

DISSERTATION
SUBMITTED TO THE
COMBINED FACULTY OF NATURAL SCIENCES AND MATHEMATICS
OF THE
RUPERTO
CAROLA
UNIVERSITY
HEIDELBERG, GERMANY
FOR THE DEGREE OF
DOCTOR OF NATURAL SCIENCES

PRESENTED BY JAKOB TRENDEL, MSc. MOLECULAR MEDICINE
BORN IN ULM, GERMANY
ORAL EXAMINATION ON THE 20TH OF MARCH 2019

PROTEOMIC EXPLORATION OF PROTEIN-RNA INTERACTIONS
IN HUMAN CELLS

REFEREES:
DR. NASSOS TYPAS
PROF. DR. ANDRES JÄSCHKE

SUMMARY

The interactions of proteins with RNA are ubiquitous in human cell biology and fundamental to most if not all of life as we know it. Proteins and RNA interact for the purpose of gene expression, where genetic information is encoded within RNA, which serves as template for protein in messenger RNA, as scaffold for the trimming of other RNA within the spliceosome, as a catalyst for the production of protein within the ribosome, and as component of various other processes, many of which we are only beginning to understand. Central to our understanding of such processes is the interplay between proteins and RNA. Protein-RNA interactions are often disrupted by conventional extraction methods and need to be stabilized through crosslinking in order to be appreciated by transcriptomic or proteomic methodology. Crosslinking typically occurs through UV irradiation, which covalently connects protein bound to RNA. However, currently there is no universal method for the extraction of protein-crosslinked RNA available, so that insights can only be gained for certain subsets of the transcriptome or individual RNA-binding proteins, respectively. In this thesis a new method for the extraction of protein-crosslinked RNA from UV-crosslinked cells is described. In order to demonstrate the comprehensive usefulness of the approach, which was termed XRNAX, it is applied for the resolution of several proteomic and transcriptomic problems. First, XRNAX is employed for the purification of ribonucleotide-crosslinked peptides identifying known and unknown protein-RNA interfaces. Next, a comprehensive draft for the human RNA-binding proteome is derived applying XRNAX to three commonly used cell lines. The same approach is applied for the differential quantification of RNA-binding during a timeline of arsenite-induced translational arrest in human cells. Additionally, XRNAX is combined with CLIPseq to monitor the RNA exosome component EXOSC2 processing pre-ribosomal RNA during arsenite stress. At last, XRNAX is used along with TMT-SILAC in order to derive protein half-lives of the human RNA-bound proteome. Half-lives of the RNA-bound proteome are found on average 1.75 fold increased in comparison to the total proteome. For validating the stabilization of protein-RNA complexes, ribosomal assemblies are purified from human cells using polysome profiling and protein half-lives assessed with TMT-SILAC. Stabilization is confirmed, however, only for ribosomal proteins in 80S ribosomes within the 80S or polysome fractions. Further experiments interfering with ribosome biogenesis, the proteasome and autophagy are distilled into a model for the turnover of ribosomal proteins and ribosomal assemblies. In summary, this thesis describes new methodology and biological insight on protein-RNA interactions in human cells.

ZUSAMMENFASSUNG

Die Interaktionen von Proteinen mit RNA sind allgegenwärtig in der menschlichen Zellbiologie und fundamental für das meiste, wenn nicht das gesamte, uns bekannte Leben. Proteine und RNA interagieren zum Zweck der Genexpression, während welcher genetische Information in RNA codiert wird, welche als Vorlage für Messenger RNA, als Gerüst zur Beschneidung anderer RNA innerhalb des Spliceosoms, als Katalysator zur Produktion von Protein innerhalb des Ribosoms und als Komponente einer Vielzahl anderer Prozessen dient, viele welche wir erst beginnen zu verstehen. Zentral für unser Verständnis solcher Prozesse ist das Zusammenspiel zwischen Proteinen und RNA. Protein-RNA Interaktionen werden durch konventionelle Extraktionsmethoden oftmals getrennt und müssen deshalb durch Crosslinking stabilisiert werden, um durch Methoden der Transcriptomic oder Proteomic erfasst zu werden. Crosslinking wird typischerweise durch UV Bestrahlung herbeigeführt, welche RNA-gebundene Proteine kovalent mit dieser verbindet. Allerdings ist derzeit keine universelle Methode zur Extraktion von RNA, die an Protein gecrosslinkt ist, verfügbar, sodass Einsichten nur für bestimmte Teile des Transcriptoms bzw. für einzelne RNA-bindende Proteine erlangt werden können. In dieser Doktorarbeit wird eine neue Methode zur Extraktion von Protein-gecrosslinkter RNA aus UV bestrahlten Zellen beschrieben. Um die allgemeine Nützlichkeit diese Methode zu demonstrieren, welche XRNAX genannt wurde, wurde sie auf verschiedene Probleme der Transcriptomic und Proteomic angewendet. Zunächst wurde sie zur Aufreinigung von Ribonukleotid-gecrosslinkten Peptiden verwendet, welche Aufschluss über bekannte und unbekannte Oberflächen zwischen Proteinen und RNA gewähren. Folglich wurde XRNAX auf drei gemeinhin verwendete Zelllinien angewendet, um einen vollumfänglichen Entwurf über das RNA-bindende Proteom abzuleiten. Die gleiche Herangehensweise wurde verwendet, um RNA Bindung differentiell in einer Zeitreihe zu messen, in welcher translationaler Arrest in menschlichen Zellen durch Arsen herbeigeführt wurde. Darüber hinaus wurde XRNAX mit CLIPseq kombiniert, um die Prozessierung von pre-ribosomaler RNA durch EXOSC2, einer Komponente des RNA Exosoms, während des Arsenstresses zu verfolgen. Zuletzt wurde XRNAX in Verbindung mit TMT-SILAC verwendet, um die Proteinhalbwertszeiten innerhalb des menschlichen RNA-bindenden Proteoms zu bestimmen. Es wurden Halbwertszeiten des RNA-gebundenen Proteoms festgestellt, welche im Durchschnitt 1.75-fach über denen des gesamten Proteoms liegen. Zur Validierung dieser Stabilisierung von Protein-RNA Komplexen wurden ribosomale Komplexe aus menschlichen Zellen durch Polysome Profiling aufgereinigt und Proteinhalbwertszeiten mittels TMT-SILAC bestimmt. Die Stabilisierung wurde bestätigt, allerdings nur für ribosomale Proteine innerhalb der 80S Ribosomen aus der 80S oder Polysomen Fraktion. Weiterführende Experimente zur Störung der ribosomalen Biogenese, des Proteasoms und von Autophagie wurden in ein Modell integriert, welches die Umsetzung von ribosomalen Proteinen und ribosomalen Komplexen beschreibt.

Zusammenfassend beschreibt diese Doktorarbeit neue Methodik und biologische Erkenntnisse zu Protein-RNA Interaktionen in menschlichen Zellen.

TABLE OF CONTENTS

SUMMARY	5
ZUSAMMENFASSUNG	7
TABLE OF CONTENTS	9
INTRODUCTION	12
PROTEIN-RNA INTERACTIONS ARE UBIQUITOUS.....	12
CURRENT METHODOLOGY FOR THE INTERROGATION OF PROTEIN-RNA INTERACTIONS	14
<i>RNA Sequencing-Based Methodology</i>	14
<i>Mass Spectrometry-Based Methodology</i>	18
CURRENT UNDERSTANDING OF RNA-BINDING PROTEOMES.....	22
<i>Many Human Proteins Interact with Polyadenylated RNA</i>	22
<i>Many Emerging Regulatory Functions Are Based on Interactions with Non-Polyadenylated RNA</i>	23
<i>Many Organisms Make Different Use of Polyadenylation</i>	24
AIM OF THIS THESIS PROJECT AND BIOLOGICAL QUESTIONS.....	24
RESULTS	26
COPYRIGHT DISCLAIMER	26
CHAPTER 1: XRNAX AS PLATFORM FOR INTERROGATING PROTEIN-RNA INTERACTIONS	27
<i>Extending TRIZOL Extraction to Protein-Crosslinked RNA</i>	27
The Behaviour of Protein-Crosslinked RNA During Conventional TRIZOL Extraction	27
Extraction of Protein-Crosslinked RNA From UV-Crosslinked Cells: XRNAX	27
The RNA Content of XRNAX Extracts	29
The Protein Content of XRNAX Extracts.....	30
<i>Interrogation of Protein Interfacing with RNA</i>	31
Purification and Detection of Nucleotide-Crosslinked Peptides from XRNAX Extracts	32
Database Search For Nucleotide-Crosslinked Peptides.....	32
Nucleotide-Crosslinked Peptides Pinpoint Protein-RNA Interfaces	34
<i>Discovery of Proteins Interacting with RNA</i>	36
SILAC-Controlled Enrichment for RNA-Crosslinked Protein from XRNAX Extracts	37
An Integrated Draft for the Human RNA-Binding Proteome From Three Cell Lines.....	39
A Novel Group of RNA-Binding Proteins Interacting with Non-Polyadenylated RNA.....	40
<i>Quantification of Changes in the Interaction of Proteins with RNA</i>	43
Background and Dynamic Range for the Differential Quantification of RNA-Binding with XRNAX	43
Differential Quantification of Protein-RNA Interaction During Translational Arrest.....	44
Excursion: Arsenite Induces Massive and Specific Autophagy on a Timescale of Minutes.....	48
<i>Interrogation of RNA Interfacing with Protein</i>	49
XRNAX CLIPseq Reveals Interactions Sites of Individual Protein Species with the Transcriptome.....	50
PEPseq Reveals Transcriptome-Wide Occupancy of Protein on RNA.....	53
CHAPTER 2: COMPARING PROTEIN HALF-LIVES INSIDE AND OUTSIDE OF PROTEIN-RNA COMPLEXES	55
<i>Protein Half-Lives of the Total Proteome</i>	55
Determining Protein Half-Lives Using TMT-SILAC.....	56
Normalizing Protein Half-Life Data for Dividing Cells	57
<i>Protein Half-Lives on RNA</i>	62

Determining Protein Half-Lives Using XRNAX and TMT-SILAC.....	62
Comparing Protein Half-Lives on RNA to the Total Proteome.....	64
<i>Protein Half-Lives in Ribosomal Assemblies</i>.....	64
Determining Protein Half-Lives Using Polysome Profiling and TMT-SILAC.....	65
Comparing Half-Lives of Ribosomal Proteins in Polysome Fractions to the Total Proteome.....	67
Exchange of Ribosomal Proteins From Translation-Competent Ribosomes.....	70
Comparing Half-Lives of Non-Ribosomal Proteins in Polysome Fractions to the Total Proteome.....	76
DISCUSSION.....	79
XRNAX ALLOWS FOR A PLETHORA OF NOVEL APPLICATIONS IN RNA BIOLOGY.....	79
FUNCTIONS OF NON-CODING RNA ARISING FROM THE INTEGRATED HUMAN RNA-BINDING PROTEOME ...	80
DIFFERENTIAL QUANTIFICATION OF PROTEIN-RNA INTERACTIONS PROVIDES UNSEEN MECHANISTIC INSIGHT.....	81
STABILIZATION OF PROTEIN IN PROTEIN-RNA COMPLEXES IS CONTEXT-DEPENDENT.....	82
RIBOSOMAL ASSEMBLIES ARE MORE DYNAMIC THAN PREVIOUSLY POSTULATED.....	86
MATERIALS AND METHODS.....	90
METHODS FOR CHAPTER 1.....	90
<i>Resource Table</i>	90
<i>Experimental Model and Subject Details</i>	93
Mammalian Cell Culture and Stable Cell Lines.....	93
<i>Method Details</i>	93
Guanidinium Thiocyanate–Phenol–Chloroform (TRIZOL) Extraction.....	93
UV-Crosslinking of Cells.....	93
Protein-Crosslinked RNA Extraction (XRNAX).....	94
Comparison of RNA Extracts Using Agarose Gel-Electrophoresis.....	94
RNA Sequencing for Quantifying the Relative Composition of RNA Extracts.....	95
Proteomic Sample Preparation.....	95
High pH Reversed-Phase Fractionation of Proteomic Samples.....	96
HPLC-MS for the Analysis of Nucleotide-Crosslinked Peptides, Discovery of RNA-Binding Proteins or the Differential Quantification of RNA-Binding.....	96
Isolation of Nucleotide-Crosslinked Peptides from XRNAX Extracts.....	97
Expression and Purification of C7orf50 and DUF2373 from E. coli.....	97
In Vitro Validation of RNA-Binding by Native Agarose Gel Electrophoretic Mobility Shift Assay.....	98
In Vivo Validation for RNA-Binding by PNK Assay or IP-MS.....	98
SILAC-Controlled Discovery of RNA-Binding Proteins From XRNAX Extracts.....	99
Quantification of Nascent Protein Upon Arsenite Stress Using Azidohomoalanine-Labeling.....	100
Differential Quantification of RNA-Binding Upon Arsenite Stress.....	101
Total Proteome Analysis of Arsenite-Induced Protein Degradation.....	101
Ethynyl-Uridine Incorporation and Confocal Microscopy for Monitoring RNA Turnover Upon Arsenite Stress.....	101
Immunofluorescence Staining and Confocal Microscopy for Monitoring EXOSC2 Localization Upon Arsenite Stress.....	102
XRNAX CLIP-seq for the Detection of EXOSC2 RNA Targets.....	102
<i>Quantification and Statistical Analysis</i>	104
MS Database Search.....	104
Processing and Analysis of MS Data.....	104
Gene-Ontology (GO) Enrichment Analysis.....	105
Visualization of the Human 80S Ribosome.....	105
Processing and Analysis of RNA Sequencing Data.....	105
Processing and Analysis of XRNAX-CLIP-seq Data.....	106
Statistical Analysis and Data Visualization.....	106

ADDITIONAL METHODS FOR CHAPTER 2	107
<i>Polysome Profiling for Proteomics</i>	107
Purification of Ribosomal Complexes By Polysome Profiling.....	107
Modified Proteomic Sample Preparation For Polysome Profiling Fractions.....	107
<i>Protein Half-Life Measurements</i>	108
Determination of Protein Half-Lives By Pulsed SILAC	108
TMT-SILAC Strategy for Determination of Protein Half-Lives	109
Pulsed SILAC for Monitoring Exchange of Ribosomal Proteins.....	110
<i>Quantification and Statistical Analysis</i>	110
MS Database Search for TMT-SILAC Data	110
Processing and Analysis of TMT-SILAC Data.....	110
Sizing Protein-RNA Interfaces within the Human 80S Ribosome.....	113
REFERENCES.....	114
ACKNOWLEDGEMENTS.....	123
SUPPLEMENTARY FIGURES	125
SUPPLEMENTARY FIGURES FOR CHAPTER 1.....	125
SUPPLEMENTARY FIGURES FOR CHAPTER 2.....	137

INTRODUCTION

PROTEIN-RNA INTERACTIONS ARE UBIQUITOUS

RNA is polymerized by a protein, i.e. RNA polymerase, and as soon as it comes into being wrapped into protein.¹ RNA-binding proteins are essential companions of messenger RNA (mRNA), which is spliced², capped², polyadenylated and exported³ from the nucleus to the cytoplasm of eukaryotic cells (for review see⁴). Within the cytoplasm more RNA-binding proteins responsible for nonsense-mediate decay check the quality of the message during the first round of translation, which is carried out by a massive assembly of more RNA-binding proteins – the ribosome (for review see⁵). Other RNA-binding proteins locate to the untranslated regions of mature mRNA and influence its stability⁶, rate of translation⁷ and even subcellular localization⁸. Messenger RNA only accounts for a low percentage of the total RNA in a cell, which in mammals is less than 5 % (per weight).⁹ Non-coding RNA (ncRNA) makes up for the most part of the transcriptome and especially ribosomal RNA (rRNA) contributes often more than 80 % (per weight). In ribosomes the intricate assembly of ribosomal proteins along with rRNA creates an enormous molecular machine, which itself binds to mRNA in order to polymerize new protein. Importantly, RNA fulfils here a structural role, however, is also responsible for the catalytic activity of the ribosome by contributing the peptidyl transferase centre.^{10,11} Another macromolecular assembly with enzymatic activity is the spliceosome, which assembles splicing factors around five highly abundant ncRNAs called small nuclear RNAs (snRNAs) U1-U5.¹² Non-coding RNAs are conventionally classified by their length, so that snRNAs fall into the category of small RNAs, i.e. smaller 200 nucleotides. The spliceosome excises intronic sequences from pre-mRNA and is therefore essential during eukaryotic gene expression. There are more small RNAs conferring cleavage, prominently small interfering RNAs (siRNAs), which recruit nucleases of the Argonaut family to RNA molecules with complementary sequence in the process of RNA interference.^{13,14} Cleavage only occurs when the siRNA exhibits perfect complementary towards its target.^{15,16} Small RNAs with imperfect complementarity can be endogenously expressed (so-called micro RNAs (miRNAs)) and regulate protein expression through a combination of translation repression and mRNA destabilization.¹⁷ All these actions are accomplished by the interplay of these RNAs with their cognate RNA-binding proteins. Other small RNAs are involved in the targeted modification of other classes of RNA. Small nucleolar RNAs (snoRNAs) recruit modifying enzymes, such as RNA methylases, to rRNA or snRNA and thereby confer RNA modifications to specific sequences motifs within these RNA species.^{18,19} RNA sequences can also be used as template for the production of DNA as in the case of telomerase. The protein component of telomerase is a reverse transcriptase that uses the RNA component of the complex for the extension of chromosome ends.²⁰ Telomerase RNA is more than 400 nucleotides long, thus, is considered a long ncRNA (lncRNA). RNA sequencing studies have revealed that beside long, protein-coding RNA eukaryotic cells produce thousands of long, non-coding RNA transcripts (for review see²¹). These lncRNAs have very little coding potential and proteomic²² or ribosome profiling studies²³ confirmed that most of them are not translated into protein. However, they

fulfil important and diverse functions by interacting with cognate protein (for review see²⁴). How many of the thousands of lncRNAs expressed in a cell are functional remains elusive. Nevertheless, fundamental and regulatory important functions have been described for some lncRNAs, which have made it very clear that even if only a small percentage of them is functional, they will still have a profound effect on the function of cells. This is especially the case because lncRNAs are primarily located in chromatin where they interact with and regulate the behaviour of chromatin modifying proteins. An illustrative example for such a lncRNA in mammals, which directly affects genomic regulation, is *Xist*. *Xist* is expressed from one of the two X chromosomes in a female cell, leading to transcriptional shutdown of the entire other X chromosome²⁵ – a process referred to as X chromosome inactivation²⁶. While the effect of *Xist* was observed early, it was not until very recently that its exact mode of action could be described. Importantly, mass spectrometry (MS) - based proteomics revealed that *Xist* recruits the chromatin modifier SPEN (also known as MINT or SHARP).^{27,28} Consequent experiments showed that SPEN recruitment to the target X chromosome leads to deacetylation and successive transcriptional shutdown. Another prominent example is the lncRNA HOTAIR, which was shown to shut down genes during development by recruiting proteins of the PRC2 complex to specific sites in the HOX locus.²⁹ These are only two early examples of profound genomic regulation mediated by lncRNAs and many more examples have been described over the last years (for reviews see^{21,24,30,31}). Already in 1975 it was recognized that purified chromatin contains twice as much RNA than DNA.³² While our growing understanding of the DNA component of chromatin has taught us about the genome, our growing understanding of the RNA component of chromatin seems to teach us more and more about its regulation. The sheer number of different RNA species leaves the number of processes potentially regulated by them unlimited. From the protein perspective this regulatory potential is not delimited. As will be discussed in the following, in human cells more than 1000 proteins have been found to interact with RNA so far (for review see³³). Many of these proteins are known to be involved in processes unrelated to classic RNA biology, which suggests functions of protein-RNA interactions far beyond the canonical understanding. An emerging concept in this regard is the formation of membrane-less organelles by liquid-liquid phase separation (for review see³⁴). Protein-RNA interactions have emerged as central in this context because prime examples for phase-separated organelles come from macroscopic protein RNA assemblies, such as nucleoli³⁵, p-bodies³⁶ or stress granules³⁷. Liquid-liquid phase separation is a very powerful concept because it is able to explain the often-observed assembly of specific complexes, which have a specific protein composition, while the individual proteins in the complex do not specifically bind to one another. Formation of these organelles appear to rely on the presence of specific sequence features called intrinsically disordered regions (IDRs) in proteins inducing phase separation.³⁸ The same regions have been shown to interact with RNA, however, proteins carrying IDRs often also carry classical, globular RNA-binding domains.³⁹ *In vitro* proteins carrying IDRs and even the IDRs themselves can phase-separate independently of RNA, however, in the presence of RNA they phase-separate at much lower concentrations.^{37,40} Recently, a novel granule type has been described, which offered an intriguing example on how phase-separation can organize

molecular assembly from specific protein and RNA components. TIS11B forms phase-separated granules in human cells, which sequester mRNAs encoding membrane proteins, thereby mediating their localization.⁴¹ The future will show in how far other new or known assemblies are phase-separated. What is clear already, however, is that protein-RNA interactions are central to the process.³⁴

In summary, these examples demonstrate the vital importance of protein-RNA interactions in most aspects of human cell biology. While this thesis project is primarily focused on human cells, we mention here that of course protein-RNA interactions are fundamental to other organisms as well. Only to mention one example with importance to human health, many viruses and especially RNA viruses rely on interactions with proteins within themselves and within the host cells they infect. All this makes clear that in order to understand any organism one has to understand protein-RNA interactions.

CURRENT METHODOLOGY FOR THE INTERROGATION OF PROTEIN-RNA INTERACTIONS

In this thesis project new methodology is explored, which serves the extraction and analysis of protein interacting with RNA or RNA interacting with protein, respectively. To get a better understanding of what the technological challenges for any such endeavours are, the following summary enumerates methodology that has been used and its individual contribution so far. Conceptually, protein-RNA interactions can be approached from a protein-centric or RNA-centric perspective, which means proteomic technologies such as MS or RNA-sequencing are used (for review see⁴²). Unified approaches, which sequence both protein and RNA at the same time do not currently exist. However, there is evidence that MS-based proteomics hold the potential for intersecting these two disciplines. Details will be discussed when we talk about the common shortcomings of all presently available methods.

RNA SEQUENCING-BASED METHODOLOGY

Protein-centric ways for investigating protein-RNA interactions try to enrich one specific protein in order to find out what its RNA interaction partner is. Most protein-centric methodologies for the investigation of protein-RNA interactions today work *in vivo* and typically in UV-crosslinked cultured monolayers of cells. However, one powerful *in vitro* approach for determining binding features of a purified RNA-binding protein is RNA bind-n-seq.⁴³ The method developed from systematic evolution of ligands by exponential enrichment (SELEX)^{44,45} and the realisation that massively-parallel sequencing offers a resolution that makes much of SELEX obsolete. During SELEX a target, for example a purified RNA-binding protein, is immobilized and subjected to a random pool of RNA sequences. Non-specific RNA is washed away and bound RNA is reversed transcribed, PCR amplified, *in vitro* transcribed into RNA and again subjected to the same immobilized protein. During many rounds of selection this evolutionary process leaves only sequences carrying the optimal binding motif for the specific protein behind (if an optimal binding motif exists and the protein does not simply bind any RNA). SELEX was developed at a time when massively-parallel sequencing did

not exist, so that the pool of RNA sequences was eventually cloned into *E.coli* and single colonies picked in order to determine the most common sequences by Sanger sequencing. One flow cell in a standard Illumina sequencing run today has the ability to detect more than one billion sequences in one experiment. RNA bind-n-seq harnesses this ability by reducing the enrichment steps to a single one, thereby allowing for the detection of secondary motifs and even semi-quantitative statements on the binding strength of some sequences over others. Therefore, RNA bind-n-seq offers a clean and parameter-reduced environment for the *in vitro* interrogation of fundamental RNA-binding features of a protein (and potentially other substances).

In vivo methods for the detection of RNA sequences bound by one particular protein go back to the observation that antibodies from patients with *lupus erythematosus* are able to precipitate RNA-binding protein along with six different species of small RNAs.⁴⁶ Biologically, this observation became pivotal in the years afterwards when those small RNAs were termed snRNAs and the protein they assemble with the spliceosome. However, on the technological side this led to the development of RNA immunoprecipitation or RIP, where one protein is enriched from a cellular lysate using a specific antibody, in order to identify the RNA species that come along with it. RNA is sticky and RNA-binding proteins tend to bind to other RNA-binding proteins⁴⁷, so that the conclusiveness of RIP was in most cases limited. A denaturing approach was required, however, obstructed by the fact that RNA-binding proteins lose their RNA cargo as soon as they are denatured. Attempts that used formaldehyde crosslinking in order to stabilize protein-RNA interactions did for the most part not improve the purity of RIP because formaldehyde also crosslinks other proteins (or DNA) to the bait. This problem was solved by the introduction of UV-crosslinking to RIP, which heralded the era of crosslinking and immunoprecipitation or CLIP.⁴⁸

As UV-crosslinking is vital not only to CLIP but large parts of the technological developments pursued in this thesis project, we take here an excursion into this field. UV-crosslinking is widely used in RNA-biology, however, many of its fundamental features remain poorly understood. That UV-light crosslinks protein to nucleic acids was first suspected in 1962 when the germicidal activity of UV was investigated.⁴⁹ Initially, the assumption was that thymidine dimers were the lethal mutation to the *E.coli* genome introduced by UV, however, a much better correlation was found with crosslinked protein.⁵⁰ In the coming years the photo-adducts of the crosslinking process towards DNA and RNA bases was interrogated. The first photo-adduct to be identified was a uracil-cysteine crosslink (5-5-cysteine-6-hydroxouracil) purified from a UV-crosslinked solution of uracil with cysteine⁴⁹ and later DNA or RNA⁵¹ with cysteine, respectively. Importantly, for a cysteine-containing peptide crosslinked to a uridine, this would mean an adduct mass identical to the mass of uridine. In experiments with cysteine and polyuridine oligomers or yeast total RNA, both crosslinked equally well and by far better than oligomers composed of any other nucleotide. Later a more systematic approach showed that all 20 naturally occurring amino acids could be UV-crosslinked to polyuridine oligonucleotides, however, some more efficient than others.⁵² Already in these early studies it was observed that purine bases do not crosslink efficiently and the main crosslinking bases are pyrimidines. This was later confirmed in studies using UV lasers⁵³. Until now a

comprehensive catalogue of crosslinking adducts does not exist. However, the photochemical process underlying UV-crosslinking of protein and nucleic acids has been further studied, so that a number of general statements can be made (for review see⁵⁴): UV light excites nucleotide bases, which convert over a singlet or triplet state to cation radicals. Reducing agents such as mercaptoethanol are able to suppress crosslinking so that the actual crosslinking reaction is expected to occur via the radical state of the base, however, the exact mechanism remains unknown. Interestingly, recently it was shown that the reducing agent dithiothreitol can bridge uridine crosslinks to cysteine-containing peptides.⁵⁵ Other recent studies on uridine indicate that two reaction mechanisms might exist, one fast and direct reaction of the radical and one slow via a photohydrate.⁵⁶ *In vitro* crosslinking yields, i.e. how much of the protein in an experiment ends up crosslinked to RNA, range between 5-20 %. This means UV-crosslinking in contrast to formaldehyde crosslinking is sparse, producing RNA that is mostly free of protein and only sporadically interrupted by protein crosslinks. Damage to nucleic acids at low intensity UV irradiation (UV lamp) are primarily uridine or thymine dimers, respectively, at high intensity irradiation (laser) the induction of single-strand breaks. Photodamage to proteins seems negligible at the wavelengths usually used for crosslinking (>240 nm). Crosslinks are stable to heat and alkaline conditions, however, some might be acid labile. Before UV-crosslinking of cultured cells, they are transferred onto ice in order to prevent any immediate cellular response against UV irradiation, which typically takes less than one minute to reach the conventional crosslinking energy with the conventional equipment. This is another advantage over formaldehyde crosslinking, which is inefficient if not performed at ambient temperatures and takes several minutes to occur.

UV-crosslinking and CLIP became very popular and widely used, so that today scores of variations to the CLIP protocol exist (for review see⁵⁷). Some of these variations are redundant or rarely used, so that we only discuss the most important additions to the protocol since its inception. Notably, there are reports for formaldehyde-based CLIP protocols⁵⁸, however, they are not commonly used because of the drawbacks described above. The initial CLIP protocol⁴⁸ crosslinked cells with UV, lysed them, fragmented RNA by partial RNase digestion, immunoprecipitated the RNA-binding protein NOVA, radiolabelled the RNA fragments that came along with NOVA, resolved the denatured protein-RNA complexes on an SDS-PAGE and blotted them onto nitrocellulose. Subsequently, the radioactive band representing the molecular weight of NOVA could be excised from the membrane, thereby excluding contaminations from any other RNA-binding protein or free RNA. RNA fragments were cloned into *E.coli* and individual clones selected for Sanger sequencing. The last step yielded several hundred sequences, which was an obvious point for improvement upon the introduction of massively-parallel sequencing. This led to the introduction of HITS-CLIP (high-throughput sequencing CLIP), which pushed the recovered sequences into the hundred thousands.⁵⁹ HITS-CLIP was very successful, however, some proteins could not be UV-crosslinked efficiently, although it was clear that they must be specific RNA-binders. In order to improve crosslinking, the photoactivatable nucleotide 4-thiouridine (4SU) was identified, which is readily taken up by cells from the culture media and incorporated into their transcriptome. 4SU crosslinks at a wavelength of 365 nm, whereas conventional crosslinking is performed at 254 nm. PAR-CLIP

(photoactivatable-nucleotide-enhanced CLIP) was able to retrieve sequences for some key proteins binding to small RNAs, which were under heavy investigation in the context of RNA interference, however, previously impossible to crosslink.⁶⁰ PAR-CLIP brought along another unanticipated advantage, which was a sequence conversion from uridine to cytidine at positions crosslinked to protein. This so-called T to C conversion, presumably introduced by an error of reverse transcriptase at these positions, could be used to locate the protein-RNA interaction site with nucleotide precision. In the case of 254 nm crosslinking reverse transcriptase most often dissociates upon encounter of the crosslinking site. This led to the development of individual nucleotide CLIP (iCLIP), which mapped these dissociations events using a special protocol for the preparation of the sequencing library.¹ While iCLIP claims to have better resolution than PAR-CLIP because it potentially resolves crosslinking at every nucleotide and not only at uridines, the observation that 254 nm UV primarily crosslinks uridine qualifies this claim. Effectively, both protocols have been very successful and widely used. Further additions to CLIP after these developments primarily concerned reproducibility and feasibility. CLIP protocols up to this point involved radioactive labelling and cumbersome gel-based, sequencing library preparations. For the encyclopaedia of DNA elements (ENCODE) a CLIP variation was developed, which tried to strip down the process to the point where CLIP experiments could be easily scaled in number while making them more reproducible. Enhanced CLIP (eCLIP) was developed, which not only simplified the procedure but also introduced an input control, so that better significance testing could be performed on potential binding sites.⁶¹ So far ENCODE holds eCLIP data for over 300 RNA-binding proteins, which is an invaluable resource and important for understanding the big picture of RNA regulation in a cell (www.encodeproject.org).

Concerning other crosslinking technologies there has been one recent development using the APEX system.⁶² APEX2 is an enzyme that can be genetically encoded and fused to any other desired protein. Upon addition of biotin-phenol and hydrogen peroxide the fusion protein will produce biotin radicals, which will diffuse and react to protein within 10 nm range before they decay. Fused to cellular localization signals the APEX system can be used to retrieve proteins with specific subcellular localization.⁶³ In combination with formaldehyde crosslinking, this can retrieve RNA as well and thereby allows for determining the subcellular composition of transcriptomes.⁶⁴ Alternatively, the APEX system has been used along with 4SU and 365 nm UV-crosslinking in a very recent approach called proximity-CLIP. Proximity-CLIP not only allows for the retrieval of the subcellular transcriptome, but also to give the occupancy of all RNA-binding proteins on the retrieved transcripts. In the latter variation of the protocol, biotinylated proteins are UV-crosslinked to RNA and partially RNase digested. Subsequently, only the RNase-protected fragment is sequenced. This is markedly different to a classical CLIP experiment, which immunoprecipitates one individual RNA-binding protein and tries to find its location on the transcriptome. Here, the entirety of RNA-binding proteins on the transcripts is retrieved and their combined binding sites collapsed to a protein occupancy profile. Before proximity-CLIP (which is a misleading name because it does not entail immunoprecipitation), only one other method had tried to determine the global occupancy of RNA-binding proteins on the transcriptome. During protein occupancy profiling, cells are subjected to 4SU and UV-

crosslinked at 365 nm wavelength.^{65,66} Their polyadenylated transcriptome is precipitated using poly(dT)-beads along with crosslinked protein. The crosslinked protein-RNA complexes are released by partial RNase digestion leaving RNase-protected fragments intact, which are then ammonium acetate precipitated, resolved by SDS-PAGE and blotted onto nitrocellulose. Precipitation and blotting should only retain protein-bound RNA thereby eliminating fragments, which were initially not occupied by protein. After sequencing the number of T to C conversions at a specific position are used to quantify their occupancy by protein. In how far T to C conversions are a quantitative measure of protein occupancy was never shown and protein occupancy profiling was never widely used. Proximity-CLIP is not limited to polyadenylated RNA and uses read counts for the quantification of protein occupancy, which is an established method for occupancy in the chromatin field. The biological insight presented so far with the method is limited, so that future studies will need to show its usefulness. Nevertheless, a method tracking the protein occupancy on the global transcriptome in a quantitative way is conceptually very appealing, especially in the context of perturbation experiments resolved in a timecourse.

MASS SPECTROMETRY-BASED METHODOLOGY

RNA-centric methods for the identification of RNA-binding proteins enrich RNA in order to identify the proteins that come along with it. Methodology here can be subdivided in a similar fashion as the sequencing methodology above. Methods exist for the identification of proteins binding to one particular RNA species and for the identification of proteins binding to the entire transcriptome, or at least parts of it, as we will see in the following. For the unbiased detection of protein interactors today MS is used, however, for validation purposes or in the case of very low abundant proteins antibody detection might be favoured. The sensitivity of MS has improved greatly over the years so that today femtomolar amounts of protein can be quantified on a standard basis. Still, MS in contrast to sequencing cannot rely on sample amplification by PCR, so that the usual limitation of MS is sample amounts. This is especially true in the case of an experiment that tries to identify proteins specifically binding to one particular species of RNA. The dynamic range within the transcriptome is tremendous, with some few RNA species, such as rRNA, being very abundant and dominant. This has important consequences for the enrichment that is necessary to detect any protein binding specifically to RNA. One HeLa cell is estimated to have several million ribosomes, for simplicity say one million. Consider a case in which one would like to find proteins binding an RNA species with 1000 copies per cell. In order to detect proteins, which do not exclusively bind this particular RNA, more than 1000 fold enrichment over the background will be required. RNA-binding proteins are rarely specific and often bind degenerate motifs⁶⁷, so that enrichment in combination with low amounts are the main limitations in such studies. We will discuss how current methods overcome this limitation, but first take a look at sample preparation strategies for the enrichment of one particular RNA (for review see⁶⁸). The first *in vitro* report for the purification of proteins from one particular RNA species again comes from the field of splicing, where HeLa lysates were incubated with an *in vitro* transcribed RNA transcript, which had been labelled with biotin-UTP.⁶⁹ Here, before streptavidin enrichment lysates had to be

fractionated by glycerol gradient centrifugation. The spliceosome is a fairly stable and abundant complex, which binds to its target sequence with high specificity, however, a double purification was required indicating how delicate such experiments are. In the years after this, *in vivo* methods were developed that introduced an RNA sequence tag, which could be specifically recognized by a protein on a solid support and thereby precipitated. Most commonly used are the MS2⁷⁰ or PP7⁷¹ system, both encoding for stem loops recognized by bacteriophage proteins. These proteins, carrying an affinity tag for later enrichment, can be expressed along with the RNA, which is tagged with one or several repeats of the stem-loop. Similar systems with RNA aptamers exist, where only the tagged RNA needs to be expressed, which itself binds to streptavidin⁷² or streptomycin⁷³ immobilized on beads. Either one of the approaches requires ectopic expression of the RNA of interest, which needs to be altered in its sequence in order to carry the affinity tag. In order to target endogenous transcripts there are currently two strategies available, the first of which is still very much in its infancy, however very promising because of its versatility. The endonuclease Cas13a from *Leptotrichia wadei* can be engineered to lose its nuclease activity and targeted RNA transcripts using conventional CRISPR guide RNAs. A study shows that targeting is highly specific and can be used to enrich individual RNA sequences.⁷⁴ However, so far the system has not been used in combination with crosslinking and for the identification of RNA-binding proteins, so that future studies will have to demonstrate its usefulness. Established and successfully applied methods for the identification of proteins binding to individual, endogenous RNA species use biotinylated oligonucleotide probes, which are annealed after lysis of the cells and captured on streptavidin beads. In 2015 three independent studies characterized the *Xist* interactome^{27,28,75} with oligonucleotide probes and MS, two of which came to similar conclusions about the mechanism by which *Xist* leads to X chromosome inactivation^{27,28}. Of the latter two, CHIRP-MS uses formaldehyde crosslinking, whereas RAP-MS uses UV-crosslinking. Conceptually, both methods use a similar approach for the identification of specific RNA-binders among RNA-binders common to other RNA species in a cell. Both methods apply their enrichment protocol to several highly expressed non-coding RNAs and compare their interactome to the one identified for *Xist*. Thereby, they overcome the above-mentioned enrichment problem by finding the common background of a couple of RNAs, in order to see specific enrichment of proteins for the individual candidate RNA. From an MS perspective, RAP-MS (RNA antisense purification coupled with MS) is the more reasonable approach, which uses SILAC-based quantification for this purpose, whereas CHIRP-MS (comprehensive identification of RNA-binding proteins by MS) works label-free and uses a seemingly arbitrary ranking system of peptide counts. CHIRP-MS has not found convincing application after its initial report, whereas more recently RAP-MS was successfully used in combination with iTRAC for mechanistic studies on the lncRNA NORAD⁷⁶.

To screen for proteins binding not only to one single transcript but the entire transcriptome, one method called interactome capture has dominated the field over the last years (for review see⁷⁷). Interactome capture goes back to a study in 1979, which showed that also mRNA crosslinks to protein. In a follow-up paper polyribosomes were purified in a sucrose gradient and UV irradiated for crosslinking. Subsequently, polyadenylated RNA was annealed to

poly(dT) probes and non-crosslinked protein washed away. Crosslinked proteins were released with RNase and an SDS-PAGE identified six of them. After the protocol had been picked up again in 2012, two groups independently identified the specific interaction of approximately 800 proteins using MS.^{65,78} Today interactome capture uses poly(T) DNA or LNA probes, which are immobilized on beads in order to capture polyadenylated RNA from lysates of UV-crosslinked cells. After stringent washing, proteins are released with RNase and subjected to MS. A non-crosslinked control is either analysed in separate^{78,79} or under the use of SILAC combined with the crosslinked sample⁶⁵. Obviously, the method does not capture protein-RNA interactions transcriptome-wide but is restricted to polyadenylated RNA. Interactome capture has been very successful in cataloguing RNA-binding proteins, as will be discussed below. However, so far only one study used the method successfully for the differential quantification of RNA-binding between conditions.⁸⁰

Several recent approaches try to overcome interactome capture's limitation to polyadenylated RNA. One method using 4SU and UV-crosslinking is identification of RNA-binding regions (RBR-ID). RBR-ID is based on the observation that MS data from UV-crosslinked samples has oftentimes significantly reduced intensities for peptides in known RNA-binding domains. This makes sense because crosslinked peptides have a different mass and are unfortunately not easily detected as we will see in the following. By inverting this observation, RBR-ID compares peptide intensities between crosslinked and non-crosslinked samples in order to predict RNA-binding. Unfortunately, differences are oftentimes only minute – an expected consequence of the low efficiency of UV-crosslinking – so that the initial study had to use very lax statistical cut-offs and did not correct for multiple testing. This would further indicate that RBR-ID does not have the potential for making quantitative statements on RNA-binding between conditions. A direct method for the identification of proteins interacting with the entire transcriptome is capture of the newly transcribed RNA interactome using click chemistry (RICK). RICK pulse-labels the transcriptome with the clickable ethynyl uridine (EU) and then uses 254 nm UV light to crosslink RNA-binding proteins to it. Subsequently, EU-labelled RNA can be captured using standard click-chemistry, thereby co-purifying crosslinked protein, which is identified by MS. If RICK could be used for differential quantification of RNA-binding is not clear from the data presented in the initial study, because the degree of enrichment is never clearly specified. However, the method works with relatively small amounts of cells (30 million) even at very short labelling pulses, so that further development might be promising. For the discovery of RNA-binding proteins another computational method was presented recently, which capitalizes on the observation that immunoprecipitation followed by MS (IP-MS) of RNA-binding proteins enriches primarily other RNA-binding proteins, even after RNase digestion.⁴⁷ This suggests that RNA-binding proteins primarily bind other RNA-binding proteins. Support vector machine obtained from neighbourhood associated RBPs (SONAR) uses this observation for machine learning on in BioPlex protein-protein interaction datasets, where it learns from known RNA-binding proteins how they are connected in order to identify new RNA-binding proteins.

We also want to mention orthogonal organic phase separation (OOPS). The method was developed back-to-back with the methodology explored by us in this doctoral thesis, which

we termed protein X-linked RNA eXtraction (XRNAX). OOPS⁸¹ and XRNAX⁸² were published in the beginning of 2019 as new methods for the identification and quantification of protein-RNA interactions in UV-crosslinked cells. Both methods are based on TRIZOL extraction, a method further explored below, so that we refer to the Results section of this thesis for further introduction to OOPS. However, we note here that both OOPS and XRNAX were able to identify similar sets of proteins as RNA-binding, which are so far the most comprehensive representations of the human subproteome binding to the transcriptome independent of polyadenylation.

The last and so far least developed aspect in the study of RNA-binding proteins using UV-crosslinking is the protein-RNA crosslinking site. This has a number of reasons, one of which has been mentioned before and concerns the chemistry of UV-crosslinking. Crosslinking adducts between ribonucleotides and amino acids are for the most part unknown, so that in an MS experiment it is unclear what masses to look for eventually. This is complicated by the multiplicity of the problem, because in principle all four bases might crosslink to all 20 amino acids. For an MS experiment that means at least four variable modifications on 20 amino acids, which overchallenges the capabilities of conventional search algorithms by far. From the perspective of crosslinking chemistry, getting a systematic understanding of the crosslinking adducts is not easy because UV-crosslinking yields are low. Consequently, even for very plain *in vitro* experiments of one pure protein incubated with pure RNA, it is not feasible to produce UV-crosslinked protein to the scale necessary for NMR. From the perspective of MS, nucleotide-crosslinked peptides are not well-behaved under the standard methodology used for normal peptides and no systematic optimization of the MS methodology has been possible so far because nucleotide-crosslinked peptides are not easily purified. On a systems level the problem remains the same, i.e. currently there is no universal approach available for the purification of protein UV-crosslinked to RNA, which could be used for benchmarking studies. Most of our knowledge on nucleotide-crosslinked peptides comes from one relatively recent study, which used interactome capture for their purification.⁸³ Here TiO₂ enrichment was used after interactome capture, when proteins crosslinked to RNA had been released from the poly(dT) beads through RNase. TiO₂ is typically used for phosphoproteomics, however, is also able to enrich nucleotide-crosslinked peptides through the phosphate on the nucleotide. Specialized software had to be developed to analyse the MS data, which identified 60 nucleotide-crosslinked peptides from 35 highly abundant human proteins. This rather sobering result demonstrates what the combined effects of the above-mentioned problems amount to. Nevertheless, it demonstrates that MS is in principle capable of sequencing peptides crosslinked to one or several nucleotides, which in the future might lead to simultaneous sequencing of both protein and RNA in UV-crosslinked peptide-oligonucleotide hybrids. Interactome capture as well as TiO₂-enrichment do not scale well, because they use fairly expensive beads, so that a systematic optimization for the purification and detection of nucleotide-crosslinked peptides using this strategy is hardly feasible. This is why some methods have tried to find ways around their purification. As mentioned above, their elusive behaviour has even prompted methods such as RBR-ID to use their absence in MS data for calling RNA-binding proteins.⁸⁴ Another such method is RBDmap, which extends interactome

capture to purify protein crosslinked to polyadenylated RNA, which is digested with LysC or ArgC, respectively.³⁹ This leaves behind a crosslinked peptide on the captured RNA, which, depending on the enzyme used, might still harbour another arginine or lysine. This peptide can be released with trypsin and is identified by MS in order to map its crosslinked neighbour, which remained on the RNA. Thereby, RBDmap does not even try to identify the crosslinked peptide, but uses its neighbours to map protein-RNA interaction sites. Another interesting *in vitro* study uses heavy isotope labelling of one particular RNA species, which is incubated and UV-crosslinked with one particular protein, in order to create a unique mass label.⁷⁷ Consequently, peptide doublets separated by a defined mass indicate in the MS RNA-crosslinked peptides, which can be specifically selected for fragmentation. If translated into an *in vivo* study, this approach could teach us a lot about the behaviour of nucleotide-crosslinked peptides during MS.

This summary of the presently available proteomic methodology for the interrogation of protein-RNA interactions illustrates chances but also shortcomings, which primarily lie on the side of sample preparation. We will discuss this observation further below, when we talk about the aims of this thesis project.

CURRENT UNDERSTANDING OF RNA-BINDING PROTEOMES

MANY HUMAN PROTEINS INTERACT WITH POLYADENYLATED RNA

RNA-binding domains are among the most frequently occurring domains in the human proteome. In fact, the RNA recognition motif (RRM) is one of the most frequently occurring domains among all eukaryotes.⁸⁵ Therefore, many proteins could be classified as RNA-binding from their sequence even before MS studies began to screen for them.⁸⁶ Specifically, about 700 proteins in the human genome can be identified as RNA interactors from their sequence features (for review see^{86,87}). However, interactome capture has added hundreds of proteins, which based on their sequence features had not been expected to bind RNA (for review see³³). Each of the initial interactome capture studies in HEK293⁶⁵ and HeLa⁷⁸ cells returned about 800 proteins, 300 of which had not been known to interact with RNA. Within this group of proteins no clear enrichment for a particular cellular function became apparent. This was later confirmed in a survey study that compared proteins identified through interactome capture in yeast and human cells, which found the function of this group enigmatic.⁸⁸ Of note, this study and previous ones implicitly suggested that metabolic enzymes were enriched within the experimentally determined human RNA interactomes. In fact, this is not the case, so that metabolic enzymes actually occur less frequently in the HeLa RNA-binding proteome⁷⁸ than their occurrence in a deep HeLa total proteome⁸⁹ would predict. Nevertheless, on the domain level two important observations were made. The first was an enrichment of DNA-binding domains also within the RNA interactome, which was especially true for proteins of the zinc-finger type.⁶⁵ This was further highlighted by protein-interaction network analysis, which found several DNA-related processes enriched for mutually interacting RNA-binding proteins.⁶⁵ Conceptually, this was an important observation because it indicated that the interaction of some proteins with DNA could be modulated, guided or disrupted by RNA. As we will see in the following, a number of these functions were later shown to occur and the

concept keeps on gaining traction. Another observation was an enrichment for IDRs, which occurred much more frequently in the RNA-binding proteomes than in the total proteome. Notably, this was the case for the group of proteins carrying canonical RNA-binding domains, but also for the ones that did not. Later RBDmap and RBR-ID showed systematically that IDRs not only occur more frequently in RNA-binding proteins but also interact with RNA, thereby constituting a new kind of non-globular RNA-binding domain.

A census of RNA-binding proteins, which curated RNA-binding proteins from a combination of *in silico* and experimental discovery approaches, including the overlap between the two initial interactome capture studies, came to an estimated number of 1542 proteins for the combined human RNA-binding proteome.⁸⁷ A current meta study, which collected human interactome capture datasets to integrate them into a superset of mRNA-binding proteomes, showed 55 % identity with this list.³³ Both, the curated census and the interactome capture superset, had more than 600 proteins unique to themselves. On the one hand, this illustrates the ability of experimental approaches, such as interactome capture, to discover protein-RNA interactions, which are not easily predicted by computation. On the other hand, it shows that interactome capture is limited in some way, so that it cannot procure the re-discovery of a fair number of computationally predicted RNA-binding proteins. As polyadenylated RNA makes up for less than 5 % of a mammalian cell⁹, it is likely that these RNA-binding proteins might interact with the other 95 %.

MANY EMERGING REGULATORY FUNCTIONS ARE BASED ON INTERACTIONS WITH NON-POLYADENYLATED RNA

As mentioned earlier, many lncRNAs are emerging with important functions in genome regulation (for review see²⁴). While many of these RNAs can be found in a polyadenylated form, a recent quantitative study showed that for the most lncRNA species polyadenylation is not efficient, so that their dominant form is present non-polyadenylated.⁹⁰ Furthermore, there are well-documented cases of very highly expressed lncRNAs, such as MALAT1, which are never polyadenylated but stabilized through 3' secondary structures.⁹¹ This indicates that in order to monitor protein interactions with these RNAs a more comprehensive approach than interactome capture is required, because interactome capture depends on polyadenylation. In general a number of non-coding, non-polyadenylated, highly expressed RNAs, such as 7SK RNA, are known to be involved in the canonical pathways controlling transcription (for review see⁹²). However, more and more reports show how RNA in general can influence the behaviour of chromatin modifying proteins in regulatory ways, too. For example, for the PRC2 complex it was shown that its RNA-binding antagonizes its chromatin binding, so that actively transcribed genes repel the complex, thereby repelling transcriptional shutdown by it.⁹³ Nascent transcripts only get polyadenylated at the very end of transcription, when the polyadenylation signal emerges, so that for the most part nascent transcripts are non-polyadenylated and cannot be interrogated through interactome capture. Another process important in this context is RNA-templated DNA repair.⁹⁴ For DNA double-strand breaks (DSBs) in transcribed genes it was shown that the non-homologous end joining pathway can overcome its error-prone function when nascent RNA is employed as a template,

which results in error-free DNA repair.⁹⁵ These are examples from the field of genome regulation, however, there are obviously many other important interactions involving non-polyadenylated RNA, especially in the context of splicing, translation or RNA interference, some of which have been mentioned above.

MANY ORGANISMS MAKE DIFFERENT USE OF POLYADENYLATION

This thesis project is about the interrogation of protein-RNA interaction in human cells. However, protein-RNA interactions are fundamental to other organisms as well, so that new methodology developed to understand the human RNA-binding proteome could also benefit the understanding of other RNA-binding proteomes. This is especially true for the RNA-binding proteomes of prokaryotes. Among eukaryotes, interactome capture has led to a relatively thorough description of the mRNA-binding proteome with studies in human, mouse, fish, fly, yeast, plants and several protists (for review see³³). Prokaryotic RNA-binding proteomes are understudied in this regard, because their transcriptome is not polyadenylated in the way eukaryotic transcriptomes are, which makes them inaccessible to interactome capture. While poly(A) tails in eukaryotes stabilize mRNA transcripts, in prokaryotes polyadenylation is believed to occur mostly on RNA breakdown products, recruiting nucleases and accelerating their degradation (for review see⁹⁶). However, their share of the total transcriptome can amount to significant proportions from 2 % in *E.coli* to 10 % in *B.subtilis*. The length of poly(A) tails is with 1-50 nt much smaller than in eukaryotes, where the average length lies around 200 nt. While longer tails do occur in prokaryotes the bulk of transcripts carries a tail smaller 10 nt. Interestingly, polyadenylation is not limited to mRNA but also occurs on rRNA, especially 23S rRNA, which in *E.coli* can carry polyadenylation on 10 % of all transcripts. Notably, similar observations have been made for organellar RNA in eukaryotes. Human mitochondrial mRNA is polyadenylated, however, through a pathway reminiscent of bacterial polyadenylation. Yeast mitochondrial mRNA only carries a very short poly(A) tail of eight nucleotides.

That interactome capture has not been successfully used in any bacterial species so far witnesses the fact that prokaryotic polyadenylation is too short for efficient enrichment. Thus, methodological developments beyond interactome capture could provide new insights into prokaryotic protein-RNA interactions, which so far remain unexplored. Considering that many extremely selective and potent antibiotics interfere with protein-RNA interactions in the bacterial ribosome (for review see⁹⁷), this should be a worthwhile endeavour.

AIM OF THIS THESIS PROJECT AND BIOLOGICAL QUESTIONS

Major efforts over the last years have been assigned to cataloguing RNA-binding proteins in various organisms and under various conditions (for review see³³). While these catalogues have given much insight on RNA-binding as a molecular function, they have given little insight on how RNA-binding influences the behaviour of proteins. The challenge ahead lies in the differential quantification of protein binding to RNA between conditions.

In order to find out how RNA-binding proteins regulate particular processes it would be important to know how dynamic their interaction with RNA is in the first place. A number of well-documented processes exist, leading to significant alterations in the subcellular

distribution of cellular RNA and RNA-binding proteins, which could be used to investigate this question. For example, arsenite stress is a well-documented inducer of translational arrest and macroscopic protein-RNA complexes called stress granules (for review see⁸¹). Therefore, quantifying protein-RNA interactions during arsenite stress should answer questions about how regulated and conditional RNA-binding is. Do RNA-binding proteins simply sit on RNA or do they increase or decrease binding depending on the cellular conditions?

As outlined above, many essential protein-RNA complexes exist, some of which are assembled from a large number of protein components. Studies on protein half-lives have suggested that proteins in complexes are more stable than in their free form.⁹⁸ A method that detects proteins on RNA quantitatively could also ask questions about the half-lives of these proteins. For example, are proteins in protein-RNA complexes more stable than in their free form?

This thesis project started out from the realisation that current progress in the field of protein-RNA interactions and RNA biology in general is hampered by the lack of a universal method for the purification of protein-crosslinked RNA. From a proteomics perspective this means that currently there is no method available, which allows for the quantification of protein-RNA interactions between conditions transcriptome-wide. As outlined above, available methods are either limited to a small percentage of the transcriptome or do not allow for differential quantification or both. Furthermore, none of those methods allows for the purification of protein UV-crosslinked to RNA in a way that is scalable. This is an important shortcoming because our knowledge about protein-RNA crosslinks is still in its infancy, mainly due to their poor accessibility. Consequently, an inexpensive and scalable way of purifying protein crosslinked to RNA would allow for much better understanding of the substance itself, apart from enabling new ways for quantitative proteomics.

Aim of this thesis project was to develop a universal extraction method for protein-crosslinked RNA from UV-crosslinked human cells. This was realized in a method termed XRNAX, which is described in the first chapter of this thesis. In this first chapter we apply XRNAX to a number of proteomic and transcriptomic applications and answer the question how dynamic binding of the RNA-binding proteome is during arsenite stress. Apart from that, we discover an exceptionally fast and severe form of ribophagy induced by arsenite.

In the second chapter of this thesis we turn to the question of protein stability in protein-RNA complexes. We apply XRNAX in order to derive protein half-lives on RNA and validate our findings in polysome profiling experiments. Apart from confirming that proteins bound to RNA are generally more stable, we find evidence for the exchange of ribosomal proteins from mature and translation competent ribosomes.

RESULTS

COPYRIGHT DISCLAIMER

*The work presented in this thesis was conducted by me, Jakob Trendel, supervised by my advisor Jeroen Krijgsveld. The pronoun “we” used in this thesis denotes to me, my supervisor and if applicable collaborators. Work performed by collaborators is specifically highlighted in the text. The first chapter of this thesis, “XRNAX As Platform for Interrogating Protein-RNA Interactions”, summarizes work that has been published in Trendel et al., *The Human RNA-Binding Proteome And Its Dynamics During Translational Arrest*, *Cell*, 2018⁸². Figures for this first part have been adapted from this publication (i.e. Figures 1-6 and S1-S6), which will be openly accessible after expiry of the embargo period on this thesis. Additionally, a preprint of the publication is available on bioRxiv under the title Trendel et al., *The Human RNA-Binding Proteome And Its Dynamics During Arsenite-Induced Translational Arrest*, *bioRxiv*, 2018⁹⁹. As this first thesis chapter describes the exact same data as presented in these two publications^{82,99}, we refrain from continuous and repetitive citation. Nevertheless, we emphasize the fact that the original intellectual effort originated and was laid out in these two papers. Data from the second chapter of this thesis, “Comparing Protein Half-Lives Inside and Outside of Protein-RNA Complexes”, has not been published and is originally presented in this thesis. However, we note here that plans exist to publish this work as well, which might or might not happen before expiry of the embargo period on this thesis.*

The first chapter of this thesis is dedicated to Martin Trendel, who taught me that it is okay to think different. And if you don't think so, that's okay.

CHAPTER 1: XRNAX AS PLATFORM FOR INTERROGATING PROTEIN-RNA INTERACTIONS

The first chapter of this thesis is devoted to XRNAX, a method for the extraction of protein-crosslinked RNA from UV-crosslinked cultured cells. The development of XRNAX is described along with benchmarking experiments and four dedicated applications. These applications are the identification of protein-sequences interfacing with RNA, the discovery of RNA-binding proteomes, the differential quantification of RNA-binding within a proteome and the identification of RNA-sequences interfacing with one particular protein. The presented findings have been originally published in Trendel et al.⁸², which serves as a reference for the entire chapter.

EXTENDING TRIZOL EXTRACTION TO PROTEIN-CROSSLINKED RNA

THE BEHAVIOUR OF PROTEIN-CROSSLINKED RNA DURING CONVENTIONAL TRIZOL EXTRACTION

The standard extraction method for total RNA from cultured cells or tissue is acid guanidinium thiocyanate–phenol–chloroform extraction^{100,101}, or “TRIZOL” extraction. In a TRIZOL extraction cells are lysed with guanidinium thiocyanate and phenol containing buffer, which denatures their protein and nucleic acid content almost completely. Chloroform is added, which leads to phase-separation, so that the phenolic phase contains the protein and the aqueous phase contains the RNA. Notably, at low pH DNA collapses into the interphase between the aqueous and organic phase because its phosphate backbone is protonated and uncharged. RNA remains in the aqueous phase because of its additional polarity from the 3' hydroxyl groups. In cases where genomic DNA has not been excessively fragmented, the RNA in the aqueous fraction of the extraction is very pure and can be separated to be alcohol precipitated. After alcohol precipitation extracts only contain RNA, so that the entire extraction and clean-up occurs in two steps of a few minutes, which has made TRIZOL very popular and widespread.

In order to extract protein-crosslinked RNA from UV-crosslinked cells we wanted to make use of the powerful extraction and separation powers of TRIZOL. Figure 1B illustrates that using the original TRIZOL protocol there was no indication that RNA extracted from UV-crosslinked cells was crosslinked to protein. Furthermore, HPLC-MS did not detect significant amounts of protein in these extracts after they had been RNase and trypsin digested. This indicated that TRIZOL extraction had effectively removed protein-crosslinked RNA from the aqueous phase.

EXTRACTION OF PROTEIN-CROSSLINKED RNA FROM UV-CROSSLINKED CELLS: XRNAX

As protein-crosslinked RNA is a hybrid molecule we suspected that it had probably been dragged into the interphase, where its RNA part had access to the aqueous phase and its protein part access to the organic phase. In order to extract the protein-crosslinked RNA we discarded the aqueous phase and transferred the interphase to a new vial, where we

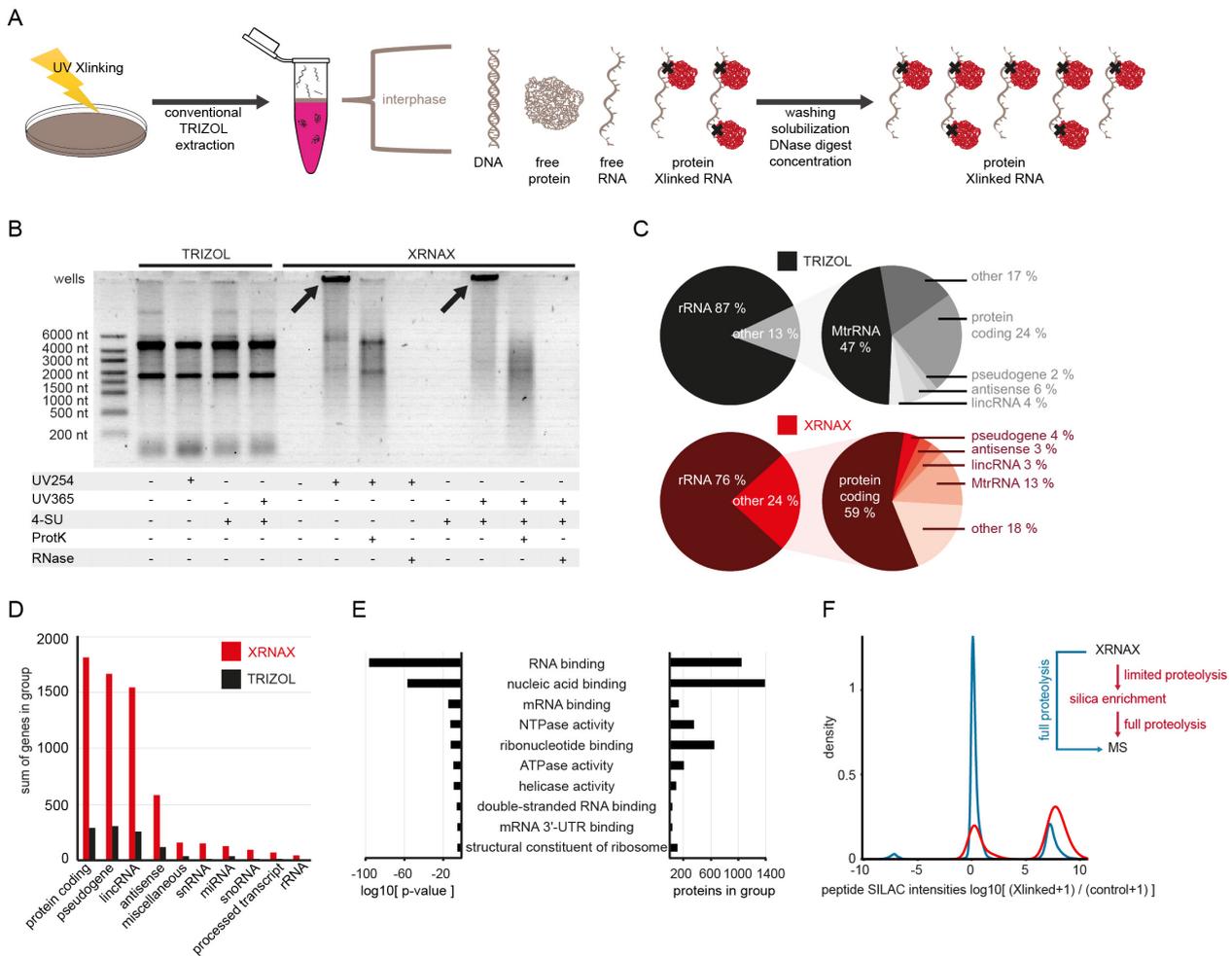


Figure 1: XRNAX extracts protein-crosslinked RNA from UV-crosslinked cells.
Adapted from Trendel et al., 2018.⁸²

A) Experimental scheme of XRNAX.

B) Comparison of classic TRIZOL and XRNAX by agarose gel electrophoresis. MCF7 cells were subjected to crosslinking with UV at 254 nm (UV254) or incubated with 100 μ M 4-thiouridine (4-SU) for 16 hours and crosslinked at 365 nm (UV365). RNA was extracted using either conventional TRIZOL or XRNAX and subsequently treated as indicated. ProtK: Digestion with proteinase K. RNase: Digestion with RNase A and I.

C) Pie diagrams comparing RNA composition of TRIZOL extracts (top) in comparison to XRNAX extracts (bottom). RNA from RNAX extracts or TRIZOL extracts was sequenced before (left) and after (right) depletion of ribosomal RNA. Reads for each GENCODE biotype were normalized to the total number of reads in one library.

D) Transcripts exclusively identified after sequencing of either TRIZOL or XRNAX-extracted RNA from MCF7 cells that were grown in the presence of 4-SU. Sum of genes are summarized within the top-10 GENCODE biotypes.

E) GO-enrichment analysis (molecular function) for proteins in XRNAX extracts from MCF7 cells. Displayed are the ten most-enriched terms.

F) Density plot showing SILAC ratios of peptides from MCF7 XRNAX-extracts before (blue) and after silica purification (red). Pseudo-counts were added to display peptides exclusively identified in either SILAC channel.

solubilized it in cycles of washes with increasing concentration of SDS (for details see methods). After solubilisation the interphase could be alcohol precipitated, DNase digested, again alcohol precipitated and subsequently taken up in any desired buffer or volume. When run on an agarose gel these extracts showed remarkably different behaviour than RNA extracted with the standard TRIZOL protocol (Figure 1B). Interestingly, RNA from these extracts showed a very extreme shift, which left most of it in the well. This shift could be reversed by proteinase K treatment, which led RNA to travel into the gel. RNase treatment did not leave any signal behind, indicating that the sole nucleic acid content of the extracts was RNA. Performing the entire extraction on non-crosslinked cells returned less than 1 % of the amount of RNA that crosslinked cells yielded. Furthermore, two different UV-crosslinking strategies led to the same result. Cells could either be crosslinked using UV at 254 nm wavelength or upon labelling of their transcriptome with the photoactivatable nucleotide 4-thiouridine (4-SU) with UV of 365 nm wavelength.

We named the method consisting of UV-crosslinking, lysis in TRIZOL, solubilization of the interphase in SDS-containing buffer and DNase digestion protein-Xlinked RNA extraction or XRNAX (Figure 1A, for details see method section).

As mentioned in the introduction, another method using TRIZOL for the purification of proteins UV-crosslinked to RNA was developed while we developed XRNAX. OOPS⁸¹ also performs conventional TRIZOL extraction on UV-crosslinked cells and transfers the interphase to a fresh tube. However, instead of preparing the interphase for DNase digestion, OOPS solubilizes the interphase in TRI reagent again in order to repeat the conventional TRIZOL extraction. This process is repeated again, so that the interphase is dissociated in TRI reagent for a total of three times. Importantly, after three rounds of TRIZOL extraction the organic phase does not contain free protein anymore because it was removed in previous rounds. RNase is added to the aqueous phase and the sample incubated at 37 °C, so that protein crosslinked to RNA is released into the organic phase. The organic phase is collected, protein precipitated, trypsin digested and analysed by MS.

THE RNA CONTENT OF XRNAX EXTRACTS

To understand the composition of the RNA in XRNAX extracts we purified it further and sequenced it. Specifically, we digested XRNAX extracts from MCF7 cells crosslinked at 254 nm UV with proteinase K in order to make the entirety of RNA accessible for reverse transcription. We compared RNA obtained in this way to RNA treated in the identical fashion from non-crosslinked MCF7 cells, which had been extracted using conventional TRIZOL extraction. In order to quantify also the contribution of rRNA to each extract we sequenced before and after rRNA depletion (RiboZero). As illustrated in Figure 1C, XRNAX doubled the proportion of non-ribosomal RNA in comparison to TRIZOL extracted RNA and significantly increased the proportion of protein-coding RNA within this share. Moreover, mitochondrial ribosomal RNA was also depleted from XRNAX extracts by more than a factor of three.

We proteinase K digested and sequenced RNA extracted with XRNAX from 365 nm UV-crosslinked MCF7 cells, which had been exposed to 4-SU for 16 hours and compared it to RNA extracted with TRIZOL from the same cells without crosslinking. As visible in Figure S1B&C the

coverage for low abundant and medium abundant transcripts was much higher in XRNAX extracted samples. Consequently, sequencing of XRNAX samples discovered more than 6000 transcripts, which had not been discovered in sequencing from TRIZOL extracted samples from the same cells including some 1500 protein-coding transcripts, 1500 pseudogenes and 1500 long intergenic non-coding RNAs (lincRNAs) (Figure 1D).

In summary our experiments showed that XRNAX extracts contained all the RNA biotypes also contained in TRIZOL-extracted RNA, which is considered the current gold standard of total RNA extractions.¹⁰⁰ Surprisingly, and not previously anticipated, XRNAX extracts were skewed in their relative composition of RNA species in a favourable way, so that highly redundant transcripts such cytosolic or mitochondrial ribosomal RNA occurred less frequently. As we intended to use XRNAX for the interrogation of protein-RNA interactions we considered this an advantage because a more diverse pool of RNA potentially allows for the detection of a larger number of protein-RNA interactions. Moreover, having more copies of otherwise rare RNA species increases the chance for detecting rare protein-RNA interactions. Thus, we found the RNA makeup of XRNAX extracts ideally suited for the exploration of its interaction with protein.

THE PROTEIN CONTENT OF XRNAX EXTRACTS

To understand the protein composition of XRNAX extracts we RNase and trypsin digested an extract from MCF7 cells and subjected its peptide content to HPLC-MS analysis. Gene ontology (GO) enrichment analysis revealed the term 'RNA-binding' most significantly enriched and next to 'nucleic-acid binding' the most populated term (Figure 1E). We used intensity-based absolute quantification (iBAQ¹⁰²) to compare the composition of XRNAX extracts to a total proteome of MCF7 cells, recorded using the identical MS method (two hour gradient QExactive). Figure S1A shows that XRNAX dramatically changed the relative protein compositions between the two samples. While proteins with the highest abundance in the MCF7 total proteome were histones, cytoskeletal proteins and metabolic enzymes, XRNAX extracts contained proteins of the HNRNP class next to splicing factors and other nuclear RNA-binding proteins (Figure S1A). As ribosomal proteins and RNA are very abundant in human cells, we were interested if ribosomal proteins also occurred in high abundance in XRNAX extracts, thereby potentially masking proteins with lower abundance during proteomic analysis. Although some ribosomal proteins were among the most abundant ones, their combined iBAQ intensity only amounted to 7.9 % of the total iBAQ intensity. In comparison, ribosomal proteins made up 13.1 % of the integrated iBAQ of MCF7 total lysates. A similar case could be made for histones, where the iBAQ intensity added up to 3.5 % in XRNAX extracts, whereas it was 13.6 % of total iBAQ intensity in MCF7 total lysates. The latter finding was important because TRIZOL interphases contain concentrated chromatin and histones reportedly bind RNA⁸⁸.

Comparing XRNAX extracts to total lysates showed that they were very different and indeed highly enriched in RNA-binding proteins. Moreover, XRNAX extracts were not dominated by one or several very abundant proteins but had a proteomic depth that was readily accessible by MS.

To control for unspecific and indirect protein-protein rather than direct protein-RNA interactions we employed stable isotope labeling in cell culture (SILAC¹⁰³). Cells of one SILAC label were UV-crosslinked, while cells of another SILAC label were left non-crosslinked before both were combined and subjected to XRNAX. When we analyzed XRNAX extracts of this kind with a standard proteomic workflow we did see enrichment for peptides from the crosslinked SILAC channel, however, a majority of peptides did not show enrichment for either channel (Figure 1F). Many of these peptides without enrichment were derived from *bona fide* RNA-binding proteins. We hypothesized that during the initial TRIZOL extraction of XRNAX RNA-binding proteins from crosslinked and non-crosslinked cells were combined in the interphase to form aggregates. In the following alcohol precipitation steps these interactions had prevented non-crosslinked proteins from being removed. RNA-binding proteins are known to form very stable interactions among themselves and with RNA.³⁸ In fact, a recently reported method for the discovery of RNA-binding proteins had made use of the observation that RNA-binding proteins interact preferably with other RNA-binding proteins.⁴⁷

Following this hypothesis, we designed a denaturing cleanup procedure downstream of XRNAX. If protein was really UV-crosslinked to RNA, purifying RNA using silica columns should allow for the co-purification of UV-crosslinked protein, while removing free, non-crosslinked protein. Silica columns retained only small amounts of protein-crosslinked RNA under standard conditions. However, when we subjected XRNAX extracts to incomplete tryptic digestion beforehand, the same columns were able to retain much better. Figure 2A outlines the procedure, by which XRNAX extracts underwent a very short tryptic digestion of 30 minutes at low trypsin/LysC concentrations. Subsequently, the digestion was stopped by addition of guanidinium chloride containing binding buffer, the sample completely denatured at 60 °C and purified on a silica column. Indeed, using this additional cleanup we found much more peptides much stronger enriched than before (Figure 1F & Figure S1D&E). These results demonstrated that the enrichment of protein via silica enrichment depended on UV-crosslinking. In order to show that it also depended on RNA we treated an XRNAX extract from heavy SILAC cells with NaOH in the presence of Mg²⁺ in order to hydrolyze all RNA. After neutralizing the pH of the sample, we combined it with the identical amount of XRNAX extract from light SILAC cells whose RNA content was intact. After silica enrichment and MS we found a strong enrichment of light peptides, indicating that the enrichment does in fact depend on RNA, too.

The combination of these experiments demonstrated that XRNAX extracts contained protein UV-crosslinked to RNA, which could be co-purified from free protein using standard silica affinity chromatography.

INTERROGATION OF PROTEIN INTERFACING WITH RNA

We employed XRNAX as a starting point for three applications, all of which we deemed impossible using established methodologies (Figure 2A). In each of those applications XRNAX served as an initial purification step, which was prefixed to a downstream proteomic or transcriptomic method. Firstly, we applied XRNAX for the purification of nucleotide-crosslinked peptides and the localization of protein-RNA interfaces. Secondly, we affinity-

purified RNA from XRNAX extracts for the co-purification of crosslinked protein under denaturing conditions. This allowed us to derive high-confidence RNA-binding proteomes from three cell lines and to quantify RNA-binding differentially upon arsenite stress in MCF7 cells. Lastly, we performed CLIP-Seq experiments from XRNAX extracts, where we substituted RNase fragmentation with high-intensity sonication.

This first paragraph will discuss how XRNAX can be used for the identification of nucleotide-crosslinked peptides, which can serve as direct evidence for the RNA interaction of a peptide sequence within a protein. Using this methodology, we discovered a new RNA-binding domain. For the bioinformatic analysis and of this domain we collaborated with Ananth Prakash of Alex Bateman's laboratory (EMBL-EBI Cambridge). For the experimental validation we received help from Rastislav Horos of Matthias Hentze's laboratory (EMBL Heidelberg), who performed the presented polynucleotide kinase (PNK) assay.

PURIFICATION AND DETECTION OF NUCLEOTIDE-CROSSLINKED PEPTIDES FROM XRNAX EXTRACTS

In order to provide more evidence that XRNAX did extract protein that was crosslinked to RNA we focused on identifying the actual crosslinking site between the two. For this purpose, we attempted to purify peptide-RNA hybrids for the identification with MS, which would not only give direct proof for UV-crosslinking but also pinpoint protein-RNA interaction sites. One earlier study had successfully identified those hybrids, which were derived from interactome capture experiments.⁸³ Although this study showed that identification is in principle possible, it also showed that it is not straightforward, when it reported 60 nucleotide-crosslinked peptides from 35 human proteins. In fact, since then detection of those peptides has turned out so challenging, that two recent approaches, RBDmap and RBR-ID, used their absence for inferring protein-RNA interaction sites.^{39,84}

RNA-binding proteins interact through a specific part of their peptide sequence – typically an RNA-binding domain – with the RNA molecule they reside on. Upon irradiation with UV-light RNA bases get activated and react to this close-by part of the peptide sequence.⁵³ As this happens only sparsely, crosslinked proteins are believed to be connected to RNA only via one or few ribonucleotides.^{83,104} We hypothesized that if all protein of an XRNAX extract was digested with trypsin, only the UV-crosslinked peptides would be left covalently attached to the RNA. We digested an XRNAX extract from MCF7 cells with large excess of trypsin and purified its RNA content using standard silica columns (Figure 2A). This way we could wash away any free peptides and recover only the ones crosslinked to RNA. Subsequent RNase digestion left behind tryptic peptides attached to ribonucleotides.

We were able to enhance the purity of our sample further by applying alkaline fractionation before standard acidic LC-MS detection. During high-pH fractionation residual RNA fragments – a major contamination during LC-MS detection – eluted very early. However, nucleotide-crosslinked peptides gained enough hydrophobicity to elute later and could be separated as such.

DATABASE SEARCH FOR NUCLEOTIDE-CROSSLINKED PEPTIDES

Very little is known about the chemical adducts formed between ribonucleobases and the amino acids in a protein upon UV-irradiation. In principle every combination of the

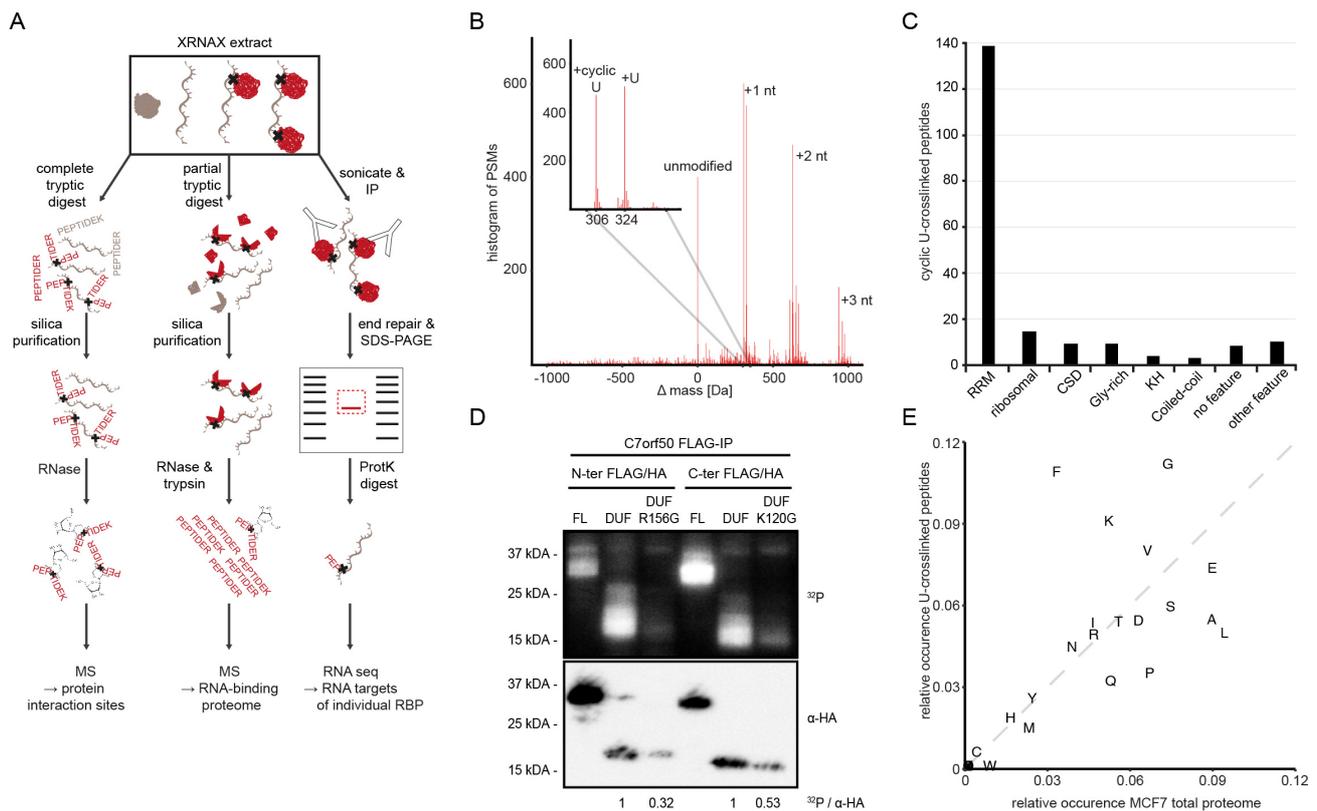


Figure 2: Identification of RNA interaction sites and domains.

Adapted from Trendel et al., 2018.⁸²

- A) Proteomic and transcriptomic applications downstream of XRNAX.
- B) Identification of adduct masses on XRNAX-purified peptides by MSFragger.
- C) Uniprot feature-annotation of cyclic U-crosslinked peptides. Bar graph displays number of cyclic-U crosslinked peptides that were found in stretches of their cognate protein with indicated feature. RRM: RNA recognition motif; KH: K homology domain; CSD: cold shock domain; Gly-rich: glycine-rich amino acid sequence; no feature: no feature deposited in Uniprot; other feature: feature other than the ones mentioned in the other categories.
- D) PNK assay for C7orf50 and its DUF2373. MCF7 cells were transiently transfected with the full-length protein (FL), DUF2373 of C7orf50 (DUF), or two mutant DUFs (R156G, K120G). Top: Representative phosphoimage showing radioactive labeling of covalently bound RNA in RNP complexes. Bottom: Anti-HA Western blot of same membrane showing efficiency of anti-FLAG IP. Values at the bottom indicate relative RNA-binding of DUF2373 to the respective DUF2373 mutant using relative intensities from the images above.
- E) Comparison of amino acid frequencies in peptides crosslinked to cyclic U and peptides found in the MCF7 total proteome.

ribonucleotides A, U, G, C and any of the 20 amino acids has to be considered, when searching MS data for this kind of modifications. Kramer et al. had addressed this problem with dedicated software that used *a priori* assumptions for the chemical amino acid-nucleotide adducts, which they were trying to identify.⁸³ We preferred to take an unbiased approach and decided to use the recently reported ultrafast, mass-tolerant search engine MSfragger¹⁰⁵ to overcome this problem. Indeed, MSfragger identified an adduct mass of 324 Dalton, corresponding to uridine monophosphate (U), as well as an adduct mass of 306 Dalton, corresponding to cyclic uridine monophosphate (cyclic U), as the most dominant modification in our sample (Figure 2B). This was in accordance with the findings by Kramer et al., who had suffixed titanium dioxide enrichment to interactome capture for the enrichment of nucleotide crosslinked peptides. As interactome capture only purifies polyadenylated RNA we were not surprised that beside most peptides described in the latter study we found many novel ones, i.e. 178 peptides from 64 proteins, all of which were annotated as RNA-binding. Interestingly, peptides were exclusively detected crosslinked to U, cyclic U or permutations of di- and trinucleotide sequences carrying at least one U. Momentarily we cannot conclude if this resulted from only U crosslinking to protein or if crosslinking sites with other ribonucleotides were instable and thereby escaped MS/MS detection. Using high pH fractionation, we were able to decrease interference from free RNA, which led to the detection of 197 cyclic U-crosslinked peptides from 93 proteins. Interestingly, in these MS runs we detected almost exclusively cyclic U crosslinked peptides and only very few peptides crosslinked to several nucleotides. This suggested that at pH=10 the charge on non-cyclic ribonucleotide crosslinked peptides had shifted them into the earlier fractions, which we had excluded because of their notorious contamination with RNA fragments.

NUCLEOTIDE-CROSSLINKED PEPTIDES PINPOINT PROTEIN-RNA INTERFACES

Most of the cyclic U-crosslinked peptides we could locate in RNA-recognition motif (RRM), K-homology (KH) or cold-shock domains (CSD) within their host proteins (Figure 2C). Ninety percent of proteins identified were annotated as RNA-binding and more than 85 % of nucleotide-crosslinked peptides mapped to *bona fide* RNA-binding domains or ribosomal proteins, confirming that these peptides represented protein-RNA interfaces. Apart from classical RNA-binding domains we found glycine-rich regions frequently crosslinked. This was in accordance with earlier RBDmap and RBR-ID studies, which had found low-complexity regions abundantly involved in RNA-binding.^{39,84} The proteins with the most crosslinked peptides identified were HNRNPA2B1 (13), NCL (10) and HNRNPAB (8), all of which located to RRM3 or glycine-rich regions (Figure S2A). Beside ribosomal proteins we found cyclic U-crosslinked peptides for proteins, which are known to interact with non-coding RNA. For example we identified the interactors of Steroid Receptor RNA Activator (SRA RNA) SLIRP and SPEN (also known as SHARP, MINT).¹⁰⁶ We could locate the RNA interaction site of SPEN to its RRM3, which reportedly is essential in SRA RNA-binding.¹⁰⁷ RBDmap had located the interaction site of SPEN with polyadenylated RNA to its RRM1³⁹, suggesting that the two methods resolve alternative binding modes depending on the RNA biotype they capture. We encountered several instances where cyclic U-crosslinked peptides located to non-canonical

RNA-binding domains. In a number of instances these putative protein-RNA interaction sites pointed out interesting structural or functional properties. For example, we found the ribosome biogenesis factor LTV1 and located its RNA interface to a C-terminal coiled-coil region. Although no RNA-binding domain is known for LTV1 cryo-electron microscopy studies had indicated this region of the protein to interact with ribosomal RNA during pre-40S ribosome assembly.¹⁰⁸ We found a nucleotide-crosslinked peptide in the SLED DNA-binding domain of SCML2A, which had been suspected to recruit polycomb repressive complex 1 (PRC1) to genomic loci under the guidance of non-coding RNA. An *in vitro* study had found a region neighboring the SLED domain capable of binding the non-coding RNA HOTAIR.¹⁰⁹ Our finding suggested that another binding mode existed, which might appear independent or in sequence of connected binding modes during RNA-guided recruitment. In yet another case we located a cyclic U-crosslinked peptide in the AAA+ ATPase domain of HNRNPU (also known as SAF-A). Only recently a study had reported the importance of this domain in HNRNPU's role in chromosome organization and *in vitro* experiments with the isolated domain indicated that its ATP hydrolyzing activity is indeed stimulated in the presence of RNA.¹¹⁰ Our findings implied that the C-terminal region of the AAA+ ATPase domain interacted with RNA, suggesting that this stimulation might occur through RNA directly.

To our great interest, we identified a nucleotide-crosslinked peptide in a domain of unknown function (DUF). Its host protein was the uncharacterized protein C7orf50, which had been reported as RNA-binding in HeLa cells before⁷⁸, however, did not carry any known RNA-binding domains. C7orf50 had not appeared in the above mentioned RBDmap or RBR-ID study.^{39,84} Interactions in the STRING database suggested that it might be involved in ribosome biogenesis, which was supported by nucleolar localization reported from the Protein Atlas database¹¹¹ (Figure S2C). Pfam annotated C7orf50 with only one domain, DUF2373, in which we identified a nucleotide-crosslinked peptide towards its C-terminal end. Using HMMER on the complete Uniprot database we were able to find 17 high confidence homologues with full-length DUF2373 from twelve vertebrate species. These uncharacterized proteins shared high sequence similarity within the domain, including the C-terminal part where we found the putative RNA interaction site (Figure S2B). Using tertiary structure prediction, we could not derive a confident structural model for DUF2373 from known structures. However, DUF2373 was predicted to fold into six alpha helices, which might assemble into an alpha-helical bundle. We expressed full-length C7orf50 and its DUF2373 in *E.coli* and performed *in vitro* agarose gel shift assays with the proteins, which had been incubated with total RNA from MCF7 cells (Figure S2F). In both cases we could confirm RNA-binding. Consequently, we transiently expressed FLAG-HA-tagged versions of the proteins in MCF7 cells and could show RNA-binding also *in vivo* with a PNK assay (Figure 2D). Notably, RNA-binding was decreased when DUF2373 was mutated at sites, which had been computationally predicted to interact with polynucleotides. Additionally, we enriched transiently expressed FLAG-HA-tagged proteins from RNase digested XRNAS extracts of MCF7 cells. Interestingly, MSfragger identified ribonucleotide crosslinked peptides for full-length C7orf50, the wild type DUF2373 but not DUF2373 mutants. Our finding suggested that DUF2373 was a conserved vertebrate RNA-binding domain for which we propose the name WKF in reference to the three most conserved

residues in its consensus sequence. This naming scheme has been adopted by the Pfam database.

A comparison of amino-acid frequencies between all peptides crosslinked to cyclic U and all peptides identified in our control MCF7 total proteome showed a clear enrichment for phenylalanine, lysine and glycine in the crosslinked sequences (Figure 2E). As RNA might obstruct tryptic cleavage of crosslinked peptide sequences even if lysine is next to a crosslinked amino acid, we could not exclude that the enrichment in lysine was due to increased miscleavage and not crosslinking. However, arginine frequencies were identical between the two samples arguing that lysine might be directly involved in crosslinking. We searched our cyclic U crosslinked peptides with the motif search engine MEME and were able to identify the conserved RNP1 or RNP2 motif in 71 % of cases. Both motifs are rich in phenylalanine and known to be vitally involved in RNA recognition of the RRM.¹¹² While this again confirmed that nucleotide-crosslinked peptides identified protein-RNA interfaces, it also pointed out the importance of phenylalanine in those interfaces during UV-crosslinking. A recent *in vitro* study had been able to identify the phenylalanine-uracil crosslinking adduct from the RRM of the splicing factor PTBP1 through isotopic labeling of RNA and protein.⁷⁷ Our data suggested that this adduct might be common and causal for many UV-induced protein-RNA crosslinks. In contrast to earlier studies that were limited to polyadenylated RNA⁸³, XRNAX extracted ribosomal RNA, too. We identified cyclic U-crosslinked peptides from seven ribosomal proteins that we could locate in the cryo-EM structure of the human 80S ribosome (Figure S2E). Notably, for four out of six peptides that contained phenylalanine, we found phenylalanine in close proximity to a uracil base in the RNA sequence.

In summary, our findings demonstrated that nucleotide-crosslinked peptides identified from XRNAX extracts could give insights to canonical and non-canonical protein-RNA interactions. These interactions suggested regulatory functions where RNA was employed for protein recruitment (in the case of SPEN, SLIRP and SCML2A) or modulation of enzymatic function (in the case of HNRNPU). Furthermore, nucleotide-crosslinked peptides gave structural insights as showcased for proteins in the ribosome and provided evidence for an entirely new RNA-binding domain.

DISCOVERY OF PROTEINS INTERACTING WITH RNA

As demonstrated in the previous paragraph nucleotide-crosslinked peptides from XRNAX extracts are the most direct MS evidence showing protein-RNA interactions. However, what becomes obvious from the number of proteins identified in our but also previous studies⁸³ is that these peptides are not easily detected and are typically derived from highly abundant proteins. In order to catalogue RNA-binding proteins interactome capture did not rely on these crosslinked peptides but on the enrichment of peptides from a UV-crosslinked sample over a non-crosslinked control.^{65,78} This way an RNA-binding protein can contribute many tryptic peptides, which do not necessarily need to be crosslinked directly to RNA, but rather indirectly by being connected to some crosslinked peptide within the complete sequence of the protein. Consequently, many more identifications can be made and many more proteins can be classified as RNA-binding, although the MS evidence is more indirect. The following

paragraph describes a SILAC-controlled strategy for cataloguing RNA-binding proteins, which we applied to three commonly used cell lines. We compare this catalogue to previous catalogues derived by interactome capture and describe commonalities and differences.

SILAC-CONTROLLED ENRICHMENT FOR RNA-CROSSLINKED PROTEIN FROM XRNAX EXTRACTS

Earlier experiments had shown that UV-crosslinked protein in XRNAX extracts could be purified from non-crosslinked protein using silica affinity chromatography. By employing this principle, we went on to derive a comprehensive draft of the RNA-binding proteome beyond the known and confirmed proteins binding to polyadenylated RNA. Again, we used SILAC to control for unspecific interactions of proteins that were not crosslinked to RNA. Therefore UV-crosslinked cells of one SILAC label were combined with non-crosslinked cells of the complementary label. After XRNAX and silica column purification the enrichment was so strong that most of the peptides did not have any intensity in the non-crosslinked SILAC channel at all (Figure 1F). To classify peptides as crosslinked or not we decided on a stringent cut-off of 1000 fold, which was intentionally beyond the SILAC dynamic range of approximately 100 fold.¹¹³ We called peptides with such an enrichment super-enriched peptides and utilized them for the classification of proteins as RNA-binding.

In order for silica columns to retain significant amounts of protein-crosslinked RNA we had established that XRNAX extracts needed to be partially digested beforehand. We optimized predigestion conditions towards detection of the maximal number of super-enriched peptides, which potentially discovered the maximal number of RNA-binding proteins. By varying the predigestion time at constant trypsin/LysC concentrations we came to the realization that there was no optimal predigestion time but that different predigestion times led to the identification of different super-enriched peptides. Nevertheless, for achieving maximum peptide diversity among super-enriched peptides there was an optimal combination of predigestion times. For the following analysis we prepared samples with 15 and 30 minutes of partial tryptic digestion, which were purified and analyzed in separate.

We noticed that another factor influencing peptide diversity was the growth state of cells used for our preparation. Confluent cells returned a different set of peptides than half-confluent cells did. While it intuitively makes sense that rapidly dividing cells have a different composition of their RNA-binding proteome than contact-inhibited cells, we wanted to integrate both these states in our analysis to derive the most general RNA-binding proteome for one cell line. Consequently, we decided to produce XRNAX extracts from half-confluent and confluent cells, each of which were subjected to silica purification after 15 or 30 minutes of partial tryptic digest, respectively. Thereby a collection of four independent samples was produced, each fractionated at high pH and analyzed in six fractions. In order to assess the reproducibility of this system we produced two independent sets of samples from MCF7 cells, swapping the SILAC-label for UV-crosslinked and non-crosslinked control cells. Figure 3A illustrates that, indeed, both sets of samples not only discovered a highly similar set of proteins through super-enriched peptides, but the number of super-enriched peptides identifying a protein was highly similar, too.

Thus, denaturing silica purification of XRNAX extracts provided a way for the controlled and

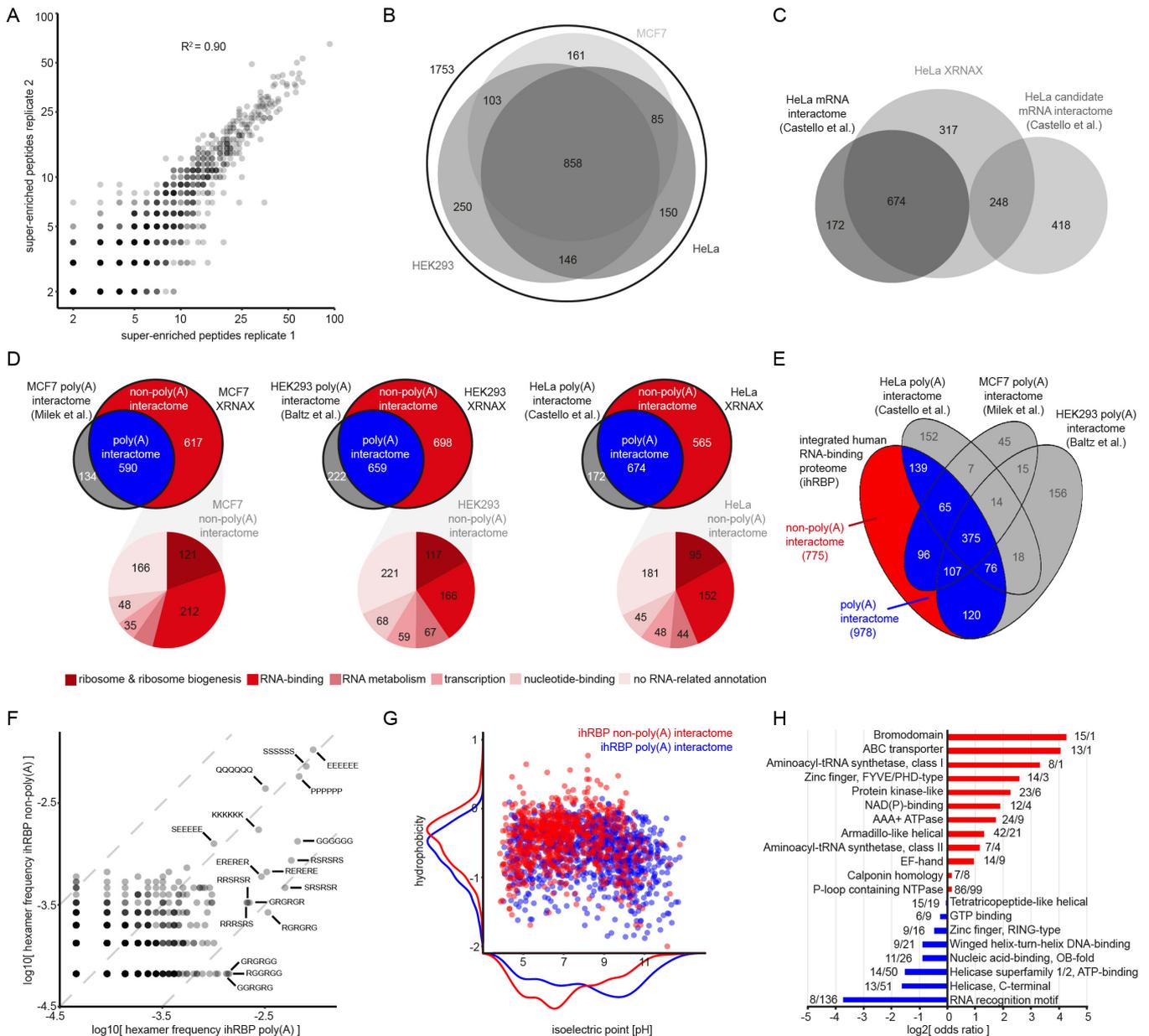


Figure 3: The integrated human RNA-binding proteome derived from 3 cell lines. Adapted from Trendel et al., 2018.⁸²

- A) Scatter plot showing reproducibility of protein identifications by super-enriched peptides. Each point represents one protein and indicates how many super-enriched peptides were found per replicate.
- B) Venn diagram for XRNAX-derived RNA-binding proteomes. Numbers indicate the proteins identified in each of the three cell lines MCF7, HeLa and HEK293.
- C) Venn diagram comparing the XRNAX-derived HeLa RNA-binding proteome to the published poly(A) RNA interactome by Castello et al., 2012⁷⁸.
- D) Composition of XRNAX-derived RNA-binding proteomes. Top: Venn diagrams comparing XRNAX-derived RNA-binding proteomes to published poly(A) interactomes. Non-poly(A) interactomes (red) were derived by subtraction of poly(A) interactomes (blue). Bottom: Pie charts for the functional annotation of proteins in the non-poly(A) interactome into five RNA-related categories. Categories were applied hierarchically in the listed order and proteins assigned to the first category they shared a GO (molecular function) annotation with.
- E) Venn diagram comparing the integrated human RNA-binding proteome (ihRBP) to published poly(A) interactomes. A non-poly(A) RNA interactome (red) was derived by subtraction of the combined known poly(A) interactomes (blue).
- F) Scatter plot comparing hexamer frequencies in poly(A) and non-poly(A) interactomes. All possible hexameric permutations of the amino acids G, S, N, Q, P, E, K and R were counted in proteins of each group and normalized to the total number of counts. Dashed lines indicate fold-changes of 1 and 10.
- G) Scatter plot comparing isoelectric points and hydrophobicity of proteins in poly(A) and non-poly(A) interactomes. Density plots outside the axes illustrate the distribution for each feature.
- H) Odds ratios of interpro domain occurrences in poly(A) and non-poly(A) interactomes. The ten most frequent domains in either group are compared.

reproducible discovery of RNA-binding proteins from SILAC-labeled cells. Furthermore, by using optimized predigestion conditions and cells of different growth states we were confident to use the system at its maximal potential.

AN INTEGRATED DRAFT FOR THE HUMAN RNA-BINDING PROTEOME FROM THREE CELL LINES

We applied our enrichment scheme to three commonly used human cells lines, which had been subject of interactome capture studies before – MCF7, HeLa and HEK293. To classify proteins identified in one cell line as RNA-binding we only considered proteins that were detected with at least two unique super-enriched peptides. This presented us with high-confidence RNA-binding proteomes of 1207 proteins for MCF7, 1239 proteins for HeLa and 1357 proteins for HEK293 cells, among of which 858 were shared by all three cell lines (Figure 3B).

For clarity we state here that in the following the term ‘poly(A) RNA’ and ‘poly(A) interactome’ refer to RNA that carries polyadenylation and the protein interactome of RNA that carries polyadenylation, respectively, which includes but is not limited to most mRNA molecules.¹¹⁴

A comparison of HeLa interactome capture data from Castello et al. to the XRNAX-derived HeLa RNA-binding proteome illustrated the powers of XRNAX over interactome capture. XRNAX rediscovered 674 of the 846 proteins found by interactome capture and added almost 600 more (Figure 3C). Although we applied a much more stringent enrichment cut-off for calling RNA-binding proteins than Castello et al. did, we were able to confirm some of the proteins that were earlier reported as ‘candidate mRNA-binder’.⁷⁸ These proteins had been identified in HeLa interactome capture experiments but had failed to reach statistical significance. Proteins which appeared in this interactome capture dataset but not in our HeLa XRNAX interactome had significantly lower confidence levels in comparison to proteins, which had been discovered by both methods (Figure S3A). This indicated that XRNAX added sensitivity and specificity in comparison to previous methodology.

We compared our XRNAX RNA-binding proteomes to other published poly(A) RNA interactomes and found extensive overlap (Figure 3D, top). Seventy five percent or more of proteins found in interactome capture studies were represented in our data, which allowed us to divide our RNA-binding proteome in two sections – the poly(A) interactome and the section that we designated the non-poly(A) interactome. Deep total proteome data for all three cell lines was readily available⁸⁹, which we used as a background control in GO analysis. Unsurprisingly, this revealed strong enrichment for RNA-related terms in the poly(A) interactome (Figure S3B). GO term enrichment for the non-poly(A) interactome was problematic because of the high prevalence of proteins annotated as RNA-binding in the total proteome of the three cell lines. Most of these annotations were assigned in reference to the two initial interactome capture studies by Baltz et al. and Castello et al., both of which identified poly(A) RNA-binding proteins. Since at this moment we were not interested in poly(A) interactors, but non-poly(A) interactors, we removed the published poly(A) interactome from the background control. Consequently, we found strong enrichment for RNA-related terms for this part of the dataset as well. Interestingly, terms relating to non-coding RNA such as rRNA, 7S RNA, tRNA, snRNA or snoRNA and the ribosome were especially

enriched. Moreover, some terms such as ‘aminoacyl-tRNA ligase activity’ or ‘ribonuclease activity’ we found only enriched in the non-poly(A) part of our interactomes. We were interested to see how much of the non-poly(A) interactomes could be explained by current GO annotations. For this purpose we collected RNA-related GO terms in 5 functional classes, i.e. ribosome & ribosome biogenesis, RNA-binding, RNA metabolism, transcription and nucleotide binding. Since many RNA-binding proteins carried annotations represented in several of those categories, we applied them hierarchically and removed proteins from our list as soon as they fit into one category (Figure 3D, bottom). Surprisingly, only approx. 30 % proteins of each non-poly(A) interactome did not carry an annotation that fit into any of the categories. In regard to the entire RNA interactomes derived by XRNAX, including the poly(A) and the non-poly(A) interactomes, this reflected approx. 15 % of proteins.

Combining our findings from all three cell lines resulted in a collection of 1753 proteins that we named the integrated human RNA-binding proteome or ihRBP, containing 978 proteins (or 70 %) of previous poly(A)-interactomes, and 775 in the non-poly(A)-interactome (Figure 3E). Comparison of the ihRBP to the census of RNA-binding proteins from Gerstberger et al., who had catalogued RNA-binding proteins by combining computational analyses with manual curation⁸⁷ showed 55 % overlap. This confirmed an observation made in earlier studies that computational prediction of RNA-binding is often incomplete, because it can be independent of protein domains and involve intrinsically disordered regions (IDRs)^{39,78,88}. Many of those proteins were known interactors of non-coding RNA, such as POLR1A – catalytic core of RNA polymerase I – or translation initiation factors not previously identified by interactome capture like ABCE1, EIF2S3, EIF3B, EIF3J. Moreover, the non-poly(A) interactome contained numerous ribosomal proteins and 17 out of 23 amino-acyl tRNA synthetases represented in the ihRBP. We found proteins involved in micro RNA (miRNA) biogenesis such as DICER1, TSN and TARBP2 and the interactor of telomerase RNA PINX1. There were numerous proteins involved in rRNA and snRNA biogenesis most prominently the nuclear exosome components SKIV2L, EXOSC2, EXOSC3 and DIS3, as well as splicing factors such as SF3B3, LSM8, ESRP1 or CWC22. Finally, we found many transcription-associated proteins some of which are known to interact with 7SK RNA like HEXIM1, CCNT1 and CDK9.

Thus, the ihRBP was a profound collection of human proteins evidently interacting with poly(A) and non-poly(A) RNA. The addition of proteins interacting with non-poly(A) RNA allowed us to interrogate questions, which earlier interactome capture studies were not able to ask. The first one of those questions was: What is the difference between the two?

A NOVEL GROUP OF RNA-BINDING PROTEINS INTERACTING WITH NON-POLYADENYLATED RNA

Growing evidence suggests that phase-separation of RNA-binding proteins organizes membrane-less macromolecular assemblies such as nucleoli³⁵ or Cajal bodies, and RNA granules^{35,37,38} such as stress granules or P-bodies (for review see^{34,115}). Feric et al. had shown that nucleolar organization is driven by different surface tensions in protein droplets contributing to different subcompartments of the nucleolus. Importantly, they had concluded surface tension and viscoelasticity of protein droplets is a ‘sequence-encoded’ feature that

crucially determines participation of proteins in higher order structures as well as their behavior inside such structures.

With the ihRBP we had the so far largest collection of experimentally derived human RNA-binding proteins at our disposal. Furthermore, by intersecting our data with interactome capture data, we could divide this collection in two parts, a poly(A) and non-poly(A) interactome. We hypothesized that members of those groups must differ in their 'sequence-encoded' information, so that they become part of different macromolecular assemblies that interact with different biotypes of RNA, i.e. poly(A) RNA or non-poly(A) RNA. Consequently, by comparing those two groups to each other we should be able to find out what the 'sequence encoded' information was, that divided them. The first question we asked was how much sequence information was necessary to find any differences at all between the two groups. We approached this problem systematically by comparing low level sequence information – amino acid frequency – then adding sequence information step-wise by comparing dipeptide and tripeptide frequencies. Figure S3C illustrates that while amino acid frequencies were virtually identical between the two groups, differences became apparent when more peptide information was added, so that dissimilarities in tripeptide frequencies turned out significant (Wilcoxon ranksum test $p=0.02$). Notably, these dissimilarities did not arise from control sets of the same size that randomly sampled proteins from the ihRBP. Interestingly, proteins of the poly(A) interactome showed strong enrichment for tripeptides carrying permutations of amino acids known to contribute to IDRs, i.e. glycine (G), serine (S), asparagine (N), glutamine (Q), proline (P), glutamic acid (E), lysine (K), and arginine (R).¹¹⁶ Some of those fairly abundant tripeptides like the motif RGG occurred more than five times more often in the poly(A) interactome than in the non-poly(A) interactome. To get a more detailed picture we limited the analysis to the eight IDR amino acids and compared hexamer frequencies between the two groups (Figure 3F & Figure S3D). Again, we found strong enrichment of some specific motifs in the poly(A) interactome. The strongest enrichment was for one with alternating occurrence of R and G, which occurred more than 10 times more frequently in the poly(A) interactome than in the non-poly(A) interactome. Motifs containing analogous alternations of either R and S or R and E were similarly enriched. Other tetramers especially often occurring in the poly(A) interactome carried an RGG motif or a poly-G stretch. Notably, the two groups did not simply differ in the appearance or absence of low-complexity motifs in general but differed in the appearance of those specific motifs. Other motifs, such as poly(E), poly(S), poly(P), poly(Q) or poly(K) stretches, occurred with very similar frequencies in both groups. Interestingly, proteins carrying those repeats were often associated with similar molecular processes independent of the biotype of RNA they bound to. For example, a common GO annotation for proteins carrying a poly(E) stretch was 'chromatin organization'. These findings showed that, indeed, 'sequence-encoded' information did distinguish poly(A) binding from non-poly(A) binding proteins. Moreover, the sequences that distinguished them the most, were those low-complexity motifs that in numerous cases had been reported to drive liquid-liquid phase transition.^{35,37,38,117,118} For clarity we state here that poly(A) binding proteins had been reported to carry low complexity motifs with higher frequency than the rest of the human proteome before (for review see¹¹⁹). However, our comparison of the

poly(A) interactome to a very specific subset of the human proteome – the non-poly(A) interactome – added a functional consequence, which earlier studies were not able to derive. Next, we asked if properties of the global amino acid sequence of proteins in the poly(A) or the non-poly(A) interactome could separate them, too. For this purpose we compared hydrophobicity, isoelectric point, mass weight and charge state at pH=7 between the two groups (Figure S3E). We found that poly(A) binders were generally more hydrophobic and had a more alkaline isoelectric point (pI), whereas non-poly(A) binders were larger and often carried more negative charges. Especially pI and hydrophobicity differed greatly between the groups, again showing that ‘sequence-encoded’ information distinguished them. Nevertheless, Figure 3G illustrates that the variance between them was great, too, so that neither pI nor hydrophobicity alone could point out one group over another.

In summary we could show that poly(A) and non-poly(A) binding proteins differed in their local and global amino acid sequence features. While those differences were highly significant, no single feature was predictive for either category suggesting that in a cell their superposition is necessary to guide the assembly of poly(A) or non-poly(A) ribonucleoprotein particles into higher order structures.

A comparison of interpro domains between the poly(A) and the non-poly(A) interactome revealed notable commonalities and differences between the groups (Figure 3H). Together with P-loop containing nucleoside triphosphate hydrolases, the RRM was the most abundant domain overall. Interestingly, proteins carrying RRMs were almost exclusively found in the poly(A) interactome. Among zinc finger proteins the RING-type domain was more prevalent with poly(A) binders, whereas the FYVE/PHD-type occurred more often in the non-poly(A) interactome. Some domains within the P-loop containing nucleoside triphosphate hydrolases were enriched in the non-poly(A) interactome, i.e. the ABC-transporter-like and the AAA+ ATPase domain. As expected, we found both classes of aminoacyl-tRNA synthetase domains strongly enriched in the non-poly(A) interactome. Surprisingly, the domain with the strongest enrichment in the non-poly(A) interactome was the bromodomain – a very rare domain among poly(A)-RNA binders. Many of these bromodomain-carrying proteins such as BRD2, BRD3, BRD4, BAZ1A, BRPF1, KMT2A, PHIP, TRIM33, ATAD2 and SMARCA4 were important histone modifiers¹²⁰. Interestingly, one very recent report had found the bromodomain itself to interact with enhancer RNA (eRNA), which leads to the recruitment of BRD4 to enhancers. We followed this lead and looked for proteins known to interact with chromatin in the non-poly(A) interactome using the STRING interaction database. Several interesting clusters occurred, the first of which included RUVBL1, member of the NuA4 complex, RSF1 and the above-mentioned BAZ1A. We identified TP53 and the DNA-damage regulators TP53BP1, RIF1, MDC1, as well as BRCA1. TP53BP1⁹⁴ and BRCA1¹²¹ had been reported to interact with non-coding RNA, whereas RNA-binding of TP53 had been controversial¹²² and for MDC1 and RIF1 not previously described. Notably, we detected 25 super-enriched peptides for TP53BP1. Thereby it was one of the proteins discovered with the highest confidence among non-poly(A) interactors. Another cluster was identified around proteins involved in spindle pole assembly, such as BUB3, AHCTF1, CKAP5, PDS5A, KIF2C, KIF11 and CENPF. Again, none of these proteins had been reported to bind RNA before. The same was true for a different cluster around the

condensin complex, which included SMC2, SMC4, NCAPD2, NCAPD3, NCAPDG and NCAPG2. Finally, we could make out a cluster of proteins around POLR1A, i.e. RNA polymerase I. This included CHD1, member of the SAGA complex, ATRX, member of the ATRX:DAXX complex, and TTF1, the RNA polymerase I transcription termination factor. Except for POLR1A, we could not find previous reports for any of the proteins in this cluster to interact with RNA.

As non-coding RNAs emerge as important regulators in genome regulation²⁴, it becomes more important to find their protein interactors in order to get mechanistic insights to their action⁴². Our non-poly(A) interactome revealed RNA-binding for many important chromatin components, whose function could potentially be guided by RNA.

QUANTIFICATION OF CHANGES IN THE INTERACTION OF PROTEINS WITH RNA

Using a non-crosslinked control and SILAC had allowed for the discovery of over 1700 RNA-binding proteins from three cell lines and well over a thousand proteins from each individual cell line. We hypothesized that instead of quantifying the enrichment of UV-crosslinked protein over a non-crosslinked control, XRNAX in combination with SILAC could quantify RNA-binding differentially between two crosslinked populations of cells. Therefore, cells of one SILAC label should undergo a certain treatment before UV-crosslinking and should then be combined with untreated cells of the complementary SILAC label, which should also be UV-crosslinked. In the first part of this paragraph we describe experimental findings demonstrating that this quantification is in fact possible and accurate if the background is considered and excluded. We then applied this quantification in a timeline experiment, where MCF7 cells were exposed to arsenite – an established stressor for the induction of translational arrest (for review see^{123,124}). Serendipitously, we found that on the total proteome level arsenite induced a fast and severe form of autophagy, which selectively eliminated proteins associated with protein biosynthesis. While this finding showed the importance of controlling changes on the RNA-level with changes on the total proteome level, its details were only tangentially involved in the differential quantification of RNA-binding and will therefore be discussed in an autophagy ‘excursion’ annexed to this paragraph.

BACKGROUND AND DYNAMIC RANGE FOR THE DIFFERENTIAL QUANTIFICATION OF RNA-BINDING WITH XRNAX

Before quantifying the dynamics of the RNA-bound proteome during differential treatment we established that XRNAX was in fact able to assess RNA-binding quantitatively. To test this, we mixed decreasing amounts of UV-crosslinked MCF7 cells with non-crosslinked cells (both grown with heavy SILAC label), and combined this with a fixed amount of UV-crosslinked cells of a light SILAC label. Thereby, unmixed, UV-crosslinked cells represented 100 % RNA-binding, whereas dilutions with non-crosslinked cells represented decreased RNA-binding, e.g. a 1:4 dilution 25 % RNA-binding (Figure S1G). For MS quantification, all samples were combined with a fixed amount of UV-crosslinked cells carrying a light SILAC label. Indeed, after processing samples via XRNAX and silica enrichment, most peptides were accurately quantified according to their dilution over two orders of magnitude. Additionally, there was a noticeable number of peptides that stayed constant, independent of how many cells were UV-

crosslinked in their parent sample. A comparison to our earlier data revealed that peptides that showed quantitative foldchanges were exclusively found among the super-enriched peptides of the ihRBP. *Vice versa*, peptides that remained constant were found among the less enriched peptides that we had discarded while generating the ihRBP.

In conclusion, this demonstrated that super-enriched peptides of the ihRBP provided a library of peptides that could be used for the accurate quantification of protein UV-crosslinked to RNA. By filtering for those peptides XRNAX could be used for the accurate quantification of changes in RNA-binding between conditions.

DIFFERENTIAL QUANTIFICATION OF PROTEIN-RNA INTERACTION DURING TRANSLATIONAL ARREST

We treated MCF7 cells with arsenite to study the effects of translational arrest on protein-RNA interactions. Arsenite stress had been reported to effectively induce translational arrest followed by the formation of stress granules (for review see ^{123,124}). In order to choose an adequate timescale, we monitored protein biosynthesis using azido homoalanine (AHA) labelling and SILAC in MCF7 cells (Figure 4A). During AHA-labelling nascent proteins incorporate the methionine analogue AHA to be captured by click-chemistry, which upon MS gives a direct measure for translation^{125,126}. We compared protein biosynthesis in MCF7 cells during one hour of arsenite stress in comparison to an untreated control. Indeed, translation was heavily decreased in arsenite-treated cells after 60 minutes to 23% (median) compared to the untreated control. The timeline indicated that translational arrest was not immediate but occurred gradually with the biggest changes arising within the first 30 minutes of treatment. In the expectation that the largest changes in RNA-binding would also occur in these 30 minutes we chose a timeline with five points within this window. We also verified that arsenite did not compromise the chemical stability or induced major degradation of RNA directly. Accordingly, we did not observe noticeable changes in the amount or degree of degradation of total RNA extracted from cells that underwent arsenite treatment (Figure S4D&E). To quantify RNA-binding differentially MCF7 cells of one SILAC label were challenged with 100 μ M arsenite and compared to untreated cells of the complementary label. In order to control for changes in the total proteome we quantified total proteomes at the same time points. After stringent filtering we obtained quantitative data for the association of 765 proteins with RNA over all time points (Figure 5B). Correspondingly, 3489 proteins in the total proteome were reproducibly quantified over all time points (Figure 4B) and the intersection between the two datasets contained 633 proteins (Figure 5D). Firstly, we were interested in general trends in RNA-binding under arsenite stress. Figure 5A&B illustrate that most proteins remained constant in their interaction with RNA over all time points measured. In fact, 90 % quantiles after 30 minutes were within 0.77 - 1.21 foldchange between arsenite treated cells over control cells. Nevertheless, several proteins showed a distinct change in RNA-binding, most of them decreasing, whereas only one protein increased RNA-binding more than twofold – TP53BP1. The kinetic profile of proteins that changed their association with RNA was typically a steady incline or decline with the exception of the exosome component EXOSC2, which already showed significantly increased binding upon five minutes of arsenite stress, and then stayed constant. To control for abundances, we checked proteins from our RNA-binding

analysis against the same proteins in our total proteome analysis (Figure 5C&D). Interestingly, in the total proteome data two groups emerged – one larger one that stayed unchanged over all time points and one smaller one that showed small but significant decrease of protein abundance over time. Figure 5C shows that this group made up approximately 25 % of proteins in the observation. We compared changes in RNA-binding and changes in the total proteome directly, which allowed us to deconvolute the contribution of either component to our measurement (Figure 5E and Figure S5A). This revealed for the two proteins with the strongest increase in RNA-binding, TP53BP1 and EXOSC2, that their increase could be entirely contributed to the association with RNA, because their absolute abundance stayed constant over all timepoints. For the group of proteins that showed decreased RNA-binding a simultaneous decrease in total protein abundance was observed. However, at no timepoint this decrease in the total proteome outweighed and explained their much stronger dissociation from RNA. Three proteins with the strongest decrease in RNA-binding after 30 minutes were the ribosomal proteins RPS28, RPS14 and RPS3. Since arsenite stress is known to induce translational arrest^{123,124} we located the proteins in the cryo-EM structure of the human 80S ribosome to find out if their changed binding might relate to any of their structural features. Indeed, we found all three proteins in the cleft that channels mRNA through the two ribosomal subunits (Figure 5F). This allowed the assumption that a decrease in RNA-binding for those proteins might have been due to polysomes disassembling from mRNA upon arsenite-induced translational arrest.^{127,128} Another protein with effectively reduced RNA-binding was USP10, which reportedly has a key role in stress granule formation.¹²⁴ To get a better understanding of translational control during our time course experiment we examined RNA-binding and proteomic changes for all eukaryotic translation initiation factors (EIFs) in our analysis (Figure S5B). A general trend within the group was a steady decrease in protein abundance, so that after 30 minutes of arsenite treatment there was a 40-50 % reduction in total protein, while for most EIFs RNA-binding stayed basically constant. Two exceptions were EIF2S2 and EIF2D, which notably are not members of the same complex (despite confusingly similar names). EIF2S2, also known as eIF2- β and core component of the EIF2 complex involved in 43S preinitiation complex formation (for review see¹²⁹), increased binding steadily over all timepoints, although its protein abundance decreased significantly. In fact, by normalizing its increase in RNA-binding with protein abundance, we found its effective increase in RNA-binding with 2.8 fold after 30 minutes stronger than for any other protein in the experiment. EIF2D, also known as Ligatin and not member of the canonical EIF2 complex¹³⁰, decreased RNA-binding. However, its effective RNA-binding after normalization to protein abundance was unchanged for all timepoints. A third protein standing out was EIF4A3, which did not show any changes in abundance or RNA-binding at any timepoint. EIF4A3 is the RNA-binding component of the exon junction complex and thereby only tangentially involved in translation initiation¹³¹, which might explain why it did not share the same behaviour as the other EIFs.

Our data illustrated that quantification of RNA-binding using XRNAX was able to recapitulated processes known to occur during translational arrest. Moreover, it identified protein players and mechanisms that future studies may address in more detail.

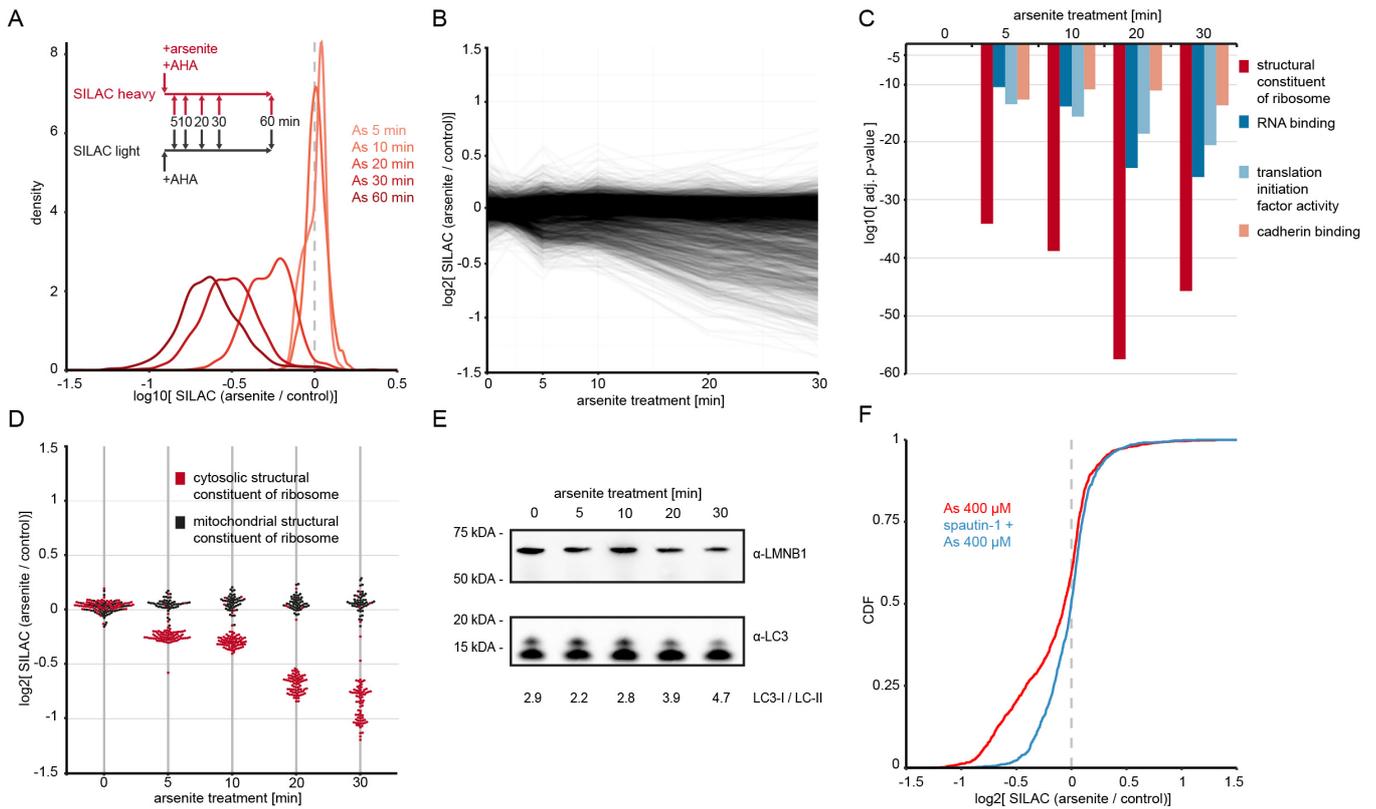


Figure 4: Impact of arsenite stress on the proteome.

Adapted from Trendel et al., 2018.⁸²

- A) Density plot showing fold-changes of nascent proteins produced under normal conditions or arsenite stress. Displayed are means of duplicate experiments with label-swap.
- B) Time course of total proteome changes during arsenite stress in MCF7 cells. Each line represents one protein. Values are means of duplicate experiments with label swap filtered for a variance of <15 %.
- C) GO-enrichment analysis (molecular function) of proteins that change in expression during arsenite stress. Shown are the top-4 GO terms with highest significances after 30 minutes.
- D) Dotplot displaying changes in total proteome upon arsenite stress for proteins under the GO term 'structural constituent of ribosome'. Each dot represents one protein. Values are means of duplicate experiments with label swap filtered for a variance of <15 %.
- E) Western blot for LC3 to monitor autophagic flux in MCF7 cells over 30 minutes of arsenite treatment. LMNB1 was used as a loading control. Numbers under images quantify the relative intensity of LC3-I and LC3-II bands.
- F) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment, with (blue) and without (red) inhibition of autophagy by spautin-1 (see also Figure S4J).

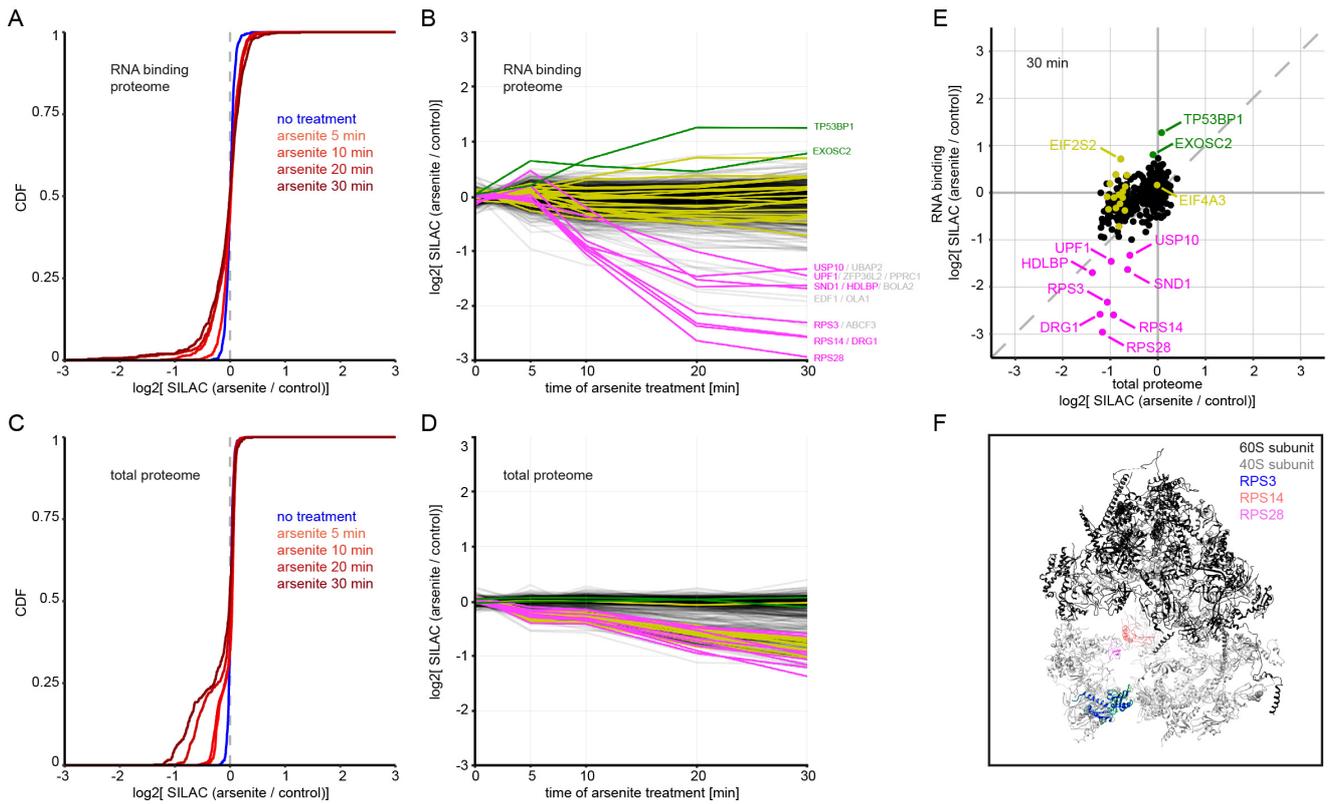


Figure 5: Changes in the total and RNA-interacting proteome induced by arsenite-mediated translational arrest. Adapted from Trendel et al., 2018.⁸²

- A) Cumulative changes in the RNA-binding proteome, assessed by XRNAX in MCF7 cells that were treated with arsenite for indicated times.
- B) As in panel A, showing temporal data for individual proteins. Each line represents one protein. Values displayed are means of duplicate experiments with label swap filtered for a variance of <15 %.
- C) Cumulative distribution of changes in the total MCF7 proteome upon arsenite stress for proteins displayed in panel A.
- D) Timeline of changes in the total proteome upon arsenite stress for proteins displayed in panel B.
- E) Scatter plots comparing changes in the total proteome to changes in RNA-binding after 30 minutes of arsenite stress. For other time points see Figure S5.
- Color-coding in panels B, D and E refers only to proteins quantified in both the RNA-binding and total proteome. The proteins TP53BP1 and EXOSC2 are displayed in green, proteins with >50 % change in RNA-binding after 30 minutes are displayed in magenta, and EIFs are displayed in yellow.
- F) Cryo-EM structure of the human ribosome highlighting the location of RPS28, RPS14 and RPS3. For visibility nucleic acids are not displayed.

EXCURSION: ARSENITE INDUCES MASSIVE AND SPECIFIC AUTOPHAGY ON A TIMESCALE OF MINUTES

Our control data for the normalization of RNA-bound protein to total protein had revealed that the MCF7 proteome split up in two parts during the response to arsenite. While most proteins retained their abundance constant a subset showed steady decrease, which resulted in significantly reduced levels of those proteins after 30 minutes. Turning away from only the RNA-binding proteome and now looking at all proteins quantified (Figure 4B), we asked which proteins were affected and performed rank-based GO enrichment analysis. For a list of proteins ranked from most decreased to least decreased upon arsenite stress we found very significant enrichment for the process 'translation initiation'. Figure 4C tracks the terms with the most extreme enrichment for molecular function over all timepoints. Notably, after 5 minutes of treatment ribosomal proteins were already affected to a degree that let them rank in a very significant way. We isolated proteins carrying the most enriched GO terms to analyse their behaviour over time (Figure 4D & Figure S4A&B). For ribosomal proteins we found two populations, only one of which changed over time. Closer examination revealed that the unaffected population were mitochondrial ribosomal proteins, whereas proteins of the cytosolic ribosome decreased in abundance without exception. As observed earlier EIFs were collectively affected with a kinetic similar to ribosomal proteins. In addition, other RNA-binding proteins decreased in abundance, most of them directly linked to protein biosynthesis.

These results showed that arsenite-induced translational arrest was followed by a specific protein degradation process. This degradation process affected the translational machinery and especially ribosomal proteins, which were reduced by up to 50 % after 30 minutes.

Arsenite exposure had been shown to induce autophagy in cultured cells^{132,133} as well as *in vivo*¹³⁴. These observations had been made on a timescale of several hours to days at high nanomolar to low micromolar arsenite concentrations. Buchan et al. had reported that stress granules are cleared by autophagy after high micromolar arsenite stress is relieved from cultured cells. We note that typical concentrations to induce translational arrest or stress granules in cultured cells range from arsenite concentrations of 40-1000 μM ^{135,136}. Our data indicated that high micromolar concentrations of arsenite induced a degradation process on the timescale of minutes. In order to get a better understanding of the arsenite-induced degradation process we performed a titration experiment from 0-1600 μM arsenite. We treated MCF7 cells for 30 minutes with increasing doses of arsenite and compared their protein content to an untreated control using SILAC. Figure S4G illustrates the relationship between arsenite concentration and the degree to which protein in arsenite treated cells was degraded. Interestingly, the effect that we had observed before was even more extreme at higher doses of arsenite. Again, we found ribosomal proteins collectively among the most affected. We used their reduction as a function of arsenite concentration to generate a dose-response model and found a good fit (Figure S4G, insert). Notably, these findings were readily reproducible in HeLa cells at very similar effective concentrations (Figure S4H). We noticed variability in the magnitude of the effect from cells that were not grown under the exact same conditions. This variance could be narrowed down to the growth state of cells, when we compared the degree to which arsenite-induced protein degradation affected cells that were

seeded at the same density but harvested in five consecutive days (Figure S4J). We chose a seeding density that produced confluent cells after 5 days for this purpose. Maximum degradation upon arsenite stress was observed on the third day, so during log-phase expansion and maximal growth.

We asked what the underlying mechanism for the degradation process was and interfered with proteasomal and autophagic degradation, respectively. MCF7 cells that were pretreated with bortezomib showed the same degree of degradation as untreated cells upon 30 minutes of arsenite stress (Figure S4F). However, degradation was markedly decreased by pretreatment with the autophagy inhibitor spautin-1¹³⁷ (Figure 4F and Figure S4F). Western blotting against the autophagy marker LC3 confirmed increased autophagic flux during the course of arsenite treatment, as indicated by increased conversion of LC3-I to LC3-II (Figure 4E). Interestingly, the baseline autophagic flux of MCF7 cells in this growth state occurred to be high, however, it got markedly increased after 30 minutes of arsenite stress. We compared the effect of arsenite-induced degradation to rapamycin-induced degradation. Rapamycin by itself induced very modest degradation within 30 minutes of treatment (Figure S4K). However, for a list of most degraded to least degraded proteins, ranked GO analysis returned enrichment for very similar terms as seen for arsenite treatment (data not shown). Rapamycin-induced degradation could be augmented by additional arsenite stress, yet, combined treatment did not exceed the maximal degradation observed for arsenite stress by itself (Figure S4K).

These findings showed that arsenite-induced degradation was a process of autophagy and independent of the proteasome. Protein degradation occurred on a timescale of minutes and depended in its intensity on the growth state of cells. Furthermore, arsenite-induced autophagy seemed to occur independent of MTOR control.

INTERROGATION OF RNA INTERFACING WITH PROTEIN

So far, we had used XRNAX for proteomic applications, however, we also wanted to explore the transcriptomic possibilities arising from the method. Conceptually, there were two approaches to this, both of which are presented in the following paragraph. The first approach was to map the interaction sites of one particular protein-of-interest with the transcriptome by prefixing XRNAX to a conventional crosslinking and immunoprecipitation and sequencing (CLIPseq) experiment. Here, the purity and concentration of XRNAX extracts give advantages, which will be discussed in more detail below. The second approach was to map the interaction sites not of one particular protein-of-interest but of the entire RNA-binding proteome with the transcriptome. In this case chemical biology was used to enrich all crosslinking sites from XRNAX extracts and their location in the transcriptome through sequencing. At the time this thesis was written this last section was still in the process of optimization so that the method will only be presented conceptually. However, from data produced in a proof-of-concept experiment our computational collaborator Michael Piechotta from the laboratory of Christoph Dieterich (University of Heidelberg) could show that the method, called PEPseq, was able to gather occupancy information of protein on the entire transcriptome.

XRNAX CLIPSEQ REVEALS INTERACTIONS SITES OF INDIVIDUAL PROTEIN SPECIES WITH THE TRANSCRIPTOME

Having successfully employed XRNAX for the proteomic analysis of protein-RNA interactions, we next aimed to address this from the other side and identify RNAs that interact with a protein-of-interest. Therefore, we combined XRNAX with a protocol for CLIPseq. Conceptually, XRNAX as a sample preparation step prior to CLIPseq is advantageous for a number of reasons: 1) Contaminants like DNA, which could physically obstruct immunoprecipitation or mask target protein in chromatin complexes, are eliminated. 2) Sample volumes are reduced from millilitres to microliters, thereby allowing for higher antibody concentrations. 3) RNA fragmentation can be supplemented by ultrasonication, thereby circumventing cumbersome optimization and potential biases of RNase treatment¹³⁸.

We selected Lamin B1 (LMNB1) as a CLIP-target in order to validate it as a novel RNA-binder, which we had identified among the proteins with the highest number of super-enriched peptides in the non-poly(A) interactome of MCF7 cells. We fragmented RNA in an XRNAX extract using ultrasonication and immunoprecipitated LMNB1 in a variation of the eCLIP protocol⁶¹ (Figure S6A). In this protocol an aliquot of the sample from which the immunoprecipitation originally occurred is later separated on an SDS-PAGE along with the immunoprecipitated sample itself. This so-called size-matched input control is treated just like the actual sample. This has the advantage that after sequencing it can be used to calculate the enrichment of specific locations in the transcriptome over this input control. Additionally, we used unique molecular identifiers (UMIs) for sequencing, which are RNA-sequencing adaptors with randomized parts. The randomized parts can be used to exclude PCR duplicates from the sequencing data and thereby identify enrichment even for very short or redundant sequences. Without UMIs it is impossible to say if identical reads were produced during PCR, i.e. during sample preparation, or if the protein-of-interest tends to bind one particular sequence especially often. This becomes important for small RNAs such as snRNAs, snoRNA etc. because there are only so many ways to fit a 50 nucleotide read into a 60-200 nucleotide sequence and make it unique. XRNAX CLIPseq on LMNB1 showcased the importance of this when RNA-sequencing identified various non-coding RNAs significantly enriched over the size-matched input control, primarily snoRNAs but also other small nuclear RNAs (Figure 6A). This was in agreement with previous reports showing that Lamin B, with a canonical function in the nuclear lamina, is also structural component of nucleoli where its presence is required to maintain nucleolar integrity during ribosome biogenesis.¹³⁹

The experiments demonstrated that XRNAX extracts could serve as input in CLIPseq experiments, and thereby could be applied for the discovery and validation of novel RNA-binding proteins by MS and RNA-sequencing from the same sample.

Our differential quantification of RNA-binding had shown that EXOSC2 (also known as RRP4) increased its interaction with RNA already after five minutes of arsenite stress (Figure 5B). Since this temporal behaviour stood out among all other proteins, we decided to investigate

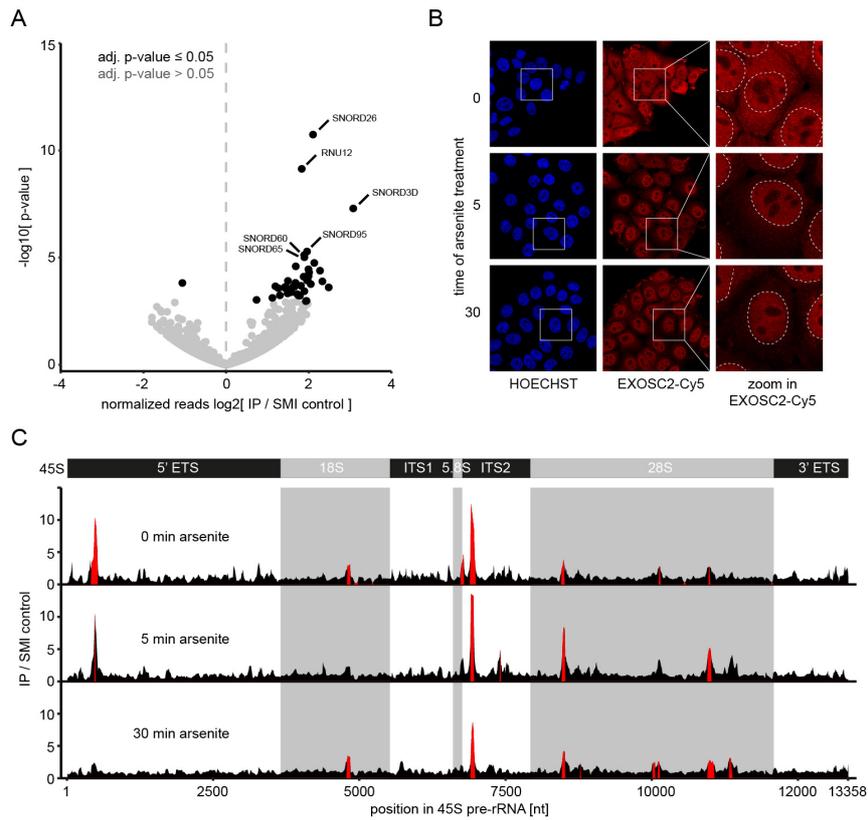


Figure 6: Combination of XRNAX with CLIP-seq to identify RNA bound to novel RNA-binding proteins.

Adapted from Trendel et al., 2018.⁸²

- A) Volcano plot showing the enrichment of non-coding nuclear transcripts in XRNAX CLIP-seq for LMNB1 in MCF7 cells.
- B) Immunofluorescence detection of EXOSC2 with confocal microscopy during translational arrest. White squares in the left and middle images mark the magnified areas displayed on the right. White, dashed lines in images on the right indicate the outline of nuclei.
- C) Gene track for the 45S pre-rRNA displaying coverage of XRNAX-CLIP-seq for EXOSC2, normalized to the SMI control. IP coverage enriched over the SMI control with an adj. p-value $<$ 0.001 is highlighted in red (for details see methods). Top scheme indicates location of transcripts that get processed from the 45S premature transcript. Grey shading refer to the mature 18S, 5.8S and 28S transcripts. 5' ETS: 5' external transcribed spacer, 3' ETS: 3' external transcribed spacer, ITS1: internal transcribed spacer 1, ITS2: internal transcribed spacer 2.

further. EXOSC2 is one component of the non-catalytic 'lid' of the exosome, which is distinct from the catalytic core that degrades RNA¹⁴⁰. In fact, we also detected the core protein EXOSC10, which exhibited a similar yet less pronounced RNA-binding kinetic than EXOSC2. We performed immunohistochemistry and confocal microscopy to locate EXOSC2 upon arsenite stress in MCF7 cells (Figure 6B). Under normal conditions EXOSC2 was located in the cytosol as well as the nucleus. After five minutes of arsenite stress much less EXOSC2 was found in the cytosol, whereas nuclear localization remained, a situation that persisted throughout the 30 minutes of the experiment. Total proteome data had shown that the overall abundance of EXOSC2 (or any other exosomal protein) was not affected by arsenite stress, from what we concluded that EXOSC2 redistributed to the nucleus. Since the change in RNA-binding (Figure 5B) coincided with nuclear localization (Figure 6B) after 5 minutes of arsenite stress, we wanted to learn more about the RNA transcripts EXOSC2 associated with during this process. Therefore, we performed XRNAX CLIPseq for untreated MCF7 cells and compared this to cells treated with arsenite for 5 minutes and 30 minutes. Indeed, we identified particularly nuclear transcripts, which increased their association with EXOSC2 upon arsenite stress. Moreover, we could confirm observations made in yeast where the exosome has been associated with tRNA degradation and snoRNA processing¹⁴¹ – both RNA biotypes with high abundance in our data. As the RNA exosome is known to participate in processing of rRNA (for review see¹⁴²) we were especially interested to find RMRP (adj. p-value = 4.8E-12) among the transcripts with the highest read count beside rRNA itself. RMRP is the RNA moiety of RNase MRP, which cleaves the internal transcribed spacer 1 (ITS1, Figure 6C) during 5.8S rRNA maturation.¹⁴³ CLIP-seq of various RNA exosome components in yeast has revealed that up to 70 % of exosome targets are pre-rRNA.¹⁴⁴ We identified similar numbers in human cells where on average 65 % ($\sigma=13$ %) of unique reads from the EXOSC2 immunoprecipitation mapped to ribosomal transcripts. Except for the 5S rRNA, eukaryotic rRNA is transcribed into one large transcript – the 45S pre-rRNA – and subsequently trimmed into 18S, 5.8S and 28S rRNA by nucleases including the RNA exosome.¹⁴² We analysed read coverage of the 45S pre-rRNA by EXOSC2 and found very significant enrichment for specific regions (Figure 6C). Most significantly enriched during all time points of arsenite stress were segments 3' of the mature 5.8S transcript, which is the main region known to be degraded by the exosome during canonical 5.8S rRNA maturation. Interestingly, in untreated cells two peaks showed significant enrichment, each indicating processing intermediates known as 7S and 6S pre-rRNA¹⁴⁵. After 30 minutes of arsenite stress and increased import of EXOSC2 into the nucleus both intermediates showed much weaker enrichment so that the peak indicating the smaller 6S species was in fact not discernible from background anymore. A similar observation could be made for the segment 5' of the mature 18S transcript, which is known to be degraded cooperatively by the exonuclease XRN2 and the exosome.¹⁴⁶ Both observations implied that import of EXOSC2 drove rRNA processing, consequently eliminating transient intermediates such as the 6S or 7S pre-rRNA.

Our data suggested that EXOSC2 was imported into the nucleus upon arsenite stress in order to promote rRNA maturation. These findings demonstrated that XRNAX CLIPseq could also be applied to examine biological questions differentially, i.e. between conditions.

PEPSEQ REVEALS TRANSCRIPTOME-WIDE OCCUPANCY OF PROTEIN ON RNA

With XRNAX CLIPseq we had shown the occupancy of individual proteins on the transcriptome. This can be powerful for finding out the function of these individual proteins, however, says little about the state of the transcriptome as a whole. Conceptually, finding the occupancy of all protein on RNA is interesting because it could indicate hotspots of protein-RNA interactions, which could only be discovered by sheer luck (or in hundreds of CLIPseq experiments⁶¹) when looking at each protein individually. In the case of DNA the problem is reversed by the fact that it is wrapped into protein, i.e. histones, over the entirety of its length. Nevertheless, for protein-DNA interactions methods such as DNase, MNase or ATACseq have become invaluable tools in assessing accessibility, thereby indicating regulation of the genome.¹⁴⁷ While RNA is believed to be covered by protein, too, the expected conformations are much more heterogenous because a histone-like protein uniformly covering RNA has not been found (for review see⁴). Furthermore, each RNA molecule can engage in protein-RNA interactions individually, in contrast to DNA, which by being a massive molecule has little flexibility in terms of location or interaction partners. Mapping and quantifying protein-RNA interactions transcriptome-wide seems especially promising between conditions, where occupancy can be linked to function. For example, in the case of arsenite-induced translational arrest protein occupancy is expected to be reduced in the coding regions of mRNA. While this intuitively makes sense, other changes in protein-RNA interactions during arsenite exposure are not easily predicted, yet, might be biologically relevant. For example, what happens to the occupancy in rRNA during the autophagic destruction of ribosomal protein? What happens to protein-occupancy during stress-granule formation?

Previously, only one method had tried to look at global protein-RNA interactions on the transcriptome level. Baltz et al. had explored protein-occupancy through sequencing in their initial interactome capture study. Here, polyadenylated RNA was captured on poly(dT) beads from lysates of 365 nm UV-crosslinked cells, RNA fragmented with RNase and protein-RNA complexes run on an SDS-PAGE before blotting onto nitrocellulose. RNA was released with proteinase K and subjected to small RNA sequencing.⁶⁵ Eventually, the quantitative measure for protein-occupancy was derived from T to C conversions, an established sequencing artefact arising from 4-SU crosslinking. The method was later applied differentially for comparing protein-occupancy between HEK293 and MCF7 cells.⁶⁶ An obvious drawback of the procedure is its limitation to polyadenylated RNA. Furthermore, the postulated quantitative properties were never experimentally established and it remained unclear why quantitation needed to be done from T to C conversions and not read counts.

Our previous experiments had shown that photo-crosslinked protein in XRNAX extracts was covalently attached to RNA. From the perspective of RNA this meant that crosslinking sites were flagged by protein. As RNA does not have primary amines, this also meant that crosslinking sites were tagged with at least one primary amine at the N-terminus of the protein, maybe several ones in case the protein contained lysine residues. We hypothesized that we could digest all the protein in an XRNAX extract with proteinase K, purify the RNA with silica columns and label the remaining peptide-snippet crosslinked to RNA with biotin using N-hydroxysuccinimid (NHS) chemistry (Figure S6B). Protein digestion was necessary to remove

excess amino residues and to allow for efficient purification of peptide-crosslinked RNA on silica columns. After fragmentation we could capture biotinylated RNA fragments with streptavidin beads, so that subsequent RNA sequencing would map all protein-RNA interaction sites present at the moment in time when cells were crosslinked. Agarose gel electrophoresis in Figure S6C illustrates that this enrichment strategy worked and enriched crosslinked RNA from MCF7 XRNAX extracts but not non-crosslinked RNA from the same cells extracted with conventional TRIZOL. However, it also illustrates that purification on beads comes with a background that is not easily removed through washing. This problem is known from CLIPseq experiments where newer variations of the protocol use a cleavable tag on the protein-of-interest, which selectively releases RNA fragments from the beads only if the protein they are attached to is degraded.^{148,149} We optimized the method, which we termed peptide-enhanced protein-occupancy sequencing or PEPseq further to decrease background and handling steps. To do so, we used 4-SU crosslinking, whose crosslinks have been shown to cause read-through and T to C conversions instead of termination as UV265 nm crosslinks do (for review see⁵⁷). Additionally, we replaced the initial proteinase K digestion with trypsin digestion. Trypsin left longer peptide sequences behind, which after coupling to beads could be cleaved with proteinase K. This made elution more specific but also possible at more gentle conditions, because releasing biotin-conjugated RNA from streptavidin beads, as initially conceptualized, required boiling in formamide. Since now elution could be performed with proteinase K, we refrained from biotin completely and coupled trypsin-digested, silica purified, ultrasonicated RNA from XRNAX extracts directly to NHS-activated beads. As a control we used the processed XRNAX extracts before reaction to NHS-beads. Thereby, and in combination with UMI sequencing, read counts could be used to compute enrichment of particular transcriptomic regions over the input control. We applied this version of PEPseq to MCF7 cells, which had been treated with arsenite for 0, 15 and 30 minutes.

At the time this thesis was completed RNA sequencing was unfortunately not completed and data could not be presented. However, we wanted to present the concept of PEPseq, which could become a valuable addition to the family of XRNAX-derived protocols.

The second chapter of this thesis is dedicated to Christina Milz, who taught me that everything in life has a half-life. So better be quick, or half of it is gone.

CHAPTER 2: COMPARING PROTEIN HALF-LIVES INSIDE AND OUTSIDE OF PROTEIN-RNA COMPLEXES

XRNAX has opened new possibilities for the proteomic quantification of protein-RNA interactions. In the following chapter XRNAX is applied to investigate if RNA-binding affects the stability of a protein. This question arose from data presented in the previous chapter, where upon arsenite-induced autophagy RNA-bound protein seemed overall less affected by degradation than proteins in the total proteome. Arsenite induced a strong autophagic effect, which led to markedly decreased total levels of specific proteins on a timescale of minutes. This reduction was induced by the addition of arsenite, i.e. a process was induced on top of the normal protein homeostasis or an out-of-equilibrium state was created. In the following chapter we want to turn to the question if proteins in protein-RNA complexes are more stable in the equilibrium state, too. A recent report by McShane et al.⁹⁸, which quantified protein half-lives in mouse and human cells, found that proteins in complexes often showed two degradation kinetics, one rapid initial one and one much slower kinetic as soon as the protein became 'old'. This supported a model by which proteins in complexes were produced super-stoichiometric in regard to the complex itself, so that protein in the complex became stabilized whereas superfluous protein became rapidly degraded. These correlative findings, combined with our observation for the arsenite-induced protein turnover of protein-RNA complexes led us to the working assumption that proteins in protein-RNA complexes are stabilized. Notably, to our knowledge nobody has been able to measure protein half-lives in protein-RNA complexes so far. Therefore, measuring protein half-lives using XRNAX could not only provide an answer to the question if RNA-binding proteins are stabilized on RNA, but additionally could provide experimental evidence for the model by McShane et al., which so far has been based on a correlation.

PROTEIN HALF-LIVES OF THE TOTAL PROTEOME

In order to have a general reference for protein half-lives in the MCF7 proteome we initially determined half-lives of proteins from total lysates. During this reference process the MS methodology and cell culture prerequisites were established so that subsequent experiments measuring half-lives in RNA complexes could be performed under reproducible conditions. In the first part of this paragraph the MS methodology for this purpose is detailed.

Measuring half-lives requires a mathematical model to be fit on the MS data so that parameters can be derived from which half-lives can be computed. Standard models for protein synthesis and degradation typically require a system to be in equilibrium, i.e. in the case of the proteome of a cell, just as much protein should be produced as is degraded. Since the system we were interested in – fast dividing MCF7 cells – was not in equilibrium the second part of this paragraph describes a normalization strategy for half-life MS data called PEN, which allows for relative comparisons of half-lives even for cells that are out of equilibrium.

Gregor Mönke from the laboratory of Alexander Aulehla (EMBL Heidelberg) helped us understand and simulate the consequences of PEN.

DETERMINING PROTEIN HALF-LIVES USING TMT-SILAC

Out-of-equilibrium states of the proteome can be followed through quantification of absolute protein amounts in cells undergoing a perturbation. For example, in the case of arsenite-induced autophagy this would be the reduction of translation-related proteins by up to 50 % over the time course of 30 minutes. For cells in a state of equilibrium, however, by definition, the amounts of protein do not change, so that measuring degradation of a protein is not possible by measuring its absolute amount. Instead, a labelling strategy has to be applied, so that old protein can be distinguished from new protein. The classic labelling strategy for determining protein half-lives uses radioactively labelled amino acids in the culture media (typically ^{35}S methionine), which are incorporated into the proteome during a pulse, followed by a chase of 'cold' media (for review see¹⁵⁰). By fitting a decay function to the radioactive signal which remains in the cells' proteome over the course of time, one can estimate the half-life of all proteins or one specific one after immunoprecipitation. A very recent adaptation of this methodology was applied by McShane et al., who used a pulse-chase approach with the non-radioactive methionine analogue azidohomoalanine (AHA), which can be specifically purified using copper-based click-chemistry¹²⁵. Detection occurred in this case not through autoradiography but through MS, thereby enabling the proteome-wide quantification of protein degradation. AHA-labelling allows for very short labelling times, and consequently for determining half-lives of very short-lived proteins. However, earlier MS-based proteomic studies for determining half-lives used SILAC-based strategies^{151–154}. In the case of SILAC a pulse-only instead of a pulse-chase setup is preferred. During a pulse-only experiment (in the following called pulsed SILAC) light cells are switched to heavy media and harvested after defined timepoints. Degradation is then assessed through the decay of one light peptide in comparison to the emerging cognate, heavy peptide^{153,154}. This approach comes with the drawback that both SILAC channels are subject to change so that none of them can be used for normalization between timepoints. In an ideal case of cells in perfect equilibrium, where synthesis equals degradation, this should not affect the measurement, because timepoints can be normalized to the combined signal of heavy and light. However, a perfect state of equilibrium is practically never achieved in normal cell culture of dividing cells. This becomes problematic for a number of reasons, which are further explored below. Methodologically, however, it led to the development of an advanced triple SILAC strategy, where medium SILAC cells are switched to a heavy label and compared to a separate standard of light SILAC cells¹⁵⁵. Here, degradation could be quantified independent of synthesis for the first time, by relating signal from the changing medium or heavy channel to the constant light channel. In either strategy the analysis required a considerable amount of measurement time, because samples are taken for up to five timepoints, which needed to be analysed in separate. Recently, multiplexing of these measurements was achieved using tandem-mass-tag (TMT) labelling on top of a two channel SILAC strategy.^{156,157} For the purpose of the work presented here we

applied this MS3-based quantification method established by Zecha et al.¹⁵⁷ as outlined in Figure 7A. In principle this strategy allows for multiplexing of up to ten timepoints in one series of measurements (in our case seven to nine timepoints combined and analysed in 16 fractions on two hour gradients). Much like earlier SILAC-based approaches, the TMT-SILAC approach allows for determining protein synthesis and degradation kinetics in the same measurement. As an example, take the time course illustrated in Figure 7A where heavy SILAC cells are switched to light SILAC media for a timeseries of seven points. For simplicity consider a protein represented by a single tryptic peptide. On the MS1 level this peptide will be represented by a light and a heavy precursor, where light represents all the protein synthesized over all timepoints and heavy all the protein that is being degraded over all timepoints. These precursors will be selected for fragmentation individually, i.e. one after the other. Upon fragmentation the TMT reporter ions allow for discrimination and quantification of the individual timepoints within one measurement (MS3 instead of MS2 quantification of the TMT reporter reduces bleed over between the TMT channels, which avoids distortion of the quantification or the so-called 'ratio compression'¹⁵⁸). Therefore TMT-SILAC has a significant advantage regarding data completeness over earlier pulsed SILAC strategies. During a pulsed SILAC experiment early timepoints of the synthesis channel or late timepoints of the degradation channel have low intensity. Therefore, when measuring timepoints separately in a pulsed SILAC experiment early timepoints will have missing values for proteins which are slowly produced, whereas late time points will have missing values for proteins which are rapidly degraded. As TMT quantification is typically very complete, this problem is resolved in a TMT-SILAC experiment where peptides of all timepoints are collapsed into one precursor, which is selected for fragmentation and quantified on the MS3 level. Zecha et al. demonstrated that a TMT-SILAC setup in comparison to a pulsed SILAC setup almost doubles the number of successfully derived protein half-lives.

For determining total proteome half-lives in MCF7 cells we decided to combine the experimental setups from McShane et al., who used seven timepoints within 32 hours, and the TMT-SILAC MS3 methodology from Zecha et al., both illustrated in Figure 7A. Ultimately, we decided for a TMT-SILAC-based approach and against the AHA-based approach since in later experiments we planned to determine half-lives of proteins on RNA using XRNAS, which had been optimized for SILAC.

NORMALIZING PROTEIN HALF-LIVE DATA FOR DIVIDING CELLS

While TMT-SILAC resolves issues of data completeness it does not resolve the more fundamental issue of data normalization and model fitting in dividing cells. These issues are based on the fact that the proteome of proliferating cells is not in equilibrium. This becomes intuitively apparent when one considers that only new cells can be produced if protein synthesis outweighs protein degradation, because new cells need to be made from new protein. The consequences of this for a pulsed SILAC experiment are illustrated in Figure 7B. In the case of equilibrium, decay equals synthesis, so that the flux of protein simply maintains a steady but functional proteome. As soon as there is growth and proliferation of cells, an

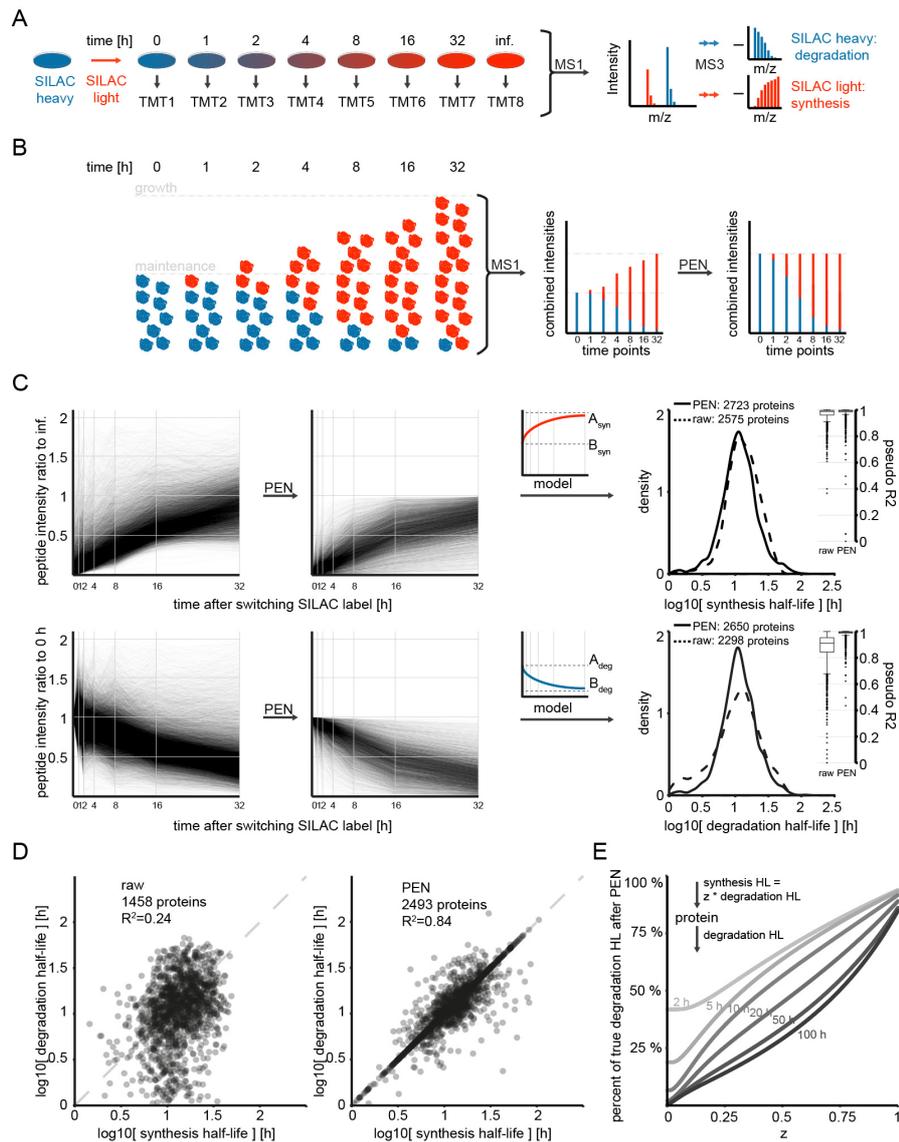


Figure 7: Pseudo-equilibrium-normalization for half-life measurements in dividing cells using TMT-SILAC.

A) MS3-based TMT-SILAC strategy for the determination of protein half-lives adapted from Zecha et al., 2018¹⁵⁷.

B) Schematic representation for the production of new protein in dividing cells over time. In non-dividing cells the amount of protein is at steady state (maintenance), whereas in dividing cells new protein also contributes to the formation of new cells (growth). The desired mathematical model for the calculation of half-lives requires steady-state conditions, which are artificially created through pseudo-equilibrium-normalization (PEN). For details see text.

C) PEN contributes to data smoothness and improves model fitting. Normalization and model fitting for synthesis (top) or degradation data (bottom), respectively, for the determination of protein half-lives within the MCF7 total proteome. Left: Lineplot of raw peptide intensities normalized to time point infinity (synthesis) or 0 (degradation). Each line represents one peptide. Middle: Same data after PEN. Right: Density plot comparing protein half-lives derived from unnormalized data (raw) or PEN-corrected data (PEN). Additional boxplots on the right compare the goodness of fit between the two groups. For details see text or refer to Materials and Methods.

D) Scatter plots comparing protein synthesis and degradation half-lives for the MCF7 total proteome (for definition see text). Each point represents one protein. R^2 values give the Spearman correlation. Half-lives were derived from unnormalized (raw) or PEN-corrected data (PEN).

E) Scatter plot for the effect of PEN on simulated protein degradation data. For details see text.

additional process is put on top of proteome maintenance. This brings about a certain distortion of TMT and SILAC data, which will be discussed in the following in order to introduce a new way of normalizing data before curve fitting.

TMT quantification as applied here works under the assumption that all ten channels contain the protein content of identical amounts of cells, so that reporter ion intensities directly quantify differences between the channels or between those ten individual populations of cells. In praxis this assumption is flawed because the total amount of peptides (representing the proteome of a certain number of cells) in each channel is never identical. Zecha et al. confronted this problem by normalizing all ten channels to a common total intensity of all reporter ions (both SILAC heavy and light) with a constant factor. This means every channel is given the same total intensity of all proteins. Considering the model in Figure 7B, this is a misrepresentation of the actual protein amounts produced over time through proliferation. The ultimate effect is a distortion of the relative reporter ion intensities between timepoints. This becomes obvious in a comparison between timepoint 0 hours and 32 hours in Figure 7B, where the true relative amount of blue protein between the two timepoints is 9 : 1. If all channels are normalized to the same intensity the relative amounts between the time points are diluted to 18 : 1. Zecha et al. corrected for that distortion in their fitting model, which considers proliferation of cells on top of protein maintenance. The model was originally described by Welle et al.¹⁵⁶, who conducted the first proof-of-concept study for half-life measurements using TMT-SILAC, which works under four prerequisites: “1.) Protein synthesis is a zero-order process with respect to protein concentration. 2.) Protein degradation occurs at a constant fractional rate that is uniform for the entire protein pool. Thus, protein degradation can be modelled as a first order process with respect to protein concentration. 3.) The total protein concentration of each cell does not change during the experimental time-course and the system is at steady-state.” Welle et al. subsequently formalize their model in the two following differential equations:

$$E1: \frac{d(\text{protein}_{old})}{dt} = -(k_{deg} + k_{div}) * \text{protein}_{old}$$

$$E2: \frac{d(\text{protein}_{new})}{dt} = k_{syn} - (k_{deg} + k_{div}) * \text{protein}_{new}$$

Where k_{deg} is the degradation constant by which old protein is degraded, k_{syn} is the synthesis constant by which new protein is produced and k_{div} is the cell division constant by which cells divide. Under the assumptions $\text{protein}_{old}(0) = \text{protein}_{steady-state}$ and under the assumption that $\text{protein}_{new}(0) = 0$, Welle et al. solve these equations to:

$$E3: \text{protein}_{old}(t) = \text{protein}_{steady-state} * e^{-(k_{deg}+k_{div}) * t}$$

$$E4: \text{protein}_{new}(t) = \text{protein}_{steady-state} - \text{protein}_{steady-state} * e^{-(k_{deg}+k_{div}) * t}$$

Where $\text{protein}_{steady-state}$ refers to the combined amount of protein, which stays unchanged over the course of the experiment. This formalism violates rather obviously their own

prerequisite 3.) in two ways. First, cells in a steady-state system will not divide, because no new material is produced. That means for a dividing system the assumption $\text{protein}_{\text{old}}(0) = \text{protein}_{\text{steady-state}}$ is false. Second, the reason why an additional constant for cell division is introduced is because of the dilution effect on the SILAC channel representing old protein. That means this constant implies that the concentration of proteins in cells does change over time through dilution.

Introducing a cell division constant that is added to the degradation constant implicitly defines cell division as a zero-order process on top of protein production necessary for maintenance. While this does not comply with the steady-state prerequisite necessary for solving the set of differential equations E1 and E2, Figure 7B shows that the intuition of the model does make sense. The dilemma can be summarized by saying that only in an equilibrium state half-lives can be calculated from the model formalized in equations E1 and E2, however, dividing cells are not at equilibrium and therefore not eligible for analysis by this model. Other studies circumvented this problem by focusing on non-dividing cells, which are arrested through contact inhibition¹⁵⁶, serum starvation¹⁵⁴ or by their nature as post-mitotic, primary cells¹⁵⁹. This comes with a number of drawbacks, which in the case of primary cells is accessibility and the question in how far findings can be generalized from one very specialized cell type. In the case of serum-starvation or contact inhibition, the treatment itself is designed to take cells into a state where they do not prosper. This raises the question in how far observations made in this state are relevant for cell line systems, which are usually used under the exact opposite conditions, i.e. where contact inhibition is generally avoided and cells are tuned to proliferate. Therefore, we sought a method for data transformation, which can be used to analyze data from dividing cells with the above model. Instead of correcting half-lives after fitting data to a model, which is not suitable for it in the first place, we asked the question, how can we transform the data beforehand, so that it becomes suitable for the model?

As mentioned earlier TMT channels usually differ in their total intensity because labeling the exact same amount of peptides is difficult. Zecha et al. corrected for this fact by multiplying each reporter intensity of a channel with a correction constant, so that the sum of all intensities was identical between channels. We took this idea further and broke it down to the single peptide level. In a system of non-proliferating cells only proteome maintenance exists as displayed in Figure 7B. This means at all timepoints the combined amounts of old peptide and new peptide add up to 100 %, thus, the summed heavy and light SILAC intensities of a peptide should add up to the same combined SILAC intensity over all timepoints. In order to transform data from dividing cells into the shape expected for non-dividing cells, we calculated correction factors for each peptide and each TMT channel. Specifically, for each timepoint this combined the light and heavy reporter ion intensities of an individual peptide and calculated a correction factor, which corrected all timepoints to the same intensity as present in the 32 hour timepoint. Correction could in principle occur towards any timepoint, however, the 32 hour timepoint had the most complete data, which allowed for deriving the most comprehensive list of correction factors. Consequently, after multiplication with the timepoint-specific correction factor, the combined reporter ion intensities of a peptide from the heavy and light channel was identical between timepoints. Also, these correction factors

could be applied to the SILAC channels independently as apparent from the following equation:

$$E5: (TMTintensity_{i,j,SILACheavy} + TMTintensity_{i,j,SILAClight}) * c_{i,j} \\ = TMTintensity_{i,j,SILAClight} * c_{i,j} + TMTintensity_{i,j,SILACheavy} * c_{i,j}$$

Where $TMTintensity_{i,j,SILAC}$ denotes to the reporter ion intensity of a peptide i at timepoint j , which was derived from selection of a SILAC heavy or light precursor, respectively. $c_{i,j}$ denotes to the correction factor of the same peptide i at the same timepoint j . This meant that after calculation of these correction factors data from heavy and light precursors, i.e. synthesis and degradation, could be separated and subsequently corrected. We called this correction strategy PEN for pseudo-equilibrium-normalization. For a detailed description of the TMT-SILAC MS data processing and PEN implementation refer to the Materials and Methods section.

Interestingly, this per-peptide correction had a profound effect on data smoothness as visible in Figure 7C. Notably, most peptides after PEN showed strictly monotonous growth or decay, respectively, which could not have been accomplished by multiplying all intensities of a timepoint by a single correction factor as Zecha et al. had done. The increased smoothness improved the quality by which model fitting could be achieved very significantly (Figure 7C). Interestingly, PEN was more beneficial for fitting decay data so that 15 % more protein degradation half-lives with overall significantly better fits were derived in comparison to fits of uncorrected data. Furthermore, Figure 7D shows that synthesis half-lives (the amount of time required to double any amount of new protein) and degradation half-lives (the amount of time required to half any amount of old protein) had much better overlap if data was corrected with PEN. Using uncorrected data, both were derived in 43 % of cases, yet, using PEN-corrected data both were derived in 87 % of cases. Importantly, the direct comparison of synthesis and degradation half-lives in Figure 7D demonstrates that PEN succeeded in transforming measurements from proliferating cells into measurements alike to non-dividing cells, where synthesis and degradation are in equilibrium. While for uncorrected data the median ratio between synthesis and degradation half-lives was 0.87 ($\sigma=0.91$), it was 1.00 ($\sigma=0.62$) for PEN-corrected data. Moreover, PEN-corrected half-lives ranged from 0.7 – 123.5 hours, whereas uncorrected half-lives only ranged from 0.14 – 67.5 hours. Hence, PEN successfully transformed half-life measurements from proliferating cells into the desired form of pseudo-equilibrium and additionally had surprisingly beneficial effects regarding data smoothness, fitting, reproducibility between the SILAC channels as well as dynamic range. Since synthesis and degradation half-lives were basically identical in our data, we regarded them as replicate measurements and for simplicity continued to use their mean for all following analysis. We denote to this value with the general term protein half-life.

The underlying aim of this study was to compare the half-life of proteins in the total proteome to their half-life in protein-RNA complexes. For this purpose, we used the identical cell line system, i.e. MCF7 cells, under identical culture conditions, i.e. a timeline of 32 hours after three days of expansion in culture. All following analysis of half-lives in the upcoming

paragraphs were relative comparisons within this system, so that we refrained from further correction of our data in respect to cell doubling times. We decided to proceed in this way because cell doubling times can vary significantly within the duration of an experiment (data not shown), so that the assumption of previous studies that cell division follows a zero-order process is not necessarily correct. Consequently, the protein half-lives presented here must not be seen as absolute values of time but rather as values of time in a coordinate system that transforms measurements of proliferating cells to measurements of cells in a steady state. In that sense, the derived synthesis and degradation constants and the resulting half-lives are the coalescence of proteome maintenance and growth as depicted in Figure 7B. Still, in order to understand the consequences of PEN better we simulated data for an ideal experiment of non-dividing cells. Therefore, synthesis and degradation were simulated for a range of half-lives between 2 – 100 hours applying the model previously used to fit our measurements. Subsequently, this simulated data was subjected to PEN and again fitted to the same model. In this way we were able to compare true half-lives to half-lives after PEN. In the equilibrium case, i.e. when synthesis was equal to degradation, PEN correction led to an underestimation of half-lives by 6.1 % for a true half-life of 2 hours up to 15.2 % for a true half-life of 100 hours. Since in proliferating cells synthesis is expected to outweigh degradation, we repeated the simulation reducing synthesis half-lives by a defined factor towards the degradation half-life. As illustrated in Figure 7E, the smaller the ratio between synthesis and degradation half-life was, the more degradation half-lives were underestimated after PEN. Since the factor by which synthesis and degradation half-lives varied was not readily predictable for a certain protein, our simulation could not be used as a correction function. However, as visible in Figure 7C, half-lives derived from PEN-corrected data was well within the range and equally distributed as half-lives derived from uncorrected data. In summary our simulation showed that PEN leads to a modest underestimation of protein half-lives, which did not seem to distort the overall distribution of half-lives in a detrimental way.

PROTEIN HALF-LIVES ON RNA

To find out if RNA-binding affected protein stability we first compared half-lives within the total proteome of MCF7 cells. Figure 8A illustrates that half-lives of RNA-binding proteins in the total proteome did not differ to the overall half-lives of all proteins in the total proteome. Since for a number of RNA-binding proteins it is known that they are produced in excess over the complex they participate in¹⁶⁰, we hypothesized that the excess of free protein over RNA-bound protein might be masking a general trend among total proteome half-lives of RNA-binding proteins (Figure 8B). The following paragraph describes how XRNAX was used to measure protein half-lives on RNA directly and how these half-lives compared to the ones found for the total proteome.

DETERMINING PROTEIN HALF-LIVES USING XRNAX AND TMT-SILAC

In order to determine half-lives of proteins bound to RNA we combined our TMT-SILAC strategy with XRNAX. Replicating the experimental setup for the determination of total

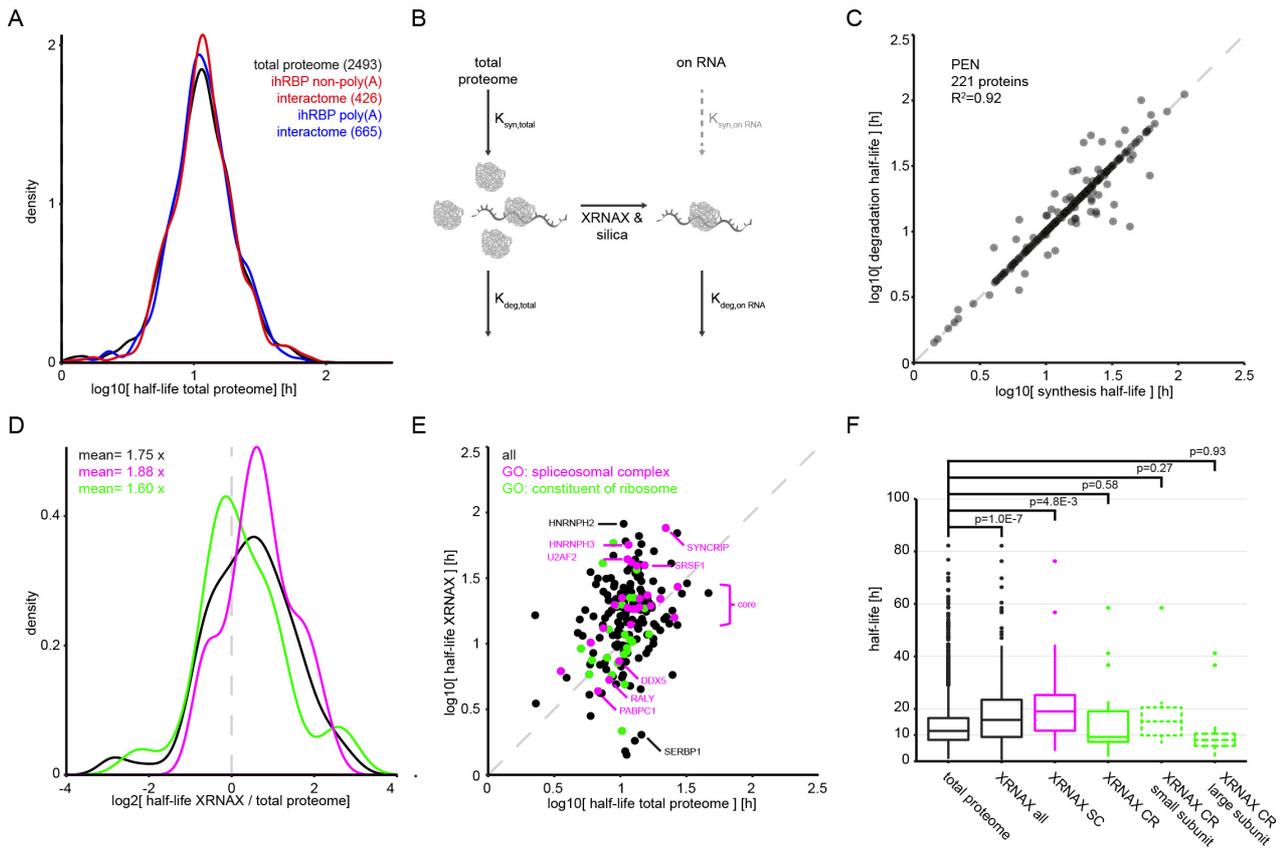


Figure 8: Protein half-lives within the total proteome and within the RNA-bound proteome.

- A) Density plot comparing protein half-lives within the total proteome of MCF7 cells. Compared are all available half-lives (total proteome) to the half-lives available for the RNA-binding subproteomes defined in Figure 3.
- B) Schematic representation of protein turnover in the total proteome or bound to RNA. The fraction of protein bound to RNA is a subset of the total proteome. XRNAX can be used to extract this subset and determine the half-lives within it.
- C) Scatter plot comparing protein synthesis and degradation half-lives for the RNA-bound subproteome of MCF7 cells. Each point represents one protein. R^2 value gives the Spearman correlation. Half-lives were derived using PEN as presented in Figure 7, however, after extraction of the RNA-bound proteome with XRNAX.
- D) Density plot comparing foldchanges of MCF7 protein half-lives between the RNA-bound proteome (XRNAX) and the total proteome. Components of the spliceosome (magenta) or the ribosome (green) are highlighted. Mean gives the average foldchange within these groups.
- E) Scatter plot comparing protein half-lives of the MCF7 total proteome to protein half-lives of the RNA-bound proteome (XRNAX). Each point represents one protein. Coloring is identical to D.
- F) Boxplots comparing MCF7 protein half-lives between different RNA-complexes. XRNAX denotes to half-lives determined for the RNA-bound proteome. All: All available half-lives; SC: spliceosomal complex; CR: cytosolic ribosome. Testing occurred towards all available half-lives for the total proteome using the Wilcoxon ranksum test.

proteome half-lives, we took seven timepoints within 32 hours after switching MCF7 cells from one SILAC media to the other (two replicates were produced with label swap). Subsequently, cells were UV-crosslinked, protein-crosslinked RNA extracted with XRNAX and further purified with silica enrichment as described above. After RNase and trypsin digestion identical amounts of peptides for each timepoint were TMT-labeled according to Figure 7A. Using PEN we derived a total of 308 synthesis and degradation half-lives 221 of which were overlapping (Figure 8C). Notably, XRNAX SILAC-TMT data was much noisier than the total proteome dataset so that without PEN only 48 degradation half-lives could be derived. This again demonstrated the powerful normalization capabilities of PEN.

COMPARING PROTEIN HALF-LIVES ON RNA TO THE TOTAL PROTEOME

A direct comparison of protein half-lives in the total proteome to the ones derived with XRNAX showed that they were in fact different. Half-lives of proteins on RNA were on average increased by a factor of 1.75 (Figure 8D), confirming our initial hypothesis that RNA-binding increases protein stability (Wilcox ranksum test $p=1.0E-7$). Many nuclear RNA-binding proteins of the HNRNP class exhibited substantial stabilization. In the most extreme case the half-life of HNRNP2 was increased almost eight-fold from 10.5 in the total proteome to 82.1 hours when bound to RNA (Figure 8E). Interestingly, we found some cases of destabilization on RNA as well, including the protein SERBP1, which is the human orthologue of the ribosome hibernation factor Stm1 in yeast.

We turned our attention towards two well-described protein-RNA complexes, the spliceosome and the ribosome. The half-lives of spliceosomal proteins were overall stabilized by a factor of 1.88, which was stronger than average (Figure 8D). Moreover, half-lives of most core spliceosome components were focused around a common half-life on RNA of approximately 20 hours, whereas more peripheral components had either shorter or longer half-lives (Figure 8E). While ribosomal proteins were on average stabilized on RNA, their half-lives span over more than one order of magnitude (Figure 8D). This observation was surprising for a complex, which is generally believed to be produced at a specific stoichiometry and degraded by subunit (i.e. 40S and 60S)¹⁶¹. Interestingly, when looking at the protein components of each ribosomal subunit separately, stabilization primarily occurred for proteins in the small subunit but not in the large subunit (Figure 8F).

In summary we were able to confirm that on average proteins in protein-RNA complexes were more stable. However, as illustrated in the case of the ribosome, this was not inevitably true for all proteins in a protein-RNA complex. Moreover, the question arose how it was possible that proteins in a well-defined complex such as the ribosome could have such a range of half-lives, assuming the complex is usually degraded in bulk.

PROTEIN HALF-LIVES IN RIBOSOMAL ASSEMBLIES

Our analysis of protein half-lives had indicated a curious variance within the half-lives of ribosomal proteins on RNA. To find out more about what governs protein stability within protein-RNA complexes the ribosome appeared as an intriguing model. That was especially

true because established and refined methodology for the purification of ribosomal complexes, i.e. polysome profiling, was readily available. Since XRNAX extracted all RNA, the half-lives we had derived for ribosomal proteins were most likely a superposition of half-lives of ribosomal proteins in different assemblies. We decided to make use of polysome profiling, in order to deconvolute this superposition and find out if half-lives differed between those assemblies. The following work was done in collaboration with Matilde Bertolini from the laboratory of Bernd Bukau (ZMBH Heidelberg), who performed the polysome profiling experiments, from which the following data was derived.

DETERMINING PROTEIN HALF-LIVES USING POLYSOME PROFILING AND TMT-SILAC

Sucrose density gradient centrifugation is a well-established method for the purification of ribosomes and their subcomplexes. For this purpose, cell lysates are loaded on a sucrose cushion and subjected to several hours of ultracentrifugation. The sucrose cushion is prepared into a gradient of 5-45 % sucrose, into which components of the lysate migrate and align according to their own density. After centrifugation the solution can be aspirated from the top of the tube and be put into fractions. Here, UV-absorbance can help to detect ribosomal RNA in the fractions and be the guide for identifying known ribosomal assemblies.¹⁶² As visible in Figure 9, this process called polysome profiling very reproducibly allowed for preparations of the 40S subunit, 60S subunit, the combined assembly of the two subunits into the 80S ribosome and a polysome fraction. The polysome fraction contains two or more 80S ribosomes assembled on one mRNA molecule. As several ribosomes only assemble on an actively transcribed transcript, the polysome fraction contains translating 80S ribosomes. The 80S fraction also contains translating 80S ribosomes on mRNA, so-called monosomes, but also empty 80S ribosomes. In yeast it was shown that the amount of empty ribosomes can be surprisingly high so that the majority of the ribosomes in the 80S fraction did in fact not carry mRNA.^{163,164} This is surprising because 80S ribosomes are generally believed to assemble from 40S and 60S subunits on mRNA only upon translation initiation and disassemble again after translation termination (for review see¹²⁹). So at present, it is unclear how these empty ribosomes assemble and what their purpose is. What is clear is that in yeast¹⁶⁵ and human cells¹⁶⁶ stress leading to translational arrest increases the pool of empty 80S assemblies, indicating that they could be derivatives of previously translating ribosomes. This is supported by the observation that puromycin collapses the polysome fraction, leading to an accumulation of 80S ribosomes in the 80S fraction, which do not disassemble further into small and large subunits.¹⁶⁶

Before determining half-lives of proteins in polysome profiling fractions we wanted to make sure that the method purified ribosomal proteins at reproducible amounts. Therefore, we performed polysome profiling with MCF7 lysates in triplicates. Figure 9B illustrates that using a modification of the SP3 protein cleanup procedure¹⁶⁷ and label free quantification (iBAQ¹⁰²) we were able to quantify ribosomal proteins for each polysome fraction in a highly reproducible way. Moreover, many non-ribosomal proteins were reproducibly detected in each fraction. With a variance of 10 % or less between replicates, proteins totaled to 1604 in

the 40S fraction, 1700 in the 60S fraction, 1841 in the 80S fraction and 2158 in the polysome fraction. Interestingly, variances for ribosomal proteins were the smallest in the 80S and polysome fractions, indicating that the stoichiometry of proteins here was especially similar between samples. A direct comparison between the fractions showed the expected allocation of ribosomal protein abundances (Figure 9C). The 40S fraction contained primarily small subunit proteins, whereas contamination of large subunit proteins had overall two orders of magnitude lower abundances. The 60S fraction contained at the foremost large subunit proteins with a somewhat stronger contamination of small subunit proteins. The 80S and polysome fractions had about equal abundances for small and large subunit proteins. Interestingly, while most proteins in the fractions had very similar abundances, each fraction had a number of significant outliers. We note here that label-free quantification with iBAQ is an intensity-based procedure, which does not use any absolute reference, i.e. a defined amount of a synthetic peptide standard. However, using such a standard it was recently shown in mouse cells that ribosomes can contain substoichiometric amounts of certain ribosomal proteins¹⁶⁸, generally referred to as 'ribosome heterogeneity'. The two most prominent examples in our analysis, which occurred reproducibly substoichiometric in the 40S, 80S and polysome fraction were RPS12 and RPS21. We tested if there were any changes in the stoichiometry between the 80S and polysome fraction and found only two cases in the small subunit, which showed significant changes between the two assemblies (Figure 9D). In summary our analysis showed that SP3 purification led to highly reproducible quantification of ribosomal proteins in polysome fractions. We proceeded in determining protein half-lives from polysome profiling fractions using the same TMT-SILAC approach as earlier. In order to increase the resolution of our analysis we collected nine timepoints within 32 hours after switching SILAC media. Applying PEN, we derived half-lives for 1486 proteins from the 40S fraction, 967 proteins from the 60S fraction, 575 proteins from the 80S fraction and 694 proteins from the polysome fraction. Among all fractions half-lives for 95 cytosolic ribosomal proteins could be determined.

COMPARING HALF-LIVES OF RIBOSOMAL PROTEINS IN POLYSOME FRACTIONS TO THE TOTAL PROTEOME

To our great interest and as illustrated in Figure 10A half-lives of ribosomal proteins differed greatly between the fractions. Especially ribosomal proteins in the 80S and polysome fraction were strongly stabilized in comparison to the total proteome, but also in comparison to the 40S and 60S fractions. While the median half-life of all ribosomal proteins in the total proteome was 10.1 hours, it was 26.9 or 23.8 hours in the 80S or polysome fractions, respectively. In the polysome fraction stabilization was stronger for ribosomal proteins of the large subunit in comparison to the small subunit, suggesting subunit-specific turnover. Minor stabilization in the subunit fractions was only significant for small subunit proteins in the 60S fraction. As small subunit proteins were assumed to be a contamination in the 60S fraction we speculate that this stabilization was artifactually created by bleed-over from the heavily stabilized 80S fraction. Interestingly, large subunit proteins in the 40S fraction were

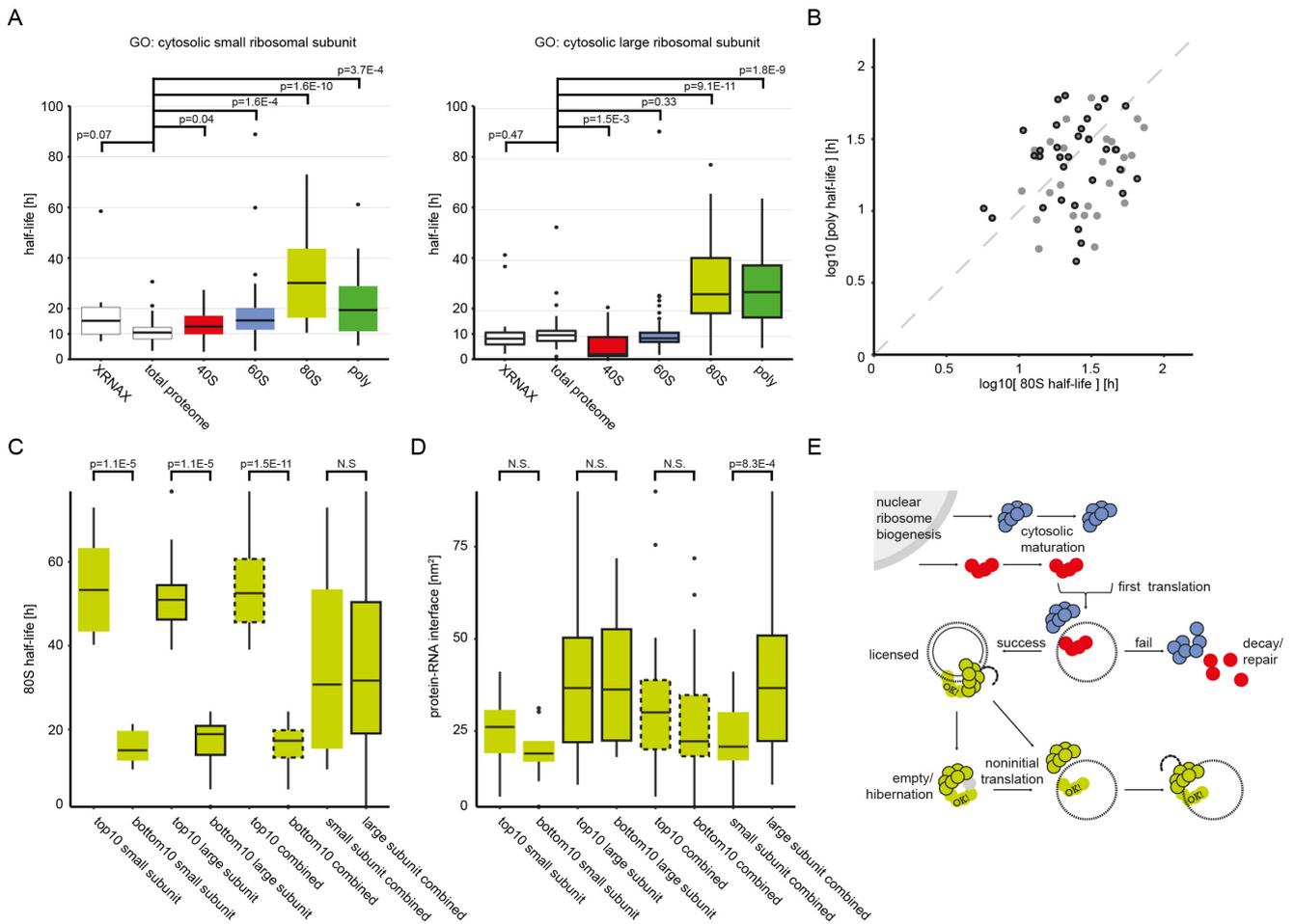


Figure 10: Comparison of protein half-lives between different ribosomal assemblies.

Color coding refers to the polysome profiling fractions introduced in Figure 9A. Proteins of the small ribosomal subunit are represented by graphical elements without margin, proteins of the large subunit with margin.

A) Boxplots comparing half-lives of ribosomal proteins between fractions of MCF7 polysome profiling. Half-lives were derived using PEN as presented in Figure 7, however, after separation of ribosomal complexes into the fractions displayed in Figure 9A. For details refer to Materials and Methods. For comparison half-lives of ribosomal proteins within the total proteome and within the RNA-bound subproteome (XRNAX) are also displayed.

B) Scatterplot comparing half-lives of ribosomal proteins within the 80S and polysome fraction.

C) Boxplots comparing the 10 most and least stable ribosomal proteins within the 80S fraction. Testing occurred with a Wilcoxon ranksum test.

D) Boxplots comparing the interface size between each of the proteins in C and ribosomal RNA. The Cryo-EM structure of the human 80S ribosome was used to compute the area between ribosomal proteins and the ribosomal RNA of each subunit. For details see Materials and Methods. Testing occurred with a Wilcoxon ranksum test.

E) Proposed model for the licensing of 80S ribosomes. After ribosome biogenesis free small and large ribosomal subunits reside in the cytosol. In case a nascent small and large subunit assembles into an 80S ribosome, which engages in translation successfully, the 80S will not dissociate into its subunits after translation but will remain in the 80S conformation. Keeping functional pairs of small and large subunits together prevents them from mixing with potentially dysfunctional nascent subunits.

significantly destabilized in comparison to the total proteome. This suggested that 60S proteins were actively removed from free small ribosomal subunits. In conclusion, half-lives of ribosomal proteins appeared to be heavily regulated depending on the assembly they participated in. This explained our initial findings for RNA-bound ribosomal proteins derived by XRNAX, which were a superposition of half-lives with contributions from all ribosomal complexes (Figure 8F).

A direct comparison of protein half-lives showed that the stabilization effect did not result from the fact that different groups of ribosomal proteins were compared between the total proteome and each fraction (Figure S10A-B). Again, the 40S and 60S fractions showed a very similar distribution of half-lives as the same proteins in the total proteome. Interestingly, the destabilization of large subunit proteins in the 40S fraction only seemed to affect a certain subset, whereas another subset seemed unaffected (Figure S10A). Proteins in the 80S and polysome fraction were almost uniformly stabilized, although, to different degrees. As visible in Figure 10A and Figure S10C&D, this was concomitant with an increased variance of half-lives within the 80S and polysome fractions. Comparing half-lives of the same ribosomal proteins between the total proteome, the 80S and the polysome fraction, the variance from the mean half-life in the total proteome was 25 %, in the 80S fraction 65 % and in the polysome fraction 42 %. We compared half-lives of ribosomal proteins in the 80S and polysome fraction directly and again found overall proteins more stable in the 80S fraction (Figure 10B). Which protein was more stable depended on the individual case, yet, proteins of the small subunit were more often more stable in the 80S fraction (70 % of small subunit proteins and 55 % of large subunit proteins were more stable in the 80S). As mentioned above and also visible in Figure S10D, this subunit effect was caused by the polysome fraction, where proteins of the small subunit were stabilized to a lesser degree than large subunit proteins. Interesting here was that this was not a uniform effect within the subunit but affected some proteins more than others. Importantly, half-lives of small subunit proteins in the 40S fraction and half-lives of large subunit proteins in the 60S fraction were not only shorter but also, they were in no apparent correlation with half-lives of proteins in 80S complexes (Figure S10E&F). That means half-lives of ribosomal proteins in the 40S and 60S assemblies were not scaled by a certain factor in comparison to 80S assemblies. Rather, some half-lives stayed the same, while others were stabilized when the ribosome was assembled from its subunits.

Since RNA is the component of the ribosome, which determines much of its structural integrity¹¹, we were interested in how far protein-RNA interaction influenced protein half-lives within the stable complex. We selected the ten most and least stable proteins from each subunit of the 80S fraction (Figure 10C) and determined the size of the interface between each protein and the RNA in its subunit from the cryo-EM structure of the human ribosome (Figure 10D). While there was no correlation between protein half-lives and RNA-interface size, there was a significant difference of RNA interface size between proteins of the large and small subunit. Both observations did not suggest that half-lives within the ribosomal subunits were governed by their interactions with RNA, but rather pointed towards an active and selective mechanism for protein exchange.

Our findings showed strong stabilization of ribosomal proteins in 80S assemblies, confirming our hypothesis that proteins are more stable as part of a protein-RNA complexes. Antithetically, this was not the case for the individual subunits in the 40S and 60S fractions. As the ribosome is a giant assembly, which follows an intricate line of biogenesis it was hard to say, which one of the two effects was the rule and which the exception. Furthermore, the question arose when is a protein-RNA complex seen as such by the cell? Is it only considered a complex when it is assembled into the 80S form and functional, i.e. is function the key to stability? The observation that the actual interface with RNA did not seem to influence stability and the fact that free small and large subunits were not stabilized at all, despite being defined protein-RNA complexes, pointed towards functionality as determining factor. This conclusion led us to derive a model for ribosome homeostasis presented in Figure 10E, which contains the subversive inference that empty ribosomes formed a pool of approved, translation-competent ribosomal subunits. This pool was distinct from free small and large ribosomal subunits, which had just emerged from the biogenesis pipeline and had never served in the act of translation. The model will be further discussed below.

EXCHANGE OF RIBOSOMAL PROTEINS FROM TRANSLATION-COMPETENT RIBOSOMES

Our analysis allowed for unseen insights into the homeostasis of the ribosome and revealed some surprising detail. Current models for ribosome homeostasis assume that ribosomes are degraded by subunit through autophagy (for review see¹⁶¹). Early indications for this were found in electron micrographs, which showed ribosomes among the major cargo of autophagosome (for review see¹⁶⁹). In yeast it was shown that starvation led to the targeted destruction of ribosomes through autophagy¹⁷⁰, which coined the term 'ribophagy' (for review see¹⁷¹). This effect was also seen in human cells, again under starvation or other stresses, whereas the autophagic flux for ribosomes under normal conditions was low.^{172,173} In *C. elegans* autophagic degradation of ribosomes was shown to occur during normal growth and was found important for nucleotide homeostasis.¹⁷⁴ Thus, on the one hand it is clear that autophagy can lead to very significant degradation of ribosomes under stress, although, its contribution to ribosomal protein turnover under normal conditions has not been quantified in a comprehensive way. On the other hand, in human cells it was shown that proteasome inhibition leads to accumulation of ribosomal proteins in the nucleus.¹⁷⁵ Moreover, in yeast it was demonstrated that an excess of ectopically expressed ribosomal protein is degraded by the proteasome.¹⁶⁰ The assumption here is that during ribosome biogenesis ribosomal proteins are produced in excess and shuttle in and out of the nucleus, while the production of rRNA is the rate-limiting factor that controls biogenesis. An excess of ribosomal proteins is then degraded in the cytosol by the proteasome.¹⁷⁵ In how far excess ribosomal proteins can exchange from functional cytosolic ribosomes is unclear, as well as if selective extraction of ribosomal proteins from the complex exists. In an *in vitro* experiment with *E.coli* ribosomal proteins it was shown that damaged ribosomes can exchange protein components with additionally added isotope-labeled protein to become functional again.¹⁷⁶ Conceptually, *in vivo* the ubiquitin-proteasome system would be more apt to this task because it degrades one protein at a time, unlike autophagy, which selectively engulfs part of the cytosol leading to

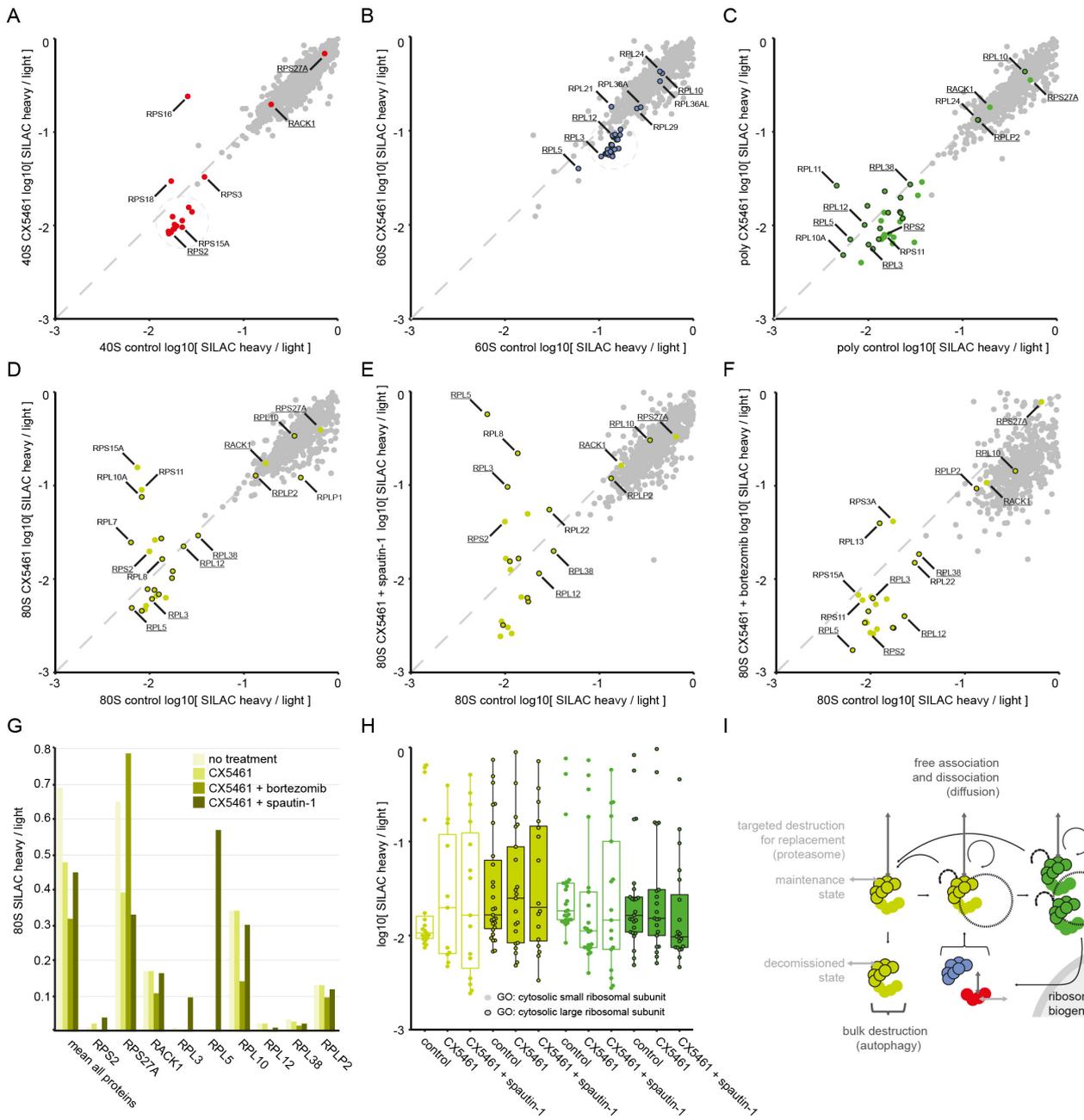
significant bystander-flux¹⁷². In conclusion, from the current state of the literature it was unclear if autophagy or the proteasome were the main contributor to turnover of ribosomal proteins from ribosomal complexes.

We were especially interested in the question, why half-lives in 80S assemblies could show such a wide range, when they were assumed to be produced and destroyed in bulk. As the existence of any 80S assembly is the outcome of a series of complex, previous events, measuring exchange of proteins from 80S ribosomes *in vivo* is not trivial. From all we know, the existence of an 80S ribosome requires that it was assembled on an mRNA during at least one successful round of translation initiation. From an experimental view this is further complicated by the fact that ribosomes are continuously assembled in the nucleus to have their biogenesis completed in the cytosol (for review see¹⁷⁷). Consequently, biogenesis and any putative exchange from already functional ribosomal complexes cannot be easily distinguished. To experimentally explore this, we decided to separate the two processes by inhibiting ribosome biogenesis. As mentioned earlier, the rate-limiting step in ribosome biogenesis is rRNA transcription.¹⁷⁵ rRNA is transcribed by RNA polymerase I, which can be inhibited with the small molecule drug CX5461 (Figure S4F).¹⁷⁸ We inhibited ribosome biogenesis in light SILAC MCF7 cells during six hours of CX5461 exposure in order to make sure that all pre-ribosomes in the biogenesis pipeline had exited. Then we switched to heavy media and continued treatment with the inhibitor for twelve hours. Any protein produced from this point on carried a heavy label and could not have been subject to ribosome biogenesis. As a control we used the same cells but switched to heavy media without continuing the inhibition with CX5461. Any protein produced in these control cells carried a heavy label and was subject to normal ribosome biogenesis and cytosolic turnover. Again, we used polysome profiling in order to resolve the different assembly states of the ribosomes and quantified their protein content using our optimized protocol. Notably, we refrained from measuring half-lives for this experimental setup, because inhibition of ribosome biogenesis by CX5461 is known to arrest cell growth¹⁷⁸. Therefore, we did not expect a meaningful comparison between our sample and the control regarding protein half-lives and decided to quantify relative protein amounts between the two only at one distinct timepoint (12 hours). Figure 11A-D displays the effect of CX5461-mediated RNA polymerase I inhibition on the amounts of new protein incorporated into each polysome-profiling fraction. In the 40S fraction the ratio of new over old protein for most ribosomal proteins was a single digit percentage (Figure 11A, grey circle). Expectedly, this ratio was somewhat higher for the control, where ribosome biogenesis was uninhibited. A very similar offset, representing more new protein in the control, was observed in the 60S fraction (Figure 10B, grey circle). Notably, here the amount of new protein was overall ten times higher than in the 40S fraction. As the stoichiometry seemed highly similar between most of those new proteins and was dependent on CX5461 inhibition we assumed that this was residual ribosome biogenesis. Interestingly, however, a number of proteins showed a very distinct deviation from this pattern, and appeared at much higher amounts than this background. These could be categorized according to their behavior towards rRNA transcription inhibition. The first group showed very high amounts of new protein incorporated into each of the

Figure 11: Exchange of ribosomal proteins from ribosomal complexes under inhibition of ribosome biogenesis.

Color coding refers to the polysome profiling fractions introduced in Figure 9A. Proteins of the small ribosomal subunit are represented by graphical elements without margin, proteins of the large subunit with margin. Underlined proteins were recurrently quantified in all experiments. For experimental details see text or refer to Material and Methods section.

- A)-D) Effects of ribosome biogenesis inhibition on the incorporation of new protein into ribosomal assemblies. rRNA transcription was inhibited in SILAC light MCF7 cells with CX5461 for 6 hours, before switching to heavy SILAC media containing no inhibitor (control) or CX5461 for 12 hours. Scatterplots compare the relative amount of new protein incorporated into 40S, 60S, 80S or polysome fractions, when ribosome biogenesis is active (control) or inhibited (CX5461).
- E) Combined effect of ribosome biogenesis and autophagy inhibition on the incorporation of new protein into the 80S fraction. Autophagy was inhibited in MCF7 cells using spautin-1 and the same experiment as described in A-D was performed. The scatterplot compares the relative amount of new protein incorporated into the 80S fraction, when ribosome biogenesis and autophagy are active (control) or inhibited (CX5461 + spautin-1).
- F) Combined effect of ribosome biogenesis and proteasome inhibition on the incorporation of new protein into the 80S fraction. The proteasome was inhibited in MCF7 cells using bortezomib and the same experiment as described in A-D was performed. The scatterplot compares the relative amount of new protein incorporated into the 80S fraction, when ribosome biogenesis and the proteasome are active (control) or inhibited (CX5461 + bortezomib).
- G) Bargraph displaying the relative amount of new protein incorporated into the 80S fraction for proteins recurrently quantified in all experiments. Shading refers to the inhibitors used.
- H) Boxplots comparing the relative amount of new ribosomal protein incorporated into the 80S (yellow) or polysome fraction (green). Small and large subunit proteins are displayed in separate.
- I) Proposed model for the turnover of ribosomal proteins. Licensed 80S ribosomes are either actively translating in monosomes/polysomes, or stored in an inactive state (empty 80S ribosomes). Within the inactive state ribosomal proteins can be extracted by the proteasome (horizontal arrow) for repair of the assembly (targeted extraction for replacement). Old or dysfunctional 80S assemblies are destroyed in bulk through autophagy. A small number of regulatory ribosomal components exchange freely from the subunits, without the need for extraction (vertical arrow).



assemblies, at similar ratios as the background of non-ribosomal proteins and independent of CX5461. These proteins, namely RPS27A, RACK1, RPL10 and RPLP2, were readily exchanged from the subunits. Notably, this was also the case for the 80S and polysome fraction. In the latter two fractions a similar behavior was also shown by RPL38, however, at a lower rate. The next group showed increased incorporation of new protein dependent on the persistent inhibition of ribosome biogenesis. In the 40S fraction continued CX5461 inhibition increased incorporation of new RPS16 and RPS18 protein (Figure 11A). In the 80S fraction the incorporation of a whole number of proteins was strongly increased by continued CX5461 treatment (Figure 11D). Notably, these were different to the ones changing in the 40S fraction. For example, there was strong exchange of RPS15A and moderate exchange of RPS2 from the 80S assembly, whereas no increased exchange for these proteins was detected in the 40S fraction (Figure 11A&D). In the 60S fraction no inhibition-dependent exchange was observed. In the polysome fraction RPL11 showed the strongest exchange in the presence of CX5461. Overall the exchange of proteins was not as strong here as in the 80S fraction, however, markedly stronger than in the control. Proteins strongly exchanging from the 80S fraction did not exchange from the polysome fraction (e.g. RPS11, RPL10A, RPS2). In conclusion the experiments showed a curious finding, i.e. upon inhibition of ribosome biogenesis, there was increased exchange of specific ribosomal proteins from 80S assemblies. The fact that for some proteins exchange from 80S ribosomes was markedly increased in comparison to cells where ribosome biogenesis was intact indicated that this exchange might have occurred especially on old 80S complexes, which usually get replaced by new ones.

A comparison between the previously assessed protein half-lives in polysome-profiling fractions and the incorporation of new protein into those fractions upon inhibition of ribosome biogenesis did not show any correlation (Figure S11A). This confirmed our initial assumption that expression of ribosomal proteins in freely proliferating cells or cells that underwent six hours of CX5461 treatment, respectively, was very different, so that protein half-lives from uninhibited cells must not necessarily inform about protein stability in cells where ribosome biogenesis was inhibited.

We wanted to understand what protein degradation mechanisms were involved in the exchange. Therefore, we combined ribosome biogenesis inhibition with the inhibition of either the proteasome through bortezomib or autophagy through spautin-1. As illustrated in Figure 11E under CX5461 and spautin-1 inhibition exchange of ribosomal proteins from the 80S fraction was even stronger than observed before. Interestingly, different protein components were exchanged as compared to the case where autophagy was uninhibited (Figure 11D). For example, the proteins RPL3 and RPL5 were unaffected in the uninhibited case, but strongly exchanged upon spautin-1 treatment. *Vice versa*, the proteins RPS15A, RPS11, RPS10A and RPL7 were only detected in their heavy form in the 80S fraction of cells that had not been treated with spautin-1. The proteins RPS2 and RPL8 were exchanged in both cases, yet, more extreme when autophagy was inhibited. Treatment with bortezomib reversed the effect we had initially observed, so that there was no increased exchange for RPS15A, RPS11 and RPS2 anymore after inhibition of the proteasome (Figure 11F). Figure 11G summarizes the findings for the 80S fractions and the proteins with the most significant

changes recurrently detected in all experiments. For the strongly exchanging group RPS27A, RACK1, RPL10 and RPLP2 inhibiting degradation mostly did not change their behavior, except for inhibition of the proteasome, where RPS27A exchanged even stronger. The only protein with a low percentage exchange rate, which was unaffected by any treatment was RPL38. RPL12 exchanged with a low percentage unaffected by ribosome biogenesis, but showed much decreased exchange upon proteasome inhibition. RPS2, RPL3 and RPL5 exchanged strongly under autophagy inhibition, however, much weaker under inhibition of the proteasome. Direct comparison of proteasome to autophagy inhibition showed that it was primarily these three proteins showing the largest differences (Figure S11B). Moreover, three other strongly exchanging proteins emerged, which had previously escaped our analysis because they did not show sufficient heavy signal in the control sample. These proteins – RPS3A, RPS26 and RPS15 – showed very significant exchange upon either proteasome or autophagy inhibition from the 80S assembly.

In the polysome fraction inhibition of autophagy led to an increased exchange of RPS3A and RPS26 (Figure S11C). In order to understand if the exchange of proteins from the 80S and the polysome fraction differed, we compared them directly within the identical treatment regimen (Figure S11D-F). Without any treatment after six hours of ribosome biogenesis inhibition the two fractions looked very similar in terms of new protein being incorporated into them (Figure S11C). If inhibition of ribosome biogenesis was continued, there was a stronger exchange of small and large subunit components from the 80S fraction (Figure S11E). The same was true if ribosome biogenesis was impaired along with autophagy, however, three strongly exchanging proteins emerged again – RPS3A, RPS26 and RPS15 (Figure S11F). These three proteins exchanged at intriguingly similar rates from both fractions. Importantly, in the 40S fraction heavy RPS15 was not detected and RPS3A exchanged to a very different degree (Figure S11G, right).

In summary our experiments showed that there was significant exchange of ribosomal proteins from mature ribosomes. Figure 11H summarizes this observation for 80S ribosomes in the 80S and the polysome fractions, where persistent inhibition of ribosome biogenesis heavily increased exchange of proteins, which was even more increased when autophagy was blocked in addition. Proteins subject to this exchange could be categorized using the degree to which it occurred and the protein degradation machinery involved. Our findings were integrated into the model presented in Figure 11I. Whereas details to this model will be discussed below, we will summarize the key findings from Figure 11, which led to it. Three ways for the turnover of ribosomal proteins were identified, the first of which was free exchange from the complex as observed in all experiments for the proteins RPS27A, RACK1, RPL10, RPLP2 and to a lesser degree RPL38. The second one was proteasomal degradation, whose inhibition had prevented exchange of a number of proteins from the 80S fraction (Figure 11D&F). The third one was bulk destruction of 80S assemblies through autophagy. Inhibition of autophagy had led to increased exchange of ribosomal proteins from 80S assemblies (Figure 11E), indicating that old ribosomes had accumulated and components had been exchanged, when under conditions of functional autophagy these 80S assemblies would have been destroyed in bulk. Comparing 80S assemblies between the 80S and polysome

fraction under inhibition of autophagy had shown remarkably similar exchange rates for the three proteins RPS3A, RPS26 and RPS15. This suggested that the pool of translating 80S assemblies in the polysome fraction had been replenished by assemblies of the 80S fraction. Additionally, there was a group of proteins, which only had exchanged from the 80S fraction (Figure S11D & E), but not from the polysome fraction. The members of this group and the degree to which they exchanged depended on the functionality of autophagy, however, the exchange pattern was not passed on to the polysome fraction. This suggested the existence of a second population of 80S ribosomes, which had lost their ability to replenish the pool of translating 80S ribosomes in the polysome fraction. We distinguished these two populations as a maintenance state, which exchanges components for repair, and a decommissioned state, which exchanges components in preparation or on the way to destruction. While we had observed a passing-on of an exchange pattern between the 80S fraction and the polysome fraction (Figure S11D-F), we had seen no such propagation between the 40S and the 80S, or the 60S and the 80S or *vice versa* (Figure S11G&H). Remarkably, this indicated that most of the translating ribosomes in the polysome fraction had re-initiated translation within their polysome or had been replenished by empty 80S ribosomes of the 80S fraction and not from free 40S or 60S assemblies.

COMPARING HALF-LIVES OF NON-RIBOSOMAL PROTEINS IN POLYSOME FRACTIONS TO THE TOTAL PROTEOME

The emerging concept of 'ribosome heterogeneity' has fostered efforts to define an extended ribo-interactome, i.e. non-ribosomal proteins, which become part of the ribosomal assemblies in order to direct or regulate their behavior (for review see¹⁷⁹). Two recent publications have used orthogonal methods for the identification of ribosome interactors. The first approach was to express ribosomal proteins carrying an affinity-tag in order to immunoprecipitate the entire 80S assembly from mouse embryonic stem cells. Interactors of the assembly were co-immunoprecipitated and detected by MS, which identified approximately 400 proteins.¹⁸⁰ The second approach was to perform polysome profiling in human cells, in order to find proteins with similar abundance profiles between the polysome fractions, which resembled the ones observed for ribosomal proteins themselves.¹⁸¹ This identified 145 interactors of polysomes, many of which had been known before, however, 84 of which were new.

Comparing protein half-lives from polysome profiling fractions to the total proteome, we had noticed a strong stabilization of ribosomal proteins in the 80S and polysome fraction (Figure 10). Following the notion that proteins in a functional complex are more stable⁹⁸, we anticipated that protein half-lives could add another dimension to the classification of proteins interacting with ribosomal assemblies. For the first dimension of separation we used deep total proteomic data from MCF7 cells as a background in order to calculate the relative enrichment of proteins within each of the polysome fractions (Figure 12). This clearly highlighted ribosomal proteins already, however, also separated some highly enriched proteins from the background. As a second dimension of separation we added how much the half-life of a protein was stabilized in comparison to its half-life in the total proteome.

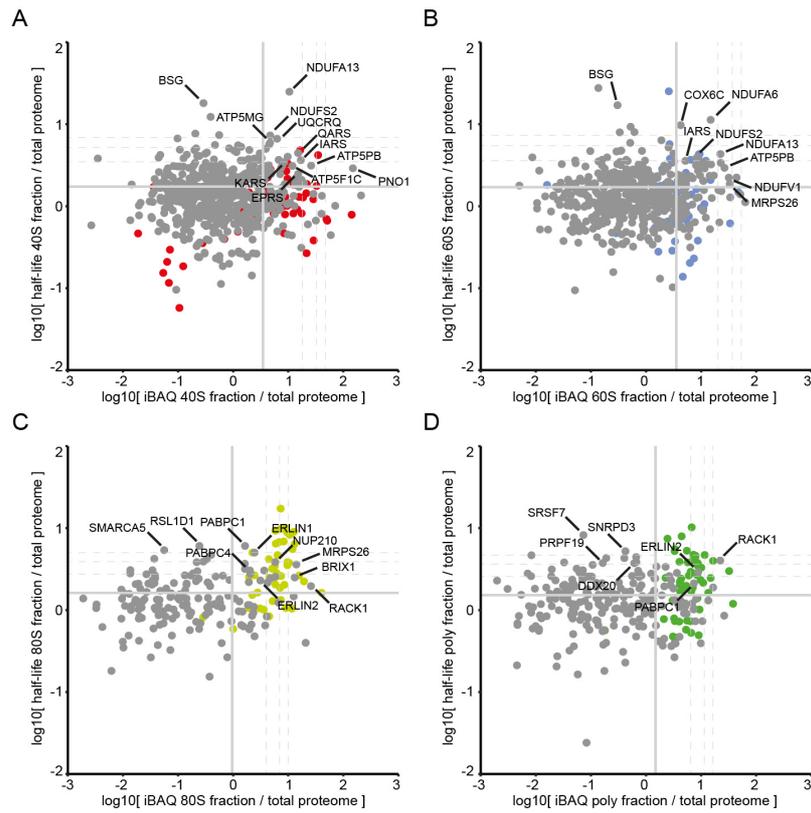


Figure 12: Double enrichment for the classification of novel ribosome interactors in MCF7 cells.

Color coding refers to the polysome profiling fractions introduced in Figure 9A. Dots in grey represent non-ribosomal proteins.

A) Scatterplot comparing ratios of protein abundance against ratios of protein stability. To test for increased abundance the iBAQ ratios between the 40S fraction and a deep MCF7 total proteome were computed. In advance, iBAQ intensities for each experiment were normalized to their combined sum. To test for increased stability the ratios of half-lives were computed for proteins in the 40S fraction compared to the total proteome (see Figure S10). Bold lines mark the mean of all points, dashed lines mark 1,2 and 3 standard deviations from the mean.

B)-C) Same as in A for the 60S, 80S and polysome fraction. For details see text.

Interestingly, this identified a small but defined set of proteins for each fraction, which showed similar behavior to the control of ribosomal proteins. Proteins enriched and stabilized in the 40S and 60S fractions were especially amino-acyl tRNA ligases, which are known interactors of the ribosome (Figure 12A&B).^{179,181} Interestingly, these proteins were enriched in the 40S and 60S, but not the 80S and polysome fractions. Furthermore, proteins especially enriched and stabilized in the 40S and 60S fraction were mitochondrial components involved in oxidative phosphorylation, which likely were contaminations of co-migrating mitochondria in our polysome profiling experiments. Still, it was interesting to see for those proteins, too, that being part of a functional complex, their half-lives were strongly increased. Remarkably, in both the 40S and 60S fractions, one of the most severely stabilized proteins was Basigin, which is known to get stabilized very significantly after an initial phase, where the protein is very unstable.⁹⁸ Furthermore we found PNO1 enriched in the 40S fraction, which has been reported to mediate methylation of the 18S rRNA¹⁸². In the 80S fraction a number of proteins associated with translation were heavily stabilized along with proteins involved in ribosome biogenesis, i.e. RSL1D1, BRIX1 and SMARCA5. However, since the 90S pre-ribosome is a precursor potentially contaminating the 80S fraction¹⁷⁷, we could not infer with certainty that these biogenesis factors associated with 80S ribosomes. Other proteins showing strong stabilization along with enrichment were the mitochondrial ribosomal protein MRPS26 and the nuclear pore complex component NUP210. The proteins PABPC1 and PABPC4, which are known to bind poly(A) tails of mRNA, were strongly stabilized and enriched along with proteins of the ERLIN complex in both the 80S and polysome fraction. ERLIN2 has been linked to cell cycle progression and is often amplified in breast cancer¹⁸³, which could link translational control to cell cycle control. Notably, neither one of the above-mentioned studies had reported ERLINs within their ribo-interactome.^{180,181} Beside the ERLIN complex we found a strongly stabilized group of proteins associated with splicing in the polysome fraction, which replicated findings of a recent report in HeLa cells¹⁸⁴.

Our analysis showed that half-life data for polysome profiling fractions could add another dimension of separation in the classification of potential ribosome interactors. While we only focused on proteins, which were enriched over the total proteome along with being stabilized in their half-lives, we suspect many of the heavily stabilized proteins, which occurred substoichiometrically compared to the total proteome also to be true interactors of ribosomal assemblies. However, the strong double-enrichment for example for ERLIN proteins implicated a direct and important biological function, which future studies should be dedicated to.

DISCUSSION

XRNAX ALLOWS FOR A PLETHORA OF NOVEL APPLICATIONS IN RNA

BIOLOGY

In the past UV-crosslinking has proven an invaluable tool for the study of protein-RNA interactions *in vitro* and *in vivo*.⁴² With XRNAX we present a new way of applying this tool with maximal force. We demonstrate that XRNAX extracts contained a distinct subset of the proteome, which was remarkably diverse and not dominated by any abundant class of RNA-binding protein. Surprisingly, the RNA content of XRNAX extracts was more diverse than that of TRIZOL extracted total RNA. This indicated that XRNAX enriched protein-RNA complexes in a way that allowed for greater proteomic and transcriptomic sequencing depth, thereby making the detection of low abundant protein-RNA interactions possible.

We used protein-crosslinked RNA extracted through XRNAX for proteomic and transcriptomic applications. These applications would have been impossible using previous methods because of their limitation to polyadenylated RNA, their inability to scale and other complications arising from the impurity of crude lysates. During the development of these applications XRNAX revealed some inherent features that underlined its general usefulness in studying RNA biology. Specifically, isolating nucleotide-crosslinked peptides proved its scalability, when XRNAX easily produced milligrams of RNA, from which RNA-crosslinked peptides could be isolated and fractionated for MS. In comparison, interactome capture, apart from being limited to polyadenylated RNA, does not scale easily because of its dependence on magnetic beads. Our experiments on RNA-crosslinked peptides showed that current standard MS is capable of detecting hybrids that carry up to three nucleotides. While in this study we used these peptides only to locate protein-RNA interfaces in a protein-centric way, we are confident that in the future XRNAX will become key technology for simultaneous sequencing of peptides and RNA from UV-crosslinked peptide-RNA hybrids. As purification of RNA-peptide hybrids and MS detection methods will advance, XRNAX will give easy access to starting material, which previously was not readily available. Next, XRNAX proved highly reproducible even in challenging applications such as the quantification of RNA-binding in a time course experiment. Despite stringent filtering the measurements quantified the association of almost 800 proteins with RNA during the response to arsenite. Additionally, XRNAX delivered robust and reproducible results when applied to different cell lines. While MCF7, HEK293 and HeLa cells did have slightly different RNA-binding proteomes, methodologically there was no problem in capturing all of them. Lastly, owing to the initial organic phase extraction and subsequent washing steps XRNAX delivered extracts that were very pure, i.e. only contained protein and RNA. Only this degree of purity allowed us to generate small RNA fragments by sonication for our CLIP-seq experiments. Furthermore, it allowed us to perform enzymatic reactions on the RNA content of XRNAX extracts. For example, we were able to label RNA in XRNAX extracts using biotinylated ribonucleotides and poly(U) polymerase (data not shown). In another application we enriched the lincRNA MALAT1 using biotinylated oligonucleotide probes from XRNAX extracts and identified its protein interaction partners using MS (data not

shown) – an application usually greatly impaired by genomic DNA and other cellular debris in crude lysates. Furthermore, our proof-of-concept experiments for PEPseq demonstrated that the purity of XRNAX extracts offers the entry point for advanced applications in chemical biology on protein-crosslinked RNA. These examples illustrated that XRNAX extracts could be used for a variety of applications, only a fraction of which we explored in this thesis. XRNAX and the above-mentioned OOPS⁸¹ are methodologically similar, however, differ in the product they produce. OOPS relies on RNase digestion to release protein crosslinked to RNA from the interphase of the TRIZOL extraction. Therefore, OOPS never presents protein-crosslinked RNA as its own entity, which ultimately makes it less versatile. However, for its proteomic applications OOPS refrains from additional silica enrichment and therefore is able to use smaller amounts of starting material (10 instead of 50 million cells). In principle, OOPS could be combined with XRNAX simply by adding two more rounds of TRIZOL extraction to the XRNAX protocol. At any rate, both protocols have proven to be a valuable addition to the toolkit, which interrogates protein-RNA interactions.

Today's challenge in RNA biology is to reconcile proteomic and transcriptomic data in a meaningful way. We demonstrated here that XRNAX is the versatile, reproducible and scalable fundament onto which future applications can build.

FUNCTIONS OF NON-CODING RNA ARISING FROM THE INTEGRATED HUMAN RNA-BINDING PROTEOME

We used XRNAX and silica enrichment to derive RNA-binding proteomes from MCF7, HEK293 and HeLa cells. Among the novel RNA binders we unexpectedly found enrichment for chromatin remodelling proteins, especially proteins carrying bromodomains. Many of those proteins, like BRD4, are heavily investigated due to their implication in cancer¹⁸⁵. Only recently it was shown that the bromodomain of BRD4 and other bromodomain proteins bind enhancer RNA (eRNA), recruiting them to enhancers¹⁸⁶. Other chromatin modifying complexes such as PRC2, CoREST or SMCX have been found to interact with a large number of lincRNAs¹⁸⁷, which in some cases were shown to modify their behaviour as described in the Introduction. In the future it will be important to quantify the interaction of chromatin modifying proteins with their cognate ncRNA in order to understand their combined function. We have shown that XRNAX can quantify the association of proteins with RNA including any known kind of ncRNA (and especially including non-polyadenylated ncRNA) thereby paving the way for a better understanding of genomic regulation. Remarkably in this context, we identified a cluster of proteins involved in the DNA-damage response, some of which had not been described as RNA-binding before. Two of those proteins, MDC1 and TP53BP1 are reportedly involved in an RNA-dependent DNA damage pathway⁹⁴. We also found the direct interactor of TP53BP1, RIF1, which is recruited to DNA-double strand breaks (DSBs) upon ATM-dependent phosphorylation of TP53BP1. Both proteins have been shown to promote non-homologous end joining. We discovered BRCA1 as RNA-binding, too, which promotes homologous recombination and expels RIF1 from DSB sites¹⁸⁸. Several recent studies have

found nascent transcripts involved in transcript-templated DNA repair^{95,189,190}, however, mechanistic details remain elusive. Our findings pinpoint, which proteins involved in DNA repair bind RNA and might make for promising targets in upcoming studies.

DIFFERENTIAL QUANTIFICATION OF PROTEIN-RNA INTERACTIONS PROVIDES UNSEEN MECHANISTIC INSIGHT

XRNAX coupled to silica enrichment allowed us to quantify RNA-binding differentially in MCF7 cells. This way we were able to follow the consequences of arsenite-induced translational arrest on the level of protein-RNA interactions. Interestingly, overall RNA-binding remained unchanged, however, there were some proteins that did show significant changes. TP53BP1 was the protein that increased its association with RNA the most after 30 minutes. As mentioned above the protein has been described to participate in an RNA-dependent repair mechanism for DSBs.⁹⁴ Arsenite has been reported to induce DSBs comparable to γ -irradiation at the concentrations used in our experiments¹⁹¹. RNA binding of TP53BP1 increased steadily over time suggesting that it followed an accumulation of DSBs and increased recruitment of the repair machinery.

Among the proteins with the most prominent decrease in RNA-binding was the stress granule marker USP10. Arsenite has been reported to induce oxidative stress, which USP10 counteracts by an unknown mechanism.¹⁹² In order to do so, USP10 needs to leave its direct interactor and inhibitor G3BP1. Our data showed that upon arsenite stress RNA binding of G3BP1 remained unaltered, whereas RNA binding of USP10 was decreased. This suggested that RNA might serve as the regulatory platform that brings the two proteins together, so that upon arsenite stress USP10 left RNA, resulting in reduced inhibition by G3BP1 and the opportunity for USP10 to exert its antioxidative effect. Notably, USP10 is next to USP13 one of the two known targets of the autophagy inhibitor spautin-1, which prevents deubiquitination of BECN1, resulting in its degradation thereby limiting induction of autophagy¹³⁷. By serendipity we discovered an autophagy-dependent degradation mechanism of the translational machinery upon arsenite-induced translational arrest, which could be inhibited by spautin-1. In yeast it has been shown that starvation can lead to a form of autophagy that eliminates the cytosolic ribosome in a process termed 'ribophagy'¹⁷⁰. Interestingly, this process was dependent on Ubp3 and Bre5 – the yeast homologues of USP10 and G3BP1. Deng et al. had shown that USP10 is part of a feed-forward loop during the induction of AMP-activated protein kinase (AMPK)¹⁹³, which can be induced by arsenite stress¹⁹⁴. As AMPK has been shown to induce translational arrest as well as autophagy (for review see¹⁹⁵), the axis between G3BP1, USP10 and AMPK could link stress granule formation and the autophagic process we observed. We propose a model in which arsenite stress releases USP10 from RNA, putatively by AMPK phosphorylation, which made it more accessible and could trigger the feedforward loop described by Deng et al.. In how far the above-mentioned antioxidative effect of USP10 was a result or an independent effect of USP10-amplified AMPK activation remains subject to future studies.

As one well-established effect of arsenite stress is translational arrest we were especially interested in the the RNA-binding behaviour of the translation initiation machinery during this process. During canonical translation initiation the EIF2 complex assembles into a ternary complex with guanosine triphosphate (GTP) and the initiator methionine transfer RNA (tRNA^{met}). This complex combines with the 40S ribosomal subunit to form the 43S preinitiation complex, which combines with the cap-binding complex on an mRNA to form the 48S preinitiation complex and scans along the mRNA in search for the initiator AUG codon.¹²⁹ The current model for translational arrest upon arsenite stress is based on the observation that EIF2 gets phosphorylated, which inhibits formation of the ternary complex, thereby inhibiting translation initiation. Unresolved, however, remains the observation that arsenite-induced stress granules accumulate all the components of 48S preinitiation complexes, although conceptually formation of 48S preinitiation complexes (and even the 43S preinitiation complex for that that matter) should be inhibited through phosphorylation of EIF2.¹²³ Nevertheless, current state of the literatures has it that arsenite-induced stress granules contain stalled 48S preinitiation complexes (for review see¹⁹⁶). This suggests that the 48S preinitiation complexes that accumulated on mRNA upon arsenite stress included EIF2. Consequently, translation initiation must have been inhibited downstream of 48S assembly, which could have happened on the level of EIF5. EIF5 activates the GTPase activity of EIF2 after recognition of the start codon, leading to its departure and joining of the 60S ribosomal subunit.¹²⁹ EIF5 is absent from stress granules, what had led Buchan and Parker to speculate that this might be how 48S preinitiation complexes could retain their stalled state in them.^{123,197} Interestingly, in our data the EIF2 component EIF2- β increased RNA-binding, which indeed suggested an additional block of translation initiation downstream of EIF2 phosphorylation.

In summary, quantifying RNA-binding recapitulated many known phenomena during translational arrest, but also revealed surprising details, which were previously unknown. Within the analysis it became apparent that the core strengths of our approach were its prior impartiality and that it was able to follow binding kinetics of proteins over time.

STABILIZATION OF PROTEIN IN PROTEIN-RNA COMPLEXES IS CONTEXT-DEPENDENT

Our comparative analysis of half-lives had revealed that proteins in protein-RNA complexes were overall more stable (Figure 8). This confirmed the observation by McShane et al., who had correlated proteins in complexes with higher stability. More specifically, McShane et al. had found protein degradation kinetics for proteins in complexes, which would be best explained by a two-state model, describing fast decay when the protein is young and slow decay when the protein has aged. Our data added experimental evidence to this correlative finding, because we found increased half-lives within some well-defined protein-RNA complexes such as the spliceosome. *Vice versa*, looking at all spliceosomal proteins in a cell, we found their half-life on average only half as long as assembled on RNA. While on average

this observation was true, i.e. on average the half-life of protein on RNA was 1.75 fold higher than in the total proteome (Figure 8D), we found some interesting variation within our data. For example, the half-life of the RNA-binding protein SERBP1 was found very much decreased on RNA from 14 hours in the total proteome to 2 hours on RNA. The yeast orthologue of SERBP1, Stf1, has been reported to sequester non-translating 80S ribosomes upon stress-induced translation termination, thereby keeping them in a state of so-called hibernation.^{198,199} Structurally, SERBP1 is known to act as a clamp fixing the 40S and 60S subunits together. In fact, the first cryo-EM structure of the human 80S ribosome was recorded stabilized through SERBP1.²⁰⁰ More recently, a cryo-EM study of ribosomes from rabbit reticulocytes showed that SERBP1-clamped ribosomes make up for the vast majority of non-translating ribosomes and actually the majority of all ribosomes in this cell type. We derived half-lives for SERBP1 in all fractions of our polysome profiling data except for the polysome fraction. Interestingly, in the 80S fraction its half-life was also heavily decreased to 4 hours from 14 hours in the total proteome. In yeast it has been reported that the majority of SERBP1 is associated with 80S ribosomes.¹⁹⁹ This could explain why the destabilization of SERBP1 in the 80S fraction recapitulated our findings for the protein on RNA in general (as assessed by XRNAX), because if interacting with RNA the protein was presumably clamping 80S ribosomes. Moreover, its instability on RNA suggests that SERBP1 might be removed as a clamp by degradation. The example of SERBP1 as an RNA-binding protein, which was destabilized on RNA, illustrated well how ribosomal RNA introduced exceptions to the rule that RNA-binding leads to increased protein stability. Moreover, it gave another interesting insight into the elusive share of empty or hibernating ribosomes within the 80S fraction, which will be further discussed in the following.

Half-lives on RNA, which we had derived through XRNAX, had shown for most proteins very robust stabilization. However, for ribosomal proteins half-lives varied seemingly at random or at least not predictive of the defined complex they were supposed to participate in. This was explained by our polysome profiling experiments, where only the 80S and polysome fractions showed stabilization of protein half-lives. Ribosomal proteins in the 40S and 60S fractions were not stabilized, so that their superposition created the inconsistent pattern observed in our XRNAX data. This observation along with the observation that the actual size of the protein-RNA interface did not inform about protein stability within the 80S complex (Figure 10C&D), led us to the conclusion that it had to be functionality, which led proteins to be more stable in the 80S complexes. Functionality in the realm of the ribosome means the ability to successfully undergo translation initiation, followed by protein biosynthesis and translation termination.¹²⁹ Translation initiation requires the small subunit to assemble with a number of translation initiation factors on the start codon of an mRNA. Only subsequent to this, the large subunit is joined and translation ensues. Anyway, prerequisite to translation initiation is that the large and small subunits are separated. That means after translation termination, the 80S assembly should fall apart and its subunits should join the pool of free subunits in the 40S and 60S fractions. Our half-life data for ribosomal proteins in the 40S and 60S fractions did not comply with this model, because the stabilization of proteins in the 80S or polysome fraction was not transferred to the 40S and 60S fraction. If after translation

termination 80S assemblies fell apart, their small subunit would end up in the 40S fraction and their large subunit in the 60S fraction. This would raise the average half-life of ribosomal proteins in those fractions, too. However, half-lives we observed for the 60S fraction were basically identical to the total proteome and half-lives in the 40S fraction only slightly increased (Figure 10A). As visible from the UV-traces in our polysome profiling experiments, whose area under the curve quantifies RNA, the amount of ribosomal subunits in the 80S or polysome fraction was very similar to the amounts of ribosomal subunits in the 40S or 60S fractions (Figure 9A). Consequently, the contribution of small and large subunits from terminating 80S assemblies would have been expected to lift the average half-lives in these fractions very significantly, if those assemblies fell apart and mixed with the 40S and 60S fraction. This suggested that – although translation initiation requires small and large subunit to be separate – two pools of subunits existed. Conceptually, it makes sense that the cell would retain subunits, which have proven functionality by successfully completing translation, as a distinct pool from subunits, which have just emerged from ribosome biogenesis and are potentially dysfunctional. Ribosome biogenesis is a well-concerted and heavily regulated process, which in dividing cells has a massive energy expenditure (for review see^{177,201}). However, to our knowledge the current literature does not report a quality control mechanism for ribosomes, which checks their functionality. Our data seemed to suggest a separation between ribosomal subunits, which had proven to be functional, and subunits which had not yet engaged in translation. Furthermore, this separation was defined by different assembly states, separating 80S from 40S and 60S subunits, which offered an intriguing role for empty 80S ribosomes. We integrated our model in the scheme presented in Figure 10E. Here, functionality of new ribosomal subunits is tested in action: If 40S and 60S subunits successfully engage in translation they remain engaged in order to signal their functionality. This creates a pool of ‘licensed’ 80S assemblies, which have proven to be functional and are kept apart from new and free subunits by virtue of their engagement. Translation initiation requires that an 80S ribosome is assembled on mRNA from a small and large subunit (for review see¹²⁹). On the basis of our model this means that new 40S and 60S subunits have to go through at least one successful round of translation initiation and 80S assembly in order to get stabilized. This makes sense in terms of quality control, where new ribosomal assemblies have to prove their functionality. As soon as functionality was proven, i.e. the 80S ribosome has successfully initiated translation, the assembly gets stabilized. This concept could also assign a role to empty, SERBP1-clamped 80S assemblies because a way to flag small and large subunits, which have successfully engaged in translation, would be to clamp them together so that they can be distinguished from free subunits, which have not yet proven functionality. This hypothesis is supported by polysome profiling in yeast¹⁶⁶ and mice²⁰², where puromycin collapses the polysome fraction, however, not to the benefit of the 40S and 60S fraction but the 80S fraction. In fact, the 40S and 60S fractions stay unchanged upon puromycin treatment, whereas there is a tremendous increase in ribosomal subunits in the 80S fraction. Puromycin is incorporated into nascent peptides, thereby leading to chain termination and premature translation termination.²⁰³ This suggests that after translation 80S ribosomes are indeed not split but remain in their 80S conformation, where they are stored as empty 80S ribosomes. As

puromycin-induced translation termination is premature the accumulation of 80S ribosomes could be an artefact of this specific drug. However, another inhibitor of protein biosynthesis, harringtonine, which does not interfere with translation elongation and leads to production of full-length protein and normal translation termination, has the same effect. Harringtonine stalls ribosomes after translation initiation, so that they cannot enter elongation and are stalled on translation start sites. Thereby subsequent rounds of translation initiation and protein production from the mRNA transcript are blocked, so that effectively no translation is going on and all ribosomes which are not stalled in initiation complexes have undergone natural translation termination as their last action.^{204–206} Polysome profiling in human²⁰⁷ or mouse²⁰⁸ cells shows that harringtonine treatment also leads to collapse of the polysome fraction into the 80S fraction, again without any increase in the 40S or 60S fractions. In yeast it has been shown that glucose starvation leads to collapse of the polysome fraction into the 80S fraction, which then primarily contains empty 80S ribosomes.¹⁶⁴ The same study demonstrates that the same ribosomes, which collapsed into the 80S fraction upon starvation, are recycled into the polysome fraction after starvation when glucose is added again. Interestingly, this recycling required the Dom34-Hbs1 complex, which splits the 80S ribosomes in order for them to re-engage in translation. This mechanistic detail is remarkable in the context of our proposed model, where an open question was in how far splitting of the 80S complex was required in order for translation to occur again. The findings in yeast suggest that, indeed, splitting is required and a specialized mechanism for this purpose exists. The mammalian homologue of Dom34 is the protein PELO (also known as pelota). Knockout of Peló in mice is embryonic lethal, however, very recently a conditional Peló knockout mouse for epidermal stem cells was described.²⁰⁹ The authors find significant changes in epidermal translation after the knockout was induced. Interestingly, their polysome profiling data suggests that Peló knockout cells did not have a pool of empty ribosomes, but only translating 80S ribosomes and polysomes. Consequently, they find their translation was heavily increased and that overall the epidermis was not viable and became scaly. This indicated that Peló was necessary to create a pool of resting empty 80S ribosomes and if it did not, 80S ribosomes were stuck in continuous translation, reading into 3' UTRs where they could not be removed. Interestingly, induction of autophagy through rapamycin partially rescued the phenotype, indicating that an alternative way for removing 80S assemblies from mRNA was ribophagy. Notably, this was different to data in yeast, where Dom34 knockout heavily increases the pool of empty 80S ribosomes.¹⁶⁴ Albeit, in yeast Dom34 knockout was not lethal, indicating that there might be differences for mammals in the exact mechanism by which 80S ribosomes are taken off the mRNA after translation termination.

Measuring half-lives of proteins on RNA gave experimental evidence to the paradigm that proteins in complexes are more stable. Seemingly, an exception to this rule was the ribosome, where simply looking at half-lives of ribosomal proteins on RNA with XRNAS showed no clear pattern for stabilization. However, an in-depth analysis of half-lives of ribosomal proteins through polysome profiling suggested that complex formation was necessary but not sufficient for proteins to become stabilized in a complex. Free, small and large ribosomal subunits were not stabilized whereas the combined 80S complex was. This suggested that a

paradigm describing stability of proteins in complexes might need an amendment introducing yet another requirement. In the case of ribosomal proteins, what seemed necessary and sufficient in order to become stabilized, was becoming part of a protein-RNA complex and demonstrating function. As the ribosome is probably one of the most complicated protein-RNA assemblies in the cell, this amendment might be unique to its case. Nevertheless, it will be interesting to see in future studies if functionality is also required for stabilization of other protein-RNA complexes such as the spliceosome.

RIBOSOMAL ASSEMBLIES ARE MORE DYNAMIC THAN PREVIOUSLY POSTULATED

Examining protein half-lives in different ribosomal assemblies had revealed an overall strong stabilization for 80S ribosomes. However, within the 80S complex half-lives differed almost one order of magnitude (Figure 10). Starting from the assumption that the assembly of an 80S ribosome and its destruction were quantized, this was impossible to explain. What could have been explained under this assumption were differences in half-lives between the subunits, because mechanistically each round of translation requires assembly and disassembly of physically separate subunits, thereby allowing for different turnover times between the two. While that seemed to be the case for ribosomal proteins in the polysome fraction, where the small subunit had an overall faster turnover than the large subunit, this was not the case for the 80S fraction, where both small and large subunit proteins were equally stabilized. Nevertheless, for proteins of either subunit, small or large, the intra-subunit variability of half-lives within the 80S assembly was similarly large (Figure 10A). The same was true for proteins in the polysome fraction. Considering the fact that 80S ribosomes are assembled from 40S and 60S subunits, this variability could have arisen from variability of half-lives within these smaller subcomplexes. In this case one would expect the same pattern of variability when comparing half-lives between the 40S or 60S and the 80S assemblies. For example, the most stable protein within the 40S fraction should be among the most stable ones in the 80S fraction and the same should be the case for the least stable protein. This kind of propagation of half-lives was not observed so that the half-life of a protein in the 40S or 60S fraction was not informative for its half-life in the 80S or polysome fraction (Figure S10E&F). This suggested that our starting assumption must have been flawed. This could have been the case in two ways, either the assumption was wrong that the 80S ribosome was always assembled from the same proteins, or that proteins in the 80S ribosome were not always destroyed in bulk. There were indications for both possibilities in the literature. The entire concept of ‘ribosome heterogeneity’ is built on the hypothesis that ribosomes can assemble into complexes of different stoichiometry, thereby allowing for specialization of ribosomes (for review see¹⁷⁹). In fact, Shi et. al have shown for a very limited number of ribosomal proteins in mouse embryonic stem cells that they were represented substoichiometrically in polysomes.¹⁶⁸ Similar observations were made in yeast.²¹⁰ Our label-free quantification of absolute abundances for ribosomal proteins in polysome profiling fractions had indicated an analogous

finding for a whole number of proteins (Figure 9C). However, we did not use spike-in peptide standards for our quantification and therefore could not exclude that this might have been an artefact of iBAQ quantification. At any rate, the total number of proteins that showed a clear stoichiometric deviation within ribosomes of the 80S or polysome fraction was small. This meant that most of the variance of protein stability within 80S assemblies was caused by differences in half-lives between proteins that were assembled in perfect stoichiometry. The other possibility for how our initial assumption could have been flawed was that proteins in 80S ribosomes were not necessarily destroyed in bulk, but that they could be exchanged from the translation-competent assembly, i.e. even after ribosome biogenesis. Conceptually, this was the more powerful hypothesis because it could explain both, sub-stoichiometric occurrence of ribosomal proteins and differences in turnover within the 80S assembly. For ribosomes purified from *E.coli* it has been shown that chemically damaged ribosomes can be rendered functional again when incubated with functional ribosomal protein *in vitro*.¹⁷⁶ This showed that in principle exchange was possible and that this exchange could serve a repair function. We mention here another, more recent study, which claimed to show protein turnover from assembled ribosomes *in vivo* using deuterium labelling in mice in a timecourse of 32 days.¹⁷⁴ Ribosomal assemblies were purified from liver tissue in bulk during 12 hours of sucrose density centrifugation. Notably, this should also purify all pre-ribosomal intermediates. The study is problematic for a number of reasons, the most pressing of which is its ignorance towards ribosome biogenesis. While its claims are far-reaching, apparently demonstrating the 'mechanism of *in vivo* maintenance' of ribosomes, we fail to extract any information on the mechanism or information on the maintenance of ribosomes from the presented data. As a detailed review of this report would go beyond the scope of this discussion we mention it here only for completeness. By inhibiting ribosome biogenesis we could show that there was in fact exchange of ribosomal proteins from translation-competent ribosomes in the 80S and polysome fractions, which could be amplified by inhibition of autophagy and reduced by inhibition of the proteasome. This indicated that exchange of ribosomal proteins from 80S assemblies was an active process driven by the proteasome. The amplification caused through inhibition of autophagy was probably caused by the accumulation of old 80S assemblies, which usually would have been subject to ribophagy. Interestingly, in MCF7 cells either inhibition of autophagy or inhibition of the proteasome significantly increased the size of the 80S fraction (data not shown), indicating that both might be necessary for the efficient disassembly and degradation of 80S ribosomes. Therefore, it is possible that the proteasome fulfils two roles in respect to 80S assemblies, one of which is disassembly in preparation for autophagic destruction. The other could be targeted extraction of proteins, which are subsequently exchanged, in order to maintain damaged ribosomes. This was supported by the increased exchange of ribosomal protein when autophagy was inhibited, presumably leading to the accumulation of old ribosomes, which were more likely damaged.

We found a small group of proteins – RPS27A, RPL10, RPLP2 and RACK1 – showing very strong exchange independent of ribosome biogenesis. RPL10 is an essential component of the ribosome, and its knockdown diminishes the assembly of 80S complexes.²¹¹ In eukaryotic

ribosome biogenesis RPL10 is only added to the large subunit in the cytoplasm during the very late steps of cytosolic maturation. Studies in yeast indicate that the protein is essential in the so-called 60S 'test-drive', where the fidelity of the complex is tested before biogenesis factors are released and the large subunit allowed to join with the small subunit.^{212,213} Our data showed frequent exchange of RPL10, suggesting that the 60S 'test-drive' could be a regular check-up, which is performed not only once during ribosome biogenesis but regularly between rounds of translation. RPS27A is transcribed from a fusion gene, representing one of the three genes in the human genome encoding ubiquitin. Consequently, RPS27A is a ribosomal protein that is born with its own ubiquitination. Interestingly, the most extreme exchange of RPS27A was under inhibition of the proteasome. We speculate that this was because higher amounts of RPS27A accumulated when the proteasome was inhibited, so that the protein exchanged more frequently. Interestingly, in yeast expression of RPS27A or its yeast paralogue RPS27B influence cells' budding behaviour, indicating that the protein has regulatory functions, too.²¹⁴ For RACK1 many functions have been reported dependent or independent of the ribosomes. RACK1 is a constitutive component of the small ribosomal subunit, that has been described as signalling hub for the regulation of translation (for review see²¹⁵). Its exchange was not as extreme as that seen for RPL10 or RPS27A but similar to that of RPLP2. RPLP2 is part of the so-called ribosomal stalk, another signalling centre of the ribosome mainly responsible for interacting with translation-associated GTPases (for review see²¹⁶) and binding site of some infamous AB toxins such as Ricin²¹⁷ or Shiga²¹⁸ toxin. The ribosomal stalk is made up from the proteins RPLP0, RPLP1 and RPLP2 of which we detected RPLP2 in all experiments and RPLP1 in only some, however, always strongly exchanging. Of all the ribosomal stalk proteins only RPLP0 interacts with rRNA, whereas RPLP1 and RPLP2 form heteromers that only interact with the 60S subunit via RPLP0.²¹⁶ Interestingly, because of their unique acidic features biochemical purification of RPLP1 and RPLP2 was established very early so that *in vivo* exchange of the two proteins from cytosolic ribosomes could be demonstrated in rat liver through radioactive pulse-chase experiments.²¹⁹ In summary these findings indicated that ribosomal proteins strongly exchanging from 80S assemblies fulfilled regulatory roles involved with functional control of the ribosome.

Apart from the small group of strongly exchanging proteins there was weaker but significant exchange of other proteins from both ribosomal subunits, especially if ribosome biogenesis was blocked (Figure 11H). The pattern of protein exchange was much more similar between the 80S and polysome fraction than between the 40S or 60S fraction and the polysome (Figure S11D-H), adding evidence to our model that empty 80S assemblies formed a reservoir of translation-approved ribosomes, which would rather engage in translation and become part of the polysome fraction, than nascent subunits in the 40S and 60S fractions (Figure 10E). Despite the similarity it is hard to say if protein exchange only occurred on empty 80S ribosomes, which were then recycled into the translating pool, or if exchange also occurred on actually translating ribosomes. However, exchange of proteins was markedly stronger from the 80S fraction. In HeLa cells it has been shown that the 80S fraction mainly contains empty 80S ribosomes, yet, not during mitosis, where the 80S fraction only contains 80S assemblies on mRNA.²²⁰ Interestingly, in this study pool size of 80S assemblies on mRNA seems constant,

so that the difference between dividing cells and non-synchronized cells in the 80S fraction is the addition of a large amount of empty 80S ribosomes. Consequently, this large amount of empty 80S ribosomes is apparently not essential for survival of a cell. As ribosome biogenesis is energetically very expensive (for review see²⁰¹) ribosomes themselves store a considerable amount of energy. We speculate here that another purpose of empty 80S ribosomes, besides providing a functional pool of translation-competent ribosomes, might be energy, amino acid and nucleotide storage. This store can be used up in times of massive energy and material consumption, such as cell division. This hypothesis is supported by findings in *C.elegans*, where autophagic degradation of rRNA was found essential to maintain nucleotide-homeostasis during growth.²²¹ Along this line, an important connection might be the one between empty 80S ribosomes and autophagy. Interestingly, another drug apart from puromycin and harringtonine interfering with translation and leading to the collapse of the polysome fraction into the 80S fraction in HeLa cells is arsenite²²². We have shown that arsenite induces ribophagy in human cells (Figure 4), suggesting that autophagy might act on empty 80S ribosomes (Figure 11I).

Our experiments propose an intriguing model, which could close the long-postulated connection between the ribosome – translation – autophagy – and energy homeostasis. This connection has found support by a lot of circumstantial evidence, however, could so far not be integrated into a unified model (for review see²²³). Future studies will have to show what happens if the pool of empty 80S ribosomes is manipulated. For example, an open issue is if arsenite-induced ribophagy specifically degrades the pool of empty 80S ribosomes created upon translational arrest. Furthermore, the question arises if induction of autophagy under these circumstances is a consequence of translational arrest or of the tremendously increased pool of empty 80S ribosomes. In this context it might be helpful to elucidate the effect of other substances like puromycin or harringtonine, which have been shown to collapse polysomes into 80S ribosomes. It will be interesting to see if these compounds trigger ribophagy as well and if yes, what population of ribosomes will be affected.

MATERIALS AND METHODS

The first part of the following methods section is adapted from the publication Trendel et al., *The Human RNA-binding Proteome and Its Dynamics During Translational Arrest, Cell, 2018*⁸². In the latter publication data produced from these methods and the methods themselves were originally presented in a peer-reviewed format. This data is presented again in the first chapter of the Results section of this thesis, i.e. “XRNAX As Platform for Interrogating Protein-RNA Interactions”. The second part of the method section has not been published and describes additional methods used in the second chapter of the Results section of this thesis, i.e. “Comparing Protein Half-Lives Inside and Outside of Protein-RNA Complexes”.

METHODS FOR CHAPTER 1

RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Lamin B1 rabbit polyclonal antibody	proteintech	12987-1-AP, RRID:AB_2136290
EXOSC2 rabbit polyclonal antibody	proteintech	14805-1-AP, RRID:AB_2101837
EXOSC2 mouse monoclonal antibody	proteintech	66099-1-Ig
Goat anti-mouse Cy5	abcam	ab6563, RRID:AB_955068
Anti-FLAG M2 Magnetic Beads	Sigma	M8823, RRID:AB_2637089
HA-tag antibody	proteintech	51064-2-AP, RRID:AB_11042321
Chemicals, Peptides, and Recombinant Proteins		
Dialysed FBS	Gibco	26400-044
Pen-Strep	Gibco	15140-122
DMEM for SILAC	Silantes	280001300
SILAC heavy L-lysine (¹³ C ₆ , ¹⁵ N ₂ -L-Lysine HCl)	Silantes	211604102
SILAC heavy L-arginine (¹³ C ₆ , ¹⁵ N ₄ -L-Arginine HCl)	Silantes	201604102
GlutaMAX	Gibco	35050061
TRI reagent	Sigma	T9424
GlycoBlue	Ambion	AM9515
4-thiouridine (4SU)	biomol	Cay-16373
SYBRSafe	invitrogen	S33102
EDTA-free protease inhibitor	Sigma (Roche)	11873580001
Benzonase	Novagen	70664
SP3 beads	GE	44152105050250
Trypsin/LysC	Promega	V5073
NEB DNase buffer 10 x	NEB	B0303S
NEB DNase	NEB	M0303L

RNASin Plus RNase inhibitor	Promega	N2615
RNase A	Thermo	EN0531
RNase I	Ambion	AM2295
RNase T1	Thermo	EN0541
Sodium arsenite (50 mM solution in water)	Santa Cruz	sc-301816
Spautin-1	Sigma	SML0440
Rapamycin (2.5 mg/ml in DMSO)	Sigma	R8781
Ethynyl-uridine (EU)	Jena Biosciences	CLK-N002
RNA polymerase I inhibitor CX5461	Millipore	509265
sulfo-Cy5-azide	Jena Biosciences	CLK-AZ118
HOECHST33342 (20 mM in water)	Thermo	62249
ProlongGold antifade mountant	invitrogen	P36934
Proteinase K	Thermo	EO0491
FastAP Thermosensitive Alkaline Phosphatase	Thermo	EF0651
T4 Polynucleotide Kinase	Thermo	EK0032
Protein G Magnetic Beads	Pierce	88847
Lipofectamine 3000	invitrogen	L3000008
Turbo DNase	Ambion	AM2238
Critical Commercial Assays		
Quiagen RNeasy Midi Kit	Quiagen	75144
NEXTflex Small RNA Sequencing Kit	Bioo Scientific	NOVA-5132
Deposited Data		
RNA sequencing data	EMBL-EBI ENA	PRJEB26441
XRNAX CLIP-seq data	EMBL-EBI ENA	PRJEB26442
Proteomics data	ProteomeXchange	PXD010520
Human proteome (search term: 'reviewed:yes AND proteome:up000005640', 20216 entries, retrieved 11 September 2017)	Uniprot	UP000005640
Human reference genome hg19	GENCODE	Release 19 (GRCh37.p13)
Human reference genome hg38	GENCODE	Release 29 (GRCh38.p12)
HeLa poly(A)-binding proteome	78	NA
HEK293 poly(A)-binding proteome	65	NA
MCF7 poly(A)-binding proteome	80	NA
Deep MCF7, HeLa, HEK293 proteomes	89	NA
Cryo-EM structure of human ribosome	11	PDB: 4UG0
Experimental Models: Cell Lines		
Human (female): MCF7 cells	ATCC	RRID:CVCL_0031
Human (female): HEK293 cells	Laboratory of Rolf Sprengel (MPI Heidelberg)	RRID:CVCL_0045
Human (female): HeLa cells	ATCC	RRID:CVCL_0030
Recombinant DNA		
pcDNA5 FRT TO c7orf50 full length FLAG-HA c-ter	This laboratory	NA
pcDNA5 FRT TO DUF2373 FLAG-HA c-ter	This laboratory	NA
pcDNA5 FRT TO DUF2373_K120G FLAG-HA c-ter	This laboratory	NA
pcDNA5 FRT TO c7orf50 full length FLAG-HA n-ter	This laboratory	NA

pcDNA5 FRT TO DUF2373 FLAG-HA n-ter	This laboratory	NA
pcDNA5 FRT TO DUF2373_R156G FLAG-HA n-ter	This laboratory	NA
Software and Algorithms		
GOrilla	224	http://cbl-gorilla.cs.technion.ac.il/
BBMap (37.68)	JGI	http://sourceforge.net/projects/bbmap
Je (version 1.2)	225	https://git.embl.de/grp-gbcs/Je
STAR (version 2.5.0a)	226	https://github.com/alexdobin/STAR , RRID:SCR_015899
Bowtie2	227	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
HTSeq	228	https://htseq.readthedocs.io/en/release_0.9.1/ , RRID:SCR_005514
UCSF Chimera (1.12)	229	https://www.cgl.ucsf.edu/chimera/ , RRID:SCR_004097
MaxQuant (1.5.1.2)	230	http://www.biochem.mpg.de/5111795/maxquant , RRID:SCR_014485
MSFragger	231	http://www.nesvilab.org/software.html
R	R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.	https://www.r-project.org/ , RRID:SCR_001905
RStudio (0.99.903, RRID:SCR_000432)	RStudio: Integrated Development for R. RStudio, Inc., Boston, MA	http://www.rstudio.com/ , RRID:SCR_000432
csaw (R)	232	http://bioconductor.org/packages/csaw/
DESeq2 (R)	233	http://bioconductor.org/packages/DESeq2/ , RRID:SCR_015687
Additional Resources		
Advanced online documentation for XRNAX protocols	82	https://www.xrnax.com/

EXPERIMENTAL MODEL AND SUBJECT DETAILS

MAMMALIAN CELL CULTURE AND STABLE CELL LINES

The cell lines MCF7, HEK293 and HeLa (all human, female) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) for SILAC supplemented with 10% dialysed FBS and Pen-Strep (100 U / ml penicillin, 100 mg / ml streptomycin) at 37 °C, 5 % CO₂. DMEM for SILAC was supplemented with 1 mM L-lysine and 0.5 mM L-arginine of the individual SILAC labels as well as 1.7 mM light L-proline and 1 x GlutaMAX. The heavy SILAC label was introduced during six passages in heavy DMEM for SILAC.

All experiments were performed on MCF7 cells, except when deriving the integrated human RNA-binding proteome (ihRBP), where MCF7, HEK293 and HeLa cells were used as indicated in the text.

METHOD DETAILS

GUANIDINIUM THIOCYANATE–PHENOL–CHLOROFORM (TRIZOL) EXTRACTION

Up to 10 million MCF7 cells were lysed in 1 ml TRI reagent by pipetting up and down. For phase-separation, 200 µl chloroform was added and samples mixed by turning tubes upside down several times. After 5 minutes incubation at room temperature, samples were spun down with 12000 g for 10 minutes at 4 °C. Approx. 400 µl of the aqueous phase was transferred to a fresh tube, NaCl was added to a final concentration of 300 mM along with 1 µl GlycoBlue. Samples were combined with 500 µl isopropanol, mixed by inversion and RNA precipitated by centrifugation with 18000 g for 15 minutes at -10 °C.

The supernatant was removed and the RNA pellet washed with 1 ml of 70 % ethanol before resuspension in the desired volume of nuclease-free water.

UV-CROSSLINKING OF CELLS

All cells were grown in 245 mm x 245 mm dishes to the desired confluence. For the incorporation of 4-thiouridine (4SU) into RNA, cells were incubated with 100 µM 4SU for 16 hours prior to UV-crosslinking. Media was decanted and cells washed with 50 ml ice-cold PBS. In order to remove as much liquid as possible dishes were propped up straight and residual PBS drained onto a paper towel through gravity. UV-crosslinking occurred on ice with 200 mJ / cm² at 254 nm wavelength with a BIO-LINK UV-crosslinker (Vilber). Cells that had incorporated 4SU were UV-crosslinked at 365 nm wavelength. Subsequently, cells were harvested into ice-cold PBS, pelleted and either directly subjected to XRNAX or stored at -80 °C for up to 14 days.

PROTEIN-CROSSLINKED RNA EXTRACTION (XRNAX)

Up to 100 million cells (typically one confluent 245 mm x 245 mm dish of UV-crosslinked MCF7, HEK293 or HeLa cells combined with one confluent 245 mm x 245 mm dish of non-crosslinked cells) were lysed in 8 ml TRI reagent by pipetting up and down. Cell clumps were disintegrated by flushing the lysate repeatedly against the wall of the tube. Lysis was further facilitated by incubation on a rotating wheel for 5 minutes at room temperature. Lysates were combined with 1.6 ml chloroform, mixed by inversion and incubated for 5 minutes at room temperature. Tubes were spun down with 7000 g for 10 minutes at 4 °C.

The aqueous phase was removed and the interphase transferred to a 2 ml tube. The interphase was gently washed twice with 1 ml low SDS buffer (tris-Cl 50 mM, EDTA 1 mM, SDS 0.1 %), flushing protein off the walls of the tube while retaining the integrity of the interphase flakes. Flakes were spun down with 5000 g for 2 minutes at room temperature and the supernatant discarded. After the washing, flakes were disintegrated by pipetting into 1 ml of low SDS buffer. The disintegrated interphase was spun down with 5000 g for 2 minutes at room temperature and the supernatant saved as interphase eluate 1. Disintegration of the interphase was repeated with another 1 ml of low SDS buffer, then twice with 1 ml of high SDS buffer (tris-Cl 50 mM, EDTA 1 mM, SDS 0.5 %) each time yielding approx. 1 ml of interphase eluates.

NaCl was added to a final concentration of 300 mM to each of the 4 interphase eluates, along with 1 µl GlycoBlue and 1 ml isopropanol before mixing by inversion. Samples were spun down for 15 minutes with 18000 g at -10 °C. The supernatants were discarded and pellets from all four elutes were combined in 2 ml of 70 % ethanol. The combined sample was again centrifuged for 1 minute with 18000 g at room temperature, supernatant discarded and all residual ethanol removed. The pellet was taken up in 1.8 ml of nuclease-free water and detached from the wall of the tube with a pipette tip. The pellet was allowed to swell for 1 hour on ice with occasional mixing by inversion and eventually dissolved by pipetting.

200 µl NEB DNase I buffer 10 x was added along with 2 µl RNasin Plus, 100 µl NEB DNase and incubated for 60 minutes at 37 °C and 700 rpm shaking. Subsequently, the sample was isopropanol precipitated as described above without further addition of GlycoBlue. Pellets were taken up in 1000 µl nuclease-free water and dissolved by pipetting. RNA concentration was estimated by UV-spectroscopy on a NanoDrop One UV photospectrometer (Thermo Scientific), neglecting adsorption by protein. Purification of protein-free RNA from XRNAX extracts after proteinase K digestion showed that this estimation was within 15 % of the actual RNA content. All amounts of XRNAX extracts mentioned in the following are given in µg of RNA referring to this estimation and do not take protein content into account.

A detailed, photo-documented version of the XRNAX protocol is included as Methods S1. For updates on XRNAX and its applications visit www.XRNAX.com.

COMPARISON OF RNA EXTRACTS USING AGAROSE GEL-ELECTROPHORESIS

To verify the integrity of RNA extracted by TRIZOL or XRNAX, agarose-gel electrophoresis was performed using 1 % agarose in TBE and SYBR Safe staining. Specifically, for Figure 1B 0.05 %

of the total yield extracted from 10 million MCF7 cells using the indicated method was subjected to the indicated treatment. Samples were denatured in RNA gel loading dye containing formamide for 2 minutes at 85 °C and run for 40 minutes with 3 W.

RNA SEQUENCING FOR QUANTIFYING THE RELATIVE COMPOSITION OF RNA EXTRACTS

For RNA sequencing, 10 µg RNA (as determined by NanoDrop UV-spectroscopy neglecting the protein content of samples) were digested for 30 minutes at 55 °C in proteinase K buffer (50 mM tris-Cl, EDTA 5 mM, NaCl 150 mM, SDS 1%) using 10 µl proteinase K. Note that both TRIZOL and XRNAX extracted samples were treated identically. Subsequently, RNA was cleaned-up using the RNeasy mini kit (Qiagen) and was ready for RNA-Seq library preparation.

Specifically, for TRIZOL and XRNAX extracted RNA derived from MCF7 cells that were crosslinked at 254 nm wavelength (Figure 1C) RNA library preparation occurred with the TruSeq RNA Library Prep Kit v2 (Illumina, not stranded) after conditional depletion of ribosomal RNA (rRNA) using the Ribo-Zero rRNA Removal Kit (Illumina). Biological duplicates extracted with TRIZOL or XRNAX (4 samples in total) were barcoded to be sequenced in one lane on a HiSeq2000.

For TRIZOL and XRNAX-extracted RNA from 4SU-labeled MCF7 cells, library preparation occurred with the TruSeq Stranded Total RNA kit (Illumina, stranded) after depletion of rRNA using the Ribo-Zero rRNA Removal Kit. Biological duplicates extracted with TRIZOL or XRNAX (4 samples in total) were barcoded to be sequenced in one lane on a HiSeq2000.

PROTEOMIC SAMPLE PREPARATION

For mass spectrometry (MS) sample preparation, a modification of the SP3 protocol described by Hughes et al.¹⁶⁷ was used. For total proteome analysis approx. 1 million cells were lysed and reduced in 1 ml lysis buffer (tris-Cl 50 mM, DTT 10 mM, SDS 0.05%) at 95 °C, 700 rpm shaking for 30 minutes. For samples other than cells, e.g XRNAX extracts, samples were brought to a total volume of 100 µl with MilliQ water and combined with 900 µl lysis buffer before reduction at 95 °C, 700 rpm shaking for 30 minutes. Magnesium chloride (final concentration of 5 mM), CAA (20 mM), and EDTA-free protease inhibitor were added and mixed before addition of 1 µl of benzonase. Subsequently, digestion of nucleic acids and alkylation occurred for 2 hours at 37 °C, 700 rpm shaking. 400 µl SP3 beads were preconditioned by washing with MilliQ water 3 times, before reconstitution in 1 ml MilliQ water. EDTA was added to 10 mM final concentration along with 1 % SDS and 20 µl SP3 beads. Samples were vortexed vigorously and subsequently combined with 1 ml acetonitrile. Samples were mixed again and incubated for 15 minutes at room temperature for protein binding to occur. The beads were collected on a magnetic stand for 2 minutes and supernatants decanted. While on the magnetic stand beads were then washed 3 times with 2 ml ethanol 70 %, which was added for 1 minute and subsequently decanted. Residual ethanol was removed and beads were taken up in the desired digestion volume of TEAB 20 mM and adequate amounts of trypsin/LysC added to the solution (for total proteomes from 1 million cells 1 µg trypsin/LysC in 100 µl TEAB). Samples were digested at 37 °C, 700 rpm shaking overnight. For

single run analysis formic acid was added to a final concentration of 1 % and samples spun down for 5 minutes with 20000 g. Supernatants were transferred to fresh tubes without disturbing the pellet and analysed by high-pressure, liquid chromatography (HPLC).

High pH reversed-phase fractionation occurred under standard settings described below. Of the 40 collected fractions the initial 8 fractions up to approx. 18 % B were discarded, the following 32 fractions were combined to 8 using the scheme 1+9+17+25/.../8+16+24+32. The combined fractions were dried by SpeedVac and taken up in 1 % formic acid before analysis by HPLC-MS.

HIGH pH REVERSED-PHASE FRACTIONATION OF PROTEOMIC SAMPLES

Fractionation at high pH occurred on an Agilent Infinity 1260 LC system (Agilent) using a Phenomenex Gemini 3 μ M C18, 100 x 1 mm column (Phenomenex). Buffer A was NH_4COOH 20 mM, buffer B was 100 % acetonitrile. The following gradient was used for all applications described in this manuscript: 0-2 minutes 0 % B, 2-60 minutes linear gradient to 65 % B, 61-62 minutes linear gradient to 85 % B, 62-67 minutes 85 % B, 67-85 minutes 0 % B. Eluates were collected in 40 fractions and combined as described in the individual paragraphs.

HPLC-MS FOR THE ANALYSIS OF NUCLEOTIDE-CROSSLINKED PEPTIDES, DISCOVERY OF RNA-BINDING PROTEINS OR THE DIFFERENTIAL QUANTIFICATION OF RNA-BINDING

Separation by HPLC prior to MS occurred on an Easy-nLC1200 system (Thermo Scientific) using an Acclaim PepMap RSCL 2 μ M C18, 75 μ m x 50 cm column (Thermo Scientific) heated to 45 °C with a MonoSLEEVE column oven (Analytical Sales and Services). Buffer A was 0.1 % formic acid, buffer B was 0.1 % formic acid in 80 % acetonitrile. The following gradient was used for all applications described in this manuscript: 0 minutes 3% B, 0-4 minutes linear gradient to 8 % B, 4-6 minutes linear gradient to 10 % B, 6-74 minutes linear gradient to 32 % B, 74-86 minutes linear gradient to 50 % B, 86-87 minutes linear gradient to 100 % B, 87-94 minutes 100 % B, 94-95 linear gradient to 3 % B, 95-105 minutes 3 % B.

Single-run, total proteome analysis was performed on a Fusion Orbitrap mass spectrometer (Thermo Scientific). MS1 detection occurred in orbitrap mode at 60000 resolution, AGC target 1E6, maximal injection time 50 ms and a scan range of 375-1500 DA. MS2 detection occurred with an HCD collision energy of 33 in ion trap top20 mode with an isolation window of 1.6 Da, AGC target 1E4 and maximal injection time of 50 ms.

Detection of XRNAX-derived nucleotide-crosslinked peptides, XRNAX-derived RNA-binding proteomes and XRNAX-derived differential analysis of RNA-binding upon arsenite stress, as well as all analysis of fractionated total proteome samples was performed on a QExactive HF mass spectrometer (Thermo Scientific). MS1 detection occurred at 120000 resolution, AGC target 3E6, maximal injection time 32 ms and a scan range of 350-1500 DA. MS2 occurred with stepped NCE 26 and detection in top20 mode with an isolation window of 2 Da, AGC target 1E5 and maximal injection time of 50 ms.

ISOLATION OF NUCLEOTIDE-CROSSLINKED PEPTIDES FROM XRNAX EXTRACTS

For the isolation of nucleotide-crosslinked peptides, 1000 µg of XRNAX extract were produced from MCF7 cells using the extraction method described above (from 2 confluent 245 mm x 245 mm dishes). Two aliquots of 500 µg of XRNAX extract were brought to 950 µl final volume containing 50 mM tris-Cl, 0.1 % SDS and 10 mM DTT. 10 µg trypsin/LysC was added to each aliquot to a final volume of 1 ml and digestion occurred for 1 hour at 37 °C, 700 rpm shaking. CAA was added to a final concentration of 20 mM and digestion continued for another hour. Purification of peptide-crosslinked RNA from the digests occurred by silica column purification using the Qiagen RNeasy Midi Kit with modified protocol (refer to kit manual for buffer descriptors). 1 ml digest was combined with 3.5 ml buffer RLT in a 15 ml falcon tube, mixed by inversion and heated to 60 °C for 15 min. The sample was allowed to reach room temperature. 2.5 ml of 100 % ethanol was added, the sample mixed by inversion and applied to an RNeasy Midi column by centrifugation with 3000 g for 5 minutes. Washing occurred twice with 2.5 ml buffer RPE, buffer RW1 was not used. Elution occurred twice with 250 µl nuclease-free water. All eluates combined to approx. 900 µl, which were transferred to a fresh tube. NaCl was added to a final concentration of 300 mM along with 1 µl glycoBlue, 1 ml isopropanol, the sample mixed by inversion and incubated for 1 hour at -20 °C. Precipitation occurred by centrifugation with 18000 g at -10 °C for 60 minutes. The supernatant was discarded and the pellet washed with 70 % ethanol. All residual ethanol was removed and the pellet taken up in 60 µl tris-Cl 10 mM. The sample was heated to 85°C for 5 minutes and cooled on ice before addition of 1.5 µl of RNase A, RNase I and RNase T1. RNA digestion occurred for 12 hours at 37 °C, 700 rpm shaking before the sample was heated to 85 °C again for 5 minutes and cooled on ice. Another 1.5 µl of RNase A, RNase I and RNase T1 was added and the sample digested for another 12 hours.

High pH reversed-phase fractionation occurred under standard settings described above. The initial peak with high adsorption up to approx. 18 % B containing RNA contaminations was discarded, the following fractions combined, completely dried by SpeedVac, taken up in 1 % formic acid and analyzed by HPLC-MS.

A detailed version of this protocol is included as Methods S2. For updates on downstream applications of XRNAX visit www.XRNAX.com/applications.

EXPRESSION AND PURIFICATION OF C7ORF50 AND DUF2373 FROM E. COLI

Synthetic genes encoding full-length C7orf50 and the C7orf50 domain-of-unknown-function (DUF2373, residues 94-194) with a C-terminal HA-tag were ordered from GenScript. Both genes were codon optimized for expression in *E. coli* and subcloned into the pETM11-Sumo3 vector (EMBL). The expression plasmids were transformed into *E. coli* BL21 (DE3) cells (Novagen). Cells were grown in LB supplemented with 30 µg / ml kanamycin at 25°C until OD600 ~ 0.6. The temperature was then lowered to 18 °C and expression was induced by the addition of 0.5 mM IPTG. The cells were grown further overnight at 18 °C and harvested by centrifugation. The cell pellets were resuspended in running buffer (tris-Cl 50 mM pH=8.0, NaCl 800 mM, imidazole 20 mM and glycerol 10 %) supplemented with 1 x protease inhibitors,

benzonase and 10 µg / ml lysozyme (Sigma). The cells were lysed via sonication and the cleared lysates were loaded onto a 5 ml Protino Ni-NTA column (Macherey-Nagel). The His6-Sumo3-c7orf50 and His6-Sumo3-c7orf50_94/194 fusion proteins were eluted in running buffer containing 300 mM imidazole. To remove the N-terminal fusion tag, His6-tagged SenP2 protease was added to the elution fractions in a 1:100 ratio. The samples were digested overnight at 4°C while being dialysed to a buffer containing tris-Cl pH=8.0, NaCl 250 mM, imidazole 20 mM and glycerol 10 %. The dialyzed samples were loaded again onto a 5 ml Protino Ni-NTA column and the C7orf50 and C7orf50_94/194 proteins without the N-terminal fusion tag were collected in the flow through. These flow through fractions were then loaded onto a 5 ml HiTrap Heparin HP column (GE Healthcare) equilibrated with tris-Cl 50 mM pH=8.0, NaCl 250 mM and glycerol 10 % in order to remove RNA. Proteins were eluted from the heparin column in a gradient to tris-Cl 50 mM pH=8.0, NaCl 1.5 M and glycerol 10 %. The full length C7orf50 eluted at approximately 800 mM NaCl in the gradient, while C7orf50_94/194 eluted around 500 mM NaCl. After SDS-PAGE analysis, the elution fractions containing full-length C7orf50 or C7orf50_94/194 were pooled and adjusted to 0.5 mg / ml concentration using tris-Cl 50 mM pH=8.0, glycerol 10 % and NaCl 800 or 500 mM, respectively. Proteins were stored at -80 °C.

IN VITRO VALIDATION OF RNA-BINDING BY NATIVE AGAROSE GEL ELECTROPHORETIC MOBILITY SHIFT ASSAY

Total RNA was extracted from MCF7 cells using the conventional TRIZOL procedure (see above). RNA was diluted to 1000 ng / µl in nuclease-free water, heated to 85 °C for two minutes and immediately transferred onto ice before use. An assay dilution of 4 µg RNA / 70 µl assay buffer (tris-Cl 50 mM, NaCl 100 mM) was prepared and stored on ice until use. Dilutions of C7orf50 and DUF2373 (C7orf50_94/194) expressed in *E. coli* (see above) were prepared in 10 µl of their respective storage buffer (tris-Cl 50 mM pH=8.0, glycerol 10 % and NaCl 800 or 500 mM, respectively). Protein dilutions were combined with 70 µl of the RNA assay dilution for a total volume of 80 µl and mixed by pipetting. RNA-binding was allowed to occur for 5 minutes at 37 °C, 300 rpm shaking before samples were transferred onto ice again. 20 µl of the samples were combined with 4 µl purple gel loading dye (SDS-free, NEB) and immediately run on a 1 % agarose gel (TBE) with SYBRsafe staining at 4 W for 45 minutes.

IN VIVO VALIDATION FOR RNA-BINDING BY PNK ASSAY OR IP-MS

Per construct approximately ten million MCF7 cells were transfected for 24 hours with 15 µg plasmid DNA using Lipofectamine 3000 according to the manufacturer's instructions. For the T4 polynucleotide kinase (PNK) assay cells were crosslinked as described above and harvested into 2 ml immunoprecipitation (IP) buffer (tris-Cl 50 mM, NaCl 100 mM, MgCl₂ 1 mM, CaCl₂ 0.1 mM, NP40 1 %, SDS 0.1 %, sodium deoxycholate 0.5 %, 1 x EDTA-free protease inhibitor) by scraping. Of the lysates 1 ml was sonicated with a Sonifier (Branson) and treated with 3 units Turbo DNase and 5 µg RNase A for 15 minutes at 37 °C. 15 µl of anti-FLAG M2 beads slurry was used for one hour of IP at 4 °C on a rotating wheel, followed by three washes

with IP buffer and two washes with PNK buffer (tris-Cl 50 mM, NaCl 50 mM, MgCl₂ 10 mM, NP40 0.5 %). The PNK labeling reaction was then performed on-bead for 15 minutes at 37 °C by addition of 30 µl hot PNK mix (27 µl PNK buffer supplemented with DTT 5 mM, 3 µl T4 PNK, 0.3 µl γ-ATP (0.03 µCi)), followed by four washes with PNK buffer. Complexes were eluted 5 minutes at room temperature into 50 µl elution buffer (glycine 0.1 M, pH=3). The pH of eluates was neutralized by addition of 7.5 µl tris-Cl 1 M, samples combined with 12.5 µl NuPAGE loading dye supplemented with DTT 200 mM and the sample run on an SDS-PAGE (NuPAGE Bis-Tris, MOPS buffer) at 180 V for 45 minutes. Protein-RNA complexes were blotted onto a nitrocellulose membrane for one hour at 1 mA / cm² and exposed a phosphorimaging screen over night before imaging on a Typhoon scanner (GE healthcare). Subsequently, anti-HA Western-blot was performed on the same membranes.

For analysis by MS cells were crosslinked and extracted through XRNAX as described above. The entire yield of the XRNAX extract was dissolved in 200 µl of 50 mM tris-Cl and digested over night at 37 °C, 700 rpm shaking, by addition of 2.5 µl RNase I and 2.5 µl RNase A. The digest was combined with 250 µl IP buffer 2 x (tris-Cl 100 mM, NP40 1 %, LiCl 300 mM, LiDS 0.2 %). 50 µl anti-FLAG M2 beads were washed twice with 1 ml IP buffer 1 x (tris-Cl 50 mM pH=7.5, NP40 0.5 %, LiCl 150 mM, LiDS 0.1 %) and reconstituted in 50 µl IP buffer 1 x before addition to the XRNAX digest in 2 ml tubes. IP occurred for 4 hours at room temperature on a rotating wheel. Beads were washed three times with 500 µl IP buffer 1 x, each time resuspending 5 minutes on a rotating wheel. All residual IP buffer was removed and elution occurred twice for 30 minutes at 37 °C, 700 rpm shaking into µl 200 elution buffer (Tris-Cl 50 mM, SDS 10 %). Eluates were combined and subjected to SP3 protein cleanup as described above with following alterations. No alkylation or reduction was performed, 10 µl SP3 beads were added to the eluates along with 500 µl ACN 100 %. Samples were digested in 20 µl TEAB 20 mM and 100 ng trypsin/LysC over night at 37 °C, 700 rpm shaking. Beads were captured on a magnet and digests transferred to a fresh tube before addition of formic acid to a final concentration of 1 %. Samples were spun down for 5 minutes at 20000 g, supernatants transferred to a fresh tube and analyzed on a QExactive HF MS as described above.

A detailed version of this protocol is included as Methods S3. For updates on downstream applications of XRNAX visit www.XRNAX.com/applications.

SILAC-CONTROLLED DISCOVERY OF RNA-BINDING PROTEINS FROM XRNAX EXTRACTS

To maximize coverage of the RNA-bound proteome, we produced XRNAX extracts from half-confluent and confluent cells (~40 million cells per condition), each of which were subjected to silica purification after 15 or 30 minutes of partial tryptic digestion. Cells of one SILAC label were crosslinked with UV-light of 254 nm wavelength as described above, while control cells of the complementary label stayed non-crosslinked. Crosslinked and non-crosslinked cells were combined and extracted by XRNAX as described above.

Per replicate, 930 µg of XRNAX extract (in 930 µl) was further processed. Therefore tris-Cl was added to a final concentration of 50 mM, DTT to 10 mM and SDS to 0.1 % before 20 minutes of incubation at 60 °C, 700 rpm shaking. CAA was added to a concentration of 20 mM and

samples incubated for 20 minutes at room temperature. For predigestion, 100 ng trypsin/lysC was added and samples were pre-digested at 37 °C, 700 rpm shaking for 15 or 30 minutes, respectively.

Purification of protein-crosslinked RNA from the digests occurred by silica column purification using the Qiagen RNeasy Midi Kit with modified protocol (refer to kit manual for buffer descriptors). Predigestion was stopped by combining the sample (approx. 1 ml) with 3.5 ml buffer RLT. The sample was mixed by inversion and heated to 60 °C for 15 minutes. 2.5 ml of 100 % ethanol was added, the sample mixed by inversion and applied to an RNeasy Midi column by centrifugation with 3000 g for 5 minutes. The flow-through was saved for additional rounds of purification. Washing occurred twice with 2.5 ml buffer RPE. Elution occurred with 250 µl nuclease-free water. The purification was repeated 3 times, each time using the saved flow through and reusing the same RNeasy Midi column for the individual sample. To the combined eluates (approximately 900 µl total volume) NaCl was added to a final concentration of 300 mM along with 1 µl glycoblue and 1 ml isopropanol. The sample was mixed by inversion and incubated for 1 hour at -20 °C. Precipitation occurred by centrifugation with 18000 g at -10 °C for 30 minutes. The supernatant was discarded and the pellet washed with 70 % ethanol and taken up in 65 µl tris-Cl 50 mM. The sample was heated to 85°C for 5 minutes and cooled on ice before 2.5 µl of RNase A, RNase I and RNase T1 was added. RNA digestion occurred over night at 37 °C, 700 rpm shaking. 500 ng trypsin/LysC were added for 16 hours digestion at 37 °C, 700 rpm shaking.

High pH reversed-phase fractionation occurred under standard settings described above. The initial fractions up to approx. 18 % B containing RNA contaminations and few peptides were discarded, the following eluate was collected in six consecutive fractions, which were subsequently dried by SpeedVac and taken up in 1 % formic acid before analysis by HPLC-MS. A detailed version of this protocol is included as Methods S4. For updates on downstream applications of XRNAX visit www.XRNAX.com/applications.

QUANTIFICATION OF NASCENT PROTEIN UPON ARSENITE STRESS USING AZIDOHOMOALANINE-LABELING

MCF7 cells with heavy and light SILAC labels were expanded for three days on 15 cm dishes until 80 % confluent. Cells of one SILAC label were treated with 400 µM sodium arsenite for 5, 10, 20, 30 or 60 minutes, while cells of the complementary label were left untreated. Azidohomoalanine-labeling (AHA-labeling) and protein purification using click-chemistry for MS quantification was performed as described before ¹²⁵. In brief, cells were deprived of methionine using methionine-free media for 30 minutes. Labeling started with the addition of AHA-containing media, which occurred simultaneously with the addition of arsenite. After labeling, cells were immediately transferred onto ice, washed with ice-cold PBS and harvested by scraping. Protein enrichment was performed using the Click-iT Protein Enrichment Kit (Thermo Scientific) using the manufacturer's instructions. Protein captured on agarose beads was subjected to tryptic digestion using 500 ng trypsin/LysC in 200 µl TEAB 50 mM and

peptides were cleaned up using an Oasis PRiME HKB μ Elution Plate (Waters). HPLC-MS detection occurred on an Orbitrap Fusion MS using the parameters described above.

DIFFERENTIAL QUANTIFICATION OF RNA-BINDING UPON ARSENITE STRESS

For the differential quantification of RNA-binding upon arsenite stress, MCF7 cells were expanded for three days to approx. 70 % confluence. 30 million cells of one SILAC-label were exposed to 100 μ M sodium arsenite for 0, 5, 10, 20, or 30 minutes, while control cells of the complementary label remained untreated. Duplicate experiments were performed for each time point, which included SILAC-label swap. Both treated and control cells were UV-crosslinked, combined and subjected to XRNAX and silica-enrichment as described above with the only difference that the predigestion time was kept constant at 30 minutes for all samples. Samples were high pH reversed-phase fractionated into 8 fractions as described above.

Total proteomes were analyzed from cells treated with arsenite at the identical time points, in duplicates and with SILAC label-swap. After treatment the media was discarded and cells were immediately put on ice and washed with ice-cold PBS. Cells were harvested by scraping and subjected to the standard proteomic workflow described above before fractionation into 8 fractions at high pH. Importantly, for the quantification only peptides were used, which previously had been discovered as super-enriched during the generation of the iHRBP. Refer to the 'Quantification and Statistical Analysis' section below for further elaboration.

A detailed version of this protocol is included as Methods S5. For updates on downstream applications of XRNAX visit www.XRNAX.com/applications.

TOTAL PROTEOME ANALYSIS OF ARSENITE-INDUCED PROTEIN DEGRADATION

For total proteome analysis of MCF7 under controlled cell culture conditions 0.5×10^6 cells were seeded in 10 cm dishes and cultured for 3 days. Inhibition of autophagy through 10 μ M spautin-1 was induced 24 hours prior to arsenite stress. Sodium arsenite was applied at 400 μ M concentration. Cells were harvested and subjected to the standard proteomic workflow described above.

We note here that the degree to which autophagic degradation was induced heavily depended on the growth state that the cells were in: During the initial lag-phase of cell culture one day after seeding, arsenite-induced protein degradation was minor compared to the effect observed in the following days of culture (data not shown). For cells seeded at a density to reach confluence after 5 days of culture the most pronounced effect was observed after 3 days.

ETHYNYL-URIDINE INCORPORATION AND CONFOCAL MICROSCOPY FOR MONITORING RNA TURNOVER UPON ARSENITE STRESS

For the visualization of nascent transcripts using ethynyl-uridine (EU), MCF7 cells were grown on glass cover slips for 3 days. EU was applied at 1 mM, sodium arsenite at 400 μ M and the RNA polymerase I inhibitor CX5461 at 10 μ M concentration. Treatment occurred for 30 or 60 minutes, the media was discarded and cells washed once with PBS. Fixation occurred with 3

% paraformaldehyde in PBS at room temperature for 10 minutes subsequent to washing with PBS. Cells were permeabilized using 0.5 % Triton-X 100 in PBS for 15 minutes at room temperature and washed again with PBS. The copper-catalyzed click reaction occurred in 100 mM HEPES pH=8, 150 mM NaCl, 5 mM sodium ascorbate, 100 μ M CuSO₄, 500 μ M THPTA and 20 μ M sulfo-Cy5-azide for 30 minutes at room temperature. This and all following steps occurred under protection from light. The reaction solution was discarded and slides washed once with TBST (50 mM tris-Cl, 150 mM NaCl, 0.1 % Tween-20). HOECHST33342 was applied at a concentration of 10 μ M in TBSP for 10 minutes, slides again washed twice with TBST and mounted using ProlongGold antifade mountant.

Imaging was performed on a Leica SP5 (Leica) using a 63x oil emersion objective. Detection of sulfo-Cy5-EU occurred using the default Leica Cy5 filter settings with excitation at 633 nm and detection at 650-750 nm wavelength.

IMMUNOFLUORESCENCE STAINING AND CONFOCAL MICROSCOPY FOR MONITORING EXOSC2 LOCALIZATION UPON ARSENITE STRESS

For the visualization of EXOSC2, MCF7 cells were grown on glass cover slips for 3 days. Sodium arsenite was applied at 400 μ M for 0, 5 or 30 minutes, the media was discarded and cells washed once with PBS. Fixation occurred with 3 % paraformaldehyde in PBS at room temperature for 10 minutes subsequent to washing with PBS. Cells were permeabilized using 0.5 % Triton-X 100 in PBS for 15 minutes at room temperature and washed again with PBS. The antibody was diluted 1:100 in TBST (50 mM tris-Cl, 150 mM NaCl, 0.1 % Tween-20) and binding allowed to occur over night at 4° C. Slides were washed twice with TBST before the secondary antibody (goat anti-mouse Cy5) was applied at a 1:500 dilution in TBST for 2 hours. Slides were stained with HOECHST33342 and imaged as described above.

XRNAX CLIP-SEQ FOR THE DETECTION OF EXOSC2 RNA TARGETS

For crosslinking and immunoprecipitation followed by sequencing (CLIP-seq) from XRNAX extracts we first validated antibodies for normal IP by MS using the IP buffer 1 x (tris-Cl 50 mM pH=7.5, 0.5 % NP40, 150 mM LiCl, 0.1 % LiDS) later used for the CLIP-seq experiment.

For each sample XRNAX extracts were prepared from MCF7 cells as described above. For RNA fragmentation 10 mM tris-Cl and 5 mM EDTA were added to 100 μ g XRNAX extract, which were sonicated in microTUBEs with AFA fiber (Covaris, 520045) using a S220 focused-ultrasonicator (Covaris) with the settings: 900 seconds, peak power 175, duty factor 50, cycles / burst 200 and average power 87.5.

For the size-matched input control (SMI-control) 5 μ g of the sonicated XRNAX extract (approx. 2 μ l) were mixed with 33 μ l MilliQ water, 5 μ l FastAP buffer 10 x and 10 μ l FastAP. Dephosphorylation occurred for 15 minutes at 37 °C, then FastAP was inactivated for 5 minutes at 80 °C and the sample transferred to ice. 5 μ l PNK buffer 10 x, 10 μ l ATP 10 mM, 25 μ l MilliQ and 10 PNK were added and incubated another 15 minutes at 37 °C. 15 μ l of the SMI-control (approx. 200 ng RNA) was combined with 5 μ l SDS-loading dye 5 x (tris-Cl pH=6.8 250

mM, SDS 10%, 0.02% bromphenol blue, glycerol 30%) and 5 μ l DTT 1 M. Samples were heated to 70 °C for 15 minutes before they were run on an SDS-PAGE along with the IP samples.

For the IP 100 μ g sonicated XRNAX extract in approx. 125 μ l was combined with 125 μ l IP buffer 2 x (tris-Cl 100 mM pH=7.5, 1% NP40, 300 mM LiCl, 0.2% LiDS). 1 μ g antibody was added for 4 hours at 4 °C on a rotating wheel before antibody capture with 100 μ l protein G beads overnight. The beads were collected on a magnetic stand and the supernatant discarded. The beads were washed three times with 1 ml IP buffer, each time carefully turning the tube upside down until the beads were completely resuspended. Subsequently, beads were washed twice with 1 ml TBST while on the magnet. For end-repair the beads were resuspended in 100 μ l dephosphorylation mix (80 μ l MilliQ, 10 μ l FastAP buffer 10 x, 8 μ l FastAP, 2 μ l RNASin) and incubated for 15 minutes at 37 °C, 1000 rpm shaking. Beads were collected on a magnetic stand, the supernatant discarded and the beads washed twice with 1 ml TBST while on the magnet. Subsequently, the beads were resuspended in 100 μ l PNK mix (70 μ l MilliQ water, 10 μ l ATP 10 mM, 10 μ l PNK buffer A 10 x, 8 μ l PNK, 2 μ l RNASin) and incubated for another 15 minutes at 37 °C, 1000 rpm shaking. Beads were collected on a magnetic stand and the supernatant discarded. Protein-RNA complexes were eluted into 5 μ l SDS loading dye 5 x (NuPAGE), 5 μ l DTT 1 M and 15 μ l MilliQ for 15 minutes at 70 °C. Beads were collected on a magnetic stand and the IP sample transferred to a fresh tube.

IP and SMI control were run alongside on a 4-12% SDS-PAGE (NuPAGE BisTris, MES buffer) and blotted onto nitrocellulose with 500 mA for one hour at 4 °C. The area corresponding to the molecular weight of the protein of interest plus 75 kDa were excised, cut into pieces and transferred to a fresh tube. RNA was released by digestion with 50 μ l proteinase K in 200 μ l proteinase K buffer (tris-Cl 50 mM, EDTA 10 mM, NaCl 150 mM, SDS 1%) at 55 °C for 30 minutes. 250 μ l PCI for RNA was added, the sample mixed by inversion, incubated 10 minutes on ice and spun down 10 minutes with 12000 g at 4 °C. 200 μ l of the aqueous phase were transferred to a fresh tube, NaCl added to a final concentration of 300 mM, combined with 1 μ l GlycoBlue and 200 μ l isopropanol. Samples were mixed and precipitated for 2 hours at -20 °C before centrifugation with 18000 g at -10 °C for 1 hour. Pellets were washed with 80% ethanol and resuspended in nuclease-free water.

RNA produced by this protocol was approx. 30-80 nt in size, carried a 5' phosphate and a 3' hydroxyl. For generation of sequencing libraries, we used the NextFlex Small RNA 3.0 kit and gel-based size-selection of RNA fragments from 30-50 nt according to the manufacturer's instructions. Samples were barcoded so that twelve samples could be run in one lane on a HiSeq2000 (Illumina).

A detailed version of this protocol is included as Methods S6. For updates on downstream applications of XRNAX visit www.XRNAX.com/applications.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS DATABASE SEARCH

All MS raw files were searched using MaxQuant, except for data of nucleotide-crosslinked peptides. The database searched was the reviewed UniProt human proteome (search term: 'reviewed:yes AND proteome:up000005640', 20216 entries, retrieved 11 September 2017) and the default Andromeda list of contaminants. All settings were used at their default value, except for specifying SILAC configurations and indicating the appropriate number of fractions per sample. For the differential quantification of RNA-binding during arsenite stress the match-between-runs option was activated, for all other searches this was explicitly not the case.

MS data of nucleotide-crosslinked peptides was searched with MSFragger using the same UniProt database as mentioned above. Precursor mass tolerance was set to 1000 Da and the export format set to tsv, otherwise all settings were used at their default value.

PROCESSING AND ANALYSIS OF MS DATA

Cyclic-U crosslinked peptides were selected from peptide-spectrum matches (PSMs) carrying a modification within 306-307 Da, as detected by MSFragger.

For the analysis of RNA-binding proteins from XRNAX extracts, the MaxQuant peptides.txt table was filtered to remove entries in 'potential contaminants' and 'reverse'. Furthermore only peptides that matched the category 'Unique Groups' were used. To derive RNA-binding proteins for the individual cell lines, peptides from the four replicates were combined and filtered with the condition $(\text{SILAC intensity crosslinked} +1) / (\text{SILAC intensity non-crosslinked} +1) > 1000$. Pseudo-counts were added to include peptides where the non-crosslinked channel had zero intensity. Proteins ('Leading razor protein') identified with two or more unique peptides were included in the RNA-binding proteome of the individual cell line.

K-mer frequencies in proteins of the ihRBP were determined using the R package biostrings. Global protein sequence features were computed using the R package 'peptides' with the scales 'Kyte-Doolittle' for hydrophobicity, 'EMBOSS' for isoelectric point and 'EMBOSS' for charge.

For differential quantification of RNA-binding upon arsenite stress, the MaxQuant peptides.txt table was filtered to remove entries in 'potential contaminants' and 'reverse'. Furthermore, only peptides that matched the category 'Unique Groups' and which occurred in the list of ihRBP super-enriched peptides were used. The latter was essential for proper quantification, as demonstrated in Figure S1G. Note that a protein can contribute super-enriched and background proteins, so that combining the SILAC ratios of all peptides, the ones with the highest intensity, or similar approaches in a differential XRNAX experiment, will lead to wrong quantification. As an example, consider a protein with one super-enriched peptide and one background peptide in a differential XRNAX experiment with identical cell numbers for each

of the two SILAC channels. If the SILAC ratio of the super-enriched peptide is 0.1 but the background peptide 0.9, the median of the two will be 0.5. As the super-enriched peptide reflects the true change, its integration with a background peptide will falsely adjust the outcome by 500 % (i.e. $0.1 \times 5 = 0.5$). A list of super-enriched peptides collated from the cell lines MCF7, HEK293 and HeLa during generation of the ihRBP can be downloaded from www.XRNAX.com/applications. Subsequently to filtering, the median of normalized SILAC Ratios was computed in order to integrate peptide ratios to individual proteins ('Leading razor protein'). Here, a minimum of two super-enriched peptides quantified over all time points was required.

Control total proteome data for the experiment was analyzed identically, except for filtering for ihRBP super-enriched peptides, which was omitted. The combined data presented in this manuscript was the mean of biological replicates for each time point filtered for a variance of 15 % or smaller.

For the analysis of the arsenite-induced degradation normalized ratios from the MaxQuant proteinGroups.txt table were used. Dose-dependence was analysed using the R package 'drc'.

GENE-ONTOLOGY (GO) ENRICHMENT ANALYSIS

GO enrichment analysis was performed using the GOrilla web interface and uniprot identifier as input. Type (ranked or background controlled) of the enrichment and background control (if applicable) are indicated in the text.

VISUALIZATION OF THE HUMAN 80S RIBOSOME

Nucleotide-crosslinked peptides from ribosomal proteins and ribosomal proteins affected in their RNA-binding upon arsenite stress were located in the cryo-EM structure of the human ribosome ¹¹ (PDB 4UG0) using UCSF Chimera.

PROCESSING AND ANALYSIS OF RNA SEQUENCING DATA

For the estimation of the rRNA content of libraries, which had not been RiboZero depleted, reads were aligned to a collection of human ribosomal sequences of the hg19 assembly retrieved using the UCSC table browser. The table used was 'rmsk' and filtering was applied so that 'repClass does match rRNA'. All reads were aligned to those sequences using bowtie2 and reads that failed to align were written to a new file. Reads in this file were aligned to the complete hg19 assembly. Percentages of the rRNA content were estimated by comparing the number of reads aligning to hg19 rRNA sequences and residual reads aligning to the complete hg19 assembly.

For estimating the content of RNA biotypes in RiboZero depleted libraries, reads were aligned to the hg19 assembly using bowtie2. Subsequently, counting was performed with HTSeq-count using the geneset annotated by GENCODE 19 (release 12.2013) and using the GTF feature 'gene' for counting.

PROCESSING AND ANALYSIS OF XRNAX-CLIP-SEQ DATA

Unique molecular identifiers (UMIs) were extracted using Je. Exact PCR duplicates were removed using BMap. Adapters were trimmed using cutadapt and processed reads sequentially aligned to the 45S pre-ribosomal RNA (NR_046235.3), Repbase²³⁴ and the hg38 genome with STAR, reusing reads which escaped previous alignment efforts. Non-exact PCR duplicates were removed with Je. The csaw library was used to calculate 20 nt coverage windows for the 45S rRNA for each individual time point. Subsequently, the DESeq2 library was used to calculate fold changes from this coverage between duplicates of the IP and duplicates of the SMI control. P-values were corrected for multiple testing with Benjamini-Hochberg.

STATISTICAL ANALYSIS AND DATA VISUALIZATION

All data handling apart from what is mentioned above was performed in RStudio and visualized using the ggplot2 library. Microscopy, Western blot and phosphor images were processed in Photoshop (CC, Adobe 2015) and figures arranged in Illustrator (CC, Adobe 2015).

ADDITIONAL METHODS FOR CHAPTER 2

POLYSOME PROFILING FOR PROTEOMICS

PURIFICATION OF RIBOSOMAL COMPLEXES BY POLYSOME PROFILING

For one polysome profiling experiment approx. 10 million MCF7 cells grown on a 15 cm dish were used. The media was discarded and residual media removed by tapping the culture dish onto a paper towel. Cells were rinsed with 10 ml ice-cold harvest buffer (10 mM MgCl₂, 100 µg/ml cycloheximide in PBS), which was subsequently discarded and again residues removed by tapping the culture dish onto a paper towel. The dish was transferred onto ice and cells scraped into 100 µl ice-cold lysis buffer (5 % NP40, 250 mM tris-Cl, 50 mM MgCl₂, 700 mM KCl, 500 µg/ml cycloheximide, 5 mM DTT, 1 x EDTA-free protease inhibitor, to 100 µl add 15 µl DNase I and 1.5 µl RNAsin Plus before use). Lysates were transferred to a fresh tube, vortexed and incubated on ice for 10 minutes before passing through a 26 gauge needle 10 times in order to shear genomic DNA. Consequently, samples were cleared by centrifugation with 20000 g at 4 °C for 5 minutes. Sucrose gradients of 5-45 % were prepared in sucrose buffer (final concentration 50 mM tris-Cl, 10 mM MgCl₂, 140 mM KCl, 1 x EDTA-free protease inhibitor, 100 µg/ml cycloheximide) on a BIOCAMP153 gradient station (biocomp instruments). Supernatants were transferred onto the sucrose gradients and subjected to 3.5 hours of ultracentrifugation with 35000 rpm in Sorvall WX90 ultracentrifuge (Beckman) and the rotor SW40Ti. Gradients were then collected into 60 fractions using a BIOCAMP153 gradient station. Subfractions representing the 40S, 60S, 80S and polysome fractions were combined according to the UV trace. For one experiment series identical subfractions were combined, however, the exact fraction numbers varied slightly between experiments series and were adapted accordingly.

MODIFIED PROTEOMIC SAMPLE PREPARATION FOR POLYSOME PROFILING FRACTIONS

As the sucrose cushion in the polysome fractionation contained KCl, which precipitates SDS, we used a modification of the SP3 protocol presented above. The total volume of the combined fractions amounted to approx. 1 ml for the 40S, 60S and 80S fractions, of which 900 µl were transferred to a 2 ml tube. 10 µl DTT 1 M was added and samples were reduced at 90 °C for 30 minutes and then cooled on ice. 20 µl CAA 1 M was added along with 1 µl benzonase and samples incubated 2 hours at 37 °C, 700 rpm shaking. 100 µl denaturation solution (EDTA 200 mM, 20 M guanidinium chloride) was added and samples vigorously vortexed. 400 µl SP3 beads were preconditioned by washing with MilliQ water 3 times, before reconstitution in 1 ml MilliQ water. 20 µl preconditioned SP3 beads were mixed to the samples by vortexing, before addition of 1 ml acetonitrile. Binding of proteins occurred for 15 minutes at room temperature. Beads were captured on a magnetic rack, supernatants discarded and beads washed three times with 2 ml EtOH 70%, while attached to the magnet. An additional round of washing was performed were beads were taken off the magnet and disintegrated into 1 ml

EtOH 70 %. Samples were put back onto the magnet in order to remove all EtOH. Beads were taken up in 100 μ l TEAB 20 mM containing 500 ng trypsin/LysC and digested for 16 hours at 37 °C, 700 rpm and peptides cleaned up using a Oasis PRiME HKB μ Elution Plate. Peptides were taken up in 20 μ l formic acid 1 % and subjected to HPLC-MS on a QExactive MS using a 2 hour gradient and the parameters outlined above.

An exception was the polysome fraction, where the combined volume of the subfractions was approx. 4 ml. These 4 ml were collected in a 15 ml falcon tube and treated analogously to the other fractions. Volumes were adjusted except for benzonase and SP3 beads, for which volumes were kept the same. After collecting the SP3 beads on a 15 ml magnetic rack and discarding the supernatant, beads were taken up in two times 1 ml EtOH 70 % in order to be transferred to a fresh 2 ml tube. From that point on samples were treated identically to the other fractions.

PROTEIN HALF-LIFE MEASUREMENTS

DETERMINATION OF PROTEIN HALF-LIVES BY PULSED SILAC

For all protein half-life measurements two replicates were generated from MCF7 cells that included a SILAC label swap. Explicitly, one replicate was produced from SILAC heavy cells switched to light SILAC media and another replicate from SILAC light cells switched to heavy SILAC media.

For determining half-lives of the total proteome and half-lives on RNA (as assessed by XRNAX) seven timepoints were generated. In the case of half-lives of the total proteome 1 million MCF7 cells of one SILAC label were seeded on 15 cm dishes and expanded for 3 days. Subsequently, the media was discarded, cells were washed twice with PBS to remove residual media before addition of new media with the complementary SILAC label. Cells were harvested after additional 0, 1, 2, 4, 8, 16 and 32 hours of culture after switching the SILAC media. Therefore, the media was discarded, cells put on ice and scraped into 10 ml ice-cold PBS and transferred to a falcon tube. Residual cells were scraped into another 10 ml ice-cold PBS, combined with the rest and spun down for 5 minutes with 1000 g at 4 °C. Proteins were purified from the cells using the normal SP3 protocol detailed above and digested in 100 μ l TEAB 20 mM containing 1 μ g trypsin/LysC. Peptides were cleaned up using an Oasis PRiME HKB μ Elution Plate (Waters) and taken up in 20 μ l MilliQ water.

In the case of half-lives of proteins on RNA two 245 mm x 245 mm dishes per timepoint were used onto which 3 million MCF7 cells of one SILAC label were seeded and expanded for 3 days. Timepoints were produced as described for the total proteomes, however, cells were UV-crosslinked before harvest as outlined above. XRNAX and silica enrichment was performed on cells from each timepoint as described above. After silica enrichment and isopropanol precipitation pellets were taken up in 100 μ l TEAB heated to 85 °C for 5 minutes and subsequently cooled on ice. 2.5 μ l of the RNases A, I and T1 were added and samples incubated over night at 37 °C, 700 rpm shaking. 500 ng trypsin/LysC were added for 16 hours

digestion at 37 °C, 700 rpm shaking, peptides cleaned up using an Oasis PRiME HKB μ Elution Plate and taken up in 20 μ l TEAB 50 mM.

For determining half-lives of proteins in polysome fractions again two replicates with label-swap were generated. However, in order to increase the resolution of our analysis nine timepoints were taken. As the quality of polysome profiling suffers if samples are frozen, all cells were harvested at the same time. This required that the SILAC media was changed at appropriate time distances towards the common harvest point. Therefore, 1.3 million MCF7 cells of one SILAC label were seeded onto 15 cm dishes and harvested after a total of 4 days in culture. Cells were washed twice with PBS and switched to the complementary SILAC label 0, 2, 4, 8, 12, 16, 20, 24 and 32 hours before those 4 days had passed. Polysome profiling and protein clean-up occurred as described above. Peptides were cleaned up using an Oasis PRiME HKB μ Elution Plate and taken up in 20 μ l TEAB 50 mM.

TMT-SILAC STRATEGY FOR DETERMINATION OF PROTEIN HALF-LIVES

In the case of the total proteome and XRNAX the first 2-7 TMT channels were occupied by the timepoints 1-32 hours, the TMT channel 8 and 10 by replicates of the infinity timepoint (for the replicate starting from SILAC light cells the infinity timepoint were SILAC heavy cells and *vice versa*) and the channels 1 and 9 by replicates of the timepoint 0. Replicate channels were produced from the same peptides in order to assess the reproducibility within one experiment. The reproducibility was found very good for all replicates so that for the polysome profiling experiments all channels were used for timepoints, i.e. for polysome profiling fractions the timepoints 0-32 hours were allocated to the TMT channels 1-9 and the infinity timepoint to TMT channel 10.

In order to label identical amounts of peptides for each TMT channel, peptide concentrations were assessed using a Quantitative Colorimetric Peptide Assay (Pierce). The amount of peptides that were labeled was oriented by the sample with the lowest peptide concentration in one particular experiment series. In the case of the total proteome 10 μ g peptides per timepoint were used, in the case of XRNAX 2 μ g per timepoint were used and in the case of the polysome fractions 5 μ g of peptides were used. Identical amounts of peptides for each timepoint were adjusted to a total volume of 20 μ l in TEAB 50 mM. Consequently, 5 μ l of TMT labeling reagent (Thermo Scientific) in acetonitrile was added and labeling allowed to occur for 1 hour at room temperature. Labeling was quenched by addition of 1.5 μ l hydroxylamine 5% and another 15 minutes of incubation. Samples for all 10 TMT channels of one experiment were combined, completely dried by SpeedVac and fractionated at high pH as described above. In the case of the polysome profiling experiments, samples were analyzed in 8 fractions, which were combined as described above. In the case of the total proteome and XRNAX samples were analysed in 16 fractions. Therefore, the first 8 of the 40 collected fractions were discarded and the following combined to 16 fractions using the scheme 1+17/.../16+32. Fractions were again dried down completely and taken up in formic acid 1% before analysis by MS. HPLC-MS occurred on a Fusion MS using a 2 hour gradient as described above, however, with the standard SPS-MS3 method provided by the manufacturer.

PULSED SILAC FOR MONITORING EXCHANGE OF RIBOSOMAL PROTEINS

To detect exchange of ribosomal proteins from different ribosomal assemblies we inhibited ribosome biogenesis by inhibiting rRNA synthesis with 10 μ M CX5461 for 6 hours in SILAC heavy MCF7 cells. Cells were washed with PBS twice and then switched to light SILAC media. Control cells received light SILAC media without any additional supplements. Continued inhibition of ribosome biogenesis was achieved by addition of 10 μ M CX5461 to the light SILAC media. Additional inhibition of autophagy was prepared earlier when spautin-1 was added at 10 μ M concentration already 12 hours before switching to light SILAC media. Spautin-1 treatment continued in the light SILAC media throughout the rest of the experiment. Additional inhibition of the proteasome was prepared earlier when bortezomib was added at 500 nM concentration 1 hour before switching to light SILAC media. Bortezomib treatment continued in the light SILAC media throughout the rest of the experiment. Cells were subjected to polysome profiling and MS as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS DATABASE SEARCH FOR TMT-SILAC DATA

TMT-SILAC data was searched with MaxQuant 1.6.0.16 and parameters adapted from Zecha et al.¹⁵⁷. TMT isotope impurities were specified in Andromeda according to the information provided by the manufacturer. In the type section reporter ion MS3 and TMT 10plex was selected and reporter mass tolerance set to 0.01 Da. lysine 8 and arginine 10 were defined in Andromeda as variable modifications and subsequently selected as such in MaxQuant modifications section next to Oxidation and N-terminal acetylation. Peptide mass tolerance in the instrument section was set to 5 ppm. In the identification section the minimum score for modified peptides and the minimum delta score for modified peptides were set to 0. In the MS/MS – ITMS section the MS/MS – ITMS match tolerance was set to 0.4 Da and water loss was deselected. All other settings were used at their default value.

PROCESSING AND ANALYSIS OF TMT-SILAC DATA

Quantitative data for the determination of all protein half-lives was extracted from the ‘evidence’ table provided by MaxQuant. PSMs in the evidence table were filtered to remove ‘Potential contaminants’ and ‘Reverse’ matches to the decoy database. Because of their uncertain TMT labeling status PSMs of acetylated N-terminal peptides (‘Acetyl’ in the ‘Modification’ column) were removed as well. For all following steps corrected reporter ion intensities were used (‘Reporter intensity corrected’). We refer to the evidence table processed in this way as ‘cleaned-up evidence’ table.

Pseudo-equilibrium-normalization (PEN) transforms the TMT intensities of each peptide, so that the intensity of each TMT channel is the same, irrespective of SILAC labels. Therefore,

correction factors need to be calculated, which normalize the intensity of every channel towards one reference channel. Which channel is the reference channel is in principle arbitrary. However, we chose the 32 hour timepoint for this purpose because it had the most complete data for either SILAC channel and therefore allowed to calculate the most complete list of correction factors. Data was processed in two independent rounds, during the first of which the cleaned-up evidence table was used to sum up reporter intensities for each TMT channel from all PSMs of one peptide, independent of SILAC label. The resulting table contained for each peptide ten reporter ion intensities, which were now used to calculate for each TMT channel a correction factor towards the TMT channel of the 32 hour timepoint. The resulting table contained for each peptide ten correction factors, which in the case of the 32 hour timepoint was always 1 (because all values had been normalized towards this timepoint). This table of correction factors was stored for later use. In the second round of data processing, the cleaned-up evidence table was split by SILAC channel. Therefore, any PSM in the cleaned-up evidence table containing a lysine 8 or arginine 10 modification was assigned to the SILAC heavy dataset, whereas any PSM that did not carry any such modification was assigned to the SILAC light dataset. Just as during the first round of processing, PSMs were collapsed into peptides by summing up the reporter ion intensities for each TMT channel from all PSMs of a peptide. However, this time the SILAC channels were treated independently. The resulting tables contained for each peptide ten reporter ion intensities, which quantified synthesis or degradation (depending on SILAC heavy or light). Additionally, each peptide was annotated with the uniprot identifier ('Leading razor protein in the cleaned-up evidence table) so that peptides from the same host protein could be merged later on. Consequently, this raw data was intersected with the previously stored correction factors. Each raw reporter ion intensity for each peptide was corrected by multiplication with its specific correction factor. Importantly, the identical set of correction factors were used for the SILAC heavy or light dataset, so that the sum of the two datasets generated the identical reporter ion intensity in each TMT channel for each peptide. Note that for the control data ("raw") we present in Figure 7C&D the table of stored correction factors was substituted by a table with the identical dimensions only containing the entries 1. Thereby, the control data could be treated identically to the PEN-corrected data in all of the following steps of the analysis. As SILAC heavy or SILAC light denotes to either synthesis or degradation, depending on the replicate in our experiment, we refer to the resulting table after these steps as 'PEN corrected synthesis' or 'PEN corrected degradation' table, which now contained for each peptide ten corrected reporter ion intensities.

PEN corrected data was further processed and filtered in order to exclude peptides, which did not follow the expected increase or decrease in intensity, respectively, over time. Note that degradation and synthesis were treated differently. In the case of degradation, the PEN corrected reporter ion intensities from the PEN corrected degradation table were normalized to timepoint 0, where the maximum amount of protein was expected to exist. Peptides with resulting ratios larger 0.99 at timepoint infinity were discarded. Subsequently, the background at timepoint infinity was subtracted from all other timepoints. Peptides with resulting ratios smaller 0 or larger 1 were discarded. In the case of synthesis, the PEN corrected reporter ion

intensities from the PEN corrected synthesis table were normalized to timepoint infinity, where the maximum amount of protein was expected to exist. Peptides with resulting ratios larger 0.99 at timepoint 0 were discarded. Subsequently, the background at timepoint 0 was subtracted from all other timepoints. Peptides with resulting ratios smaller 0 or larger 1 were discarded. The resulting tables contained for each peptide ten normalized ratios of reporter ion intensities and additionally the uniprot identifier of their host protein.

Normalized peptide ratios were collapsed into protein ratios in order to fit degradation or synthesis functions to them. For each timepoint the normalized peptide ratios from all peptides of a protein ('Leading razor protein') were collapsed into one ratio using their median. The resulting table contained the uniprot identifier of a protein and 8 (in the case of the total proteome and XRNAX) or 10 (in the case of the polysome profiling data) normalized protein ratios. In the synthesis case, entries for timepoint 0 were always 0 and entries for timepoint infinity always 1. In the degradation case, entries for timepoint 0 were always 1 and entries for timepoint infinity always 0. We refer to these tables in the following as 'normalized protein synthesis ratio' or 'normalized protein degradation ratio' table.

For curve fitting we used the model introduced by Welle et al.¹⁵⁶, which Zecha et al.¹⁵⁷ had adapted. For curve fitting we used the library `minpack.lm` (Elzhov et al., <https://CRAN.R-project.org/package=minpack.lm>), which overcomes the zero-residual problem of exceptionally well-fitting data. Synthesis was modelled using the function: $y \sim (B_{syn} - A_{syn}) * \exp(-K_{syn} * x) + A_{syn}$, with the starting values $A_{syn} = 1$, $B_{syn} = 0.3$, $K_{syn} = 0.01$, the lower modelling constraints $A_{syn} = 0$, $B_{syn} = -2$, $K_{syn} = -1$ and the upper modelling constraints $A_{syn} = 3$, $B_{syn} = 2$, $K_{syn} = 5$. Degradation was modelled using the function: $y \sim (A_{deg} - B_{deg}) * \exp(-K_{deg} * x) + B_{deg}$ with the starting values $A_{deg} = 1$, $B_{deg} = 0.3$, $K_{deg} = 0.01$, the lower modelling constraints $A_{deg} = 0$, $B_{deg} = -2$, $K_{deg} = -1$ and the upper modelling constraints $A_{deg} = 3$, $B_{deg} = 2$, $K_{deg} = 5$. Pseudo R_2 for the fitted functions was assessed with the library `rcompanion` (Mangiafico et al., <https://CRAN.R-project.org/package=rcompanion>), using the Nagelkerke method for comparing a non-linear model towards a null model. The null model was created by fitting a null function with the `minpack.lm` library for the same timepoints as for the actual experiment: $y \sim \text{function}(x, m)\{m\}$ with the starting value $m = 1$. Fitting results were discarded if the fitting parameters did not meet the constraints $0.5 \leq A \leq 2$, $-1 \leq B \leq 1$, $K > 0$ and $R^2 \geq 0.8$ for either synthesis or degradation. The quality of the curve fits, i.e. R^2 values, often differed greatly between replicates and some proteins could not be fitted at all in one replicate or the other. Instead of compromising very good fits from one replicate by averaging them with mediocre fits from another replicate we decided to use fits from the replicate with the best quality. Therefore, we computed for each protein in each replicate the mean R^2 from the R^2 for fitting the synthesis function and the R^2 for fitting the degradation function. If a fit for one of the two did not exist, the existing R^2 was averaged with 0.8. For each protein the synthesis and degradation parameters were only kept for the replicate with the better mean R^2 . Synthesis and degradation half-lives were calculated from the parameter K using the formula: $HL = \log(2) / K$. Synthesis and degradation half-lives of a protein were regarded as technical replications (Figure 7D) and their mean was used for all following analysis.

SIZING PROTEIN-RNA INTERFACES WITHIN THE HUMAN 80S RIBOSOME

For determination of the interphase size between ribosomal proteins and rRNA Pymol 2.2.3.1 (Schrödinger) and the plugin PDiviz²³⁵ was used. Subunits were analyzed separately. The entire rRNA of one subunit was defined as one distinct molecule object and each ribosomal protein as another. Subsequently, the 'Backbone and Bases Surfaces' function in PDiviz was used to compute the 'Interface Area' between rRNA and one ribosomal protein at a time.

REFERENCES

1. König, J. *et al.* ICLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
2. Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **313**, 552–7 (1985).
3. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B. & Mattaj, I. W. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *J. Cell Biol.* **133**, 5–14 (1996).
4. Müller-McNicoll, M. & Neugebauer, K. M. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* **14**, 275–87 (2013).
5. Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* **13**, 246–259 (2012).
6. Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352 (2011).
7. Damgaard, C. K. & Lykke-Andersen, J. Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes Dev.* **25**, 2057–2068 (2011).
8. Driever, W. & Nüsslein-Volhard, C. A gradient of bicoid protein in *Drosophila* embryos. *Cell* **54**, 83–93 (1988).
9. Hastie, N. D. & Bishop, J. O. The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9**, 761–774 (1976).
10. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science (80-.)*. **289**, 920–930 (2000).
11. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640–645 (2015).
12. Zhang, X. *et al.* An Atomic Structure of the Human Spliceosome. *Cell* **169**, 918–929.e14 (2017).
13. Tuschl, T. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
14. Carmell, M. A. *et al.* Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science (80-.)*. **305**, 1437–1441 (2004).
15. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–8 (2001).
16. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–11 (1998).
17. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–40 (2010).
18. Cavaille, J., Nicoloso, M. & Bachellerie, J.-P. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* **383**, 732–735 (1996).
19. Kiss-László, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077–88 (1996).
20. Feng J *et al.* The RNA component of human telomerase. *Science (80-.)*. **269**, 1236–1241 (1995).
21. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
22. Bánfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–57 (2012).
23. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–51 (2013).
24. Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
25. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
26. LYON, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–3 (1961).
27. McHugh, C. a. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* (2015). doi:10.1038/nature14443
28. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416 (2015).
29. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by

- Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
30. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
 31. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7–21 (2014).
 32. Paul, I. J. & Duerksen, J. D. Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol. Cell. Biochem.* **9**, 9–16 (1975).
 33. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* (2018). doi:10.1038/nrm.2017.130
 34. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science (80-.)*. **357**, eaaf4382 (2017).
 35. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
 36. Decker, C. J., Teixeira, D. & Parker, R. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* **179**, 437–449 (2007).
 37. Mollieux, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
 38. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
 39. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* 1–15 (2016). doi:10.1016/j.molcel.2016.06.029
 40. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015).
 41. Ma, W. & Mayr, C. A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions. *Cell* **175**, 1492–1506.e19 (2018).
 42. McHugh, C. A. *et al.* Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol.* **15**, 203 (2014).
 43. Lambert, N. *et al.* RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Mol. Cell* **54**, 887–900 (2014).
 44. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–22 (1990).
 45. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: Chemi-SELEX. *Science (80-.)*. **249**, 505–510 (1990).
 46. Lerner, M. R. & Steitz, J. A. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Pnas* **76**, 5495–9 (1979).
 47. Brannan, K. W. *et al.* Article SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Article SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol. Cell* **64**, 1–12 (2016).
 48. Ule, J., Jensen, K. B., Ruggiu, M., Mele, A. & Darnell, R. B. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (80-.)*. **302**, 1212–1215 (2003).
 49. Smith, K. C. & Aplin, R. T. A Mixed Photoproduct of Uracil and Cysteine (5-S-Cysteine-6-hydrouracil). A Possible Model for the in Vivo Cross-Linking of Deoxyribonucleic Acid and Protein by Ultraviolet Light. *Biochemistry* **5**, 2125–2130 (1966).
 50. Smith, K. C. & O'Leary, M. E. Photoinduced DNA-protein cross-links and bacterial killing: a correlation at low temperatures. *Science* **155**, 1024–6 (1967).
 51. Smith, K. C. & Meun, D. H. Kinetics of the photochemical addition of [35S] cysteine to polynucleotides and nucleic acids. *Biochemistry* **7**, 1033–7 (1968).
 52. SHETLAR, M. D., CARBONE, J., STEADY, E. & HOM, K. Photochemical Addition of Amino Acids and Peptides To Polyuridylic Acid. *Photochem. Photobiol.* **39**, 141–144 (1984).
 53. Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R. & von Hippel, P. H. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *J. Biol. Chem.* **261**, 3512–8 (1986).
 54. Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* **16**, 323–6 (1991).
 55. Zaman, U. *et al.* Dithiothreitol (DTT) Acts as a Specific, UV-inducible Cross-linker in Elucidation of Protein–RNA Interactions. *Mol. Cell. Proteomics* **14**, 3196–3210 (2015).
 56. Shetlar, M. D., Hom, K. & Venditto, V. J. Photohydrate-mediated reactions of uridine, 2'-deoxyuridine and 2'-deoxycytidine with amines at near neutral pH. *Photochem. Photobiol.* **89**, 869–877 (2013).

57. Lee, F. C. Y. & Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol. Cell* **69**, 354–369 (2018).
58. G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, 28 (2016).
59. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
60. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
61. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
62. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
63. Rhee, H. *et al.* Proteomic Mapping of Mitochondria. *Science (80-.)*. **339**, 1328–1331 (2013).
64. Steube, A., Schenk, T., Tretyakov, A. & Saluz, H. P. High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells. *Nat. Commun.* **8**, 1–9 (2017).
65. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
66. Schueler, M. *et al.* Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol.* **15**, R15 (2014).
67. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70**, 854–867.e9 (2018).
68. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W. J. Identifying proteins that bind to specific RNAs - Focus on simple repeat expansion diseases. *Nucleic Acids Res.* **44**, 9050–9070 (2016).
69. Gibowsiu, P. J. & Sharp, P. A. Nuclear Ribonucleoprotein Particles in the Spliceosome Affinity. **233**,
70. Bardwell, V. J. & Wickens, M. Purification of RNA and RNA–protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.* **19**, 1980–1980 (1991).
71. Lim, F., Downey, T. P. & Peabody, D. S. Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7. *J. Biol. Chem.* **276**, 22507–22513 (2001).
72. Srisawat, C. & Engelke, D. R. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7**, 632–41 (2001).
73. Bachler, M. & Schroeder, R. StreptoTag: A novel method for the isolation of RNA-binding proteins. *Microbiology* 1509–1516 (2002).
74. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017).
75. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (80-.)*. 1–19 (2015).
76. Munschauer, M. *et al.* The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* (2018). doi:10.1038/s41586-018-0453-z
77. Dorn, G. *et al.* Structural modeling of protein–RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat. Methods* **14**, (2017).
78. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–406 (2012).
79. Perez-perri, J. I. *et al.* Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* **4**, (2018).
80. Milek, M. *et al.* DDX54 regulates transcriptome dynamics during DNA damage response. *Genome Res.* gr.218438.116 (2017). doi:10.1101/gr.218438.116
81. Queiroz, R. M. L. *et al.* Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* (2019). doi:10.1038/s41587-018-0001-2
82. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* (2018). doi:10.1016/j.cell.2018.11.004
83. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, (2014).
84. He, C. *et al.* High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol. Cell* **64**, 416–430 (2016).
85. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
86. Anantharaman, V. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464 (2002).

87. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
88. Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, 10127 (2015).
89. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **11**, M111.014050 (2012).
90. Schlackow, M. *et al.* Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65**, 25–38 (2017).
91. Wilusz, J. E. *et al.* A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392–407 (2012).
92. Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. How do lncRNAs regulate transcription? *Sci. Adv.* **3**, eaao2110 (2017).
93. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
94. Francia, S. *et al.* Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature* **488**, 231–235 (2012).
95. Chakraborty, A. *et al.* Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nat. Commun.* **7**, 13049 (2016).
96. Mohanty, B. K. & Kushner, S. R. Bacterial/archaeal/organelle polyadenylation. *Wiley Interdiscip. Rev. RNA* **2**, 256–276 (2011).
97. Matzov, D., Bashan, A. & Yonath, A. A Bright Future for Antibiotics? *Annu. Rev. Biochem.* **86**, 567–583 (2017).
98. McShane, E. *et al.* Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **167**, 803–806.e21 (2016).
99. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics During Arsenite-Induced Translational Arrest. *bioRxiv* (2018). doi:329995
100. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nat. Protoc.* **1**, 581–585 (2006).
101. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
102. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).
103. Ong, S.-E. *et al.* Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
104. Darnell, R. B. HITS-CLIP: Panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286 (2010).
105. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger : ultrafast and comprehensive peptide identification in mass spectrometry – based proteomics. *Nat. Publ. Gr.* **293**, (2017).
106. Hatchell, E. C. *et al.* SLIRP, a Small SRA Binding Protein, Is a Nuclear Receptor Corepressor. *Mol. Cell* **22**, 657–668 (2006).
107. Arieti, F. *et al.* The crystal structure of the split end protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Res.* **42**, 6742–6752 (2014).
108. Larburu, N. *et al.* Structure of a human pre-40S particle points to a role for RACK1 in the final steps of 18S rRNA processing. *Nucleic Acids Res.* **44**, 8465–8478 (2016).
109. Bonasio, R. *et al.* Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. *Elife* **3**, e02637 (2014).
110. Nozawa, R. S. *et al.* SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* **169**, 1214–1227.e18 (2017).
111. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, 202–208 (2009).
112. Elliott, D. & Lodomery, M. *Molecular Biology of RNA*. (Oxford University Press, 2015).
113. Hawkridge, A. *Quantitative Proteomics*. (Ro, 2014). doi:10.1039/9781782626985
114. Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G. & Chen, L.-L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.* **12**, R16 (2011).
115. Mitrea, D. M. & Kriwacki, R. W. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal.* **14**, 1 (2016).

116. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
117. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).
118. Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789–802.e12 (2016).
119. Calabretta, S. & Richard, S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).
120. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).
121. Ganesan, S. *et al.* BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell* **111**, 393–405 (2002).
122. Riley, K. J.-L. & Maher, L. J. p53 RNA interactions: new clues in an old mystery. *RNA* **13**, 1825–33 (2007).
123. Buchan, J. R. & Parker, R. Eukaryotic Stress Granules: The Ins and Outs of Translation. *Mol. Cell* **36**, 932–941 (2009).
124. Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **10**, 430–6 (2009).
125. Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S. & Krijgsveld, J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* **30**, 984–990 (2012).
126. Eichelbaum, K. & Krijgsveld, J. Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. *Mol. Cell. Proteomics* **13**, 792–810 (2014).
127. Kedersha, N. & Anderson, P. Stress granules : sites of mRNA triage that regulate mRNA stability Stress granules (SGs): a historical perspective. 963–969 (2002).
128. Anderson, P. & Kedersha, N. Stress granules : the Tao of RNA triage. 141–150 (2007). doi:10.1016/j.tibs.2007.12.003
129. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
130. Dmitriev, S. E. *et al.* GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor. *J. Biol. Chem.* **285**, 26779–26787 (2010).
131. Shibuya, T., Tange, T. Ø., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.* **11**, 346–351 (2004).
132. Zhang, T. *et al.* Autophagy is a cell self-protective mechanism against arsenic-induced cell transformation. *Toxicol. Sci.* **130**, 298–308 (2012).
133. Bolt, A. M., Byrd, R. M. & Klimecki, W. T. Autophagy is the predominant process induced by arsenite in human lymphoblastoid cell lines. *Toxicol. Appl. Pharmacol.* **244**, 366–373 (2010).
134. Bai, J. *et al.* Taurine protects against As₂O₃-induced autophagy in pancreas of rat offsprings through Nrf2/Trx pathway. *Biochimie* **123**, 1–6 (2016).
135. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife* **2015**, 1–21 (2015).
136. Buchan, J. R., Kolaitis, R. M., Taylor, J. P. & Parker, R. Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. *Cell* **153**, 1461–1474 (2013).
137. Liu, J. *et al.* Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13. *Cell* **147**, 223–234 (2011).
138. Haberman, N. *et al.* Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* **18**, 7 (2017).
139. Martin, C. *et al.* Lamin B1 maintains the functional plasticity of nucleoli. *J. Cell Sci.* **122**, 1551–1562 (2009).
140. Zinder, J. C. & Lima, C. D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **31**, 88–100 (2017).
141. Gudipati, R. K. *et al.* Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Mol. Cell* **48**, 409–421 (2012).
142. Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A. & Gleizes, P. E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* **6**, 225–242 (2015).
143. Goldfarb, K. C. & Cech, T. R. Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing. *Genes Dev.* **31**, 59–71 (2017).
144. Schneider, C., Kudla, G., Wlotzka, W., Tuck, A. & Tollervey, D. Transcriptome-wide Analysis of Exosome Targets. *Mol. Cell* **48**, 422–433 (2012).
145. Tafforeau, L. *et al.* The complexity of human ribosome biogenesis revealed by systematic nucleolar

- screening of pre-rRNA processing factors. *Mol. Cell* **51**, 539–551 (2013).
146. Sloan, K. E., Bohnsack, M. T., Schneider, C. & Watkins, N. J. The roles of SSU processome components and surveillance factors in the initial processing of human ribosomal RNA. *Rna* **20**, 540–550 (2014).
 147. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
 148. Hocq, R., Paternina, J., Alasseur, Q., Genovesio, A. & Le Hir, H. Monitored eCLIP: high accuracy mapping of RNA-protein interactions. *Nucleic Acids Res.* **46**, 11553–11565 (2018).
 149. Gu, J. *et al.* GoldCLIP: Gel-omitted Ligation-dependent CLIP. *Genomics, Proteomics Bioinforma.* **16**, 136–143 (2018).
 150. Goldberg, A. L. & Dice, J. F. Intracellular Protein Degradation in Mammalian and Bacterial Cells. *Annu. Rev. Biochem.* **43**, 835–869 (1974).
 151. Pratt, J. M. *et al.* Dynamics of Protein Turnover, a Missing Dimension in Proteomics. *Mol. Cell. Proteomics* **1**, 579–591 (2002).
 152. Milner, E., Barnea, E., Beer, I. & Admon, A. The Turnover Kinetics of Major Histocompatibility Complex Peptides of Human Cancer Cells*. *Mol. Cell. Proteomics* **5**, 357–365 (2006).
 153. Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J. & Beynon, R. J. Turnover of the human proteome: Determination of protein intracellular stability by dynamic SILAC. *J. Proteome Res.* **8**, 104–112 (2009).
 154. Cambridge, S. B. *et al.* Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10**, 5275–5284 (2011).
 155. Boisvert, F.-M. *et al.* A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11**, M111.011429 (2012).
 156. Welle, K. A. *et al.* Time-resolved Analysis of Proteome Dynamics by Tandem Mass Tags and Stable Isotope Labeling in Cell Culture (TMT-SILAC) Hyperplexing. *Mol. Cell. Proteomics* **15**, 3551–3563 (2016).
 157. Zecha, J. *et al.* Peptide level turnover measurements enable the study of proteoform dynamics. *Mol. Cell. Proteomics* mcp.RA118.000583 (2018). doi:10.1074/mcp.RA118.000583
 158. Mcalister, G. C. *et al.* MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. (2014).
 159. Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science (80-)*. science.1260793- (2015). doi:10.1126/science.1259038
 160. Sung, M.-K., Reitsma, J. M., Sweredoski, M. J., Hess, S. & Deshaies, R. J. Ribosomal proteins produced in excess are degraded by the ubiquitin–proteasome system. *Mol. Biol. Cell* **27**, 2642–2652 (2016).
 161. Lafontaine, D. L. J. A ‘garbage can’ for ribosomes: how eukaryotes degrade their ribosomes. *Trends Biochem. Sci.* **35**, 267–277 (2010).
 162. Chassé, H., Boulben, S., Costache, V., Cormier, P. & Morales, J. Analysis of translation using polysome profiling. *Nucleic Acids Res.* **45**, e15 (2017).
 163. Martin, T. E. & Hartwell, L. H. Resistance of active yeast ribosomes to dissociation by KCl. *J. Biol. Chem.* **245**, 1504–6 (1970).
 164. Van Den Elzen, A. M. G., Schuller, A., Green, R. & Séraphin, B. Dom34-Hbs1 mediated dissociation of inactive 80S ribosomes promotes restart of translation after stress. *EMBO J.* **33**, 265–276 (2014).
 165. Uesono, Y. & Toh-e, A. Transient inhibition of translation initiation by osmotic stress. *J. Biol. Chem.* **277**, 13848–13855 (2002).
 166. Liu, B. & Qian, S.-B. Characterizing inactive ribosomes in translational profiling. *Translation* **4**, e1138018 (2016).
 167. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).
 168. Shi, Z. *et al.* Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell* **67**, 71–83.e7 (2017).
 169. Eskelinen, E. L., Reggiori, F., Baba, M., Kovács, A. L. & Seglen, P. O. Seeing is believing: The impact of electron microscopy on autophagy research. *Autophagy* **7**, 935–956 (2011).
 170. Kraft, C., Deplazes, A., Sohrmann, M. & Peter, M. Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. *Nat. Cell Biol.* **10**, 602–610 (2008).
 171. Kraft, C., Peter, M. & Hofmann, K. Selective autophagy: Ubiquitin-mediated recognition and beyond. *Nat. Cell Biol.* **12**, 836–841 (2010).
 172. An, H. & Harper, J. W. Systematic analysis of ribophagy in human cells reveals bystander flux during selective autophagy. *Nat. Cell Biol.* **20**, 135–143 (2018).
 173. Wyant, G. A. *et al.* Nufip1 is a ribosome receptor for starvation-induced ribophagy. *Science (80-)*. **360**,

- 751–758 (2018).
174. Mathis, A. D. *et al.* Mechanisms of *In Vivo* Ribosome Maintenance Change in Response to Nutrient Signals. *Mol. Cell. Proteomics* **16**, 243–254 (2017).
 175. Lam, Y. W., Lamond, A. I., Mann, M. & Andersen, J. S. Analysis of Nucleolar Protein Dynamics Reveals the Nuclear Degradation of Ribosomal Proteins. *Curr. Biol.* **17**, 749–760 (2007).
 176. Pulk, A. *et al.* Ribosome reactivation by replacement of damaged proteins. *Mol. Microbiol.* **75**, 801–814 (2010).
 177. Peña, C., Hurt, E. & Panse, V. G. Eukaryotic ribosome assembly, transport and quality control. *Nat. Struct. Mol. Biol.* **24**, 689–699 (2017).
 178. Haddach, M. *et al.* Discovery of CX-5461, the first direct and selective inhibitor of RNA polymerase I, for cancer therapeutics. *ACS Med. Chem. Lett.* **3**, 602–606 (2012).
 179. Genuth, N. R. & Barna, M. The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell* **71**, 364–374 (2018).
 180. Simsek, D. *et al.* The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **169**, 1051–1065.e18 (2017).
 181. Imami, K. *et al.* Phosphorylation of the Ribosomal Protein RPL12/uL11 Affects Translation during Mitosis. *Mol. Cell* 1–15 (2018). doi:10.1016/j.molcel.2018.08.019
 182. Zorbias, C. *et al.* The human 18S rRNA base methyltransferases DIMT1L and WBSR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. *Mol. Biol. Cell* **26**, 2080–2095 (2015).
 183. Zhang, X. *et al.* A novel ER-microtubule-binding protein, ERLIN2, stabilizes Cyclin B1 and regulates cell cycle progression. *Cell Discov.* **1**, 5–9 (2015).
 184. Aviner, R. *et al.* Proteomic analysis of polyribosomes identifies splicing factors as potential regulators of translation during mitosis. *Nucleic Acids Res.* **45**, 5945–5957 (2017).
 185. Ciceri, P. *et al.* Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nat. Chem. Biol.* **10**, 305–312 (2014).
 186. Rahnamoun, H. *et al.* RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat. Struct. Mol. Biol.* **25**, 687–697 (2018).
 187. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11667–11672 (2009).
 188. Escribano-Díaz, C. *et al.* A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. *Mol. Cell* **49**, 872–883 (2013).
 189. Keskin, H. *et al.* Transcript-RNA-templated DNA recombination and repair. *Nature* **515**, 436–9 (2014).
 190. Wei, L. *et al.* DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination. *Proc. Natl. Acad. Sci.* **112**, E3495–E3504 (2015).
 191. Okayasu, R. Induction of DNA double strand breaks by arsenite: comparative studies with DNA breaks induced by X-rays. *DNA Repair (Amst)*. **2**, 309–314 (2003).
 192. Takahashi, M. *et al.* Stress Granules Inhibit Apoptosis by Reducing Reactive Oxygen Species Production. *Mol. Cell. Biol.* **33**, 815–829 (2013).
 193. Min Deng, A. *et al.* Deubiquitination and Activation of AMPK by USP10. (2016). doi:10.1016/j.molcel.2016.01.010
 194. Mahboubi, H., Barisé, R. & Stochaj, U. 5'-AMP-activated protein kinase alpha regulates stress granule biogenesis. *Biochim. Biophys. Acta* **1853**, 1725–37 (2015).
 195. Herzig, S. & Shaw, R. J. AMPK: guardian of metabolism and mitochondrial homeostasis. *Nat. Rev. Mol. Cell Biol.* (2017). doi:10.1038/nrm.2017.95
 196. McCormick, C. & Khapersky, D. A. Translation inhibition and stress granules in the antiviral immune response. *Nat. Rev. Immunol.* (2017). doi:10.1038/nri.2017.63
 197. Poirier, M. G., Eroglu, S. & Marko, J. F. The bending rigidity of mitotic chromosomes. *Mol. Biol. Cell* **13**, 2170–2179 (2002).
 198. Van Dyke, N., Chanchorn, E. & Van Dyke, M. W. The *Saccharomyces cerevisiae* protein Stm1p facilitates ribosome preservation during quiescence. *Biochem. Biophys. Res. Commun.* **430**, 745–750 (2013).
 199. Van Dyke, N., Baby, J. & Van Dyke, M. W. Stm1p, a Ribosome-associated Protein, is Important for Protein Synthesis in *Saccharomyces cerevisiae* under Nutritional Stress Conditions. *J. Mol. Biol.* **358**, 1023–1031 (2006).
 200. Anger, A. M. *et al.* Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80–85 (2013).
 201. Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
 202. Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* **12**, 147–153 (2015).
 203. NATHANS, D. PUROMYCIN INHIBITION OF PROTEIN SYNTHESIS: INCORPORATION OF PUROMYCIN INTO

- PEPTIDE CHAINS. *Proc. Natl. Acad. Sci. U. S. A.* **51**, 585–92 (1964).
204. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
 205. Huang, M.-T. Harringtonine, an Inhibitor of Initiation of Protein Biosynthesis. *Mol. Pharmacol.* **11**, 511–519 (1975).
 206. FRESNO, M., JIMÉNEZ, A. & VÁZQUEZ, D. Inhibition of Translation in Eukaryotic Systems by Harringtonine. *Eur. J. Biochem.* **72**, 323–330 (1977).
 207. Huang, M.-T. Harringtonine, an Inhibitor of Initiation of Protein Biosynthesis. *Mol. Pharmacol.* **11**, 511–519 (1975).
 208. Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* **12**, 147–53 (2015).
 209. Liakath-Ali, K. *et al.* An evolutionarily conserved ribosome-rescue pathway maintains epidermal homeostasis. *Nature* **556**, 376–380 (2018).
 210. Slavov, N., Semrau, S., Airoldi, E., Budnik, B. & van Oudenaarden, A. Differential Stoichiometry among Core Ribosomal Proteins. *Cell Rep.* **13**, 865–873 (2015).
 211. Xue, S. & Barna, M. Specialized ribosomes: A new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012).
 212. Bussiere, C., Hashem, Y., Arora, S., Frank, J. & Johnson, A. W. Integrity of the P-site is probed during maturation of the 60S ribosomal subunit. *J. Cell Biol.* **197**, 747–759 (2012).
 213. Sulima, S. O. *et al.* Eukaryotic rpl10 drives ribosomal rotation. *Nucleic Acids Res.* **42**, 2049–2063 (2014).
 214. Ni, L. & Snyder, M. A Genomic Study of the Bipolar Bud Site Selection Pattern in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **12**, 2147–2170 (2001).
 215. Nielsen, M. H., Flygaard, R. K. & Jenner, L. B. Structural analysis of ribosomal RACK1 and its role in translational control. *Cell. Signal.* **35**, 272–281 (2017).
 216. Wahl, M. C. & Möller, W. Structure and function of the acidic ribosomal stalk proteins. *Curr. Protein Pept. Sci.* **3**, 93–106 (2002).
 217. May, K. L. *et al.* The P1/P2 proteins of the human ribosomal stalk are required for ribosome binding and depurination by ricin in human cells. *FEBS J.* **279**, 3925–3936 (2012).
 218. McCluskey, A. J. *et al.* The Catalytic Subunit of Shiga-like Toxin 1 Interacts with Ribosomal Stalk Proteins and is Inhibited by Their Conserved C-Terminal Domain. *J. Mol. Biol.* **378**, 375–386 (2008).
 219. TSURUGI, K. & OGATA, K. Evidence for the Exchangeability of Acidic Ribosomal Proteins on Cytoplasmic Ribosomes in Regenerating Rat Liver. *J. Biochem.* **98**, 1427–1431 (1985).
 220. Sivan, G., Kedersha, N. & Elroy-Stein, O. Ribosomal Slowdown Mediates Translational Arrest during Cellular Division. *Mol. Cell. Biol.* **27**, 6639–6646 (2007).
 221. Liu, Y. *et al.* Autophagy-dependent rRNA degradation is essential for maintaining nucleotide homeostasis during *C. elegans* development. *Elife* **7**, e36588 (2018).
 222. Dang, Y. *et al.* Eukaryotic initiation factor 2 α -independent pathway of stress granule induction by the natural product pateamine A. *J. Biol. Chem.* **281**, 32870–32878 (2006).
 223. Frankel, L. B., Lubas, M. & Lund, A. H. Emerging connections between RNA and autophagy. *Autophagy* **13**, 1–21 (2016).
 224. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
 225. Girardot, C., Scholtalbers, J., Sauer, S., Su, S. Y. & Furlong, E. E. M. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics* **17**, 4–9 (2016).
 226. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
 227. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 228. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
 229. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
 230. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–72 (2008).
 231. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Publ. Gr.* **293**, (2017).
 232. Lun, A. T. L. & Smyth, G. K. Cseq: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, (2015).
 233. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data

- with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
234. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 4–9 (2015).
 235. Ribeiro, J., Melo, F. & Schüller, A. PDIviz: Analysis and visualization of protein-DNA binding interfaces. *Bioinformatics* **31**, 2751–2753 (2015).

ACKNOWLEDGEMENTS

This section enumerates people that helped to bring this thesis about. The listing is strictly hierarchical and appreciates the helpfulness of a person by rank. High ranking people can live in the expectation to have a fulfilled and truly wonderful life, whereas medium and low-ranking people should not be so sure about that. Jokes aside, the entire presented work would not have been possible without the fantastic help I received. My gratitude goes out to all of you: Thank you very much.

I want to thank Jeroen Krijgsveld for his supervision and mentorship. Jeroen has granted me great liberties, some of which have led to great things. The work in his lab has made me grow as a scientist and grow as a person. Thank you Jeroen for the chances and the patience.

I want to thank the entire Krijgsveld lab, current and previous members. Thanks, Christian Frese for telling me over and over again I should let XRNAX go. There was no better motivation. Also, thanks for teaching me MS and valuable scientific and non-scientific discussions. In that respect, thanks also to Gertjan Kramer for teaching me more MS and for keeping the engines running. Thanks, to Gianluca Sigismondo for a reasonable mind in scientific and social questions. Thanks, to Sophia Föhr for keeping the shakers shaking, the orders in order and for closing the order sheet (if you could do that, that would be great, thanks). Thanks, to Daria Fijalkowska for interesting and reasonable thoughts. Thanks, to Dimitrios Papageorgiou and Daniele Colombo for the same thing. Thanks, to Mandy Rettel, Sina Rafiee, Nade Abazova, Torsten Müller, German Monogarov, Bianca Kuhn, Karim Aljakouch, Gesa Durniock-Thierschmann and Marco Jochem for creating a great place to work.

I want to thank Sara Fahs for potassium acetate and for being the chemist I love. Thanks, for putting up with me during the writing process of this thesis. And thanks, for irradiating me with your happiness to the point of crosslinking (that turned out weirder than expected).

I want to thank my Parents, my brother Simon Trendel, my brother Christoph Trendel and my stepfather Claus Hagmayer for continuous support.

I want to thank all the collaborator that contributed to this thesis project. For their help on the Cell manuscript I want to thank Matthias Hentze, Tom Schwarzl, Rasti Horos, Alex Bateman and Ananth Prakash. Thanks, to Vladimir Benes, Nayara Azevedo and Bettina Hase from EMBL GeneCore and Kim Remans and Jacob Scheurich from EMBL PepCore for their advice and help. Thanks, to Markus Landthaler and Miha Milek for access to their MCF7 interactome capture data.

For their help on the ribosome turnover I want to thank Bernd Bukau, Günter Kramer and Matilde Bertolini. Thanks to Bernhard Küster and Jana Zecha for their help with the MaxQuant version and settings for the search of TMT-SILAC data. Thanks, to Gregor Mönke for his help

through the differential equations around PEN. I want to thank Ignacio Ibarra del Rio from the laboratory of Judith Zaugg, who pointed me towards and helped me to install the PDViz tool.

Last but not least I want to thank the members of my thesis advisory committee, Judith Zaugg, Nassos Typas and Andres Jäschke, for their valuable input and advice over the years. Especially the recommendation to coordinate the publication of XRNAX with competing groups has saved me a lot of nerve and maybe even a Cell paper. Thanks very much for that.

SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURES FOR CHAPTER 1

Figure S1: Proteomic and transcriptomic features of XRNAX-extracts.

Adapted from Trendel et al., 2018.⁸²

- A) Relative abundance of proteins in the total proteome and XRNAX-extracts of MCF7 cells, estimated from iBAQ values. Bar graphs display contribution of the top-30 proteins to the combined iBAQ intensity of all detected proteins.
- B) Scatterplot comparing normalized read counts for all GENCODE-annotated transcripts in RNA obtained by XRNAX and TRIZOL. MCF7 cells were exposed to 4SU for 16 hours before UV-crosslinking at 365 nm and processing via XRNAX, or without crosslinking and conventional TRIZOL extraction. Reads were counted per gene and normalized to the total number of counts. Each point represents one gene and displays the mean of two replicates. Contour lines indicate highest density of the points in the plot. Dashed lines indicate fold-changes of 1, 10 and 100. Sequencing for all replicates was performed in one lane and read count for all libraries was within 10 % deviation from the average read-count.
- C) Comparison of normalized read counts for all GENCODE genes between XRNAX and TRIZOL-extracted RNA. Same data as in B shown as cumulative distribution.
- D) Ranked scatterplot (left) and boxplot (right) for RNA-crosslinked protein in XRNAX extracts before and after silica enrichment. Percentages of crosslinking were computed using peptide SILAC intensities and the formula $\text{percent Xlinked} = (\text{Xlink/control}) / (\text{Xlink/control} + 1)$.
- E) Dotplot showing the enrichment of five well-studied RNA-binding proteins through XRNAX. Each dot represents one peptide of the indicated protein and how much of it occurs in the crosslinked SILAC channel (using the formula from D). Blue coloring shows this percentage after XRNAX only, whereas red coloring after XRNAX and subsequent silica enrichment.
- F) Density plot showing the enrichment of peptides from XRNAX-extracts with intact RNA over XRNAX-extracts where RNA was degraded. For details see text.
- G) Proof-of-concept for the differential quantification of RNA-binding using XRNAX and silica enrichment. Heavy SILAC-labeled MCF7 cells were UV-crosslinked and mixed with non-crosslinked heavy MCF7 cells in 5 defined ratios. These mixtures of heavy cells were combined with the identical amount of UV-crosslinked, light cells and subjected to XRNAX followed by silica enrichment and MS quantification. Histogram displays SILAC ratios without normalization. Peptides that were found super-enriched in previous experiments using a non-UV-crosslinked control (Figure 1F) showed discrete fold-changes corresponding to mixing ratios (grey arrows), whereas peptides that were not super-enriched before showed a 1:1 ratio (black arrow).

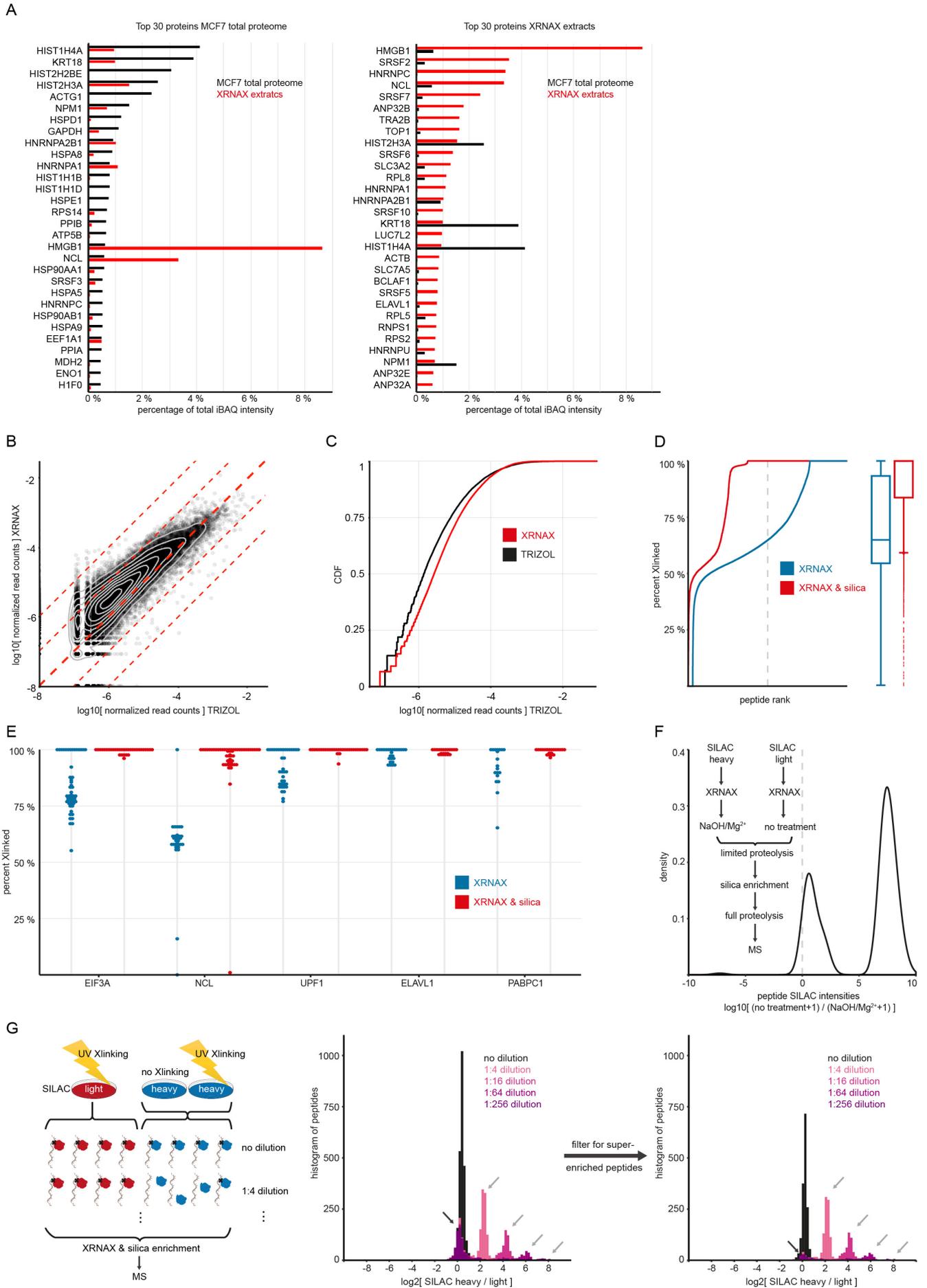


Figure S2: Sequence analysis of C7orf50 and the putative RNA-binding domain DUF2373.

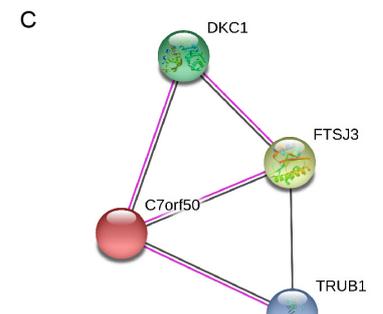
Adapted from Trendel et al., 2018.⁸²

- (A) Proteins with the largest number of identified cyclic U-crosslinked peptides, indicating their annotated domain structure (dark grey) and the position of the detected cyclic U-containing peptides.
- (B) Sequence alignment for 17 high-confidence hits from a HMMER search for human C7orf50. The identified cyclic U-crosslinked peptide, ELTVQKAEALMR, is highlighted in red.
- (C) STRING interaction network for C7orf50. DKC1 and TRUB1 are involved in pseudourylation of rRNA, FTSJ3 in methylation of rRNA.
- (D) Immunofluorescence image of C7orf50 in MCF7 cells. Red: microtubules; green: C7orf50. Image courtesy: www.proteinatlas.org.
- (E) Protein-RNA interactions in crystal structure of the human ribosome for four exemplary proteins. The amino-acid backbone is indicated in blue, peptides found crosslinked to cyclic U are indicated in magenta and RNA is indicated in grey. Red circles highlight phenylalanine in the vicinity of a uracil base. Upper left: GFVKVVK in RPL5, upper right: IHGVGFKK in RPL31, lower left: MKFNPFVTSDR in RPL26L1, lower right: VAYVSFGPHAGK in RPL14.
- (F) Native agarose gel electrophoretic mobility shift assay (EMSA) for C7orf50 and its DUF2373. Proteins were expressed and purified from *E. coli*. Increasing amounts of protein were incubated with constant amounts of MCF7 total RNA and run on an agarose gel without prior denaturation. Left: Addition of denaturing running buffer and heating (den.) demonstrates that RNA under assay conditions (assay) is intact.

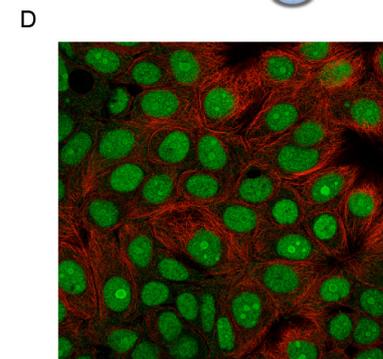


B

Q9BRJ6	CG050_HUMAN	1	MAKQKRKVEPV-----	11
Q510B3	CG050_RAT	1	MAKQKRKVEPV-----	11
Q9CXL3	CG050_MOUSE	1	MAKQKRKVEPV-----	11
E7FH60	E7FH60_DANRE	1	MDKSKADEFLKRRKRSQAQEKTKNRKAEDDELVTLQEDETNRMTVEEKTKKQKR--	57
AA0AGGJY4	AA0AGGJY4_MOUSE	1	MAKQKRKVEPV-----	11
K7AV53	K7AV53_PANTR	1	MAKQKRKVEPV-----	11
K7D4B2	K7D4B2_PANTR	1	MAKQKRKVEPV-----	11
I7GKL7	I7GKL7_MACFA	1	MAKQKRKVEPV-----	11
H9FM2	H9FM2_MACMU	1	MAKQKRKVEPV-----	11
K91B47	K91B47_DESRO	1	MAKQKRKVEPV-----	11
V9L6J3	V9L6J3_CALMI	1	MAKQKRKVEPV-----	13
V9L8K5	V9L8K5_CALMI	1	MAKQKRKVEPV-----	13
V9L8I2	V9L8I2_CALMI	1	MAKQKRKVEPV-----	13
V9L5Y3	V9L5Y3_CALMI	1	MAKQKRKVEPV-----	13
B9DQ82	B9DQ82_SALSA	1	MAKQKRKVEPV-----	13
ASR461	ASR461_BOVIN	1	MAKQKRKVEPV-----	11
AA0AF7ZEH5	AA0AF7ZEH5_CROAD	1	MAKQKRKVEPV-----	0



Q9BRJ6	CG050_HUMAN	12	-----	33
Q510B3	CG050_RAT	12	-----	33
Q9CXL3	CG050_MOUSE	12	-----	33
E7FH60	E7FH60_DANRE	58	-----	102
AA0AGGJY4	AA0AGGJY4_MOUSE	12	-----	33
K7AV53	K7AV53_PANTR	12	-----	33
K7D4B2	K7D4B2_PANTR	12	-----	33
I7GKL7	I7GKL7_MACFA	12	-----	33
H9FM2	H9FM2_MACMU	12	-----	33
K91B47	K91B47_DESRO	12	-----	33
V9L6J3	V9L6J3_CALMI	14	-----	57
V9L8K5	V9L8K5_CALMI	14	-----	57
V9L8I2	V9L8I2_CALMI	14	-----	56
V9L5Y3	V9L5Y3_CALMI	14	-----	56
B9DQ82	B9DQ82_SALSA	12	-----	34
ASR461	ASR461_BOVIN	12	-----	33
AA0AF7ZEH5	AA0AF7ZEH5_CROAD	1	-----	0



Q9BRJ6	CG050_HUMAN	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
Q510B3	CG050_RAT	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	86
Q9CXL3	CG050_MOUSE	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	86
E7FH60	E7FH60_DANRE	103	-----	156
AA0AGGJY4	AA0AGGJY4_MOUSE	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
K7AV53	K7AV53_PANTR	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
K7D4B2	K7D4B2_PANTR	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
I7GKL7	I7GKL7_MACFA	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
H9FM2	H9FM2_MACMU	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
K91B47	K91B47_DESRO	34	APSGGAGSGKGAIVRPLDAPLSPFEEQVLRERLKKRKKKEERQRIEAGLVAQH--	91
V9L6J3	V9L6J3_CALMI	58	-----	109
V9L8K5	V9L8K5_CALMI	64	-----	115
V9L8I2	V9L8I2_CALMI	57	-----	108
V9L5Y3	V9L5Y3_CALMI	65	-----	116
B9DQ82	B9DQ82_SALSA	94	-----	149
ASR461	ASR461_BOVIN	34	-----	110
AA0AF7ZEH5	AA0AF7ZEH5_CROAD	1	-----	0

E

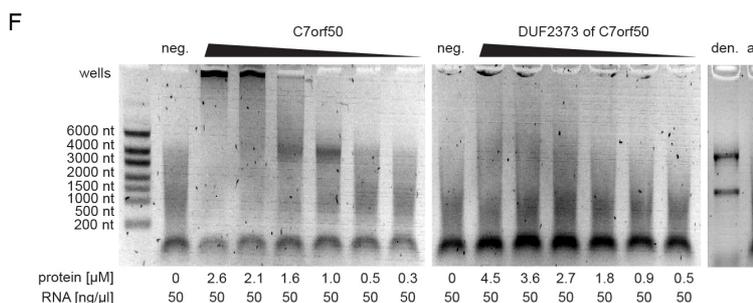


Figure S3: Properties of XRNAX-derived RNA-binding proteomes.

Adapted from Trendel et al., 2018.⁸²

- A) Boxplot comparing adjusted p-values as reported by Castello et al. for proteins discovered through both interactome capture (IAC) and XRNAX, or interactome capture only.
- B) GO-enrichment analysis (molecular function) for XRNAX-derived RNA-binding proteomes from three cell lines. Displayed are adjusted p-values for ten RNA-related terms with especially strong enrichment in either group. Note that enrichment analysis was performed against two different background sets. The poly(A) interactome was analyzed using the combined proteome of MCF7, HeLa and HEK293 cells reported by Geiger et al. as background control. A similar analysis was problematic for the non-poly(A) interactome because of the high prevalence of RNA-binding proteins in the total proteome of the three cell lines, annotated as such in reference to the two initial poly(A) interactome studies by Baltz et al. and Castello et al.. Since for this particular analysis we were exclusively interested in non-poly(A) interactors, we removed the published poly(A) interactome from the background control. Inf: infinity, NS: not significant with a p-value < 10E3.
- C) Scatter plots comparing amino acid, dipeptide and tripeptide (k-mer) frequencies between poly(A) and non-poly(A) interactomes in the ihRBP (see Figure 3E). Top: All possible permutations of the 20 amino acids for each k-mer were counted in proteins of each group and normalized to the total number of counts. Bottom: Control analysis for two groups of the same size but containing randomly selected proteins from the ihRBP.
- D) Control analysis referring to Figure 3F. Scatter plot comparing hexamer frequencies for control groups with randomly selected proteins as described in panel C.
- E) Cumulative distributions of physicochemical properties in the poly(A) and non-poly(A) interactomes (see Figure 3E) and the entire UniProt human proteome (uniprot).

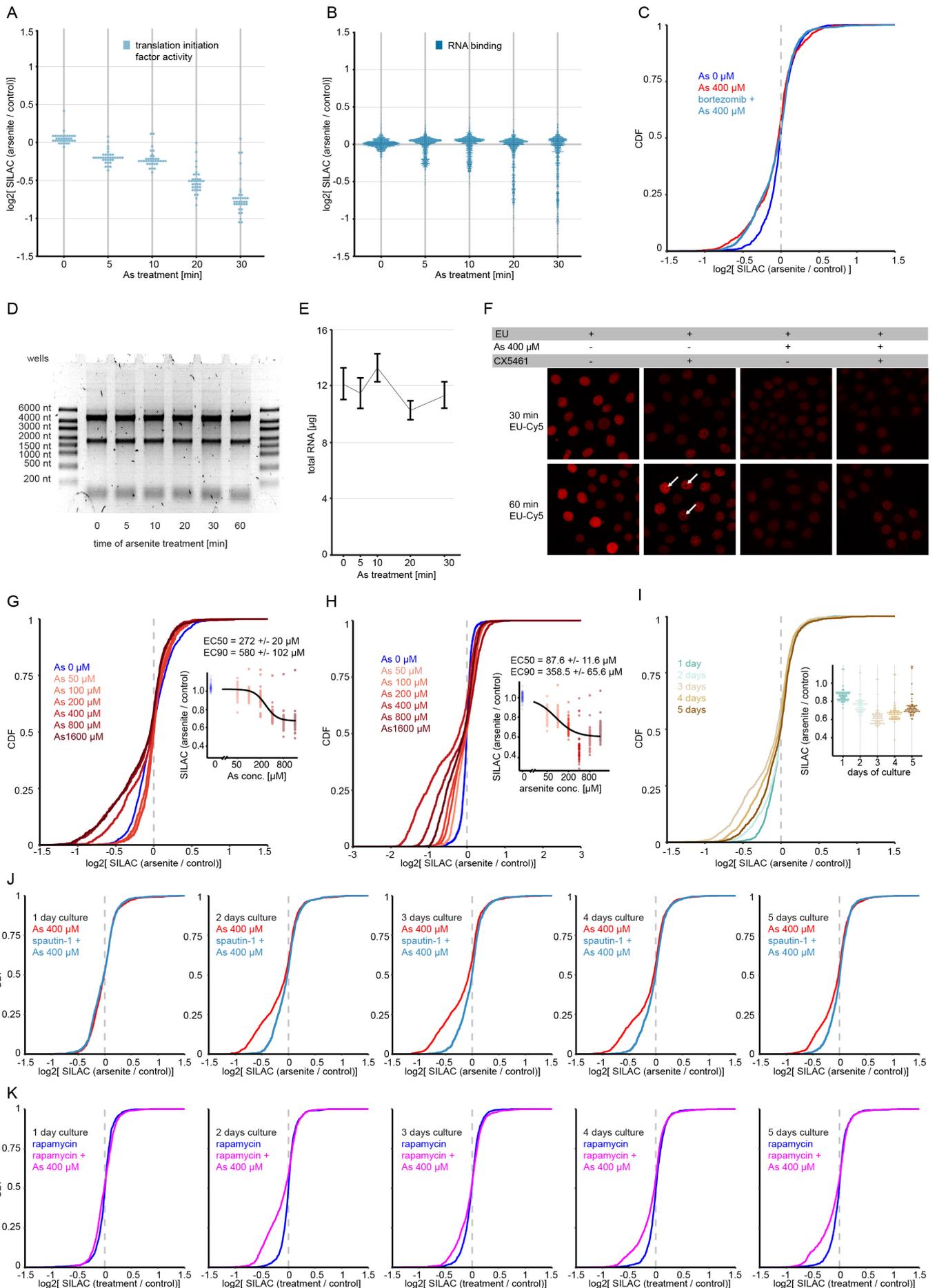
Figure S4: Changes in the total proteome upon arsenite stress.

Adapted from Trendel et al., 2018.⁸²

- A) Dotplots displaying changes in total proteome of MCF7 cells upon arsenite stress for proteins with the GO term 'Translation initiation factor activity'. Each dot represents one protein. Values displayed are means of duplicate experiments with label swap filtered for a variance of 15 % or smaller.
- B) Same as panel A for proteins with the GO term 'RNA binding'.
- C) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment and proteasome inhibition by bortezomib. Cells were treated for 1 hour with 500 nM bortezomib (green) or left untreated (red) before heavy SILAC labeled cells were treated with 400 μ M arsenite for 30 minutes.
- D) Agarose gel electrophoresis of total RNA extracted from MCF7 cells upon 400 μ M arsenite stress for indicated time.
- E) Timeline showing yield of total RNA extracted from MCF7 cells after arsenite stress. Identical number of cells were treated with 400 μ M arsenite for indicated time, total RNA was extracted and quantified using UV-spectroscopy. N=6, error bars indicate standard error of the mean (SEM).
- F) Confocal microscopy of MCF7 cells incorporating EU upon arsenite stress. To \neg exclude the possibility that RNA turnover may be increased upon arsenite stress, ethenyl-uridine (EU) labelling was applied to visualize newly synthesized RNA using click-chemistry. After applying the RNA-polymerase I inhibitor CX5461 discrete areas (white arrows), presumably representing nucleoli, were not stained anymore demonstrating specific labelling of nascent transcripts. Under arsenite stress incorporation of EU was lower, indicating reduced transcription. Exchange of cytosolic RNA with nascent RNA was not apparent during the chosen time window.

Panels G-K are not adapted from Trendel et al., 2018⁸² but are originally presented in this thesis.

- G) Cumulative distribution of changes in total proteome of MCF7 cells upon increasing doses of arsenite and 30 minutes of treatment. Insert: Log-logistic model fit against foldchanges of ribosomal proteins at different arsenite concentrations. Each dot represents one ribosomal protein.
- H) Same experiment presented in G, here presented for HeLa cells.
- I) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment and increasing time in cell culture. Cells were seeded at a density that allowed for full confluency after 5 days. Arsenite treatment occurred at 400 μ M for 30 minutes. Insert: Dot plot showing changes of ribosomal proteins only. Each dot represents one ribosomal protein.
- J) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment and inhibition of autophagy by spautin-1. Cells were treated for 24 hours with 10 μ M spautin-1 (teal) or left untreated (red) before heavy SILAC labeled cells were treated with 400 μ M for 30 minutes. Cells were seeded at a density that led to confluence after 5 days and the same experiment performed after indicated time in culture. Figure 4F displays results shown here for 3 days of culture.
- K) Cumulative distribution of changes in total proteome of MCF7 cells upon rapamycin (blue) or combined rapamycin and arsenite treatment (magenta). Heavy SILAC-labeled cells were treated with 1 μ M rapamycin for 30 minutes, while light cells were left untreated (blue). Cells of both labels were treated with 1 μ M rapamycin for 30 minutes, while heavy SILAC-labeled cells underwent additional 400 μ M arsenite stress (magenta). Cells were seeded at a density that led to confluence after 5 days and the same experiment was performed after indicated time in culture.



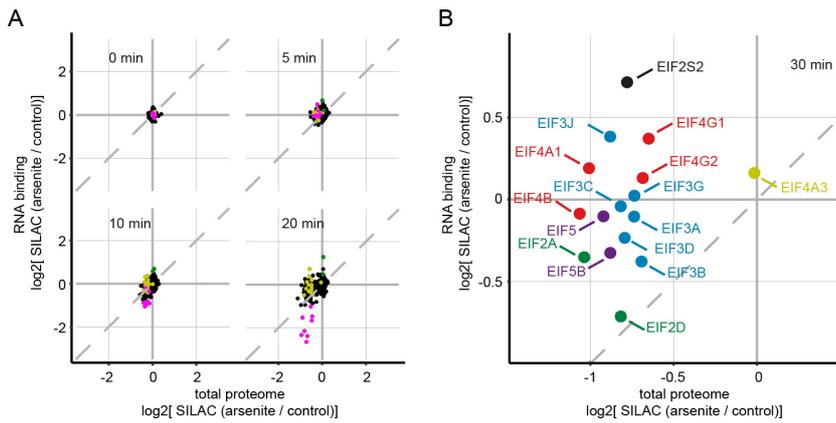


Figure S5: Quantification of changes in RNA-binding proteins.

Adapted from Trendel et al., 2018.⁸²

- A) Scatter plots comparing changes in the total proteome to changes in RNA-binding after 0, 5, 10 and 20 minutes of arsenite stress. See also Figure 5.
- B) Scatter plot comparing RNA-binding versus total proteome changes after 30 minutes of arsenite treatment. Magnification of data in Figure 5E, only showing eukaryotic translation initiation factors (EIFs). Color coding refers to complexes these proteins are known to be part of. Black: EIF2 complex, blue: EIF3 complex, red: EIF4 complex, magenta: EIF5 complex, yellow: exon junction complex, green: auxiliary EIFs without membership in any core complex.

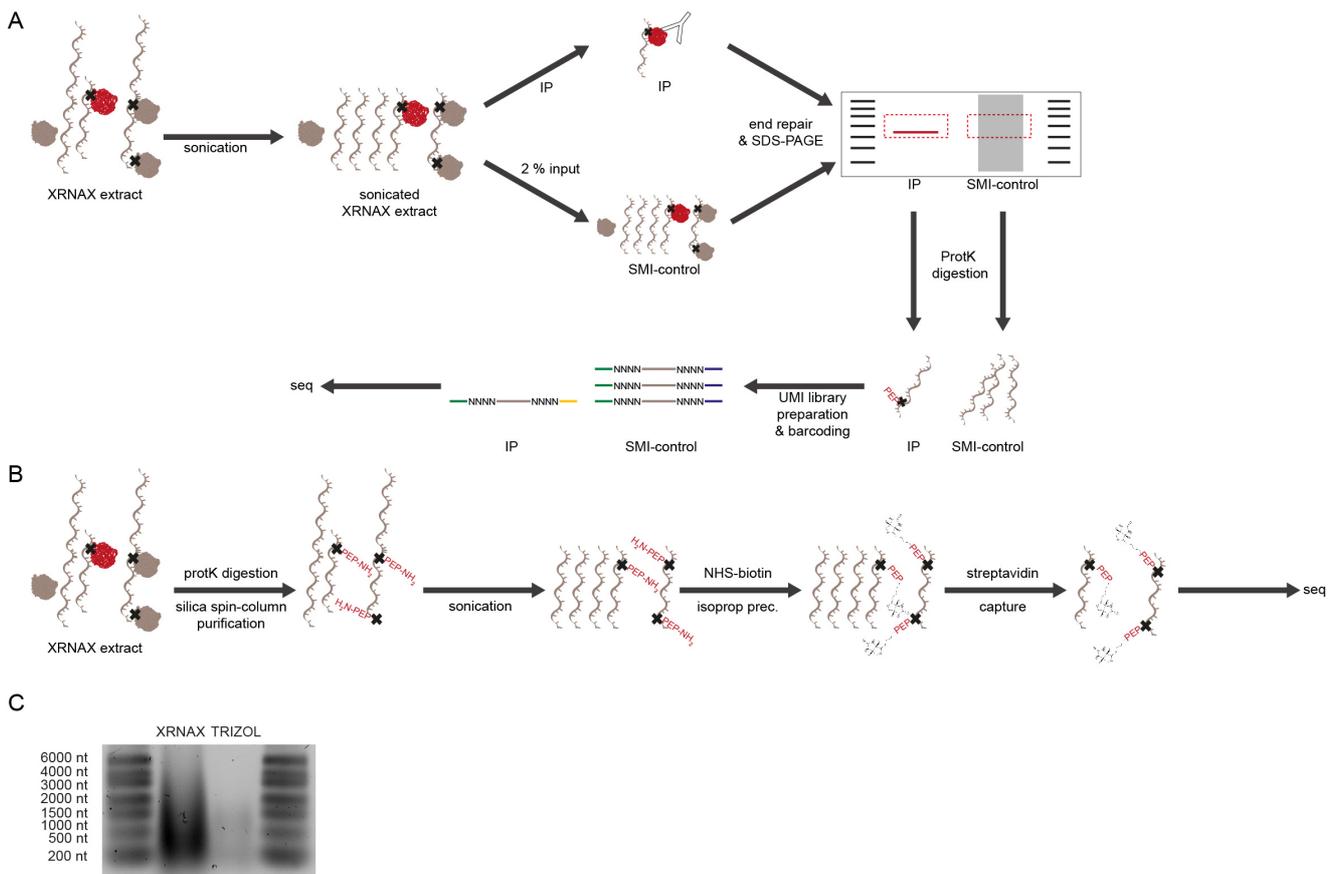


Figure S6: Combination of XRNAX and CLIP-seq.
Adapted from Trendel et al., 2018.⁸²

A) Experimental scheme for an XRNAX CLIP-Seq experiment. After initial fragmentation of RNA through ultrasonication, IP against a protein of interest co-precipitates crosslinked RNA fragments. Both input and IP are resolved on an SDS-PAGE, blotted, and membrane pieces cut out in a region corresponding to the adequate molecular mass. RNA is released by proteinase K digestion and prepared into a sequencing library using conventional small RNA library preparation with unique molecular identifiers (UMIs). For further details refer to method section.

Panels B and C are not adapted from Trendel et al., 2018⁸² but are originally presented in this thesis.

B) Experimental scheme for a PEPseq experiment. An XRNAX extracts is produced from UV-crosslinked cells. The extract is subjected to complete proteolytic digestion with proteinase K and RNA cleaned up using silica columns. Clean RNA is ultrasonicated and the fragments carrying crosslinked peptide remnants reacted to NHS-biotin. Free NHS biotin is removed through alcohol precipitation, biotin-labeled fragments captured on streptavidin beads and free fragments removed by washing. Fragments are released by boiling in formamide, cleaned up with phenol-chloroform extraction and subjected to sequencing.

C) Proof-of-concept experiment for PEPseq. Protein-crosslinked RNA was extracted with XRNAX from UV-crosslinked MCF7 cells (XRNAX) or non-crosslinked RNA with conventional TRIZOL from non-crosslinked cells (TRIZOL). Same amounts (as estimated by UV spectroscopy for RNA neglecting protein in XRNAX extracts) of either extraction was subjected to the protocol presented in B. After elution RNA fragments were resolved on an agarose gel and visualized with SYBRsafe staining. Note that normal sonication (Bioruptor) and not ultrasonication (Covaris) was used, so that the recovered fragments are larger than anticipated for sequencing.

SUPPLEMENTARY FIGURES FOR CHAPTER 2

Figure S10: Comparison of protein half-lives between the total proteome and fractions of polysome profiling.

All color coding refers to the polysome profiling fractions introduced in Figure 9A. Proteins of the small ribosomal subunit are represented by graphical elements without margin, proteins of the large subunit with margin.

- A)-C) Scatterplots comparing half-lives of proteins in the 40S, 60S, 80S or polysome fraction, respectively, and their half-lives in the total proteome. Grey dots represent non-ribosomal proteins.
- D) Scatterplot comparing half-lives of small subunit ribosomal proteins in the 40S fraction to their half-lives in the 80S (left), or polysome fraction (right), respectively.
- E) Scatterplot comparing half-lives of large subunit ribosomal proteins in the 60S fraction to their half-lives in the 80S (left), or polysome fraction (right), respectively.

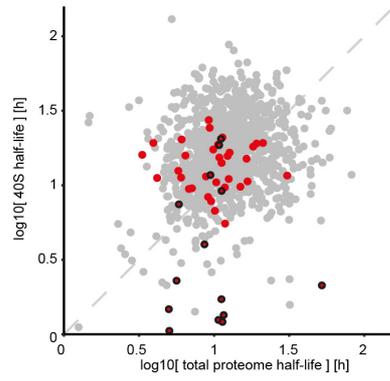
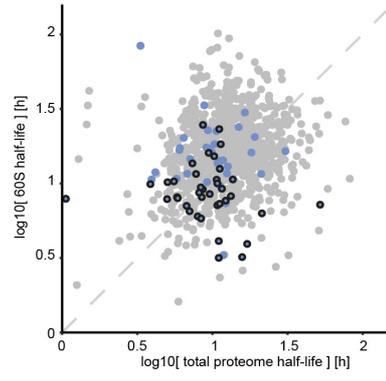
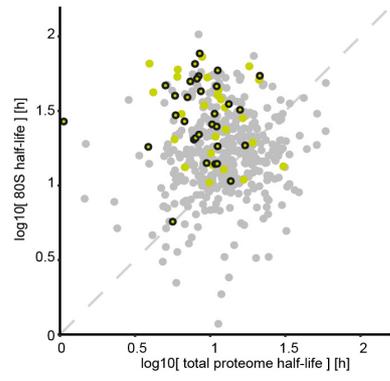
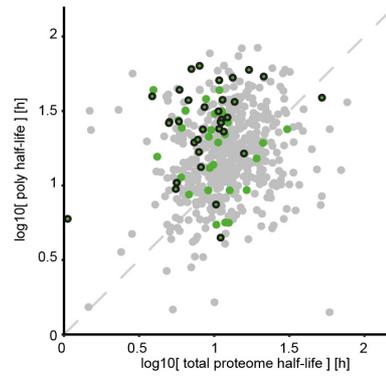
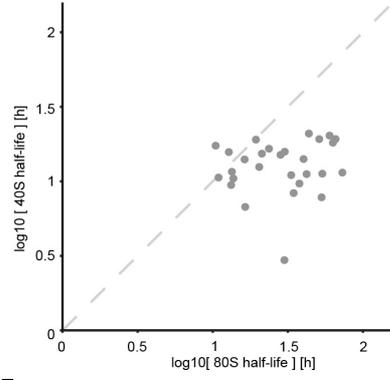
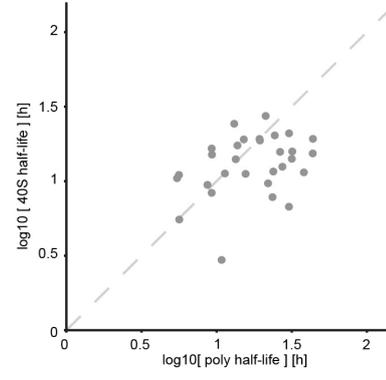
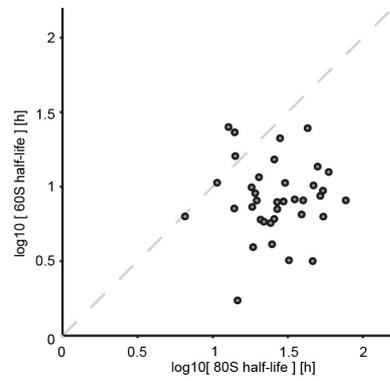
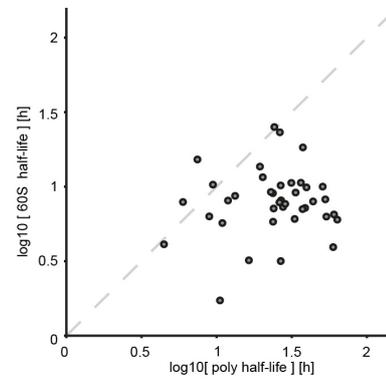
A**B****C****D****E****F****F****F**

Figure S11: Exchange of ribosomal proteins from ribosomal complexes under inhibition of ribosome biogenesis.

Color coding refers to the polysome profiling fractions introduced in Figure 9A. Proteins of the small ribosomal subunit are represented by graphical elements without margin, proteins of the large subunit with margin. Underlined proteins were recurrently quantified in all experiments.

- A) Scatterplots comparing protein half-lives of ribosomal proteins (see Figure 10A) to the relative incorporation of new protein into the 40S, 60S, 80S and polysome fraction under inhibition of ribosome biogenesis (Figure 11A-D). Small and large subunit proteins are not identified individually (no margins).
- B) Combined effect of ribosome biogenesis and autophagy (CX5461 + spautin-1) or the proteasome (CX5461 + bortezomib) inhibition, respectively, on the relative incorporation of new protein into the 80S fraction. Experiments described in Figure 11E&F are compared to each other directly.
- C) Combined effect of ribosome biogenesis and autophagy inhibition on the incorporation of new protein into the polysome fraction. Autophagy was inhibited in MCF7 cells using spautin-1 and the same experiment as described in Figure 11 A-D was performed. The scatterplot compares the relative amount of new protein incorporated into the polysome fraction, when ribosome biogenesis and autophagy are active (control) or inhibited (CX5461 + spautin-1).
- D) Scatterplot comparing the relative incorporation of new protein between the 80S and polysome fractions. Compared are cells under control conditions after release of ribosome biogenesis inhibition.
- E) Same as in D but under continued inhibition of ribosome biogenesis.
- F) Same as in D but under continued inhibition of ribosome biogenesis and inhibition of autophagy.
- G) Same as in D-F but comparing 40S and polysome fractions.
- H) Same as in D-F but comparing 60S and polysome fractions.

