DETERMINATION OF PROTEIN LOCALIZATION AND RNA KINETICS IN HUMAN CELLS

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

7 September 2022

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Roberto Arsiè

Kommissarischer Präsident der
Humboldt-Universität zu Berlin:
Prof. Dr. Peter Frensch

Dekan der Lebenswissenschaftlichen
Fakultät: Prof. Dr. Dr. Christian Ulrichs

Gutachter der Dissertation:

Prof. Dr. Markus Landthaler
Dr. Benedikt Beckmann
Dr. Tuğçe Aktaş

TABLE OF CONTENTS

STATEMENT OF CONTRIBUTIONS

The projects presented in this thesis were possible thanks to collaborations with many different people and laboratories. Prof. Dr. Markus Landthaler supervised the development of all the projects. Dr. Igor Minia co-developed with me the method to identify RNA-binding proteins in different subcellular compartment (f-XRNAX) and the SLAM-Drop-seq approach. The bioinformatical analysis for the SLAM-Drop-seq project was carried out by Haiyue Liu, with great help from Marcel Schilling and Jonathan Alles supervised by Prof. Dr. Nikolaus Rajewsky, and by Dr. Daniel Schwabe supervised by Prof. Dr. Martin Falcke. The mass spectrometry and MaxQuant runs for the different BirA* cell lines were performed by Dr. Guido Mastrobuoni from the laboratory of Dr. Stefan Kempa (MDC). The mass spectrometry and MaxQuant runs for the f-XRNAX project were performed by Dr. Marieluise Kirchner from the laboratory of Dr. Philipp Mertins. Ouidad Benlasfer (Landthaler laboratory) helped with the creation of the different BirA* HEK293 stable cell lines.

SUMMARY

The intracellular distribution of a specific protein is essential for its function. Proper protein localization is therefore an essential biological process, regulated at many levels and central for eukaryotic compartments organization. Consequently, several methods have been established to globally profile proteomes at different subcellular regions. Many of these approaches rely on the different biochemical features of eukaryotic compartments to separate them and analyse their relative protein composition in parallel. In this study we assessed the protein distribution of HEK293 cells using BirA* proximity labelling activity and restricting its localization at cellular regions difficult to purified with traditional methods: the cytosol-facing sides of the endoplasmic reticulum, mitochondria, and plasma membranes. We obtained high-quality datasets for each of the compartment examined, both recapitulating the protein distribution known from literature and discovering new candidates at the investigated subcellular regions. Interestingly, we also identified proteins localizing at more than one subcellular compartment, providing useful insights on the protein composition shared by different organelles.

Proteins with different biological activity were captured by our proximity labelling approach, such as metabolic enzymes and integral receptors. RNA-binding proteins (RBPs) were also found in our datasets, but the direct interactions between them and RNA molecules could not be validated using only BirA* biotinylation. To obtain an activity-based RBP profiling and assess the composition of protein binding to RNA in different cellular compartments, we developed the f-XRNAX protocol. It consists of UV-crosslinking, cellular fractionation and organic-phase enrichment of RBPs coupled with mass spectrometry. We recovered background-corrected proteomes for nuclei, cytoplasm and membranes of HEK293 cells. Protein domain analysis of peptides crosslinked to ribonucleotides showed an enrichment in known RNA-binding domains. We also determined the protein domains where the RNA-protein contacts occurred *in vivo* using enriched peptides profiles and reported some cases of differential domain usage of the same protein in two different compartments. Surprisingly, many non-canonical RBPs were identified in the membrane fraction, and their peptide profiles were enriched in regions with high density of intrinsically disordered regions, indicating a possibly weak interaction with RNA mediated by these non-structural motives. Lastly, we provided evidence of the differential binding to RNA of the same protein in different HEK293 compartments.

In the second part of this thesis, we focused on the determination and quantification of newly transcribed RNAs at the single-cell level. The kinetics of RNA transcription, processing and degradation were until recently not measurable at the single-cell level. Thus, we have developed a novel approach (called SLAM-Drop-seq) by adapting the published SLAM-seq method to single cells. We used SLAM-Drop-seq to estimate time-dependent RNA kinetics rates of transcription and turnover for hundreds

of oscillating transcripts during the cell cycle of HEK293 cells. We found that genes regulate their expression with different strategies and have specific modes to fine-tune their kinetic rates along the cell cycle.

ZUSAMMENFASSUNG

Die intrazelluläre Verteilung eines bestimmten Proteins ist für seine Funktion von wesentlicher Bedeutung. Die korrekte Proteinlokalisierung ist daher ein wesentlicher biologischer Prozess, der auf vielen Ebenen reguliert wird und für die Organisation eukaryontischer Kompartimente von zentraler Bedeutung ist. Infolgedessen wurden mehrere Methoden entwickelt, um ein globales Profil der Proteome in verschiedenen subzellulären Regionen zu erstellen. Viele dieser Ansätze beruhen auf den unterschiedlichen biochemischen Merkmalen eukaryontischer Kompartimente, um diese zu trennen und ihre relative Proteinzusammensetzung parallel zu analysieren. In dieser Studie untersuchten wir die Proteinverteilung in HEK293-Zellen mit Hilfe der BirA*-Näherungsmarkierungsaktivität und beschränkten ihre Lokalisierung auf zelluläre Regionen, die mit herkömmlichen Methoden nur schwer zu reinigen sind: die dem Zytosol zugewandten Seiten des endoplasmatischen Retikulums, Mitochondrien und Plasmamembranen. Wir erhielten für jedes der untersuchten Kompartimente qualitativ hochwertige Datensätze, die sowohl die aus der Literatur bekannte Proteinverteilung rekapitulierten als auch neue Kandidaten in den untersuchten subzellulären Regionen entdeckten. Interessanterweise haben wir auch Proteine identifiziert, die in mehr als einem subzellulären Kompartiment lokalisiert sind, was nützliche Einblicke in die gemeinsame Proteinzusammensetzung der verschiedenen Organellen liefert.

Proteine mit unterschiedlicher biologischer Aktivität wurden durch unseren Proximity-Labelling-Ansatz erfasst, wie z. B. Stoffwechselenzyme und integrale Rezeptoren. Auch RNA-bindende Proteine (RBPs) wurden in unseren Datensätzen gefunden, aber die direkten Wechselwirkungen zwischen ihnen und RNA-Molekülen konnten nicht allein durch BirA*-Biotinylierung validiert werden. Um ein aktivitätsbasiertes RBP-Profil zu erstellen und die Zusammensetzung der Proteinbindung an RNA in verschiedenen zellulären Kompartimenten zu bewerten, haben wir das f-XRNAX-Protokoll entwickelt. Es besteht aus UV-Vernetzung, zellulärer Fraktionierung und Anreicherung der RBPs in der organischen Phase in Verbindung mit Massenspektrometrie. Wir gewannen hintergrundkorrigierte Proteome für Kerne, Zytoplasma und Membranen von HEK293-Zellen. Die Analyse der Proteindomänen von Peptiden, die mit Ribonukleotiden vernetzt sind, zeigte eine Anreicherung in bekannten RNA-bindenden Domänen. Wir bestimmten auch die Proteindomänen, in denen die RNA-Protein-Kontakte in vivo stattfanden, indem wir Profile angereicherter Peptide verwendeten, und berichteten über einige Fälle unterschiedlicher Domänenverwendung desselben Proteins in zwei verschiedenen

Kompartimenten. Überraschenderweise wurden viele nicht-kanonische RBPs in der Membranfraktion identifiziert, und ihre Peptidprofile waren in Regionen mit hoher Dichte an intrinsisch ungeordneten Regionen angereichert, was auf eine möglicherweise schwache Interaktion mit RNA hinweist, die durch diese nicht-strukturellen Motive vermittelt wird. Schließlich konnten wir die unterschiedliche Bindung desselben Proteins an RNA in verschiedenen HEK293-Kompartimenten nachweisen.

Im zweiten Teil dieser Arbeit konzentrierten wir uns auf die Bestimmung und Quantifizierung von neu transkribierten RNAs auf Einzelzellebene. Die Kinetik der RNA-Transkription, -Prozessierung und -Degradation war bis vor kurzem auf Einzelzellebene nicht messbar. Daher haben wir einen neuen Ansatz (SLAM-Drop-seq genannt) entwickelt, indem wir die veröffentlichte SLAM-seq-Methode an Einzelzellen angepasst haben. Wir haben SLAM-Drop-seq verwendet, um die zeitabhängigen RNA-Kinetikraten der Transkription und des Umsatzes für Hunderte von oszillierenden Transkripten während des Zellzyklus von HEK293-Zellen zu schätzen. Wir fanden heraus, dass Gene ihre Expression mit unterschiedlichen Strategien regulieren und spezifische Modi zur Feinabstimmung ihrer kinetischen Raten entlang des Zellzyklus haben.

INTRODUCTION

Protein localization in eukaryotic cells

Cells are dynamic and complex entities. To properly maintain cellular homeostasis, these seemingly chaotic and crowded environments need to be highly organized and coordinated. One way to regulate different types of activities is to assign specific tasks to diverse spatial locations. Like in human corporation system, where different teams have separate areas in a building to work on precise assignments, cells are composed of separate intracellular compartments, each of which having its set of specialized roles. For example, in eukaryotic cells the endoplasmic reticulum (ER) is the cellular central storage for calcium ions (Baumann and Walz, 2001) and the place where synthesis and glycosylation of secreted and membrane-attached proteins occurs (Reily et al., 2019). In the nucleus, the genome is stored and RNA is transcribed and processed. Mitochondria, on the other hand, produce ATP molecules and have the potential to activate the apoptosis intrinsic pathway if stressed (Spinelli and Haigis, 2018). Most of these compartments are physically separated between each other with lipid membranes bilayers, while other are in cellular continuity with the surrounding milieu and are formed by large protein complexes (e.g. centrosomes) or condensates formed by phase separation of their constituents (e.g. stress granules) (Helder et al., 2016). Membrane-bound organelles offer different chemical properties: the cytosol is a reducing environment, while within the ER is an oxidative one (Hwang et al., 1992). Functionally different organelles and compartments are one of the hallmarks of eukaryotic cells and ensure partition of biological processes.

The described membrane separation does not only define distinct physical spaces, but also allow for flexibility and integration of responses to stimuli. In fact, these different compartments and cellular functions should not be seen as independent or isolated entities inside the eukaryotic cell, but as modular and integrative networks. For example, a signalling pathway may start with an extracellular stimulus, delivered to the intracellular space by conformational changes of a specific receptor integrated in the plasma membrane, which leads to cytosolic protein modifications, rearranging their structure and promoting their re-localization to another subcellular compartment. Membrane-enclosed organelles communication is central for responding to stimuli and maintaining cellular homeostasis. Physical contacts between them remodel their structure and induces transfer of metabolites and propagation of signals (Cohen et al., 2018).

Most of the functions at these specialized cellular compartments is carried out by proteins. Thus, the presence or absence of proteins at a restricted area of the cell enables the acquisition and implementation of specific cellular activities. One way to regulate the presence of proteins is to regulate them at the level of abundance, making or degrading more of them, thus increasing or

decreasing their local concentration. However, this mechanism is time and energy consuming, since *de novo* synthesis of polypeptides involves all the steps of gene expression regulation and it cannot be implemented when fast cellular responses are needed. A much more energy-efficient way to regulate cellular function is to selectively localize proteins at the cellular compartment of interest with specific targeting peptides and modulate the protein activity by means of post-translational modifications or substrate binding. This allows for a fast and effective response to changes in environmental circumstances, like stress conditions or extracellular signalling. In order to do so, each protein must contain a certain kind of information in their polypeptide sequence which allows for its correct cellular localization. Another way to obtain subcellular protein localization and rapid response to stimuli is to halt RNA translation and localize the transcript where the protein needs to be produced (Das et al., 2021; Engel et al., 2020). These mechanisms and cellular controls defining RNA localization are not subject of discussion here.

The various cellular compartments adopt different mechanisms to preferentially localize polypeptides of interest. Protein translation usually starts in the cytosol, with the first amino acids emerging from the ribosome determining the destination of a translated protein. In fact, if the first 16-30 codons translate a sequence that is recognized by the Signal Recognition Particle (SRP) protein (Zimmermann, 2009), translation is halted and the complex of mRNA-ribosome-SRP is relocated to the endoplasmic reticulum membrane (ERM), where translation continues and the newly synthetized protein is "threaded" in the lumen of the endoplasmic reticulum (ER). In this case, the primary regulatory targeting information resides in the N-terminal amino acids of the protein, but once in the ER lumen or integrated in the ER membrane other factors will determine the further localization of the protein. For example, two well known signals for retention in the ER of luminal and membrane proteins are coded close to their C-terminus, namely the KDEL (Munro and Pelham, 1987) and KXXX (Jackson et al., 1990) target peptides. The nuclear localization signal (NLS) is instead composed by stretches of basic amino acids and it is recognized by importin proteins which bind to the NLS-containing protein and transport it into the nucleus (Kosugi et al., 2009). Conversely, proteins containing nuclear export signal (NES) are excluded from the nucleus by active transportation to the cytosol (Fischer et al., 1995; Stewart, 2019). Finally, a 15-70 amino acids long alpha helix with positively charged recognized by the TOM complex is what determines the restricted localization of a protein in the mitochondria (Bolender et al., 2008), while a variety of sequences allow for protein lipidation and a subsequent targeting to the plasma membrane (Chen et al., 2018).

The described targeting peptides can however be quite variable, and many more signals exist for preferential localizations of a protein to other compartments (lysosomes, Golgi, peroxisomes, extracellular environment). Bioinformatic tools are of great help for the identification of sequence motifs coding for targeting peptides (Fukasawa et al., 2015; Savojardo et al., 2018) and the analysis of

localized proteins (Gatto et al., 2014; Jiang et al., 2021). Nonetheless, these computational methods have limitations: they are not accurate, for example, in assigning the correct localization to polypeptides found in more than one compartment and in quantitatively estimating their abundance (Jiang et al., 2021). These tools are also inefficient to build a strong predictive model for proteomes in sub-organellar regions with little information available and are unable to discern the distribution of polypeptides affected by a mutations (Nielsen, 2017). Thus, *in vivo* biochemical validation of protein localization must always be performed to confirm bioinformatical predictions. Moreover, proteins deployed of targeting signals can also be temporary relocated to a specific cellular region (through phosphorylation, for example) and perform there a precise function. Therefore, different strategies and methods should be applied to map protein localization in eukaryotic cells and their dynamics upon stimuli or changes of environmental conditions.

The importance of an accurate and complete catalogue of protein localization stems also from the fact that disruption or impairment in protein trafficking is a leading cause for various diseases (Bridges and Bradbury, 2018; Cheung and Ip, 2012; Guardia et al., 2018). The mislocalization of proteins is implicated for example in cystic fibrosis (Cheng et al., 1990), amyotrophic lateral fibrosis (Guo and Shorter, 2017) and pulmonary atrial hypertension (Sehgal and Lee, 2011). An accurate understanding of protein localization in health and disease can therefore guide therapeutic strategies.

Approaches to study protein compartmentalization

Different methods have been developed to study the localization of proteins in living or fixed cells. Immunofluorescence (IF) is a widely used approach to detect the cellular expression and distribution of a particular protein of interest in fixed cells or tissues. It is straight-forward, inexpensive and works at individual cell resolution. Downsides of IF include the limited number of proteins traceable per experiment, the requirement of cell fixation and permeabilization which introduces artifacts in cellular morphology (Schnell et al., 2012), and the necessity to rely on antibodies with varying affinities and specificities depending on the sampled protein(s). Still, thousands of independent academic works have used IF to study the cellular distribution of proteins of interest. Even more impressive, an international scientific collaboration used immunofluorescence to create the Human Protein Atlas, a database with subcellular information for 12003 proteins in human cell lines (Thul et al., 2017). The online and downloadable data provide an excellent source of information for a great part of human proteome in specific cell lines (Cho et al., 2022).

The use of fluorescent proteins (like Green or Red Fluorescent Protein, GFP or RFP) fused by genetic manipulation to a gene of interest has also been broadly used to map the protein localization *in vivo* in different biological systems (Huh et al., 2003; Leonetti et al., 2016), without the issue of fixation and

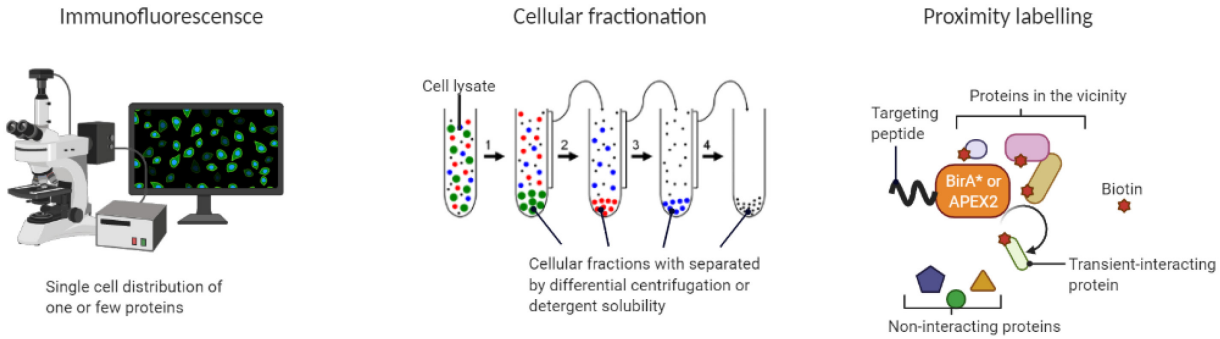

Fig. I Three widely used approaches to study protein localization. (Image created in BioRender.com)

with a single cell resolution. Nonetheless, this approach requires genetic manipulation of the cells and the exogenously introduced fluorescent polypeptides may influence the behaviour and distribution of the investigated protein. Moreover, the number of different proteins analysed per experiment is rather low.

Instead of focusing on few proteins, different approaches have been developed in the last decades to retrieve entire proteomes localized at specific subcellular regions. These methods do not rely on microscopy and results are obtained analysing mass spectrometry data, *i.e.* averaging the information among many cells. To obtain more extensive and unbiased information about the compartment-specific protein repertoire within one experiment, the two main approaches are cellular fractionation and proximity labelling (PL). Cell fractionation relies on the distinct biochemical properties of the different cellular compartments and is based on sequential detergent extractions (Baghirova et al., 2015), or gradient density centrifugation, where a differential centrifugation is applied to homogenized cells in a continuous or discontinuous gradient (usually of sucrose), allowing the separation of organelles and sub-cellular particles based on their density (Duve, 1975). Fractionation approaches are generally inexpensive, reproducible and with no requirements of genetically manipulate the samples to be analysed. However, the similarity of biochemical composition between compartments does not always allow to obtain pure fractions and damage can lead to artefacts (like the microsomes formed by fragmented endoplasmic reticulum), which reduces the specificity of cell fractionation-based approaches.

A more recent approach is proximity labelling (PL), which takes advantage of enzymes attaching biotin moieties to proteins in their vicinity, followed by selective avidin/streptavidin purifications to recover the modified proteins (Sunbul and Jäschke, 2019). The two most frequently used enzymes for PL are BirA* and APEX, as well as variations thereof. BirA is a biotin-ligase enzyme found in *E. coli* and mutated ($BirA^{R118G}$ = BirA*) to render its biotinylation activity not substrate-specific, *i.e.* promiscuously biotinylating all proteins in its vicinity (Roux et al., 2012). Ascorbate peroxidase (APEX) (Rhee et al., 2013) and its improved version APEX2 (Lam et al., 2015) are engineered peroxidases developed in the laboratory of Alice Ting (Stanford University) relying on H_2O_2-driven creation of biotin-phenol radicals, which are very reactive and conjugate with nearby proteins. The biotinylation induced by APEX is fast

(30-60 seconds), but it has the downsize to require a stress-inducing agent (H_2O_2) and to produce radicals able to diffuse from the location where they are produced. On the other hand, BirA* directly biotinylates proteins in its vicinity, but with a much slower kinetic (around 16-18 hours). In order to solve this problem, a faster version of BirA* was developed called TurboID, which is able to induce a fast biotinylation (5-10 minutes) and can also be applied *in vivo* (Branon et al., 2018). In addition, APEX2 biotinylates ribonucleic acids in its vicinity, and this property was used to provide subcellular distribution maps of transcripts (Fazal et al., 2019).

The described proximity labelling approaches have been applied in recent years to investigate proteomes at different cellular compartments, from membrane enclosed environments like mitochondria, ER and nuclei (Hung et al., 2017; Rhee et al., 2013), to subcellular regions without a lipid bilayer surrounding them like the mitochondrial nucleoid, the stress granules and the synaptic cleft (Han et al., 2017; Loh et al., 2016; Markmiller et al., 2018).

Still not fully characterized is the composition of proteomes located at compartment interfaces (in contrast to the compartment themselves), like for example the cytosol-facing surfaces of ER and mitochondria, which cannot be enriched by fractionation approaches. These subcellular regions are in fact engaged in a wide variety of cellular functions, from lipid biosynthesis, protein translation and secretion to cell death and calcium ion regulation. Obtaining a high confidence protein map of these cellular locations could shed more light on the functional units regulating the processes taking place there and identify new candidates with unknown or unspecific cellular localization.

Proximity labelling with APEX2 revealed the protein composition at these specialized compartments (the cytosol-facing membranes of ER and mitochondria) and validated the molecular function for some of the proteins at the contact sites between them (Hung et al., 2017). In order to do so, APEX2 was fused to a peptides which targeted its trafficking to the cytosol facing membranes of the ER and the mitochondria. Thanks to a carefully designed, SILAC-based mass spectrometry experiment, the authors identified proteins at the cytosolic surfaces of these membranous compartments and also the one which co-localized between the ER and mitochondria. Some of these proteins did not have any previous annotation to the compartment they were found: the authors found 22 mitochondria and 72 ER ‘orphans’, one of which (C2 Domain Containing 3 Centriole Elongation Regulator, C2CD3) was validated by immunofluorescence (Hung et al., 2017).

This work provided the first evidence of the protein composition at subcellular regions difficult to annotate with standard approaches. Although informative, the true protein content might be underestimated or misidentified. The fold change thresholds to assign a protein to one compartment respect to background were as low as 0.08. The short biotinylation time with APEX2 might be a cause for the low enrichment of proteins in these experimental conditions. A longer biotin labelling of nearby proteins might enhance the efficiency of resident proteins recovery.

Furthermore, the proteome mapping of other subcellular compartment difficult to purify would be informative, like for example the cytosol-facing side of the plasma membrane. This region is involved in major roles for cellular integrity and response to stimuli, since most of receptors and proteins for extracellular contacts reside here (Serna et al., 2016). Although much is known about their composition and interactions, a systematic approach with a focus on the proteome in near contact to the cytosolic surface of the human plasma membrane has never been applied.

RNA-binding protein function & localization

RNA-binding proteins (RBPs) have an essential role in living cells and are evolutionary well conserved in eukaryotic organisms. They are the main actors for RNA metabolism, regulating the processing, nuclear export, localization, translation, stability and degradation of virtually all transcripts produced in a cell. Proteins and transcripts together form ribonucleoprotein particles (RNP), which finally coordinate the output of gene expression, *i.e.* regulating cellular homeostasis and responses to stimuli (Singh et al., 2015). Therefore, the study and characterization of proteins interacting with RNAs is fundamental to understand the RNA regulatory processes and the modulation of cell activity.

The first high-throughput studies to identify the RNA-binding proteome (RBPome) were based on screens with immobilized RNAs or proteins (Butter et al., 2009; Tsvetanova et al., 2010). However,

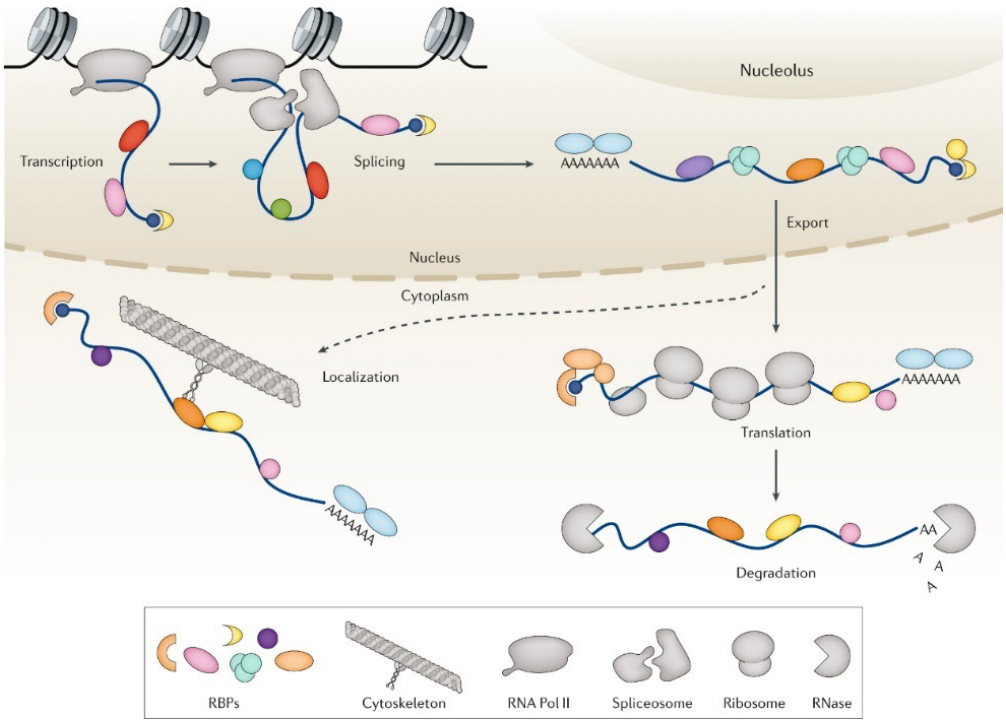

Fig. II RNA-binding proteins are involved in every step in the life cycle of mRNA transcripts. RBPs fine tune the splicing, processing, export, localization, translation and degradation of virtually every RNA in human cells. [Image from (Gebauer et al., 2021)]

these approaches were applied *in vitro*, lacking the full complexity of the cellular environment and requiring more experiments to validate the results. In the last decade, several high-throughput approaches have been developed to identify the set of proteins binding RNA *in vivo.* They all rely on the ultraviolet light (UV) property to form covalent bonds between RNA nucleotides and proteins found in close proximity to them, around 2 Å (Brimacombe et al., 1988; Burger et al., 2013). The so-called RNA interactome capture (RIC) protocols were the first established for the unbiased detection of RBPs and consist of UV crosslinking between RNA and contacting proteins followed by purification with oligo(dT)-coated beads (Baltz et al., 2012; Castello et al., 2012). The enriched proteins are then digested to peptides and analysed with mass spectrometry. These ground-breaking approaches showed that the amount of annotated RBPs had been massively underestimated, with many novel RBPs lacking any known RNA-binding domain (RBD) or roles associated to RNA biology (Beckmann et al., 2016; Hentze et al., 2018). RIC was applied to many different model systems and conditions (Beckmann et al., 2015; Kwon et al., 2013; Reichel et al., 2016; Sysoev et al., 2016; Wessels et al., 2016) and was further optimized by using locked nucleic acids as probes (Perez-Perri et al., 2018). Nonetheless, a main disadvantage of these techniques is the bias towards polyadenylated transcripts, overlooking the binding of proteins to RNA lacking a poly(A) tail, such as nascent pre-mRNAs, rRNAs, tRNAs and ncRNAs.

To address this issue, two methods were concurrently developed, named CARIC and RICK (Bao et al., 2018; Huang et al., 2018), in which RNA is labelled with 5-EU (and 4sU in CARIC), cells are then irradiated with UV light and enrichment of labelled transcripts is achieved via click chemistry. These approaches led to the identification of many proteins bound to transcripts without a poly(A) bias. Because of the use of synthetic nucleotides and their reported toxicity over long period of incubation with cells (Burger et al., 2013), these protocols are more suited to study the RBPome restricted to nascent RNA. Furthermore, the incorporation of ribonucleotide analogues in animal model is challenging, limiting the systems where these approaches can be applied.

In 2019, three groups independently published similar methods, named PTex (Urdaneta et al., 2019), XRNAX (Trendel et al., 2019) and OOPS (Queiroz et al., 2019), tackling the described challenges. In all studies, the authors took advantage of the physical-chemical properties of UV crosslinked protein-RNA complexes and organic phase separation. This resulted in the identification of RBPomes without synthetic nucleotides labelling and with no bias towards polyadenylated transcripts, while reducing the amount of input material needed per experiment. In all three publications, the number of detected proteins interacting with RNA was larger than the number of annotated RBPs in human, highlighting once again how the composition of human RBPome has been so far underestimated. It has been showed that many of these unconventional RBPs have roles in the intermediate metabolic pathways

and for some of these proteins their moonlighting activity as RBP has been explored in more details. Although, the function and role for the majority of them remains still to be explored.

One poorly studied aspect of RBPs functional characterization is their comprehensive localization in cellular compartments adopting high-throughput techniques. Nonetheless, RNA-binding protein compartmentalization is of central importance during many biological processes requiring RNA molecules localization at a specific area of the cell. The localization of transcripts is more cost-effective from an energetical point of view than the transport of proteins, and is a cellular mechanism conserved from bacteria to multi-cellular organisms (Das et al., 2021).
Recently, two approaches have been independently developed to bridge this gap of knowledge and, by coupling proximity labelling with APEX or cellular fractionation with organic phase separation, they were able to map the relative abundance of human proteins binding to RNA in different subcellular regions (Qin et al., 2021; Yan et al., 2021). This resulted in not only the identification of new proteins with RNA-binding potential, but also in their characterization in a spatio-temporal fashion.
These findings show the importance of combining different methods to study proteins actively binding to RNAs and to determine their relative cellular distribution. How the protein-RNA affinities change between compartments and which RBDs mediate these interactions are still unanswered questions.

Approaches to study temporal RNA dynamics

As previously described, the primary role of RBPs is to regulate the post-transcriptional life of RNA molecules by influencing their processing, localization, translation and decay. This is of crucial importance because the cellular transcriptome represents a central output of gene expression and its post-transcriptional regulation shapes the fate and function of a cell. All the mentioned RNA-linked processes regulated by RBPs are not statically determined for every cell, but have kinetics associated to them, which can be dynamically modulated by the cellular machinery.
The homoeostasis of cells and their responses to intrinsic and extrinsic signals rely on the regulation of gene expression. Therefore, the global measurement of transcripts abundance is central to understand the biology of any cellular system. Even if sometimes the amount of RNA molecules measured in a sample is used as a proxy for transcriptional activity, we now know that its regulation is driven by the complex interplay of RNA synthesis, processing and turnover. The cellular RNA levels are dynamically fine-tuned according to the environmental conditions, and their proper regulation is central for an appropriate response to stimuli (Reilly and Noonan, 2014).

The total amount of a specific transcript can be post-transcriptionally regulated by fine tuning its synthesis and decay rates as well as splicing. The kinetics related to these processes are also associated to dynamic responses according to the cellular state and environmental conditions. One central aim for cell biology is to better understand the kinetic rates associated to RNA regulation, and to have quantitative information about how fast a transcript is synthetized or degraded and how changes in environmental conditions lead to the fine-tuning of these parameters. In fact, understanding how gene regulation is coordinated at the cellular level has the potential to unravel how cells maintain homeostasis and drive differentiation, and how aberrant regulation leads to diseases (Lee and Young, 2013; de-Leon and Davidson, 2007).

The well established and widely used bulk RNA-sequencing (RNA-seq) experiments are very informative regarding the steady state level of thousands of genes in a cell population, but they cannot determine the processes leading to the accumulation or depletion of transcripts. Several independent approaches have been developed in the last 20 years to measure global changes of RNA synthesis and decay, focusing on the dynamics between different cell states. Inhibition of transcription with different compounds (alpha-amanitin, actinomycin D, flavopiridol or DRB) coupled to microarray or RNA-seq has been the first genome-wide method to measure RNA decay rates (Chen et al., 2015; Lam et al., 2001; Raghavan et al., 2002; Yang et al., 2003). Although informative, these approaches are highly invasive, impairing cellular physiology and the processes controlling transcripts stability and localization (Bhattacharyya et al., 2006; Shyu et al., 1989; Tani et al., 2012). Furthermore, the global block of transcription does not stop precursor RNA molecules processing, particularly interfering with the half-life measurements of transcripts with a slow splicing rate.

Enriching transcripts associated with chromatin or active RNA polymerase II accurately identified the *de novo* synthesized RNAs and revealed new features of gene expression (Core et al., 2008; Kwak et al., 2013; Mayer and Churchman, 2016), but with little quantitative information. In fact, these approaches cannot disentangle the presence of transcripts or the velocity of RNA polymerase II from the efficiency of RNA synthesis, lacking the information to measure the transcriptional output (Furlan et al., 2020a). These protocols could also be influenced by binding of non-specific RNAs or proteins. Moreover, using these approaches does not reveal any information regarding transcript stability.

The first quantitative, transcriptome-wide informative data about RNA kinetics were achieved by RNA metabolic labelling. In these approaches, cells are incubated with nucleotides analogue (the most frequently used being 4-thiouridine – 4sU, 5'-bromouridine – BrU, 5-ethinyluridine – 5EU), which are incorporated into newly synthetized RNAs. The labelled transcripts are then biochemically separated from the unlabelled ones, resulting in two RNA populations: 'total and 'new'. These two pools of transcripts are then treated and analysed independently. Using different experimental designs (single-pulse, pulse or pulse-chase experiments (Uvarovskii et al., 2019)) and times of incubation, the

sequencing of labelled and unlabelled transcripts enables a reliable estimation of transcription and/or decay rates genome-wide (Dölken et al., 2008; Miller et al., 2011; Rabani et al., 2011; Schwalb et al., 2016; Schwanhäusser et al., 2011). Noteworthy, these methods are not only used to study RNA kinetics in cell culture under steady-state conditions, but they can also be applied to different cellular systems and dynamic biological processes, as for example zebrafish development or cellular stimulation with inflammatory molecules (Rabani et al., 2011, 2014).

All the described approaches are relatively laborious and require biochemical separation, thus being very sensitive to the accuracy of labelled molecules enrichment. They also require relatively high amount of starting material and might contain up to 30% unlabelled RNA contaminants in the labelled fraction (Furlan et al., 2020b). To circumvent these limitations, new metabolic labelling protocols were recently developed. SLAM-seq (Herzog et al., 2017), TimeLapse-seq (Schofield et al., 2018) and TUC-seq (Riml et al., 2017) are all based on *in vivo* labelling of RNAs with 4sU and biochemical *in vitro* conversion of the thiolated nucleotides into a cytosine-analogue using different compounds (iodoacetamide, TFEA/$NaIO_4$ or OsO_4/NH_4Cl, respectively). This treatment modifies the pairing property and leads to an accumulation of T to C conversions during RNA-seq library preparation at the position where 4sU was incorporated. These conversions are then used to detect and quantified newly transcribed RNAs by computational analysis. Thus, it is possible to distinguish *in silico* the 'new' and 'pre-existing' transcripts populations present and calculate the associated kinetic rates of RNA synthesis and decay for hundreds of genes, without performing pull-down or immunoprecipitation enrichment steps (Baptista and Dölken, 2018). Recently, a labelling strategy using a combination of nucleotide analogues has been applied to obtain RNA synthesis and decay rates that are more reliable (Gasser et al., 2020; Kawata et al., 2020).
Different computational pipelines were developed in parallel to the experimental optimization of metabolic labelling approaches to quantify RNA dynamics: cDTA (Sun et al., 2012), DRiLL (Rabani et al., 2014), INSPEcT (Pretis et al., 2015) and GRAND-SLAM (Jürges et al., 2018) are some of them. Ultimately, these methods began to reveal how the fine tuning of temporal RNA dynamics contributes to gene-specific regulation and the outcome of complex transcriptional cellular responses.

The described methods are based on the analysis of bulk RNA-seq data, averaging the transcriptome expression over the whole cell population. An approach that could distinguish the RNA kinetics for every single cell present in a sample would be advantageous and more informative. Recently, single cell RNA sequencing (scRNA-seq) has become the method of choice to investigate the complexity and heterogeneity of cell populations (Alles et al., 2017; Hashimshony et al., 2012; Macosko et al., 2015; Pepe-Mooney et al., 2019). It provides a snapshot of the cellular RNA abundance for individual cells in

the analysed samples. Typically, scRNA-seq focuses on the mature RNA expression for the identification of cell identity and cell states, without direct measurement of the transcript kinetics. RNA velocity analysis (Manno et al., 2018) estimates gene-specific transcription and degradation rates using both spliced and unspliced transcripts retrieved from scRNA-seq data. Although crude estimates of kinetic parameters of a subset of genes are used to identify the overall direction of the expression state (Bergen et al., 2020), the analysis opens the door to the study of RNA kinetics using single cell data.

More recently, single cell methods have been established to investigate dynamic gene-specific temporal RNA kinetic (scSLAM-seq (Erhard et al., 2019), NASC-seq (Hendriks et al., 2019), scEU-seq (Battich et al., 2020), sci-fate (Cao et al., 2020) and scNT-seq (Qiu et al., 2020)). While these approaches are similar to each other in some aspects, there are important differences: SMART-seq based methods (scSLAM-seq, NASC-seq, scEC-seq, sci-fate) can perform deep coverage of transcripts in each cell, but can only assay hundreds of cells. When many cells in heterogeneous populations need to be analysed, droplet-based approaches such as scNT-seq are more advantageous, recovering thousands of cells per experiment. Moreover, methods that are devoid of any cell sorting or pull-down procedures would be favourable, since they would diminish cellular stress, introduction of biases and sample handling time. To date, the described single-cell approaches have been exclusively applied to investigate gene-specific changes on transcriptome kinetics between states of cell activation, differentiation or infection. Since biological phenomena are continuous processes, it would be more informative to investigate the RNA-associated kinetic rates in a time-dependent manner, assigning quantitative values throughout the course of the biological event such as the cell cycle.

RNA kinetic rates analysis during the cell cycle

Cell cycle progression is tightly regulated at the of transcriptional (Dynlacht, 1997), post-transcriptional (Schafer, 1998) and post-translational levels, in order to avoid unwanted and possibly harmful cell divisions. Cell cycle checkpoints and regulatory genetic circuits survey the progression through the cell cycle and are tuned to prevent unwanted cell divisions when DNA damage or other types of stressors threaten it (Cho et al., 2001; Pines, 1999). Hundreds of genes regulate these sophisticated biological processes and many of them are expressed at precise time points of the cell cycle, *i.e.* their relative RNA amount oscillates respect to the cell cycle phase. Of note, the disruption of their coordinated regulation is a major cause of several disorders, including cancer (Hanahan and Weinberg, 2011; Matthews et al., 2022), emphasizing the importance of understanding how they are regulated. In the last decades, the expression profiles for genes oscillating during the cell cycle have been determined in several cellular systems and with different approaches (Breeden, 2003; Gauthier et al., 2008; Liu et

al., 2017; Whitfield et al., 2002). On the other hand, the processes underlying gene expression (*i.e.*, transcription, splicing and degradation) and their relative kinetic rate dynamics throughout the cell cycle are still poorly characterized. Because of its importance in every eukaryotic cell, a global, quantitative and time-depended profiling of RNA kinetics during the cell cycle would be of extreme value.

Previous approaches to study RNA kinetics of the cell cycle synchronized cell population by blocking cells around a phase of the cell cycle using drugs or serums. Subsequent RNA expression profiling of synchronized cells was analysed along specific cell cycle states, revealing hundreds of periodically expressed genes in human cells (Grant et al., 2013; Whitfield et al., 2002). Although informative, synchronization procedures introduce biases which are difficult to account for (Cooper, 2019; Darzynkiewicz et al., 2011; Ligasová and Koberna, 2021). Fluorescent reporters have been engineered to overcome some of these limitations and discern the cell cycle phase of individual cells from a population of unsynchronized cells (Koh et al., 2016; Sakaue-Sawano et al., 2008), but they require genetic manipulation. Moreover, both synchronization and fluorescent reporters recover cell populations from discrete cell cycle phases, partially losing the temporal resolution of RNA regulation dynamics.

Assigning a precise cell cycle time to unperturbed single cells using their transcriptional information (*i.e.*, oscillating cell cycle marker genes) would overcome this temporal inaccuracy. Recent approaches have enabled the *in silico* reconstruction of cell cycle phases from scRNA-seq data (Schwabe et al., 2020; Wolf et al., 2019). Integrating this accurate cell cycle temporal assignment of single cells with a method profiling their transcriptional activity would allow the study of RNA synthesis and degradation kinetics in the continuous cell cycle at an unprecedent temporal resolution.

AIMS OF THE THESIS

The unbiased assessment and quantification of the determinants of biological entities is a central goal for system biology. In this work, we aimed to develop novel approaches to identify different biochemical characteristics of human cells.

The proper compartmentalization of proteins determines their function and activity. New methods and datasets aiming to define their localization are useful to have a more comprehensive understanding of cellular biology. Moreover, the development of activity-based approaches to annotate the RNA-bound proteome at a specific cellular region could also offer insights about the distribution of RBPs and their modality of binding to RNAs in different cellular compartments.
Lastly, improving the resolution to study temporal kinetic information of gene regulation not only in cell population but also in single cells has the potential to unravel new gene-specific mechanisms regulating gene expression in continuous processes.

Overall, the aim of this thesis is to provide a more detailed description of the spatial and temporal regulation of gene expression in a human cell line.

More specifically, the aims of this PhD thesis are:

1) Determine the repertoire of proteins localized on the cytosolic surface of the endoplasmic reticulum, mitochondria and plasma membranes in HEK293 cells.

2) Establish an approach to characterize proteins actively binding to RNA in different cellular compartments and categorize some of their binding properties.

3) Develop a method to study RNA temporal dynamics in single cells and apply it to study the kinetics of RNA synthesis and turnover during the cell cycle of HEK293 cells.

RESULTS

Localization of proteins in human cells by proximity labelling with BirA*

The identification of human proteomes resident in compartments difficult to purify (i.e., the cytosol-facing membranes of ER, mitochondria and plasma membrane) was addressed by implementing a compartment-specific proximity labelling by BirA and mass spectrometry analyses of enriched proteins.*

Generation and validation of BirA stable cell lines*

A key problem in the previous research on protein localization is the identification of the interface-bound proteome. We therefore implemented an approach aiming to determine the protein composition at selected cellular compartments namely the cytosolic sides of ER, mitochondria and plasma membranes.). As discussed in the introduction, the proteomes resident in these cellular locations are involved in many different biological processes and pathways (apoptosis, secretion, extracellular response to stimuli, etc.), but they are also difficult to purified and therefore understudied. HEK293 were selected as a model cell line. These cells offer the advantage to be quickly genetically modifiable: any gene of interest can be stably integrated in their genome and transcribed through the induction of a doxycycline-controlled promoter.

The promiscuous biotin ligase BirA* was chosen as the enzyme to perform proximity labelling (Figure 1). Even with a slow labelling kinetic (around 16 hours), this enzyme does not require toxic compounds

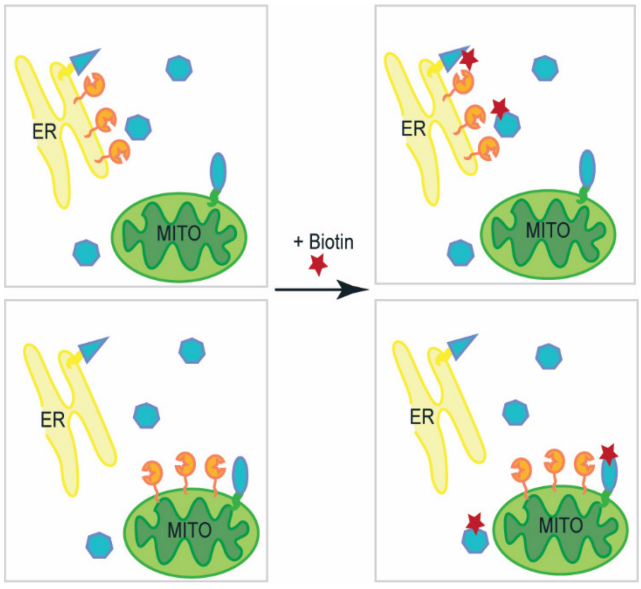

Figure 1 ***Schematic representation of the experimental workflow to identify the proteome composition of different subcellular compartments in HEK293 cells.***

BirA compartment-specific fusion proteins (indicated in orange color) are expressed upon doxycycline induction and localize at specific cellular compartments via a genetically fused target peptide. BirA* at the cytosol facing side of ER (above) and mitochondria (below) is shown, but cell lines were also created with BirA* at the plasma membrane and diffusing in the cytosol. Biotinylation of nearby proteins (indicated in blue) is then induced by addition of biotin for 16 hours.*

to start the biotinylation reaction, as the H_2O_2 required by APEX (Lam et al., 2015) and its expression does not induce cellular stress.

Stable HEK293 cell lines expressing BirA* in different subcellular compartments were created using plasmids coding for the promiscuous biotin ligase linked to a specific peptide targeting to the desired cellular localization (Figure 1). BirA* was localized at the cytosol-facing membrane surfaces of three different compartments: endoplasmic reticulum (ER-BirA*, via fusion to the N-terminal first 27 amino acids of the human protein P450 2C1), outer mitochondrial membrane (MITO-BirA*, via fusion to the C-terminal 31 amino acids of the MAVS protein) and plasma membrane (PM-BirA*, via fusion to the prenylation motif CaaX, derived from the Rev protein of the HIV-1 virus). Two control cell lines were also created, with soluble cytoplasmic BirA* to distinguish the resident proteins at a particular location from proteins with short interactions with the analysed compartments: cyto1-BirA*, by fusing a NES motif, and cyto2-BirA*, by mutating the cytosine in CaaX motif of PM-BirA* a serine residue. The cytoplasmic cell lines were important as background controls: they were used to distinguish the resident proteins at a particular location (ER, mitochondria or plasma membrane) from proteins with short interactions with the analysed compartments.

The doxycycline-inducible expression of BirA* in them was checked using Western blotting analysis with antibodies against the N-terminally fused HA epitope tag. The correct localization of the enzyme

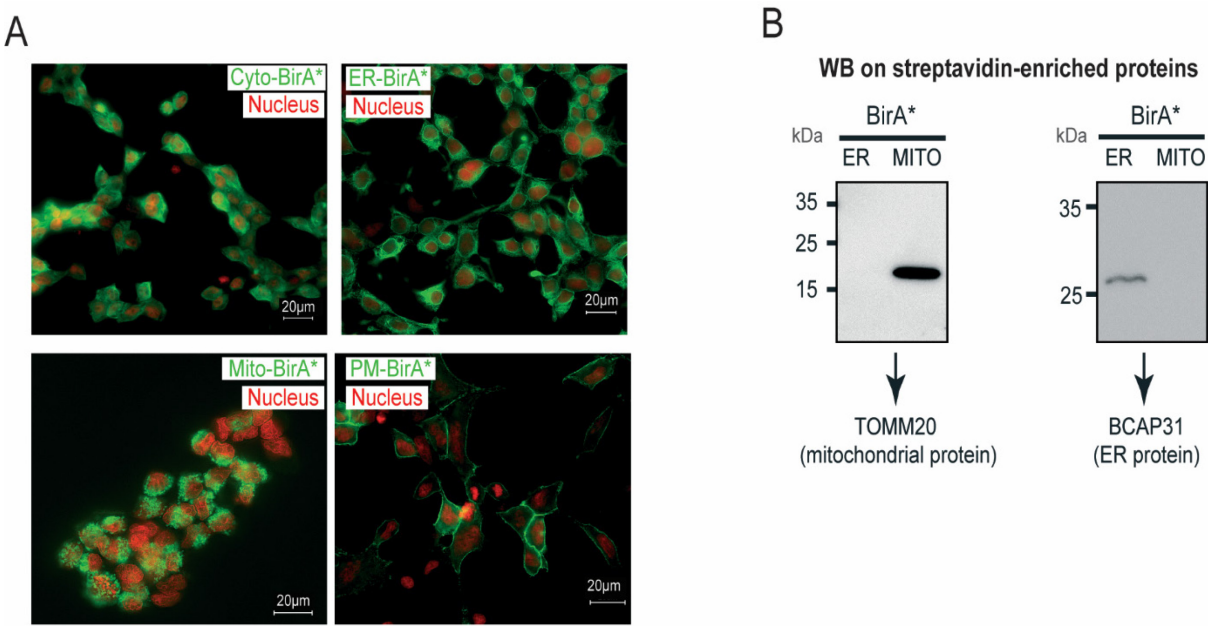

Figure 2 ***BirA* cell-line specific localization and biotinylation is validated.***

A) Immunofluorescence showing the cellular localization of the BirA enzyme in four generated cell lines. Nuclei are in red and BirA* in green. The enzyme is freely diffusing in the cytosol in the cyto-BirA* cell line, while it forms the typical ER structures in the ER-BirA* cell line. Cells expressing the fusion protein of PM-BirA* localize the biotin ligase on the cell surfaces, while it can be visualized in the mitochondria in the MITO-BirA* cell line.*

B) Western blotting analysis for protein markers of mitochondria and ER (TOMM20 and BCAP31, respectively) show a specific enrichment only if BirA localized to the respective compartment. Proteins were purified with streptavidin beads after biotinylation by BirA* was induced for 16 hours and equal volumes were run on a polyacrylamide gel.*

at the addressed cellular region was confirmed by immunofluorescence microscopy (Figure 2A). The proteins biotinylated by BirA* were also visualized by immunofluorescence with streptavidin conjugated to a fluorophore, showing a preferential distribution in the different cell lines (not shown). The accuracy of compartment-specific biotinylation was then tested as following. BirA* biotinylation activity was induced in the different stable cell lines and the biotin-labelled proteins were enriched via streptavidin beads pull-down. Western blots against known protein-markers of the different cellular compartments showed specific protein enrichment in the expected sample, confirming the specificity of the biotinylation (Figure 2B).

In summary, we created five HEK293 cell lines expressing the promiscuous biotin-ligase BirA* linked to different targeting peptides. BirA* was properly localized at the addressed cellular compartment and its enzymatic activity was reliably restricted to the expected area of the cell.

Mass spectrometry reveals proteomes compartmentalization

To identify the population of biotinylated proteins at the selected cellular regions, we performed an enrichment experiment coupled with liquid chromatography tandem mass spectrometry (LC-MS/MS) in collaboration with Dr. Guido Mastrobuoni (in the laboratory of Dr. Kempa at the MDC / Berlin). The different BirA* cell lines and a control sample without the biotin-ligase (HEK293) were induced with doxycycline and biotin was added to the cell culture media to stimulate the biotinylation for 16 hours. Proteins with a covalently attached biotin were enriched from lysed cells using streptavidin beads, digested by proteases and submitted for label-free mass-spectrometry (Figure 3). LFQ and iBAQ intensities of four biological replicates were calculated with the software MaxQuant (Cox and Mann, 2008).

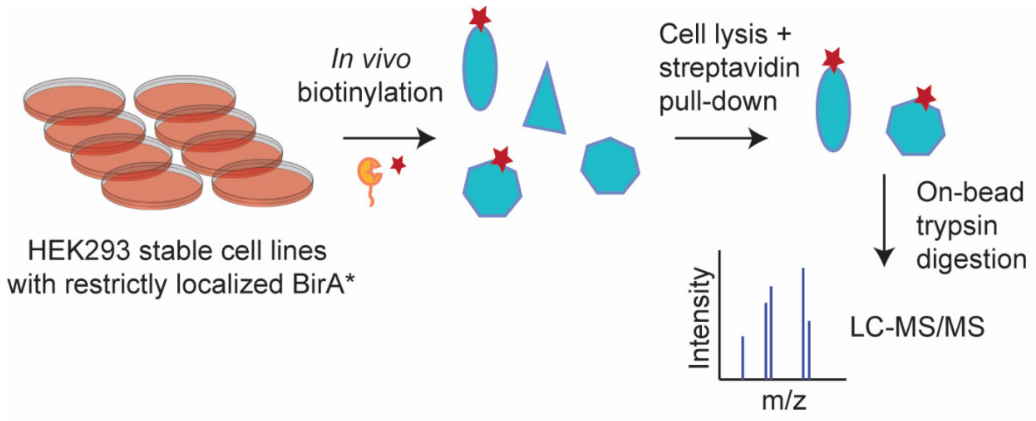

Figure 3 ***Scheme representing the experimental outline to study proteome compartmentalization using locally-restricted BirA* and MS.*** *BirA* expression and biotinylation is induced in the different cell-lines (with doxycycline and biotin). The subcellular-specific proteomes are enriched after cell lyses with streptavidin pull-downs. The proteins are digested and analyzed with LC-MS/MS. m/z = mass to charge ratio. Four biological replicates were prepared.*

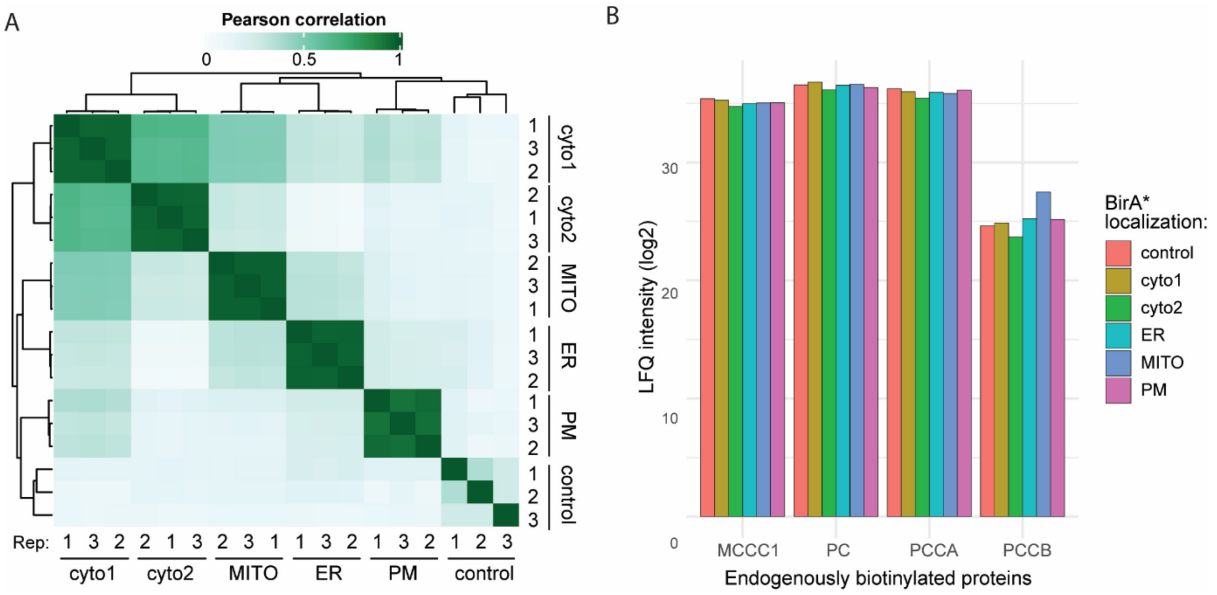

Figure 4 ***Replicates reproducibility and equal samples loading is recapitulated in MS data.***

A) Correlation matrix of pair-wise comparisons between proteins' LFQ intensities from BirA cell lines. It shows high correlation (Pearson coefficient > 0.9) between replicates of the same condition, except for the control HEK293 sample.*

B) Histogram of LFQ intensities for four endogenously biotinylated proteins. MCCC1, PCC1, PC and PCCB are physiologically biotinylated proteins in human cells and their LFQ intensities are proportional between all samples analyzed (mean between replicates), indicating a comparable loading of the samples for MS.

A minimum of two unique peptides and an absence of missing values for LFQ intensities in at least one condition were the requirements for a protein to be considered in the analysis. After these and other filtering steps (see Material and Methods), 3706 different proteins were detected in our data. The scripting language R was used to perform quality check and data analysis.

The correlation of LFQ intensities between the different replicates was good, except for replicate 4 which was therefore excluded. Figure 4A shows the high correlation of LFQ intensities between replicates of the same condition (Pearson correlation coefficient > 0.9), except for the control HEK293 sample, which is not surprising since these proteins can be mostly regarded as random background attached to the beads. The correlation diminishes when comparing different conditions between each other, with higher correlation between the two cytosol samples.

As a control for the equal loading and pull-down efficiency of the different samples, the intensities of four proteins known to be endogenously linked to a biotin moiety (*i.e.*, not added by BirA*) were compared. Their LFQ intensities were very similar between all samples, also in the control HEK293 sample which does not express a BirA* fusion protein (Figure 4B). This indicates that the amount of protein lysate per sample was comparable and, once the background proteins from the control condition were filtered out, the differences in LFQ intensities of proteins between samples is due to a localized BirA* mediated biotinylation. The LFQ intensities were then normalized to trypsin, which was added in equal amount in the different samples.

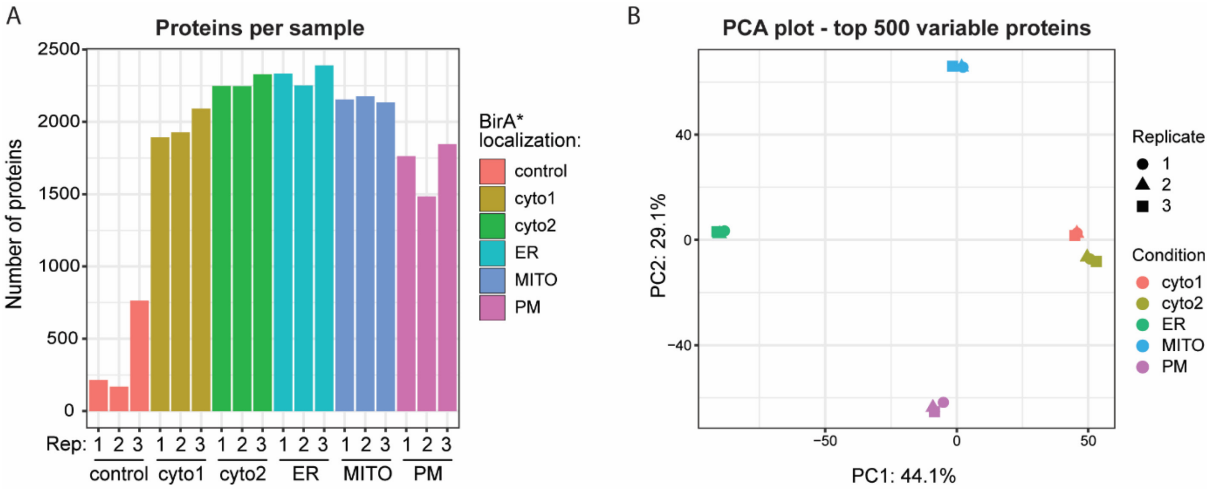

Figure 5 ***Quality control of proteins enriched in different BirA* samples***

A) Histogram showing the total number of proteins identified in the different samples and replicates.

B) PCA analysis of the top 500 variable proteins in the BirA datasets. Replicates of the same condition cluster together, while conditions are well separated between each other along the first two principal components, except for the two cytosolic proteomes.*

The number of proteins detected in each sample is indicated in Figure 5A, and ranges from a minimum of 1439 to a maximum of 2339 between BirA* samples, while in the control HEK293 sample much less proteins were identified (only 121 were identified in all the replicates). These proteins were defined as background and were filtered out from the data analysis.

The principal component analysis of the top 500 variable proteins (without HEK293 sample, Figure 5B) shows a clear separation of the different samples and clustering of replicates for each condition. Of note, the two cytosolic BirA* samples cluster in close proximity, while the other conditions have a higher variability between each other in both principal components. This highlights how the proteomes recovered are dependent on the cellular localization of the BirA* enzyme and the reproducibility of the detection is very high between the replicates of the same sample.

The R-package “limma” (Ritchie et al., 2015) was applied to identify the differential localization of proteins in the selected compartments. This package uses protein-wise linear models and empirical Bayes statistics to calculate LFQ intensity fold changes and p-values. Pair-wise comparisons of the LFQ intensities for each protein in the different conditions were tested and we considered a protein to be specifically localize to one compartment if it had a log2 fold change of LFQ intensities greater than 1 and an adjusted p-value (Benjamini-Hochberg correction for multiple comparisons) lower than 0.01. If a protein did not meet these thresholds, it was assigned as non-compartment specific. The samples with BirA* localized in the cytoplasm were used as a background control and proteins enriched in them were not considered for further analysis. Since all the other cell lines have BirA* facing the cytosol, an abundant cytoplasmatic protein like GAPDH would be biotinylated and enriched in every sample. Using two cytosol-background samples, allowed to filter these proteins from the datasets.

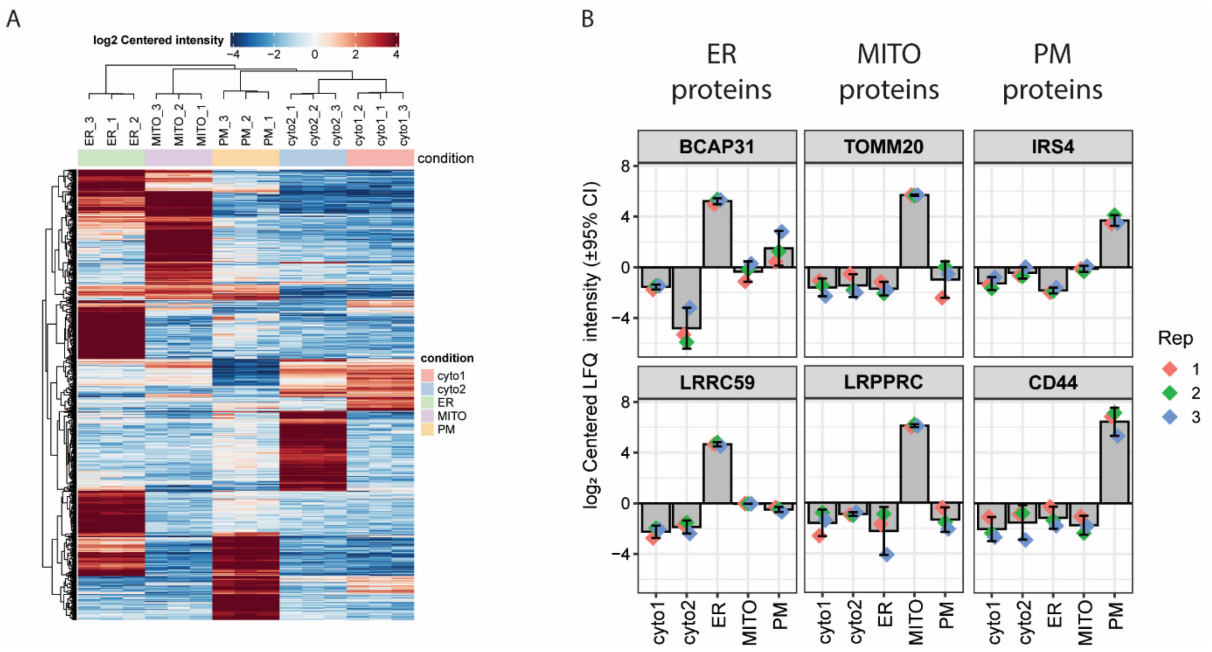

Figure 6 ***MS data analysis recapitulates subcellular compartmentalization of proteins.***

A) Unsupervised hierarchical clustering separates proteins in different groups, according to their LFQ intensities.

B) Protein markers for the different compartments (BCAP31 & LRRC69 for ER, TOMM20 &LRPPRC for the mitochondria, IRS4 & CD44 for the plasma membrane) are successfully detected and enriched for the expected compartment.

The heatmap of significantly localized proteins (Figure 6A) shows the distribution of LFQ intensities among the different samples and highlights clusters of proteins enriched in specific compartments. Colocalization of proteins can also be visualized, like for example between plasma membrane and ER (highlighting the connection of these compartments in the secretory pathway) or between mitochondria and ER (possible candidates for mitochondria-ER contact sites proteins). The two cytosolic controls show similar intensities in the majority of proteins detected, apart from a cluster enriched in the cyto2-BirA* cell line. This result was expected, due the difference of the targeting peptide sequences in the two cell lines: while cyto1-BirA* is strictly localized to the cytosol via a nuclear-export signal, cyto2-BirA* is diffusing in both cytosol and nuclei by a mutation introduced in its prenylation sequence. Therefore, the proteins enriched only in the cyto2-BirA* cell line are mostly nuclear proteins, including e.g. 36 zinc-finger transcription factors and 13 dead box helicases.

When looking at some specific protein markers of ER, mitochondria or plasma membrane, the LFQ intensities distribution indicated an enrichment of the respective protein in the compartment where it would be expected to be (Figure 6B). These results gave us confidence that the enrichment and detection of localized proteins following the applied protocol was biologically meaningful.

After filtering and enrichment analysis as described above, 475 proteins were found preferentially localized at the cytosolic surface of the ER membrane, 450 at the plasma membrane and 281 at the

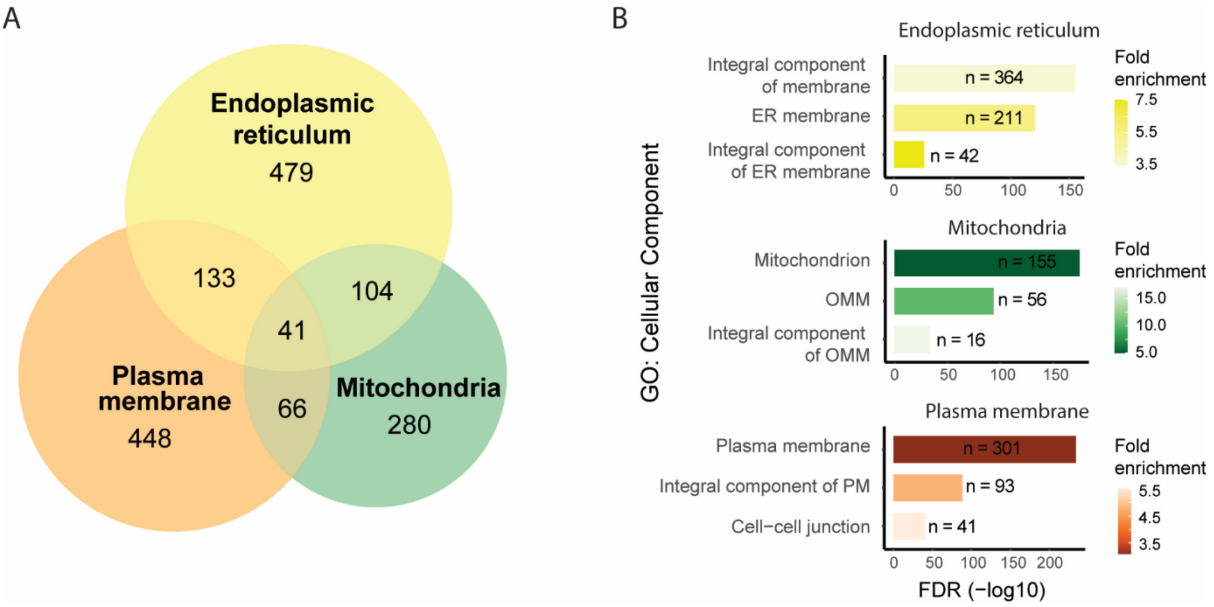

Figure 7 ***Proteomes retrieved from the different BirA* cell lines***

A) Venn diagram of protein distribution found to be enriched in the indicated compartments through our BirA proximity labelling approach and mass spectrometry analysis.*

B) Gene ontology (GO) terms enrichment analysis for cellular component for proteins identify to exclusively localize at the ER, plasma membrane or mitochondria membranes. Fold enrichments and FDR are calculated by DAVID. The number of proteins for each term is indicated by 'n'.

outer membrane of the mitochondria (Figure 7A). Overlaps between the datasets indicate multiple locations for some proteins and/or a close proximity of the two compartments. Only 39 proteins were identified to be enriched in all three compartments. The Gene Ontology analysis for cellular compartment resulted in terms strongly related to the location investigated (Figure 7B), indicating accuracy of the enriched proteomes and computational analysis.

As examples, the following proteins were found at the endoplasmic reticulum: the translocon SEC61/62/63, the lectin associated proteins CNX, ERP57 and CRT, the COP-II vesicles associated SEC13/31, SEC23/24 and SEC12, the ER-stress receptors PERK, IRE1 and ATF6 and its associated protein BiP, the p180 receptor and many others.

Within the mitochondrial enriched proteins, we found for example the SAM complex, composed by the SAM50, MTX1/3 and MTX2 proteins, and the TOM complex, from which the MITO-BirA* cell line efficiently biotinylated TOMM40, 20, 22, 7, 70A and 5. The KEGG pathway enriched with the proteins found at the plasma membrane with BirA* were instead related to cell-to-cell tight junction (e.g. CLDN, OCLN, JAM1, JAM2 and JAM3 proteins) and to neuronal pre- and post-synaptic contacts (e.g. PVRL1, PVRL3, IGSF4, PTPRF, SDC, NGL3 and ITGB1 proteins). This analysis gave us further confidence that the proteome enriched were accurately resembling the *in vivo* protein distribution in HEK293 cells.

After having identified these proteins at the investigated cellular compartments, we compared our results with published datasets. The coverage of proteins enriched at the ER was assessed by

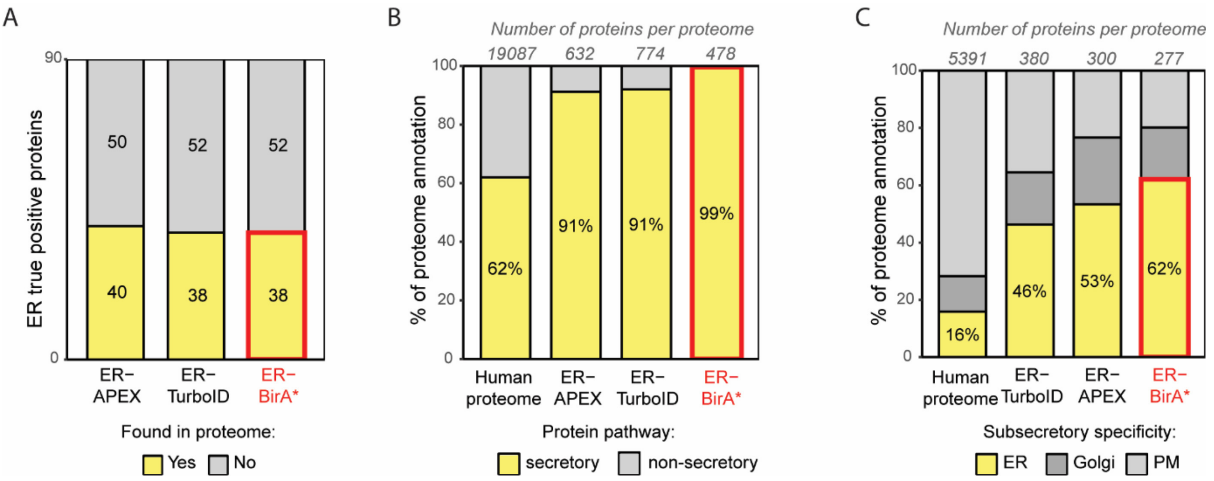

Figure 8 ***Bar plots indicating the sensitivity and specificity of the proteome enriched by ER-BirA*.***

A) Identification of ER annotated proteins (true positive) in proteomes from previous studies (ER-APEX and ER-TurboID) and this study (ER-BirA). B) Annotation for secretory pathways in proteins composing the human proteome or detected by ER-APEX, ER-TurboID and ER-BirA*. C) Sub secretory protein annotation of indicated datasets based on Gene Ontology cellular components terms*

comparing it to a list of manually annotated true positive proteins located at the endoplasmic reticulum (Hung et al., 2017). Our proteome enriched for around 42% of the annotated ER proteins and performed very similarly to the one previously detected with similar approaches (APEX2 and TurboID, Figure 8A). To define the specificity of the proteome enriched by ER-BirA*, we determined the fraction of proteins with a previous secretory annotation (according to GOCC and Phobius). In the ER-BirA* dataset 98.5% of the proteins are annotated to be in the secretory pathway, against 62% of the total human proteome and 91% of the ER-APEX and ER-TurboID datasets (Figure 8B). This result suggests that, even if our dataset is smaller than the one published in the APEX or TurboID paper, the specificity for secretory proteins enrichment was higher. When comparing the sub-secretory annotation of the proteins retrieved, my ER-BirA* proteome is enriching for more proteins with an ER annotation respect to the other dataset (Figure 8C), confirming its high specificity and accuracy for ER resident proteins.

Of the entire human proteome, only about 8% localize at the mitochondria. Our mitochondrial dataset instead consists of more than 60% proteins with previous mitochondrial annotation (Figure 9A). The majority of the remaining proteins are cytosolic, suggesting a transient interaction these proteins might have with the mitochondria. The comparison of my MITO-BirA* proteome with a list of well-annotated mitochondrial proteins did retrieve more than half of them, a higher score respect to the previously annotated datasets using APEX2 and TurboID (Figure 9B). The sub-mitochondrial specificity was also checked: between the different compartments of the mitochondria (outer mitochondrial membrane — OMM, inner mitochondrial membrane — IMM, intermembrane space — IMS, and matrix), the proteins localized at the mitochondrial outer membrane in human are around 18%,

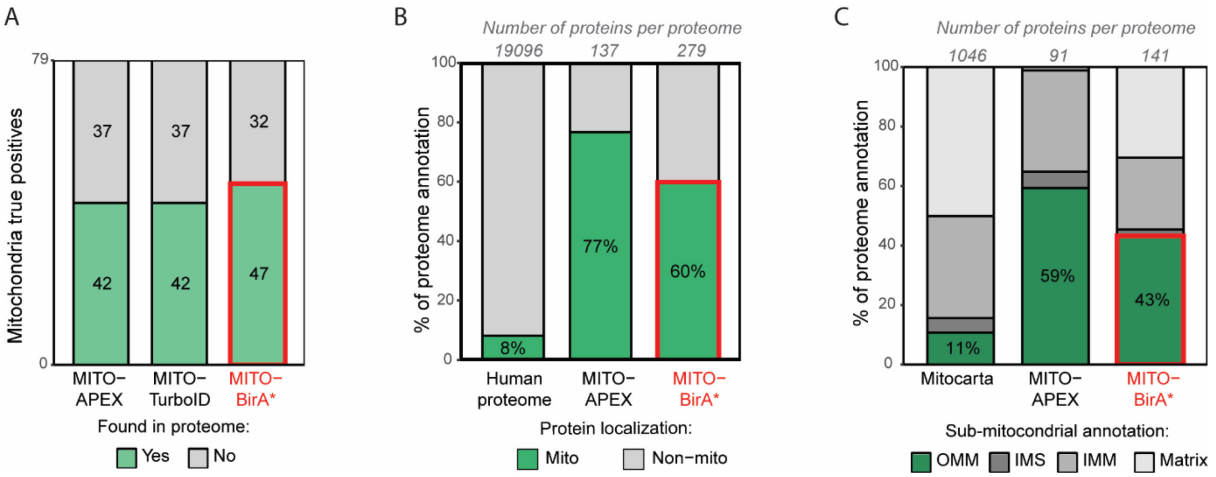

Figure 9 ***Bar plots indicating the sensitivity and specificity of the proteome enriched by MITO-BirA*.*** *A) Coverage of known localized mitochondrial proteins (true positives) in 3 datasets. B) Mitochondrial protein annotation for human, MITO-APEX and MITO-BirA* proteomes. C) Sub-mitochondrial specificity of localization between the different datasets. OMM = outer mitochondrial outer, IMS = intermembrane space, IMM = inner mitochondrial membrane.*

according to GOCC. In our dataset, with BirA* localized at the cytosol facing side of the outer mitochondrial membrane, we counted around 40% of proteins localized at the OMM (Figure 9C), underlying a compartment-specific enrichment of the proteins retrieved in my MITO-BirA* dataset. Compared to the MITO-APEX dataset, MITO-BirA* detects more proteins at the outer mitochondrial membrane (141 versus 91), but with less sub-mitochondrial specificity.

The specificity of the proteins reported in the PM-BirA* dataset is shown in Figure 10. As expected an enrichment of protein located on the plasma membrane was observed in respect to their composition in the human proteome or according to their sub-secretory distribution.

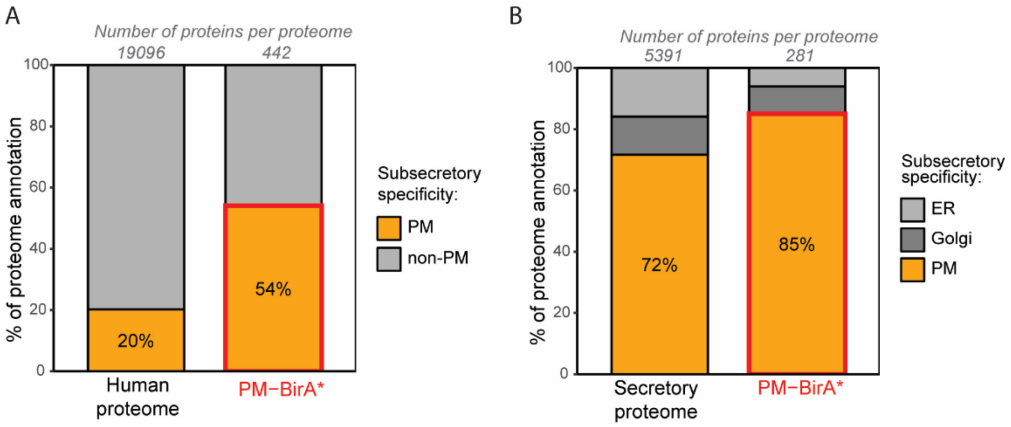

Figure 10 ***PM-BirA* enriches for plasma membrane annotated proteins.*** *A) Bar plot showing the enrichment of the PM-BirA* dataset for proteins with plasma membrane annotation or (B) with sub-secretory annotation to it, compared with the total human or secretory proteomes.*

Our approach and computational analysis detected polypeptides not restricted to only one compartment. Proteins co-localizing in more than one cellular region provide information about

potential physical contacts between two compartments. In fact, we detected proteins previously identified at mitochondrial-ER interaction sites to share the reported localization in our approach too (*i.e.,* VDAC1, INF2 and others, see Discussion). 133 proteins co-localize at the plasma membrane and ER in our dataset and, while many of them are part of the secretory pathway as expected, some of them are known to mediate the physical contacts between the two compartments, like the adenylyl cyclases 3 and 8 (Okeke et al., 2016).

In summary, our compartment-specific proximity labelling approach in combination with mass spectrometry created high quality datasets of localized proteomes and described for the first time the localization of proteins in HEK293 cells at these subcellular regions.

Localization of RNA-binding proteins in HEK293 cells

We coupled cellular fractionation with organic phase enrichment (XRNAX) to detect proteins actively binding RNA in vivo in the nucleus, membranes and cytoplasm of HEK293 cells. The global identification of these proteomes through LC-MS/MS revealed their localization, their RBD composition and the compartment-specific RNA-binding preference.

Mapping RBPs at different subcellular compartments with f-XRNAX

To identify the subcellular distribution of human RNA-binding proteins and to have a better understanding of RBPs actively engaging with RNA molecules, we developed a novel approach that enables the identification of proteins binding to transcripts from a cellular compartment of choice. We combined sequential detergent extraction fractionation with XRNAX (Trendel et al., 2019), an organic-aqueous phase separation method enriching proteins crosslinked to nucleic acid (Figure 11A). I developed this new approach in collaboration with Dr. Igor Minia, a post-doctoral scientist in the laboratory of Markus Landthaler (MDC/Berlin).

Dr. Minia first optimized the fractionation protocol described previously for HEK293 cells (Jagannathan et al., 2011), testing different buffer compositions and timing of compartment dissociation (see Material and Methods). This optimized approach allowed us to obtain high-quality fractionated samples of HEK293 cytoplasm, membranes and nuclei. The purity of each fraction was confirmed by Western blot analysis with antibodies against known proteins of these compartments: the left panel of Figure 11B shows the detection of cytosolic protein Beta-tubulin only in the cytoplasmic fraction, of the ER-localized BCAP31 only in the membrane fraction and of histone H3 in the nucleus of HEK293 cells.

After obtaining good quality cellular fractions, we attempted to enrich RBPs from them. We applied UV light at 254 nm wavelength on intact cells to crosslink proteins to nucleic acids in their proximity, performed the fractionation protocol while blocking translation with cycloheximide and applied XRNAX from these fractions. This led to an enrichment of proteins linked to nucleic acids while retaining spatial information. The efficiency of RBPs recovery was validated using Western blotting analysis with specific antibodies against known protein which bind nucleic acids. Furthermore, protein samples from cells with and without UV crosslinking were compared: Figure 11B right panel Western blots show how the applied protocol specifically enriches for proteins binding to RNA in the UV-irradiated experimental condition, conserving the information of their cellular localization in the cytosol, membranes and/or nuclei. The proteomes recovered in the different fractions and conditions were separated by SDS-PAGE and visualized using silver staining, clearly showing enrichment in the UV crosslinking condition of the same compartment. When comparing between different fractions, the

amount and identity of proteins identified by silver staining is varying widely, pointing to a differential proteome composition dependent on the cellular compartment. These results confirmed our goal to obtain compartment-specific enrichment for RBPs. The novel protocol of cell fractionation in

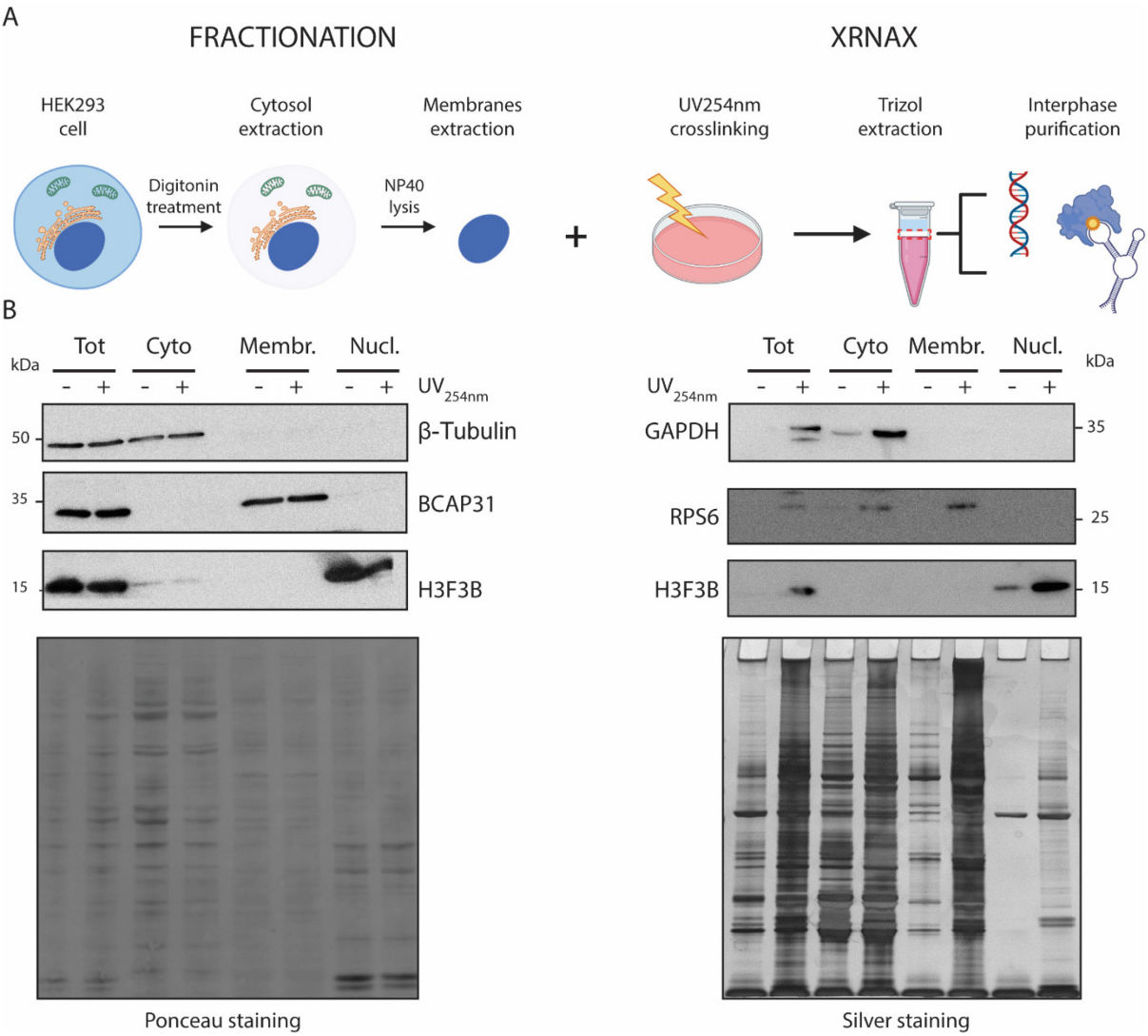

Figure 11 ***Mapping RBPs localization with fractionation coupled to XRNAX.***

A-B) Scheme of the experimental procedure (A) and its specificity assessed by Western blots (B). Briefly, cells are UV-irradiated and treated with different buffers to obtain cytosolic, membrane and nuclear fractions. The purity of the fractions is assessed by detection of compartment protein markers (β-Tubulin for cytosol, BCAP31 for membrane and histone H3 for nucleus). The XRNAX protocol was performed on the different fractions and its accuracy checked with Western blots with antibodies against RBPs with a preferential compartment localization (FUS for nucleus, RPS6 for membrane, GAPDH and H3 for cytosol). Enrichment of RBP is only in the +UV condition. Loading controls are shown below (Ponceau and silver staining).

combination with the isolation of crosslinked protein-RNA complexes was termed 'f-XRNAX'.

Having established reproducible conditions for RBP recovery from specific cellular compartments, we designed an experiment to identify the overall composition of proteins binding to RNA in the cytosol, membranes and nucleus of human cells. We applied f-XRNAX to HEK293 cells with or without UV_{254nm} crosslinking and treated the RBP-enriched fractions with partial enzymatic digestion, column purification and complete Trypsin/Lys-C digestion (see Material and Methods). The obtained peptides

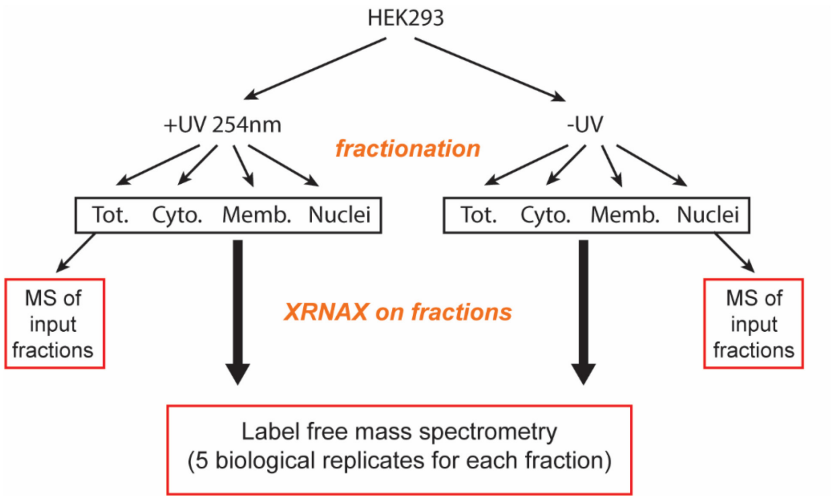

Figure 12. ***Experimental outline of f-XRNAX coupled to LC-MS/MS.***

UV-crosslinked (+UV) and non-crosslinked (-UV) samples were separated in different cellular fractions, which were then subjected to XRNAX and LC-MS/MS. Input samples were also analysed by mass spectrometry. 5 replicates were prepared.

were submitted for liquid chromatography tandem mass spectrometry (LC-MS/MS) to Dr. Marieluise Kirchner (collaborator in Philipp Mertens laboratory). Input samples from the different fractions without the XRNAX enrichment step were also analysed in order to detect total protein abundance before RBP-enrichment. In addition to cytosol, membrane and nuclei fractions, a sample of unfractionated HEK293 cells was included (total). Five biological replicates of the experiment were prepared and label-free quantification was applied to analyse protein samples (Figure 12)

RBPs compartmentalization

MaxQuant (Cox and Mann, 2008) was used to analyse data from the LC-MS/MS experiment and to calculate the LFQ and iBAQ intensities for proteins detected in all analysed samples. We applied stringent thresholds to filter for high-confidence data: hits with less than two unique peptides were discarded and only proteins with no missing values in all replicates in at least one condition were kept

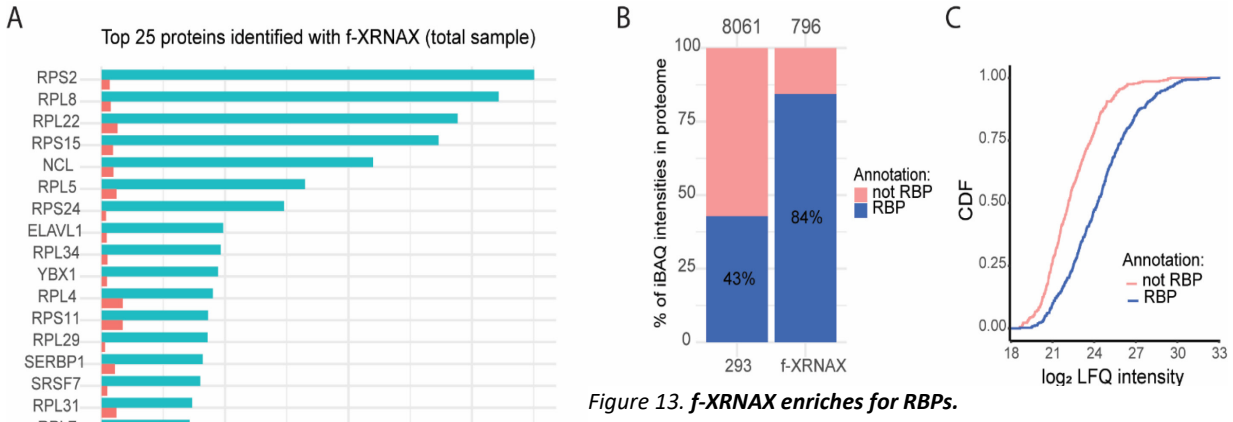

Figure 13. ***f-XRNAX enriches for RBPs.***

A) The iBAQ fraction for the 25 most enriched proteins with f-XRNAX is shown in comparison with the one from the HEK proteome.

B) The percentage of iBAQ intensities for annotated RBPs is almost double in the f-XRNAX extract respect to whole cell lysate.

C) Not annotated RBP are less enriched respect annotated ones in the total f-XRNAX sample.

(see Material and Methods for more details). To confirm the high quality of the data, we calculated the relative abundance of the top 25 proteins enriched by f-XRNAX compared to HEK293 total lysate: all 25 proteins are annotated as RBPs and are all more represented in the experimental condition compared to the whole cell lysate (Figure 13A). Furthermore, the proteomes obtained with f-XRNAX have a higher fraction of proteins previously annotated as RBP compared to the HEK293 total proteome (Figure 13B). When looking at the total sample proteome obtained with f-XRNAX, the fraction of proteins without a previous annotation of RNA-binding are less enriched (lower cumulative intensity values) respect to established RBPs (Figure 13C). These results confirmed that the applied protocol specifically enriched for RBPs and allowed us to focus on the detection of RBPs differentially localized in the analysed compartments.

To filter out background noise from the data, we first normalized the LFQ intensities with trypsin (added equally to each sample before LC-MS/MS) and then compared them gene-wise between the +UV and -UV conditions of the same fraction. Replicates of the same sample not correlating well with each other (Pearson correlation coefficient < 0.85) were discarded: this left 5 replicates for the membrane and nuclear fractions, 4 for the total sample and 3 for the cytosolic one.

We applied R packages limma and eBayes to calculate fold changes of LFQ intensities and determine p-values of the determined contrasts. The Benjamini-Hochberg correction was used to test for false

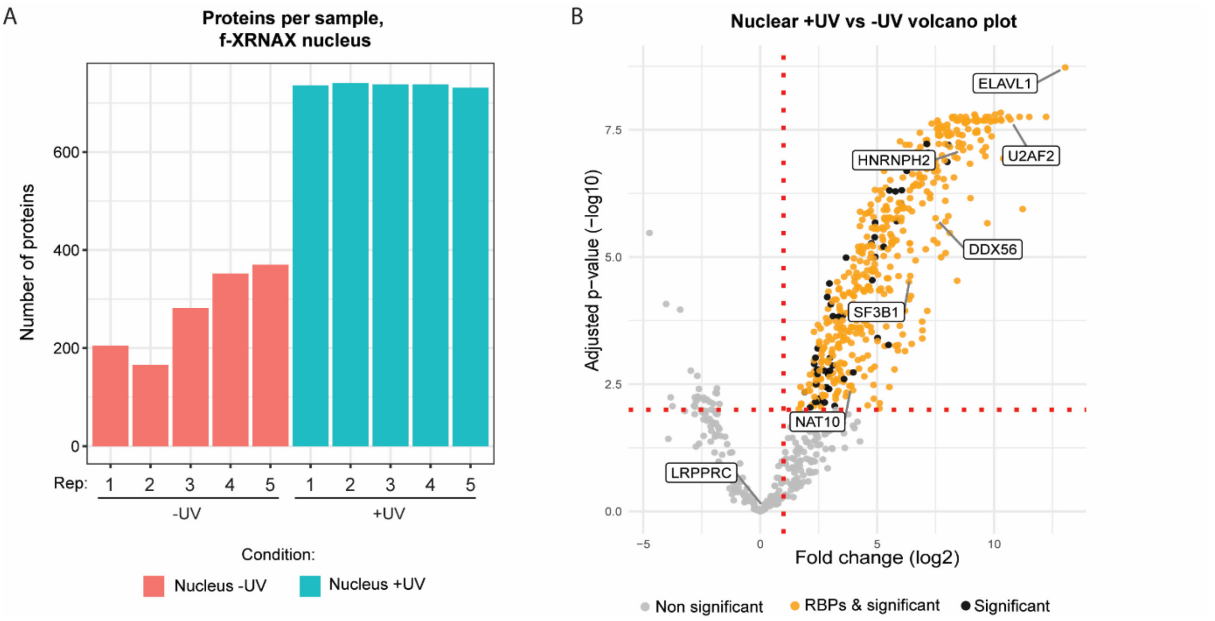

Figure 14 ***Removing background proteins from the nuclear fraction of f-XRNAX.***

A) Total number of identified proteins per replicate in the -UV and +UV conditions of nuclear f-XRNAX experiment are indicated.

B) Volcano plot of LFQ intensities (log2) against adjusted p-values (-log10) between the +UV and -UV conditions of nuclear f-XRNAX. Proteins with a fold-change higher than 1 and adjusted p-value lower than 0.01 are categorized as significantly enriched in the nucleus (yellow = RBP annotated, black = not annotated as RBP). The rest of proteins (gray) were removed from the analysis. Some known RBP are indicated.

discovery rate and calculate adjusted p-values. Proteins were called to be enriched in the +UV condition of a specific cellular fraction if their log2 fold change was greater than 1 compared to the -UV condition and their adjusted p-value lower than 0.01. Figure 14 shows two steps of the analysis to remove background proteins from the nuclear fraction: on the left, the total number of identified proteins per replicate is depicted. This number remains consistent in the +UV condition, while it is more variable in the -UV one. On the right side, the volcano plot shows the enriched proteins in the +UV condition, with the thresholds for LFQ intensity fold change and adjusted p-values. Most of the proteins removed from background are annotated as RBPs, with some examples indicated. The mitochondrial RBP LRPPRC is not a nuclear protein and was, as expected, filtered out from the analysis. We applied this filtering process to all the different conditions. In this way, we obtained a high-confidence datasets of localized RBPs in cytosolic, membrane and nuclear fractions of HEK293 cells.

The total number of proteins detected in each replicate after filtering and +UV enrichment analysis is displayed in Figure 15A. In the membrane condition, we consistently identified a higher number of proteins (>1700) compared to the other samples, even the total. This can be explained by the fact that the performed fractionation specifically enriches for a cellular compartment, making the detection of proteins with weak or few RNA interactions easier, which would not be identified in the analysis of whole cell lysate (see later section on a more in-depth analysis on the membrane-enriched RBPs).
The correlation plot of LFQ intensities in Figure 15B shows a very good agreement between replicates of the same condition (> 0.9 Pearson correlation coefficient), while the value decreases between different samples. As expected, the correlation of the 'total' sample versus other conditions is slightly higher, because most proteins identified in the total sample should also be present in cytosol, membrane and nucleus.
The PCA analysis of the 500 most variable proteins shows also a clear separation of the different conditions and cluster of the replicates for the same sample (Figure 15C), emphasizing reproducibility and enrichment for different sets of proteins in the respective subcellular locations.
Examples of proteins with RNA-binding annotation and known localization preference are plotted in Figure 15D: the centered LFQ intensity distribution (*i.e.*, mean samples' intensity set to zero) between the analysed compartments for GAPDH, LRRC59 and NAT10 proteins reflects their expected cellular distribution, with an enrichment in the cytosolic, membrane and nuclear fractions, respectively. This finding suggested that differences between samples' enriched proteomes is based on the cellular fractions obtained, and their purity was recapitulated in the high-throughput mass spectrometry experiment.
Interestingly, the protein isoelectric point (pI) distribution is comparable to published data (Kurotani et al., 2019), with acidic cytosolic polypeptides and more basic nuclear ones. While RBPs captured with

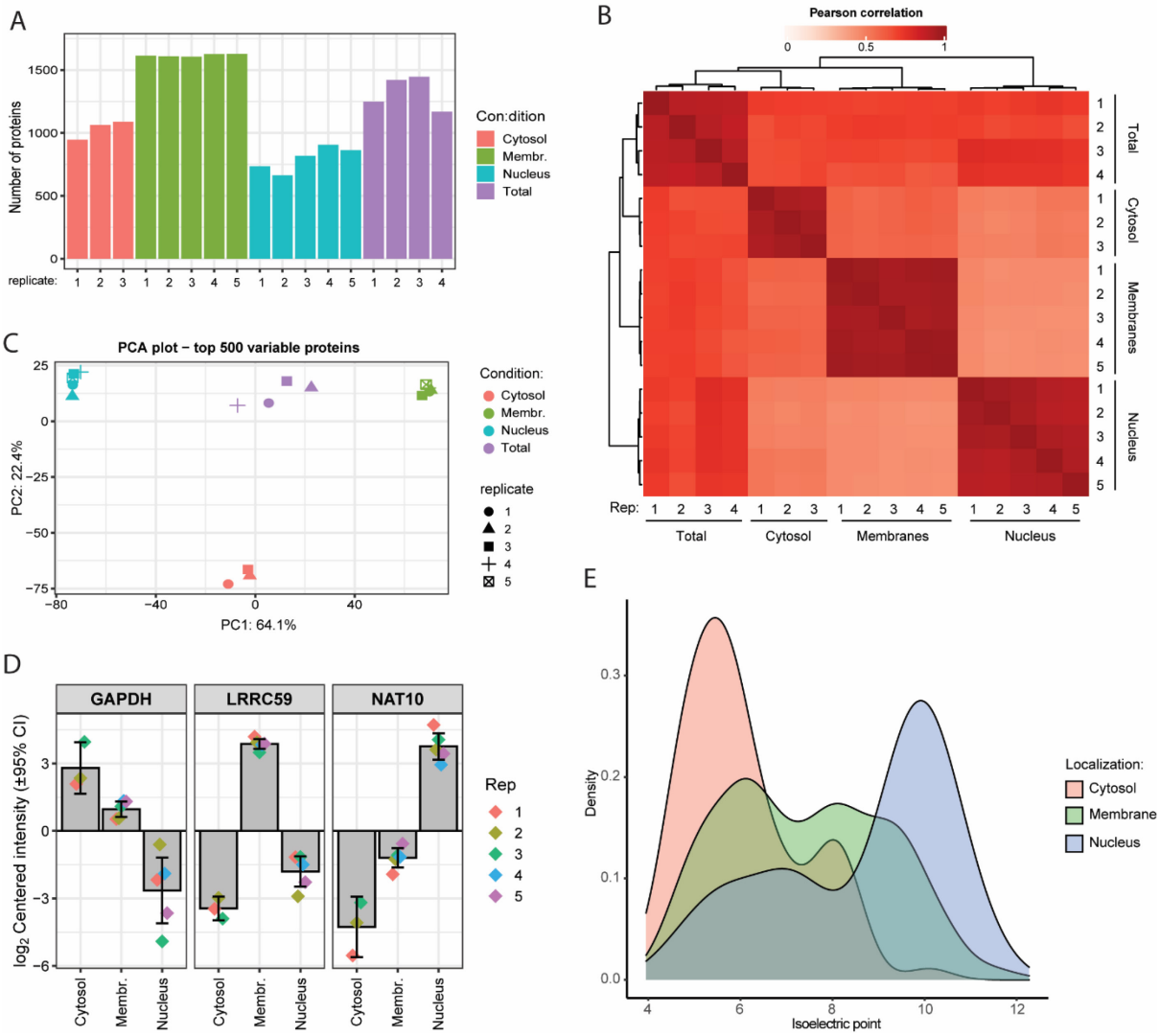

Figure 15. ***Compartmentalization of RNA-binding proteins is detected by mass spectrometry data analysis.***

A) Total number of proteins detected per in the replicates analyzed for the samples after background proteins filtering (+UV vs -UV comparison).

B) LFQ intensities correlation shows high reproducibility between replicates of the same sample.

C) PCA plots of the 500 most variable proteins.

D) Distribution of LFQ intensities (centered on the mean, gene-wise) between three RBPs with a preferential localization in human cells.

E) Isoelectric point distribution for proteins enriched with f-XRNAX and separated by their compartment localization.

RIC methods show on average a high pI (Beckmann et al., 2015; Castello et al., 2012), novel approaches have demonstrated that many proteins binding to RNA can also display a lower isoelectric point (Urdaneta et al., 2019)

Protein enrichment in one compartment respect to another, or co-localization between more than one fraction was determined using the R packages limma and eBayes and comparing conditions between each other. We used a mixed imputation approach (see Material and Methods) to determine the LFQ intensities in replicates with missing values. Proteins were compared pair-wise in two conditions and called to be enriched in one compartment if the log2 fold change was greater than 1 and the adjusted

p-value (Benjamini-Hochberg correction) lower than 0.01. In this way, we created a dataset of proteins binding to RNA which are preferentially located in the cytosol, membrane or nuclear compartment in HEK293 cells. The proteins that did not pass these thresholds were assigned to have a shared localization between more than one compartment. In Figure 16A the assignment of proteins to the different cellular compartments is represented: 138 and 78 proteins specifically localized in the nucleus and cytosol, and most of them have a previous annotation as RNA-binding. An unexpected result are the 630 proteins assigned to the membrane compartment, of which the majority has so far no assigned RNA-binding potential: only 46% has been annotated as RBPs in other studies using HEK293 cells (Baltz et al., 2012; Trendel et al., 2019), compared to the 96% and 78% in the nucleus and cytosol, respectively. 173 proteins do not have a preferential localization in one of the three compartments, while there are 160 and 468 proteins which are shared between the membrane-nucleus and membrane-cytosol compartments, respectively.

The Gene Ontology terms enrichment analysis for the localized proteins revealed biological processes related to RNA metabolism and recapitulated the specific compartmentalization (Figure 16B). For example, the nuclear localized proteins have term related to rRNA processing and RNA unwinding,

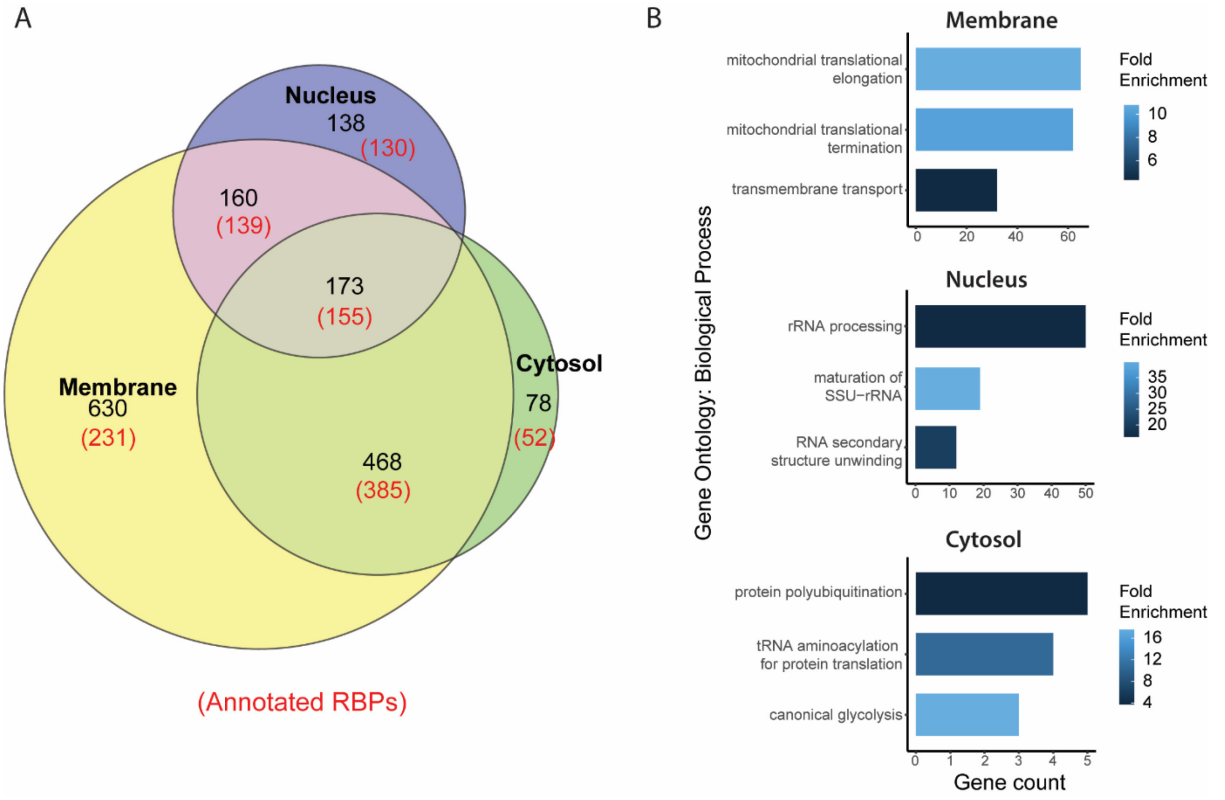

Figure 16. ***RBPome classification and characterization.***

A) Number of proteins distributed between and among the compartments studied with f-XRNAX. Total count of annotated RNA-binding proteins is indicated in red.

B) Gene ontology annotation for biological process for proteins found specifically enriched in cytosol, membrane and nuclei in the f-XRNAX experiment. P-value for the terms shown are < 0.01.

while the membrane localized ones are more associate to translation in the mitochondria and transmembrane transport (Figure 16B).

RNA-binding domains profiling

After the identification and compartment-classification of proteins enriched with f-XRNAX, we focused on the determination of the structural properties of the identified polypeptides. The characterization of the protein domain guiding the contact with ribonucleic acid molecules (the so-called RNA-binding domain or RBD) is of particular interest for RBP classification.

The peptides detected by mass spectrometry following our experimental approach provided additional information about the protein domains interacting with RNA *in vivo*. In fact, a minority of peptides were identified as crosslinked to RNA nucleotides and can be used to better characterize the location where the protein-RNA contacts occurred. Two ribonucleotides have been previously identified to be UV-crosslinked to peptides: cyclic-uridine monophosphate (cU) and uridine monophosphate (U) (Trendel et al., 2019). It is possible to detect these ribonucleotide-peptides adducts in mass spectrometry analysis by searching for peptides with a mass shift corresponding to the modification of cU and U (324 and 306 Dalton, respectively. See Material and Methods). We identified 175 and 57 peptides conjugated to cyclic uridine and uridine in samples irradiated with UV light, respectively, distributed in 73 and 29 proteins, with 26 proteins bearing peptides with both the modifications. Many proteins had multiple peptides with a crosslinked monophosphate nucleotide, like different heterogeneous nuclear ribonucleoproteins (e.g. HNRNPA1, HNRNPA2B1 and HNRNPA3), ribosomal proteins (e.g. RPL14, RPL27 and RPL4), translation regulating factors (e.g. EIF4B and EIF4H) and others. Analysis of the protein domain composition of the different peptides with crosslinked cyclic-uridine and uridine revealed a clear enrichment of protein domains related to RNA-binding, particularly of the RRM and S1 superfamily (Figure 17A).

The amino acid composition of peptides with crosslinked cyclic-uridine and uridine was also analysed and it showed an enrichment for the amino acid lysine (K) compared to the amino acid composition of peptides from total HEK lysate (Figure 17B). This result suggestedd a possible crosslink between the reported amino acid and the uracil base on the RNA in HEK293 cells, as it was already previously showed in another cell line (Trendel et al., 2019). Furthermore, based on the identification by MS/MS

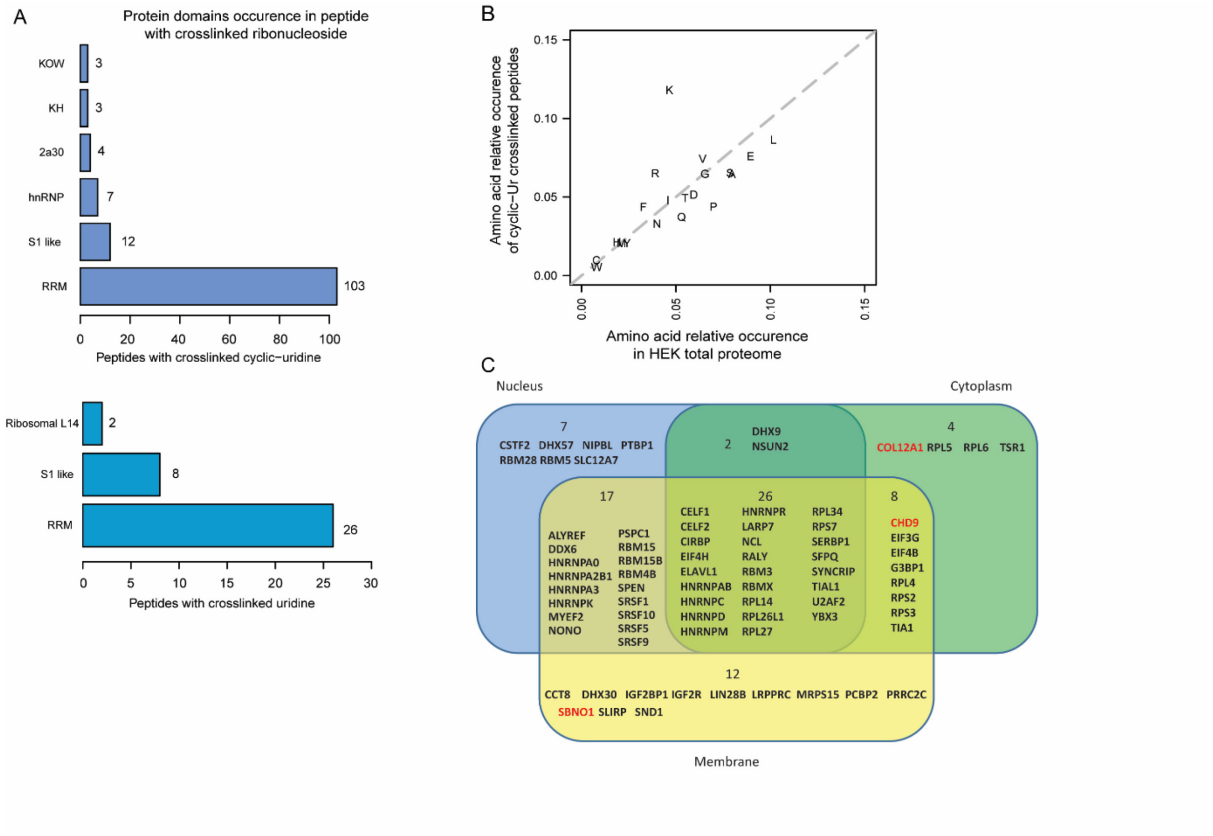

Figure 17 ***Peptides crosslinked to ribonucleotides are used to identify the domains of protein-RNA contacts.***

A) Number of peptides crosslinked to cyclic-uridine (above) or uridine (below) with an assigned protein domain.

B) Relative occurrence of amino acids in peptides with crosslinked cU or U respect to the HEK proteome.

C) Cellular distribution of proteins with at least one crosslinked ribonucleoside. Unannotated RBPs are shown In red

or by using matching between runs (Yu et al., 2020), the software MaxQuant assigns probabilities to which amino acid the cyclic-uridine / uridine was bound. According to these calculated probabilities, four amino acids are forming covalent contacts with RNA uridine nucleotides: lysine, glycine, valine and phenylalanine. Between them, lysine is the most represented in my dataset, with an enrichment of over 6 folds compared to the other amino acids, further supporting the evidence of a probable crosslink of the amino acid with nucleic acids.

Almost the totality of proteins containing peptides crosslinked to a ribonucleotide were already annotated as RBPs. Their cellular distribution following the f-XRNAX protocol is illustrated in Figure 17C.

In addition to the RNA-peptide adducts identified for a relatively small fraction of proteins, we aimed to retrieve additional information about RBDs by looking at the mass spectrometry data generated with our approach. The XRNAX protocol includes a partial tryptic digestion step before silica columns purification and complete tryptic digestion for MS ((Trendel et al., 2019), Figure 18A). The limited digestion is performed to reduce the mass of proteins attached to RNAs and to not clog the silica

columns, which purify peptides only if they are bound to nucleic acids. Because of the partial tryptic digestion, only the protein domains in the proximity of covalently attached RNAs are purified through silica columns and identified by LC-MS/MS. Therefore, we expected an enrichment of peptides close to the RNA-protein contacts. Moreover, peptides with crosslinked cyclic uridine or uridine could also be mapped to their respective protein regions and give more confidence about statements on RNA-protein contacts. Peptides from the input samples were used to normalize the data and reduce background signal (see Material and Methods). Thus, we obtained protein specific profiles of enriched peptides, which can be used to locate RNA-binding protein domains.

The profiles in the top panel of Figure 18B show the coverage of enriched peptides for three well known RBPs (FUS, YBX1 and RBM3), indicating their distribution along the primary sequence and the presence of crosslinked cyclic-uridine / uridine. Each profile is compared to the data from pCLAP, a protocol designed to specifically enrich RNA-binding domains applied in HEK293 cells (Mullari et al., 2017) and to the protein domain annotation from UniProt (Consortium et al., 2020). The clustering of enriched peptides for the analysed proteins correlates well with the data from pCLAP and it matches the areas where RBDs are annotated in UniProt. Our approach detected enriched peptides corresponding to the cold shock domain of YBX1, which are missed by pCLAP. The C-terminal portion of the protein is consistently enriched in both approaches, emphasising a potential RNA-binding

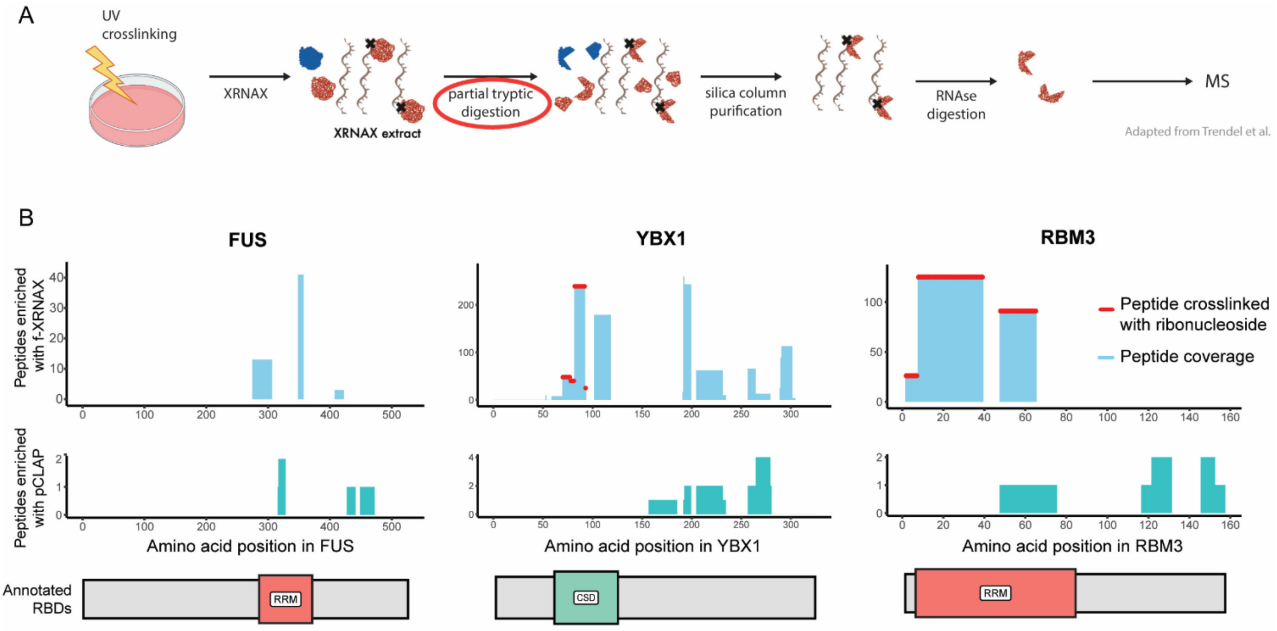

Figure 18. ***Enriched peptides profiles from f-XRNAX recapitulate the position of known RNA-binding domains in annotated RBPs.***

A) Scheme of the XRNAX protocol coupled to mass spectrometry (MS). The partial tryptic digestion is highlighted, the critical step leading to an enrichment of peptides in close proximity of RNA-protein contacts after silica column purification.

B) Top panel: examples of enriched peptides profiles for the three human proteins FUS, YBX1 and RBM3 obtained with f-XRNAX and normalized with input sample. Enriched peptides are in blue and regions with crosslinked uridine or cyclic-uridine monophosphate are in red. Peptides detected by the RBD-specific enrichment protocol pCLAP are shown in the middle panel. UniProt protein domain annotation in the bottom part of the figure.

domain of YBX1 not annotated as such. The presence of peptides with crosslinked ribonucleotides supports the confidence of calling a region of the protein binding to RNA and it correlates with the RBDs annotated by UniProt (as seen in the previous analysis and illustrated in Figure 16A).

The clustering of enriched peptides on and in the vicinity of RNA-binding domains can also be analysed in relation to the cellular compartmentalization of the protein in question. The peptide profiles showed in Figure 18B can be separated according to how much the cytosolic, membrane and nuclear compartments contribute to its abundance. In this way, it is possible to gain information about how much a protein binds to RNA in the different compartments and, more specifically, which protein domains are more involved in these interactions. The spatial and domain protein-specific information can then be used to investigate if proteins have a preferential domain usage depending on the compartment where they are localized. Even if the data is noisy for low abundant proteins and a global conclusion cannot be drawn from this analysis, we found examples where RBDs of the same protein are preferentially used in one cellular compartment respect to another. In Figure 19 is illustrated the distribution of enriched peptides for the two paralog proteins ELAVL1 and ELAVL2, according to their

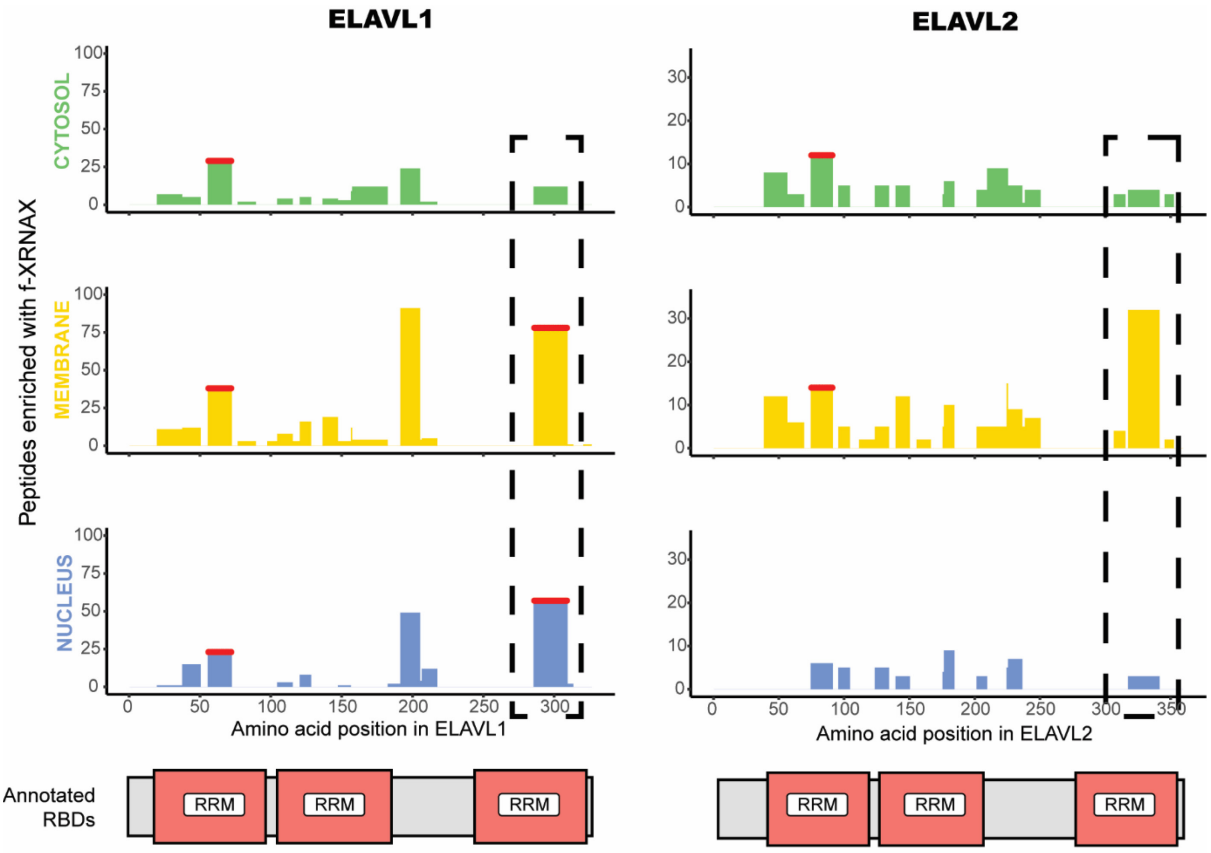

Figure 19 ***Compartment-specific RBDs differential binding in ELAVL1 and ELAVL2.*** *The enriched peptides profiles obtained with f-XRNAX for ELAVL1 and ELAVL2 are shown according to the protein cellular localization in HEK293 cells. In red are highlighted peptides that were found crosslinked with cyclic uridine or uridine monophosphate. The dashed black lines indicates the peptide coverage in the different compartment of the C-terminal RRM domain, showing a conserved enrichment in the membrane respect to the cytosolic compartment between the two paralog proteins.*

cellular distribution in HEK293 cells. The structural similarity between the two proteins is reflected in their protein domains composition and their enriched peptides profiles, but differences are present. The distribution of peptides between cytosol and membrane in the region associated to the more N-terminal RNA-recognition motif (RRM) domains is comparable. On the other hand, the C-terminal RRM domain is clearly enriched in the membrane (and nucleus for ELAVL1) compartment, while it is less represented in the cytoplasmic fraction. The fact these findings were reported for two paralog proteins points to a possible evolutionary conserved function of the RRM in ELAV1 and ELAV2, where the localization of the protein plays a role in its binding to transcripts.

Characterization of RBPs in the membrane compartment

The number of proteins identified by f-XRNAX in the membrane compartment without a previous RNA-binding annotation is higher than expected and for this reason was further scrutinized.

Figure 20A shows the volcano plot of LFQ intensity fold changes versus adjusted p-value for proteins in the membrane and nuclear compartments. Of note, 399 out of 620 membrane localized proteins does not have a previous annotation as RNA-binding protein (Figure 16A). These proteins have been identified in all five biological replicates of the +UV membrane fraction of the f-XRNAX experiment, so they are unlikely to represent background or experimental artefacts. The majority of proteins enriched

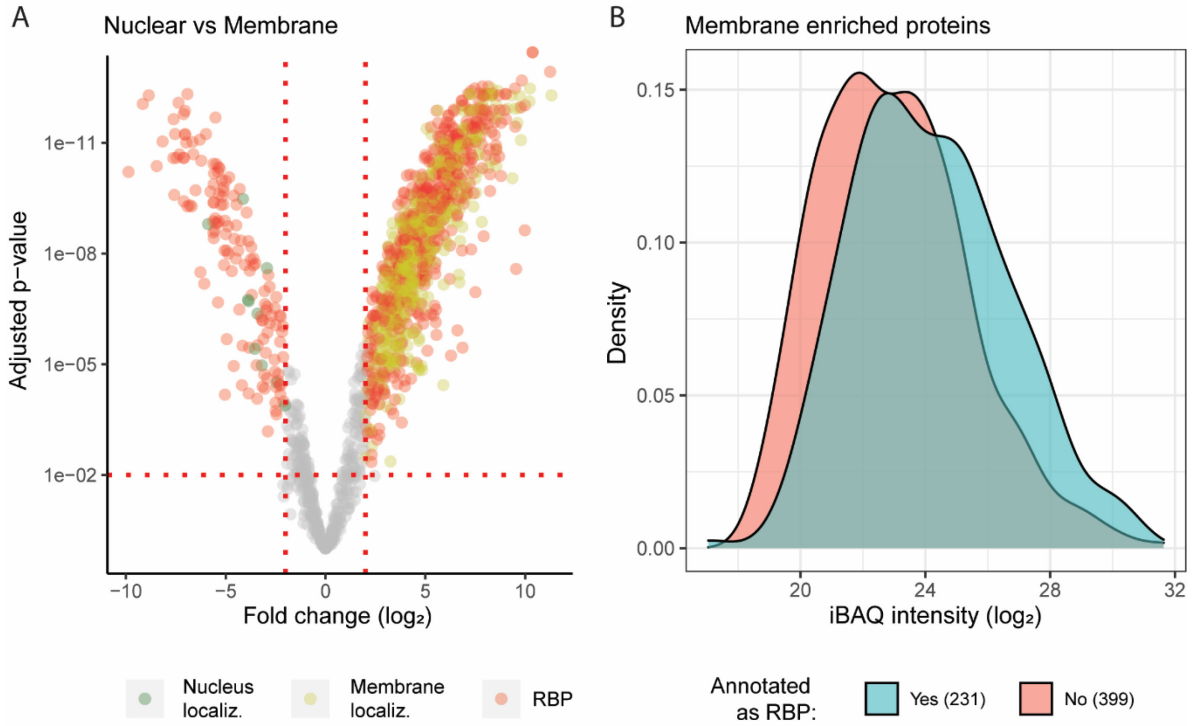

Figure 20 ***Not annotated RBPs compose a large fraction of the proteome in the f-XRNAX membrane fraction.***

A) Volcano plot of LFQ intensities fold change between nuclear and membrane compartments. Red-dashed lines delimit the thresholds set to call a protein specifically enriched to one compartment (adj. p-value < 0.01, |log2 (LFQ fold change) | > 1).

B) Density plot of iBAQ intensities for proteins enriched in the membrane fraction, colored by RBP annotation.

in the membrane compartment is not glycosylated, a post-translational modification that resemble the physical-chemical properties of RBPs crosslinked to RNAs during TRIzol extraction. Furthermore, if a protein would be capture due to glycosylation, it should be detected in the -UV condition too and filtered during computational analysis.

However, the iBAQ intensities distribution of the non-annotated RBPs on the membrane fraction is overall lower than the one for the annotated ones (Figure 20B), indicating less stable RNA-protein interactions. These contacts could be transient and captured only upon the compartment enrichment by fractionation performed in f-XRNAX. Nonetheless, their presence in the membrane sample was remarkable and it contributed to 44% of the total iBAQ intensity of the sample.

In total, we identified 399 new potential RBPs that reside and bind to RNA on membranes of HEK293 cells. A GO term enrichment analysis for biological process revealed how these proteins are associated with different types of transport, particularly of ions and molecules (Figure 21A). The protein domain composition of the membrane fraction proteins shows a significant increase in transmembrane and signal peptide containing domains proteins not annotated as RBP (Figure 21B). In fact, in this class of proteins enriched with f-XRNAX on human membranes, 71 belong to the solute carrier (SLC) superfamily and 21 to the TMEM family, both of which are groups of proteins with known transmembrane domains, and both with localization at the ER and plasma membranes. Some of the components of the SLC and TMEM families have already been described to bind RNAs (SLC3A2,

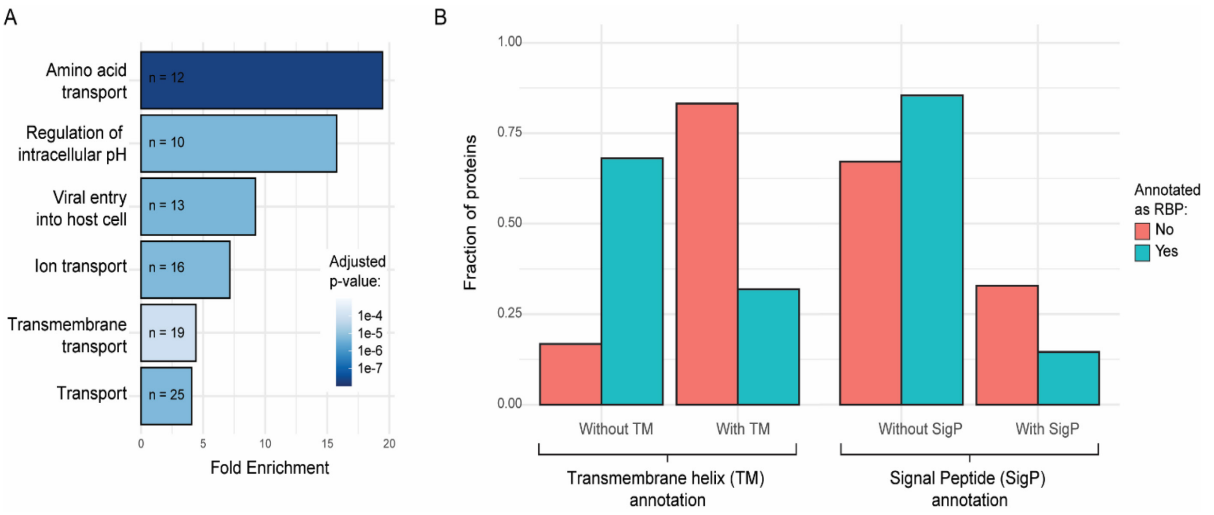

Figure 21 ***Characterization of not annotated RNA-binding proteins at the membrane compartment.***

A) Gene Ontology term enrichment analysis for biological function for proteins detected at the membrane compartment with f-XRNAX lacking RBP annotation. Fold enrichment is indicated on the abscissa and GO term name of the top six enriched terms on the ordinate. Adjusted p-value (Benjamini-Hochberg correction) is shown by the color bar.

B) Protein domain distribution of transmembrane helices and signal peptides for proteins found in the membrane compartment of the f-XRNAX experiment. The groups of protein with and without annotation as RBPs are compared.

SLC25A3, TMEM165 and TMEM214 for example (Baltz et al., 2012; Trendel et al., 2019)), but many more proteins seem to have this potential according to the f-XRNAX data.

Although transmembrane domains and signal peptides are more represented in these RBP candidates, protein domain enrichment analyses to find a common motif which binds RNAs did not lead to any conclusive results. One possibility is that these proteins do not have a structurally stable domain which binds to nucleic acid, but they might have intrinsically disordered regions (IDR) with the potential of interacting with RNA. In fact, in the last decade, there have been several works where the role of IDR in RNA-binding has been highlighted and discussed (Basu and Bahadur, 2016; Protter et al., 2018; Varadi et al., 2015).

Using the MobiDB database (Piovesan et al., 2020), we defined the number of proteins with IDRs in the f-XRNAX data set. We found that proteins in the membrane fraction without RBP annotation are much more likely to contain IDRs (Figure 22A), while for proteins in the cytosol or nucleus this is not the case. We subsequently compared the position of IDRs within each protein to its peptide profile from our data. For many of the non-annotated RBPs at the membrane compartment there was a

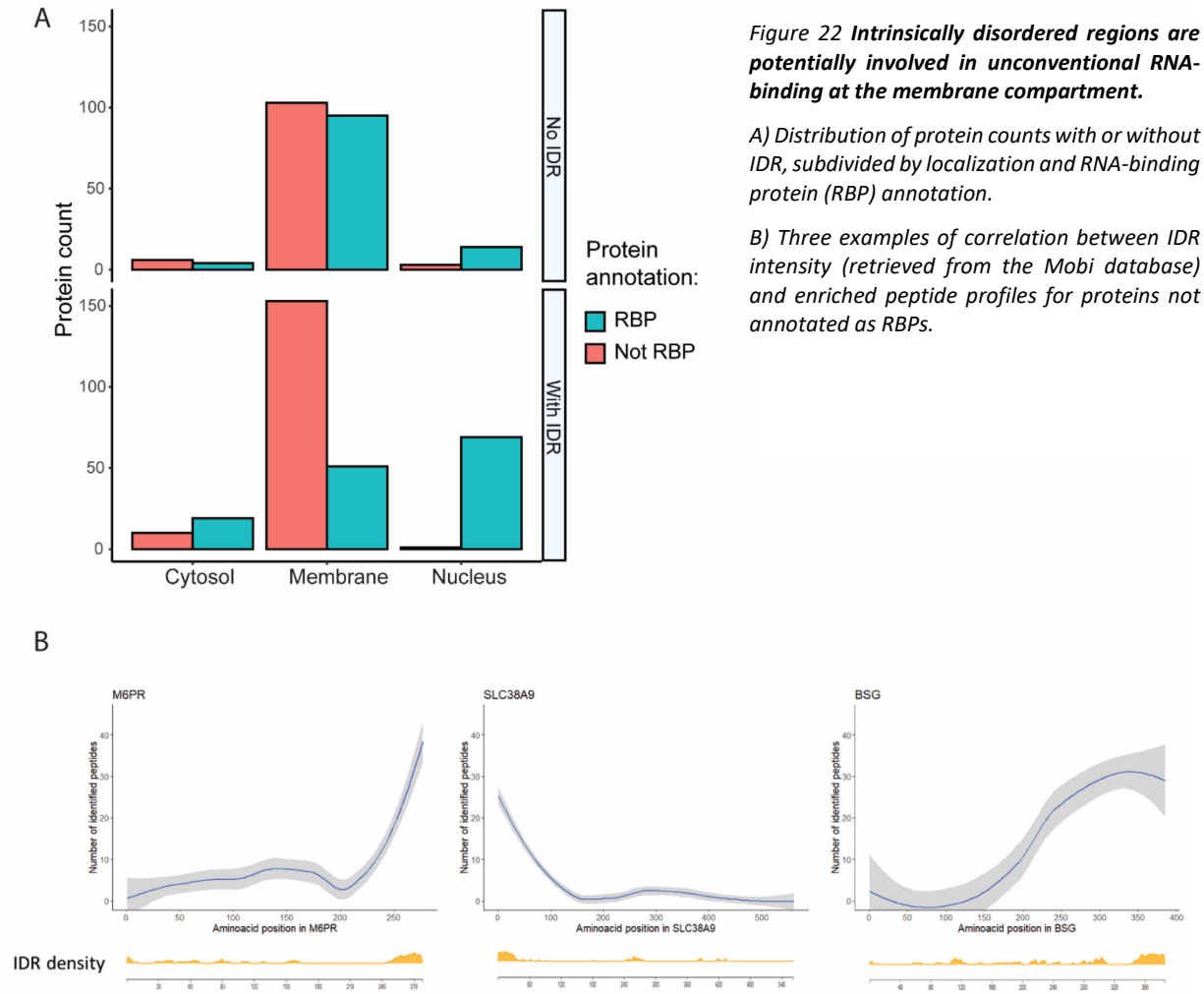

Figure 22 ***Intrinsically disordered regions are potentially involved in unconventional RNA-binding at the membrane compartment.***

A) Distribution of protein counts with or without IDR, subdivided by localization and RNA-binding protein (RBP) annotation.

B) Three examples of correlation between IDR intensity (retrieved from the Mobi database) and enriched peptide profiles for proteins not annotated as RBPs.

correlation between the probability of IDRs and the number of f-XRNAX detected peptides. As shown in Figure 22B, the number of identified peptides for the protein M6PR clusters with the annotated IDR regions towards at its C-terminus. Similar observations can be made for the proteins SLC38A9 and BSG (Figure 22B). These results suggest a possible function for intrinsically disordered regions in contacting RNA, particularly in the membrane compartment of human cells.

Differential RNA-binding of proteins in different compartments

The proteins identified by f-XRNAX can be found co-localizing in more than one of the compartments analysed (Figure 16A). Nevertheless, the f-XRNAX data does not take into consideration the total amount of protein present at the different cellular compartments (binding + not binding to RNA). We took advantage of the mass spectrometry data generated by the "input" samples (cellular fractions before the XRNAX purification) to normalize the f-XRNAX data and account for protein abundance in the different fractions. In this way, we obtained protein-wise normalised values of LFQ-intensities that could be used to evaluate the different RNA-binding potential of different proteins in the respective compartments. In fact, if a protein is reported to be localized aspecifically between the membrane and cytosol in the f-XRNAX data, but it is also described as having a greater abundance in the membrane "input" before the RBP-enrichment step, this can represent a higher potential of that protein to bind RNA in the cytosol compared to the membrane compartment. Figure 23 shows the results of this analysis between the nucleus and membrane compartments and indicates how the same protein can bind to RNA with different affinity depending on its localization. The protein taking into consideration for this analysis had to be found in both analysed fraction in the f-XRNAX experiment and in the input fractions. Also, the difference of protein abundance (measured in log2 LFQ intensities) between the two compartments in the input samples was set to be not larger than 2. In this way, it was avoided to compare proteins which localization in HEK293 cells is severely skewed to one compartment.
Therefore, the f-XRNAX approach and the data generated can used to evaluate the differential RNA-binding of RBPs in multiple compartments.

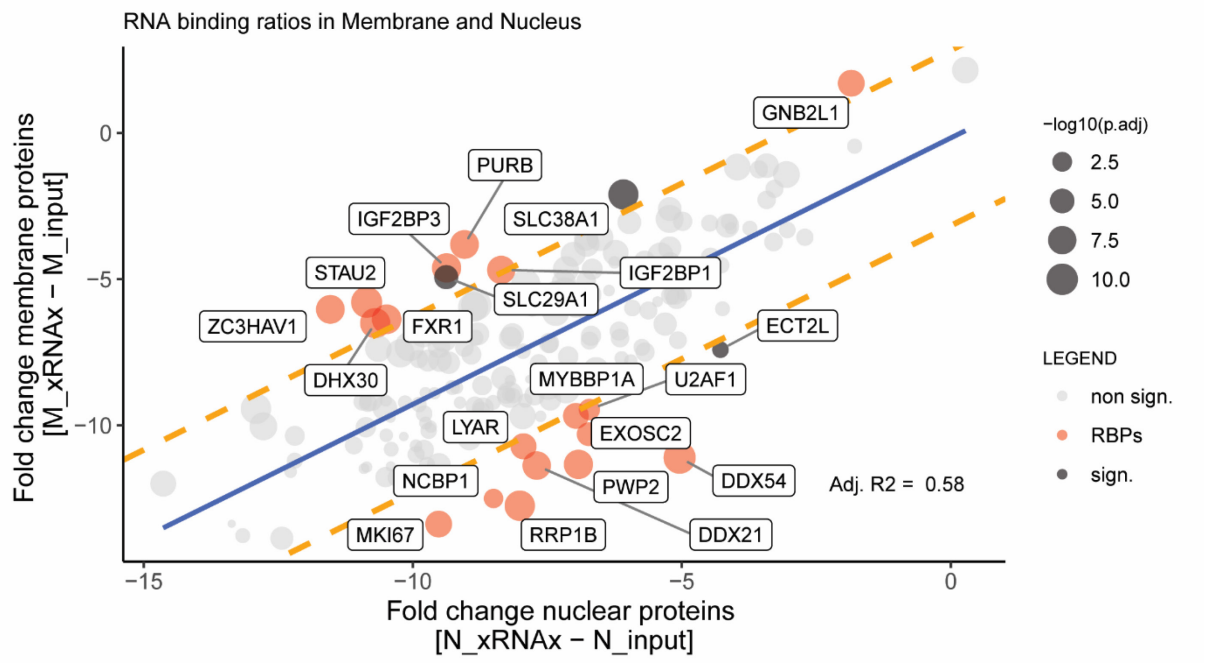

Figure 23 Proteins detected in more than one compartment show differntial binding to RNA depending on their localization. Proteins identified in both nuclear and membrane compartments of f-XRNAX experiment are plotted. LFQ intensities are normailzed protein-wise by the intensity from the fraction-specific input MS data.

RNA dynamics in HEK293 single cells

RNA expression is coordinated by the rates of its synthesis, processing and degradation. We set up a novel experimental and computational approach to investigate these parameters at the single cell level. We focused on their temporal dynamics during a fundamental biological process: the cell cycle.

Development of the SLAM-Drop-seq protocol

RNA metabolic labelling in living cells using ribonucleotide analogues is a widely applied method to study gene expression kinetics in cell populations (Baptista and Dölken, 2018). For a better understanding of the temporal dynamics regulating RNA production, splicing and degradation at the single cell level, we developed a new approach, based on the published SLAM-seq protocol (Herzog et al., 2017) and on microfluidic systems for single cell sorting (Davey et al., 2021; Macosko et al., 2015). In SLAM-seq, cells are incubated for a determined amount of time with the nucleotide analogue 4sU, which is incorporated into newly synthetized transcripts; RNA is then purified and treated with iodoacetamide (IAA). This last step induces the alkylation of 4sU residues and leads to the incorporation of T to C transitions during the reverse transcription step of RNA-seq library preparation. These transitions are detected and quantified by computational analysis, identifying the 'new' and 'old' populations of transcripts in the sample.

We wanted to apply the tracking the temporal dynamics at the single cell level. In order to do so, we introduced cellular fixation with methanol after 4sU labelling of HEK293 cells, similar to the one used in some scRNA-seq protocols (Alles et al., 2017). Next, we optimized the conditions for the RNA alkylation with iodoacetamide within fixed cells. In SLAM-seq, purified RNA is treated with IAA for 5 minutes at 55 °C, in slightly basic buffer containing 50% DMSO. However, we had to achieve efficient alkylation in cell suspensions, while maintaining the integrity of the individual cells and their transcripts.

Several conditions were tested. We optimized the temperature, time, pH and composition of buffers, in which the alkylation reaction occurred. As a readout for the efficiency of 4sU alkylation, we set up a "biotinylation blocking assay" (Figure 24A). The thiol group of 4sU can react with the commercial compound MTS-XX-biotin (Duffy et al., 2015), which can subsequently be conjugated with streptavidin-HRP and detected by chemiluminescence. In case 4sU was alkylated by IAA, MTS-XX-biotin would be unable to react with free thiol groups, resulting in a decrease of chemiluminescent signal by streptavidin-HRP. Using this semi-quantitative assay, the ratio between the chemiluminescent readout of IAA treated versus untreated samples determined the efficiency of alkylation in the different applied

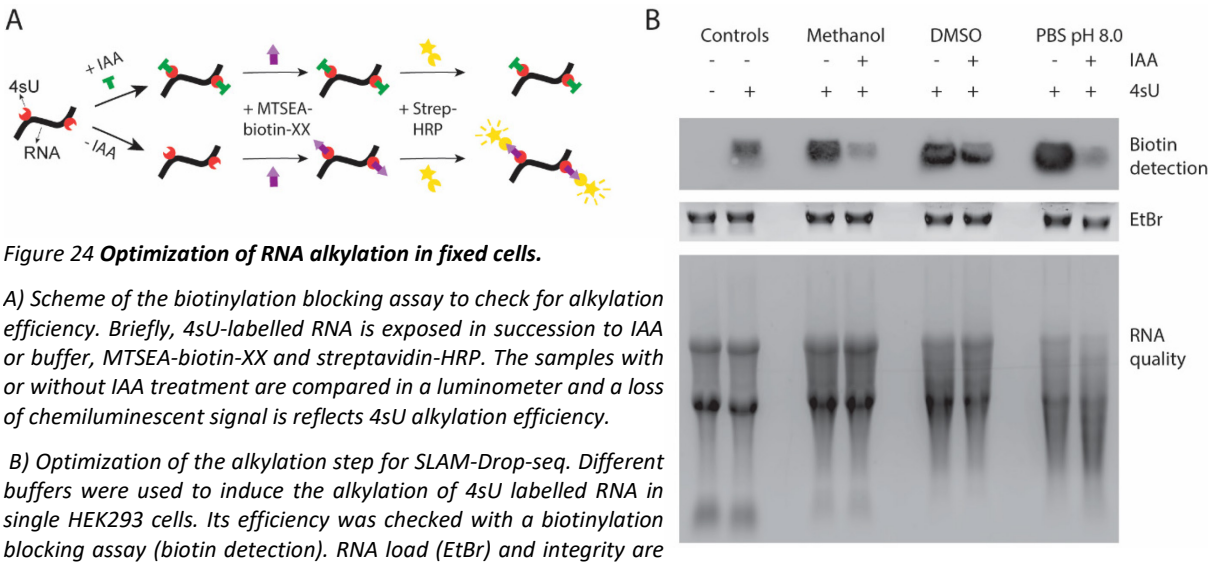

Figure 24 ***Optimization of RNA alkylation in fixed cells.***

A) Scheme of the biotinylation blocking assay to check for alkylation efficiency. Briefly, 4sU-labelled RNA is exposed in succession to IAA or buffer, MTSEA-biotin-XX and streptavidin-HRP. The samples with or without IAA treatment are compared in a luminometer and a loss of chemiluminescent signal is reflects 4sU alkylation efficiency.

B) Optimization of the alkylation step for SLAM-Drop-seq. Different buffers were used to induce the alkylation of 4sU labelled RNA in single HEK293 cells. Its efficiency was checked with a biotinylation blocking assay (biotin detection). RNA load (EtBr) and integrity are also showed.

conditions. Moreover, the RNA integrity was determined in parallel, since degraded transcripts reduce data quality.

Figure 24B shows three examples of conditions tested and the readout obtained in order to optimize the protocol: cells were incubated with 4sU, fixed with methanol and resuspended in the three different buffers indicated (80% methanol, 50% DMSO or DPBS pH 8.0) with or without IAA. Cells were then lysed and RNA extracted to perform the biotinylation blocking assay and check for efficiency of IAA-induced alkylation ('Biotin detection'). RNA integrity was assessed using agarose gels. The buffer with the best efficiency of alkylation and potential to keep RNA intact was the solution of 80% methanol and 20% DPBS (same composition of the fixation buffer). Several other conditions were tested, such as different concentration of IAA and 4sU, and timing of the alkylation, and finally we found the optimal ones for an efficient 4sU alkylation and purification of not degraded RNA, as described in the next paragraph.

The fixation of 4sU-labelled single cell suspension was performed in 80% methanol and 20% DPBS solution (*i.e.*, fixation buffer). As mentioned previously, this step was fundamental for the preparation of single cell RNA libraries, because it blocks cell metabolism but leaves cellular compartments mostly intact. In addition, these conditions inactivate RNases and permeabilizes cell membranes, allowing the IAA-induced alkylation of labelled transcripts to take place within fixed cells, a novel step which reduces the hands-on time and complexity of similar protocols (Qiu et al., 2020). After fixation, cells are incubated with 10 mM IAA (final concentration) in fixation buffer overnight at room temperature, with slight agitation and in the dark. The cell suspension is then washed, rehydrated and IAA quenched with DTT (free IAA would impair the function of reverse transcriptase and other enzymes in the downstream protocol). At this point, cells could be used for single cell encapsulation with one of the available microfluidic system. The newly developed approach has been termed "SLAM-Drop-seq".

SLAM-Drop-seq protocol was used on HEK293 cells, which were 4sU-labeled for 30, 60 and 120 minutes and compared to unlabelled cells. The efficiency of IAA alkylation was shown by the dot-blot analysis (Figure 25A). In RNA samples from cells where IAA was added, only a background signal was present, indicating efficient alkylation of 4sU residues. The quality of the RNA was checked with a 2100 Bioanalyzer Instrument (Agilent), showing a RIN value greater than 8 for all the samples.

To quantitively check the efficiency of 4sU alkylation in cell suspension, we constructed bulk poly(A)+ RNA-seq libraries. As expected, we detected increased T to C transitions when cells were incubated with 4sU and RNA alkylated with IAA, and this rise correlated with the time of 4sU labelling (Figure 25B). Nucleotide changes different from T to C did not show this correlation and remained constantly low (median around 0.0002 per gene, Figure 25B). More interestingly, our protocol generated a number of T to C conversions very comparable to the published SLAM-seq protocol, using purified RNA instead of cell suspension to perform the alkylation step (Figure 25C, green bars versus blue bars

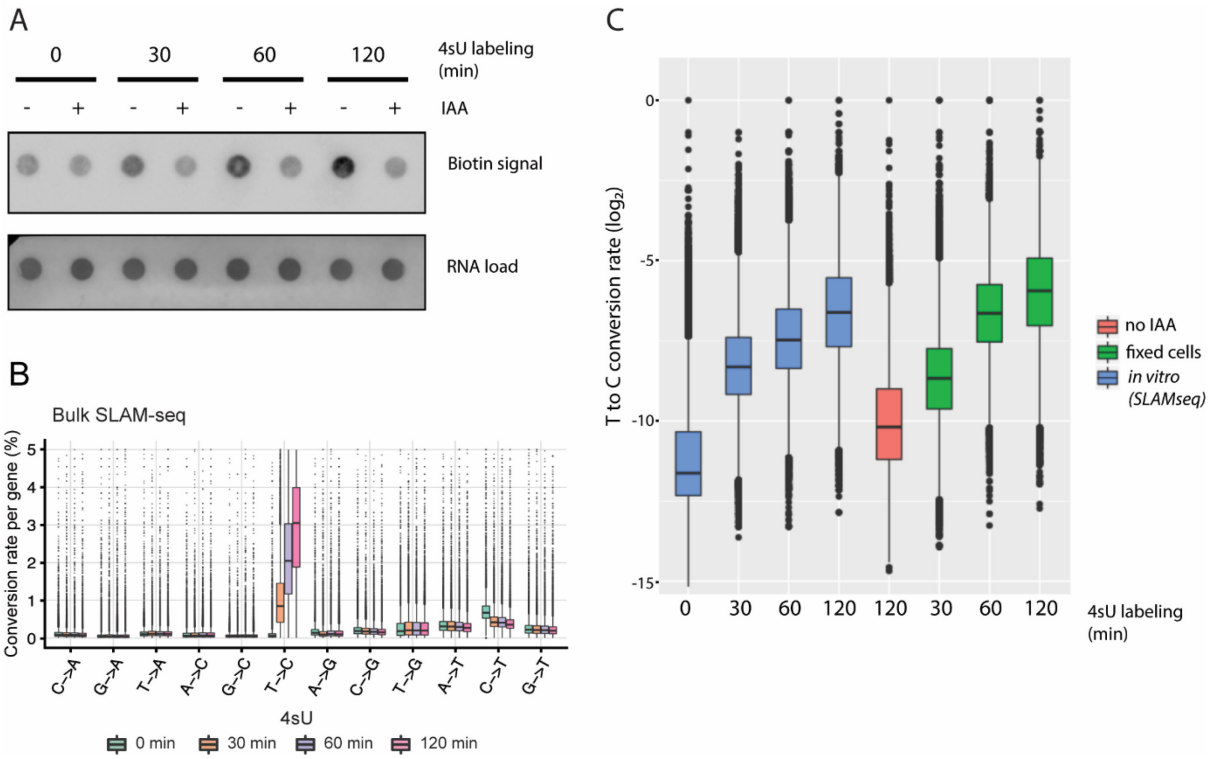

Figure 25 ***SLAM-Drop-seq efficiently alkylates labelled transcripts in cell suspension.***

A) Representative dot blot of biotinylation blocking assay showing the efficiency of 4sU incorporation with different labelling times and the efficiency of alkylation by IAA in fixed cell suspensions. Biotin signal was detected with streptavidin-HRP by chemiluminescence, methylene blue staining was used for RNA loading control.

B) Nucleotide mutations per genes detected in a bulk polyA+ RNA-seq experiment upon 4sU labelling following the SLAM-Drop-seq protocol. Only T to C conversions correlate with the incubation time with 4sU. The conversion rate is calculated by dividing the number of mismatches by the number of the original nucleotide (i.e., T to C conversion rate = number of T to C divided by total number of Ts).

C) T-to-C conversion rates of bulk SLAM-seq polyA+ libraries. RNA was alkylated with IAA either in fixed cells (green bars) or in vitro (blue bars, i.e. purified RNA was derivatized with IAA in vitro following the published SLAM-seq protocol); no IAA alkylation control (pink bar) served for estimation of T > C conversion rate caused by 4sU alone and to check the efficiency of IAA-derivatization in fixed cells.

comparison). This gave us the confidence to couple our protocol with microfluidics for single cell library preparation.

SLAM-Drop-seq detects newly synthetized transcripts in single cells.

The schematic workflow of the SLAM-Drop-seq protocol is represented in Figure 26. In our experiment, HEK293 cells were labelled for 0, 15, 30 or 60 minutes with 300 µM 4sU, fixed in methanol-based buffer and then collected for SLAM-Drop-seq. Two biological replicates were prepared for each time point. The alkylation efficiency of 4sU residues in cell suspensions was checked with a biotinylation-blocking experiment and the RNA quality assessed using a TapeStation device (Agilent). All samples showed a satisfactory IAA-induced alkylation of 4sU-labelled RNA and a RIN value higher than 8. We then used the Nadia platform (Dolomite Bio) to encapsulate single cell together with beads attached to oligonucleotides containing a poly(dT) tails, specific cellular barcode (CB, to distinguish the different cells captured) and unique molecular identifier (UMI, to distinguish the different RNA molecules captured within a cell). PolyA+ RNAs attached to the beads were purified, reverse transcribed and amplified via PCR. Transposase-driven tagmentation was used to fragment the libraries and ligate sequences necessary for deep sequencing. The 15 minutes 4sU-labelled samples were first sequenced on a NextSeq machine (Illumina), and later on NovaSeq instrument (Illumina).
As declared in the 'Statement of Contributions', the computational analysis related to SLAM-Drop-seq

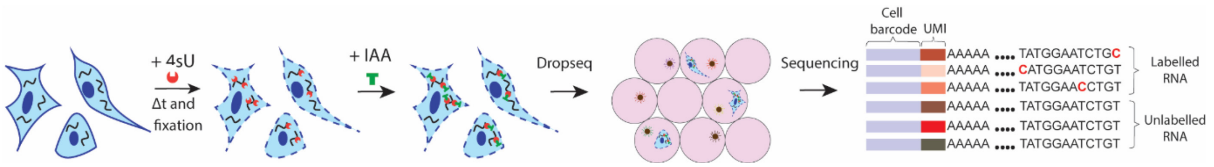

Figure 26 ***Scheme of the SLAM-Drop-seq experiment****. Briefly, HEK293 cells are incubated with 4-thiouridine (4sU) and fixed. 4sU residues in the newly synthesized RNAs are alkylated by IAA in situ, which results in T->C transitions during the reverse transcription step after encapsulation of single cells in oil droplets using Drop-seq or Nadia devices. Single-cell RNA libraries are prepared after droplet lysis and the T->C conversions in barcoded reads detected by computational analysis.*

was carried out by H. Liu, with help from M. Schilling and Dr. Schwabe.

We identified a total of 7280 single cells expressing at least 200 genes. The average gene expression correlation between the different replicates and time points was high (Pearson correlation coefficient $r > 0.99$, Figure 27A), indicating that 4sU incubation did not introduce any apparent change in gene expression.

To have a deeper sequencing coverage, we developed a novel method that computationally extends the reads containing the same UMI. As shown previously, 4sU is incorporated in 1 out of 40 uridines on average (Herzog et al., 2017; Jürges et al., 2018), which lowers the probability of detecting a T to C

transition due to a read length of only around 130 nucleotides. By merging reads generated from the same RNA molecule (*i.e.*, same UMI) we lowered the probability of such false negatives: our approach

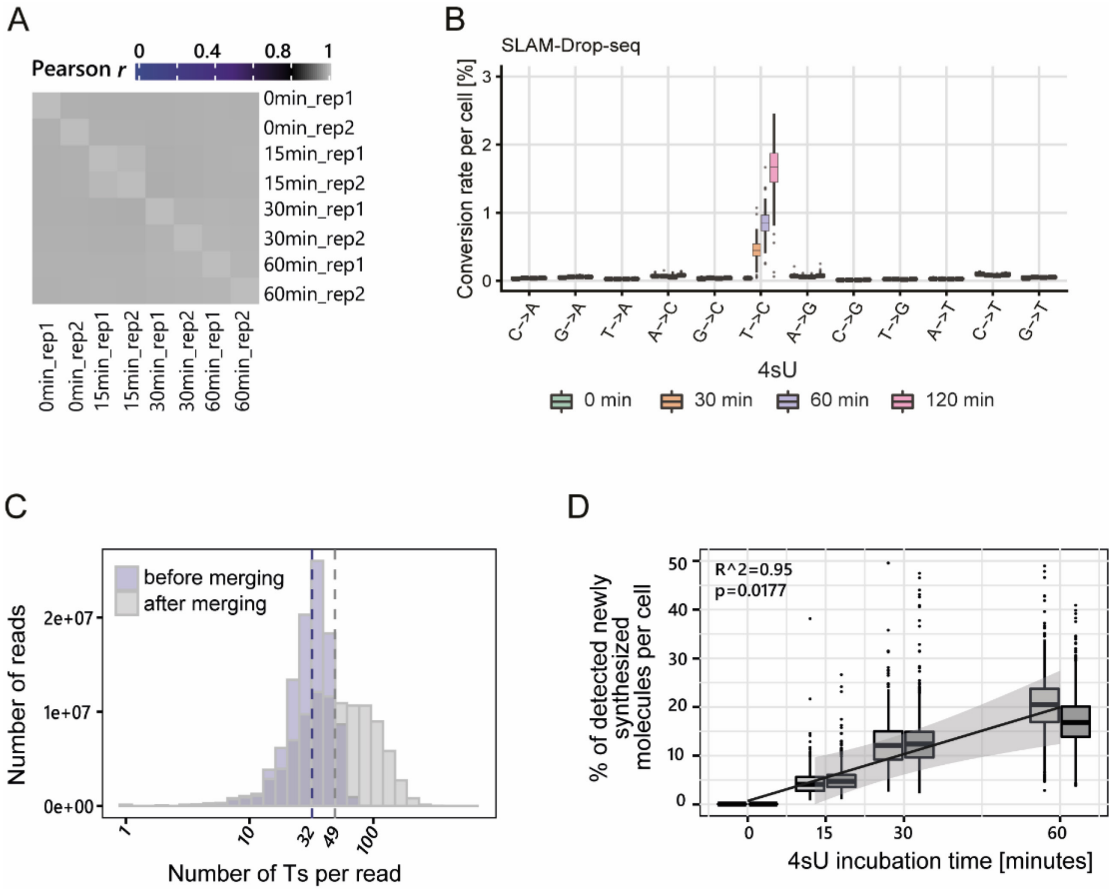

Figure 27 ***SLAM-Drop-seq identifies newly synthetized transcripts in single cells.***

A) Gene expression pairwise Pearson correlation analysis of single cell SLAM-Drop-seq experiment shows high reproducibility between replicates and samples. The four samples differ for incubation times with 4sU (0, 15, 30 and 60 minutes).

B) Conversion rate per cell for different mutational pattern following the SLAM-Drop-seq protocol upon different timing of 4sU incubation.

C) Histogram showing the increased number of thymidines (Ts) obtained by merging reads with the same UMI.

D) Correlation between the number of newly transcribed molecules per cell and the metabolic labelling time. The linear regression is based on the mean of the mean value between replicates of one time point. R^2 = coefficient of determination, p = p-value of the regression.

increased the median number of thymidines in a read by 52% (from 32 to 49, Figure 27C). As it was already shown for the bulk SLAM-seq data, the number of T to C transitions per cell increases proportionally to the time of incubation with 4sU (Figure 27B), while the other possible nucleotide changes were constantly low between the different samples. RNA molecules containing T to C transitions (*i.e.*, 4sU-labelled) were categorized as newly transcribed, while the other as pre-existent (*i.e.*, 4sU-unlabelled). We observed a linear correlation between the mean number of newly synthetized molecules per cell and the 4sU-labelling time (Figure 27D), meaning that diagnostic T to C conversions were detected with high sensitivity.

To calculate RNA kinetic rates, not only we had to separate the newly transcribed and pre-existing RNA populations, but also detect their splicing status, *i.e.* whether the sequenced reads contained intron sequences. Based on the 4sU-labelling and splicing status, we were able to assign each RNA molecule in each single cell to one of four RNA types (labelled-spliced, labelled-unspliced, unlabelled-spliced, unlabelled-unspliced), instrumental for the calculation of the RNA kinetic rates.

Cell cycle reconstruction

To give a temporal value to the calculation of the RNA kinetic rates, we focused on a dynamic biological process. Since we applied the SLAM-Drop-seq protocol to a population of unsynchronized and exponentially proliferating HEK293 cells, we decided to study changes of synthesis, processing and degradation during the cell cycle. In order to do so, we first assigned to each cell its corresponding cell cycle time based on their gene expression using Revelio (Schwabe et al., 2020). Figure 28 shows the expression profiles of known house-keeping and cycling genes as determined by Revelio.

Based on the cell cycle time assignment and identification of the four RNA types, we developed the

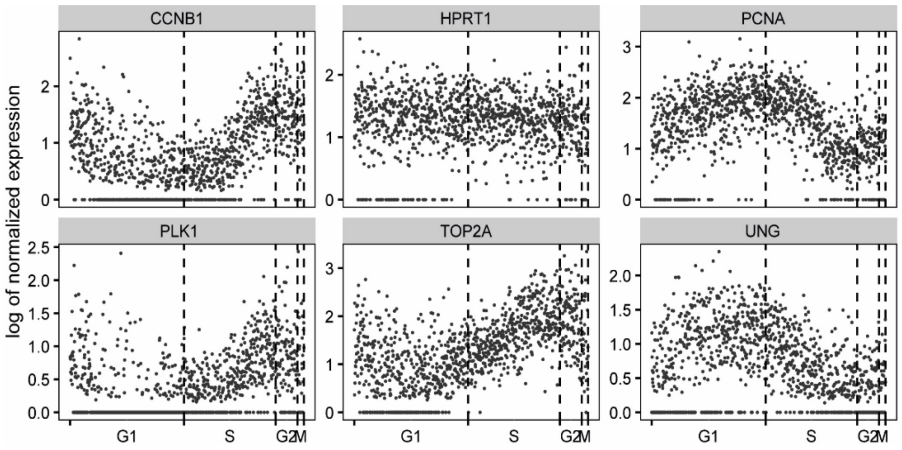

Figure 28 ***Examples of gene expression dynamics along the cell cycle for six genes****. Each dot represents a cell, assigned to that cell cycle time with Revelio. HPRT1 is a non-cycling house-keeping gene. UNG and PCNA are G1/S markers. TOP2A,CCNB1, PLK1 are G2/M markers.*

R-package Eskrate to calculate the time-dependent transcriptome-wide RNA kinetic rates. Eskrate solves a system of four ordinary differential equations (ODEs), describing the RNA abundance changes over the cell cycle time for the described four types of RNAs (labelled-spliced, labelled-unspliced, unlabelled-spliced, unlabelled-unspliced) and calculates their dynamic profiles along the cell cycle. Notably, we solved the time-dependent version of the ODEs: we thus obtained the estimation of kinetic rates as a continuous function of the cell cycle time. Since the amount of unspliced RNA in our data was relatively low, we simplified the kinetic model assuming that splicing is much faster than degradation (Alpert et al., 2017; Herzog et al., 2017). Eskrate was then able to estimate the cell cycle time-dependent RNA synthesis and degradation rates transcriptome wide from SLAM-Drop-seq data.

To investigate the regulation of temporal RNA kinetic rates during the cell cycle, we focused on genes whose expression oscillates along the cell cycle (*i.e.,* cycling genes). After some filtering steps, we identified 399 genes with expression peaking at a specific phase of the cell cycle (Figure 29A). The vast majority of them were well-predicted by comparing gene-wise predictions (from our kinetic model) to the observations (data from SLAM-Drop-seq experiment). Their estimated transcription and

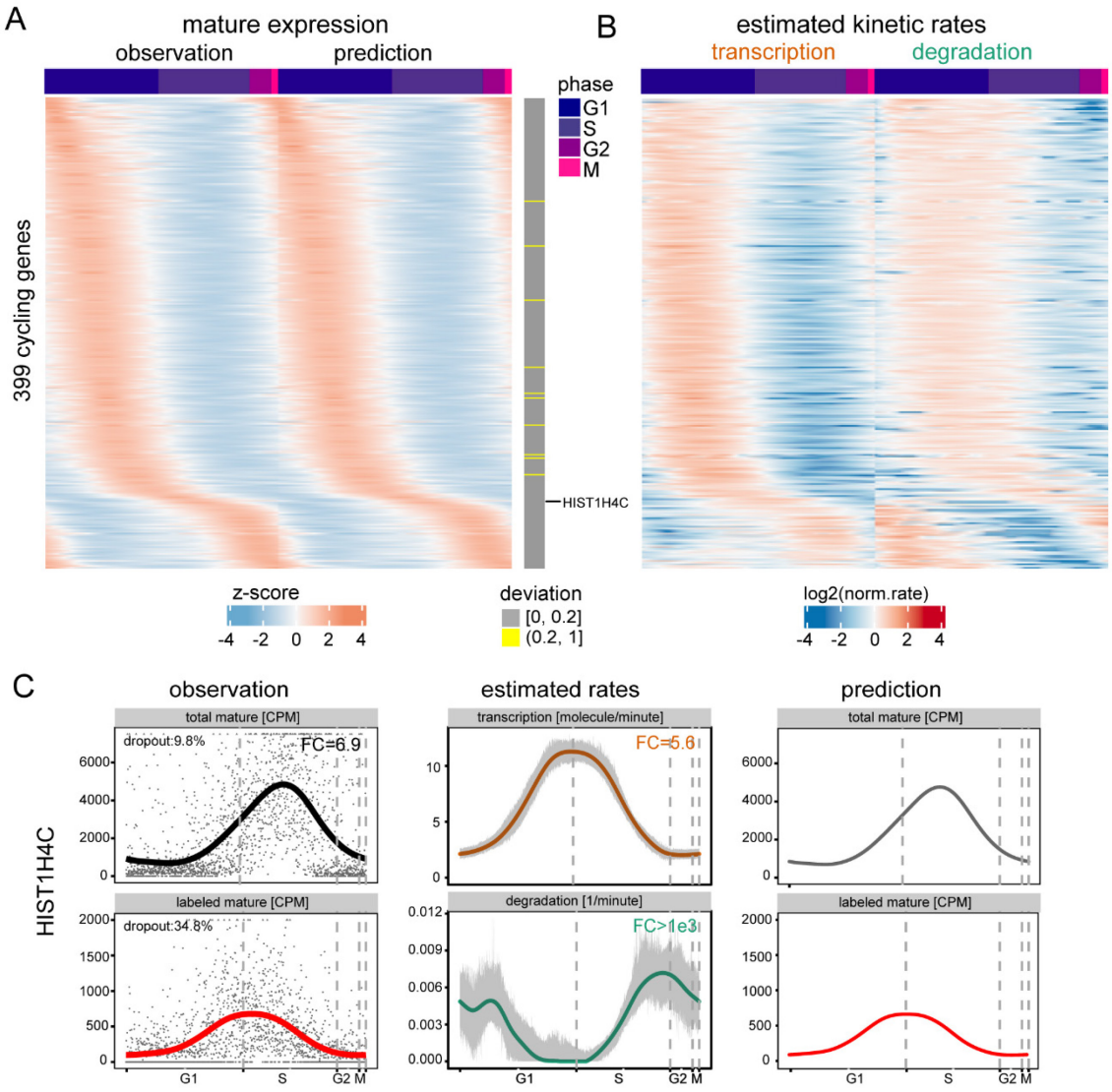

Figure 29 ***Gene expression and kinetic rates for cycling genes are highly dynamic along the cell cycle.***

A) Heatmap representation of the expression fluctuation throughout the cell cycle for 399 cycling genes (each row represents one gene). The 'observed' is the expression from single-cell sequencing, while the 'preticted' is the expression derived from our kinetic model. On the right the gene-wise deviation between observed and predicted is shown. A deviation lower than 0.2 classifies a gene as well-predicted. HIST1H4C is the gene displayed.

B) Gene-wise profiles of estimated transcription and degradation rates calculated by Eskrate.

C) Profiles of the observed gene expression (total and 4sU-labelled), estimated kinetic rates of transcription and degradation, and predicted expression for the HIST1H4C gene. On the left panels, each point represents the normalized read counts for each cell at the assigned cell cycle time. The points are smoothed to a profile. In the middle panel, the estimated transcription and degradation rates are displayed with the 90% confidence intervals in gray. CPM = counts per million; FC = fold change (ratio between max and min values of the profile).

degradation rates are also shown in Figure 29B, and display a dynamic regulation throughout the cell cycle.

To validate our estimations, we looked at the profiles from the histone gene HIST1H4C. This gene is a well-known S-phase marker and was previously reported to be transcribed just before its expression peak, while its stability decreases during the G2 phase of the cell cycle (Harris et al., 1991). These findings were confirmed from our estimates of HIST1H4C RNA kinetics (Figure 29C).

As a more global analysis, we also compared the estimates of our degradation rates to independent studies on RNA half-life (Murakawa et al., 2015; Schofield et al., 2018). Even if these publications calculated the degradation rates at steady state, we obtained positive correlation when comparing our results with them (Spearman correlation coefficients were 0.60 and 0.46, respectively).

Different regulation strategies for different cycling genes

To understand which component between transcription and degradation had a bigger importance in defining the expression profile of cycling genes, we classified them according to their dependency on the dynamics of transcription and degradation by checking the prediction changes upon constant kinetic rate setting. After filtering for well-predicted genes exhibiting too many peaks for their synthesis and/or degradation rates profiles (Figure 30A, see Material and Methods), we obtained a set of 377 genes which we called “robust-cycling”. We then fixed the transcription or degradation to a constant value (mean between all cells) and calculated how much the new predicted expression would differ from the original one. If a gene was highly depended on RNA synthesis to shape its expression along the cell cycle, this difference would be relatively large, and the same is true if the expression of a gene was more dependent on its stability. From this analysis, genes were classified in three different groups: regulated by dynamic transcription, dynamic degradation, or both (Figure 30B). The majority (337) of cycling genes showed a dependence on both kinetic rates to shape their expression peak, implicating that the regulation of RNA stability is highly involved in the cell cycle control of HEK293 cells. Moreover, we identified 16 and 21 genes applying a dynamic degradation or dynamic transcription strategy, respectively. As expected, the profiles (without fixing any value to a constant) of genes regulated by dynamic degradation show an approximately constant transcription rate throughout the cell cycle, while genes classified as regulated by dynamic transcription displayed a relatively steady time-dependent profile of degradation rate Figure 30C.

In conclusion, SLAM-Drop-seq coupled with the bioinformatical tool Eskrate enabled the quantitative and time-dependent profiling of synthesis and kinetic rates along the cell cycle of HEK293 cells.

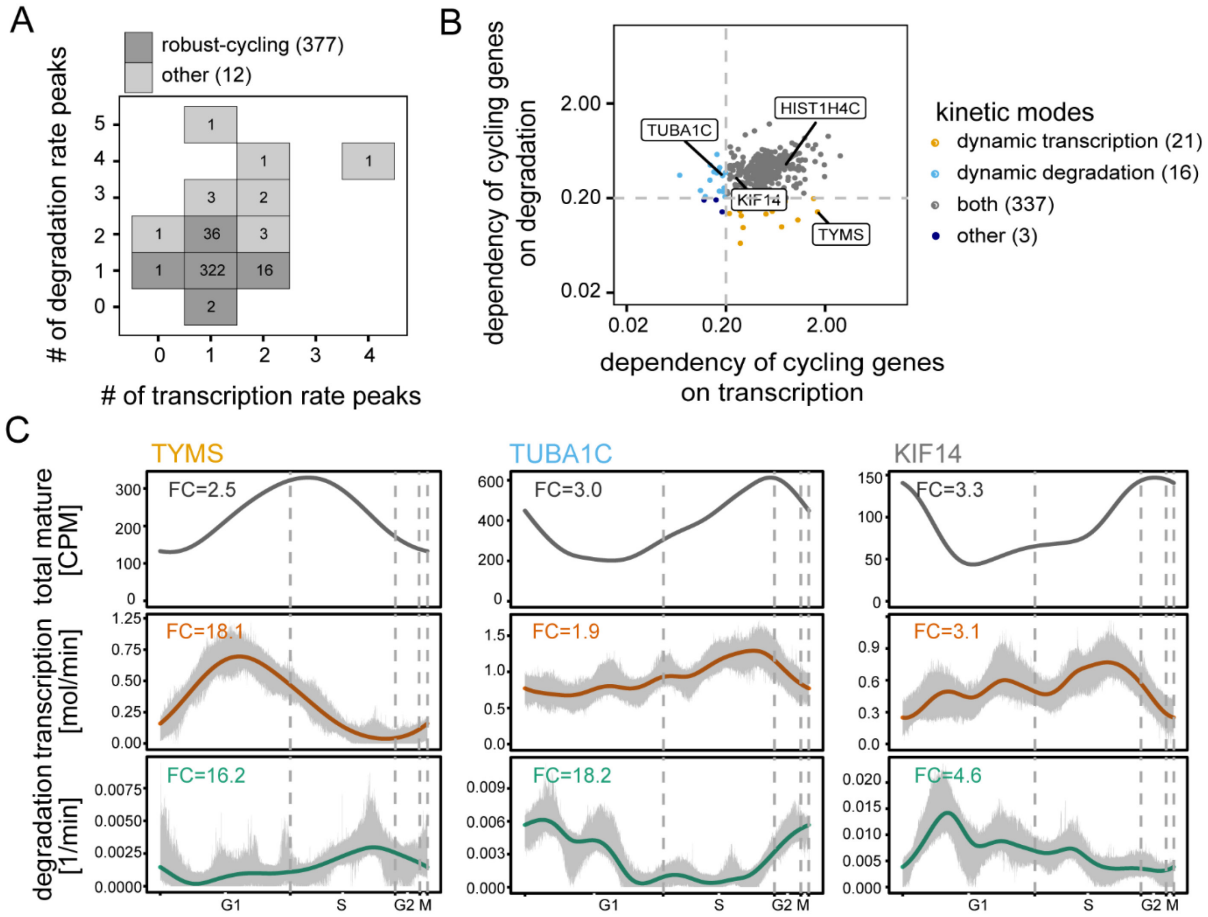

Figure 30 ***Different strategies regulate the expression of cycling genes.***

A) Representation of the number of transcription and degradation peaks in well-predicted cycling genes. In dark-gray are shown the set of genes which were selected for further analysis.

B) Robust-cycling genes are classified based on their dependency on dynamic transcription and dynamic degradation. By forcing either degradation or synthesis to a fixed value and after calculating the similarity (i.e., dependency) between old and new expression profiles, we classified the robust-cycling genes in three kinetic modes. Genes indicated are the one shown in the next figure.

C) Examples of expression, transcription and degradation rates profiles along the cell cycle for three genes belonging to the three different categories of kinetic regulation.

DISCUSSION

In this PhD thesis, several biological challenges have been addressed by implementing a combination of novel biochemical methods and computational approaches. The high-quality datasets created within this work describe protein localization, RNA-binding proteins compartmentalization and RNA kinetic rates temporal dynamics in HEK293 cells, and report some of the biological implications these findings are contributing to form.

The aim at the foundation of the developed approaches is to better describe the spatial and temporal determinants of gene regulation in human cells. Since many biological processes are involved in the fine-tuning of gene expression, we investigated different cellular mechanisms with a combination of newly developed high-throughput techniques. The identification of proteins and RBPs localization, and the profiling of time-dependent gene expression dynamics increased our understanding of the cellular environment and its regulation. From this information, new predictive models can be created on how eukaryotic cells maintain homeostasis and respond to stimuli.

Proximity labelling at the cytosolic side of ER, mitochondria and plasma membranes

We investigated the global localization of proteins in cellular regions difficult to assess with traditional high-throughput methods (such as cellular fractionation or differential centrifugation). We focused on the characterization of proteins localized at the cytosolic side of endoplasmic reticulum, outer-mitochondrial and plasma membranes. These subcellular regions play central roles in several essential biological processes. The endoplasmic reticulum membrane is involved in protein translation and secretion, lipid synthesis, calcium signalling, etc. The outer mitochondrial membrane mediates apoptosis, proteins import and interaction with other organelles. The plasma membrane is critical for the regulation of the flow of material, information and energy between intra and extracellular spaces. Despite their vital roles for a functional cellular environment, the protein composition at these subcellular regions is still not fully characterized, which would be essential for their in-depth understanding.

To detect the resident proteomes at the selected compartments, we created stable HEK293 cell lines expressing the promiscuous biotin ligase BirA* fused to specific targeting peptides. Mass spectrometry data analysis revealed that our approach reproducibly purified proteomes specific to one localization and confirmed the accuracy of the compartment-specific enrichment. We were able, for example, to recapitulate the expected localization of compartment-specific protein markers by looking at their intensities in the different conditions (Figure 6B).

The proteomes obtained with the different BirA* cell lines recapitulated the expected functional processes of the compartment where the enzyme was targeted. Furthermore, new candidate proteins have been identified which were previously unknown to localize in these cellular regions, mostly at the mitochondrial and plasma membranes. Therefore, our approach can be informative to further characterize and understand the proteomes located at the cytosol facing ER, mitochondria and plasma membranes.

We identified 479 proteins localized at the cytosolic side of the ER membrane, including many members involved in protein processing pathways at the ER. For example, the proteins composing the coat protein complex II (COPII) vesicles are well represented in our dataset: the activating and cargo-selecting SEC13/31 complex, the inner-coat SEC23/24 heterodimer and the SAR1A-acivating guanidine-exchange factor SEC12 are found enriched in the ER-BirA* dataset. The SAR1A protein however, which is also a component of COPII vesicles, was found to co-localize between ER and mitochondria. These results might reflect the function of SAR1A, which is freely diffusing in the cytosol when bound to GDP (*i.e.*, inactive) and only when activated by SEC12 exposes a hydrophobic tail, fusing to the ER membrane. The ambiguity of SAR1A localization and the numerous contacts between endoplasmic reticulum and outer mitochondrial membrane could explain why SAR1A is found in both of these compartments. This indicates that mixed localization are biologically meaningful. Still, more than 99% of the proteins enriched by ER-BirA* are annotated as components of the secretory-pathway, indicating a strong enrichment of the dataset created here.

More than 480 proteins showed a preferential localization to the cytosolic part of the plasma membrane. Many of them are known to localize at the plasma membrane or in its vicinity. This includes G-protein coupled receptors (like GRK6, GNA11, GNA13, GNAI2, GPRIN1 and GPRIN3 proteins), solute carrier proteins (like the leucine heterodimer transporter SLC3A2/ SLC7A5 (Nguyen et al., 2018), the glucose transporter SLC2A1, the zinc transporter SLC30A1, and 27 more), proteins associated to phospholipid transport and processing (PPAP2B, ATP11A, ATP11C and ATP8B2), and many others.

Proteins not integrated in the cellular membrane, but in close contact with its cytosolic side, were also detected: the Rho GTPase-activating proteins ARHGAP21, ARHGAP5, ARHGAP32, ARHGAP5 and the associated guanine exchange factors FARP1, FARP2, FGD1, FGD6, or phosphodiesterases PLCB1, PLCB3, PLCB4, PLCH1 mediating intracellular cascades of signal transduction. These results indicate that proteins do not require a transmembrane helix to be characterized as membrane-anchored in our approach, indicating that BirA* biotinylation is able to also enrich for proteins which are in close proximity to the localization of the enzyme.

The MS analysis detected 280 proteins localized at the cytosolic surface of the mitochondria outer membrane. Our data recapitulated the expected localization of proteins mediating mitochondrial

fission and fusion (MFN1, MFN2, MFF, MTFR1, MTFR1L, MTFR2, MIEF2, etc.), apoptosis (BCL2L11, BCL2L2, BCL2L13, BNIP3, BNIP3L, VDAC2, AIFM1, BAX, MCL1, CYC1, etc.) and cellular respiration, *i.e.* the respiratory chain complex (NDUFA8, TIMMDC1, AIFM1, NDUFS5, NDUFAF4, ACAD9, NDUFS1 and BCS1L).

Nonetheless, several proteins without any prior annotation at the mitochondrial membrane were also identified. For example, many peroxisomal proteins are biotinylated by MITO-BirA*, indicating the close proximity of these two organelles in HEK293 cells. Previous studies have shown that mitochondria and peroxisomes can indeed be close to each other (Fan et al., 2016; Fransen et al., 2017), and share the regulation of some metabolic processes, such as fatty acids oxidation and homeostasis of reactive oxygen species (Wanders et al., 2016). Since we did not target BirA* at the membrane of peroxisomes, we cannot have a full characterization of the proteins more likely to mediate the interactions with mitochondria, but our approach provides a first comprehensive insight into potential candidates at the interfaces of the two compartments. Further studies, including a proximity labelling enzyme at the cytosolic-facing membrane of peroxisome will better characterize the proteome at these specialized cellular regions.

Proteins found to co-localize in two compartments are valuable to understand the functional properties and communications between different regions of human cells. For example, when looking at proteins similarly enriched at both the outer mitochondrial and ER membranes, we not only found proteins previously detected at these locations, like VDAC1 (Shoshan-Barmatz et al., 2004), FKBP8 (Kwak et al., 2020), INF2 (Chakrabarti et al., 2018) and DFCP1 (Hamasaki et al., 2013), but also many unannotated ones, for a total of 102 proteins at ER-mitochondrial contacts. Interestingly, more than 10% of proteins without previous annotation at these specialized compartments (for example ABCD3, ALDH3A2, PTPN1) were recently also detected by a similar approach focusing on the identification of the proteome composition at ER-mitochondrial contacts (Cho et al., 2020), indicating a potential underestimation of the proteins shared by these cellular regions.

Many of the proteins involved in the communication and regulation between different cellular compartments have still to be clearly characterized, and our datasets provide insights on their potential composition in human cells.

A similar study to the one presented here was published by the Ting laboratory (Hung et al., 2017), where the ascorbate peroxidase APEX2 was used instead of BirA*. Although the kinetic of biotinylation of APEX2 is much faster than BirA* (seconds versus hours), its activation requires the use of high concentration (1 mM) of H_2O_2, a well know cell stressor that could quickly affect protein localization. TurboID is an evolutionary selected version of BirA*, displaying a faster biotinylation kinetics (minutes)

without the need of toxic reagents to induce proximity labelling (Branon et al., 2018). Although, when the experiment is planned with proper cytosolic controls such as the one used in this study, the slow kinetic of the BirA* permits a better characterization of proteins, which have stable interactions with the compartment interrogated. Proteins with transient contacts are instead filtered out. Therefore, BirA* proximity labelling should not be utilized to investigate cellular processes occurring with fast dynamics and short time periods, but it is a potent approach to study the resident composition of a cellular region.

RNA-binding protein compartmentalization in HEK293 cells

Our BirA*-based and other similar approaches (Geladaki et al., 2019; Hung et al., 2017; Thul et al., 2017) to study protein localization in a high-throughput manner do not take into consideration the functional properties of the captured polypeptides. For example, RNA-binding proteins are represented in these datasets, but the method does not assess their binding to transcripts. Since the interaction with RBPs influences every step of the life of an RNA molecule, the location of actively binding RBPs would be useful to better understand where these proteins perform their functions and compare their RNA-binding affinities dependent on the compartment. Moreover, by separating the analysis of RBP-RNA interactions to different cellular locations, the complexity of the cellular environment is reduced and the identification of these contacts increases. Nucleus, cytosol and membranes compose an eukaryotic cell. These three compartments carry out different cellular functions, thus the RBPs repertoire is expected to differ between each other.

For these reasons, we developed a novel activity-based protocol to purify proteins binding to RNA in the nucleus, the cytosol and at the membranes of HEK293 cells. We coupled cellular fractionation and acidic guanidinium thiocyanate-phenol-chloroform biphasic extraction (XRNAX (Trendel et al., 2019)) with mass spectrometry. The data obtained allowed us to characterize the protein-RNA interactome in different locations, to analyse the protein domains mediating the interactions and to compare how the RNA-binding affinities change in the different compartments.

Development of f-XRNAX

When compared to proximity labelling approaches, biochemical fractionation methods are regarded as less specific due to incomplete separation between the different samples. Still, proximity labelling is useful to gain insights about proteomes in specific regions of the cell, but is less adapted to give a global picture of the whole cell compartmentalization. Proximity labelling also requires cell-line genetic manipulation and more hands-on time compared to fractionation-based methods.

For these reasons, we opted to develop a protocol which does not rely on proximity labelling and optimized a detergent-based fractionation until pure cytosolic, membrane and nuclear fractions were obtained, with no samples cross-contamination (Figure 11B). We also chose to lose in resolution of the compartments analysed (*i.e.*, we cannot distinguish nucleolar RBPs from spliceosomal ones), but retain the full information regarding the cellular environment of the analysed compartment (*i.e.*, nucleus, membranes and cytosol). On top of these fractions, we applied and adapted a protocol for RBP enrichment called XRNAX (Trendel et al., 2019) and detected the purified proteins through mass spectrometry. Respect to previously established protein occupancy profiling on transcripts (Baltz et al., 2012; Bao et al., 2018; Castello et al., 2012; Huang et al., 2018), XRNAX is not biased to capture RNAs with a poly(A) tail and does not require ribonucleoside analogues.

As for the majority of biochemical techniques, f-XRNAX suffers of background noise. That's why polypeptides are identified also in the control condition (the non UV-irradiated samples) polypeptides are identified. Although this might be seen as a drawback of our approach, the presence of these proteins makes the comparison between control and experimental conditions easier. In fact, it is more quantitative and reliable to calculate enrichment over some low intensity value than over something that is imputed or not detected.

RBPome localization in HEK293 cells

The analysis of the proteins detected with f-XRNAX showed a clear enrichment for RBPs and recapitulated the expected cellular localization for known cell-compartment markers. After stringent filtering (e.g. adjusted p-values < 0.01) and careful comparisons, we obtained high-quality datasets for proteins actively binding to ribonucleic acids in the nucleus, membranes and cytosol of HEK293 cells, and the intersections between these cellular regions. Although the majority of proteins were previously annotated as RNA-binding, our approach revealed new potential candidates RBPs. For example, LSM5, RBIS and NUP85 are binding to RNA in the nucleus according to our data. While these proteins have not been previously identified with RIC or XRNAX approaches in human cells, they are involved in cellular processes related to regulation of RNA metabolism. LSM5 is a Sm-like protein involved in splicing, RBIS in ribosome biogenesis in the nucleolus and NUP85 is a component of the nuclear pore, where all mRNA targeted to the cytosol have to pass through. Thus, these proteins are likely binding or in close contact to RNA in the nucleus and these interactions were previously missed. Similarly, in the cytosolic fractions, some of the proteins which lack evidence of RNA-binding activity, are involved in RNA regulation: EAF1, for example, stimulates RNA polymerase II elongation (Kong et al., 2005). On the other hand, many cytosolic proteins not identified by RIC or XRNAX in HEK293 are

metabolic enzymes (PFKP, NRD1, PLK2, PGD, TPI1), a class of polypeptides which has been repeatedly identified to 'moonlight' as RBPs (Hentze et al., 2018).

Differently from the cytosolic and nuclear samples, the membrane fraction enriched for a large number of proteins not annotated as RBPs. This result likely arises from the applied procedure: instead of recovering RNA-binding proteins from whole cell lysate as in previous approaches, the applied fractionation in f-XRNAX separates the different cellular compartments. As a result, the sample complexity was reduced and the sensitivity of RBP detection was increased, *i.e.* for low abundant proteins or those with weak/transient interactions. These findings are in accordance with the ones from another organic-phase enrichment approach (Queiroz et al., 2019), which identified many unannotated RBPs and predicted their localization at the membrane compartment.

We have also provided evidences that these potential RBPs are usually integrated in the membranes (*i.e.*, they contain a transmembrane domain) and their contacts with transcripts mostly occur in region depleted of protein domains. These IDRs have the potential to bind RNA, and an approach aimed to define the RBD in human cells showed that more than half of the identified domains contained IDRs (Castello et al., 2016).

In addition, we identified the position of RNA-protein contacts with amino acid resolution for 76 polypeptides, by analyzing peptides crosslinked to ribonucleosides. Thus, we determined the protein domains interacting *in vivo* with RNA in HEK293 cells for a subset of proteins. Unsurprisingly, many of the peptides with crosslinked ribonucleosides were found in well annotated RNA-binding domains, such as RRM and S1-like. The ability to infer which part of a protein is in contact with RNA *in vivo* provides a mean to define RBDs. In the near future, the development of faster and more accurate algorithms to informatically search for peptides crosslinked to ribonucleotides will provide a powerful tool to identify novel RBDs and with a deeper protein coverage.

Although highly informative, only a minority of peptides in our datasets were crosslinked with cyclic uridine or uridine monophosphate. To determine the protein domains in proximity to RNA for the rest of the polypeptides, we observed how peptides identified by MS were distributed along the protein sequence. With f-XRNAX, accumulation of peptides in one region of the protein is a good indication that RNA was crosslinked in its vicinity. Computational analysis confirmed that the obtained peptides profiles recapitulated the RBD positions in well know RNA-binding proteins (Figure 18).

The same analysis can further be exploited to discover novel and unannotated protein domains potentially binding to RNA. As an example, the protein YBX1 in Figure 18 displayed a peptides enrichment at its C-terminal, a region devoid of known structural domains. These results indicate a possible involvement of that part of the protein in interactions with ribonucleic acids. YBX1 is a well known RBP, with cellular function spanning from RNA splicing (Raffetseder et al., 2003) and stability

(Chen et al., 2019), to RNA loading of exosomes (Shurtleff et al., 2016) and binding to methylated ribonucleotides (Chen et al., 2019). While the cold shock domain of YBX1 has been extensively studied (Zhang et al., 2020), the regulatory activity of its C-terminal part is less characterized. Nevertheless, there have been evidence on its involvement in RNA translation inhibition *in vitro* (Nekrasov et al., 2003) and many positively charged residue are found in this part of YBX1, facilitating the interaction with negatively charged molecules as RNAs. This information, together with our data, suggests a potential and not characterized RNA-binding domain within YBX1 C-terminus.
The same analysis can be performed to find other uncharacterized RBDs, integrating the data from our datasets with the one from previous works (Castello et al., 2016; Maticzka et al., 2018; Mullari et al., 2017). The enriched peptides analysis was a qualitative measurement of the RNA-protein contacts and its findings should be confirm experimentally.

The uniqueness of our datasets consists in the distinction of the RBPome retrieved at different cellular environments. Integrating the enriched peptides profiles in our analysis, we were able to distinguish the differential domain usage for the same protein according to its cellular localization.
We found that the two paralogue proteins ELAVL1 and ELAVL2 exhibit a domain- and compartment-specific RNA-binding behaviour, with the most C-terminal RRM (RRM3) displaying higher binding capacity to ribonucleic acids in the membrane region.
ELAV protein family is a well described class of RBPs in mammals, binding to adenylate/uridylate-rich elements within untranslated regions of transcripts. Interestingly, RRM3 is the least characterized of the three RRMs in ELAV proteins. It has been recently shown that it preferentially binds to polyU stretches, versus the UAUUA binding motif recognized by RRM1-2 and it is prone to dimerization, which increases its affinity of RNA-binding (Ripin et al., 2019). The difference in binding discrimination and the ability to increase its local concentration through oligomerization may likely be two important reasons why RRM3 in the two ELAV paralogues is binding with higher affinity to RNA in the membrane compartment. The membrane compartment localizes a different transcript set than the cytosol or the nucleus and could provide a suitable surface enhancing the dimerization of RRM3. The fact that we found comparable results for two paralogue proteins provides further support for this hypothesis.
Nevertheless, only proteins with many peptides provided enough information for this analysis and deeper dataset will provide a more comprehensive understanding of RBDs usage in different subcellular compartment.

As every biochemical protocol, also the f-XRNAX approach suffers of limitations. Because of the large amount of cells and replicates required to obtain valuable data, a SILAC approach was disregarded because it would have been too expensive. Normalization with control samples (-UV) and

quantification with LFQ intensities applied in this work is doable, but retains a lower quantification resolution compared to a SILAC experiment. An RBP-enrichment approach like OOPS (Queiroz et al., 2019), which requires less cells as starting material would allow an experimental design including amino acid labelling with stable isotopes.

The RBPome enriched are also slightly biased for proteins interacting with one or more uracil. It has been previously shown, in fact, that UV-crosslinking of protein-RNA contacts is more efficient on uracil bases (Smith and Meun, 1968). Since the UV-driven covalent bond between RBP and RNA is a necessary pre-requisite for the purification of RNP complexes, proteins not in direct contact with uracil ribonucleotides have a minor probability to be enriched

A common limitation of organic-phase RBP enrichment protocols is that they poorly distinguish RBPs from glycosylated proteins, due to their phyco-chemical similarities. Thus, parallel methods have to be employed to better characterize the RNA-binding properties of proteins with covalently attached sugar groups and their pattern of subcellular distribution.
Because of the depth of the data, the computational analysis of protein composition in the different subcellular compartments from the f-XRNAX experiment was conducted without considering distinct protein isoforms generated from differentially spliced RNA. Thus, our analysis is not considering a key factor contributing to protein heterogeneity in human cell, namely the different composition of polypeptides generated from the same gene but with different splicing pattern.

RNA kinetics during the cell-cycle

To accurately distinguish newly synthetized and pre-existing RNA molecules in single cells, we developed a novel protocol, named SLAM-Drop-seq. To achieve this goal, we metabolically labelled proliferating cells with the ribonucleotide 4sU and optimized the conditions for IAA-driven alkylation within fixed individual cells. Single cell encapsulation is then performed with one of the available droplet-based method, and libraries containing cell barcodes and UMI are prepared. Newly synthetized and pre-existent molecules are computationally identified and quantified by detecting T to C transitions in the sequenced reads. In addition, we have also developed the R package Eskrate to estimate the transcriptome-wide and time-dependent RNA kinetic rates of transcription, processing and degradation over the course of a dynamic biological event.
As a first application, we implemented our novel approaches to study the RNA kinetic rates associated with the cell cycle of unsynchronized HEK293 cells. We found 377 genes with an oscillating expression

along the cell cycle, with a defined peak of expression and a low amount of dropout (*i.e.*, cells with no information on that gene). Using the kinetic model and solutions of the ODEs calculated by Eskrate, we were able to identify the underlying transcription and degradation rates giving rise to the observed expression. Importantly, these estimations are time-dependent and change continuously accordingly to the cell cycle time. This means that this novel approach not only assigned single cells to a continuous cell cycle time, without relying on the strict definition of cell cycle phases as in previous synchronization studies (Grant et al., 2013; Whitfield et al., 2002), but is also superior to other single cell analysis relying on pseudo-time assignments or RNA velocity to determine RNA kinetics (Manno et al., 2018). Furthermore, the temporal information retrieved with our approach led us to assign physical units to the parameters of RNA synthesis (molecules per minute) and decay (1/ minute, inversely proportional to half-life), which is highly valuable for accurate quantifications of transcripts produced or degraded in a defined timeframe.

We found that the gene expression profiles form vast majority of the identified 377 cycling genes rely not only on temporal changes of RNA synthesis but also on degradation. This is in contrast with previous reports (Cho et al., 2001; Liu et al., 2017) which stated that the main driver for human cycling gene expression is RNA transcription. We have shown that the expression of only 6% cycling genes is regulated during the cell cycle largely by transcription. Therefore, at least in the case of HEK293 cells, our approach was useful to define the expression strategies that each gene employs along the cell cycle. Not only we have identified degradation to be highly involved in the regulation of gene expression together with transcription, but we have also found 16 genes that solely rely on temporal changes of decay to shape their expression profiles, with an approximately constant RNA synthesis rate. Our findings emphasize the different strategies that HEK293 cells exhibit to regulate gene expression throughout the cell cycle, and how the assignment of temporal units to these parameters allow a better comprehension and quantification of the underlying biological phenomena. While the temporal changes of transcription and degradation rates have been previously revealed in the cell cycle with low time resolutions (Battich et al., 2020; Eser et al., 2014), we infer these rates in much higher time resolution (*i.e.*, less than one minute).

During the course of this dissertational work, different approaches to study RNA kinetics at the single cell resolution were published (Battich et al., 2020; Cao et al., 2020; Erhard et al., 2019; Hendriks et al., 2019; Qiu et al., 2020). scSLAM-seq (Erhard et al., 2019) and NASC-seq (Hendriks et al., 2019) couple 4sU metabolic labelling with Smart-seq2 (Picelli et al., 2014), a sequencing protocol recovering a high number of full-length RNA molecules per single cell. Nonetheless, these approaches capture at least an order of magnitude less cells respect to droplet-based sorting. The published scSLAM-seq and NASC-seq studies sequenced less than 50 and 75 cells per condition, respectively, making them less reliable

to track the full dynamics of complex biological processes. In scEU-seq (Battich et al., 2020) FACS-sorting of single cells and pull-down were performed to enrich for labelled RNAs, which is laborious and prone to introduce biases in the procedure. A complex protocol was also applied to obtain the combinatorial indexes in sci-fate (Cao et al., 2020). On the other hand, scNT-seq (Qiu et al., 2020) is conceptually and methodologically similar to SLAM-Drop-seq, except the already discussed alkylation in fixed cells, which reduces the hands-on time and experimental complexity. Moreover, Qui et al. do not focus on time-dependent estimation of RNA decay rate, but more on their differences between cellular states (e.g. differentiation from mESC to 2C cells), losing on temporal and quantitative resolution of the transcriptional degradation rate. This is also true for NASC-seq, scSLAM-seq and sci-fate, where the decay rate was fit under steady state assumptions. In scEU-seq, Battich and colleagues introduced the florescent reporter FUCCI (Sakaue-Sawano et al., 2008) to assign cell cycle time to each cell and estimated the RNA kinetic rates by pooling cells at close cell cycle times (accordant to the FUCCI reporter fluorescence). Thus, they not only genetically manipulated a cell line to infer the time variable, but also calculated the synthesis and degradation rates from small cell populations, instead of assigning each single cell to a specific cell cycle time and to RNA kinetics rates, as in SLAM-Drop-seq. The implementation of a fluorescent reporter would also be less suitable for studies in primary cells or experimental conditions where genetic manipulation might be difficult or not feasible.

One limitation of SLAM-Drop-seq is the enrichment with oligo-dT beads of polyadenylated transcripts (or, more rarely, long poly(A) stretches in CDS or UTRs). To reduce the described 3'-end bias and increase the coverage of sequenced nucleotides per molecule, we developed a computational approach to merge *in silico* reads containing the same UMI. In this way, we accomplished a more reliable quantification of newly synthetized and pre-existing RNAs in thousands of single cells, without relying on full-transcript and low cell-throughput approaches as Smart-seq2.
Labelling RNA with 4sU could also been addressed as a limiting step of the developed protocol, because of its reported toxicity (Burger et al., 2013) and the limited application in living animals. Computational analysis did not reveal any change in gene expression and no alteration of unspliced transcripts upon 4sU treatment, indicating that incubation with the ribonucleotide analogue did not cause cellular toxicity. Furthermore, even if not discussed in the presented thesis, the SLAM-Drop-seq protocol, described in this thesis, has also been successfully applied to other biological systems in collaborations with colleagues from MDC and Charite, like patient-derived colon organoids (Uhlitz et al., 2021) and *in vivo* in developing zebrafish (Holler et al., 2021). This highlights the feasibility of the approach, which can be applied to different cellular contexts and processes.

A constraint of the SLAM-Drop-seq approach and its data analysis was that we had to adjust our kinetic model and introduced simplifications, because of the low amount of unspliced transcripts detected per cell. This made the estimation of transcription and degradation rates more accurate, even at the expenses of not being capable of calculating the processing rates (*i.e.*, splicing rate). Nonetheless, the mathematical framework used in our approach would be able to estimate the processing rate if deeper data were feed to it. The SLAM-Drop-seq protocol is adjustable to many different single cell sequencing platforms, since the alkylation step is preceding the cell encapsulation / sorting and was successfully used in combination with a Dolomite Nadia system, a self-built Drop-seq device and to 10x Genomics Chromium sequencer. Probably in the near future, innovative single cell sequencing protocols can be coupled with the SLAM-Drop-seq approach to achieve a deeper coverage of single cell transcriptomes and, by using the Eskrate package, achieve a reliable estimation also for the splicing rates. Moreover, Eskrate is capable to implement times different from the cell cycle one, opening the doors to study different dynamical biological processes with time-dependent RNA kinetics.

MATERIAL AND METHODS

Cell culture

Human HEK293 Flp-In T-Rex cells (HEK293, Thermo Fisher Scientific) were cultured at 37 °C supplied with 5 % CO_2 and were grown in Dulbecco's Modified Eagle Medium (Gibco, 41965039) supplemented with 10 % fetal bovine serum (Gibco, 10270106) and 2 mM L-Glutamine (Gibco, 25030081). TrypLE Express Enzyme (Gibco, 12605036) was used to detach cells and split them 2-3 times per week. Contamination with Mycoplasma was checked every week with a specific PCR kit (ITW Reagents, 38221900). All the samples submitted for sequencing weren't contaminated with Mycoplasma.

BirA cell lines generation*

All the DNA plasmids used to generate BirA* cell lines were derived from the pcDNA5-FRT-TO vector and contained the FLAG-HA epitope tag. BirA* was synthetized by BioCat and cloned with a plasmid already present in the laboratory containing the N-terminal first 27 amino acids of the P450 2C1 protein (MDPVVVLGLCLSCLLLLSLWKQSYGGG) and cloned into the pcDNA5-FRT-TO vector. This plasmid was used to create the ER-BirA* cell line. A similar procedure was used to clone the plasmid for MITO-BirA*: the C-terminal 31 amino acids of the protein MAVS (RPSPGALWLQVAVTG VLVVTLLVVLYRRRLH) were inserted in the C-terminus of BirA* (with a GS linker between them) and the construct was cloned into a pcDNA5-FRT-TO vector containing the FLAG-HA tag. The plasmid for PM-BirA* was cloned using a DNA oligonucleotide containing the sequence of the CaaX domain from the KRAS4B protein (KMSKDGKKKKKKSKTKCVIM) and cloning it (using restriction enzymes) in the MITO-BirA* plasmid substituting the C-terminal MAVS peptide. For the cytosolic controls, one plasmid was obtained by mutational PCR of the PM-BirA* construct, substituting the cytosine in the CaaX domain (essential for the protein prenylation and insertion in the inner leaflet of the plasmatic membrane) with a serine residue; the second cytosolic control was created by adding via PCR the nuclear export signal of the Rev protein from HIV-1 virus (LQLPPLERLTLD) to the C-terminus of BirA*. Stable HEK293 cell lines were created by plasmids transfection and colony selection with hygromycin (Gregersen et al., 2014). The expression of the integrated gene of interest was achieve by adding 1µg/mL of doxycycline to the culture medium for around 16 hours.

Immunofluorescence

Stable HEK293 cell lines were grown on a glass coverslip in 6-well plates. BirA* expression was induced for 16 hours with 1 µg/mL doxycycline. Cells were fixed in 3.7% formaldehyde (dissolved in DPBS) for 10 minutes at room temperature, washed 3 times with DPBS and permeabilized with ice-cold methanol for 5 minutes at -20 °C. Cells were then washed 3 times with DPBS, blocked with 3% BSA (dissolved in

DPBS) for 3 hours at room temperature and incubated overnight at 4 °C with anti-HA antibody (1:1000 dilution). The excess of primary antibody was washed three times with 0.2% Tween20 - DPBS (DPBS-T) and the anti-mouse Alexa fluorophore 488 (Thermo Fisher Scientific) secondary antibody was incubated for 1 hour at 4 °C (1:750 dilution, in blocking buffer). After 3 more washes with DPBS-T, cells were incubated for 5 minutes with 10 µg/µL Hoechst 33342 (stock 10 mg/mL, dilution 1:1000), mounted on a glass surface using ProLong Gold Antifade Mountant (Thermo Fisher Scientific) and imaged with Keyence BZ-X700 microscope.

BirA cell lines biotinylation and streptavidin pull-down*

The protocol set in the Gingras lab for BioID was followed (Roux et al., 2012), with some modifications. One 15 cm plate was used for each biological replicate. HEK293 cell lines were splitted two days before the experiment and induced with 1 µg/ml doxycycline to express the locally-restricted BirA* and 50µM biotin to promote proximity biotinylation for 16 hours. Cells were then carefully washed three times with 10mL warm DPBS to remove all residual free biotin, which would impair the streptavidin beads enrichment. They were afterwards collected in DPBS using a cell scraper, collected by centrifugation (3 minutes, 300g, 4°C) and the pellet snap frozen with liquid nitrogen and kept at -80°C. When samples were ready for cell lysis and affinity purification, they were thaw on ice and lysed in RIPA buffer (50mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 1mM EDTA, 1 mM EGTA, 0.1% SDS, Sigma protease inhibitors P8340 1:500, and 0.5% Sodium deoxycholate) and 250 units of benzonase for 1 hour, 750rpm, at 4°C. After sonication, lysates were cleared for 30 minutes at > 20 000g, 4°C. Streptavidin agarose beads (GE Healthcare System) were washed three times with RIPA buffer and 30µL of bed volume was transferred to each cleared sample. The affinity purification took place at 4°C, for three hours with mild rotations. Beads were then washed two times in RIPA buffer, two times in TAP lysis buffer (50 mM HEPES-KOH pH 8.0, 100 mM KCl, 10% glycerol, 2 mM EDTA, 0.1% NP-40) and an additional 3 times in ABC buffer (50mM ammonium bicarbonate pH 8.0). Samples in ABC buffer are ready for mass spectrometry.

Mass spectrometry of BirA samples*

Streptavidin beads with biotinylated proteins attached to them were resuspended in 200 µl of 50 mM ammonium bicarbonate containing 1 µg trypsin (Promega V511A). The samples were incubated for 16 hours in a Thermomixer (Eppendorf) at 37 °C and 750 rpm shaking. After this time, one µg of trypsin was again added and sample incubated further for 2 hours. Samples were then centrifuged at 400 g for 2 minutes and the supernatant transferred to a new vial. To ensure complete beads removal, the samples were centrifuged again at 16000 g for 10 minutes and the supernatants were transferred to a

new vial containing 2.5 µl of trifluoroacetic acid. Each sample was loaded on two StageTips (Rappsilber et al., 2003) for desalting. Eluates for each sample were pooled together prior to mass spectrometry.
For all the samples of the experiment, 5 µL were injected in duplicate on a LC-MS/MS system (Ekspert NanoLC 415 [Eksigent] coupled to Q Exactive HF [Thermo Fisher Scientific]), using a 240 minutes gradient ranging from 2% to 45% of solvent B (80% acetonitrile, 0.1% formic acid; solvent A= 5% acetonitrile, 0.1% formic acid). For the chromatographic separation, 30 cm long capillary (75 µm inner diameter) was packed with 1.8 µm C18 beads (Reprosil-AQ, Dr. Maisch). On one end of the capillary nanospray tip was generated using a laser puller (P-2000 Laser Based Micropipette Puller, Sutter Instruments), allowing fretless packing.
The nanospray source was operated with a spay voltage of 2.2 kV and an ion transfer tube temperature of 260 degrees. Data were acquired in data dependent mode, with a top10 method (one survey MS scan with resolution 120,000 at m/z 200, followed by up to 10 MS/MS scans on the most intense ions, resolution 30,000, intensity threshold 5,000). Once selected for fragmentation, ions were excluded from further selection for 30 seconds, in order to increase new sequencing events.
Raw data were analysed using the MaxQuant proteomics pipeline (v1.5.3.30) and the built in the Andromeda search engine (Cox et al., 2011) with the Uniprot Human database. Carbamido-methylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. The search engine peptide assignments were filtered at 1% FDR and the feature match between runs was not enabled; other parameters were left as default.

Data analysis of BirA samples*

The "ProteinGroups" file from the MaxQuant software was analysed with R (version 4.0.4) RStudio (version 1.4.1106). Proteins detected with two or more unique peptides were included in the analysis (5842 out of 7069). If a protein was detected as 'Potential contaminant' or 'Reverse', it was also filtered out (36 in total). We used the "DEP" package (Zhang et al., 2018) to filter the dataset for proteins without any missing values for LFQ intensities in at least one condition and to impute the remaining missing value using a manually defined left-shifted Gaussian distribution (1.8 left-shift from median and 0.3 width relative to original LFQ-distribution). The package was also used to plot the total number of proteins, Pearson-correlation, PCA, heatmap and single-protein centered-intensity analysis. Limma (Ritchie et al., 2015) was then used to perform protein-wise enrichment tests between conditions and compute empirical Bayes statistics. The resulting p-values were corrected for multiple comparison using the Benjamini-Hochberg FDR approach (Benjamini and Hochberg, 1995). Proteins were called significantly enriched in one condition respect to a second one if the adjusted p-value was lower than

0.01 and the log2 LFQ-intensity fold change greater than 1. A protein was assigned to be specifically localize to one compartment when these conditions were met against all possible contrasts. When two or more conditions met the set thresholds against other conditions, the proteins were called to co-localize. The analysis using HEK293 without BirA* detected enriched background proteins (68), which were filtered out.

GO-term analysis was carried out with DAVID (Huang et al., 2009) using the desired list of genes and the total amount of proteins experimentally identified as background.

The true positive proteins of ER and mitochondrial were taken from a manually curated lists of annotated localized proteins (Hung et al., 2017). The secretory and sub- secretory protein annotation was extracted from the Human Protein Atlas (Thul et al., 2017), Phobius (Käll et al., 2004) and Gene Ontology (Consortium, 2015). The dataset of TurboID and APEX2 were retrieved from their respective publication (Branon et al., 2018; Hung et al., 2017). The mitochondrial and sub-mitochondrial annotation was assessed with Mitocarta 3.0 (Rath et al., 2020). The Human Protein Atlas was used to check the annotation of a protein in the plasma membrane compartment.

Optimized fractionation protocol for HEK293 cells

The fractionation protocol was derived by the published paper from Jagannathan et al. and optimized for HEK293 cells. The main changes from the previous protocol are: increased incubation time in both lysis buffers for 10 min; lysis occurring while rocking at 4 °C instead of incubation in ice; in the second lysis buffer the concentration of NP-40 was decreased to 0,5% instead of 1% and sodium-deoxycholate was not added (this helped to preserve nuclei); an extra step to the purification of nuclei was added through a 10% sucrose cushion.

HEK293 cells were splitted the day before the fractionation experiment and reached ~80% confluency the day after. Cells were washed with warm DPBS, without detaching them. Cold DPBS was used to scrape them from plates. Cells were then spin down and resuspended in DPBS + 50ug/ml CHX for 10 min, on ice (200µL for one 10cm plate). ¼ of the volume was taken and collected as the total fraction. Cells were spun down again and resuspended in 'Permeabilization buffer' (110mM KOAc, 25mM K-HEPES pH 7. 2, 2.5mM $Mg(OAc)_2$, 1mM EGTA, 1mM DTT, 0.015% digitonin , 50µg/ml CHX, 1× Complete Protease Inhibitor Cocktail (Roche), 20U/mL SUPERase•In™ (Thermo Fisher Scientific)) for 10 minutes, rotating slowly at 4°C. The supernatant after the spin was collected as the cytosolic fraction. The cell pellet was then washed once in 'Washing Buffer' (110mM KOAc, 25mM K-HEPES pH 7.2, 2.5mM Mg(OAc)2, 1mM

EGTA, 0.004% digitonin, 1mM DTT, 50µg/ml CHX) and then resuspended in 'Lysis buffer' (400mM KOAc, 25mM K-HEPES pH 7.2, 15mM Mg(OAc)2, 0.5% NP-40, 1mM DTT, 50µg/ml CHX, 1× Complete

Protease Inhibitor Cocktail, 20U/mL SUPERase•In™). for 5 minutes on ice. After the spin, the supernatant was collected as the membrane fraction. The nuclei were collected after being passed through a 10% sucrose cushion. The membrane and cytosolic fractions were further clarified with a centrifugation 7500g, 4°C, 10 min.

Western blotting analysis

The Western blots were performed using pre-casted 4-12% acrylamide Bis-Tris gel (Thermo Fisher Scientific). Samples were resuspended in Laemmli loading buffer (32mM Tris HCl pH 6.8, 10% glycerol, 1% SDS, 0.002 % Bromophenol blue) with 2-mercaptoethanol (final concentration 355 nM), sonicated (5 seconds at 80% intensity) and boiled for 5 minutes at 95°C before loading equal volumes on the gel, which was run with 1x SDS MOPS buffer (prepared from 20x NuPAGE™ MOPS SDS Running Buffer). The proteins were then transfer to a nitrocellulose membrane (Whatman) using a semi-dry blotting apparatus (1 hour, 20 Volts) in transfer buffer (25mM Tris pH 8.3, 192mM glycine and 20% (v/v) methanol). The transferred proteins were then visualized with Ponceau staining (0.1% Ponceau S, 5% acetic acid) and the membrane was blocked for 1 hour at room temperature with 5% milk diluted in TBST (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1% Tween20). The primary antibody was incubated overnight in agitation, at 4°C. The membrane was then washed 3 times for 5 minutes at room temperature with TBST and incubated for another hour at room temperature with the secondary antibody. After washing 3 times for 5 minutes in TBST, the membrane was visualized using ECL detection reagent (GE Healthcare) and imaged with an Amersham Imager 680 imaging system or a LAS-4000 imaging system or an (GE Healthcare).

Silver staining

The polyacrylamide gel containing the proteins separated by molecular weight was incubated for 1h in fixation solution (30mL ethanol, 15 mL acetic acid and 55 mL ddH_2O) and one additional hour at room temperature in 0.5 M NaOAc, 12 mM Na2S2O3, 0.125% glutaraldehyde, 25% ethanol. It was then washed three times for 10 minutes in ddH_2O and left for 30 minutes in staining solution (0.1% $AgNO_3$ and 0.011% formaldehyde in ddH_2O). After the gel was briefly washed in ddH_2O, the developer solution was added (236 mM Na2CO3, ,0.011% formaldehyde in ddH_2O) and the reaction was stopped when bands saturation was nearly reached by 50mM EDTA. The gel was washed in ddH_2O and imaged with an Amersham Imager 680 imaging system or a LAS-4000 imaging system or an (GE Healthcare).

f-XRNAX protocol

For each biological replicate of the f-XRNAX experiment, we used 250 million of exponentially growing HEK293 cells (~70% confluency on the day of the experiment). The +UV samples were crosslinked with 150mJ/cm^2 UV_{254nm}, without detaching them from the plate. The fractionation protocol follows the above described optimized protocol, adjusting the volumes accordingly and with minor changes: cells were left in permeabilization buffer for 20 minutes, instead of 10 and samples were snap frozen in liquid nitrogen and kept at -80°C. Input samples from the different fractions were saved.

The XRNAX procedure followed the published protocol (Trendel et al., 2019), adjusting for volumes. Each sample had five biological replicates and +UV and -UV conditions. Since the different fractions were diluted in their respective buffers, we used a self-made version of Trizol LS (50% ROTI-phenol for RNA, 1M guanidinium thiocyanate, 0.5M ammonium thiocyanate, 0.12M sodium acetate pH 5.0, 10% glycerin) to separate the organic and aqueous phase, using a 3:1 volume ratio between Trizol LS and sample. After adding chloroform (one fifth of the volume) and centrifugation (7500g, 10', 4°C), the interphase containing the proteins crosslinked to RNA is carefully captured and dissociated twice with TE + 0.01% SDS and two more times with TE + 0.05% SDS. Samples were then precipitated with isopropanol (1 volume 2-pronol, 150mM NaCl, 1µL GlycoBlue™ Coprecipitant (Thermo Fisher Scientific)) and the pellet washed with 70% ethanol. The pellet was then rehydrated in distilled water and chromatin was digested for 90 minutes with DNAse I. Samples were again precipitated with isopropanol and washed with 70% ethanol and resuspended in water. The RNA concentration was measured with Nanodrop.

For mass spectrometry preparation, equal volumes of f-XRNAX samples were reduced at 60 °C for 30 minutes at 700 rpm in 5 mM Tris-HCl pH7.5, 0.04% SDS and 10 mM DTT. Alkylation occurred for 30 minutes at room temperature using 20 mM CAA. 100 ng Trypsin/ LysC (Promega) were added for 15 minutes at room temperature for partial digestion to occur. The reaction was stopped by adding 3.5 mL RTL buffer. The silica column purification was performed using Quiagen RNeasy Midi Kit as described in the published XRNAX paper (Trendel et al., 2019). Samples were then heated for 5 minutes at 65 °C and cooled on ice before adding 1.5 µL of RNase I (Ambion, AM2295), RNase A (Thermo Fisher Scientific, EN0531) and RNase T1 (Thermo Fisher Scientific, EN0541). The RNA digestion lasted for 12 hours at 37 °C. The complete protein digestion was achieved by adding 500 ng Trypsin/ LysC for 16 hours at 37° C, shaking. The digested peptides were submitted to mass spectrometry

Mass spectrometry of f-XRNAX

f-XRNAX sample preparation and LC-MS analyses:

Sampled were acidified with formic acid (final concentration 1%), cleaned up and stored on C18 stage tips (Rappsilber et al., 2003). Peptides were eluted from stage tips (80% acetonitrile, 0.1% formic acid), and after evaporating the organic solvent, they were resolved in sample buffer (3% acetonitrile/ 0.1% formic acid). From each sample one fourth of peptide material was injected into the LC-MS system. Peptide separation was done on a 20 cm reversed-phase column (75 µm inner diameter, packed with ReproSil-Pur C18-AQ (1.9 µm, Dr. Maisch GmbH)) using a 200 min gradient with a 250 nL/min flow rate of increasing Buffer B concentration (from 2% to 60%) on a High-Performance Liquid Chromatography (HPLC) system (Thermo Fisher Scientific). Peptides were measured on a Q Exactive HF-X instrument (Thermo Fisher Scientific). The mass spectrometer was operated in the data dependent mode with a 60K resolution, 3 x 106 ion count target and maximum injection time 10 ms for the full scan, followed by Top 20 MS2 scans with 15K resolution, 1 x 105 ion count target and maximum injection time of 22 ms.

Total cell lysate protein extraction ("input"), digest and LC-MS analyses:

Cell pellets were lysed in SDS buffer (2% sodium dodecyl sulphate, 100 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl, 10 mM dithiothreitol, 40 mM chloroacetamide), heated for 10 minutes at 95°C, cooled down to room temperature and incubated with 25 U Benzonase (Merck) for 30 min, followed by addition of 20 mM dithiothreitol. Insoluble parts were removed by centrifugation for 20 min at 14,000 rpm, and the supernatant (containing the proteins) was collected. Protein concentration was measured (Bio-Rad DC Protein assay) and 100 µg of each sample were further processed using SP3 clean up and digestion protocol (Hughes et al., 2019). Briefly, 1mg of paramagnetic bead mix (containing 1:1 hydrophilic and hydrophobic beads) was added to the sample (ratio protein:beads 1:10). Acetonitrile was added to a final concentration of 70% and samples were incubated for 20 minutes on a rotator. Samples were washed 2 times with 70% ethanol and ones with 100% acetonitrile. 50ul 100 mM ammonium bicarbonate containing 2 µg sequence grade trypsin (Promega) and 2 µg LysC (Wako) were added and samples were incubated over night at 37°C. After collecting the supernatant, beads were incubated with another 50 µl 100 mM ammonium bicarbonate and both supernatants were combined (containing the peptides). Samples were acidified with formic acid (final concentration 1%) and desalted on C18 material (Rappsilber et al., 2003).

Peptide samples were eluted from stage tips (80% acetonitrile, 0.1% formic acid), and after evaporating the organic solvent peptides were resolved in sample buffer (3% acetonitrile/ 0.1% formic acid). For each sample two analytical runs with 2ug of peptide material per run were performed. Peptide separation was done on a 20 cm reversed-phase column (75 µm inner diameter, packed with ReproSil-Pur C18-AQ (1.9 µm, Dr. Maisch GmbH) using a 200 min gradient with a 250 nL/min flow rate of increasing Buffer B concentration (from 2% to 60%) on a High Performance Liquid Chromatography

(HPLC) system (Thermo Fisher Scientific). Peptides were measured on a Q Exactive HF-X instrument (Thermo Fisher Scientific). The mass spectrometer was operated in the data dependent mode with a 60K resolution, 3 x 106 ion count target and maximum injection time 10 ms for the full scan, followed by Top 20 MS2 scans with 15K resolution, 1 x 105 ion count target and maximum injection time of 22 ms.

MaxQuant Data Analyses:

Raw data were processed using MaxQuant software package (v1.6.0.1). The internal Andromeda search engine was used to search MS2 spectra against a decoy human UniProt database (HUMAN.2019-01) containing forward and reverse sequences. The search included variable modifications of oxidation (M), deamidation (N,Q), N-terminal acetylation and fixed modification of carbamidomethyl cysteine. For f-XRNAX sample analyses the variable modification of cyclic uridine and uridine monophosphate (K, F, G, V) was added. Minimal peptide length was set to seven amino acids and a maximum of three missed cleavages was allowed. The FDR (false discovery rate) was set to 1% for peptide and protein identifications. Unique and razor peptides were considered for quantification. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS2 identifications were transferred between runs with the “Match between runs” option, in which the maximal retention time window was set to 0.7 min. The integrated iBAQ algorithm was applied. The resulting text files were filtered to exclude reverse database hits, potential contaminants, and proteins only identified by site.

Data analysis of f-XRNAX samples

Proteins with at least two unique peptides were kept for the analysis, while reverse hits and potential contaminant were filtered out. As a first step of the analysis, background proteins from non UV-irradiated samples (-UV) were filtered out. To achieve this, samples from the same compartment (cytosol, membrane, nucleus or total) were compared between the two conditions (with or without UV_{254nm} crosslinking). Proteins without any missing value for LFQ intensities in at least one condition were kept, and the remaining missing values imputed with a manual imputation approach (left-shifted Gaussian) using the DEP package (Zhang et al., 2018). The differences between -UV and +UV conditions for the same compartment were then tested with limma (Ritchie et al., 2015), which performed a protein-wise enrichment test. False discovery rates were estimated using Benjamini-Hochberg correction. Proteins with log2 fold change of LFQ intensity greater than 1 and an adjusted p-value lower than 0.05 were considered enriched in the +UV condition and considered for further analysis. The other proteins were filtered out.

After obtaining compartment-enriched and background-depleted proteomes from the different cellular location interrogated, they were merged together. The remaining missing values were imputed with a "mixed" approach: values missing in more than 80% of replicates in one condition, were considered as MNAR (missing not at random), while the rest were labelled as MAR (missing at random). MNAR values were imputed using the default 'MinProb' variables from the MSnbase R package (Gatto and Lilley, 2012), while MAR with a left-shifted Gaussian distribution as before. LFQ intensities fold changes and false discovery rates were calculated with limma and Benjamini-Hochberg correction, respectively. Proteins were assigned to be significantly enriched in one compartment respect to another if their fold change was greater than 2 and adjusted p-value smaller than 0.01. If these thresholds were not met, the protein was labelled as co-localizing in both compartments. The DEP package was used to plot the total number of proteins per replicates, their intensity Pearson correlation and PCA, and the centered LFQ-intensity for protein examples. Isoelectric point data was obtained from 'Proteome-pI' (Kozlowski, 2017).

The profiles of enriched peptides were calculated using the "evidence" files from MaxQuant of the f-XRNAX and input samples. Identified peptides were aligned and counted along the protein length. Position-specific peptides from the input condition were subtracted from the one of the f-XRNAX experiment, leading to the identification of enriched peptides. Negative values (*i.e.*, more peptides identified in input respect to f-XRNAX) were set to zero.

Peptides identified to be crosslinked to uridine or cyclic uridine residues were further filtered with an FDR of 0.05 and a score > 40. Protein domains from the resulting peptides were computed and investigated using the Batch-CD search from NCBI (Marchler-Bauer et al., 2011). Their relative amino acid abundance was also calculated and compared with the one from peptides of HEK lysates (*i.e.*, 'total input').

The transmembrane domains and signal peptides annotation was downloaded from Ensembl Biomart using the human genome assembly hg38 while the intrinsically disordered regions were mapped using the mobiDB database(Piovesan et al., 2020). GO term analysis was performed using DAVID (Huang et al., 2009) and having as background proteins identified in the input samples.

For the analysis of RBPs differential RNA-binding between two cellular compartments, proteins retrieved in both localization in the f-XRNAX experiment and with LFQ intensity fold change (log2) less than 2 in the relative input samples were taken into consideration. Protein-wise LFQ intensities from the f-XRANX samples were normalized with the relative ones from the input samples, and plotted. A linear regression analysis was performed and if proteins had a distance of +/-2 from the linear fit, they were called as significantly differentially binding to RNA. The adjusted p-value from the f-XRNAX compartments comparison data are also plotted.

Bulk RNA alkylation and sequencing

HEK293 cells were incubated with 300 µM 4sU for 30, 60 and 120 minutes; control sample without 4sU incubation. RNA was isolated and 5µg of purified total RNA was alkylated with iodoacetamide (IAA, Sigma-Aldrich, I6125) following the published SLAM-seq protocol (Herzog et al., 2017). 1µg of alkylated total RNA was processed for stranded mRNA library preparation with the Illumina TruSeq kit (20020595) and sequenced on a NextSeq 500/550 HO.

Biotinylation blocking assay and SLAM-Drop-seq procedure

Cells reached ~ 60% confluence on the day of the experiment (*i.e.*, they were in exponential growth phase) and were incubated with 300 µM 4sU (ChemGenes, RP-2304; from a 0.5 M stock, dissolved in DMSO) for different times (15, 30, 60, 120 minutes). Cells were washed with DPBS and dissociated with TrypLE Express Enzyme (Gibco, ref. 12605028) for 2 minutes.

After a wash in cold DPBS, cells were resuspended at a final concentration of 1-2 million cells/mL in cold DPBS. Methanol was added (four volumes) to perform the fixation (around 30 minutes at -20 °C). After re-equilibration of the samples at room temperature, they were separated in half: in one IAA was added at a final concentration of 10 mM (+IAA), in the other was added DPBS. IAA-induced alkylation was performed overnight, in the dark, at room temperature and with a mild agitation. After washing the samples with fixation buffer two times, they were split again: half was used for RNA extraction dot-blot assay, while the other half was conserved at -80 °C for scRNA-seq. . Samples were then centrifuged (300 g, 2 minutes, 4 °C), supernatant removed and washed once with the fixation buffer. Cell pellets were resuspended in fixation buffer. 2/3 of the sample was used for RNA extraction

We used TRIzol Reagent (Thermo Fisher Scientific, 15-596-018) to purify the total RNA and we used 1-10 µg of purified RNA to be biotinylated with MTSEA-XX-biotin (Biotium, 90066, 10 µg/ml final concentration) in biotinylation buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, in ddH2O, 100 uL final volume) for > 1 hour in the dark at room temperature. Phenol/Chloroform extraction with Phase Lock Gel Heavy tubes (Quantabio, 2302830) removed the excess of MTS-XX-biotin. The samples were then blotted on a nylon membrane (Amersham Hybond-N+, RPN203B) and crosslinked with 2400 µJ UV 254 nm. The membrane was blocked in blocking solution (DPBS, 10% SDS, 1 mM EDTA) for 20 minutes at RT and probed with 1:10000 dilution of 1 mg/mL streptavidin-HRP (Pierce, 21130) in blocking solution. The membrane was washed six times with a blocking solution containing decreasing concentration of SDS (10%, 1%, 0.1%, applied twice each) for 10 minutes. The biotin signal was visualized by ECL detection reagent (GE Healthcare, RPN2209).

Single cell library preparation and sequencing

Single cell were resuspended in rehydration buffer (DPBS - BSA 0.01% + 1 U/µL 1:100 Superasein RNAse Inhibitor (Thermo Fisher)) and IAA was quenched wit 100 mM DTT for 5 minutes at RT. Cell suspensions were washed once, resuspended in rehydration buffer and passed through a cell strainer (40 µm). The obtained single cells were counted with a TC20 automated cell counter (Bio-Rad) and encapsulated using a Nadia device (Dolomite Bio) and following the version 1.8 of the protocol from the manufacturer. 250 µL of rehydration buffer containing ~ 75 000 cells were mixed with 3 mL QX200 Droplet Generation Oil (Bio-Rad #1864006) and 250 µL lysis buffer (, 0.2% Sarkosyl, 6% Ficoll PM-40020 mM EDTA, 200 mM Tris pH 7.5, 50 mM DTT). The encapsulated transcripts were reverse transcribed (Maxima H- RT enzyme, Thermo Fisher EP0753) and RNA was removed with exonuclease NEB (NEB #M0293). No second strand synthesis was carried out.

SMART PCR amplification was carried out in 6 replicates, each containing 4 000 beads, for 14 cycles. AMPure XP beads (Beckman Coulter, A63881) were used to purify the PCR reactions, which were then quantified with QuBit (Thermo Fisher Scientific) and Bioanalyzer DNA HS chip. Tagmentation of the libraries was carried out using the Nextera XT v2 DNA sample preparation kit (Illumina) and using 1000 pg of DNA and 11 amplification steps. After two purification with AMPure XP Beads, the libraries were quantified and pooled.The Illumina Nextseq500 sequencer was used for the 15 minutes labelling samples (library concentration 1.8 pM; Nextseq 500/550 High Output v2 kit (150 cycles) in paired-end mode; read 1 = 20/21 nt using the custom primer Read1CustSeqB (Macosko et al., 2015) , read 2 = 133/132 nt). The 0, 30 and 60 minutes 4sU-labeled samples were sequenced on a Illumina NovaSeq 6000 device (SP configuration, pair-end; read 1 = 20 nt using the custom primer Read1CustSeqB, index 1 (i7) = 8 nt, read 2 = 150 nt)

Data processing for SLAM-Drop-seq

Bcl2fasq v2.20.20 and FastQC v0.11.15 demultiplexed raw sequencing data and checked for its quality, respectively. Drop-seq tools v2.2.0 identified molecular and cellular barcodes. STAR v2.6.0a (Dobin et al., 2013) aligned reads to HEK293 genome and we filtered out non-uniquely mapped reads. Bedtools (Quinlan, 2014) assigned the exonic and/or intronic composition of reads using gtf annotation files.

T to C conversion per read were counted using an adapted version of NASC-seq analysis pipeline (Hendriks et al., 2019). To increase the coverage per read, we merged reads with the same UMI in each cell by collapsing overlapping regions, while keeping the annotation of the read as labelled or not labelled.

The splice status of a merged molecule was assigned “spliced” if all read were mapped to exonic regions, “intronic” if there was at least one read from an intro, “ambiguous” if reads did not fit one of the two criteria. Ambiguous reads were not used to estimate kinetic rates.
Gene expression was normalized to counts per million (CPM) and smoothed along the cell cycle using GAM (Wood, 2017). Gene expression correlation between samples was calculated comparing the average of gene-specific CPM over all single cells and using the common set of genes between the compared samples.

In silico cell cycle sorting

We first filtered cells containing more than 5% mitochondrial RNA and cells with high content of stress related genes.
Revelio (Schwabe et al., 2020) sorted the cells to a continuous cell cycle using the single cell data information. The intersection of default Revelio and variable genes were used to calculate PCA. The algorithm transform the high-dimensional data from single cell sequencing to a two dimensional circular trajectory and the cell-cycle progression is approximated using the angular component of it. Cell cycle phases boundaries and durations were determined using knowledge from previous studies (Cheng and Solomon, 2008) and the cell cycle was simplified to a circle, with a phase Φ values varying between 0 and 2π.

Time-dependent RNA kinetic model

Using 4sU labelling and splice annotation information, we obtained 4 RNA types: p_u - unlabelled precursors (pre-existing precursors), m_u - unlabelled matures (pre-existing matures), p_l - labelled precursors (newly synthesized precursors), m_l - labelled matures (newly synthesized matures)
The total amounts of precursor RNA p(Φ) and the total amounts of mature RNA m(Φ) are defined by:

$$p(\Phi) = p_u(\Phi, \mathrm{t}) + p_l(\Phi, \mathrm{t})$$
$$m(\Phi) = m_u(\Phi, \mathrm{t}) + m_l(\Phi, \mathrm{t})$$

Extending the commonly applied RNA kinetic rate model for steady state analysis (Manno et al., 2018; Zeisel et al., 2011), we obtained a system of four ordinary differential equations, which describes the changes in abundance over time of the four different RNA types, depending on the rate parameters of synthesis alpha, splicing beta and degradation gamma.

$$\frac{dp_l}{d\Phi} = \alpha(\Phi) - \beta(\Phi)p_l(\Phi, t)$$

$$\frac{dm_l}{d\Phi} = \beta(\Phi)p_l(\Phi, t) - \gamma(\Phi)m_l(\Phi, t)$$

$$\frac{dp_u}{d\Phi} = \beta(\Phi)p_u(\Phi, t)$$

$$\frac{dm_u}{d\Phi} = \beta(\Phi)p_u(\Phi, t) - \gamma(\Phi)m_u(\Phi, t)$$

Transcript splicing is usually much faster than degradation. This assumption led to a simplification of the model and, since in our data the average of precursor reads are less than 2%, we decided to further approximate the model and solve it to obtain the approximations:

$$\begin{cases} \hat{\gamma}(\Phi) = -\frac{1}{t}\log\left(\frac{m_u(\Phi, t)}{m(\Phi - \omega t)}\right), \\ \hat{\alpha}(\Phi) = -\frac{1}{t} \cdot \frac{m(\Phi - \omega t) \cdot m_l(\Phi, t)}{m(\Phi - \omega t) - m_u(\Phi, t)} \cdot \log\left(\frac{m_u(\Phi, t)}{m(\Phi - \omega t)}\right). \end{cases}$$

Profile peaks identification

Profiles of expression and kinetic rates were smoothed by penalized splines. The fold change between the maximum and minimum values of the profile was defined as fold change (fc).

$$fc(global) = \frac{\max(global)}{\min(global)}$$

A profile was considered to peak if the fc was greater than 1.5. The total number of peaks per profile were counted and genes with one peak were defined as cycling.

Deviation and half-life calculation

We defined the mean absolute deviation parameter (dev) in order to compare the gene expression calculated from estimated synthesis and degradation rates (pred) to the one calculated from scRNA-seq data (obs), with m indicating the cell number:

$$dev = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{|pred_i - obsv_i|}{obsv_i}$$

To check how the prediction changed upon setting one kinetic rate to a constant value, we calculated the deviation from constant (dev_c) as follows:

$$dev_c = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{|pred(constant\ rate)_i - pred_i|}{pred_i}$$

The half-life of transcripts was determined using the RNA decay rate. The averaged time-dependent decay rate over cells was calculated for each gene and derived the half life calculating:

$$t_{1/2} = \frac{ln(2)}{\bar{\gamma}}$$

Antibodies

Antibody	Vendor	Dilution
BCAP31	Proteintech, 11200-1-AP	1:1000
GAPDH	Cell Signaling, #2118	1:5000
Goat Anti-Mouse Ig/HRP	Agilent-Dako, P0447	1:2000
Goat Anti-Mouse Alexa 488	Thermo Fisher, A-11001	1:750
Goat Anti-Rabbit Ig/HRP	Agilent-Dako, P0448	1:2000
H3	Cell Signaling, #9715	1:5000
HA	Covance, MMS-101P-1000	1:2000
RPS6	Abcam, ab40820	1:4000
TOMM20	Scbt, SC-11415	1:2000
Beta-Tubulin	Sigma-Aldrich, T8328	1:4000

BIBLIOGRAPHY

Alles, J., Karaiskos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. Bmc Biol *15*, 44.

Alpert, T., Herzel, L., and Neugebauer, K.M. (2017). Perfect timing: splicing and transcription rates in living cells. Wiley Interdiscip Rev Rna *8*, e1401.

Baghirova, S., Hughes, B.G., Hendzel, M.J., and Schulz, R. (2015). Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. Methodsx *2*, 440–445.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Molecular Cell *46*, 674–690.

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat Methods *15*, 213–220.

Baptista, M.A.P., and Dölken, L. (2018). RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. Nat Methods *15*, 171–172.

Basu, S., and Bahadur, R.P. (2016). A structural perspective of RNA recognition by intrinsically disordered proteins. Cell Mol Life Sci *73*, 4075–4084.

Battich, N., Beumer, J., Barbanson, B. de, Krenning, L., Baron, C.S., Tanenbaum, M.E., Clevers, H., and Oudenaarden, A. van (2020). Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. Science *367*, 1151–1156.

Baumann, O., and Walz, B. (2001). Endoplasmic reticulum of animal cells and its organization into structural and functional domains. Int Rev Cytol *205*, 149–214.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun *6*, 10127.

Beckmann, B.M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. Pflugers Archiv *468*, 1029–1040.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Royal Statistical Soc Ser B Methodol *57*, 289–300.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol *38*, 1408–1414.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. Cell *125*, 1111–1124.

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N. (2008). Multiple pathways for sorting mitochondrial precursor proteins. Embo Rep *9*, 42–49.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nature Biotechnology *36*, 880–887.

Breeden, L.L. (2003). Periodic Transcription: A Cycle within a Cycle. Curr Biol *13*, R31–R38.

Bridges, R.J., and Bradbury, N.A. (2018). Targeting Trafficking in Drug Development. Handb Exp Pharmacol 385–425.

Brimacombe, R., Stiege, W., Kyriatsoulis, A., and Maly, P. (1988). [19] Intra-RNA and RNA—protein cross-linking techniques in Escherichia coli ribosomes. Methods Enzymol *164*, 287–309.

Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. Rna Biol *10*, 1623–1630.

Butter, F., Scheibe, M., Mörl, M., and Mann, M. (2009). Unbiased RNA–protein interaction screen by quantitative proteomics. Proc National Acad Sci *106*, 10626–10631.

Cao, J., Zhou, W., Steemers, F., Trapnell, C., and Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. Nat Biotechnol 1–9.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., and Steinmetz, L.M. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell *63*, 696–710.

Chakrabarti, R., Ji, W.-K., Stan, R.V., Sanz, J. de J., Ryan, T.A., and Higgs, H.N. (2018). INF2-mediated actin polymerization at the ER stimulates mitochondrial calcium uptake, inner membrane constriction, and division. J Cell Biology *217*, 251–268.

Chen, B., Sun, Y., Niu, J., Jarugumilli, G.K., and Wu, X. (2018). Protein Lipidation in Cell Signaling and Diseases: Function, Regulation, and Therapeutic Opportunities. Cell Chem Biol *25*, 817–831.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015). Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. Mol Syst Biol *11*, 781.

Chen, X., Li, A., Sun, B.-F., Yang, Y., Han, Y.-N., Yuan, X., Chen, R.-X., Wei, W.-S., Liu, Y., Gao, C.-C., et al. (2019). 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. Nat Cell Biol *21*, 978–990.

Cheng, A., and Solomon, M.J. (2008). Speedy/Ringo C regulates S and G 2 phase progression in human cells. Cell Cycle *7*, 3037–3047.

Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'Riordan, C.R., and Smith, A.E. (1990). Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell *63*, 827–834.

Cheung, Z.H., and Ip, N.Y. (2012). Cdk5: a multifaceted kinase in neurodegenerative diseases. Trends Cell Biol *22*, 169–175.

Cho, K.F., Branon, T.C., Rajeev, S., Svinkina, T., Udeshi, N.D., Thoudam, T., Kwak, C., Rhee, H.-W., Lee, I.-K., Carr, S.A., et al. (2020). Split-TurboID enables contact-dependent proximity labeling in cells. Proc National Acad Sci *117*, 12143–12154.

Cho, N.H., Cheveralls, K.C., Brunner, A.-D., Kim, K., Michaelis, A.C., Raghavan, P., Kobayashi, H., Savy, L., Li, J.Y., Canaj, H., et al. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. Science *375*, eabi6983.

Cho, R.J., Huang, M., Campbell, M.J., Dong, H., Steinmetz, L., Sapinoso, L., Hampton, G., Elledge, S.J., Davis, R.W., and Lockhart, D.J. (2001). Transcriptional regulation and function during the human cell cycle. Nat Genet *27*, 48–54.

Cohen, S., Valm, A.M., and Lippincott-Schwartz, J. (2018). Interacting organelles. Curr Opin Cell Biol *53*, 84–91.

Consortium, T.G.O. (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res *43*, D1049–D1056.

Consortium, T.U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., et al. (2020). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res *49*, D480–D489.

Cooper, S. (2019). The synchronization manifesto: a critique of whole-culture synchronization. Febs J *286*, 4650–4656.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845–1848.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J Proteome Res *10*, 1794–1805.

Darzynkiewicz, Z., Halicka, H.D., Zhao, H., and Podhorecka, M. (2011). Cell Cycle Synchronization, Methods and Protocols. Methods Mol Biology *761*, 85–96.

Das, S., Vera, M., Gandin, V., Singer, R.H., and Tutucci, E. (2021). Intracellular mRNA transport and localized translation. Nat Rev Mol Cell Bio *22*, 483–504.

Davey, K., Wong, D., Konopacki, F., Kwa, E., Ly, T., Fiegler, H., and Sibley, C.R. (2021). A flexible microfluidic system for single-cell transcriptome profiling elucidates phased transcriptional regulators of cell cycle. Sci Rep-Uk *11*, 7918.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Molecular Cell *59*, 858–866.

Duve, C. de (1975). Exploring Cells with a Centrifuge. Science *189*, 186–194.

Dynlacht, B.D. (1997). Regulation of transcription by proteins that control the cell cycle. Nature *389*, 149–152.

Engel, K.L., Arora, A., Goering, R., Lo, H.G., and Taliaferro, J.M. (2020). Mechanisms and consequences of subcellular RNA localization across diverse cell types. Traffic *21*, 404–418.

Erhard, F., Baptista, M.A., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C.S., Theis, F.J., Saliba, A.-E., and Dölken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. Nature 1–6.

Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirkl, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. Mol Syst Biol *10*, 717.

Fan, J., Li, X., Issop, L., Culty, M., and Papadopoulos, V. (2016). ACBD2/ECI2-Mediated Peroxisome-Mitochondria Interactions in Leydig Cell Steroid Biosynthesis. Mol Endocrinol *30*, 763–782.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. Cell *178*, 473-490.e26.

Fischer, U., Huber, J., Boelens, W.C., Mattajt, L.W., and Lührmann, R. (1995). The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475–483.

Fransen, M., Lismont, C., and Walton, P. (2017). The Peroxisome-Mitochondria Connection: How and Why? Int J Mol Sci *18*, 1126.

Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites*[S]. Mol Cell Proteomics *14*, 1113–1126.

Furlan, M., Pretis, S. de, and Pelizzola, M. (2020a). Dynamics of transcriptional and post-transcriptional regulation. Brief Bioinform bbaa389-.

Furlan, M., Galeota, E., Gaudio, N.D., Dassi, E., Caselle, M., Pretis, S. de, and Pelizzola, M. (2020b). Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling. Genome Res *30*, 1492–1507.

Gasser, C., Delazer, I., Neuner, E., Pascher, K., Brillet, K., Klotz, S., Trixl, L., Himmelstoß, M., Ennifar, E., Rieder, D., et al. (2020). Thioguanosine Conversion Enables mRNA-Lifetime Evaluation by RNA Sequencing Using Double Metabolic Labeling (TUC-seq DUAL). Angewandte Chemie Int Ed Engl *59*, 6881–6886.

Gatto, L., and Lilley, K.S. (2012). MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics *28*, 288–289.

Gatto, L., Breckels, L.M., Burger, T., Nightingale, D.J.H., Groen, A.J., Campbell, C., Nikolovski, N., Mulvey, C.M., Christoforou, A., Ferro, M., et al. (2014). A Foundation for Reliable Spatial Proteomics Data Analysis*. Mol Cell Proteomics *13*, 1937–1952.

Gauthier, N.P., Larsen, M.E., Wernersson, R., Lichtenberg, U. de, Jensen, L.J., Brunak, S., and Jensen, T.S. (2008). Cyclebase.org—a comprehensive multi-organism online database of cell-cycle experiments. Nucleic Acids Res *36*, D854–D859.

Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. Nat Rev Genet *22*, 185–198.

Geladaki, A., Britovšek, N.K., Breckels, L.M., Smith, T.S., Vennard, O.L., Mulvey, C.M., Crook, O.M., Gatto, L., and Lilley, K.S. (2019). Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. Nature Communications *10*, 331.

Grant, G.D., Brooks, L., Zhang, X., Mahoney, J.M., Martyanov, V., Wood, T.A., Sherlock, G., Cheng, C., and Whitfield, M.L. (2013). Identification of cell cycle–regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. Mol Biol Cell *24*, 3634–3650.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5′ to 3′ RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3′ UTRs. Molecular Cell *54*, 573–585.

Guardia, C.M., Pace, R.D., Mattera, R., and Bonifacino, J.S. (2018). Neuronal functions of adaptor complexes involved in protein sorting. Curr Opin Neurobiol *51*, 103–110.

Guo, L., and Shorter, J. (2017). Biology and Pathobiology of TDP-43 and Emergent Therapeutic Strategies. Csh Perspect Med *7*, a024554.

Hamasaki, M., Furuta, N., Matsuda, A., Nezu, A., Yamamoto, A., Fujita, N., Oomori, H., Noda, T., Haraguchi, T., Hiraoka, Y., et al. (2013). Autophagosomes form at ER–mitochondria contact sites. Nature *495*, 389–393.

Han, S., Udeshi, N.D., Deerinck, T.J., Svinkina, T., Ellisman, M.H., Carr, S.A., and Ting, A.Y. (2017). Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. Cell Chem Biol *24*, 404–414.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Harris, M.E., Böhni, R., Schneiderman, M.H., Ramamurthy, L., Schümperli, D., and Marzluff, W.F. (1991). Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. Mol Cell Biol *11*, 2416–2424.

Helder, S., Blythe, A.J., Bond, C.S., and Mackay, J.P. (2016). Determinants of affinity and specificity in RNA-binding proteins. Curr Opin Struc Biol *38*, 83–91.

Hendriks, G.-J., Jung, L.A., Larsson, A.J.M., Lidschreiber, M., Forsman, O.A., Lidschreiber, K., Cramer, P., and Sandberg, R. (2019). NASC-seq monitors RNA synthesis in single cells. Nat Commun *10*, 3138.

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

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., Haeseler, A. von, Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nature Methods *14*, 1198–1204.

Holler, K., Neuschulz, A., Drewe-Boß, P., Mintcheva, J., Spanjaard, B., Arsiè, R., Ohler, U., Landthaler, M., and Junker, J.P. (2021). Spatio-temporal mRNA tracking in the early zebrafish embryo. Nat Commun *12*, 3358.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Huang, R., Han, M., Meng, L., and Chen, X. (2018). Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. Proc National Acad Sci *115*, 201718406.

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc *14*, 68–85.

Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Hung, V., Lam, S.S., Udeshi, N.D., Svinkina, T., Guzman, G., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2017). Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. ELife *6*.

Hwang, C., Sinskey, A.J., and Lodish, H.F. (1992). Oxidized Redox State of Glutathione in the Endoplasmic Reticulum. Science *257*, 1496–1502.

Jackson, M.R., Nilsson, T., and Peterson, P.A. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. Embo J *9*, 3153–3162.

Jagannathan, S., Nwosu, C., and Nicchitta, C.V. (2011). Analyzing Subcellular mRNA localization via cell fractionation. Methods in Molecular Biology (Clifton, NJ) *714*, 301.

Jiang, Y., Wang, D., Wang, W., and Xu, D. (2021). Computational methods for protein localization prediction. Comput Struct Biotechnology J *19*, 5834–5844.

Jürges, C., Dölken, L., and Erhard, F. (2018). Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinformatics *34*, i218–i226.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol *338*, 1027–1036.

Kawata, K., Wakida, H., Yamada, T., Taniue, K., Han, H., Seki, M., Suzuki, Y., and Akimitsu, N. (2020). Metabolic labeling of RNA using multiple ribonucleoside analogs enables the simultaneous evaluation of RNA synthesis and degradation rates. Genome Res *30*, 1481–1491.

Koh, S.-B., Mascalchi, P., Rodriguez, E., Lin, Y., Jodrell, D.I., Richards, F.M., and Lyons, S.K. (2016). A quantitative FastFUCCI assay defines cell cycle dynamics at a single-cell level. J Cell Sci *130*, 512–520.

Kong, S.E., Banks, C.A.S., Shilatifard, A., Conaway, J.W., and Conaway, R.C. (2005). ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. P Natl Acad Sci Usa *102*, 10094–10098.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin alpha*. J Biol Chem *284*, 478–485.

Kozlowski, L.P. (2017). Proteome-pI: proteome isoelectric point database. Nucleic Acids Res *45*, D1112–D1116.

Kurotani, A., Tokmakov, A.A., Sato, K.-I., Stefanov, V.E., Yamada, Y., and Sakurai, T. (2019). Localization-specific distributions of protein pI in human proteome are governed by local pH and membrane charge. Bmc Mol Cell Biology *20*, 36.

Kwak, C., Shin, S., Park, J.-S., Jung, M., Nhung, T.T.M., Kang, M.-G., Lee, C., Kwon, T.-H., Park, S.K., Mun, J.Y., et al. (2020). Contact-ID, a tool for profiling organelle contact sites, reveals regulatory proteins of mitochondrial-associated membrane formation. Proc National Acad Sci *117*, 12109–12120.

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

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol *20*, 1122–1130.

Lam, L.T., Pickeral, O.K., Peng, A.C., Rosenwald, A., Hurt, E.M., Giltnane, J.M., Averett, L.M., Zhao, H., Davis, R.E., Sathyamoorthy, M., et al. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. Genome Biol *2*, research0041.1.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K., and Ting, A.Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. Nature Methods *12*, 51–54.

Lee, T.I., and Young, R.A. (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell *152*, 1237–1251.

de-Leon, S.B.-T., and Davidson, E.H. (2007). Gene Regulation: Gene Control Network in Development. Biophysics Biomol Struct *36*, 191–212.

Leonetti, M.D., Sekine, S., Kamiyama, D., Weissman, J.S., and Huang, B. (2016). A scalable strategy for high-throughput GFP tagging of endogenous human proteins. Proc National Acad Sci *113*, E3501–E3508.

Ligasová, A., and Koberna, K. (2021). Strengths and Weaknesses of Cell Synchronization Protocols Based on Inhibition of DNA Synthesis. Int J Mol Sci *22*, 10759.

Liu, Y., Chen, S., Wang, S., Soares, F., Fischer, M., Meng, F., Du, Z., Lin, C., Meyer, C., DeCaprio, J.A., et al. (2017). Transcriptional landscape of the human cell cycle. Proc National Acad Sci *114*, 3473–3478.

Loh, K.H., Stawski, P.S., Draycott, A.S., Udeshi, N.D., Lehrman, E.K., Wilton, D.K., Svinkina, T., Deerinck, T.J., Ellisman, M.H., Stevens, B., et al. (2016). Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. Cell *166*, 1295-1307.e21.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Manno, G.L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res *39*, D225–D229.

Markmiller, S., Soltanieh, S., Server, K.L., Mak, R., Jin, W., Fang, M.Y., Luo, E.-C., Krach, F., Yang, D., Sen, A., et al. (2018). Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. Cell *172*, 590-604.e13.

Maticzka, D., Ilik, I.A., Aktas, T., Backofen, R., and Akhtar, A. (2018). uvCLAP is a fast and non-radioactive method to identify in vivo targets of RNA-binding proteins. Nat Commun *9*, 1142.

Matthews, H.K., Bertoli, C., and Bruin, R.A.M. de (2022). Cell cycle control in cancer. Nat Rev Mol Cell Bio *23*, 74–88.

Mayer, A., and Churchman, S.L. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. Nature Protocols *11*, 813–833.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol *7*, 458–458.

Mullari, M., Lyon, D., Jensen, L.J., and Nielsen, M.L. (2017). Specifying RNA-Binding Regions in Proteins by Peptide Cross-Linking and Affinity Purification. J Proteome Res *16*, 2762–2772.

Munro, S., and Pelham, H.R.B. (1987). A C-terminal signal prevents secretion of luminal ER proteins. Cell *48*, 899–907.

Murakawa, Y., Hinz, M., Mothes, J., Schuetz, A., Uhl, M., Wyler, E., Yasuda, T., Mastrobuoni, G., Friedel, C.C., Dölken, L., et al. (2015). RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-κB pathway. Nat Commun *6*, 7367.

Nekrasov, M.P., Ivshina, M.P., Chernov, K.G., Kovrigina, E.A., Evdokimova, V.M., Thomas, A.A.M., Hershey, J.W.B., and Ovchinnikov, L.P. (2003). The mRNA-binding Protein YB-1 (p50) Prevents Association of the Eukaryotic Initiation Factor eIF4G with mRNA and Inhibits Protein Synthesis at the Initiation Stage*. J Biol Chem *278*, 13936–13943.

Nguyen, N.N.T., Lim, Y.-S., Nguyen, L.P., Tran, S.C., Luong, T.T.D., Nguyen, T.T.T., Pham, H.T., Mai, H.N., Choi, J.-W., Han, S.-S., et al. (2018). Hepatitis C Virus Modulates Solute carrier family 3 member 2 for Viral Propagation. Sci Rep-Uk *8*, 15486.

Nielsen, H. (2017). Protein Sorting Prediction. Methods Mol Biology *1615*, 23–57.

Okeke, E., Dingsdale, H., Parker, T., Voronina, S., and Tepikin, A.V. (2016). Endoplasmic reticulum-plasma membrane junctions: structure, function and dynamics: Endoplasmic reticulum-plasma membrane junctions. J Physiology *594*, 2837–2847.

Perez-Perri, J.I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., and Hentze, M.W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nat Commun *9*, 4408.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc *9*, 171–181.

Pines, J. (1999). Four-dimensional control of the cell cycle. Nat Cell Biol *1*, E73–E79.

Piovesan, D., Necci, M., Escobedo, N., Monzon, A.M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., et al. (2020). MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res *49*, gkaa1058-.

Pretis, S. de, Kress, T., Morelli, M.J., Melloni, G.E.M., Riva, L., Amati, B., and Pelizzola, M. (2015). INSPEcT: a computational tool to infer mRNA synthesis, processing and degradation dynamics from RNA- and 4sU-seq time course experiments. Bioinformatics *31*, 2829–2835.

Protter, D.S.W., Rao, B.S., Treeck, B.V., Lin, Y., Mizoue, L., Rosen, M.K., and Parker, R. (2018). Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. Cell Reports *22*, 1401–1412.

Qin, W., Myers, S.A., Carey, D.K., Carr, S.A., and Ting, A.Y. (2021). Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery. Nat Commun *12*, 4980.

Qiu, Q., Hu, P., Qiu, X., Govek, K.W., Cámara, P.G., and Wu, H. (2020). Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. Nat Methods 1–11.

Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat Biotechnol *37*, 169–178.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinform *47*, 11.12.1-11.12.34.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nature Biotechnology *29*, 436–442.

Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N., et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell *159*, 1698–1710.

Raffetseder, U., Frye, B., Rauen, T., Jürchott, K., Royer, H.-D., Jansen, P.L., and Mertens, P.R. (2003). Splicing Factor SRp30c Interaction with Y-box Protein-1 Confers Nuclear YB-1 Shuttling and Alternative Splice Site Selection*. J Biol Chem *278*, 18241–18248.

Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl, M., and Bohjanen, P.R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. Nucleic Acids Res *30*, 5529–5538.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. Anal Chem *75*, 663–670.

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2020). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res gkaa1011-.

Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A.-M., Horos, R., Hentze, M.W., Preiss, T., and Millar, A.A. (2016). In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. Plant Cell *28*, 2435–2452.

Reilly, S.K., and Noonan, J.P. (2014). Evolution of Gene Regulation in Humans. Annu Rev Genom Hum G *17*, 1–23.

Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. Nat Rev Nephrol *15*, 346–366.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. Science *339*, 1328–1331.

Riml, C., Amort, T., Rieder, D., Gasser, C., Lusser, A., and Micura, R. (2017). Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. Angewandte Chemie Int Ed *56*, 13479–13483.

Ripin, N., Boudet, J., Duszczyk, M.M., Hinniger, A., Faller, M., Krepl, M., Gadi, A., Schneider, R.J., Šponer, J., Meisner-Kober, N.C., et al. (2019). Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. P Natl Acad Sci Usa *116*, 2935–2944.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47–e47.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biology *196*, 801–810.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. Cell *132*, 487–498.

Savojardo, C., Martelli, P.L., Fariselli, P., and Casadio, R. (2018). DeepSig: deep learning improves signal peptide detection in proteins. Bioinformatics *34*, 1690–1696.

Schafer, K.A. (1998). The Cell Cycle: A Review. Vet Pathol *35*, 461–478.

Schnell, U., Dijk, F., Sjollema, K.A., and Giepmans, B.N.G. (2012). Immunolabeling artifacts and the need for live-cell imaging. Nat Methods *9*, 152–158.

Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat Methods *15*, 221–225.

Schwabe, D., Formichetti, S., Junker, J.P., Falcke, M., and Rajewsky, N. (2020). The transcriptome dynamics of single cells during the cell cycle. Mol Syst Biol *16*, e9946.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science (New York, N.Y.) *352*, 1225–1228.

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*.

Sehgal, P.B., and Lee, J.E. (2011). Protein Trafficking Dysfunctions: Role in the Pathogenesis of Pulmonary Arterial Hypertension. Pulm Circ *1*, 17–32.

Serna, J.B. de la, Schütz, G.J., Eggeling, C., and Cebecauer, M. (2016). There Is No Simple Model of the Plasma Membrane Organization. Frontiers Cell Dev Biology *4*, 106.

Shoshan-Barmatz, V., Zalk, R., Gincel, D., and Vardi, N. (2004). Subcellular localization of VDAC in mitochondria and ER in the cerebellum. Biochimica Et Biophysica Acta Bba - Bioenergetics *1657*, 105–114.

Shurtleff, M.J., Temoche-Diaz, M.M., Karfilis, K.V., Ri, S., and Schekman, R. (2016). Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. Elife *5*, e19276.

Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989). The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Gene Dev *3*, 60–72.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem *84*, 1–30.

Smith, K.C., and Meun, D.H.C. (1968). Kinetics of the photochemical addition of cysteine- 35 S to polynucleotides and nucleic acids. Biochemistry-Us *7*, 1033–1037.

Spinelli, J.B., and Haigis, M.C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol *20*, 745–754.

Stewart, M. (2019). Polyadenylation and nuclear export of mRNAs. J Biol Chem *294*, 2977–2987.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res *22*, 1350–1359.

Sunbul, M., and Jäschke, A. (2019). Proximity Labeling, Methods and Protocols.

Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A., and Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat Commun *7*, 12128.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res *22*, 947–956.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403.e19.

Tsvetanova, N.G., Klass, D.M., Salzman, J., and Brown, P.O. (2010). Proteome-Wide Search Reveals Unexpected RNA-Binding Proteins in Saccharomyces cerevisiae. Plos One *5*, e12671.

Uhlitz, F., Bischoff, P., Peidli, S., Sieber, A., Trinks, A., Lüthen, M., Obermayer, B., Blanc, E., Ruchiy, Y., Sell, T., et al. (2021). Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. Embo Mol Med *13*, e14123.

Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., et al. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. Nat Commun *10*, 990.

Uvarovskii, A., Vries, I.S.N., and Dieterich, C. (2019). On the optimal design of metabolic RNA labeling experiments. Plos Comput Biol *15*, e1007252.

Varadi, M., Zsolyomi, F., Guharoy, M., and Tompa, P. (2015). Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. PLOS ONE *10*, e0139731.

Wanders, R.J.A., Waterham, H.R., and Ferdinandusse, S. (2016). Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. Frontiers Cell Dev Biology *3*, 83.

Wessels, H.-H.H., Imami, K., Baltz, A.G., Kolinksi, M., Beldovskaya, A., Selbach, M., Small, S., Ohler, U., and Landthaler, M. (2016). The mRNA-bound proteome of the early fly embryo. Genome Research.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., et al. (2002). Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. Mol Biol Cell *13*, 1977–2000.

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol *20*, 59.

Wood, S.N. (2017). Generalized Additive Models.

Yan, S., Zhao, D., Wang, C., Wang, H., Guan, X., Gao, Y., Zhang, X., Zhang, N., and Chen, R. (2021). Characterization of RNA-binding proteins in the cell nucleus and cytoplasm. Anal Chim Acta *1168*, 338609.

Yang, E., Nimwegen, E. van, Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J.E. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res *13*, 1863–1872.

Yu, S.-H., Kyriakidou, P., and Cox, J. (2020). Isobaric Matching between Runs and Novel PSM-Level Normalization in MaxQuant Strongly Improve Reporter Ion-Based Quantification. J Proteome Res *19*, 3945–3954.

Zeisel, A., Köstler, W.J., Molotski, N., Tsai, J.M., Krauthgamer, R., Jacob-Hirsch, J., Rechavi, G., Soen, Y., Jung, S., Yarden, Y., et al. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. Mol Syst Biol *7*, 529.

Zhang, J., Fan, J.-S., Li, S., Yang, Y., Sun, P., Zhu, Q., Wang, J., Jiang, B., Yang, D., and Liu, M. (2020). Structural basis of DNA binding to human YB-1 cold shock domain regulated by phosphorylation. Nucleic Acids Res *48*, 9361–9371.

Zhang, X., Smits, A.H., Tilburg, G. van, Ovaa, H., Huber, W., and Vermeulen, M. (2018). Proteome-wide identification of ubiquitin interactions using UbIA-MS. Nat Protoc *13*, 530–550.

Zimmermann, R. (2009). Protein Transport into the Endoplasmic Reticulum. 21–36.

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben.

Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad.

Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde.

Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin vom 5. März 2015.

Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsbearbeiterinnen/ Promotionsberatern stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Datum ..... Unterschrift .....

ABBREVIATIONS

4sU: 4-thiouridine
APEX : ascorbate peroxidase
BirA*: BirA-R118G
CDS: coding sequence
COPII: coat protein complex II
CPM: counts per million
DMEM: Dulbecco's modified Eagle medium
DNA : deoxyribonucleic acid
ER : endoplasmic reticulum
FBS : fetal bovine serum
GO : gene ontology
GOCC: GO cellular component
HEK : human embryonic kidney
HEK293: HEK293 Flp-In T-Rex
HRP : horseradish peroxidase
IAA: iodoacetamide
iBAQ : intensity-based absolute
IF: immunofluorescence
IMM: inner mitochondrial membrane
IMS: intramitochondrial space
LC-MS/MS: liquid chromatography tandem MS
MITO: mitochondria
MS: mass spectrometry
mRNA : messenger RNA
ODE : ordinary differential equation
OMM: outer mitochondrial membrane
PBS: phosphate buffer saline
PCA: principal component analysis
pI: isoelectric point
PL: proximity labelling
PM: plasma membrane
RBD: RNA-binding domain
RBP: RNA-binding protein
RIC: RNA-interactome capture
RIN: RNA integrity number
RNA: ribonucleic acid
RRM: RNA-recognition motif
T to C: thymidine to cytosine transitions
UMI: unique Molecular Identifier
UTR: untranslated region
UV: ultraviolet light

PUBLICATIONS

1. Holler, K., Neuschulz, A., Drewe-Boß, P., Mintcheva, J., Spanjaard, B., **Arsiè, R.**, Ohler, U., Landthaler, M. & Junker, J. P. Spatio-temporal mRNA tracking in the early zebrafish embryo. Nat Commun 12, 3358 (2021).

2. Wyler, E., Mösbauer, K., Franke, V., Diag, A., Gottula, L. T., **Arsiè, R**., Klironomos, F., Koppstein, D., Hönzke, K., Ayoub, S., Buccitelli, C., Hoffmann, K., Richter, A., Legnini, I., Ivanov, A., Mari, T., Giudice, S. D., Papies, J., Praktiknjo, S., Meyer, T. F., Müller, M. A., Niemeyer, D., Hocke, A., Selbach, M., Akalin, A., Rajewsky, N., Drosten, C. & Landthaler, M. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. Iscience 24, 102151 (2021).

3. Uhlitz, F., Bischoff, P., Peidli, S., Sieber, A., Trinks, A., Lüthen, M., Obermayer, B., Blanc, E., Ruchiy, Y., Sell, T., Mamlouk, S., **Arsie, R.**, Wei, T., Klotz-Noack, K., Schwarz, R. F., Sawitzki, B., Kamphues, C., Beule, D., Landthaler, M., Sers, C., Horst, D., Blüthgen, N. & Morkel, M. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. Embo Mol Med 13, e14123 (2021).

4. Liu H.*, **Arsiè R.***, Schwabe D., M. Schilling, I. Minia, J. Alles, A. Boltengagen, C. Kocks, M. Falcke, N. Friedman, M. Landthaler, N. Rajewsky SLAM-Drop-seq reveals transcriptome-wide RNA kinetic rates throughout the cell cycle (in preparation)(Chen et al., 2019)

ACKNOWLEDGMENTS

A famous quote cites: *'It's difficult to make predictions, especially about the future'*. Well, it is... The title of my poster at the first conference I attended as PhD student was 'Transcriptome-wide mapping of *in vivo* RNA-RNA interactions using biotin-psoralen'. Well, if you have read/skimmed this thesis, you might have realized something changed along the way…

Many *Homo Sapiens* have to be acknowledged for the person and scientist I have become. First, I have to thank Markus Landthaler, who believed and supported me throughout these years. He showed me the good (and also the bad) of the scientific world. Thank you and I won't forget: 'Always look forward!'.

Next, I have to acknowledge the scientific companions who shared this journey with me in the laboratory. I have to start with the 'primordial' PhD group: Ulrike, the most organized, reliable and modest PhD I know. Thank you for trying to make me a more organized and less chaotic person. You failed. My mate Karol, such a great scientist and funny lad. Thank you for the infinite and scientifically thrilling conversations on the balcony in Buch. We still miss you (hire me!). The grumpy Nico: we know that under that fur you have the biggest heart! And a lot of scientific skills too. Thank you for always being up for talking about science (and also not-science!). Marcin, who taught me how to do PAR-CLIP and run 1 meter gels. Rubensito, for the fun time together and more. Jonas and the socio-political questions always interesting to discuss. Orsalia for trying to make sense of my RNA structure project. And thanks also to my Italian buddy Simone, who passed a lot of time with me in 'my second office' of BIMSB-Mitte. I also have to acknowledge the post-doc team in Landthaler lab. The great Miha, who has always helpful feedback to give, even without an elevating desk. Ema: I will never forget the lockdown months spent with you in the lab and the cous-cous lunches, so epic. Thank you for your help throughout all these years. The biochemical monster Igor, whose knowledge and creativity never stops to surprise me. Thank you for teaching me so much throughout these years, without judgement on my lack of understanding on topics I should know. One suggestion though: use the Trizol under the hood like me! And I also have to acknowledge Gustavo, with his great energy and positivity, he brought light to the lab.

More people in other laboratories contributed greatly to my PhD. Foremost, Haiyue, Marcel and Daniel, trying to make sense of the mole of single cell data dumped on them. Alessandra and the long pauses together on the BIMSB terrace. I miss not seeing you every day! Thank you JJ for always being so friendly, knowledgeable and reliable. It has been great having you around these years!

And of course I need to thank all the people outside the work environment, who have been great in supporting me (and in asking 'So, are you finished with your PhD or what?'). Grazie alla mia famiglia, che mi ha sempre sostenuto durante il lungo percorso del dottorato e che ha creduto in me. Grazie, sapere che ci siete (anche se un po' distanti) mi dà sempre forza. Thanks to my cani: Jan, Niccoló and Alexito. I spent some of the best and most funny moments with you. And many have yet to come! And finally, I want to thank the most special person of my life, who stood by me when I was excited and when I was crying, who gave me hope, strength and happiness. Thank you Aoife for the infinite amount of love you keep pouring into me, you are the brightest light of my life.