function, one can postulate that the Tax PBM is implicated in the enhanced T-cell proliferation observed *in vivo* in WT hu-mice.

Such a possibility was further investigated in assaying the proliferation of T-cells collected from WT or  $\Delta$ PBM hu-mice and *in vitro* seeded in growth medium supplemented with IL2. We periodically verified that these T-cells contained integrated copies of the provirus used at

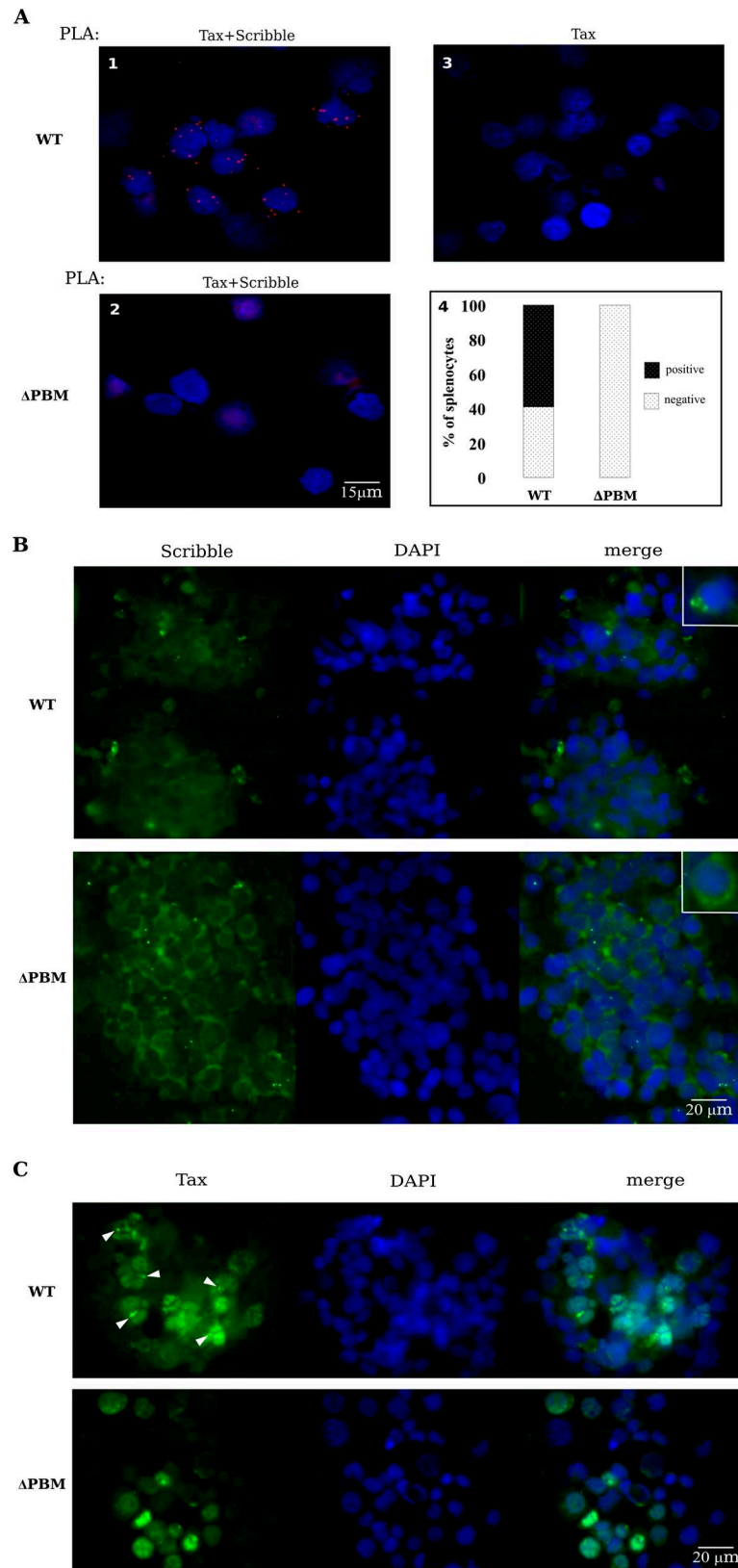


**Fig 3. Tax PBM increases the frequency of human CD4+CD25+ T-cells in lymphoid organs of infected hu-mice.** (A) Number of human CD3+ T-cells among human cells in the spleen of infected mice (WT, n = 13; ΔPBM, n = 13; M22, n = 3; mock, n = 3). (B) Summary of human CD4+ CD25+ T-cell expansion in the spleen of infected mice. (C) Composite data from 13 WT (black), 13 ΔPBM (white), 3 M22 (grey) and 3 mock (crossed) infected mice showing the frequency of human CD3+ T-cells subpopulations: DN (CD4-CD8-), DP (CD4+CD8+) and SP4 and SP8 in the spleen of infected hu-mice. Data are represented as mean± SEM. Statistical significance was determined using the ANOVA test with \*\*\*, *P* < 0.005. (D) Frequency of human CD4+ CD25+ T-cells among human cells in lymphoid organs from infected mice (WT, n = 13; ΔPBM, n = 13; M22, n = 3; mock, n = 3). Lymph nodes were not detected in M22 and mock infected hu-mice. Data are represented as mean± SEM. Statistical difference was calculated with Mann-Whitney *U* test with \*, *P* < 0.05.

<https://doi.org/10.1371/journal.ppat.1006933.g003>

infection (S5 Table). Interestingly, the patterns of Tax/Scribble interaction (Fig 5A) and sub-cellular localization of both Tax and Scribble (Fig 5B and 5C) in these cultured human T-cells were identical to those observed in *ex vivo* splenocytes. We also observed that they expressed similar amounts of Tax (Fig 6A). Likewise, periodic FACS analyses of both types of cells revealed the presence of a majority of CD25+, GITR+, CCR4+ and CADM-1+ T-cells (Fig 6B). Interestingly, cell enumeration performed during several weeks showed that WT T-cells were actively proliferating, in contrast to ΔPBM T-cells that displayed a restrained growth (Fig 6C). Overall, it is important to note that, contrary to ΔPBM T-cells, the proliferation of which was regularly in crisis, WT T-cells constantly proliferated and became immortalized. Taken together, these observations proposed that the Tax-PBM is enhancing HTLV-1-mediated T-cell proliferation and is necessary to immortalize T-cells isolated from hu-mice.





**Fig 4. Tax PBM interacts with Scribble and induces its mis-localization *ex vivo*.** (A) Tax PBM interacts with endogenous Scribble in splenocytes extracted at sacrifice (*ex vivo*) of WT (1, 3) and  $\Delta$  PBM (2) infected-hu mice. A

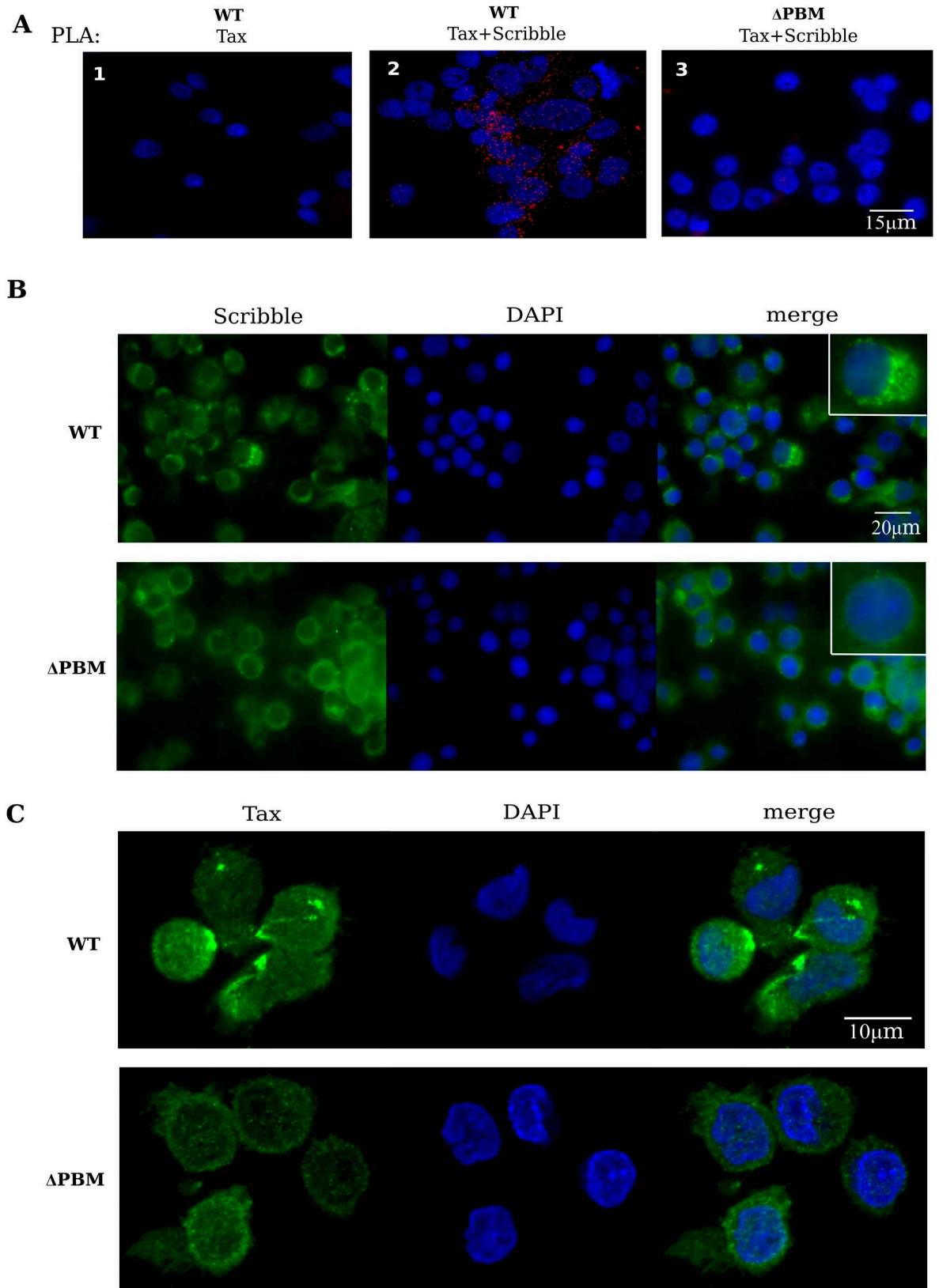
direct quantification of Tax/Scribble interactions (red dots) performed by *in situ* Proximity Ligation Assay (PLA) is shown in panel 4. Primary anti-Tax, anti-Scribble antibodies were combined with secondary PLA probes (Olink Bioscience). Nuclei are stained in blue (DAPI). Negative control (3) was performed in the absence of anti-Scribble antibodies. (B-C) Tax PBM alters subcellular localization of endogenous Scribble in infected hu-mice. Splenocytes collected from WT and  $\Delta$ PBM infected hu-mice were stained at sacrifice with anti-Scribble (B) and anti-Tax (C), and with DAPI (blue) for nuclear staining. Arrows indicate the presence of condensed aggregates of Tax.

<https://doi.org/10.1371/journal.ppat.1006933.g004>

It is well known that T-cells infected by HTLV-1 are forming large clumps when cultivated *in vitro* in presence of IL2. We observed that the clumps of WT T-cells were regular and concentric whereas those of  $\Delta$ PBM T-cells were irregular and eccentric (Fig 6D). These observations are reminiscent of the FACS data showing differences in the size and granularity of *ex vivo* CD4+CD25+ T-cells isolated from the spleens of either WT hu-mice or  $\Delta$ PBM hu-mice (S5 Fig). Furthermore, the WT T-cells displayed long protrusions (filopods) of actin while  $\Delta$ PBM T-cells showed shorter actin filopods, suggesting that PBM might be involved in cell migration (Fig 6D). We also observed that in WT cells, the nuclei were oval (ratio L/l = 1.8) while they were spherical in  $\Delta$ PBM cells, suggesting that the cytoskeleton in WT cells exerts a distortion force on the nuclei (Fig 6D and 6E). In summary, *in vivo* and *in vitro* data indicate that Tax PBM mislocalizes Scribble leading to morphological and cytoskeletal modifications that correlate with a sustained proliferation of WT T-cells.

### Tax PBM impacts transcriptional pathways: A genome-wide transcriptomic analysis of T-cells isolated from infected hu-mice

As PDZ proteins have been directly linked to the control of processes such as cytoskeletal organization, cell polarity and signal transduction pathways, we next investigated global transcriptional pathways that might be dysregulated by the PBM-PDZ interaction [17]. This was achieved by analyzing the transcriptome of cytoplasmic mRNA levels of both WT- and  $\Delta$ PBM T-cells by RNA-sequencing (RNA-seq), at 5 months *in vitro* culture. Differential gene expression analysis using the DESeq2 package (adjusted *P*-value <0.01) resulted in the identification of 629 transcripts downregulated in WT T-cells, 503 of them displaying a fold change of at least 5.6 (Log2 fold-change <-2.5), whereas 400 transcripts were found to be upregulated, 337 of them displaying a fold change of at least 5.6 (Log2 fold-change >2.5) compared to the  $\Delta$ PBM T-cells (Fig 7A, 7B and 7C). We looked for the transcriptional expression of genes coding for the PDZ proteins known to be involved in T-cell homeostasis, such as Scribble, MAGI-1, MAGI-3 and DLG1. We did not observe a significant difference in their expression levels, indicating that expression of these PDZ proteins is PBM-independent. In addition, the number of HTLV-1-related reads was identical under both conditions suggesting that the viral expression was not impaired in  $\Delta$ PBM cells. Finally, we performed a gene ontology analysis of differentially expressed mRNAs (adjusted *P*-value <0.01; log2 fold-change of 2.5). Among the genes upregulated in WT cells we identified genes involved in cell proliferation such as *IL9* (fold change of 57) and cell activation such as *LCK* (fold change of 172) (Figs 7D and S6B). In contrast, genes downregulated in WT T-cells consisted of genes involved in inhibition of cell proliferation such as *CD9* (fold change of 129) and in apoptotic processes such as *RHOB* (fold change of 27). Furthermore, genes related to cytoskeleton organization were also identified as dependent on the Tax/PDZ interactions, some of them upregulated such as *CDC42BPA* while others showed decreased expression such as *FLNB*. Interestingly, the expression of class I regulatory PIK3R6 and PIK3CD subunits was upregulated in WT cells (fold change of 32 and 5.7 respectively). These proteins are implicated in the activation of the Akt/mTOR pathway, involved in cell proliferation and survival. We also identified a gene involved in the non-canonical Wnt pathway (*WNT5B*; increased expression; fold change of 16.7) and two genes





**Fig 5. Tax PBM interacts with Scribble and induces its mis-localization *in vitro*.** (A) Interaction of Tax PBM and endogenous Scribble in cultured T-cells obtained from the spleen of WT and  $\Delta$ PBM hu-mice, by PLA as described in Fig 4A. (B) Subcellular localization of endogenous Scribble in cultured T-cells obtained from the spleen of WT and  $\Delta$ PBM infected hu-mice. (C) Subcellular localization of Tax in cultured T-cells obtained from the spleen of WT and  $\Delta$  PBM infected hu-mice.

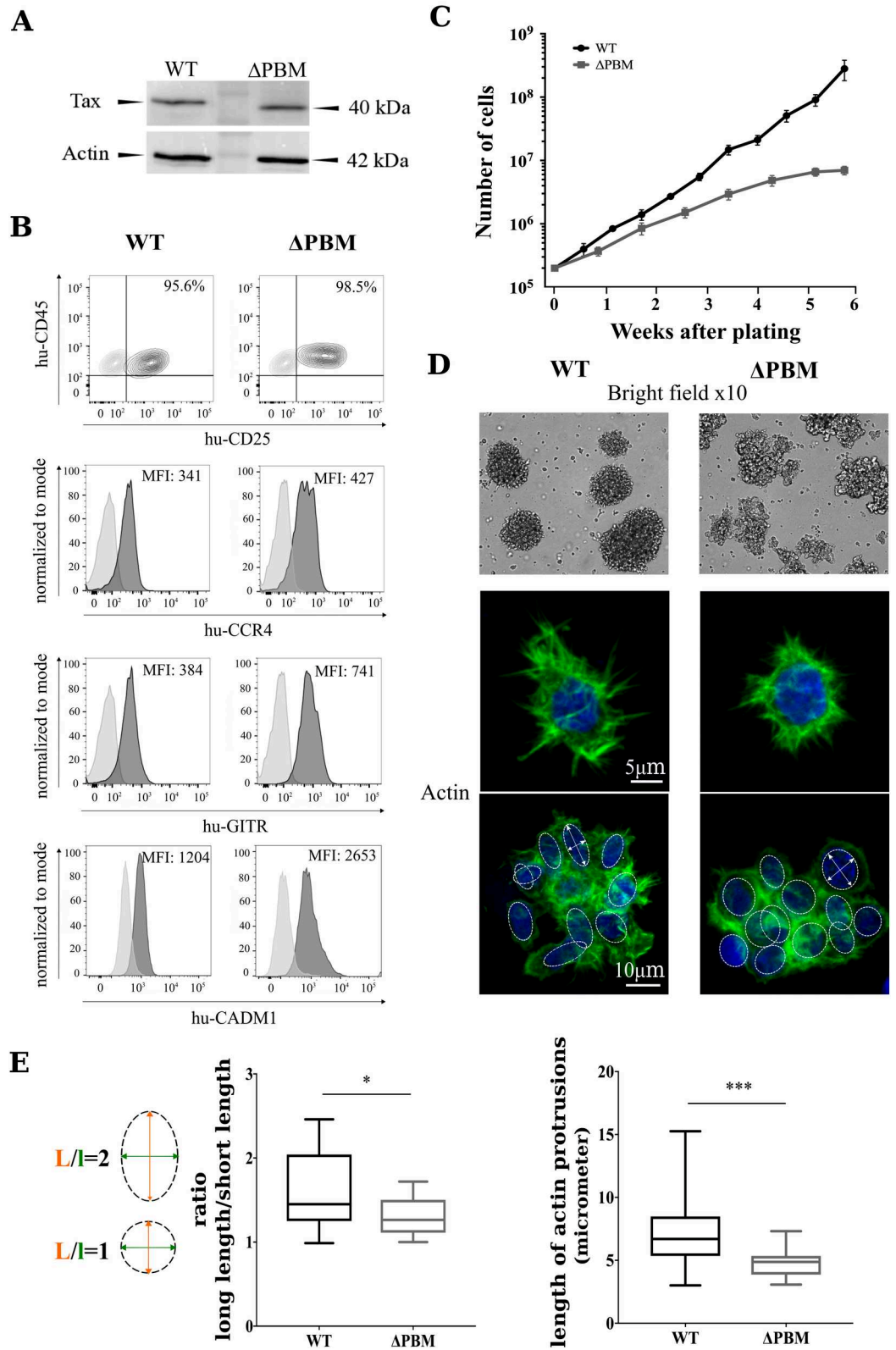
<https://doi.org/10.1371/journal.ppat.1006933.g005>

associated with the canonical Wnt pathway (*WNT2* and *WNT1*; decreased expression; fold change of 16.7 and 147 respectively). This indicates that both canonical and non-canonical Wnt pathways might be modified by the interaction of Tax with PDZ proteins. In conclusion, these results suggest that the Tax PBM is involved in the transcriptional regulation of multiple genes implicated in T-cell proliferation, in the inhibition of apoptosis, as well as in cell polarity, in cytoskeletal and in morphological changes. Taken together, these *in vivo* and *in vitro* findings underline that PBM/PDZ recognition may be required for sustaining HTLV-1 mediated T-cell proliferation, for inducing cell polarity and cytoskeletal modifications and for triggering the immortalization of T-cells in hu-mice.

## Discussion

The generation of hu-mouse models, capable of multi-lineage human hematopoiesis has paved the way for the *in vivo* study of infection by human specific pathogens. Several human viruses and among them lymphotropic viruses have been extensively used in these models [22, 23, 30]. Thus, hu-mice have been used to approach the pathogenic activity of HTLV-1 Tax in a more biological model than cultured cells [25–27, 31]. Here we have investigated the role played by the Tax PBM in the lymphoproliferation triggered by HTLV-1 infection of hu-mice. To achieve this objective, we validated a new procedure to infect these animals with cloned proviruses. Thus, hu-mice were infected either by WT virus or by  $\Delta$ PBM virus carrying a PBM-deleted genome, both produced after transfection of 293T cells with the corresponding provirus.

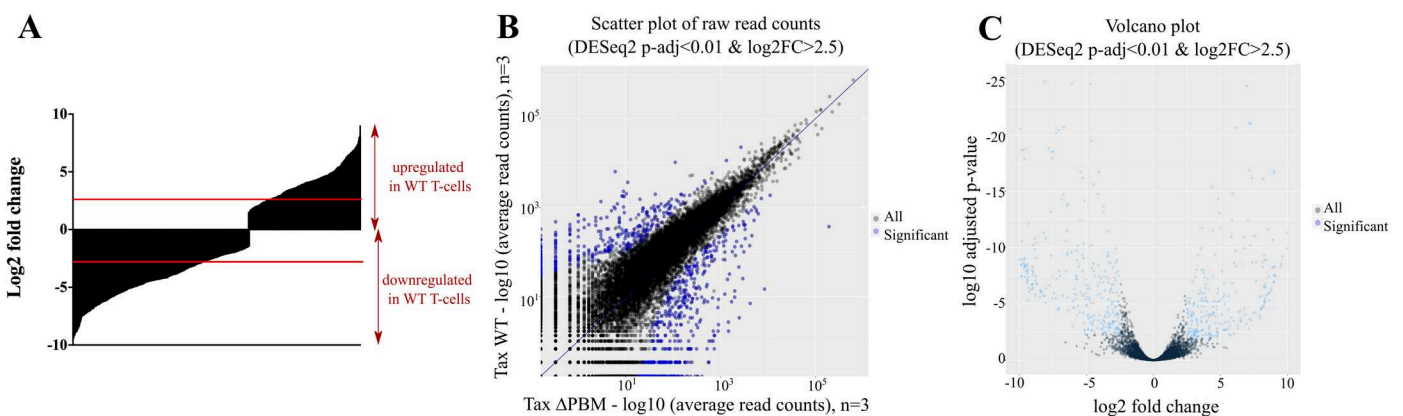
After examining the response of hu-mice to infection by WT or  $\Delta$ PBM HTLV-1, we have characterized T-cells either freshly isolated from the spleen of these infected hu-mice or *in vitro* cultured. We observed an increased number and frequency of activated CD4+CD25+ T-cells in the peripheral blood of WT HTLV-1 hu-mice, albeit at a lower level in  $\Delta$ PBM HTLV-1 infected mice. Splenomegaly, which was observed in both infected hu-mice, is mainly caused by a similar accumulation of CD3+ T-cells. But, once again, the number and frequency of CD4+CD25+ T-cells is significantly higher in WT than in  $\Delta$ PBM hu-mice. Proviral loads and clonality in both types were similar, indicating that the PBM does not impact any of these parameters. Furthermore, we report that T-cell proliferation was severely impaired in hu-mice infected with the ACH-M22 provirus that carries the PBM, but unable to activate the NF- $\kappa$ B pathway. These data indicate for the first time that *in vivo* the PBM alone is unable to induce HTLV-1-mediated proliferation, but is only able to sustain the NF- $\kappa$ B-mediated proliferation. We next analyzed the proliferation of cells collected from the spleen of infected hu-mice and *in vitro* cultivated in presence of IL2. Short-term assays underlined that WT T-cells constantly proliferated over at least more than one year and were therefore considered as immortalized. In contrast,  $\Delta$ PBM T-cells showed a restrained growth and experienced several death crisis suggesting that they were not immortalized. These data provided from infected hu-mice underline that the Tax-PBM is enhancing HTLV-1-mediated T-cell proliferation and is required for T-cell immortalization. A previous study has reported that the Tax PBM significantly increased HTLV-1-induced T-cell proliferation after *in vitro* cocultivation of human PBMCs (Peripheral Blood Mononuclear Cells) with irradiated cell lines producing WT or  $\Delta$ PBM HTLV-1 [21]. These authors report that Tax PBM promotes HTLV-1-induced



**Fig 6. Tax PBM sustains the proliferation of T-cells from HTLV-1 infected hu-mice.** (A) Expression of Tax in the cultured T-cell lines isolated from WT and  $\Delta$ PBM infected hu-mice. Western blot analysis was performed using anti-Tax and anti-actin antibodies. (B) Phenotypic characterization of the cultured T-cell lines isolated from WT and  $\Delta$ PBM infected hu-mice by FACS analysis of the CD25, CCR4, GITR and CADM-1 markers. (C) Growth curves. Human T-cells ( $2 \times 10^5$ ) isolated from the spleen of a WT (black line) or a  $\Delta$ PBM (grey line) infected hu-mice were cultured in growth medium supplemented with IL2 in 24-well plates. Cells were split as indicated and counted. The mean and sem of each time point was determined from triplicate counts from one of three representative experiments performed at 2-, 5- and 8-months of *in vitro* culture. (D) Clumps of WT and  $\Delta$ PBM T-cells and the actin cytoskeleton shown by IF staining. (E) Quantification of the ratio long/short length of the nuclei (Mann-Whitney one-paired:  $P = 0.0267$ ) (left panel), and the length of the protrusions (Mann-Whitney one-paired:  $P = 0.0001$ ) (right panel).

<https://doi.org/10.1371/journal.ppat.1006933.g006>

proliferation of human PBMCs, but that it is not required for virus-mediated immortalization of these cells, as they did not detect any difference in immortalization potential between WT Tax and  $\Delta$ PBM Tax. As this conclusion concerning the implication of the Tax PBM in



**D**

UP-REGULATED IN WT T-CELLS (total of 337 genes)				
	GO-Term	GO-Term id	P-value	# Genes (observed)
Signal transduction	signal transduction	GO:0007165	7,60E-07	129
Proliferation	positive regulation of cell proliferation	GO:0008284	2,52E-06	32
	positive regulation of T cell proliferation	GO:0042102	2,79E-06	10
	positive regulation of JAK-STAT cascade	GO:0046427	2,42E-04	7
Adhesion	positive regulation of leukocyte cell-cell adhesion	GO:1903039	1,89E-06	15
Activation	positive regulation of T cell activation	GO:0050870	1,16E-06	15
Differentiation	positive regulation of leukocyte differentiation	GO:1902107	1,84E-07	13
Cytoskeleton	actin filament-based process	GO:0030029	6,20E-03	19

DOWN-REGULATED IN WT T-CELLS (total of 503 genes)				
	GO-Term	GO-Term id	P-value	# Genes (observed)
Signal transduction	signal transduction	GO:0007165	2,75E-10	201
Inhibition of proliferation	negative regulation of cell proliferation	GO:0008285	5,85E-05	33
Apoptosis	positive regulation of apoptotic process	GO:0043065	3,32E-03	25
Cell shape	cell morphogenesis	GO:0000902	8,86E-04	48

**Fig 7. Genome-wide expression patterns of WT and  $\Delta$ PBM T-cells by RNA-seq.** (A) Graphic representation of transcript expression in WT T-cells compared to  $\Delta$ PBM T-cells expressed as Log2 fold change. (B) Scatterplots comparing the per transcript read count values between WT and  $\Delta$ PBM cells. The vast majority of differentially expressed transcripts (blue dots) in  $\Delta$ PBM showed up-regulation while a smaller number of transcripts showed down-regulation. (C) Volcano plot comparing the per transcript fold change versus adjusted P-value. (D) Differential expression of transcripts (adjusted P-value < 0.01) by GO annotation according to the biological process category, calculated using Genomatix GeneRanker tool.

<https://doi.org/10.1371/journal.ppat.1006933.g007>

immortalization is differing from ours, it is tempting to speculate that such a difference may be related to the specific experimental approach of each study.

Previous studies have shown that the Tax PBM induced the mislocalization of Scribble [14, 15]. However, most if not all of the reported observations have been obtained through the use of transfected cell lines or HTLV-1 T-cell lines. Our study was performed with T-cells from hu-mice clearly indicates that the PBM/PDZ interactions are involved in the distribution pattern of Tax. Among the PDZ proteins, we focused our attention on the Scribble protein known to interact with Tax and to be involved in cell polarity, in T-cell development and proliferation. Based on the results of IF and PLA assays, a comparative analysis of *ex vivo* (immediately after their collection from the spleen) and *in vitro* (cultured in the presence of IL2) WT and  $\Delta$ PBM T-cells underline that the Tax PBM plays a prominent role in the re-localization of both proteins. In WT cells, both Tax and Scribble were observed in large aggregates, mainly at the cell membrane. In contrast, such aggregates were not detected in  $\Delta$ PBM cells, confirming the ability of Tax PBM to sequester Scribble.

The mislocalization of Scribble upon binding to the Tax PBM may be linked to morphological changes that differentially affect the actin cytoskeleton and the polarity of these WT and  $\Delta$ PBM cells. In addition, the binding of the Tax PBM to Scribble may be responsible for the sustained proliferation of WT infected T-cells by negatively interfering with the tumor suppressor property of that PDZ protein. In contrast, the lack of interaction of Tax with Scribble may result in the decreased proliferation of  $\Delta$ PBM T-cells. Further studies in hu-mice will aim at characterizing other PDZ proteins that interact with the Tax PBM, such as DLG-1, to unravel the possible link between these interactions and T-cell proliferation. It will also be of interest to test whether the acetylation of Tax lysine K10 (amino acid 346) located immediately upstream of the PBM could have an impact on the PBM/PDZ interactions, and finally on the localization of Tax and its function in T-cell proliferation.

It has been demonstrated that overexpression of Scribble attenuated NFAT reporter activity in anti-CD3/anti-CD28-stimulated Jurkat cells. By interacting with Scribble, Tax could counteract this negative effect on NFAT activation and thus stimulate T-cell proliferation [14]. Consequently Tax PBM association with PDZ proteins represents an essential event during the development and maintenance of the lymphoproliferative process. Moreover, this association appears to be linked to the exclusive HTLV-1-induced genetic instability observed in human PBMCs infected with WT-HTLV-1 [21]. In that context, to further explore the differences between cells expressing either Tax WT or Tax  $\Delta$ PBM, we performed a comparative transcriptomic analysis that enabled the identification of a set of genes that are differentially expressed in either type of cells. Genes coding for PDZ proteins normally expressed in T lymphocytes such as Scribble, MAGI-1, MAGI-3 and DLG1 were found to be similarly expressed in WT and  $\Delta$ PBM T-cells. In contrast, we noticed the deregulated transcription of genes involved in T-cell signaling and proliferation, in apoptosis induction and in cytoskeleton organization. The expression of PIK3 subunits retained our attention as these kinases activate Akt [32, 33]. We observed that the expression of PIK3 subunits is significantly higher in WT cells than in  $\Delta$ PBM cells, suggesting a direct effect on Akt activation. It has been reported that Tax by binding to the PDZ DLG-1 protein counteracts the negative effects of the PTEN and PHLPP phosphatases and thus activates Akt [16]. Further studies will be needed to determine whether Scribble by interacting with Tax will have such an effect on Akt activation.

Finally, results from the genome wide transcriptome analysis, by supporting our *in vivo* and *in vitro* observations, propose that the Tax PBM plays an eminent role in the pathogenicity of HTLV-1. Further work is needed to establish that such an activity is dependent on PBM-PDZ interactions and to precisely determine which PDZ protein deregulates which set of genes.

As stated in the Introduction, the presence of the PBM was identified not only in Tax but also in several viral oncoproteins such as in the protein E4-ORF1 of adenovirus type 9 and in E6 proteins of several human papilloma viruses [7, 34]. It is worth noting that the presence of a PBM is linked to the ability of HPV-16 and HPV-18 to induce malignant tumors and that such a motif is absent in HPV subtypes that induce benign tumors (for example HPV-9 or-11) [35]. Our data converge to a similar conclusion and stress that the PBM/PDZ recognition is perturbing the regulation of processes such as cytoskeletal organization, cell polarity, cell proliferation. Consequently, the PBM represents a Tax domain endowed with an auxiliary activity essential in the induction and the maintenance of the HTLV-1-induced T-cell proliferation leading to a malignant proliferation. Thus targeting the PBM/PDZ nodes by small peptides is offering a novel strategy to slowdown the T-cell proliferation in HTLV-1 infected hu-mice [36].

## Materials and methods

### Cells and plasmids

The human embryonic kidney (HEK) 293T cells (American Type Culture Collection CRL-3216) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) (Sigma-Aldrich, France), 50 µg/ml streptomycin, 100 U/ml penicillin (Invitrogen, France).

The ACH plasmid is an infectious molecular clone of HTLV-1 [37]. The HTLV-1 provirus deleted from the PBM of Tax (ACH-ΔPBM) was constructed by introducing a TAA stop codon instead of the GAA codon in the Tax C-terminus resulting in loss of the last four amino-acids of Tax (ETE:V: consensus PBM). This mutation was confirmed by DNA sequencing. The ACH-M22 plasmid encoding for a mutated Tax protein that do not activate the NF-κB pathway is a kind gift of Dr. F. Bex (Belgium).

### Cell transfection and Gag p19 ELISA

293T cells were plated at  $5 \times 10^5$  cells in a 6-well plate the day prior to transfection. Plasmid DNA (3.3 µg) was applied to the cells as calcium phosphate coprecipitates. Medium was changed 6h after transfection. One day later, supernatants were collected and analyzed for HTLV p19 antigen content by using the Retrotek ELISA-kit (ZeptoMetrix Corp., USA).

### Ethics statement

Anonymized human umbilical cord samples from the Maternity Ward of Hôpital Femme-Mère-Enfant (Bron, France) were obtained from healthy full-term newborns with written parental informed consent according to the guidelines of the medical and ethical committees of Hospices Civils de Lyon and of Agence de la Biomédecine, Paris, France. Experiments using cord blood were approved by both committees and were performed in full compliance with French law. Animal experimentation was performed in strict accordance with the French "Comité National de Réflexion Ethique sur l'Expérimentation Animale, n° 15" and the ethical guidelines for the care of these mice of the Plateau de Biologie Expérimentale de la Souris (PBES, UMS 3444) at Ecole Normale Supérieure (ENS) de Lyon. This protocol has been approved by the Committee on the Ethics of Animal Experiments of ENS de Lyon (approval number: ENS\_2014\_043). All efforts were made to minimize animal suffering.

### Isolation of human CD34<sup>+</sup> cells from cord blood samples

After density gradient centrifugation of human cord blood, CD34<sup>+</sup> cells were enriched twice using immunomagnetic beads according to the manufacturer instructions (CD34<sup>+</sup> MicroBead



Kit, Miltenyi Biotec, Bergisch-Gladbach, Germany). Purity ( $\geq 95\%$ ) was evaluated by FACS analysis using human PE-CD34 antibody (Miltenyi Biotec). Cells were frozen before the transplantation when newborn mice were available.

### Generation and infection of humanized mice

NOD.Cg-Prkdc<sup>scid</sup> Il2ry<sup>tm1Wjl</sup> Tg(HLA.A2.1)1Enge/SzJ (NSG-HLA-A2/HDD) were obtained from the Jackson Laboratory (Bar Harbor, USA) and bred and maintained under pathogen-free conditions at the PBES. Newborn males and females NSG mice (2 to 5 days old) were sublethally irradiated with 1.1 Gray (320 kV, 12.5 mA) from a X-ray irradiator (XRad-320, PXI Precision XRay, France) and intra-hepatically injected with  $2 \times 10^5$  human CD34<sup>+</sup> hematopoietic stem cells isolated from cord blood samples [38]. After 10 weeks, humanized mice ( $\geq 30\%$  hu-CD45<sup>+</sup> cells in peripheral blood) were infected by HTLV-1. Infection and mice monitoring were performed in a Biosafety Level 3 Laboratory in accordance with the PBES guidelines. Lethally irradiated 293T cells (50 Gray, 320kV, 12.5 mA) transfected with full length or truncated HTLV-1 molecular clone were intra-peritoneally injected: the amount of irradiated cells to inject per mouse corresponds to the number of cells producing 70 ng of p19 in 24h-culture. Mock infected mice were injected with the same amount of irradiated non-transfected 293T. Hu-mice were daily monitored for signs of obvious suffering, such as weight loss, back arches and prostrated behavior. Peripheral blood was collected from the retro-orbital venous sinus under Isoflurane anesthesia. When mice were either suffering or displaying more than 10% circulating CD25<sup>+</sup> T-cells, they were sacrificed after anesthesia. Tissue specimens (spleen, mesenteric lymph node and tibia bone) were collected and gently minced in PBS to obtain a single-cell suspension and immediately frozen in FCS containing 10% DMSO and kept at  $-80^\circ\text{C}$ .

### Cell preparation and flow cytometry analysis

Spleens from infected hu-mice were harvested and analyzed at indicated time points following infection. To obtain a single-cell suspension, spleens were minced and passed through a nylon mesh. Red cell lysis was performed in red cell lysis buffer (Sigma, France) for 10 min. Cells were then washed and enumerated. For flow cytometry, single-cell suspensions were stained with the appropriate monoclonal antibody (S1 Table) or the respective isotype control antibody for 30 min in the dark at  $4^\circ\text{C}$ . Human lymphocytes first gated as hCD45<sup>+</sup> cells were then defined as CD3<sup>+</sup> T-cells containing the following subsets: DN, CD4-CD8<sup>-</sup>; DP, CD4+CD8<sup>+</sup>; and SP, CD4<sup>+</sup> CD8<sup>-</sup> or CD4-CD8<sup>+</sup>. Absolute numbers of cells were determined by multiplying the number of nucleated cells by the percentage of positive cells for the indicated cell surface marker(s). For CADM-1 expression analysis, cells were stained with the primary antibody (Rabbit polyclonal antibody H-300 Santa Cruz Biotechnology, USA) for 2h at  $4^\circ\text{C}$  in PBS containing 5% FCS then washed twice and incubated 30 min at  $4^\circ\text{C}$  in the dark with a secondary fluorescent antibody anti-rabbit (A-11008 Molecular Probes, France). After two wash steps in FACS-buffer, fluorescence was measured on a flow cytometer (FACSCanto II, BD, San Jose, CA, USA). Cells were always gated to exclude doublet. Compensations were realized using Miltenyi MACS Comp Beads. Data were evaluated with BD Diva software (Becton Dickinson Immunocytometry Systems, Mountain View, CA) and analyzed using FlowJo software (TreeStar, Ashland, USA). Results are expressed as the mean of % positivity of surface expression  $\pm$  SEM or as the Mean Fluorescence Intensity (MFI).

To obtain cell lines derived from infected hu-mice, single cell suspensions were cultured in complete RPMI 1640 medium supplemented with 10% FCS (Sigma-Aldrich, France). Recombinant IL2 (20 U/ml) (Peprotech, France) was added to the cultures every 3 days. After one

month of culture, selected cell lines were tested for the relevant proviral sequence and weekly monitored for their proliferation.

### Histology and immunohistochemistry

Spleen samples were fixed with 4% paraformaldehyde in PBS, embedded in paraffin, sectioned and stained with H&E solution. An indirect immunoperoxidase technique with commercially available monoclonal antibodies to CD3 and rabbit polyclonal antibody to Tax (kind gift of Dr. B. Cullen) was applied to the tissue sections as previously described [27]. Pictures were analyzed with ImageJ software.

### Immunofluorescent (IF) staining and proximity ligation assay (PLA)

Slides or ibiTreat  $\mu$ -dishes (IBIDI) were pre-coated with poly-L-lysine (Sigma-Aldrich, France) 20  $\mu$ g/ml for IF and 10  $\mu$ g/ml for PLA. To detect Tax and Scribble proteins, T-cells isolated from spleens were added to the slides before being fixed with 3% paraformaldehyde in PBS for 10 min at room temperature. Cells were then permeabilized with TritonX-100 or methanol at  $-20^{\circ}\text{C}$ . For IF staining, cells were blocked with 5% FCS then stained with primary and secondary antibodies and mounting medium with DAPI (Duolink, Sigma, France). Polyclonal goat antibodies to Scribble (C-20; Santa Cruz Biotechnology), rabbit antibodies to Tax (kind gift of Dr. B. Cullen) and appropriate controls were used. Phalloidin, fluorescein isothiocyanate labeled (Sigma-Aldrich) interacts with polymeric actin. PLA was carried out with Duolink In Situ-Fluorescence Red kit (Sigma-Aldrich) according to the manufacturer's instructions. Negative controls were performed on T-cells in the absence of antibodies to Scribble. Microscopic examination was performed using a Zeiss LSM710 confocal microscope (Carl Zeiss Jena, inc, Germany) or an Axioimager Z1 epifluorescence microscope (Carl Zeiss Jena, inc, Germany). Images were analyzed using ImageJ software.

### Statistical analysis

Statistical tests were performed using GraphPad PRISM software. When  $n \leq 5$  in one or both groups, they are one-tailed and the non-parametric Wilcoxon-Mann-Whitney test was performed. When  $n > 5$  in one or both groups, the parametric Student t-test was performed if variance are equal (F-test with  $P$ -value  $> 0.05$ ). If not (F-test with  $P$ -value  $< 0.05$ ), the non-parametric Wilcoxon-Mann-Whitney  $U$  test was performed. Statistical analysis of hu-CD3 subpopulations was performed with one-way ANOVA test. The results were considered statistically significant when  $P$ -value  $< 0.05$ .

### Western blot analysis

Cells were lysed with sample buffer (10 mM HEPES, pH 7.9, 500 mM NaCl, 3 mM  $\text{MgCl}_2$ , 1 mM DTT, 1 mM PMSF, and 0.5% Triton X-100 supplemented with protease inhibitors). After incubation on ice for 60 min, whole cell lysates were centrifuged at 15,000 g for 10 min at  $4^{\circ}\text{C}$  to remove the debris. Protein concentration of the cleared lysates was determined using the Bradford assay. Cell lysates (15  $\mu$ g) were size-separated by electrophoresis on a 12% SDS-polyacrylamide gel (3h migration at 20 mA) and transferred onto PVDF membranes. The blot was blocked in PBS-5% milk and incubated with anti-Tax antibodies (1:1,000; kind gift of B. Cullen), anti- $\beta$ -actin (1:5,000) obtained from Sigma-Aldrich (clone AC-15). After several washes, horseradish peroxidase (HRP)-conjugated anti-rabbit IgG (1:10,000; Cell Signaling, The Netherlands) or HRP-linked anti-mouse IgG (1:10,000; GE Healthcare, France) were added to the membranes which were washed again several times and subsequently incubated with the

Western Lightning ECL solution (Thermo Fisher Scientific, France). Images were captured using a ChemiDoc Imaging system (Biorad, France).

### DNA and RNA extraction, quantitative real-time PCR (qPCR) and proviral load

Genomic DNA was extracted from the single cell suspension using the Nucleospin Blood kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. PVL was measured by qPCR with HTLV-1 *tax*-specific primers. The PVL was calculated as previously described [27] and expressed as the number of *pX* copies per 100 human cells.

RNA was extracted from the single cell suspension using RNAzol RT (Sigma Aldrich, France) according to the manufacturer's instructions, and resuspended in 10  $\mu$ l of RNase-free water and treated with 10 U of RNase-free DNase I (Qiagen, Hilden, Germany) for 15 min at 30°C and then for 15 min at 60°C. 500 ng of total RNA were then retro-transcribed at 42°C during 50 min in a total volume of 20  $\mu$ l reaction buffer containing 100 U of SuperScript II reverse transcriptase (RT; Invitrogen, CA, USA). A reaction without RT was performed as a control for genomic DNA contamination. The mRNA levels were normalized using 3 different housekeeping genes (*ACTB*, *RSP11* and *RSP14*) chosen to be the most stable in our model with BestKeeper and NormFinder algorithms. Quantitative real-time PCR (qPCR) was performed using the FastStart Universal SYBR Green Master (Roche, Mannheim, Germany) on a StepOnePlus system (Applied Biosystem, CA, USA). The initial denaturation step at 95°C for 10 min was followed by 40 cycles with one cycle consisting of 10s at 95°C, 30s at 60°C, and 15s at 72°C.

### Primers

The nucleotide sequences of the primers were used for RT-PCR, proviral load measurement and DNA sequencing of the mutation of *Tax* gene are shown in [S2 Table](#).

### High-throughput sequencing (HTS) of HTLV-1 integration sites

To determine the number and abundance of HTLV-1 infected clones in humanized mice, we used an improved quantitative HTS method to map the proviral integration sites in the human genome [39]. Libraries were prepared starting from 500 ng DNA and sequenced on an Illumina MiSeq instrument. 150 bp paired-end reads were acquired and sequencing reads that supported either the 5' or the 3' LTR-host junctions were retained. The number of unique integration sites (UIS) was determined as previously described [40].

### Cell lysis and RNA-seq analysis

T lymphocytes isolated from the spleen of WT or  $\Delta$ PBM infected hu-mice were cultured in complete RPMI medium containing IL2. Cells ( $5 \times 10^6$ ) were then washed with ice-cold PBS, centrifuged at 500 g for 5 min at 4°C and lysed in 1 ml of lysis buffer (10mM Tris-HCl pH7.5, 5mM MgCl<sub>2</sub>, 100mM KCl, 2mM DTT, protease inhibitor EDTA-free (Roche, Mannheim, Germany), 1% Triton X-100). Lysates were gently homogenized and incubated at 4°C for 10 min, centrifuged at 1,300 g for 10 min at 4°C and the supernatant was recovered. Total RNA was extracted from cellular extracts using Trizol and subjected to cDNA library construction using the smartseq2 protocol [41].



## Mapping of high-throughput sequencing reads

Reads were first split with respect to their 5'-barcode sequence. After this, 5'-barcode and 3'-adaptor sequences were removed from reads. Reads were then aligned to a custom set of sequences corresponding to ribosomal RNA and tRNA sequences using Bowtie [42] in order to remove contaminants. Remaining reads were aligned to the human genome and transcriptome (hg19 assembly) using TopHat2 [43].

## Transcript-level quantification and differential gene expression

The alignment files obtained from TopHat2 were used to count reads mapping to the 5'UTR coding sequence and 3'UTR of human transcripts using HTSeq [44] and the UCSC hg19 gene annotation file. Differential gene expression was performed using the R package DESeq2 [45].

## Data analysis for HTS

Only genes with an adjusted *P*-value  $\leq 0.01$  and a log<sub>2</sub> foldchange superior to 2.5 were selected. GeneOntology was done using the GeneRanker tool of Genomatix software.

## Supporting information

**S1 Table. Monoclonal antibodies and isotype controls used in flow cytometry.** IgG indicates immunoglobulin G; FITC, fluorescein isothiocyanate; PE, phycoerythrin; APC, allophycocyanin; V450, BD Horizon V450, a coumarin dye excited by the violet laser.  
(DOCX)

**S2 Table. Primers used for PCR and RT-PCR.**  
(DOCX)

**S3 Table. Human T-cell subsets in the peripheral blood (PB) of infected hu-mice, seven weeks after HTLV-1 infection.**

<sup>a</sup>hu-mice were intraperitoneally inoculated with 293T cells transfected with ACH-WT (n = 5), ACH-ΔPBM (n = 8), ACH-M22 (n = 3) or mock infected (n = 3) and then X-irradiated.

Peripheral blood samples were collected at 7 weeks after infection.

<sup>b</sup>Frequencies of the CD3+, CD4+CD25+, and CD8+CD25+ cells in the peripheral blood were calculated out of hu-CD45+ cells.

(DOCX)

**S4 Table. Proviral load, Tax mRNA expression and pathological features in hu-mice infected with HTLV-1.** Hu-mice were intraperitoneally inoculated with irradiated 293T cells transfected with the indicated plasmids. They were sacrificed at indicated times. Proviral load is expressed as number of proviral copies per 10<sup>5</sup> splenocytes. LN = lymph node; nd = not determined. Levels of Tax mRNA in splenocytes isolated from HTLV-1-infected hu-mice were measured by RT-qPCR as indicates in Fig 2B.

(DOCX)

**S5 Table. Sequencing analysis of representative samples.**

<sup>a</sup>DNA from ACH plasmids, from mouse splenocytes and from cultured T-cells were extracted as indicated in Materials and methods, subjected to PCR amplification and sequenced by using the primers listed in S1 Table.

<sup>b</sup>Nucleotide sequence of Tax in *italic* and of PBM in **bold**. Note the mutation of **GAA** into **TAA** (stop codon) in ACH ΔPBM plasmid, in the spleen of ΔPBM infected hu-mice and in

$\Delta$ PBM T-cells cultured *in vitro*.  
(DOCX)

**S6 Table. Human T-Cell subsets in the bone marrow (BM), lymph nodes (LN) and the spleen (SPL) of infected hu-mice at the autopsy.**

<sup>a</sup>hu-mice were intraperitoneally inoculated with 293T cells transfected with ACH-WT (n = 13), ACH- $\Delta$ PBM (n = 13), ACH-M22 (n = 3) or mock infected (n = 3) and then X-irradiated.

<sup>b</sup>Bone Marrow (BM) from tibia, Spleen (SPL) and mesenteric lymph nodes (LN) were collected and analyzed by FACS for indicated surface markers.

<sup>c</sup>Frequency of the CD3+, CD4+CD25+, and CD8+CD25+ cells were calculated out of the number of hu-CD45+ cells.

(DOCX)

**S1 Fig. Tax transcriptional activation of CREB/ATF- and NF- $\kappa$ B-dependent reporter genes.**

293T cells ( $9 \times 10^4$ / 24 well) were cotransfected with the indicated ACH plasmid (100 ng), TK-*Renilla* (5 ng) reporter plasmid together with the HTLV-1 LTR-luc (A), or the  $\kappa$ B-luc (B) as calcium phosphate coprecipitates. Cell lysates were harvested 48h after transfection and luciferase activity was determined using the Dual Luciferase Assay System (Promega). The histogram presents the average fold activation over control values for 2 independent experiments in triplicate; data are presented as mean  $\pm$  SEM.

(TIF)

**S2 Fig. Number of unique integration sites (UIS) in both types of infected hu-mice.** The number of independent HTLV-1-infected clones was determined by HTS clonality analysis in splenocytes (8 WT and 9  $\Delta$ PBM). Bar represents mean. Student *t*-test,  $P = 0.3021$ .

(TIF)

**S3 Fig. Immunohistochemistry of spleen sections of different WT and  $\Delta$ PBM infected hu-mice.** Staining with anti-Tax antibodies revealed an infiltration of T-lymphocytes with a nuclear localization of Tax.

(TIF)

**S4 Fig. FACS analysis of splenic T-cells in HTLV-1 infected hu-mice.** Splenocytes from WT or  $\Delta$ PBM-infected hu-mice were harvested 7 weeks after infection. Representative profile for CD4, CD8, and CD25 expression on gated hu-CD3+ cells.

(TIF)

**S5 Fig. (A) Size (FSC for Forward Scatter) and (B) Granularity (SSC for Side Scatter) of CD4+CD25+ T-cells in the spleen of WT and  $\Delta$ PBM hu-mice.**

(TIF)

**S6 Fig. Gene Ontology Analysis.** (A) Reads were mapped on the human genome (hg19). They are specific of gene exons and do not map on repeated sequences. Shown is the number of reads in the WT cells (in purple) and  $\Delta$ PBM cells (in orange). (B) Detailed list of the differential expression of transcripts (adjusted  $P$ -value < 0.01) by GO annotation according to the biological process category, calculated using Genomatix GeneRanker tool.

(TIF)

## Acknowledgments

We thank Pr. Pascal Gaucherand, Virginie Finat (Hôpital Femme-Mère-Enfant, Bron) and Dominique Rigal (Etablissement Français du sang, Lyon) for kindly providing us with cord blood units, Nicolas Gadot (Anipath, Lyon) and the members of animal facility (PBES/

UMS3444, ENS Lyon) for technical support, and Vincent Gache (ENS, Lyon) for helpful advices. We are also grateful to Françoise Bex (Brussels, Belgium) for helpful discussions. We thank the Pôle Scientifique de Modélisation Numérique (PSMN ENS-Lyon, France) for computer resource. Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the ‘France Génomique’ consortium (ANR-10-INBS-0009). We thank Wouter Coppeters, Latifa Karim, Manon Deckers and the GIGA Genomics Platform (Liège, Belgium) for sequencing services and excellent technical support.

## Author Contributions

**Conceptualization:** Eléonore Pérès, Louis Gazzolo, Pierre Jalinot, Madeleine Duc Dodon.

**Data curation:** Eléonore Pérès.

**Formal analysis:** Eléonore Pérès, Emiliano P. Ricci, Anne Van den Broeke, Louis Gazzolo, Madeleine Duc Dodon.

**Funding acquisition:** Madeleine Duc Dodon.

**Investigation:** Eléonore Pérès, Juliana Blin, Maria Artesi, Louis Gazzolo, Madeleine Duc Dodon.

**Methodology:** Eléonore Pérès, Juliana Blin, Maria Artesi.

**Project administration:** Eléonore Pérès, Louis Gazzolo, Madeleine Duc Dodon.

**Resources:** Lee Ratner, Madeleine Duc Dodon.

**Software:** Eléonore Pérès, Emiliano P. Ricci, Vincent Hahaut, Antoine Corbin.

**Supervision:** Anne Van den Broeke, Louis Gazzolo, Pierre Jalinot, Madeleine Duc Dodon.

**Validation:** Eléonore Pérès, Emiliano P. Ricci, Louis Gazzolo, Lee Ratner, Pierre Jalinot, Madeleine Duc Dodon.

**Visualization:** Eléonore Pérès, Louis Gazzolo, Madeleine Duc Dodon.

**Writing – original draft:** Eléonore Pérès, Louis Gazzolo, Madeleine Duc Dodon.

**Writing – review & editing:** Anne Van den Broeke, Madeleine Duc Dodon.

## References

1. Ishitsuka K, Tamura K. Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma. *Lancet Oncol.* 2014; 15(11):e517–26. [https://doi.org/10.1016/S1470-2045\(14\)70202-5](https://doi.org/10.1016/S1470-2045(14)70202-5) PMID: [25281470](https://pubmed.ncbi.nlm.nih.gov/25281470/).
2. Matsuoka M, Jeang KT. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer.* 2007; 7(4):270–80. <https://doi.org/10.1038/nrc2111> PMID: [17384582](https://pubmed.ncbi.nlm.nih.gov/17384582/).
3. Baydoun HH, Bai XT, Shelton S, Nicot C. HTLV-I tax increases genetic instability by inducing DNA double strand breaks during DNA replication and switching repair to NHEJ. *PLoS One.* 2012; 7(8):e42226. <https://doi.org/10.1371/journal.pone.0042226> PMID: [22916124](https://pubmed.ncbi.nlm.nih.gov/22916124/); PubMed Central PMCID: [PMCPMC3423393](https://pubmed.ncbi.nlm.nih.gov/PMC3423393/).
4. Haoudi A, Semmes OJ. The HTLV-1 tax oncoprotein attenuates DNA damage induced G1 arrest and enhances apoptosis in p53 null cells. *Virology.* 2003; 305(2):229–39. PMID: [12573569](https://pubmed.ncbi.nlm.nih.gov/12573569/).
5. Matsuoka M, Jeang KT. Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy. *Oncogene.* 2011; 30(12):1379–89. <https://doi.org/10.1038/onc.2010.537> PMID: [21119600](https://pubmed.ncbi.nlm.nih.gov/21119600/); PubMed Central PMCID: [PMCPMC3413891](https://pubmed.ncbi.nlm.nih.gov/PMC3413891/).
6. Lodewick J, Lamsoul I, Bex F. Move or die: the fate of the Tax oncoprotein of HTLV-1. *Viruses.* 2011; 3(6):829–57. <https://doi.org/10.3390/v3060829> PMID: [21994756](https://pubmed.ncbi.nlm.nih.gov/21994756/); PubMed Central PMCID: [PMC3185767](https://pubmed.ncbi.nlm.nih.gov/PMC3185767/).

7. Lee C, Laimins LA. Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. *J Virol.* 2004; 78(22):12366–77. <https://doi.org/10.1128/JVI.78.22.12366-12377.2004> PMID: [15507623](https://pubmed.ncbi.nlm.nih.gov/15507623/); PubMed Central PMCID: PMCPMC525055.
8. Lee SS, Weiss RS, Javier RT. Binding of human virus oncoproteins to hDlg/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. *Proc Natl Acad Sci U S A.* 1997; 94(13):6670–5. PMID: [9192623](https://pubmed.ncbi.nlm.nih.gov/9192623/); PubMed Central PMCID: PMCPMC21216.
9. Rousset R, Fabre S, Desbois C, Bantignies F, Jalinot P. The C-terminus of the HTLV-1 Tax oncoprotein mediates interaction with the PDZ domain of cellular proteins. *Oncogene.* 1998; 16(5):643–54. <https://doi.org/10.1038/sj.onc.1201567> PMID: [9482110](https://pubmed.ncbi.nlm.nih.gov/9482110/).
10. Fanning AS, Anderson JM. Protein modules as organizers of membrane structure. *Curr Opin Cell Biol.* 1999; 11(4):432–9. [https://doi.org/10.1016/S0955-0674\(99\)80062-3](https://doi.org/10.1016/S0955-0674(99)80062-3) PMID: [10449334](https://pubmed.ncbi.nlm.nih.gov/10449334/).
11. Subbaiah VK, Kranjec C, Thomas M, Banks L. PDZ domains: the building blocks regulating tumorigenesis. *Biochem J.* 2011; 439(2):195–205. <https://doi.org/10.1042/BJ20110903> PMID: [21954943](https://pubmed.ncbi.nlm.nih.gov/21954943/).
12. Javier RT, Rice AP. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J Virol.* 2011; 85(22):11544–56. <https://doi.org/10.1128/JVI.05410-11> PMID: [21775458](https://pubmed.ncbi.nlm.nih.gov/21775458/); PubMed Central PMCID: PMCPMC3209276.
13. James CD, Roberts S. Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? *Pathogens.* 2016; 5(1). <https://doi.org/10.3390/pathogens5010008> PMID: [26797638](https://pubmed.ncbi.nlm.nih.gov/26797638/); PubMed Central PMCID: PMCPMC4810129.
14. Arpin-Andre C, Mesnard JM. The PDZ domain-binding motif of the human T cell leukemia virus type 1 tax protein induces mislocalization of the tumor suppressor hScrib in T cells. *J Biol Chem.* 2007; 282(45):33132–41. <https://doi.org/10.1074/jbc.M702279200> PMID: [17855372](https://pubmed.ncbi.nlm.nih.gov/17855372/).
15. Okajima M, Takahashi M, Higuchi M, Ohsawa T, Yoshida S, Yoshida Y, et al. Human T-cell leukemia virus type 1 Tax induces an aberrant clustering of the tumor suppressor Scribble through the PDZ domain-binding motif dependent and independent interaction. *Virus Genes.* 2008; 37(2):231–40. <https://doi.org/10.1007/s11262-008-0259-4> PMID: [18661220](https://pubmed.ncbi.nlm.nih.gov/18661220/).
16. Cherian MA, Baydoun HH, Al-Saleem J, Shkriabai N, Kvaratskhelia M, Green P, et al. Akt Pathway Activation by Human T-cell Leukemia Virus Type 1 Tax Oncoprotein. *J Biol Chem.* 2015; 290(43):26270–81. <https://doi.org/10.1074/jbc.M115.684746> PMID: [26324707](https://pubmed.ncbi.nlm.nih.gov/26324707/); PubMed Central PMCID: PMCPMC4646275.
17. Ludford-Menting MJ, Oliaro J, Sacirbegovic F, Cheah ET, Pedersen N, Thomas SJ, et al. A network of PDZ-containing proteins regulates T cell polarity and morphology during migration and immunological synapse formation. *Immunity.* 2005; 22(6):737–48. <https://doi.org/10.1016/j.immuni.2005.04.009> PMID: [15963788](https://pubmed.ncbi.nlm.nih.gov/15963788/).
18. Humbert PO, Grzeschik NA, Brumby AM, Galea R, Elsum I, Richardson HE. Control of tumorigenesis by the Scribble/Dlg/Lgl polarity module. *Oncogene.* 2008; 27(55):6888–907. <https://doi.org/10.1038/onc.2008.341> PMID: [19029932](https://pubmed.ncbi.nlm.nih.gov/19029932/).
19. Hirata A, Higuchi M, Niinuma A, Ohashi M, Fukushi M, Oie M, et al. PDZ domain-binding motif of human T-cell leukemia virus type 1 Tax oncoprotein augments the transforming activity in a rat fibroblast cell line. *Virology.* 2004; 318(1):327–36. <https://doi.org/10.1016/j.virol.2003.10.006> PMID: [14972558](https://pubmed.ncbi.nlm.nih.gov/14972558/).
20. Higuchi M, Tsubata C, Kondo R, Yoshida S, Takahashi M, Oie M, et al. Cooperation of NF-kappaB2/p100 activation and the PDZ domain binding motif signal in human T-cell leukemia virus type 1 (HTLV-1) Tax1 but not HTLV-2 Tax2 is crucial for interleukin-2-independent growth transformation of a T-cell line. *J Virol.* 2007; 81(21):11900–7. <https://doi.org/10.1128/JVI.00532-07> PMID: [17715223](https://pubmed.ncbi.nlm.nih.gov/17715223/); PubMed Central PMCID: PMCPMC2168800.
21. Xie L, Yamamoto B, Haoudi A, Semmes OJ, Green PL. PDZ binding motif of HTLV-1 Tax promotes virus-mediated T-cell proliferation in vitro and persistence in vivo. *Blood.* 2006; 107(5):1980–8. <https://doi.org/10.1182/blood-2005-03-1333> PMID: [16263794](https://pubmed.ncbi.nlm.nih.gov/16263794/); PubMed Central PMCID: PMCPMC1895710.
22. Rongvaux A, Takizawa H, Strowig T, Willinger T, Eynon EE, Flavell RA, et al. Human hemato-lymphoid system mice: current use and future potential for medicine. *Annu Rev Immunol.* 2013; 31:635–74. <https://doi.org/10.1146/annurev-immunol-032712-095921> PMID: [23330956](https://pubmed.ncbi.nlm.nih.gov/23330956/); PubMed Central PMCID: PMCPMC4120191.
23. Legrand N, Ploss A, Balling R, Becker P, Borsotti C, Brezillon N, et al. Humanized mice for modeling human infectious disease: challenges, progress, and outlook. *Cell host & microbe.* 2009; 6(1):5–9. <https://doi.org/10.1016/j.chom.2009.06.006> PMID: [19616761](https://pubmed.ncbi.nlm.nih.gov/19616761/)
24. Banerjee P, Crawford L, Samuelson E, Feuer G. Hematopoietic stem cells and retroviral infection. *Retrovirology.* 2010; 7:8. <https://doi.org/10.1186/1742-4690-7-8> PMID: [20132553](https://pubmed.ncbi.nlm.nih.gov/20132553/)
25. Duc Dodon M, Villaudy J, Gazzolo L, Haines R, Lairmore M. What we are learning on HTLV-1 pathogenesis from animal models. *Front Microbiol.* 2012; 3:320. Epub 2012/09/13. <https://doi.org/10.3389/fmicb.2012.00320> PMID: [22969759](https://pubmed.ncbi.nlm.nih.gov/22969759/); PubMed Central PMCID: PMCPMC3431546.

26. Tezuka K, Xun R, Tei M, Ueno T, Tanaka M, Takenouchi N, et al. An animal model of adult T-cell leukemia: humanized mice with HTLV-1-specific immunity. *Blood*. 2014; 123(3):346–55. <https://doi.org/10.1182/blood-2013-06-508861> PMID: 24196073.
27. Villaudy J, Wencker M, Gadot N, Gillet NA, Scoazec JY, Gazzolo L, et al. HTLV-1 propels thymic human T cell development in "human immune system" Rag2(-)/(-) gamma c(-)/(-) mice. *PLoS Pathog*. 2011; 7(9):e1002231. <https://doi.org/10.1371/journal.ppat.1002231> PMID: 21909275; PubMed Central PMCID: PMC3164654.
28. Tsubata C, Higuchi M, Takahashi M, Oie M, Tanaka Y, Gejyo F, et al. PDZ domain-binding motif of human T-cell leukemia virus type 1 Tax oncoprotein is essential for the interleukin 2 independent growth induction of a T-cell line. *Retrovirology*. 2005; 2:46. <https://doi.org/10.1186/1742-4690-2-46> PMID: 16042787; PubMed Central PMCID: PMC1199618.
29. Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius KJ, Jarvius J, et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods*. 2006; 3(12):995–1000. <https://doi.org/10.1038/nmeth947> PMID: 17072308.
30. Akkina R, Berges BK, Palmer BE, Remling L, Neff CP, Kuruville J, et al. Humanized Rag1-/- gammac-/- mice support multilineage hematopoiesis and are susceptible to HIV-1 infection via systemic and vaginal routes. *PLoS One*. 2011; 6(6):e20169. <https://doi.org/10.1371/journal.pone.0020169> PMID: 21695116; PubMed Central PMCID: PMC3114781.
31. Percher F, Curis C, Peres E, Artesi M, Rosewick N, Jeannin P, et al. HTLV-1-induced leukotriene B4 secretion by T cells promotes T cell recruitment and virus propagation. *Nat Commun*. 2017; 8:15890. <https://doi.org/10.1038/ncomms15890> PMID: 28639618; PubMed Central PMCID: PMC5489682.
32. Peloponese JM Jr., Jeang KT. Role for Akt/protein kinase B and activator protein-1 in cellular proliferation induced by the human T-cell leukemia virus type 1 tax oncoprotein. *J Biol Chem*. 2006; 281(13):8927–38. <https://doi.org/10.1074/jbc.M510598200> PMID: 16436385.
33. Thorpe LM, Yuzugullu H, Zhao JJ. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat Rev Cancer*. 2015; 15(1):7–24. <https://doi.org/10.1038/nrc3860> PMID: 25533673; PubMed Central PMCID: PMC34384662.
34. Kiyono T, Hiraiwa A, Fujita M, Hayashi Y, Akiyama T, Ishibashi M. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proc Natl Acad Sci U S A*. 1997; 94(21):11612–6. PMID: 9326658; PubMed Central PMCID: PMC23554.
35. Ganti K, Broniarczyk J, Manoubi W, Massimi P, Mittal S, Pim D, et al. The Human Papillomavirus E6 PDZ Binding Motif: From Life Cycle to Malignancy. *Viruses*. 2015; 7(7):3530–51. <https://doi.org/10.3390/v7072785> PMID: 26147797; PubMed Central PMCID: PMC4517114.
36. Ivarsson Y, Arnold R, McLaughlin M, Nim S, Joshi R, Ray D, et al. Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc Natl Acad Sci U S A*. 2014; 111(7):2542–7. <https://doi.org/10.1073/pnas.1312296111> PMID: 24550280; PubMed Central PMCID: PMC3932933.
37. Kimata JT, Wong FH, Wang JJ, Ratner L. Construction and characterization of infectious human T-cell leukemia virus type 1 molecular clones. *Virology*. 1994; 204(2):656–64. <https://doi.org/10.1006/viro.1994.1581> PMID: 7941334.
38. Shultz LD, Saito Y, Najima Y, Tanaka S, Ochi T, Tomizawa M, et al. Generation of functional human T-cell subsets with HLA-restricted immune responses in HLA class I expressing NOD/SCID/IL2r gamma (null) humanized mice. *Proc Natl Acad Sci U S A*. 2010; 107(29):13022–7. <https://doi.org/10.1073/pnas.1000475107> PMID: 20615947; PubMed Central PMCID: PMC2919921.
39. Artesi M, Marçais A, Durkin K, Rosewick N, Hahaut V, Suarez F, et al. Monitoring molecular response in adult T-cell leukemia by high-throughput sequencing analysis of HTLV-1 clonality. *Leukemia*. 2017; 31(11):2532–5. <https://doi.org/10.1038/leu.2017.260> PMID: 28811663; PubMed Central PMCID: PMC5668493.
40. Rosewick N, Durkin K, Artesi M, Marçais A, Hahaut V, Griebel P, et al. Cis-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. *Nat Commun*. 2017; 8:15264. <https://doi.org/10.1038/ncomms15264> PMID: 28534499.
41. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014; 9(1):171–81. <https://doi.org/10.1038/nprot.2014.006> PMID: 24385147.
42. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174; PubMed Central PMCID: PMC2690996.
43. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013; 14(4):R36.

<https://doi.org/10.1186/gb-2013-14-4-r36> PMID: [23618408](https://pubmed.ncbi.nlm.nih.gov/23618408/); PubMed Central PMCID: [PMC4053844](https://pubmed.ncbi.nlm.nih.gov/PMC4053844/).

44. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638> PMID: [25260700](https://pubmed.ncbi.nlm.nih.gov/25260700/); PubMed Central PMCID: [PMC4287950](https://pubmed.ncbi.nlm.nih.gov/PMC4287950/).
45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/); PubMed Central PMCID: [PMC4302049](https://pubmed.ncbi.nlm.nih.gov/PMC4302049/).

## Abstract

The dynamic regulation of the protein synthesis process participates in the cell adaptation to a constantly evolving environment. Despite its critical role in gene expression regulation, the understanding of translational control in fundamental biological processes, such as immune responses, is still incomplete. The implementation of new approaches based on deep sequencing can be used to fill the gap in the knowledge of protein synthesis regulation. Notably, monosome vs polysome footprinting is an innovative approach derived from ribosome profiling that allow the characterization of 80S footprints derived either from monosomes or polysomes associated ribosomes. In this work, I identified the key parameters required to obtain a robust picture of ribosomal densities across cellular mRNAs using monosome vs polysome footprinting in murine primary bone-marrow derived macrophages (pBMDM). These immune cells are particularly interesting to study protein synthesis regulation in evolving conditions as they display a high sensitivity towards their environment and have the ability to trigger different gene expression programs depending on external cues. Their high phenotypic plasticity is in fact essential to ensure their protective functions in the organism such as the triggering and the resolution of the inflammatory response. As monosome vs polysome footprinting was initially developed in yeast, the adaptation of this method to study murine immune cells required extensive optimizations. The resulting protocol developed in this work was used to confirm that, contrary to a long lasting belief in the scientific community, murine pBMDM monosomes are actively involved in the translation process. Interestingly, we were able to recapitulate similar observations to what was previously observed in yeast regarding the features of mRNAs preferentially bound to monosomes or polysomes in murine pBMDM. This could suggest that the differential trafficking of ribosomes depending on specific features of the cellular mRNAs is a conserved mechanism of translational control. Importantly, the distribution of ribosomes across the different mRNAs is not random and the proper ribosome allocation pattern could be critical to adapt protein synthesis levels to the cellular needs. Here we developed a robust strategy to study this overlooked transcript-specific mechanism of translational control. Moreover, our optimized protocol can now be used to study the impact of translation through monosomes or polysomes at different stages of the inflammatory response in murine macrophages.

Key-words : Inflammation, Macrophage, Translation, Ribosome profiling, Monosome vs Polysome footprinting

## Résumé

La régulation dynamique de la synthèse des protéines en fonction des besoins de la cellule facilite son adaptation face aux fluctuations de l'environnement. Malgré l'importance de la régulation de la traduction au cours du processus d'expression des gènes, l'impact de ce mécanisme sur des processus biologiques fondamentaux, comme la mise en place d'une réponse immunitaire, reste mal compris. Grâce au développement de nouvelles technologies basées sur l'utilisation du séquençage à haut débit, comme le *ribosome profiling*, il est désormais possible d'étudier en détails la façon dont la synthèse des protéines est contrôlée. Le *monosome vs polysome footprinting* est une nouvelle méthode qui permet d'étudier la traduction des ARN messagers (ARNm) selon leur association avec un seul ribosome (monosome) ou avec plusieurs ribosomes (polysomes). Au cours de ma thèse, j'ai identifié les paramètres essentiels pour la mise en place d'une expérience de *monosome vs polysome footprinting* donnant des résultats fiables en utilisant des macrophages primaires dérivés de la moëlle osseuse de souris. Je me suis intéressée à ce type de cellules immunitaires particulier car elles présentent une grande capacité à détecter des modifications dans leur environnement et à modifier leur taux d'expression de protéines en fonction des signaux reçus. Leur grande plasticité est notamment essentielle pour assurer leurs diverses fonctions de protection de l'organisme, comme le déclenchement et la résolution de la réponse inflammatoire. La méthode de *monosome vs polysome footprinting* ayant été initialement développée chez la levure, son utilisation avec un modèle d'étude différent a nécessité de nombreuses modifications du protocole. Suite à cette phase de développement technologique, j'ai pu confirmer que les monosomes, une population de ribosomes historiquement considérés comme inactifs, sont activement impliqués dans le processus de traduction dans les macrophages primaires de souris. Les données obtenues ont également permis d'identifier des caractéristiques communes entre les ARNm enrichis dans les monosomes chez la levure et dans les macrophages murins. La régulation de la synthèse des protéines via l'association à des monosomes ou à des polysomes pourrait donc être un mécanisme conservé chez les organismes eucaryotes. Enfin, le travail d'optimisation réalisé dans les macrophages primaires murins ouvre la possibilité d'étudier l'effet de la régulation de la traduction sur la mise en place et la résolution de la réponse inflammatoire de façon très détaillée.

Mots-clés: Inflammation, Macrophage, Traduction, Ribosome profiling, Monosome vs Polysome footprinting