Inaugural dissertation

for

obtaining the doctoral degree

of the

combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht-Karls-University

Heidelberg

presented by Toman Deniz Borteçen

(M.Sc. Molecular Biosciences)

born in Bergisch Gladbach, Germany

oral examination: 17.04.2024

Analyzing the individual contributions of translation initiation factors to shaping the proteome of cancer cells

Referees:

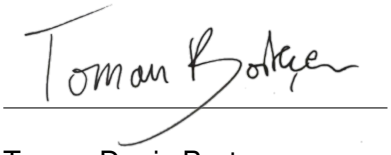
Prof. Dr. Dr. Georg Stoecklin

Prof. Dr. Michaela Frye

Declaration

Herewith I declare that I have written and submitted this dissertation myself and, in this process, have not used any other sources than those indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used this dissertation in this or any other format any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Date: 31.01.2024



Toman Deniz Borteçen

Copyright Disclaimer

The work presented in this thesis was carried out by me, Toman Deniz Borteçen, under the supervision of Prof. Jeroen Krijgsveld. Collaborations and their contribution are specifically highlighted in the main text as well as in the method sections.

Parts of the results outlined in section 5.1, involving the development and optimization of the workflow for quantitative newly synthesized proteome analysis have been published in Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023.
Moreover, a pre-print under the same title was published on bioRχiv.

The results described in section 5.2 and 5.3 have not yet been published and are originally presented in this thesis. However, there are plans to publish these results, irrespective of the embargo period on this thesis, and with significant input from our collaborators regarding the analysis of the acquired data.

1 Abstract

In living cells, proteins are synthesized through the complex multi-step process process of mRNA translation. The initiation step of translation is believed to be rate limiting step of the process and is subject to diverse regulatory mechanisms. A network of proteins, referred to as (eukaryotic) translation initiation factors (eIF), facilitate the recruitment and activation of the translation machinery to initiate protein synthesis. Individual eIF and protein synthesis in general are frequently deregulated in various human cancers to enable aberrant cellular growth and proliferation. Increasing evidence suggests that individual eIF proteins regulate the translation of specific mRNA and do not simply contribute to protein synthesis globally. Changes in protein synthesis are frequently studied using mRNA sequencing-based approaches, however, direct measurements of newly synthesized proteins can be advantageous. However, the majority of established proteomic methods for analyzing changes in protein synthesis have severe limitations.

In this thesis, a new proteomics workflow was developed to simplify and improve the analysis of changes in the newly synthesized proteome across numerous conditions. This workflow was then employed to carry out a systematic analysis of changes in protein synthesis, induced by the depletion of 40 different eIF proteins in a human cancer cell line model. The analysis demonstrated that depletion of different eIF lead to distinct changes in protein synthesis, suggesting that several eIF regulate widespread preferential mRNA translation. Acquired data and findings from the research conducted in this thesis could potentially help to characterize specific signatures of eIF-mediated preferential translation of specific mRNA, which could aid in the development of novel therapeutic strategies.

In summary, this thesis describes an optimized proteomics workflow for the analysis of newly synthesized proteins was developed and employed to carry out a systematic analysis of how individual eIF shape the proteome of cancer cells.

2 Zusammenfassung

Proteinsynthese erfolgt in lebenden Zellen durch einen komplexen mehrstufigen Prozess, der als mRNA-Translation bezeichnet wird. Die Initiationsphase der mRNA Translationunterliegt verschiedenen komplexen Regulierungsmechanismen. Ein Netzwerk von Proteinen, die als (eukaryotische) Translationsinitiationsfaktoren (eIF) bezeichnet werden, regulieren die Rekrutierung und Aktivierung der Translationsmaschinerie zur Einleitung der Proteinsynthese. Einzelne eIF und die Proteinsynthese im Allgemeinen sind bei verschiedenen menschlichen Krebsarten häufig dereguliert, um ein abnormales Zellwachstum und Zellproliferation zu ermöglichen. Immer mehr Hinweise deuten darauf hin, dass einzelne eIF-Proteine die Translation spezifischer mRNA regulieren und nicht außschließlich zur globalen Proteinsynthese beitragen.

Veränderungen in der Proteinsynthese werden häufig mithilfe von mRNA-Sequenzierungsverfahren untersucht, doch direkte Messungen von neu synthetisierten Proteinen können von Vorteil sein. Die meisten etablierten Proteomik-Methoden zur Analyse von Veränderungen in der Proteinsynthese verfügen jedoch über erhebliche Anwendungsgrenzen. In dieser Arbeit wurde ein neuer Proteomik-Workflow entwickelt, der die Analyse von Veränderungen im neu synthetisierten Proteom unter zahlreichen Bedingungen vereinfacht und verbessert. Dieser Arbeitsablauf wurde eingesetzt, um eine systematische Analyse der Veränderungen in der Proteinsynthese durchzuführen, die durch die Deletion von 40 verschiedenen eIF-Proteinen in einem menschlichen Krebszelllinienmodell ausgelöst wurden. Die Analyse zeigte, dass die Deletion verschiedener eIF zu unterschiedlichen Veränderungen in der Proteinsynthese führt, was darauf hindeutet, dass mehrere eIF i großen Maßstäben die Translation spezieller mRNA regulieren.Die in dieser Arbeit generierten Daten und Erkenntnisse könnten dazu beitragen, spezifische Signaturen der eIF-vermittelten präferentiellen Translation bestimmter mRNA zu charakterisieren, was bei der Entwicklung neuer therapeutischer Strategien hilfreich sein könnte.

Zusammenfassend beschreibt diese Arbeit einen optimierten ProteomiK-Workflow für die Analyse neu synthetisierter Proteine, der entwickelt und eingesetzt wurde, um systematisch zu analysieren, wie einzelne eIF das Proteom von Krebszellen prägen.

Contents

3 Introduction

3.1 Mass spectrometry-based proteomics

3.1.1 Fundamentals of proteins and proteomics

Proteins are biopolymers and macromolecules that consist of 20 different alpha amino acids, that are linked by amide bonds of the α-carboxyl and α-amino groups. Differences in the sequence of linked amino acids and chemical modifications of specific amino acids, also referred to as post-translational modifications (PTM), cause proteins to fold into distinct three-dimensional structures which determine their function and activity. Proteins are involved in virtually all biochemical processes of living cells and are synthesized in a process referred to as mRNA translation. The human genome encodes approximately 20000 different proteins [1], which are expressed at varying levels in different cells and tissues [2]. Protein concentrations span a wide dynamic range from several hundred to tens of millions of copies per single cell [3] and spanning up to ten orders of magnitude in body fluids [4]. Through alternative splicing and PTMs multiple different proteoforms can be produced from a single gene, further complicating the determination of the exact number of different proteins within a cell [5].

Proteomics describes the large-scale analysis of proteins through their identification and quantification in complex biological samples. Different technologies and methods are employed in proteomics research, however, mass spectrometry (MS) is the most-widely used technology and enables the most comprehensive analysis of the proteome and its complexity. Mass spectrometry-based proteomics can be divided into two major disciplines. “Top-down” proteomics described the analysis of intact full-length proteins and proteoforms [5][6]. In contrast, “bottom-up” proteomics (also referred to as “shotgun” proteomics) refers to the analysis of peptides, which are produced by intentional hydrolysis of proteins via protease digestion. Identified peptides are aligned to protein sequences and provide evidence to infer the presence and quantity of proteins in bottom-up proteomics analyses. Despite the potential advantage of top-down proteomics, in capturing the true complexity of the proteome with respect to PTMs and the resulting multitude of different proteoforms [7], bottom-up proteomics remains the predominantly applied technology, due to the multiple

technical challenges of intact protein analysis which lead to lower proteome coverage [8].

Proteolysis describes the hydrolytic cleavage of full length proteins into peptides, which are analyzed in bottom-up proteomics workflows. Proteases, a group of enzymes that catalyze the hydrolytic cleavage of proteins, are conventionally used in the preparation of samples for bottom-up proteomics analyses. Ideally, peptides with a length of between 7-35 amino are produced through proteolysis, since peptides with longer sequences may engage in strong non-covalent binding to surfaces, due to high hydrophobicity, and are more challenging to fragment [9], while shorter peptides can often not be aligned to a unique protein sequence. Trypsin is the most-widely used protease, due to the high specificity (for C-terminal cleavage after lysine- or arginine residues) [10], efficiency and relatively low cost. A large number of tryptic peptides (peptides derived from proteins through digestion with trypsin) are below 20 amino acids in length and are well suited for chromatographic separation and analysis by mass spectrometry. However, a large number of peptides with $<$6 amino acids are produced through tryptic digestion, which limits the number of proteins that can be identified. The analysis of multiple samples, generated with the use of different proteases with differing specificity have been used to increase proteome coverage significantly [11][12].

3.1.2 Fundamentals of mass spectrometry

Mass spectrometers are analytical instruments that measure the mass over charge ratio (m/z) of molecules. Due to the high sensitivity of modern mass spectrometers, which is the result of combined progress in the fields of physics, chemistry, engineering and computer science, mass spectrometry has become the most-widely used tool for analysis of the proteome.

Mass spectrometry measurements encompass the generation of gas-phase ions, separation of ions according to their m/z, measurement of the ions m/z and their abundance. Five fundamental components are involved in mass spectrometry measurements. Devices that create gas-phase ions for mass spectrometry are referred to as “ion source“. Different ion sources and mechanisms of ionization are employed in proteomics research, which are briefly described in the next subsection. Separation of m/z ratios of gas-phase ions is

carried out by devices called “mass analyzer”. Measurements of m/z ratios is carried out via “detectors”. Different types of mass analyzers and detectors have been developed and multiple different combinations of both are commonly employed in mass spectrometry based proteomics. The efficient transmission of gas-phase ions in a mass spectrometer requires the use of vacuum systems. Finally, sophisticated control systems coordinate the different elements and their respective control modules to enable efficient end-to-end operation of the mass spectrometer and additional connected devices.

Important metrics, for mass spectrometers in proteomics applications, are resolving power and scan speed. According to the “peak width definition”, Resolution in mass spectrometry is defined as the fraction of the mass and width of the mass spectral peak at full width at half maximum (FWHM). Scan speed describes the number of mass spectra which can be acquired by the instrument per second. High resolving power is required for many proteomics applications due to the acquisition of complex spectra, whereas scan speed is a crucial factor for designing acquisition methods for quantitative analyses and high-throughput applications.

The majority of bottom-up proteomics makes use of “hybrid mass spectrometers”, which combine two mass analyzers. Most commonly, quadrupole mass analyzers (Q) are integrated into hybrid mass spectrometers together with a time of flight (TOF) analyzer (Q-TOF) or Orbitrap analyzers (Q-Orbitrap). As its name suggests, quadrupoles consist of four parallel metallic rods. Radio frequency (RF) voltage with direct current offset (DC) between the two pairs of quadrupole rods enables manipulation of the electrical field within the quadrupole. The manipulations of the electrical field allows for the selection of ions with defined m/z that can travel through the center of the quadrupole. The movement of ions in the electrical field can be described through the Mathieu differential equations [13]. The quadrupole analyzer therefore primarily serves as a mass filter in hybrid mass spectrometers. Time of flight mass spectrometry (TOF-MS), as the name implies, is based on the measurement of gas-phase ion m/z using different flight times in a flight tube. Through acceleration of the ion population with equal initial kinetic energy inputs, ions are separated according to different flight times, where ions with a higher m/z travel slower than ions with

lower m/z [14]. The relationship between flight time and m/z in TOF-MS is described through (1):

$$\frac{m}{z} = 2eEd\left(\frac{t}{l}\right) \tag{1}$$

where: e = charge of the electron [Coulombs]; E = electrical field strength [$Vm^{-}1$]; d = distance from the source (of acceleration) [m]; l = length of the flight tube [m]; t = flight time [s]

Resolution of a TOF analyzer is determined by the error of the flight time measurement. Increasing the flight path length can increase the precision of the flight time measurements, and in turn the resolution of TOF analyzers [14]. TOF analyzers have several strengths, which include high scan speed (up to 300 Hz), wide mass range and relatively high dynamic range. TOF-MS used for proteomics applications, feature a reflectron and achieve relatively high resolution $>$ 20000 [15].

Orbitrap mass spectrometers, include an orbitrap mass analyzer. The orbitrap analyzer consists of an inner spindle-like electrode and an outer barrel-shaped electrode. Trapped ions rotate around the inner electrode with characteristic axial frequencies, which are dependent on their mass and charge state eq. (2):

$$\omega = \sqrt{k\left(\frac{z}{m}\right)} \tag{2}$$

where: m = mass [kg]; z = charge; k = force constant; ω = axial frequency of ions [$s^{-}1$]

In contrast to TOF-MS, where ions hit a detector for measurement, the periodic motion of trapped ions induces an image current (“transient”) in the outer electrode of the orbitrap. Mass over charge ratios are derived from Fourier transformed axial frequency measurements of the trapped ions [16][17]. Orbitrap mass spectrometers can achieve very high resolution $>$480000 but are limited in scan speed. To achieve high mass resolution, high accuracy of the frequency measurements is essential which requires longer transient measurements [18]. In contrast to TOF-MS, the resolution of orbitraps can be changed and adapted for different acquisition methods. Modern orbitrap mass spectrometers, equipped with high

filed orbitrap mass analyzers can acquire spectra at up to 44 Hz, although this relatively high speed is only possible at 7500 resolution (at 200 m/z) [19].

Both orbitrap and TOF mass spectrometers have strengths and limitations, but both provide powerful analytical instruments, that enable the identification and quantification of thousands of peptides and proteins in bottom-up proteomics analyses.

In the last decade an additional component for ion separation has been coupled with, and in several cases even been integrated into, mass spectrometers for bottom-up proteomics applications. Ion mobility separation describes the mechanism, by which gas phase ions are separated according to their interaction with a buffer gas in an electrical field. Apart from charge state, the ion mobility of peptides is determined by size and geometry. Different methods and devices for ion mobility separation have been developed. In bottom-up proteomics the most commonly used ion mobility separation techniques are “trapped ion mobility spectrometry” (TIMS) [20][21] and “high-field asymmetric-waveform ion mobility separation” (FAIMS) [22][23].

3.1.3 Separation and ionization of peptides

The complexity of peptide mixtures, derived from proteolysis of proteins extracted from cells or tissues, poses a challenge to mass spectrometry-based analysis, due to the wide dynamic range of different protein concentrations. Detection of low abundant peptides, and large numbers of peptides in general, therefore requires their physical separation to reduce the complexity of the sample for mass spectrometry measurements. Liquid chromatography (LC) is commonly used to achieve the separation of peptides for bottom-up proteomics analyses. In liquid chromatography a mixture of analytes, dissolved in the mobile phase, is separated through different binding affinities to the solid stationary phase of a column. Pressure is applied to generate a flow of the mobile phase through the column, where analytes that have a high affinity for the stationary phase are retained. Different methods for separating peptides via chromatography have been developed, however, the majority of bottom-up proteomics workflows make use of reverse-phase liquid chromatography (RPLC) and gradient elution [24] [25]. Stationary phases made of superficially porous silica

beads (1-3 μm diameter), with covalently bound hydrophobic C18-alkyl chains, are most commonly used for RPLC-based peptide separation. For gradient elution of peptides in RPLC, the mobile phase consists of water (acting as "weak solvent"), a "strong solvent" with lower polarity (in most cases organic solvent such acetonitrile (ACN) that can be mixed with water) and an acidifier such as formic acid (FA) [26]. The programmed increase of organic solvent in the mobile phase over the course of the chromatographic separation is called "gradient". Peptides can be efficiently separated through the increasing percentage of non-polar organic solvents in the course of the gradient, depending on the strength of their hydrophobic interactions with the stationary phase of the column. Weakly hydrophobic peptides typically have lower retention times (RT) than strongly hydrophobic peptides, which elute towards the end of the gradient. Acidifiers in the mobile phase for RPLC-based separations, affect the separation and retention time stability of peptides by ensuring reproducible protonation states, forming more hydrophobic ion-pairs and by protonating unmodified silanol groups of the stationary phase to prevent ionic interactions with peptides [27]. Numerous additional factors and characteristics of the stationary phase, such as column length and paricle diameter, influence the separation of peptides in RPLC [28] [29]. The performance of RPLC-based separations can be evaluated through different metrics, such as the peak capacity and chromatographic resolution [30][31].

Peptides and any other molecule to be analyzed by mass spectrometry need to be ionized. For the analysis of peptides and intact proteins, "soft" ionization methods are required to prevent unwanted fragmentation. The most widely used soft ionization methods in the field of bottom-up proteomics are matrix-assisted laser desorption (MALDI) [32] and electrospray ionization (ESI) [33]. RPLC-based bottom up proteomics workflows primarily make use of ESI technology for peptide ionization. A high voltage in the range of 1-4 kV is applied to emitter capillary at front of the mass spectrometer. Liquid which exits the emitter form droplets which are charged at the surface. Evaporation of the droplets leads to a concentration of charges which results in "Coulomb fission" once the static repulsion overcomes surface tension [34]. Through the use of acidic mobile phases, the excess of protons produces positively charged peptide ions.

Mass spectrometry-based measurements of peptides frequently suffer from a phenomenon that is referred to as “ion suppression”. Ion suppression describes the reduction or loss signal for low abundant analytes, either due to the presence of molecules that reduce ESI efficiency or highly abundant ions that mask the presence of the low abundant species [35]. Contaminants such as inorganic salts and detergents are known ion suppressants, but also mobile phase acidifier trifluoroacetic acid (TFA) has been shown to act as an ion suppressant by increasing surface tension of aqueous solutions [36] and is therefore avoided for high sensitivity applications.

3.1.4 Tandem mass spectrometry

A fundamental part of bottom-up proteomic analysis is the determination of peptide sequences, through the matching of peptide fragment ions. Intact peptide ions (‘precursors‘) can be fragmented through highly reproducible mechanisms, which enables the determination of peptides peptide sequences via the differences in fragment ion masses. A common nomenclature describing the different types of peptide fragment ions (also referred to as “product ions”) has been developed and widely adopted [37]. The main peptide fragment ions are those that contain the original N-terminus (a-,b-,c-ions) or C-terminus (x-,y-,z-ions) (fig. 1, panel A). Multiple different methods for the fragmentation of peptides have been developed and are commonly employed in bottom-up proteomics. One of the first and most widely used methods is collision-induced dissociation (CID), where collisions with inert gas induce fragmentation of peptides that yield b- and y-ions [38]. Higher energy collisional dissociation (HCD), also referred to as “beam-CID”, is commonly used in orbitrap mass spectrometers and can be carried out in the C-trap. Fragmentation efficiency in HCD is very high, however, low stability b-ions are frequently decomposed so that y-ions are the dominant fragment ion species. HCD fragmentation was also shown to produce immonium ions, which can serve as diagnostic fragment ions for the analysis of PTMs such as tyrosine phosphorylation [39]. Electron transfer dissociation (ETD) relies on the gas-phase reaction of precursor ions with electrons, which primarily yields c- and z-fragment ions. An advantage of ETD is the increased detection of labile PTM, such as phosphorylation which often are retained at a higher rate, compared to CID where neutral loss of phosphate etc is common

(fig. 1, panel B) [40].

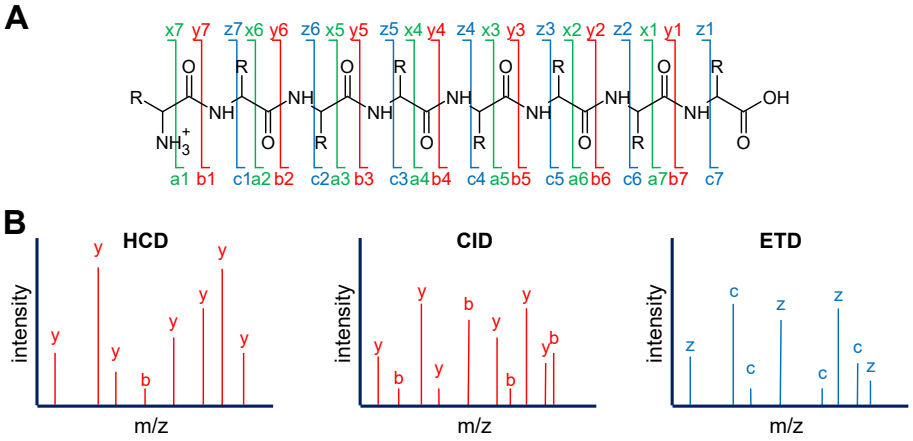


Figure 1: Schematic representation of peptide fragmentation in bottom-up mass spectrometry analyis. A) Sequence informative peptide fragment ions are named a/x, b/y and c/z, depending on the specific bond which breaks in the peptide backbone. B) Schematic depiction of peptide fragmentation spectra, generated with different fragmentation methods. HCD produces primarily y-ions, CID produces b- and y-ions, whereas ETD produces c- and z-ions.

MS acquisition cycles in bottom-up proteomics usually consist of a full scan (also referred to as MS1), in which a wide m/z range of intact precursors are measured, and subsequent fragment ion scans (also referred to as MS2). The width of the quadrupole isolation windows is a crucial factor in tandem MS acquisition. Narrow quadrupole isolation windows ($\leq$2 m/z in Th) enable the isolation and fragmentation of individual precursor ion species, which greatly simplifies the peptide sequence determination from MS2 spectra and increases sensitivity. Wide quadrupole isolation windows result in the isolation and fragmentation of multiple precursor ion species and produces “chimeric” fragment spectra. The De-convolution of highly complex MS spectra, that are obtained from wide isolation window acquisition methods, requires the use of sophisticated software algorithms which use empirical or in-silico predicted information on peptide retention time and fragment ion intensities to determine peptide sequences [41]. The advantage of wide window acquisition lies in the ability to sample a wide m/z range of precursors in a comprehensive manner, which is otherwise impossible using narrow window acquisition methods due to limited scan speed of most mass spectrometers (fig. 2).

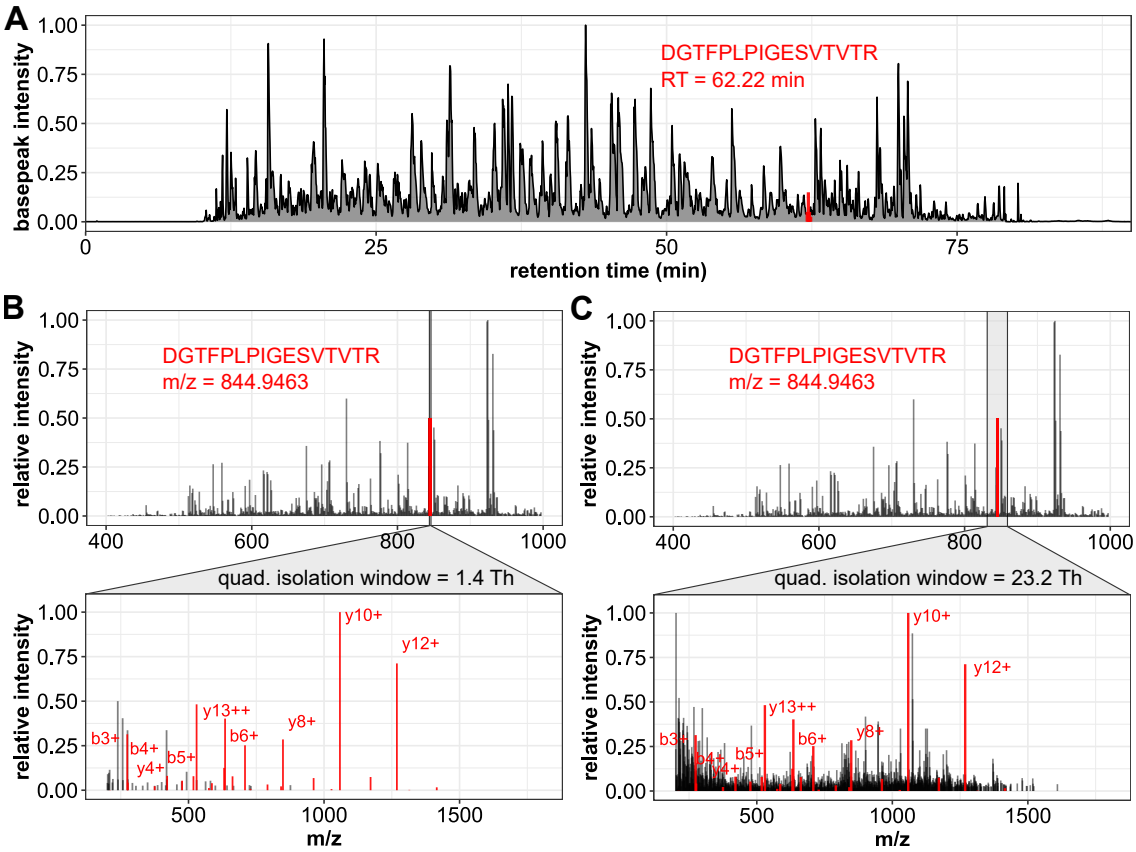


Figure 2: Schematic representation of peptide separation and analysis in bottom-up tandem mass spectrometry. A) Complex proteolytic peptide mixtures are separated using RP-HPLC. Individual peptides, such as the example peptide “DGTFPLPIGESVTVTR”, elute at given retention times in peak shaped elution profiles. Intact precursor ions are measured in MS1 scan. Subsequent MS2 spectra are acquired using fragmented precursors which are isolated from either narrow quadrupole isolation windows (B) or wide isolation windows, which produce chimeric fragment spectra (C). In this example HCD fragmentation was applied, yielding primarily y- and b-ions that are used for the determination of the peptide sequence.

Fine-tuning of different parameters, such as quadrupole isolation window size and number of MS and MS/MS scans, has lead to the development of different types of methods for MS-based data acquisition. These different acquisition methods are employed for various types of bottom-up proteomics analyses.

3.1.5 Mass spectrometry acquisition methods used in bottom-up proteomics

Hybrid mass spectrometers, can be operated using different types of acquisition methods, that enable targeted and un-targeted bottom-up preoteomics analysis.

Data-dependent acquisition (DDA), is based on the sequential isolation and fragmentation

of the “top N” (e.g. top 10) precursors which were measured with the highest intensity in a previous MS1 full scan [42]. Conventional DDA methods feature narrow quadrupole isolation windows to ideally fragment and measure single peptide species in the MS2 scans. Since the majority of DDA-based analyses utilizes MS2 scans exclusively for the identification of peptides, so called “dynamic exclusion” is commonly included in DDA method design to prevent repeated measurement of the same precursors over their chromatographic elution. Due to the lack of repeated MS2 scans over a common m/z, DDA data can only be used for MS1-based quantification. A notable exception, however, are MS2- or MS3-based reporter ion peak height quantification, which is employed in specialized isobaric labeling workflows. Deviations in retention time precursor intensity, at the given MS acquisition cycle, can lead to different identifications across multiple DDA measurements. The semi-stochastic nature of DDA often results in numerous missing values in the produced data. To increase data completeness, software algorithms for the transfer of identifications, also referred to as “match-between runs” have been developed [43][44][45].

Parallel reaction monitoring (PRM) is a targeted proteomics approach in which predefined peptide targets are repeatedly isolated and fragmented using narrow quadrupole isolation windows, Targeted MS2 spectra in PRM are either acquired over the complete method duration, or over scheduled retention time windows of the peptide targets. The latter enables the inclusion of more peptide targets due to minimizing acquisition cycle time [46]. In contrast to other targeted acquisition methods, such as selective reaction monitoring (SRM) where individual fragment ions are selected and measured at a time using triple quadrupole mass spectrometers, in PRM acquisition all fragment ions are analyzed in parallel at high resolution using hybrid mass spectrometers [47].

Un-targeted data-independent acquisition (DIA) methods, in contrast to DDA, feature wide isolation windows, which are predefined to enable the isolation and fragmentation of all precursor ions across the entire m/z range of interest (fig. 3). There have been multiple implementations of un-targeted DIA methods on different hybrid MS platforms [48][49][50]. DIA data commonly contains higher degrees of data completeness than DDA data, but requires the use of highly sophisticated software algorithms for the de-convolution of the

complex chimeric MS2 spectra [41][51][52]. However, modern DIA analysis software frequently enables the identification and quantification of larger numbers of proteins/peptides than un-targeted DDA analysis, explaining the rising popularity of DIA [53].

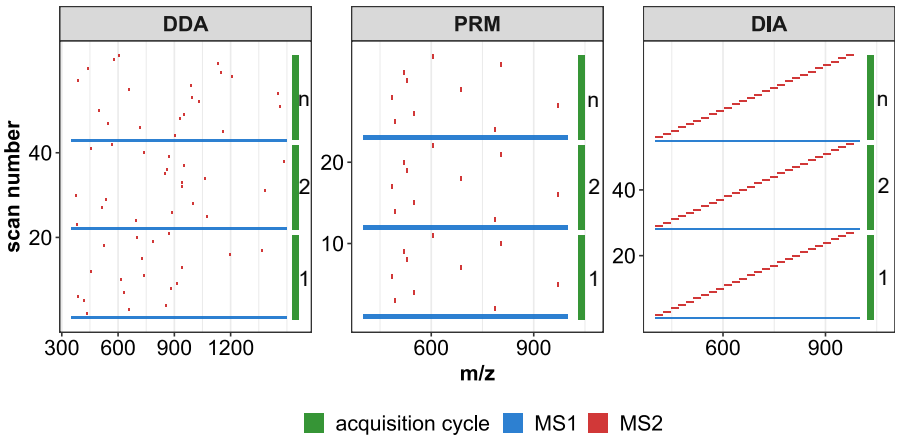


Figure 3: Schematic representation of commonly used hybrid MS acquisition methods, which are used in bottom-up proteomics analyses. A) Top 20 data-dependent acquisition (DDA). Precursors with the top 20 highest intensities are isolated using narrow quadrupole isolation windows and fragmented. Due to dynamic exclusion, different isolation window m/z are employed in each acquisition cycle. B) Parallel reaction monitoring (PRM). In the given example 10 predefined peptide targets are isolated using narrow quadrupole isolation windows and fragmented across multiple acquisition cycles. C) Data-independent acquisition (DIA). Precursors are isolated using wide quadrupole isolation windows (26 in this example), resulting in a mixture of different precursor species being fragmented and analyzed. Through th use of multiple wide isolation windows, all precursors from a large range of 400-1000 m/z are fragmented and analyzed.

Additional ion mobility separation in modern hybrid mass spectrometers adds increased sensitivity to all of the listed acquisition methods and was shown to be highly beneficial [21][23][54].

3.1.6 Mass spectrometry-based methods for peptide quantification

Different quantification methods and strategies have been developed for quantitative bottom-up proteomics analysis and are applied in combination with the different acquisition methods.

Label-free quantification is based on the relative quantification of peptides in separately measured samples and in contrast to label-based quantification does not rely on the addition of stable-isotope labeled amino acids or chemicals. Specialized label-free quantification algorithms have been developed for the normalization of peptide intensity values, that are

derived from separate LC-MS measurements [55][56]. Label-free analysis is compatible with all of the previously described acquisition methods.

Label-based quantification usually features multiplexing of several samples, which can be distinguished through stable isotope compositions by creating distinct m/z shifts for peptide precursors or fragment ions. Two main types of labeling approaches are non-isobaric and isobaric labels. Samples with isobaric- and non-isobaric labels are conventionally analyzed using DDA methods.

Isobaric labeling technology requires the use of chemical labels, which consist of an amine-reactive group (such as an N-Hydroxysuccinimide ester (NHS)), a balancer and reporter group. The most frequently used isobaric tags are TMT [57] and iTRAQ labels [58]. Through the combination of different stable isotope combinations in the reporter and balancer groups of the isobaric tags, multiplexing of up to 18 samples is possible [59]. Fragmentation of precursors in tandem MS also leads to the cleavage of the balancer and reporter group of the isobaric tag, producing reporter ions with defined m/z shifts, whose peak height can be used for relative quantification of the multiplexed samples.

Non-isobaric labels (also referred to as MS1-labels) differ in their number of heavy isotopes and therefore lead to multiplexed precursor and fragment ion spectra through different masses (and m/z). Non-isobaric labeling is carried out either via metabolic labeling of living cells or via chemical labels, coupled to extracted proteins or proteolytic peptides. Stable isotope labeling with amino acids in cell culture (SILAC) [60] as the name implies makes use of the incorporation of stable isotope containing essential amino acids into proteins. Combining cell lysates with different SILAC labels enables the highly accurate relative quantification of protein expression changes. The incorporation of SILAC amino acids into newly synthesized proteins (NSP) of living cells, has enabled the quantitative proteome-wide analysis of protein synthesis and degradation through pulsed SILAC labeling (pSILAC) [61]. The most frequently used non-isobaric chemical labels are amine-reactive di-methyl- [62] and mTRAQ labels [63].

In contrast to isobaric tags, quantification in label-free and non-isobaric labeling methods is

based on the integration of peak areas of the “extracted ion chromatogram” (XIC). XIC elution profiles can be generated from MS1 precursor intensities of the individual isotopomers, or MS2 fragment ion intensities. Co-elution of the fragment- and precursor ions is expected and provides additional confidence for identification and quantification of peptides in tandem MS. Integration of the XIC peak areas, for the calculation of peptide intensities, requires a sufficient number of “data points across the peak” for tracing elution profiles accurately (fig. 4).

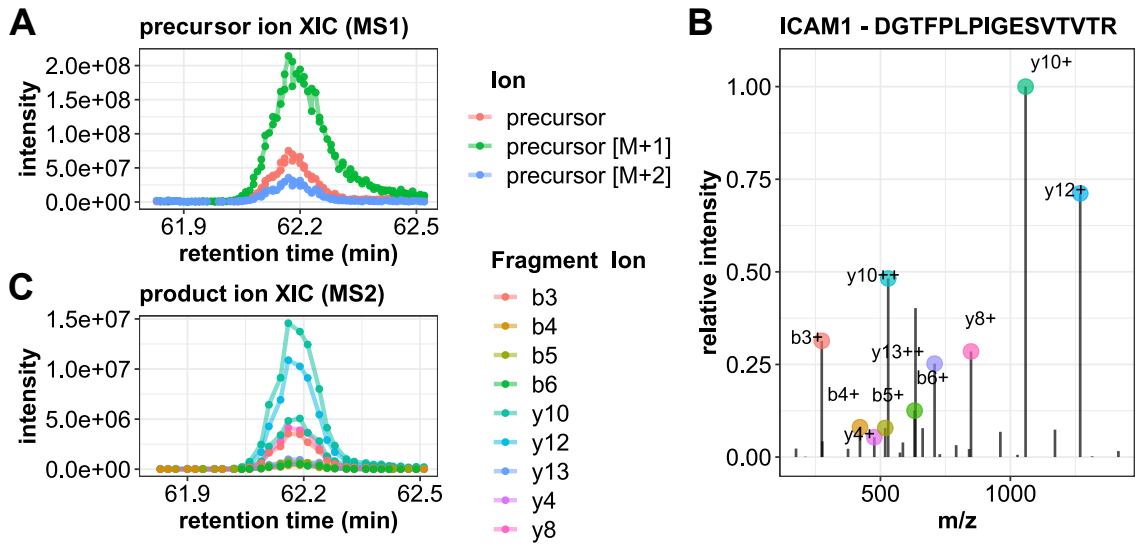


Figure 4: Representation of an extracted ion chromatogram (XIC) of the peptide “DGTFPLPIGESVTVTR”, derived from the human ICAM1 protein. A) MS1 XIC of the doubly charged monoisotopic precursor and two isotopomers. (844.9463 m/z (monoisotopic) 2+ charge). B) MS2 fragmentation spectrum of the doubly charged “DGTFPLPIGESVTVTR” peptide precursor at 27 % HCD collision energy. Colored dots indicate intensities of fragment ions that were identified with high confidence. C) MS2 XIC of the previously highlighted fragment ions. The fragment ions and precursor ions co-elute at the indicated retention time. Each acquired spectra in which the respective precursor- or fragment ion is measured is indicated as a colored dot, representing a data point for the integration of the XIC elution profile.

In published data analyses, MS2-based quantification was shown to achieve higher quantitative precision and sensitivity than MS1-based quantification. However, combinations of MS1- and MS2- quantitative information was shown to simplify the identification and removal of interfering signal and hybrid quantification methods have been shown to outperform both in label free analysis [64][65][66]. Accuracy and precision of MS1- and MS2-based quantification was also evaluated in this thesis.

3.1.7 Multiplexed data-independent acquisition

The high complexity of fragmentation spectra, that are obtained through wide quadrupole isolation windows, pose a major computational challenge for the analysis of samples that were measured via DIA. Multiplexing with non-isobaric labels adds an additional layer of complexity to the de-convolution of chimeric fragmentation spectra, due to the higher number of distinct fragment ions and thus increased probability of interference. However, recent advances in analysis software have demonstrated the feasibility of processing SILAC-, di-methyl- and mTRAQ-labeled samples, while archiving high quantification accuracy and proteomic depth [67][68][69].

The use of DIA could proof highly beneficial for increasing throughput, depth and quantification accuracy in the analysis of SILAC-labeled newly synthesized proteins, and has therefore be evaluated in this thesis. Among the software, capable of processing multiplex DIA data, the recently developed plexDIA features of DIA-NN were shown to achieve similar performance as label-free analyses, due to the implementation of novel algorithms for the processing of multiplexed DIA data [52][69]. These novel software features have therefore been evaluated in the work described in this thesis.

In order to improve quantification accuracy the so called “peak translation algorithm” was developed. This algorithm first identifies the precursor across the different labeling channels which has the highest quality signal and integrates its elution profile for quantification. Subsequently, possible retention time shifts are corrected, to force co-elution of the labeled peptides in the remaining channels and the apex peak ratio of channel-specific fragment ions are calculated and multiplied with the peak area of the reference precursor to calculate “translated intensity” values. Moreover, a decoy-based false-discovery rate (FDR) correction for labeled peptide identification across multiple labeling channels was implemented via addition of an in-silico generated “decoy channel”, to ensure confident identifications in the multiplexed samples [69].

In summary, mass spectrometry-based proteomics is a powerful technology that can be applied for quantitative measurements of thousands of peptides and proteins in complex mixtures. Advances in instrumentation, analysis software and development of acquisition

methods continue to increase the ability of bottom-up proteomics to uncover the complexity of the proteome, for the study of cellular mechanisms and pathological conditions. More specifically, the application of pulsed SILAC labeling-based methodologies has proven to be a powerful tool for the analysis of cellular protein synthesis, which has therefore also been in the focus of this thesis.

Before describing technical developments and the application of specialized methods for the analysis of changes in protein synthesis, that are covered in this thesis, a brief introduction on gene expression and the regulation of protein synthesis will be given in the next subsection.

3.2 Initiation of mRNA translation in human cells

The process of gene expression consists of two major steps. During the initial step, referred to as “transcription”, genetic information encoded in DNA is transferred to a messenger RNA (mRNA). “Translation” describes the second major step in gene expression in which template mRNA sequence information is decoded by a multitude of factors, to produce a polypeptide chain (protein). The different macromolecules, including ribosomes, transfer RNA (tRNA) and initiation- and elongation factor proteins, are involved in mRNA translation and are partaking in multi-layered regulatory mechanisms, controlling the synthesis of proteins. The process of mRNA translation can be divided into four main phases: initiation, elongation, termination and recycling of ribosomes.

The process of translation initiation is believed to be the rate limiting step of the process. The estimated rate of initiation in eukaryotic cells lies between 0.5-3.6 s^{-1}, where as the average rate of elongation is in the range of 3-10 amino acids s^{-1} [70][71][72]. In eukaryotes, initiation of translation is facilitated by a group of proteins referred to as eukaryotic translation initiation factors (eIF), which form complexes and recruit ribosomal subunits to the template mRNA. There are $>$40 eIF proteins sub-units and numerous eIF-interacting proteins which have been studied extensively over the last decades. The complexity of eukaryotic translation initiation is also exemplified by the fact that there is not a single but multiple different mechanisms of initiation that are invoked depending on the cellular state [73]. However, under non-stressed conditions, the predominant mechanism of translation initiation is the so

called “canonical (5′-)cap-dependent translation initiation” and has been studied extensively (fig. 5, panel A) [74].

3.2.1 Canonical 5′-cap-dependent translation initiation

At the start of the cap-dependent initiation, the 40S ribosomal subunit is bound by eIF1, eIF1A, the eIF3 sub-complex, the ternary complex, which consists of the eIF2 proteins eIF2S1-3, Guanosine-5′-triphosphate (GTP) and methionine bound transfer RNA (tRNA), and eIF5. Together these components form the so called “43S pre-initiation complex” (43S-PIC) [74]. The initiation factors eIF1 and eIF1A stabilize the 40S ribosomal sub-unit and enable correct binding of the remaining factors for the 43S-PIC formation. Initiation factor sub-complex eIF3 is a highly complex structure, which in humans and other mammals consists of 13 eIF3 protein sub-units (eIFA-M) (fig. 5, panel B). The protein sub-units eIF3A and eIF3C are crucial for the recruitment of the 43S-PIC to mRNA and for the translation start site selection. They are therefore often referred to as the “core” components of eIF3. Sub-units eIF3B eIF3G and eIF3I have been shown to also be involved in the recruitment of mRNA and later scanning of the 5′-UTR [75][76]. An additional role of the eIF3 sub-complex is the stabilization of the ternary-complex and 43S-PIC [77] and facilitation of ribosomal recycling following translation termination [75]. Interestingly, individual eIF3 sub-units were also reported to be involved in other mechanisms of non-canonical translation initiation [78][74].

In parallel to the 43S-PIC formation, the 5′-cap of mRNA is bound by the eIF4F complex which consists of 3 sub-units cap-binding protein eIF4E, DEAD-box RNA-helicase eIF4A and the scaffold protein eIF4G. There are multiple known paralogues that can constitute the eIF4F protein complex in humans. These include eIF4G1, eIF4G2 (also referred to as DAP5), eIF4G3, eIF4A1, eIF4A2, eIF4E, eIF4E1B, eIF4E2 and eIF4E3. Several of the listed protein paralogues are believed to have separate functions, since they do not interact with all components of the eIF4F sub-complex. EIF4G2 for example lacks eIF4E and PABP interaction sites but interacts with eIF4A and eIF3 [79]. It has been proposed that eIF4G2 is involved in the regulation of cap-independent translation initiation mechanisms, such as internal ribosome entry site (IRES)-mediated initiation [80]. Several other proteins bind

and interact with components of the eIF4F complex. EIF4B and its homolog eIF4H bind eIF4A and have been reported to modulate helicase activity [81]. The interaction of the eIF4F complex with poly-A binding protein (PABP) at the 3′-untranslated region of the mRNA (3′-UTR) creates circular structures which are also referred to as "activated mRNA" [82][83]. Binding of the 43S-PIC to the eIF4F complex on activated mRNA leads to the formation of the "48S complex". Upon formation, the 48S complex scans the 5′-UTR of the activated mRNA in an Adenosine-5′-triphosphate (ATP)-dependent manner until the start codon sequence (AUG) is located [84]. Start codon recognition is sequence dependent and usually occurs in context of a "Kozak sequence" motif (5′-A/G-NN-AUG-G-3′) [85]. In the final step of translation initiation the ribosomal 60S sub-unit joins with the 40S sub-unit in the 43S-PIC, leading to the formation of the "80S initiation complex". The ribosomal subunit joining is regulated by initiation factor proteins eIF1A and eIF5B [86]. During the elongation process a nascent polypeptide chain is generated by the 80S ribosome, which is moving along the coding sequence (CDS) of the template mRNA until termination of translation occurs at a stop codon sequence (UAA, UAG, UGA). Proteins involved in the regulation of translation elongation, through aminoacylated tRNA delivery and translocation, are referred to as (eukaryotic) elongation factors (eEF) [87]. Despite its name eIF5A acts as an elongation factor, which in its hypusinated form prevents ribosomal stalling at repeating CCA codon sequences, encoding poly-proline tracts in proteins [88][89]. The eIF3 subcomplex was also reported to remain associated with the elongating ribosome during the initial phase of elongation [90][91].

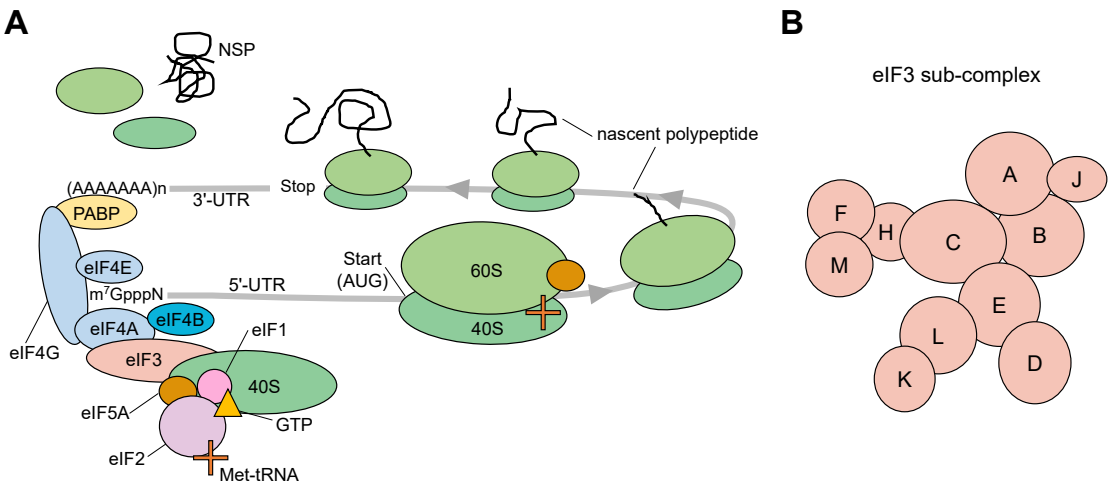


Figure 5: Regulation of mRNA translation initiation by eukaryotic translation initiation factor (eIF) proteins. A) Schematic representation of the canonical 5′-cap-dependent initiation of mRNA translation. B) Schematic representation of the structure of the human eIF3 sub-complex and it 13 sub-units (eIF3A-M).

3.2.2 (De)regulation of translation initiation

Several key signaling pathways involved in cellular metabolism, stress and proliferation such as the mTOR-, the integrated stress response (ISR) and MAPK pathway, regulate the process of translation initiation in response to intracellular and extracellular signals [92][93]. Multiple different components of the eIF complexes, especially the eIF2 and eIF4F, are subject to regulatory mechanisms. Phosphorylation of eIF2S1 lies at the basis of the ISR pathway and inhibits dissociation from the interaction with the guanine nucelotide exchange factor (GEF) protein complex eIF2B, which consists of 5 sub-units (eIF2B1-5), limiting ternary complex levels [94]. PDCD4, was reported to bind and inhibit helicase activity of eIF4A, acting as a tumor supressor. PDCD4 is phosphorylated by the protein kinase S6K1, triggering the degradation of PDCD4 in response to mitogenic stimuli [95]. Cap-binding initiation factor eIF4E can be phosphorylated at Serine position 209 by the kinase MNK1 and MNK2 which reduces the cap-binding affinity and inhibits mRNA activation through eIF4F [96][97]. Moreover, a group of eIF4E-binding proteins (4E-BP) binds eIF4E, reducing cap-binding of eIF4E and thus inhibiting canonical translation initiation. 4E-BPs are phosphorylated and inhibited by the anabolic kinase mTORC1 kinase. Inactivation of mTORC1 in response to low growth factor-, ATP-, amino acid and other nutrient availability, leads to activation of 4E-BP and leads to inhibition of canonical translation initiation [98][99].

Protein paralogues of eIF4E such as eIF4E2 have also been shown to regulate cap-dependent translation initiation through competing with eIF4E for the binding to 5′-cap structures. In contrast to eIF4E, eIF4E2 does not interact with eIF4G and thus inhibits eIF4F formation. However, the measured 5′-cap-binding affinity of eIF4E2 is approximately 100-fold lower than eIF4E. It has been postulated that it is responsible for regulating the translation of specific mRNA subsets [100].

Due to the inhibition of canonical 5′-cap-dependent translation initiation upon cellular stress, alternative cap-independent translation initiation mechanisms have been reported to play a crucial role in the cellular stress-responses. Interestingly, it has been speculated that these alternative initiation mechanisms are responsible for the translation of specific mRNA subsets. Indeed, selective translation of several regulatory proteins through cap-independent initiation mechanisms have been described in recent years [101][73]. IRES-mediated translation is driven by cis-acting RNA elements, which contain complex secondary RNA structures in the mRNA, and so called IRES trans acting factor (ITAF) proteins. IRES features were intially discovered in viruses, however, several IRES elements have also been described in human transcripts [102][103]. The initiation factor protein eIF4G2 has been reported to act as an ITAF [104]. N^6-methyladenosine (m^6A), the most frequent internal modification of mRNA, was shown to not only regulate mRNA stability via the 3′-UTR, but also to serve as a regulatory factor for cap-independent translation initiation. Components of the eIF3 sub-complex, more specifically the sub-units eIF3A and eIF3B were shown to be crucial factors in m^6A-dependent translation initiation [105]. The initiation factor protein eIF3D was also determined to be a 5′-cap binding protein, which regulates translation initiation for specific subsets of mRNA together with eIF4G2 [106][107]. Mechanisms of eIF2-independent translation initiation have also been described for a number of mRNA. It has been speculated that upon phosphorylation and inhibition of eIF2, in response to stress, non-canonical initiation factors such as eIF2D and eIF2A substitute eIF2(S) for translation initiation [108]. Moreover, eIF2D has been reported to play a role in the re-initiation of translation following the translation of up-stream open reading frames (uORF) [109]. The presence of short 5′-UTR sequences and so called “translation initiator of short 5’ UTR” (TISU) sequence motifs (5′-SAASATGGCGGC-3′, in which S is C or G) have also

been reported as potential regulatory factors that can induce eIF4-independent initiation of translation [110].

Deregulation of mRNA translation and protein synthesis has frequently been observed in multiple diseases [111]. Especially in the context of cancer, aberrant protein synthesis contributes to the ability of cancer cells to proliferate and invade other tissues [112]. Different mechanisms of deregulated translation initiation have been described in various cancer entities and have fueled the development of novel therapeutic agents [113]. In general, the different stresses that cancer cells experience in the tumor enviroment, such as hypoxia and nutrient deprivation are believed to active cap-independent translation initiation mechanisms that regulate synthesis of specific proteins with functions in cell survival and proliferation [73]. For example, activity and abundance of the eIF4F complex is frequently up-regulated in human cancers [114]. The protein PDCD4, an inhibitor of eIF4A helicase activity, acts as a tumor suppressor and was reported to be down-regulated in several cancer sub-types, including lung- and breast cancer [115][116]. Increased activity of the PI3K-Akt-mTORC1 pathway is a common feature of multiple cancers, which leads to phosphorylation of PDCD4 and eIF4E-binding proteins [112]. Initiation factor proteins eIF1, eIF2B, eIF5A and eIF6 have also been reported to be up-regulated in distinct cancer entities [117]. Moreover, multiple sub-units of the eIF3 complex are frequently deregulated in different cancer sub-types. Interestingly, both up- and down-regulation of eIF3 proteins has been observed in distinct cancer entities, further indicating that they potentially have additional functions outside of canonical 5′-cap-dependent translation initiation or in facilitating the translation of specific mRNA subsets via unknown determinants [118].

Increasing evidence points to the ability of individual eIF proteins and their sub-units to regulate the translation of specific mRNA subsets. It remains unclear, however, to which extent parts of the proteome are regulated through “selective mRNA translation” and by what mechanism mRNA subsets interact with individual eIF proteins.

3.3 Analytical methods for measuring protein synthesis

Cells respond to external stimuli and stress by remodeling their proteome, through protein synthesis and degradation, which drives a change of the cellular phenotype. Analyzing changes in protein synthesis can provide a snapshot of the initial proteome remodeling process, before overall changes in protein abundance or cellular phenotype are apparent, enabling the differentiation of primary and secondary responses. Multiple different methodologies have been developed to measure protein synthesis. These methods rely on different technologies and perform measurements of either the translated mRNA, the nascent polypetides or newly synthesized proteins (NSP) [119].

The process of protein synthesis via mRNA translation marks a crucial intersection between nucleic acids and proteins. Therefore, high sensitivity genome-wide measurements of mRNA are frequently employed to gain insights into changes in gene expression and to infer changes in protein synthesis. However, increasing evidence has revealed that the measured correlation between mRNA and protein abundance is relatively low [120][121]. Ribosome profiling methodologies are based on the sequencing of actively translated mRNA, which is bound by ribosomes [122]. Ribosome occupancy data is often combined with global RNA sequencing data to calculate translation efficiency (TE) [123]. Among the main advantages of ribosome profiling are single codon and nucleotide resolution and the possibility to identify transcript variants as well as upstream and downstream open reading frames (u/dORFs) [122]. However, ribosome profiling methods perform indirect measurements by analyzing mRNA templates and not the NSP product. In addition, for the analysis of cellular perturbations, sophisticated analysis is required to calculate changes in translation efficiencies, by discarding inactive ribosomes [124][125].

Analysis of nascent polypeptides and NSP directly measures the intermediates and products of protein synthesis. These analyses are either carried out via fluorescence-based measurements of bulk NSP amounts per cell, or via mass spectrometry-based bottom-up proteomics workflows that enable the identification and quantification of thousands of NSP. Measurement of NSP and nascent polypetides requires metabolic labeling to enable the distinction of otherwise chemically identically pre-existing proteins. Several metabolic labeling

strategies have been developed and are based on the incorporation of either Puromycin (or derivates of Puromycin), non-canonical amino acids, stable-isotope containing amino acids (SILAC) and combinations thereof.

Pulsed SILAC (pSILAC) labeling enables relative quantification of NSP, which can be distinguished from pre-existing proteins using high resolution mass spectrometry [61].

Puromycin is a naturally occuring aminonuclesoide antibiotic which resembles the 3′-end of aminoacylated tRNA linked to tyrosine and is non-specifically incorporated into elongating nascent polypetide chains leading to termination of translation [126]. Metabolic labeling with Puromycin(-derivates) is usually carried out in short time frames to avoid cellular stress and reduced viability. Truncated Puromycin-containing polypepides are rapidly cleared by the protein degradation machinery of the cells to avoid proteotoxic stress [127].

Non-canonical amino acid (NAA) labeling methods commonly employ azide or alkyne containing amino acids, such as the methionine analogues L-azidohomoalanine (AHA) [128][129], L-azidonorleucine (ANL) [130], L-homopropargylglycine (HPG) [128] and threonine analogue L-β-ethynylserine (β-ES) [131], which are incorporated into full length proteins that can be conjugated using click chemistry. Non-canonical amino acids are incorporated a lower rate than their natural counterparts, causing their toxicity at high concentrations and over prolonged exposure [128][132]. Metabolic labeling periods with NAA are therefore also usually kept relatively short.

For microscopy or flow cytometry-based analysis of bulk protein synthesis, O-propargylpuromycin (OPP)-containing truncated polypeptides and NAA-containing proteins can be coupled to azide/alkyne-conjugated fluorophores. Quantitative proteomic analysis of NAA-labeled NSP and puromycin-containing polypeptides can be carried out using several different methodologies that have been developed in recent years [119][133]. The majority of these proteomic methods rely on the enrichment of labeled NSP.

Puromycin-containing polypeptides can be enriched using biotin-Streptavidin affinity purification [134], click chemistry-based coupling [135][136], or through use of Puromycin-directed antibodies [137]. Similarly, NAA-labeled NSP can be enriched through click chemistry-based

coupling to azide/alkyne beads [138][131], coupling to biotin-conjugates and subsequent Streptavidin pull-down [139][140] and phosphonate alkynes [141]. Strategies combining metabolic labeling with NAA and SILAC amino acids do not only profit from highly accurate relative quantification of NSP, but can distinguish SILAC-labeled NSP from pre-existing unlabeled proteins, that are retained in small quantities even with stringent washing [142][143]. Newly synthesized proteome analysis have been performed in a range of model systems ranging from cultured cells [142], primary cells [139], tissue samples [144] and in-vivo [145]. It has been shown in a study describing activation of macrophages, that changes in NSP abundance could be detected in shorter time scales using an enrichment-based workflow compared to pSILAC-based analysis without NSP enrichment, highlighting a gain in sensitivity via NSP enrichment [146].

Despite the multiple advantages of enrichment-based newly synthesized proteome analysis, several major limitations exist and hinder its routine application in research. Established protocols require relatively large amounts of input material for the enrichment of NSP, ranging from $\geq$500 μg to several mg protein [141], which are harvested from several million cultured cells [139][134]. In addition to large-scale cell culture work, manual multi-step sample preparation limits the throughput of newly synthesized proteome analysis. Two automated enrichment protocols have been developed recently, which however still have large input requirements ($\geq$500 μg protein or cells grown in 15 cm dishes), include lengthy off-deck dephosphorylation and dialysis steps and report relatively low proteome coverage [141][147]. Furthermore, the vast majority of analysis workflows make use of data-dependent acquisition (DDA) methods that in order to increase proteomic depth require offline fractionation or long HPLC gradients, leading to increases in measurement times and decreased throughput [132][143][148].

An improved newly synthesized proteome analysis workflow with lower input requirements, automated sample processing, high LC-MS measurement throughput and proteomic depth would enable the systematic analysis of changes in protein synthesis across numerous conditions or time points, which would be extremely challenging and demanding using established protocols.

3.4 Objectives

Direct measurements of protein synthesis, via newly synthesized proteome analysis, offers distinct advantages over indirect measurements of transcript levels and ribosome occupancy in the context of investigating rapid cellular proteome remodeling and studying regulatory mechanisms of mRNA translation. Multiple bottom-up proteomics workflows have been developed in recent years, which specifically quantify newly synthesized proteins. However, the majority of these established protocols suffer from one or multiple severe limitations, which prevent their use in large-scaled analyses across numerous experimental conditions. Therefore, one of the main objectives of this thesis was the development of optimized sample preparation and tandem mass spectrometry methods for the quantitative analysis of newly synthesized proteins.

Secondly, newly synthesized proteome analysis was applied to study regulatory factors of mRNA translation initiation in a systematic manner. Specifically, increasing evidence indicates that several eIF regulate the translation of specific mRNA subsets. However, it remains largely unclear to what extent and by what mechanism individual eIF proteins regulate the translation of specific mRNA subsets. To determine the contributions of individual initiation factors to shaping the newly synthesized proteome, a systematic analysis with induced depletion of 40 different factors was performed. Finally, measured changes in protein synthesis were analyzed to obtain potential clues towards regulatory mechanisms of selective mRNA translation, driven by individual eIF proteins.

4 Materials and Methods

Several different methodologies were applied in the course of the project, which are described in the following sections and subsections.

4.1 Cell culture

Hela cells were grown in DMEM high glucose medium (Gibco) supplemented with 2 mM L-glutamine, 10 % (v/v) fetal bovine serum (FBS) (Gibco) and additional 2 mM GlutaMAX (Gibco). For the interferon gamma (IFNg) stimulation experiments, Hela cells were treated with 10 ng/mL recombinant IFNg (Cell signalling), diluted in 0.5 % (w/v) bovine serum albumin (Serva). For the preparation of the SILAC benchmark samples, Hela cells were grown in high glucose DMEM, with the previously listed supplements, and heavy- ($^{13}C_6{}^{15}N_4$-Arg, $^{13}C_6{}^{15}N_2$-Lys), intermediate ($^{13}C_6$-Arg, D_4-Lys) or light isotope-containing Lysine, Arginine for 10 days. Hela cells were grown in 15-cm dishes for the preparation of newly synthesized proteome samples used for protocol optimization. For the SILAC benchmark sample preparation, the cells were grown in 10-cm dishes and for the preparation of the samples of IFNg treated Hela cells the cells were grown in 6-well plates. RKO cells were grown in RPMI 1640 medium, supplemented with 4 mM GlutaMAX and 10 % FBS. Upon initial thawing, the RKO cells were expanded with culture medium, supplemented with 500 μg/ML Geneticin (G418). Induction of the spCas9-mediated KO of the targeted EIF genes and AAV1 control was carried out by addition of 200 ng/mL doxycycline (dox). For the preparation of samples for newly synthesized proteome-, proteome- and RNA sequencing analysis the RKO cells were seeded in 6 well plates, 24 h prior to the start of the 48 h dox induction.

4.1.1 Cell viability assay

Cell viability assays were performed using the CellTiter Glo-2.0 kit (Promega), according to the manufacturers instructions. 1000 RKO cells per well were seeded in white plastic 96-well plates (Nunclon, Thermo Fisher Scientific). Cellular ATP-content was measured 0 h, 24 h, 48 h and 72 h after seeding using a 96-well luminescence plate reader (Mithras LB 940, Berthold Technologies).

4.1.2 Metabolic labeling of newly synthesized proteins

A metabolic labeling approach, combining pulsed stable isotope-labeling (pSILAC) and L-Azidohomoalanine (AHA)-based labeling of newly synthesized proteins was used for the analysis of the newly synthesized proteome [138]. Prior to the labelling step, the cells were washed with warm PBS and incubated with DMEM high glucose medium deprived of Methionine, Arginine and Lysine for 45 min. The pulsed SILAC and AHA labeling was carried out with either Methionine-free DMEM high glucose medium (for HeLa cells) or RPMI 1640 medium (for RKO cells), containing heavy- ($^{13}C_6{}^{15}N_4$-Arg, $^{13}C_6{}^{15}N_2$-Lys) or intermediate ($^{13}C_6$-Arg, D_4-Lys) Lysine, Arginine and 100 μM AHA. The labeling period for the initial experiments with HeLa and RKO cells ranged from 2 h, 4 h or for a maximum of 6 h (section 5.1 and 5.2), whereas a uniform 4 h labeling duration was used for the RKO cells in the final experiments (section 5.3). The specific labeling periods for the initial method developments and optimizations are specified in the respective sections that are describing the experiments.

4.2 Newly synthesized proteome analysis

4.2.1 Preparation of magnetic alkyne agarose beads

Epoxy-activated magnetic agarose beads (Cube Biotech) were coupled with propargylamine (Santa Cruz Biotechnology) to produce magnetic alkyne agarose (MAA) beads. 5 mL epoxy-activated magnetic agarose beads were washed with 10 mL milliQ water and resuspended in the coupling solution of 1 M propargylamine in 0.5 M di-potassium phosphate solution (pH 10.5). The handling of the propargylamine and coupling solution were carried out under a fume hood with appropriate safety precautions. The beads and coupling solution were incubated in a thermo shaker at 45 °C for 16 h, and beads were then washed with 25 mL milliQ water. To ensure the complete quenching of remaining epoxy groups on the beads, they were incubated with 1 M Tris-HCl buffer (pH 8.0) for 4 h. The beads were subsequently washed with 45 mL milliQ water and stored in 20 mM sodium acetate buffer (pH 6.5) with 20 % ethanol at 4 °C. The MAA beads are stable at 4 °C for multiple months.

4.2.2 Cell lysis and ultrasound sonication

Two different lysis buffers were used for the preparation of samples for proteome and newly synthesized proteome analysis. In the initial stages of the project, during the early developments of the workflow for newly synthesized proteome analysis, lysis buffer containg 5 % (w/v) 3-((3-Cholamidopropyl) dimethylammonio)-1-propansulfonat (CHAPS), 750 mM NaCl, 200 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES) (pH 8.0) and cOmplete EDTA-free protease inhibitor cocktail (Merck) was used. Lysis buffer containing 1 % Sodium-dodecylsulfate (SDS), 300 mM HEPES (pH 8.0) and cOmplete EDTA-free protease inhibitor cocktail (Merck) was used for the preparation of newly synthesized proteome and full proteome samples of HeLa and RKO cells.
Two different methods of sonication for DNA fragmentation were used, depending on the starting amount of sample material. Cell lysates form cells, grown in 10-15 cm dishes, were sonicated with a probe sonicator (Branson) at 10 % power for 1 min. Lysates from cells grown in 6-well plates were sonicated in AFA-tube TPX strips or 96-well AFA plates (Covaris), using a Covaris LE220R-Plus for 300 s at 325 peak power with a duty factor of 50 %, 200 cycles per burst, average power of 162.5. The dithering parameters were set to $\pm$ 5 mm in x and z direction and 4.5 mm in y direction at a speed of 20 mm/s. Protein concentrations of the sonicated lysates were determined using a BCA assay (Pierce) according to the instructions of the kit.

4.2.3 Enrichment of newly synthesized proteins

Enrichment of AHA-containing NSP was initially performed using a manual protocol, which was used as a template to develop a semi-automated protocol.

The automated newly synthesized proteome enrichment protocol was programmed to enable the processing of up to 96 samples in parallel in a PCR plate, using the Agilent Bravo robotic liquid handling platform (equipped with a 96 channel disposable tip head “LT96”). By setting the number of columns on the sample plate, incubation times, volumes of buffers and reagents as variables, adaptations to the protocol can easily be introduced. Unless specified for the particular application, 100 μg protein were used as input for the semi-automated enrichment of NSP. Cell lysates were combined and diluted to a total

volume 150 μL using lysis buffer. In order to prevent the coupling of proteins containing strongly nucleophilic Cysteine to the beads [149], the samples were alkylated by addition 3.4 μL of 600 mM iodoacetamide (IAA) for 20 min at room temperature. Subsequently, 20 μL of magnetic alkyne agarose (MAA) beads, diluted in lysis buffer and the Copper(I)-catalyzed Azide Alkyne Cycloaddition (CuAAC) reaction [150] mixture were added, containing 21.62 mM CuSO4, 108.11 mM Tris-hydroxypropyltriazolylmethylamine (THPTA), 216.22 mM pimagedine hydrochloride and 216.22 mM sodium ascorbate. Next, the plate was removed from the Bravo platform, sealed using VersiCap Mat 96-well flat cap strips (Thermo Fischer Scientific) and incubated for 2 h at 40 °C in a thermal shaker. Following the coupling of the AHA-containing newly synthesized proteins (NSP), the plate was unsealed and moved back onto the orbital shaker (fig. 9) on the Bravo platform, and the supernatant was removed, by placing the sample plate on a magnetic rack (ALPAQUA MAGNUM FLX enhanced universal magnet) for 30 s and aspirating the supernatant in two steps using tips from position 6 and dispensed in the waste plate (position 2). The beads were subsequently washed with 150 μL milliQ water. After addition of the liquid to the beads on the orbital shaker, the plate was moved to the heating station (position 4) where the beads were mixed by pipetting up and down 8 times with a constant flow rate of 300 μL/s, to prevent aggregation of the beads. Next, the plate was transferred onto the magnetic rack and the supernatant was removed. The NSP bound to the beads were subsequently reduced and alkylated by addition of 150 μL of 10 mM Tris(2-carboxylethyl)phosphine (TCEP) and 40 mM 2-chloroacetaminde (CAA), dissolved in 100 mM Tris-HCl buffer (pH 8.0), containing 200 mM NaCl, 0.8 mM Ethylendiamintetraacetic acid (EDTA), 0.8 % SDS and incubating on the heating station at 70 °C for 20 min and subsequent incubation at 20 °C for 15 min on the orbital shaker. The beads were subsequently washed three times with 1 % SDS dissolved in 100 mM Tris-HCl (pH 8.0), 250 mM NaCl and 1 mM EDTA buffer, once with milliQ H2O, three times with 6 M Guanidine-HCl in 100 mM Tris-HCl (pH 8.0) and three times with 70 % ethanol, in consecutive washing steps of 150 μL each. Following the washing steps, the beads were resuspended in 50 μL 100 mM Ammonium bicarbonate buffer (pH 8.0). Proteins were digested off the beads by adding 6 μL of 1 μg/μL sequencing grade Trypsin (Promega), diluted in 50 mM acetic acid, for 16 h at 37 °C, which was performed in a thermal shaker

after sealing the plate with VersiCap Mat 96-well flat cap strips.

The manual NSP enrichment protocol differs in the amount of input material and scale. Manual NSP enrichment was carried out with either commercial alkyne agarose or MAA beads in 2 mL test tubes using 1 mg protein input from combined pulse-labeled cell lysates. All individual steps in the manual enrichment protocol are identical to the semi-automated enrichment protocol, with the exception of bead- (50 μL) and washing buffer volumes (5 mL), and are described in detail in the previous paragraph. The development

4.2.4 Automated SP3 peptide purification

Individual steps of the automated SP3 protocol were used for the purification of peptides by processing of 16-96 samples in parallel on a Bravo liquid handling robot [151]. Following protein digestion, the peptide-containing supernatant was transferred from the sample plate on position 7 onto a new plate on position 8 (fig. 9), and peptides were lyophilyzed using a UNIVAPO-150H vacuum concentrator, coupled to a UNICRYO MC2 cooling trap and UNITHERM 4/14 D closed circuit cooler (UNIEQUIP). On the Bravo platform, magnetic carboxylate Sera-Mag Speed Beads (Fischer Scientific) were diluted to 100 μg/μL in 10 % formic acid and 5 μL were added to each lyophilized sample. Aggregation of the peptides onto beads was induced via addition of 195 μL acetonitrile and incubating for 18 min, while shaking at 100 rpm on the orbital shaker. Next, the supernatant was removed from the magnetic rack in two steps. The beads were washed 2 times with 180 μL acetonitrile and subsequently dried. In the final steps the beads were resuspended in 20 μL 0.1 % formic acid in water and sonicated in an Ultrasonic Cleaner USC-T (VWR) for 10 min and the supernatant was transferred to a new plate. The purified peptides were dissolved in 0.1 % formic acid and used for LC-MS/MS analysis.

4.3 Preparation of samples for (global) proteome analysis

Samples for the analysis of the full proteome (without enrichment of NSP) were prepared using either the manual [152] or automated SP3 protocol [151]. Samples for the evaluation of the semi-automated NSP enrichment protocol, benchmarking of SILAC plexDIA analysis and targeted proteomic analysis of IFNg-treated HeLa cells were prepared via the manual

SP3 protocol. For manual SP3 protocol, cell lysate with 50 μg were used as input material for the preparation of proteome samples. Proteome samples of the RKO cell lines were prepared via the automated SP3 protocol, using 15 μg protein input, from SILAC and AHA pulse-labeled cells, which were also used for newly synthesized proteome analysis. For both automated and manual SP3 protocols, protein aggregation on the magnetic carboxylate beads was induced via addition of acetonitrile to a final concentration of 50 % (v/v), while incubating at 20 °C for 18 min. The beads were subsequently washed with 80 % ethanol and twice with acetonitrile. Tryptic digestion was performed by re-suspending the beads in 50 μL 100 mM Ammonium bicarbonate (Ambic) (pH 8.0) and adding sequencing-grade modified Trypsin (Promega), disolved in 50 mM acetic acid, with a final protease to protein ratio of 1/50 (w/w). Samples were sealed and placed in a thermo shaker for incubation at 37 °C for 16 h. Following the digestion step, the peptides in the supernatant were transferred twice from a magnetic rack (either manually or by the liquid handling robot), and acidified via addition of 10 % trifluoroacetic acid (TFA) to a final percentage of 1 % TFA.

4.4 Liquid chromatography and tandem mass spectrometry (LC-MS/MS)

Quantitative proteomic measurements, described in this dissertation, were carried out using an EASY-nLC 1200 system (Thermo Fischer Scientific) coupled to a QExactive HF mass spectrometer (Thermo Fischer Scientific). The peptides were separated by reverse-phase liquid chromatography using 0.1 % formic acid (solvent A) and 80 % acetonitrile (solvent B) as mobile phases. Peptide separation occurred on an Acclaim PepMap trap column (Thermo Fischer Scientific, C18, 20 mm x 100 μm, 5 μm C18 particles, 100 Å pore size) and a nanoEase M/Z peptide BEH C18 analytical column (Waters, 250 mm x 75 μm 1/PK, 130 Å, 1.7 μm). The samples were loaded onto the trap column with constant of solvent A at maximum pressure of 800 bar. The analytical column was equilibrated with 2 μL solvent A at maximum pressure of 600 bar and heated to 55 °C using a HotSleeve+ column oven (Analytical SALES & SERVICES). The peptides were eluted at a constant flow rate of 300 nL/min. For the elution of peptides, the concentration of solvent B was gradually increased.

A number of different HPLC gradients were used in this study for the analysis of the different sample sets. The gradient, used for the optimizations of the semi-automated NSP

enrichment protocol, started with 4 % solvent B and was increased to 6 % in the first 1 min, increased to 27 % at 70 min and further increased to 44 % after 85 min. After 85 min the percentage of solvent B was raised to 95 %. After 95 min the system was re-equilibrated using 5 % solvent B for 10 min. The gradient, used for the initial comparisons of newly synthesized proteome sample preparation protocols and for the the SILAC labelled Hela benchmark samples, started with 3 % solvent B for the first 4 min, increased to 8 % after 4 min and to 10 % after 6 min. After 68 min the percentage of solvent B was raised to 32 % and after 86 min to 50 %. From 87 min till 94 min of the gradient the percentage of solvent B increased to 100 %. After 95 min the system was re-equilibrated using 3 % solvent B for 10 min. The gradient, used for the analysis of the proteome and newly synthesized proteome of IFNg- or cycloheximide (CHX) treated HeLa cells and RKO cells, started with 4 % solvent B and was increased to 6 % in the first 1 min, increased to 27 % at 51 min and further increased to 44 % after 70 min. After 70 min the percentage of solvent B was raised to 95 %. After 80 min the system was re-equilibrated using 5 % solvent B for 10 min. The gradient, used for the analysis of the full proteome of RKO cells, started with 5 % solvent B and was increased to 9 % in the first 1 min, increased to 27 % at 43 min and further increased to 44 % after 57 min. After 57 min the percentage of solvent B was raised to 100 %. After 65 min the system was re-equilibrated using 5 % solvent B for 10 min.

Eluting peptides were ionized and injected into the mass spectrometer, using the Nanospray flex ion source (Thermo Fischer Scientific) and a Sharp Singularity nESI emitter (ID = 20 μm, OD = 365 μm, L = 7 cm, α = 7.5°) (FOSSILIONTECH), connected to a SIMPLE LINK UNO-32 (FOSSILIONTECH). A static spray voltage of 2.5 kV was applied to the emitter and the capillary temperature of the ion transfer tube was set to 275 °C.

The QExactive HF mass spectrometer was operated in data-dependent- and data-independent mode for the measurement of different samples. The full scan range for data-independent acquisition (DIA) methods was set to 400-1000 m/z with Orbitrap resolution of 60000 FWHM, 3000000 AGC target and maximum injection time of 20 ms. Only for the MS1-optimized DIA method (DIA m2), the full scan Orbitrap resolution was set to 120000. Data-independent MSMS spectra were acquired with an Orbitrap resolution of 30000 FWHM, maximum injection time of 40 ms and AGC target of 1000000. The number and size of the quadrupole

isolation windows was varied for different HPLC gradients and methods, which were used for the acquisition of the different samples. For the measurement of the SILAC plexDIA benchmark (DIA m1) and RKO proteome samples, 26 equally sized, 1 Th overlapping isolation windows with a width of 23.3 m/z, covering the scan range from 402-982.8 m/z were used. DIA method 2, which was used for exclusively in the SILAC plexDIA benchmark, 27 equally sized, 1 Th overlapping isolation windows with a width of 32.2 m/z, covering the scan range from 399.4-999.2 m/z were used. For the measurement of newly synthesized proteome samples, generated from HeLa and RKO cells, 28 equally sized, 1 Th overlapping isolation windows with a width of 22.0 m/z, covering the scan range from 400-989 m/z were used.

The normalized collision energy for the fragmentation of precursor ions was set to 27 and a fixed first mass of 200 m/z was set for the acquisition of the MSMS spectra in all DIA methods. Data-dependent acquisition (DDA) only differed in their HPLC gradient, but used otherwise identical parameters. A full scan range of 375-1500 m/z, Orbitrap resolution of 60000 FWHM, automatic gain control (AGC) target of 3000000 and maximum injection time of 32 ms were set and data-dependent MSMS spectra were acquired using a Top 20 scheme, using a fixed scan range from 200-2000 m/z and fixed first mass of 110 m/z. The quadrupole isolation window was set to 2.0 m/z and normalized collision energy was set to 26. The Orbitrap resolution was set to 15000 FWHM with an AGC target of 100000 and maximum injection time of 50 ms. MSMS spectra were acquired in profile mode and a charge state exclusion of 1, 5-8 & $>$8 was defined. An intensity threshold of 20000 and a minimum AGC target of 1000 was set. Targeted proteomic analysis of selected IFNg target protein candidates was carried out via parallel reaction monitoring (PRM). The previously described 90 min gradient, used for the analysis of newly synthesized proteome samples of IFNg-treated HeLa cells, was applied. Targeted precursors were isolated using 1.4 m/z wide isolation windows and fragmented using various normalized collision energies ranging from 20-27. The full scan resolution was set to 60000, scan range was set from 370-1015 m/z with target AGC of 3000000 and the maximum inject time to 30 ms. Subsequent targeted MS/MS scans were carried out using 1.4 Th wide quadrupole isolation windows, with a maximum inject time of 200 ms and Orbitrap resolution of 60000. Retention time scheduling

of the targeted precursors was designed using information from the un-targeted newly synthesized proteome analysis.

4.5 Extraction, amplification and analysis of genomic DNA

4.5.1 Extraction and amplification of genomic DNA

In order to validate and visualize the genomic editing and induced KO of targeted EIF genes in the RKO cell lines, genomic DNA was extracted and analyzed for RKO cells expressing 2 sgRNA targeting EIF4E. Genomic DNA (gDNA) was extracted from RKO cells using the PureLink Genomic DNA kit (Invitrogen), according to the manufacturers instructions for DNA extractions from cultured cell lines. Concentations and purity of extracted gDNA was measured using a nanoDrop spectrophotometer (Thermo Fisher). DNA surrounding the sgRNA cutsides was amplified using polymerase chain reaction (PCR) [153], with 5′-GATGGTCAGGAAGTTCATCG-3′ acting as forward primer and 5′-GTCACTTCGTCTCTGCTGTT-3′ as reverse primer. 100 ng extracted gDNA was used as template DNA for he amplification, which was performed using the Phusion High-Fidelity DNA Polymerase (M0530) enzyme (1.0 u/50 μL) with addition of 0.2 mM dNTPs, Phusion HF buffer (by New England Biolabs) and 0.5 μM forward and reverse primer. Initial denaturation was carried out for 30 s at 98 ℃, followed by 30 cycles of 10 s denaturation at 98 ℃, 30 s annealing at 61 ℃ and 30 s extension at 72 ℃. A final extension step was included in the amplification protocol at 72 ℃ for 5 min. The PCR products were subsequently cooled at 4 ℃ for 10 min and purified using the QIAquick PCR purification kit (Qiagen), following the manufacturers instructions. DNA concentration and purity of the PCR product was measured using a nanoDrop spectrophotometer (Thermo Fisher). 100 ng amplified DNA was loaded onto and separated on a 1 % (w/v) agarose gel, containing SYBR Safe (Thermo Fisher) staining solution.

4.5.2 Sanger sequencing

Sanger sequencing [154] of the amplified EIF4E DNA was carried out via the ONT sequencing service of the company Microsynth SeqLab GmbH. 300 ng purified DNA mixed with 4 μM reverse primer (5′-GTCACTTCGTCTCTGCTGTT-3′) were used as input for the analysis.

Sequencing traces (in .ab1 format) were visualized using a custom R script.

4.6 RNA sequencing analysis

RKO cells were grown in 6-well plates and KO of the targeted eIF or AAV1 controls were induced through addition of 200 ng/mL doxycyline for 48 h. The cells were scraped off the plates and washed with cold PBS, and 600 μL buffer RLT (Qiagen) was to the pellets for lysis and RNA stabilization. Two experimental replicates of the different eIF KO samples, and four replicates of the AAV1 KO samples were generated. The samples were stored at -80 ℃ and shipped to the IMP in Vienna. Robert Kalis, processed the samples using an in house developed semi-automated RNA extraction and sequencing library preparation protocol and RNA sequencing analysis was performed by the Next-generation sequencing Core-facility of the Vienna Biocenter using the QuantSeq workflow [155]. The raw sequencing data was further processed using an in house generated pipeline for the alignment and quantification of the sequencing reads. Read counts of identified transcript isoforms were aggregated under a single Entrez gene identifier and the resulting data was used for determining differential mRNA expression levels. Differential expression analysis was performed by Robert Kalis using the DEseq2 R/Bioconductor package [156].

4.7 Data analysis

Analysis of processed raw data, from various experimental measurements was carried out using the R software environment (version 4.0.3) with additional software packages, which are specified in the following subsections. No statistical method was used to predetermine sample size. The number of experimental and technical replicates was based on considerations from previous experiments.

4.7.1 Processing of raw data from LC-MS measurements

Raw files from LC-MS measurements were processed with different softwares, depending on the used acquisition scheme. Raw files from DDA measurements were processed using Maxquant (version 2.0.3) [43] and the Andromeda search engine [157], using a human proteome fasta file, retrieved from the SwissProt database (version from February 2021 with

20934 entries). The enzymatic digestion was set to Trypsin/P and a maximum of 2 missed cleavages per peptide were allowed. For the analysis of NSP data, raw files of both the newly synthesized proteome and proteome samples were processed together, using Maxquant. The multiplicity was set to 3, comprising of a light channel, an intermediate channel with Arg6 and Lys4 and heavy channel with Arg10 and Lys8. Cysteine carbamidomethylation was set as fixed modification, whereas Methionine oxidation, N-terminal acetylation, and deamidation of Asparagine and Glutamine were set as variable peptide modifications. The Re-quantify function was enabled, match-between-runs was disabled and other search functions were left with default parameters. Minimum peptide length was set to 7 and max peptide mass was set to 4600 Da. PSM-, Protein- and site decoy fraction FDR were set to 1 %. The minimum delta score threshold for unmodified peptides was set to 6, and 40 for modified peptides. Only for the analysis of the SILAC labeled Hela benchmark samples, the match-between-runs function was enabled. Unique and razor peptides were used for quantification and normalized SILAC ratios and iBAQ values were calculated. The minimum ratio count was set to 0 to not exclude identifications in single SILAC channels. Raw files from DIA measurements were analyzed using DIA-NN (version 1.8.1) [52]. A predicted spectral library was generated from the fasta file, which was also used in the Maxquant searches. Additionally, a fasta file containing common protein contaminants was added for the spectral library prediction [158]. Default settings were used for the spectral library prediction, with the addition of Methionine oxidation as variable modification. For the processing of the raw files, the default settings of DIA-NN were used with additional functions from the plexDIA module enabled [69]. Three SILAC channels with mass shifts corresponding to Lys, Lys4 (+4.025107 Da), Lys8 (+8.014199 Da), Arg, Arg6 (+6.020129 Da), Arg10 (+10.008269 Da) and an additional decoy channel with Lysine (+12.0033 Da) and Arginine (+13.9964 Da) were registered. Translation of retention times between peptides within the same elution group was enabled. The first 13C-isotopic peak and monoisotopic peak was included for the quantification and the MS1 de-convolution level was set to 2. Peptide length range was set from 7-30, precursor charge rate was set from 1-4, mass of charge (m/z) range of the precursors was set from 300-1800 and fragment ion m/z range was set from 200-1800. Precursor FDR was set to 1 %. Precursor matrix output tables

were filtered for FDR $<$ 0.01 and additionally for channel q-value $<$ 0.01 and translated q-value $<$ 0.01. Raw PRM data was analyzed using Skyline (version 22.2.0.527) [159] [160]. A predicted spectral library, which was generated by Prosit was used for the analysis in Skyline [161]. Relative quantification and statistical analysis were carried out using the sum of TIC-normalized fragment ion peak areas via MSstats (version 4.2.2.0) [162] [163].

4.7.2 Analysis of proteomic data

The output tables from Maxquant ("ProteinGroups.txt", "evidence.txt") and DIA-NN ("report.pr_matrix_channels_translated.tsv" and "report.pr_matrix_channels_ms1_translated.tsv") were processed in the R software environment (version 4.0.3) using custom scripts. Identified contaminants were removed and protein abundance was calculated using the MaxLFQ algorithm, applied to the individual SILAC channels, using the iq (version 1.9.6) R package function "process_long_format()" [164]. For MS1- and MS2-based quantification, the "Ms1.translated" and "precursor.translated" quantity was used for the MaxLFQ calculation, respectively. Protein-group SILAC ratios were calculated for each sample using the LFQ values. With the exception of the SILAC DIA benchmark, only MS2 intensity values were used for LFQ normalization. Intensity-based absolute quantification (iBAQ) normalization [120] was carried out by dividing the un-nomalized summed protein intensity values ("PG.Quantitiy" metric in DIA-NN report) by the number of tryptic peptides (with sequence length $\geq$7 and $\leq$30). For newly synthesized proteome data, the summed translated intensities ("Precursor.Translated" metric in DIA-NN) of each protein group were used for the calculation of iBAQ intensities. Principle component analysis (PCA) of the log2 transformed SILAC ratios was performed using the "prcomp" function of the stats (version 4.0.3) R package. For differential expression analysis and correlation with RNA data, only unique protein groups (single Uniprot identifier) with a minimum of 2 SILAC ratios values in 3 replicates were used. Differential expression tests were carried out using the Limma (version 3.46.0) [165] and DEqMS (version 1.8.0) [166] R/Bioconductor packages, by fitting the data onto a linear model and performing an empirical Bayes moderated t-test. The number of precursor SILAC ratios of each protein group was included as a factor for the variance estimation in DEqMS. Overrepresentation enrichment analyisis of significantly deregulated protein

groups ($|\log2 \text{ fold change}| > 0.585$ and adjusted p-value < 0.05), from the IFNg time course experiments, was carried out using a hypergeometric test via the “enricher” function of the clusterProfiler (version 3.18.0) [167] R/Bioconductor package. Gene set enrichment analysis was carried out using the sorted log2 fold change values of all quantified protein groups, using the “GSEA” and “gseGO” function of the clusterProfiler (version 3.18.0) [167] R/Bioconductor package. Gene lists of the Molecular Signatures Database were retrieved and analysed using the msigdbr (version 7.5.1) R package of the CRAN software repository [168]. Gene sets of the Hallmark (H) subset were included in the analysis. Additionally, the top 500 significantly up-regulated genes from the 24 h IFNg-treated Hela cells of the deposited RNA-seq dataset (GSE150196, more specifically “GSE150196_RNA-seq_DESeq2_priming_vs_naive.tab”) [169] and the top 500 target genes from a public STAT1 ChIP-seq dataset from 30 min IFNg treated Hela S3 cells (ENCSR000EZK, more specifically “ENCFF039MZH.bed”) [170] were included in the enrichment analysis of IFNg-treated HeLa cells. ChIP-seq data were processed using a custom R-script. Only genes within 2000 bp upstream or downstream of the annotated transcription start sites of the hg38 reference genome were included in the final STAT1 target gene list. Quantified protein groups in the respective condition were included as background gene list for the enrichment and p-values were adjusted using the Benjamini-Hochberg approach [171]. The number of datasets in which differential expression of selected proteins or genes was reported in response to IFNg in human cells or tissues, was retrieved from the “interferome.org” database [172]. Translation efficiency (TE), as quantitative metric for protein synthesis, was calculated by dividing the NSP fold change (obtained from newly synthesized proteome analysis) by the fold change of the respective RNA (obtained from RNA sequencing). RNA sequence information of transcripts with canonical annotation of the Ensembl database were retrieved using the ensembldb (version 2.14.1) [173] and GenomicFeatures (version 1.42.3) [174] R/Bioconductor package . More specifically, untranslated region (UTR) sequences were retrieved via the “fiveUTRsByTranscript” and “threeUTRsByTranscript” functions and coding sequence (CDS) information was retrieved via the “cdsBy” function of the GebomicFeatures R/Bioconductor Package. Data with characteristics of human uORF elements, was obtained from the supplementary material of Chew et al. [175]. Data listing evidence for

N6-methyladenosine modification site on human transcripts (in HeLa, HepG2 and HCT116 cell lines) was obtained from the “m6A atlas” [176].

5 Results

5.1 Developing a semi-automated workflow for the enrichment of newly synthesized proteins

Newly synthesized proteome analysis is a powerful methodology that is well suited for the analysis of changes in protein synthesis, in response to different eIF depletions. The relatively large number of eIF proteins to be targeted for analysis pose a challenge for conventional workflows. As indicated in the introduction (section 3.3), established protocols for the enrichment of NSP, require large numbers of cultured cells in the metabolic labelling step to produce the necessary amounts protein extracts that serve as input for the enrichment [134][139]. Furthermore, manual sample preparation limits the number of samples that can be processed in general. Processing of the samples on differing dates could introduce additional technical variation and complicate analysis of the proteomic data. Finally, in order to achieve relatively high proteome coverage, newly synthesized proteome samples are conventionally analysed using data-dependent acquisition (DDA) methods with additional offline fractionation or long LC gradients. For a large number of samples, as is the case in this project, this would lead to very long LC measurement time, which in turn poses additional challenges in ensuring stability of the HPLC and mass spectrometer for multiple weeks without major changes or drifts in performance. Therefore, initial efforts in this project were focused on adapting workflows to simplify the sample preparation and increase the throughput of quantitative newly synthesized proteome analysis.

To address the aforementioned limitations, the initial goal was to develop an automated sample processing with reduced protein input requirements for the NSP enrichment. To achieve this goal, a click chemistry-based enrichment protocol, using magnetic beads, was developed and optimized for automated sample preparation with a robotic liquid handling platform. Additionally, to increase proteomic depth in single LC-MS measurements, data-independent acquisition (DIA) methods and software were evaluated and adapted for newly synthesized proteome analysis.

5.1.1 creating magnetic alkyne beads for the enrichment of newly synthesized proteins

Established protocols for newly synthesized proteome analysis primarily make use of non-magnetic alkyne-conjugated agarose beads [138] [139] [134]. Cross-linked agarose beads are not well suited for adapting protocols for automated sample preparation, since they do not have magnetic properties and require centrifugation. However, these non-magnetic cross-linked agarose beads offer a number of key advantages over alternative comercially available magnetic beads. The high surface area and binding capacity of conjugated cross-linked agarose, enables the capture of large amounts of NSP. The amount of NSP that can be bound and released by beads in the enrichment step is a major factor determining the sample quality in the different workflows for newly synthesized proteome analysis. Furthermore, agarose beads are also highly durable and are stable in a wide range of pH and temperature conditions, solutions with high salt or chaotrope concentrations and organic solvents [177]. These characteristics are crucial for the extensive and stringent washing steps needed to remove unlabeled pre-existing proteins in the NSP enrichment. To enable automated sample processing through click chemistry-based enrichment of AHA-containing NSP, a magnetic bead type with alkyne conjugation, high stability in regards to salt, chaotrope and organic solvent concentrations as well as thermal stability is needed. For initial testing, a number of NSP enrichments were carried out using different commercial and in-house modified magnetic- and non-magnetic bead types. The NSP enrichment was carried out with 1 mg total protein input obtained from RKO cells, which were pulse labelled for 6h with AHA and SILAC Lysine and Arginine containing media. The protocol followed the general large scale NSP enrichment protocol outlined in method description (section 4.2). In the case of biotin-alkyne conjugation of AHA-containing NSP, proteins were precipitated using methanol/chloroform precipitation to remove excess biotin-alkyne, which would inhibit NSP binding to Streptavidin coupled to the magnetic beads. In addition to the various NSP enrichment protocols, a sample without enrichment of NSP was prepared using the SP3 protocol [152] indicated in method description (section 4.3). The samples were then analyzed via LC-MS, using a data-dependent acquisition method. To create magnetic beads with similar characteristics to the established and commercially available

alkyne agarose beads, magnetic cross-linked agarose beads containing amine (-NH2) and carboxy (-COOH) functional groups were chemically modified to produce magnetic alkyne agarose (MAA) beads. Initially two separate approaches based on EDC-NHS cross-linking chemistry were applied (fig. 6).

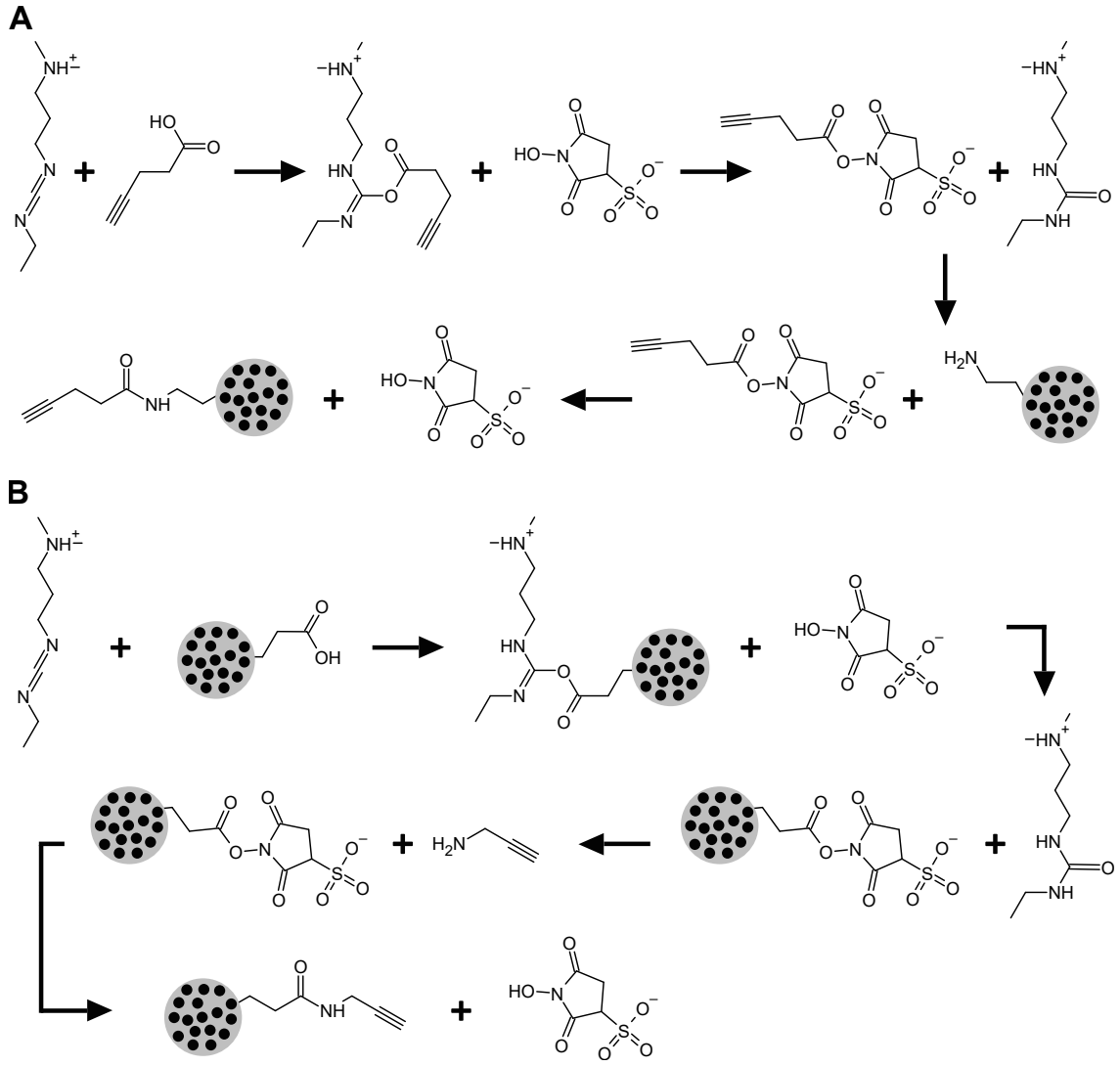


Figure 6: Proposed reaction mechanism for the coupling of amine- and carboxy-activated magnetic agarose beads, to produce magnetic alkyne agarose beads. A) Coupling mechanism of amine-activated magnetic agarose beads with EDC, NHS and 4-Pentynoic acid. B) Coupling mechanism of carboxy-activated magnetic agarose with EDC, NHS and propargylamine.

Two main metrics were assessed in the comparison of the different beads, to assess their suitability for the enrichment of NSP and LC-MS-based analysis. First, the number of quantified peptides and proteins, which have detectable levels in both heavy and intermediate SILAC channels and are thus originating from NSP, are compared. Second, to assess the

efficiency of the NSP enrichment, the intensity ratios of the heavy and intermediate SILAC labeled proteins (which originate from the pulse-labeled NSP) over the unlabeled (light) proteins, which represent pre-existing proteins, are calculated.

The enrichment of biotin-alkyne conjugated NSP with magnetic Streptavidin beads resulted in the quantification of only very few NSP in contrast to the other protocols. This could most-likely be explained by the low binding capacity of these beads. Although the other methods yielded higher numbers of identified and quantified protein groups, the commercial magnetic alkyne beads did not lead to an increase in the intensity of SILAC-labeled NSP, indicating low enrichment efficiency. Intensity ratios in samples prepared with the commercial magnetic alkyne beads are comparable to the ratios which were observed in the sample without enrichment of NSP. The commercial alkyne beads also have much lower surface density of terminal alkyne groups, indicating that unspecific protein binding to the beads occurs to large extends even with extensive and stringent washing. This unspecific binding of pre-existing proteins out-weights the covalent NSP binding. The chemically modified magnetic agarose beads, produced samples which are comparable to commercially available non-magnetic alkyne agarose beads, in respect to the number of quantified peptides and proteins as well as SILAC intensity ratios (fig. 7).

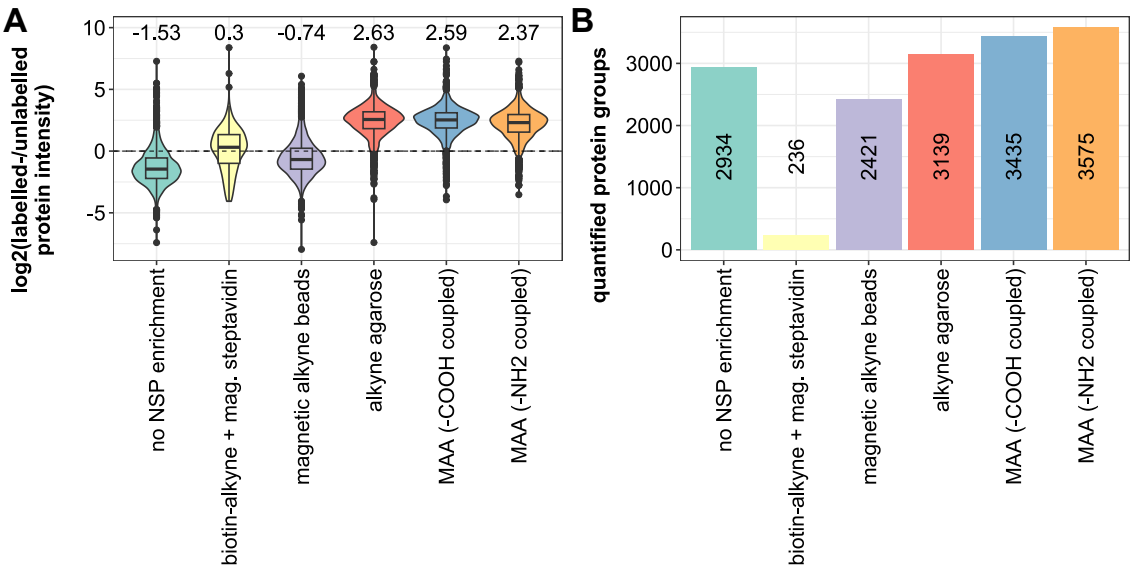


Figure 7: Comparison of multiple different methods and beads for the enrichment of AHA-containing newly synthesized proteins. In addition to the multiple samples which were produced through enrichment of NSP, a sample without NSP enrichment was also included in the analysis. A) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labeled NSP, over unlabeled pre-existing proteins. B) Number of quantified protein groups, determined through the number of proteins with heavy- over intermediate SILAC ratios.

The results of the initial test therefore indicate that chemically modified magnetic alkyne agarose (MAA) beads are well suitable for the preparation of high quality enriched newly synthesized proteome samples.

To further simplify the AHA-containing NSP enrichment as well as the metabolic labeling of cultured cells, scaling down the enrichment reactions is crucial. Lower input requirements are needed to scale the NSP enrichment procedure down to a volume < 200 μL in order to be adapted to Agilent Bravo robotic liquid handling platform. Additionally, lower protein input requirements also reduce the number of cells that need to be grown and labeled prior to the NSP enrichment step. Conventional workflows require >500 μg, which requires cells to be grown and harvested in large culture flasks (10-15 cm diameter). The ability to grow and label cells in smaller dimensions further increases throughput in sample preparation and greatly simplifies the handling for larger sets of conditions in the analysis. To assess the feasibility of a PCR tube scaled NSP enrichment, and to determine the ideal range of protein input, NSP enrichments were carried out using MAA beads, produced via propargylamine cross-linking carboxy-activated magnetic agarose beads, in PCR tubes and with standard

large-scale conditions using 1 mg protein input and a 2 mL tube. The amount of MAA beads was kept constant at 5 μL bead volume in the PCR tube and 50 μL bead volume in the 2 mL tube. Peptide purification, following the NSP enrichment in PCR tubes was performed with SP3, instead of desalting with OASIS plates. In this experiment, relatively high numbers of proteins and peptides were identified in samples that were enriched in PCR tubes even 50 μg input. However, the highest numbers of protein and peptide identifications were still observed in the large-scale enrichment. From this experiment it can be concluded that higher protein input be beneficial for NSP sample quality. However, even with relatively low amounts of protein input ($\leq$50 μg) the number of quantified protein groups is not decreased below 2500. Importantly, the intensity ratio of labeled NSP over unlabeled pre-existing proteins is comparable across the tested range of protein input, suggesting that the NSP enrichment is efficient even with low protein input and overall volume (fig. 8).

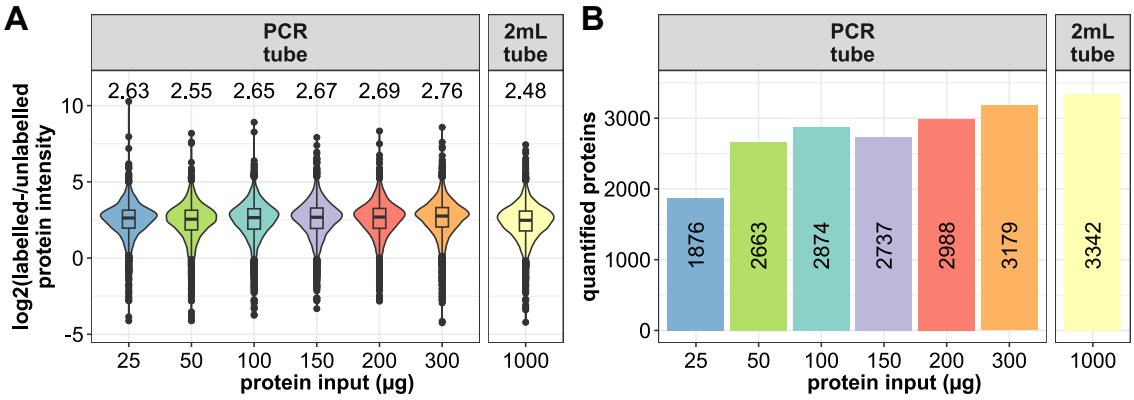


Figure 8: Initial tests of PCR-tube scaled click chemistry-based enrichment of AHA-containing NSP enrichments via MAA beads. To assess the feasibility of scaling down the NSP enrichment protocol to a volume $<$ 200 μL (PCR tube scale), samples were generated with MAA beads in PCR tubes and a 2 mL tube with different amounts of protein input material. A) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labelled precursors, originating from NSP, over unlabeled precursors which originate from pre-existing proteins. B) Number of quantified protein groups in samples prepared with different protein input amounts and test tube dimensions.

The promising results of the experiment, with NSP enrichments in PCR tubes using MAA beads and different protein input amounts, indicates that it is feasible to perform the sample preparation for newly synthesized proteome analysis in a format that can be adapted on a robotic liquid handling platform.

5.1.2 developing a semi-automated protocol for the click chemistry-based enrichment of newly synthesized proteins

In addition to the programming of the liquid handler there are several considerations and challenges for the creation of an automated NSP enrichment protocol. These include the unique characteristics of the MAA beads and exposed surface of the robotic liquid handling platform. The MAA beads have an average diameter of 30 μm and are thus orders of magnitude larger and heavier than conventional solid magnetic beads, such as SeraMag beads which on average have a 1 μm diameter and are used in SP3 protocols [152] [178]. Multiple magnetite spheres inside the MAA beads, add to their high mass and strong ferromagnetic properties. Therefore, MAA beads diffuse to the ring magnet very rapidly within a few seconds, but also rapidly settle on the well bottom within a few minutes. Due to the relatively large volumes in the 96-well conical shaped plates, in which the enrichment takes place, rapid shaking to re-suspend the MAA beads after settling is not possible without spilling liquid into adjacent wells. The automated SP3 protocol for proteomics sample preparation served as a template for the creation of the automated NSP enrichment protocol. For the initial development of a prototype NSP enrichment protocol all steps involving liquid addition, washing of beads incubation at various temperatures and time periods were copied from the autoSP3 protocol. In addition, a 2 h incubation step at 40 ℃ for the copper catalysed azide alkyne cycloaddition (CuAAC) click reaction was created. This step was carried out by placing the sample plate in position 4 of the liquid Bravo platform (fig. 9). To prevent settling of the MAA beads, regular intervals of mixing through pipetting the reaction mixture up- and downwards were included. Similar to conventional cross-linked agarose beads, MAA beads to not stick to pipet tips. This characteristic of the beads, enables resuspending of the beads via mixing through up- and downwards pipetting with minimal loss of beads. Although all additions of beads, reagents and washing steps were carried out through automated pipetting, the tryptic digestion step was carried out off the liquid handling platform, to avoid evaporation issues. Tryptic digestion of the NSP, bound to the MAA beads, was performed by sealing the sample plate and incubating it in a thermoshaker with heated lid at 37 ℃ for 16 h. Peptide purification of the enriched NSP samples was carried out by adapting the autoSP3 protocol. Prior to subsequent peptide

purification via autoSP3, the sample volume needs to be reduced to enable sufficient dilution of the sample with acetonitrile (>95 %). Therefore, the samples were dried using either a SpeedVac vacuum desiccator. Magnetic SeraMag carboxyl beads are then diluted in 10 % formic acid and added to the dried sample. Additional adjustments included, changing the depth of the pipet tips relative to the positioning of the well bottom from 0.5 to 0.95, to ensure that no acetonitrile is spilled out of the wells by displacement of the pipet tips, and reducing the speed of the orbital shaker in position 9 to 100 rpm to avoid contaminations across wells (fig. 9).

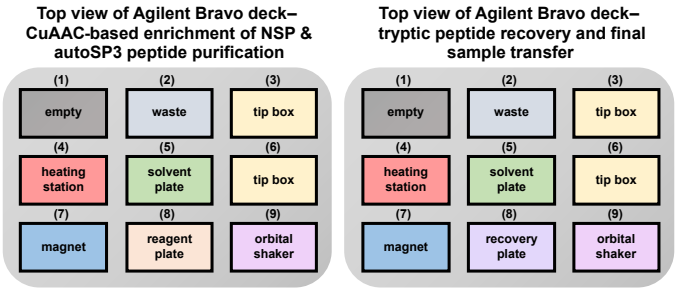


Figure 9: Schematic overview of the Bravo liquid handling platform setup for the automated NSP enrichment and autoSP3-based peptide purification (A) and recovery step for the transfer of tryptic peptides and separation of magnetic beads (B). This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

Samples produced in initial tests with the first version of the automated enrichment protocol only lead to the quantification of relatively few protein groups and primarily detected unlabeled peptides, originating from pre-existing proteins (fig. 10). In order to determine the causes for the low sample quality of samples produced with the first version of the automated protocol, the procedure was repeated with a single modification to the protocol: In order to exclude oxidation of the Cu(I) catalyst in the presence of high levels of oxygen caused by repeated pipet mixing on the open robotic liquid handling robot, the CuAAC-based click reaction was carried out with the sealed sample plate in a thermo shaker with heated lid. This modification to the protocol, which requires the plate to be manually removed for the CuAAC-based NSP coupling step, lead to an increase in the number of quantified protein groups while also decreasing the intensity ratio of SILAC labeled peptides, indicating a lower NSP enrichment efficiency (fig. 10, panel A). To exclude that this issue is not caused by aggregation of proteins to the MAA beads during the reduction-, alkylation and washing

steps in the protocol, additional pipet mixing steps were included in each washing step, to ensure resus pension of the beads. The additional mixing and re-suspension of the beads, together with the CuAAC-based coupling of the NSP with a sealed plate off the liquid handler in the third version of the protocol, resulted in increased numbers of quantified protein groups and an increased SILAC intensity ratio (fig. 10).

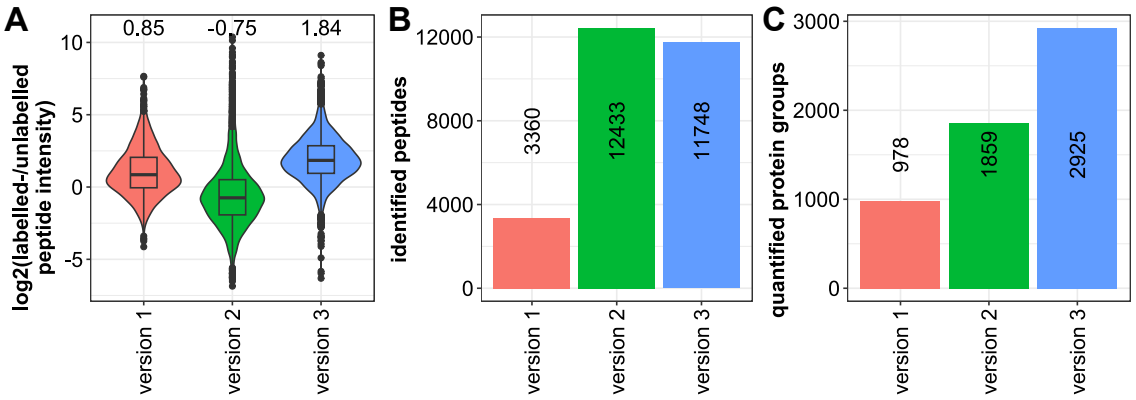


Figure 10: Assessing the performance of three prototype versions of the semi-automated protocol for the enrichment of AHA-containing NSP. A) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labeled NSP over unlabeled pre-existing proteins for samples produced with the different semi-automated enrichment protocol versions. B) Number of identified peptides in samples produced with the different versions of the semi-automated enrichment protocol. C) Number of quantified protein groups in samples produced with the different protocol versions.

Although the third version of the automated NSP enrichment protocol produces samples with improved quality, final optimizations were carried out to yield enrichment efficiencies and proteomic depth comparable to manually prepared samples. Final modifications included the increased flow speed to 300 μL/s and number of repeated pipet-mixing in the washing steps of the protocol, to reduce the unspecific binding of unlabeled proteins to the beads, and switching to a different lysis buffer. In addition, the preparation of magnetic alkyne agarose (MAA) beads was optimized by using a more streamlined coupling protocol through the use of epoxy-activated magnetic agarose beads and propargylamine. In contrast to the EDC-NHS based cross-linking reactions, the coupling with epoxide groups can be carried out in a single step through the addition of propargylamine. In addition to the simplified preparation of the beads, the epoxy-derived MAA beads perform better in comparison to carboxy-derived MAA beads, in terms of peptide identifications although they on average yield slightly lower SILAC intensity ratios (fig. 11).

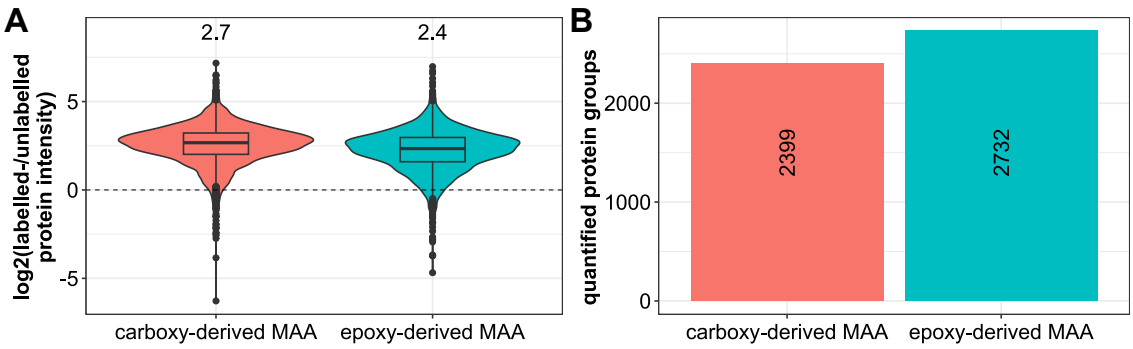


Figure 11: Comparative analysis of newly synthesized proteome samples, produced with MAA beads derived from either carboxy-activated- or epoxy activated magnetic agarose beads. A) Quantifying the efficiency of NSP enrichment via intensity ratios of heavy- plus intermediate SILAC labeled NSP over unlabeled pre-existing proteins for samples produced with the two different MAA beads. B) Number of quantified protein groups in samples produced with the two different MAA beads.

Furthermore, the lysis buffer in which the CuAAC-based coupling of the NSP to the MAA beads is performed was exchanged from a buffer containing 8 M urea and 5 % (w/v) 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) detergent to a 1 % SDS-containing lysis buffer. This change was primarily driven by the high amounts of CHAPS contamination, observed in samples prepared with MAA beads. However, comparative analysis of samples prepared with either CHAPS and 8 M urea-, 8 M urea- and SDS-containing lysis buffer revealed that the latter performs best in terms of peptide identifications and NSP enrichment efficiency (fig. 12).

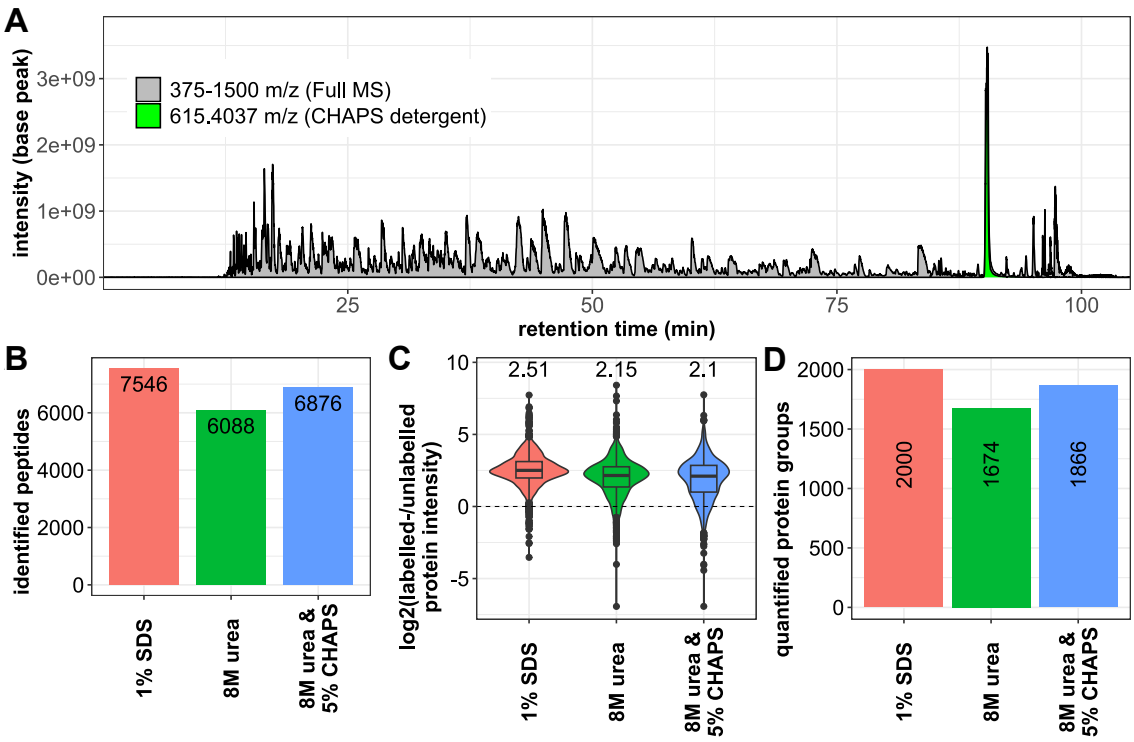


Figure 12: Comparative analysis of newly synthesized proteome samples, generated with the enrichment of MAA beads using different lysis buffer. A) Contamination of CHAPS detergent in samples generated through NSP enrichment with MAA beads, using lysis buffer containing 5 % (w/v) CHAPS and 8 M urea. Base peak chromatogram of a representative sample, featuring the characteristic peak of singly-charged, protonated CHAPS at 615.4037 m/z. B) Number of identified peptides in samples produced through NSP enrichment with different lysis buffers. C) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labeled NSP over unlabeled pre-existing proteins for samples produced with different lysis buffers. D) Number of quantified protein groups.

Additional experiments were carried out with the aim of determining the optimized conditions of the semi-automated enrichment protocol, for the preparation of future samples. The high mass and density of the MAA beads causes them to precipitate quickly out of suspensions. In the context of automated liquid handling, this can lead to differences in bead concentrations within the sample plate upon addition of beads to the samples. These differences in bead and alkyne concentration could potentially impact the efficiency of the NSP enrichment and reduce reproducibility. To prevent precipitation, the MAA bead suspension was mixed through alternating up- and downwards pipetting in the reagent plate (position 8) prior to addition to the sample plate (position 9) (fig. 9). Additionally, to assess the influence of differing MAA bead amounts in the semi-automated NSP enrichment, samples were prepared with equal protein input (100 μg) and a range of different MAA bead volumes. Lysate of HeLa cells, which were pulse labeled with methionine-free, AHA and SILAC amino

acid containing media for 4 h, were used for this experiment. To contrast samples that were prepared with NSP enrichment to samples that have only been pulse labeled, an additional sample was prepared via manual SP3 [152]. Within the tested range of MAA bead volumes, no major changes in the number of quantified protein groups could be determined. However, the number of quantified protein groups is slightly increased in samples prepared with the NSP enrichment, compared to samples prepared without enrichment (nE) ((fig. 8), panel A). A minor decrease in the ratio of labeled NSP precursor intensities over unlabeled precursors could be measured in negative correlation to MAA bead volumes. According to the median intensity ratios of the samples prepared with and without NSP enrichment, NSP intensities are increased $>$25 fold through the procedure (fig. 13, panel B). Since the samples were obtained from HeLa cells without any perturbations, no changes in protein expression are expected. The measured SILAC (heavy over intermediate) ratios therefore indicate accuracy of the relative NSP quantification. Narrow distributions of log2 SILAC ratios, centred around 0, can be observed in all enriched NSP samples, whereas the sample prepared without enrichment contains a wide range of fold changes (fig. 13, panel C). Furthermore, the quantitative precision was compared between the different samples, through calculating the coefficient of variation (CV) of SILAC ratios between replicates. Low CV values averaging around 5 % can be observed in the enriched NSP samples. The CV value of the SILAC ratios in the sample without NSP enrichment is significantly higher at 22.8 % (fig. 13, panel D).

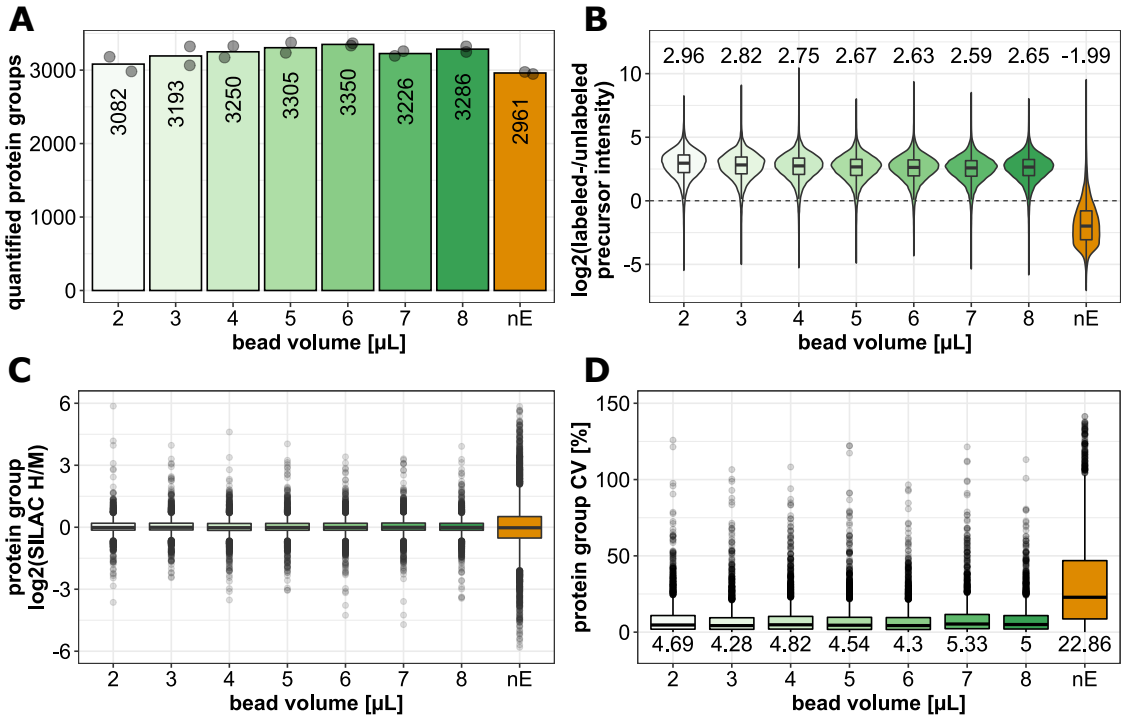


Figure 13: Determining and optimizing the range of MAA bead volume for the semi-automated NSP enrichment through comparative analysis of samples generated with differing volumes of MAA beads for the semi-automated NSP enrichment. Additionally, a sample which was prepared without NSP enrichment is included in the comparative analysis. A) Number of quantified protein groups, determined via the number of proteins with heavy- over intermediate SILAC ratios. B) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labeled precursors, originating from NSP, over unlabeled precursors which originate from pre-existing proteins. C) Distribution of log2 heavy over intermediate protein SILAC ratios. Since samples were generated from HeLa cells without perturbations, the expected log2 SILAC ratios are equal to 0. D) Precision of quantification, measured via the coefficient of variation (CV) of the heavy- over intermediate SILAC ratios. Data are based on two experimental replicates. This figure was taken and modified from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

The results imply that the established NSP enrichment protocol is robust and can produce samples with high reproducibility, even with differing amounts of MAA beads. The quality of the samples produced by the semi-automated NSP enrichment protocol is very high and yields rich chromatograms during measurement, which are comparable to conventional proteomics samples (fig. S1). Moreover, the contrast to samples prepared without NSP enrichment highlight that the main advantage of the protocol is not primarily in improving proteome coverage, but in allowing for analyzing changes in newly synthesized protein expression with high quantitative accuracy and precision.

In order to determine the semi-automated NSP enrichment protocols limitations with respect

to protein input, samples were prepared with equal amounts of MAA beads (4 μL volume) and different amounts of protein lysate. For this purpose, HeLa cells were pulse labeled with methionine-free, AHA and SILAC amino acid containing media for 4 h without any perturbations. Increased intensity ratios of labeled peptides, originating from NSP, were determined in all tested samples across the range of tested protein input from 1-300 μg. Even with 1 μg protein input as starting material, NSP are on average 14.5-fold enriched, in comparison to the sample prepared without NSP enrichment. The intensity ratios increase with higher amounts of protein input, but do not significantly differ upwards of 25 μg (fig. 13, panel B and fig. 14, panel A). The number of quantified NSP correlates with protein input. However, no significant increase in the number of quantified NSP could be measured with protein input $\leq$100 μg. Indicating that the MAA beads are saturated or cannot bind more NSP at this concentration of labeled NSP so that increases in protein input are not beneficial (fig. 14, panel B). No major differences in the quantitative accuracy and precision of the NSP SILAC ratios could be observed across the range of samples with differing protein input (fig. 14, panel C-D).

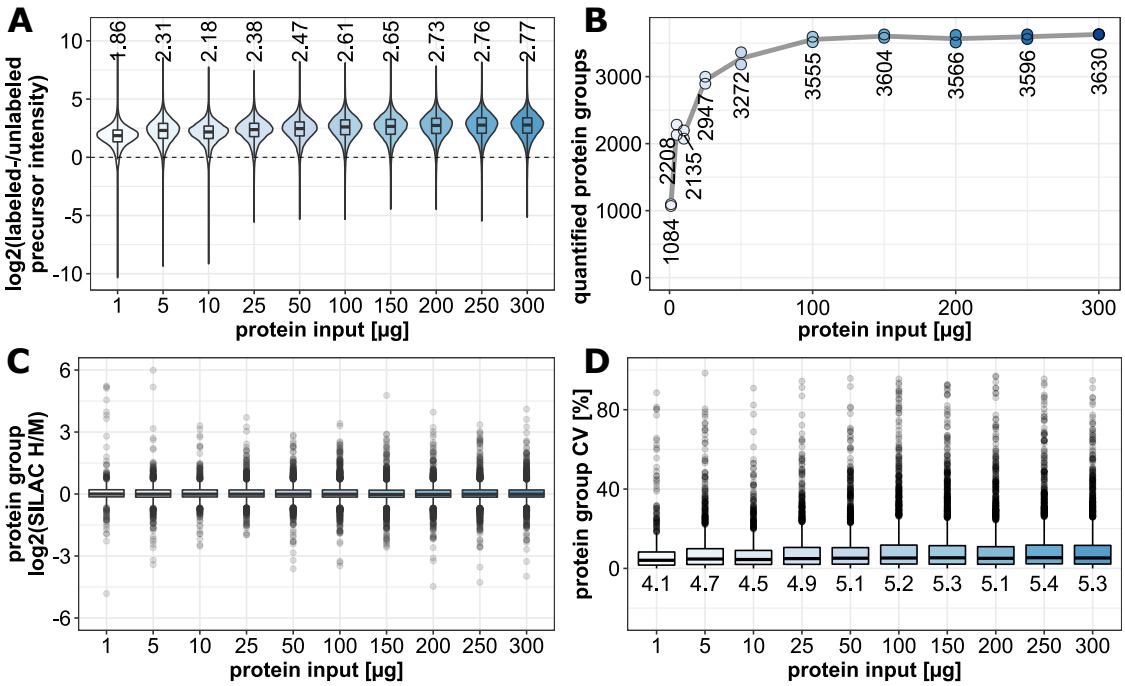


Figure 14: Determining and optimizing the range of protein input for the semi-automated NSP enrichment through comparative analysis of samples generated with differing amounts of protein input for the semi-automated NSP enrichment. A) Quantifying the efficiency of NSP enrichments via intensity ratios of heavy- plus intermediate SILAC labeled precursors, originating from NSP, over unlabeled precursors which originate from pre-existing proteins. B) Number of quantified protein groups in samples prepared with different protein input amounts for the semi-automated NSP enrichment. C) Distribution of log2 heavy over intermediate protein SILAC ratios. Since samples were generated from HeLa cells without perturbations, the expected log2 SILAC ratios are equal to 0. D) Precision of quantification, measured via the coefficient of variation (CV) of the heavy- over intermediate SILAC ratios. Data are based on two experimental replicates. This figure was taken and modified from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

The results of the dilution series experiment indicate that the semi-automated NSP enrichment protocol works efficiently even with limited starting material. High quality data can be generated with as little as 25 μg of protein input material, which can be obtained from low cell numbers, thus permitting to scale down cell culture and pulse-labelling and rendering quantitative analysis of the newly synthesized proteome more manageable.

5.1.3 Adapting multiplexed data-independent acquisition (plexDIA) for the analysis of SILAC labeled newly synthesized proteome samples

Samples that are generated with the semi-automated NSP enrichment protocol are of high quality but are limited in terms of sample amounts, yielding enough material for only a

single sample injections for LC-MS measurements (using an Orbitrap MS platform). The manual and large scale NSP enrichment protocol yields enough material for up to 10 sample injections and can be offline fractionated using a high pH reverse-phase chromatography system. The measurement of fractionated NSP samples increases proteome coverage, while also greatly increasing measurement time. Data-independent acquisition (DIA) methods and software tools were shown to increase proteome coverage in label free analysis in single injection measurements [179]. Recently, multiplexed DIA analysis functions, for the analysis of peptides with non-isobaric labels, were also implemented in the DIA-NN software environment [69][52]. Through the use of plexDIA, increased proteomic depth and high quantitative accuracy and precision were reported for the analysis of mTRAQ labeled samples. To test whether the use of DIA methods and the plexDIA analysis framework are beneficial for increasing proteomic depth in SILAC labeled samples, such as the NSP samples which are generated with the semi-automated enrichment protocol, lysates of SILAC labeled HeLa cells were mixed in defined ratios and analyzed using either DDA or DIA methods. Two different samples were created, to either resemble a sample produced through NSP enrichment that primarily consists of similar amounts of heavy and intermediate SILAC labeled proteins (SILAC mix 1), and a sample that mimics the composition of a pulse labeled sample without NSP enrichment and is thus primarily composed of unlabeled light proteins (SILAC mix 2) (fig. 14, panel A). To compare MS1 and MS2-based quantification in DIA data, two different DIA methods were created and used for the analysis of the mixed SILAC samples. DIA method 1 is optimized for short cycle time to enable robust MS2-based quantification, while DIA method 2 includes 3 high resolution MS1 scans per measurement cycle to enable improved MS1-based quantification (fig. 14, panel B).

In this benchmark, DDA analysis consistently leads to lower precursors identifications (fig. S2) and quantified protein groups. In contrast to DIA data, the number of proteins quantified across the different SILAC channels is constant in the DDA data. This finding indicates that within the tested range, the number of quantified proteins is independent of the abundance differences across the channels. Using plexDIA analysis, fewer protein groups were quantified in the mix 1 sample compared to mix2. The number of identified precursors correlates with the differences in relative abundance of proteins across the

SILAC channels. These observations suggest that the detected proteins are primarily of high abundance and easier to quantify. Since it is more difficult to quantify signals at low S/N ratio, this observation matches expectations. Using both the conventional- (method 1) MS1-optimized (method 2) DIA methods, MS1-based quantification (shown in blue) and MS2-based quantification (shown in red) did not produce large differences in the numbers of quantified proteins. Furthermore, the number of H/M-quantified proteins in mix2 more than doubled from approximately 3000 in DDA to well over 6000 in DIA. Since the mix2 sample was created to mimic the composition of an enriched newly synthesized proteome sample, plexDIA appears to be very well suited for the analysis of NSP enriched samples. To investigate the quantification of the known SILAC ratios across the different methods, comparative analysis of two different figures of merit was carried out. Precision of the quantified protein groups was compared using the coefficient of variation metric (CV), and the deviation of the expected SILAC ratio was calculated for every quantified protein.

For the relative protein quantification across all methods, the median CVs were in a relatively low and within a relatively narrow range between 5-15 %, indicating high quantitative precision. However, some minor trends can be observed in the comparisons. DDA-based MS1 quantification was more precise in the mix2- than in the mix1 sample. The higher precision in the mix2 samples could possibly be caused by SILAC ratios that are close to or equal to 1, which can be quantified more accurately. In both DIA methods, MS1-based quantification resulted in comparable precision for both SILAC mix samples. Minor improvements in quantitative precision, using the MS1-optimized DIA method (DIA m2), can most-likely be attributed to the 2 additional high resolution MS1 scans per instrument cycle (fig. 14, panel B). Surprisingly, for both DIA methods, MS2-based quantification results in superior quantitative precision for both SILAC mix samples. To assess the quantification accuracy, the measured SILAC ratios were compared to the expected ratios in the 2 mix samples. MS1-based quantification accuracy was shown to be comparable across all methods. In contrast to the MS2-based quantification, MS1-based quantification data, in both DDA and DIA methods, includes a larger spread of fold change values with numerous outliers. In this benchmark, MS2-based quantification resulted in the most accurate quantification across all tested conditions (for both DIA methods) (fig. 14, panel

D-E). The lowest average deviation from the expected SILAC ratios can be observed in the SILAC mix 2 samples, which have similar labeled protein abundances and also resulted in the highest number of quantified protein groups (fig. 14, panel C). Although this benchmark primarily focused on the comparison of protein level quantification data, the results are also reflected on the precursor level (fig. S2).

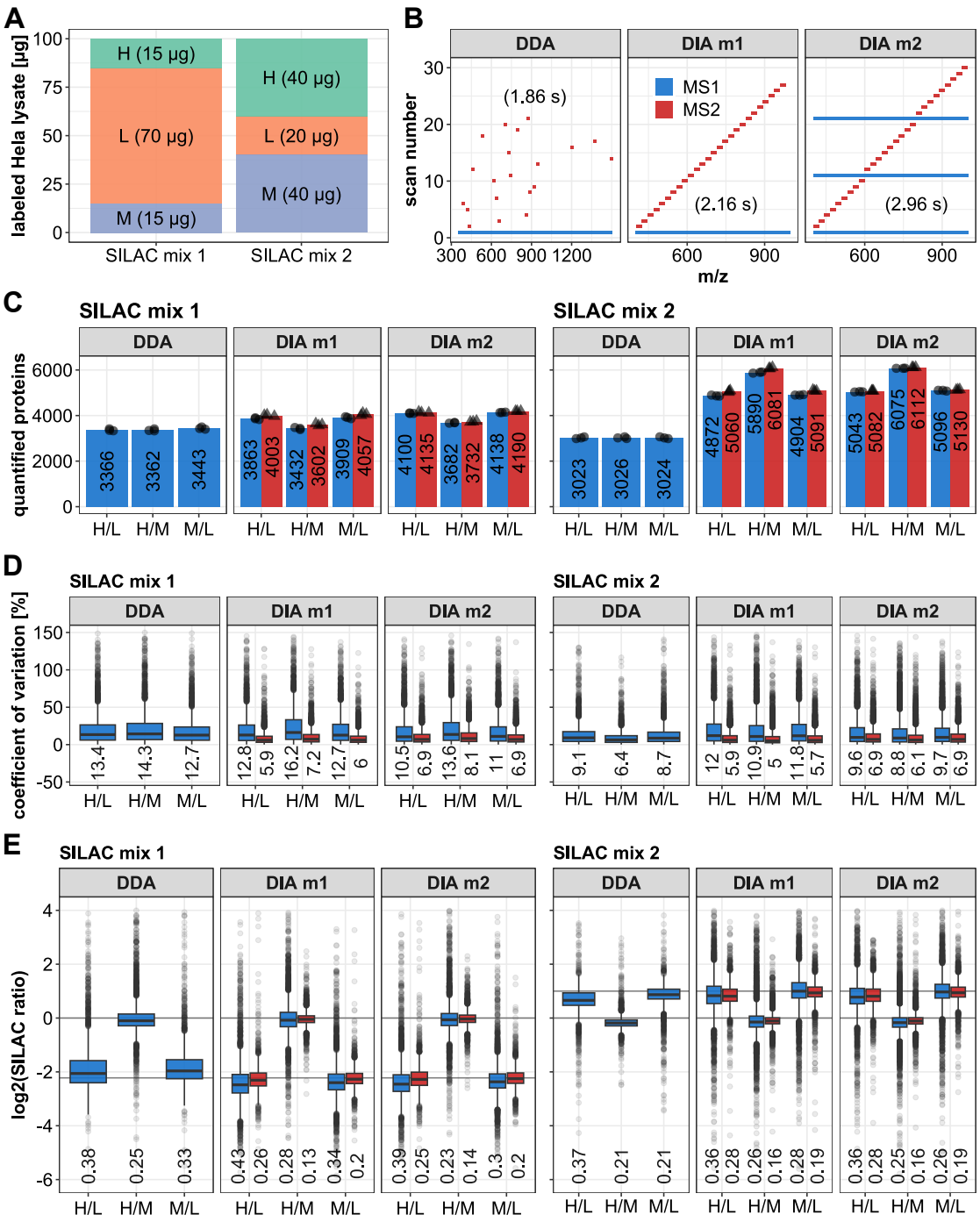


Figure 15: Comparative analysis of samples prepared with known ratios of SILAC labeled HeLa cells, using data-dependent acquisition (DDA) and data-independent acquisition (DIA) mass spectrometry. For all comparisons both MS1- and MS2-based quantification data are included. Metrics from MS1-based quantification are indicated in blue and MS2-based quantification in red. A) Schematic representation of the composition of the samples, generated from stable SILAC labeled HeLa cells. B) Schematic representation of the DDA method (with top 20 precursor isolation) and the two DIA methods, which were used for the comparative analysis. Maximum cycle time for each method is indicated in each panel. C) Comparison of the number of quantified protein groups for the different methods. D) Box plots indicating the quantitative precision, measured in the coefficient of variation (CV) values of the SILAC ratios of the quantified protein groups. E) Box plots indicating the distribution of log2-transformed SILAC ratios of the quantified protein groups. The median difference from the theoretical log2 SILAC ratios are indicated below the box plots. This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

In summary, this benchmark dataset proves that DIA measurements of SILAC labeled samples, combined with the plexDIA functionalities of the DIA-NN software, result in a significant increase in proteomic depth while preserving high quantification accuracy and precision. According to the results of the benchmark, SILAC plexDIA analysis is most effective for the analysis of samples, that primarily consist of similar fractions of labeled peptides, making it highly suitable for the analysis samples generated through NSP enrichment.

To ensure that the benefits of plexDIA analysis, that were observed in a benchmark with labeled cell lysates, translate to samples generated through the semi-automated enrichment of NSP, an additional experiment was carried out. Samples were once again generated with different amounts of protein input and analysed using plexDIA. High intensity ratios of labeled precursors were detected in all samples across the tested range of protein input, indicating efficient enrichment of NSP, even at low overall protein input (fig. 16, panel A). However, similarly to the previous measurements with DDA methods, increased protein input also resulted in an increase in protein and peptide identifications. No significant increase in the number of quantified protein groups (around 6000) could be achieved with $\geq$100 μg. Interestingly, even with as little as 25 μg protein input for the semi-automated NSP enrichment more than 5000 protein groups could be quantified (fig. 16, panel B).

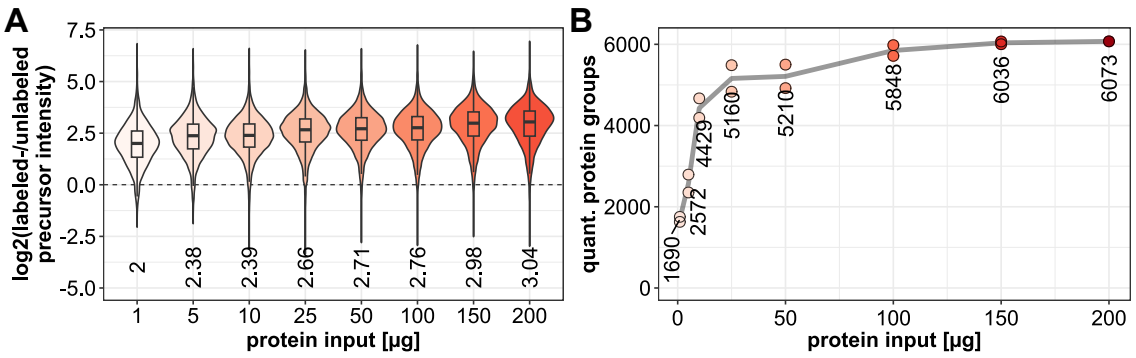


Figure 16: Applying SILAC plexDIA for the analysis of NSP enriched samples generated via the semi-automated enrichment protocol, and determining the optimized range of protein input for newly synthesized proteome analysis. A) SILAC intensity ratios of samples prepared via the semi-automated NSP enrichment protocol with differing amounts of protein input. To quantify the efficiency of the NSP enrichment, intensities of heavy- and intermediate SILAC labeled precursors (originating from newly synthesized proteins) are contested to light precursor ion intensities (originating from pre-existing proteins). B) Number of quantified protein groups in enriched NSP samples prepared with differing amounts of protein input. Data are based on 2 technical replicates, except the 200 μg sample which consists of a single replicate. This figure was taken and modified from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

Given the relatively low protein input requirements for the preparation of samples with the semi-automated enrichment protocol, cells can be grown and labelled in smaller culture dishes, such as 6-well plates. To further simplify the sample preparation protocol, cell lysates were sonicated using adaptive-focused acoustic (AFA) technology using a Covaris LE220R. Through the use of AFA-based ultrasonication, up to 96 lysates can be processed in parallel. Moreover, probe sonication of lysates with $<$200 μL volume can prove challenging with regards to sample losses and spillage, marking another advantage of the AFA-based sonication.

5.1.4 An integrated workflow for the quantitative analysis of the newly synthesized proteome (QuaNPA)

Taken together, the semi-automated click chemistry-based enrichment of AHA-containing NSP with MAA beads and SILAC plexDIA analysis can be combined into an optimized workflow for quantitative analysis of the newly synthesized proteome (QuaNPA). The workflow can be divided into four main steps. The first step includes the preparation of magnetic alkyne agarose (MAA) beads, through the coupling of epoxy-activated magnetic agarose

with propargylamine. In the second step, combined metabolic labeling of cultured cells is performed with AHA and SILAC amino acid-containing media. Cell lysis and protein extraction is subsequently performed using AFA-based ultrasonication with a Covaris LE220R. For comparative analysis and relative quantification, cell lysates with heavy- or intermediate SILAC labels are mixed. In the third step of the workflow, semi-automated enrichment of NSP and subsequent SP3-based peptide purification is carried out with an Agilent Bravo liquid handling platform. In the fourth and final stage of the workflow, LC-MS measurements of the samples is carried out using either DDA or DIA acquisition methods. The generated data can be analyzed using Maxquant or other DDA compatible search engines, or via the plexDIA features of DIA-NN for DIA data (fig. 17) [180].

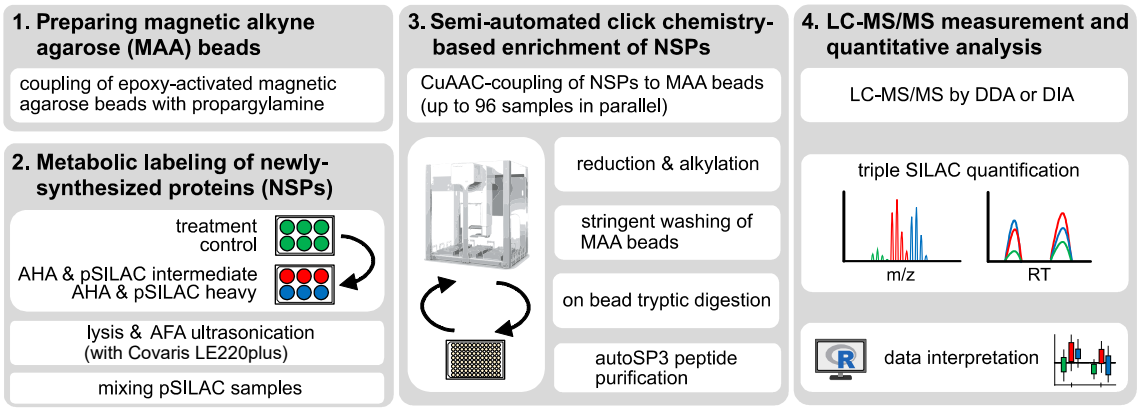


Figure 17: Schematic representation of the integrated, semi-automated workflow for the quantitative analysis of the newly synthesized proteome (QuaNPA). This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

5.1.5 Applying the QuaNPA workflow for the analysis of the newly synthesized proteome in response to IFNg stimulation

To ensure that the QuaNPA workflow is capable of generating biological insights with high sensitivity, an experiment was performed in which Hela cells were treated with either 10 ng/mL recombinant interferon gamma (IFNg) or 0.5 % (w/v) bovine serum albumin (BSA). The increased throughput of the QuaNPA workflow in terms of sample preparation and LC-MS measurement via plexDIA, simplifies the analysis of large numbers of samples, enabling time-resolved analysis of cellular responses to perturbations, such as stimulation with IFNg. Since the IFNg signaling pathway and its multiple downstream targets have been

extensively studied, the results of the quantitative newly synthesized proteome analysis can be compared to reported results, obtained using different methodologies. Due to the different duration of IFNg stimulation in the time course experiment, the metabolic labeling steps were also carried out at with different lengths. HeLa cells treated with IFNg for 2-, 4- and 6 h were labeled with AHA- and SILAC amino acid containing medium in parallel to the treatment, whereas cells treated with IFNg for 9 h and 24 h were pulse-labeled in the last 6 h of the treatment period (fig. 18, panel A). Interestingly, the different metabolic labeling periods did not lead to significant differences in the number of quantified proteins and quantitative precision of the samples (fig. 18, panel B-C). Using Principal component analysis, the individual replicates for each time point cluster together and are arranged according to the duration of IFNg treatment along component 2, indicating a reproducible and progressive effect of IFNg on the newly synthesized proteome. Differential expression analysis revealed that up-regulation of several well characterized targets of IFNg, such as STAT1, ICAM1 and TAP1 can be measured with IFNg stimulation of only 2 h (fig. 18, panel D). Over-representation analysis of the subset of up-regulated NSP revealed significant enrichments of annotated targets of IFNg signaling, proteins which are directly involved in the interferon- and innate immunity signaling networks, proteins whose promoters were shown to be bound by STAT1 in response to IFNg stimulation and proteins whose mRNAs were shown to be significantly up-regulated in HeLa cells following IFNg stimulation (CITE PAPER WITH RNAseq) (fig. 18, panel E).

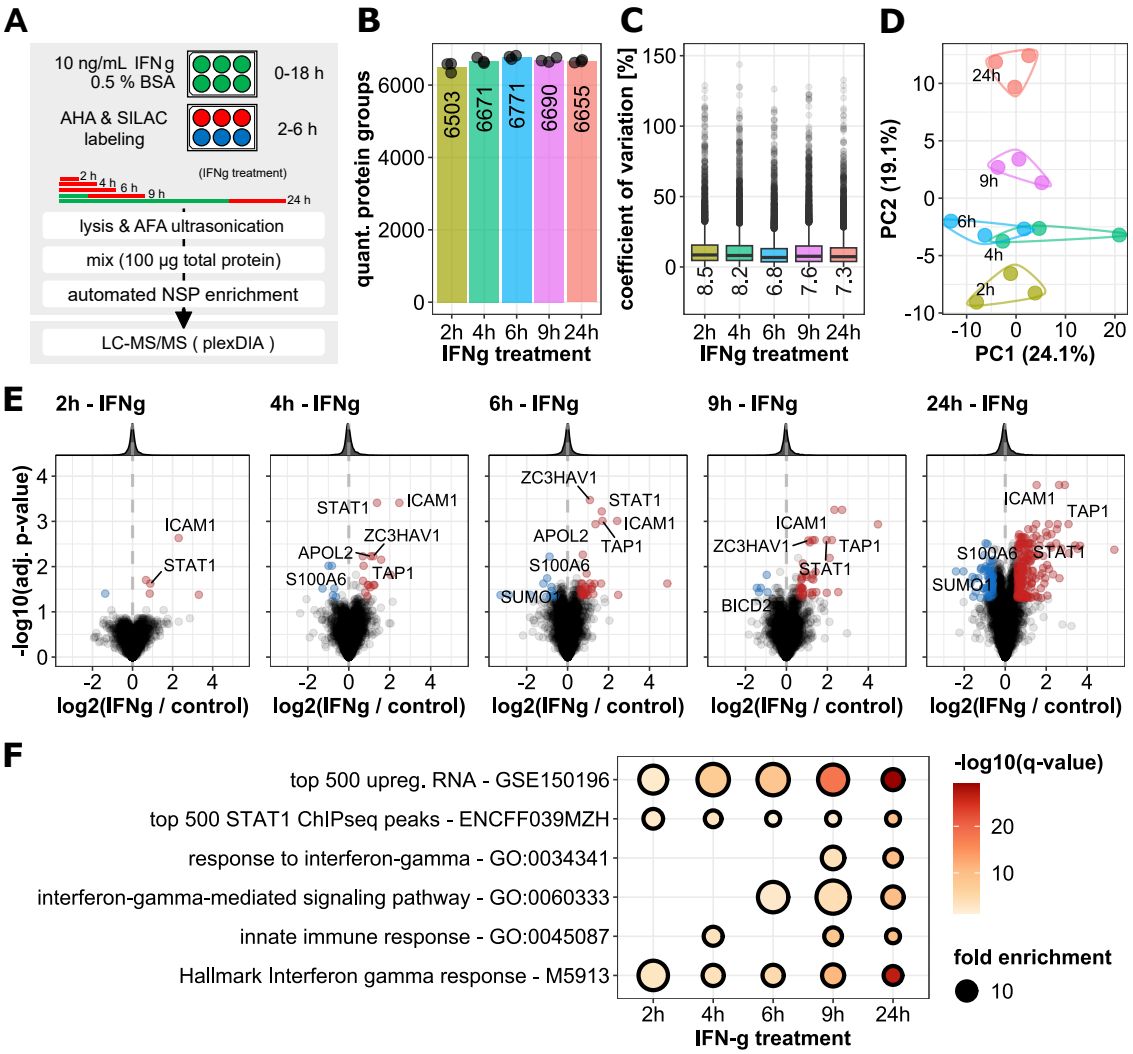


Figure 18: Newly synthesized proteome analysis in HeLa cells treated with interferon gamma (IFNg). A) Schematic representation of the experimental design and QuaNPA workflow. The IFNg treatment time is indicated using text labels and via the length of the colored bars. B) Average number of quantified protein groups in the samples with the indicated IFNg treatment time points. C) Box plots indicating the coefficient of Variation (CV) values of the quantified protein groups for each time point. D) Principal component analysis (PCA) of the NSP samples. Colors in panels B-D indicate treatment time points. E) Volcano plots of the IFNg-induced changes in the newly synthesized proteome. Significantly up-regulated proteins (adjusted p-value < 0.05 and log2 fold change > 0.585) are highlighted in red and significantly down-regulated proteins (adjusted p-value < 0.05 and log2 fold change $<$ -0.585) are highlighted in blue. F) Dot plot highlighting over-represented sets of proteins, which are significantly up-regulated in response to IFNg at the indicated time points (q-value < 0.05). This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

By combining the newly synthesized proteome data from the five different time points the temporal profiles of the differentially regulated NSP can be evaluated. The subset of differentially regulated proteins ($|\log2 \text{ fold change}| > 1$ and adj. p-value < 0.05, at any of

the 5 time points and CV $<$ 20 % across all time points) were divided into three subgroups to distinguish early (2 h), intermediate (4-9 h) and late (24 h) responses to IFNg stimulation. Additionally, to assess the level of evidence for the differentially regulated proteins to be targets of IFNg signaling, data was retrieved from the interferome database [172] which lists in how many reports the individual proteins was found to be differentially expressed in the context of IFNg stimulation in human cells or tissues. For the majority of proteins, especially for those of the early and intermediate response subgroups, numerous instances of differential expression have been reported in response to IFNg stimulation. Interestingly, several proteins which have not previously been described as targets of IFNg signaling were found to be differentially regulated. Among these potentially novel candidates, SIN3B and CRK were up-regulated, whereas SUMO1, S100A6 and BICD2 were down-regulated in response to IFNg (fig. 19).

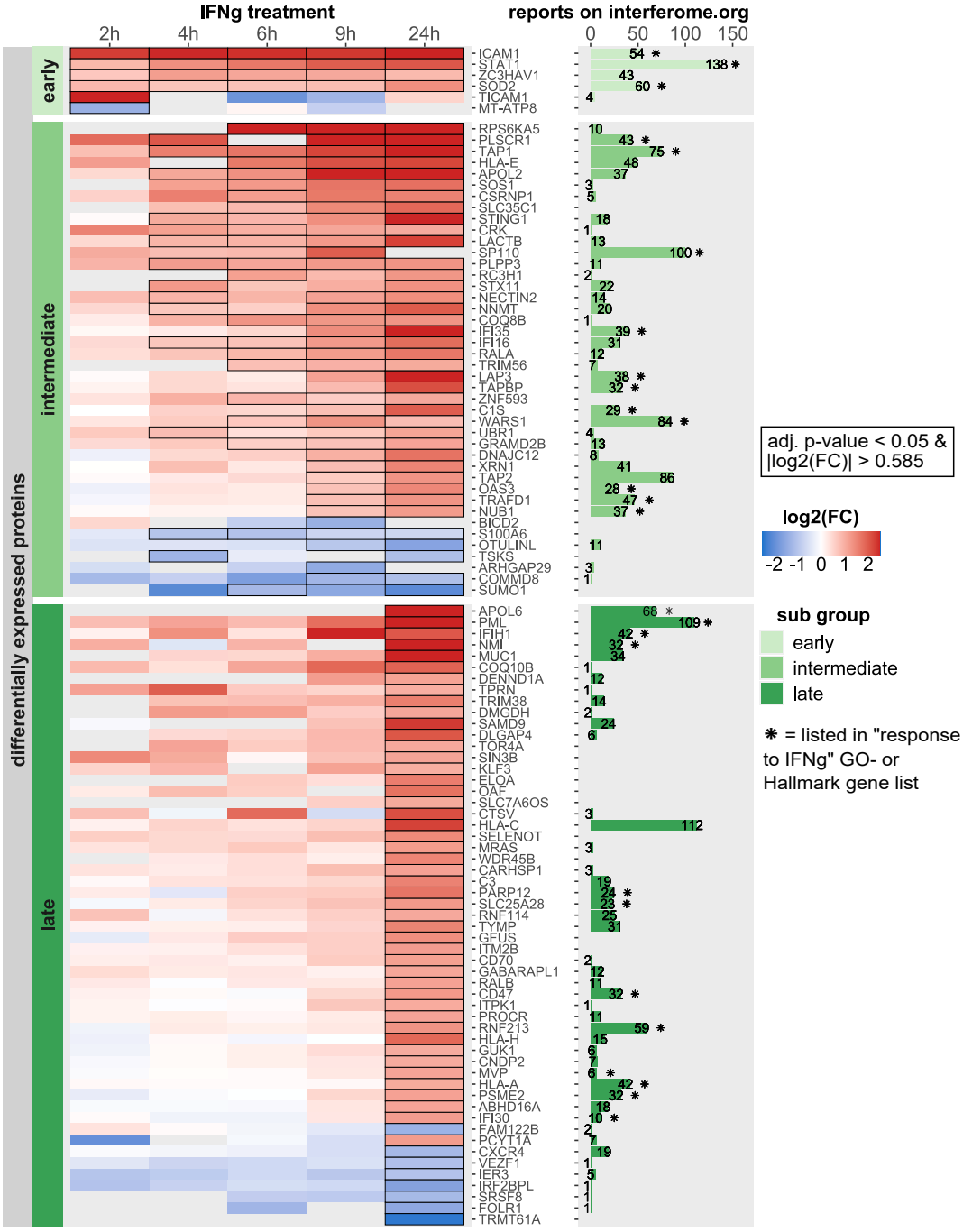


Figure 19: Heatmap of NSP with differential expression in response to IFNg treatment. The subset of differentially expressed proteins ($|\log2 \text{ fold change}| > 1$ and adj. p-value < 0.05, at any of the 5 time points and CV < 20 % across all time points) were selected and divided into three groups depending on the earliest time point in the time course their change in expression reached statistical significance (early: 2h; intermediate: 4-9h; late: 24h). Significant changes in NSP abundance are indicates with black outline in the cells of the heatmap. The colour gradient indicates the log2 fold change values of the NSP. The bar graph on the right-hand side of the heatmap indicated the number of datasets in the interferome database, in which a fold change >2 or <0.5 was reported in human cells or tissues for the respective protein. The asterisk sign ($*$) on top of bars indicates whether the protein is listed in the response to IFNg Hallmark or GO-terms. This figure was taken and modified from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

To validate these potentially novel candidates of IFNg signalling, targeted proteomic analysis via label-free parallel-reaction monitoring (PRM) was performed. Relative changes in protein abundance were analyzed in HeLa cells after 4 h and 24 h long treatments with either IFNg or BSA. In addition to the previously listed novel candidates, canonical target proteins ICAM1, STAT1 and TAP1 were included in the analysis as positive controls. Interestingly, only ICAM1 protein abundance was found to be moderately up-regulated (with borderline significance) after 4 h treatment with IFNg. In contrast to the measurement of total protein abundance, strong up-regulation of ICAM1 could be detected in the newly synthesized proteome as early as 2 h after exposure to IFNg (fig. 18, panel D). After 24 h treatment with IFNg, strong up-regulation of TAP1, STAT1 and ICAM1 were detected. Moreover, moderate up-regulation of SIN3B and down-regulation of BICD2 protein abundance could be measured via PRM, providing additonal evidence for the results of the newly synthesized proteome analysis. However, no significant changes in protein abundance could be measured for the remaining candidates SUMO1, CRK and S100A6 (fig. 20).

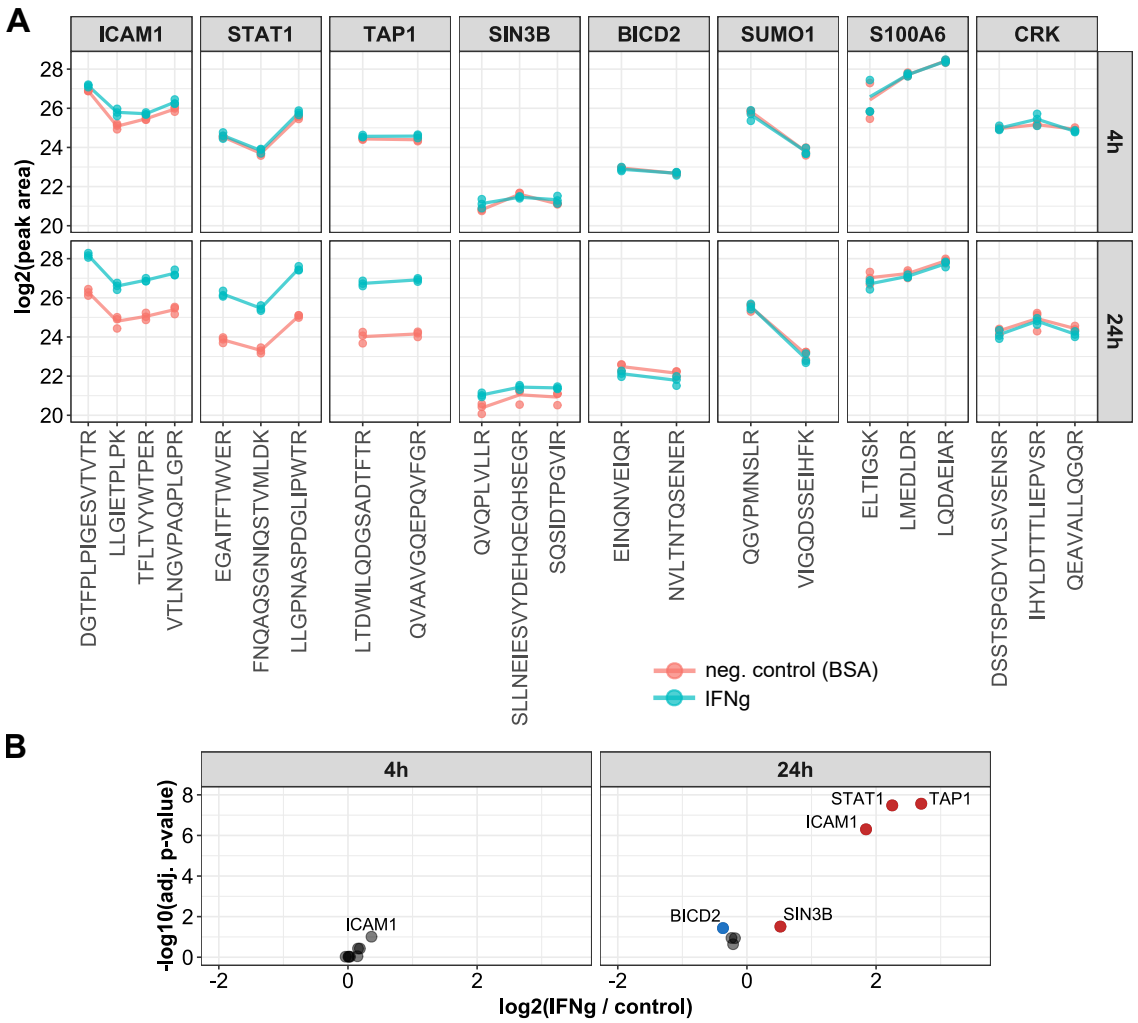


Figure 20: Targeted proteomic analysis of the novel candidate IFNg target proteins. A) Overview of the TIC normalized peak areas of the targeted peptides, analyzed in HeLa cells treated with IFNg or 0.5 % BSA for 4 h and 24 h, using PRM measurements. B) Volcano plots of the IFNg-induced protein abundance changes, after 4 h and 24 h treatment. Significantly up-regulated proteins (adjusted p-value $<$ 0.05 and log2 fold change $>$ 0.585) are highlighted in red and significantly down-regulated protein (adjusted p-value $<$ 0.05 and log2 fold change $<$ -0.585) are highlighted in blue. Data are based on 3 experimental replicates. This figure was taken and modified from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

In summary, the QuaNPA workflow is capable of quantifying proteome-wide changes in NSP levels in response to IFNg-treatment, in a time-resolved manner. Significant changes in the newly synthesized proteome could be detected at various time points of IFNg treatment, distinguishing immediate and delayed responses, indicating primary and secondary regulatory targets of IFNg signaling. Differentially regulated NSP included induction or repression of established IFNg target proteins as well as several novel candidates. Using targeted proteomic analysis additional evidence could be obtained for two of these novel targets,

demonstrating the power of QuaNPA to infer novel and meaningful biology.

5.2 Quantifying inhibition of translation in newly synthesized proteome analysis

Inhibition or depletion of crucial regulatory factors of mRNA translation, such as essential initiation factors (eIF) are likely to lead to global down-regulation of protein synthesis and therefore the majority of NSP. Using conventional normalization methods this global shift in NSP levels and protein synthesis is removed, which does not accurately represent the biological effects. In order to quantify the global inhibition of protein synthesis in newly synthesized proteome analysis, an alternative normalization strategy was developed.

Newly synthesized proteome analysis and most conventional quantitative proteomic analysis workflows use equal amounts of protein or peptide starting material for their sample preparation and apply normalization methods, which operate under the assumption that the majority of proteins are unchanged between the compared samples [55] [181]. For the analysis of newly synthesized proteome data this assumption might not be suitable for certain experiments, since under stress conditions canonical cap-dependent mRNA translation mechanisms are inhibited, resulting in down-regulation of protein synthesis in cells [182].

Although equal amounts of protein input are used as starting material for the enrichment of NSP, if the rate of protein synthesis is reduced in one of the tested conditions, these unequal fractions of NSP can also be detected after the enrichment (fig. 21, panel A). However, by default SILAC-based quantification workflows apply median normalization to the calculated precursor or protein ratios, inadvertently masking the global shift through subtraction of the median ratio value. Due to this centering of the SILAC ratio distribution to a value 1, conventional median normalization cannot be applied, if the goal is to quantify global up- or down-regulation of mRNA translation in newly synthesized proteome analysis. Avoiding median- or other normalization approaches, raises additional challenges, since extremely high accuracy in the mixing step of lysates prior to the enrichment of NSP is required to accurately quantify down-regulation of the majority of proteins. Especially when working with low volumes of lysates as input material, inaccuracies in the pipetting steps for the protein

quantification via BCA assay, or the mixing of lysates are common. The potential deviation of the relative differences in total protein and NSP abundance, through mixing pSILAC-labeled cell lysates, is propagated and increases in a linear relationship with the errors in mixing step (fig. 21, panel B). The inaccuracy for the mean of heavy- over intermediate SILAC ratios (ΔR), is described through eq. (3):

$$\Delta R = \sqrt{\left(\frac{1}{H}\Delta H\right)^2 + \left(\frac{-H}{M^2}\Delta M\right)^2} \tag{3}$$

where:

ΔR = uncertainty for the mean of the heavy over intermediate SILAC ratio; H = abundance of the heavy labeled proteins; M = abundance of the intermediate labeled proteins; ΔH = inaccuracy of the heavy labeled protein abundance; ΔM = inaccuracy of the heavy labeled abundance.

A potential solution for the accurate determination of the average ratio of NSP, in the two compared conditions, is the measurement of unlabeled, pre-existing proteins and their heavy- or intermediate SILAC labeled NSP counterpart without mixing of lysates and performing an enrichment of NSP. For each sample, ratios between the SILAC pairs can be determined and averaged to serve as metric for the rate of protein synthesis in the respective condition. Although the depth and quantification accuracy for the individual SILAC ratios is slightly lower in the sample without NSP enrichment, the median values are nonetheless accurate, as was demonstrated in the plexDIA benchmark (fig. 15, panel E). Moreover, plexDIA-based SILAC quantification spans a wide range of linear quantification accuracy for both heavy- and intermediate SILAC channels, enabling accurate determination of median SILAC ratios even in the low percentage range (fig. S3). Assuming that the majority of pre-existing protein levels are unchanged between the tested conditions, the calculated median ratios of previously- and newly synthesized proteins can be compared between samples to determine the average rate by which NSP are up- or down-regulated, and can serve as normalization factor. Normalization of the individual fold change or SILAC ratios can then be applied by shifting their median to the determined average ratio (normalization factor), which was determined in the previous step (fig. 21, panel B and C).

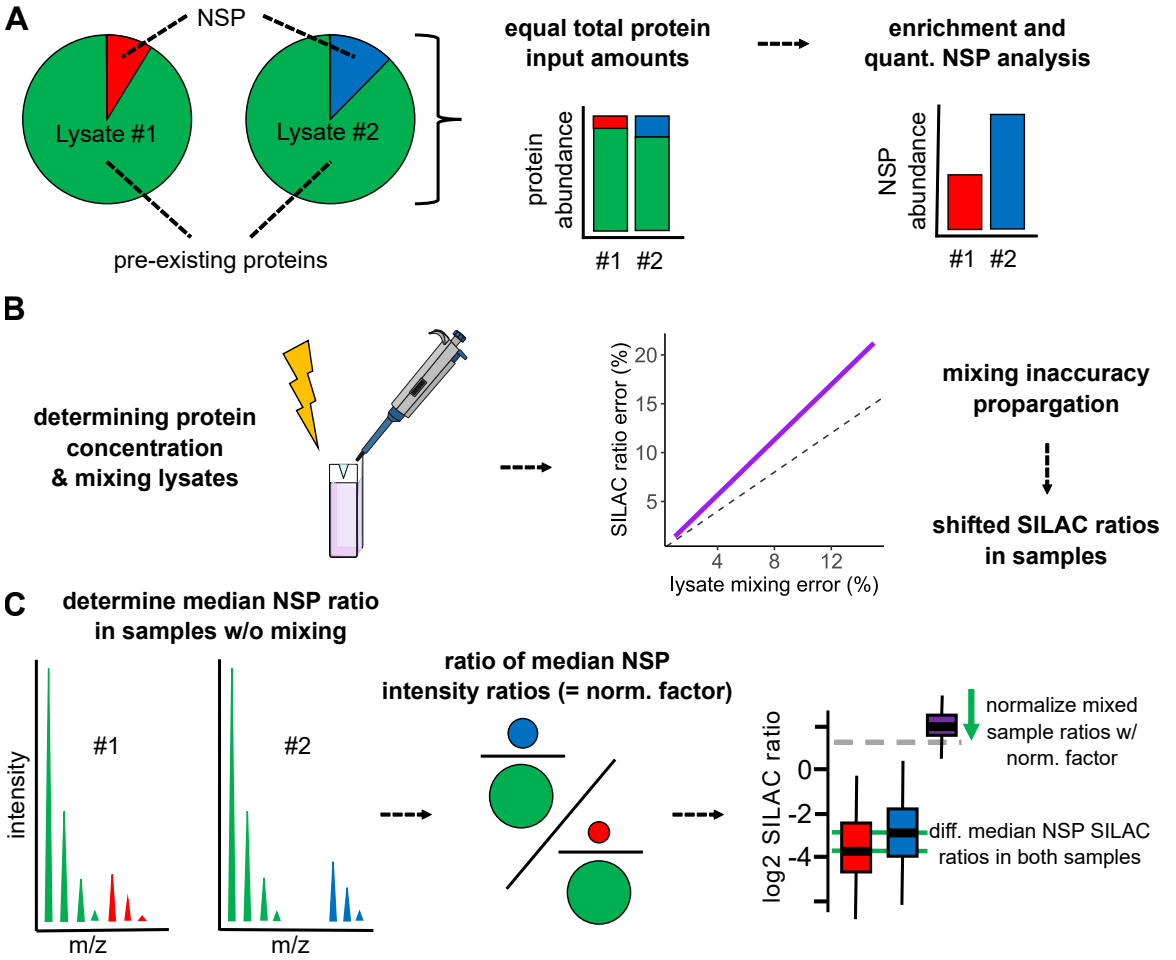


Figure 21: Schematic representation of the proposed normalization method for newly synthesized proteome analysis. A) Equal amounts of total protein content serve as input for the generation of newly synthesized proteome samples, prior to the click chemistry-based enrichment of AHA-containing NSP. In the case of inhibition or up-regulation of protein synthesis in on of the tested conditions, the fraction of NSP in the lysate differ. The differing overall abundance of NSP in the tested conditions are preserved in the click chemistry-based enrichment step. B) Without the use of conventional median normalization, the determined ratios of SILAC labeled NSP, the mixing of lysates prior to the enrichment step poses a potential source of errors. Propagation of errors in the mixing of cell lysates, caused by inaccuracies in protein concentration measurements, can cause shifted distributions of SILAC ratios in mixed samples. Propagation of the median SILAC ratio error can be described through a linear relationship with the error in the mixing step. C) By measuring the intensity ratios of SILAC labeled proteins and pre-existing (unlabeled) proteins, in samples generated without NSP enrichment, the average rate of protein synthesis can be inferred and serves as a normalization factor, on which the log2 fold change distributions are centered. The vector graphic of the micro pipette was created by Nicolàs Palacio-Escat (CC-BY 4.0). The vector graphic of the UV cuvette was created by “Servier” (CC-BY 3.0) and retrieved from “bioicons.com”.

In order to not mask biological effects of translation inhibition in newly synthesized proteome analysis, it is crucial to ensure that through the proposed normalization approach, the QuaNPA workflow is capable of quantifying global inhibition of protein synthesis.

Therefore, an experiment was performed in which HeLa cells were treated with different concentrations of either translation inhibitory compound cycloheximide (CHX) (0.1-100 μg/mL) or dimethylsulfoxide (DMSO) (0.0001-0.1 % v/v), during the metabolic labeling of NSP. Since CHX acts as potent inhibitor of mRNA translation elongation [183], no NSP should theoretically be detected and quantified in cells treated with high doses (fig. 22, panel A). In addition to the analysis of enriched newly synthesized proteome samples, global proteomic profiling was carried out using the same cell lysates of treated cells.

Surprisingly, even with the highest tested dose of 100 μg/mL no significant changes in protein abundance could be determined after 4 h treatment, using conventional proteome profiling (fig. 22, panel B).

To assess the rate of NSP identification in the CHX treated cells, the newly synthesized proteome samples were measured both DIA and DDA acquisition methods and processed using the DIA-NN or Maxquant software. The plexDIA-based analysis proved to be highly effective in determining high numbers of protein groups in the DMSO treated samples, whereas not a single NSP was identified in cells treated with a concentration range of 1-100 μg/mL CHX for 4h. Only in cells which were treated with the lowest dosage of 0.1 μg/mL CHX, approximately 4600 NSP were identified. DDA based analysis and processing with Maxquant identified $>$3500 protein groups in 0.1 % (v/v) DMSO treated HeLa cells, and 117 NSP in cells treated with 100 μg/mL CHX. The use of the “re-quantify” function, which is frequently used in the analysis of SILAC data to increase data completeness by calculating ratios with background intensities, raises the number of identified protein groups to 3047 (fig. 22, panel C). These results indicate that the QuaNPA workflow and especially plexDIA based analysis are capable of detecting the global inhibition of protein synthesis and capturing the absence of NSP in cells in which protein synthesis is inhibited.

Furthermore, the newly synthesized- and global proteome data of HeLa cells treated with 0.1 μg/mL was used to determine the degree of protein synthesis inhibition, as described in the previous section (fig. 21, panel C). The average intensity ratio of NSP over pre-existing proteins in CHX-treated cells was determined at 0.074 (-3.74 when log2 transformed) and 0.157 in DMSO treated cells, translating into an average ratio of 0.48. The average NSP

ratio metric indicates that the median NSP intensity, and protein synthesis rate by proxy, are halved upon treatment with 0.1 μg/mL CHX for 4 h (fig. 22, panel D-E). Normalization of the quantitative newly synthesized proteome data, through centering the fold change values on the determined average NSP ratio value, enables accurate NSP quantification between the two conditions (fig. 22, panel F). Moreover, it is possible to distinguish proteins which are truly up-regulated and those whose expression levels are comparable between the two conditions, whereas through conventional median normalization it remains unclear if observed up-regulation is caused via normalization artifacts.

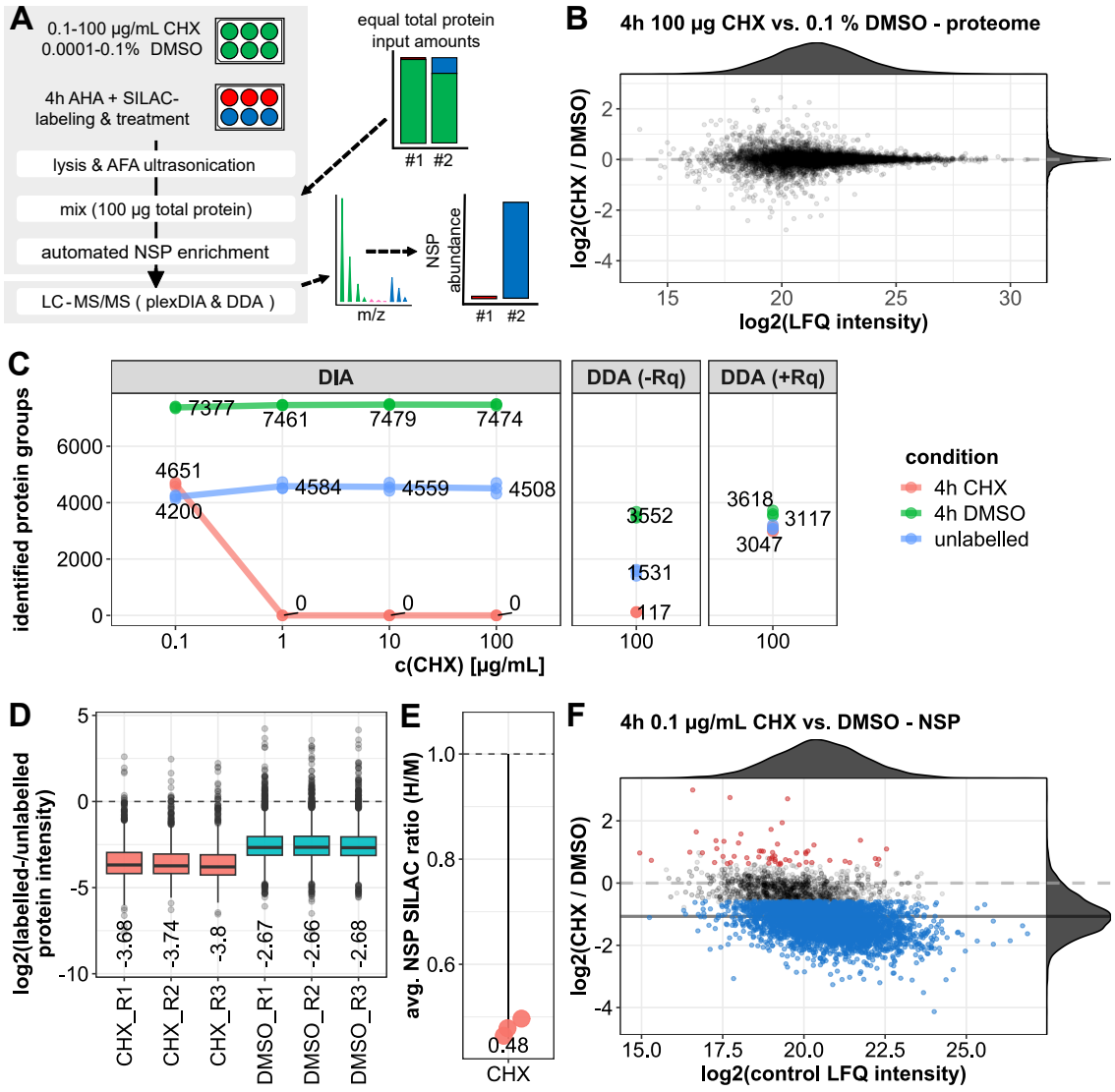


Figure 22: Quantitative newly synthesized proteome analysis of HeLa cells treated with different doses of cycloheximide. A) Schematic representation of the experimental design and analysis workflow. B) MA-plot of proteomic profiling data, generated from HeLa cells treated with either 0.1 % DMSO or 100 μg/mL CHX. 6282 protein groups, quantified in at least 2 out of 3 replicates per treatment condition, were included in the differential expression analysis. C) Number of NSP that were identified in the respective treatment condition, using the indicated acquisition method and analysis software. DIA data were analyzed using DIA-NN version 1.8.1, DDA data were processed using Maxquant version 2.0.3. D) Log2 transformed ratio of heavy or intermediate SILAC labeled NSP over unlabeled pre-existing proteins in lysates of cells treated with either 0.1 μg/mL CHX or 0.0001 % DMSO for 4h. E) Average NSP ratio (normalization factor), indicating a 52 % reduction of protein synthesis in the CHX treated cells. F) MA plot of the newly synthesized proteome data of HeLa cells treated with 0.1 μg/mL CHX versus 0.0001 % DMSO for 4h. Using the previously described normalization approach it can be demonstrated that the majority of all protein groups are down-regulated.

The subset of newly synthesized proteins, which can be quantified in samples without NSP enrichment, are highly abundant proteins with high rates of synthesis. This conclusion can

be drawn from the high iBAQ intensity in both the proteome and newly synthesized proteome of DMSO treated HeLa cells (fig. 23, panel A). Furthermore, the subset of NSP that is used for normalization did not show differential expression patterns in the respective proteomic data. This observation indicates that the assumption of equal levels of pre-existing proteins for the normalization of the two experimental conditions is valid (fig. 23, panel B).

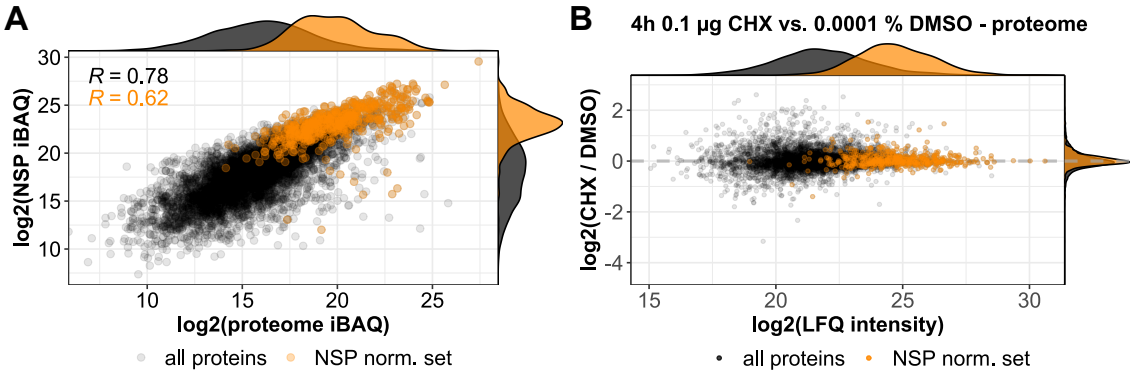


Figure 23: Abundance and fold change values of the subset of NSP, whose average SILAC ratios can be used for the normalization of newly synthesized proteome data in cycloheximide treated HeLa cells. A) Protein abundance, measured as iBAQ intensity, in the proteome and newly synthesized proteome data of DMSO treated HeLa cells. The subset of NSP, which can be quantified in samples without click chemistry-based NSP enrichment ("NSP norm. set") are highlighted in orange. The Pearson correlation coefficient (R) of the iBAQ intensities is indicated in the top left corner. B) MA-plot of proteomic profiling data, generated from HeLa cells treated with either 0.0001 % DMSO or 0.1μg/mL CHX. The subset of NSP, which can be quantified in samples without click chemistry-based NSP enrichment and are used for normalization ("NSP norm. set") are highlighted in orange.

In summary, newly synthesized proteome analysis via the QuaNPA workflow is capable of accurately determining the loss of protein expression, and performing relative quantification upon reduced global protein synthesis, caused by inhibition of mRNA translation which is to be expected for the depletion of essential eIF proteins.

Through the additional measurement of samples, generated from the pulse-labeled lysates that serve as input for the NSP enrichment, average intensity ratios can be calculated to assess the relative change in the rate of protein synthesis between tested conditions. This NSP intensity ratio metric serves as a normalization factor for the newly synthesized proteome data.

5.3 Systematic analysis of changes in protein synthesis, induced by depletion of initiation factor proteins (eIF)

Having established the QuaNPA workflow, which greatly simplifies quantitative newly synthesized proteome analysis, and demonstrating the ability to quantify global inhibition of protein synthesis, large scale systematic analysis of changes in mRNA translation through perturbations of individual initiation factor proteins (eIFs) becomes feasible. In this study, 40 core eIF and eIF-interacting proteins were analyzed.

5.3.1 Engineered RKO cell line model system for the inducible depletion of eIF

To determine the contributions of individual eIFs to protein synthesis in cancer cells, a model cell line with inducible depletion of the targeted eIF was created by the main collaborator of this project, Robert Kalis, a PhD student in the group of Prof. Johannes Zuber of the Institute of Molecular Pathology (IMP) in Vienna. A clonal cell line of genetically modified, human RKO colorectal cancer cells [184], with inducible spCas9 and eGFP expression, served as template for the generation of a cell line model with inducible eIF depletion. The RKO cell lines were further modified through retroviral transduction, introducing a vector containing a neomycin resistance gene, Thy1.1 selection marker and two sgRNA which target the selected eIF gene. Selection of the genetically modified RKO cell lines was carried out through antibiotic selection with geneticin (G418) and fluorescence activated cell sorting (FACS) selecting Thy1.1-positive cells [185]. The panel of cell lines were shipped to Heidelberg and expanded, by culturing with G418-containing RPMI-1640 media. Aliquots of these expanded cells were subsequently used for experiments. In contrast to the previous selection and expansion of the RKO cells, no antibiotics such as G418 were included in the culture media to avoid additional stress on the mRNA translation machinery of the cells. Depletion of the targeted eIF proteins was triggered through the addition of 200 ng/mL doxycycline, inducing the expression of spCas9 and the editing and genetic knockout of the target eIF gene. Newly synthesized- and global proteome analysis was carried out, using the QuaNPA workflow. Additionally, cells were also harvested and shipped to Vienna for RNA sequencing (RNA-seq) analysis, which was also carried out by Robert Kalis with the help of the next generation sequencing (NGS) core facility of the IMP Vienna (fig. 24).

In order to control for the effects of doxycycline treatment and sgRNA transduction and expression, RKO cells expressing sgRNA targeting the integration site of Adeno-associated virus DNA on chromosome 19 q13.3-qter (AAV1) [186] were included as controls for all experiments.

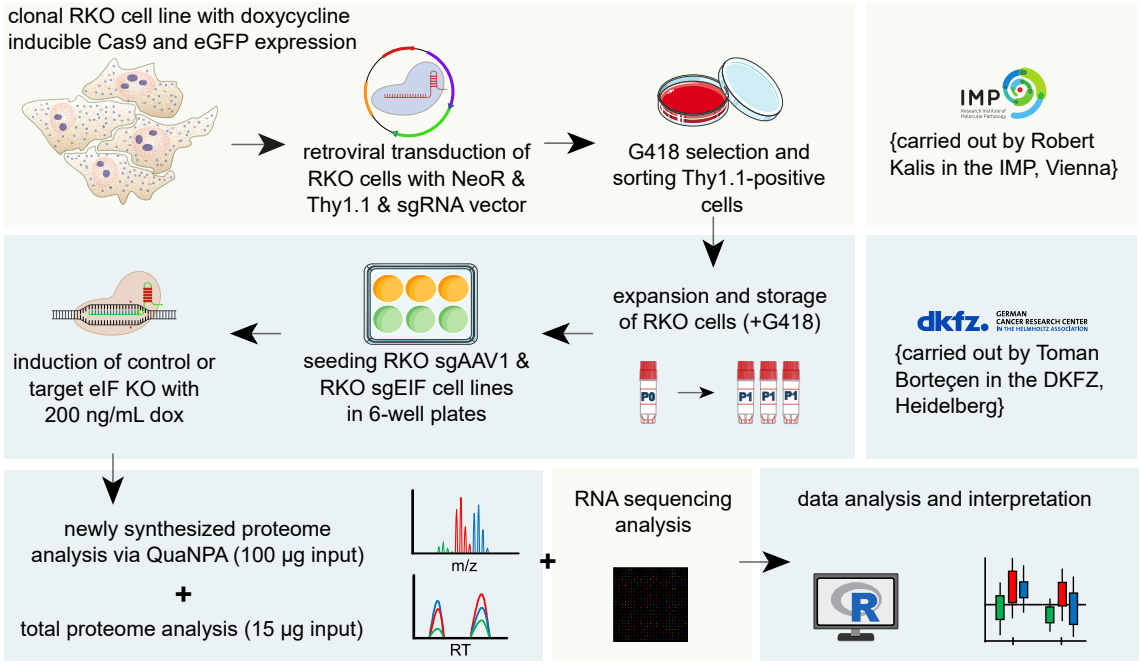


Figure 24: Schematic representation of the experimental design for the systematic analysis of changes in protein synthesis, induced by the depletion of individual translation initiation factors (eIF). Clonal RKO cell lines with inducible expression of spCas9 and eGFP were genetically modified to express two sgRNA, targeting the selected EIF genes to serve as model system for the analysis. The RKO cell lines, which were prepared by Robert Kalis in the Institute of Molecular Pathology in Vienna, were shipped to the laboratory of Prof. Krijgsveld, where all experiments and analysis (except for RNA sequencing) were performed. The systematic analysis of protein synthesis changed integrates data from newly synthesized- and global protein abundance and RNA abundance to calculate changes in mRNA translation. The icon depicting cells of a human cell line was created by “DBCLS” (CC-BY 4.0), Crispr-plasmid and Cas9 icon was created by “Marcel Tisch”, the 6-well plate and petri dish icon were creaed by “Servier“ (CC-BY 3.0) and the cryo vial icon was created by “Pauline Franz” (CC0). All listed icons were retrieved, as vector graphics, from “bioicons.com”.

Upon induction of spCas9 expression via addition of 200 ng/mL doxycycline, the DNA double strand breaks are induced between the two sgRNA cut sides leading to deletions within the EIF genes. These DNA deletions can be visualized using agarose gel-electrophoresis, of PCR amplified DNA of the targeted EIF gene, or through Sanger sequencing. Partial spCas9-mediated genetic editing of the targeted EIF gene can be detected as early as 8 h after induction, but appear to be more efficient at later time points such as 24 h post

induction, in the example of RKO cells with induced eIF4E KO (fig. 25, panel A). Sanger sequencing electropherograms of amplified EIF4E DNA show loss of interpretable sequence traces after the cut side of one of the sgRNA (fig. 25, panel B).

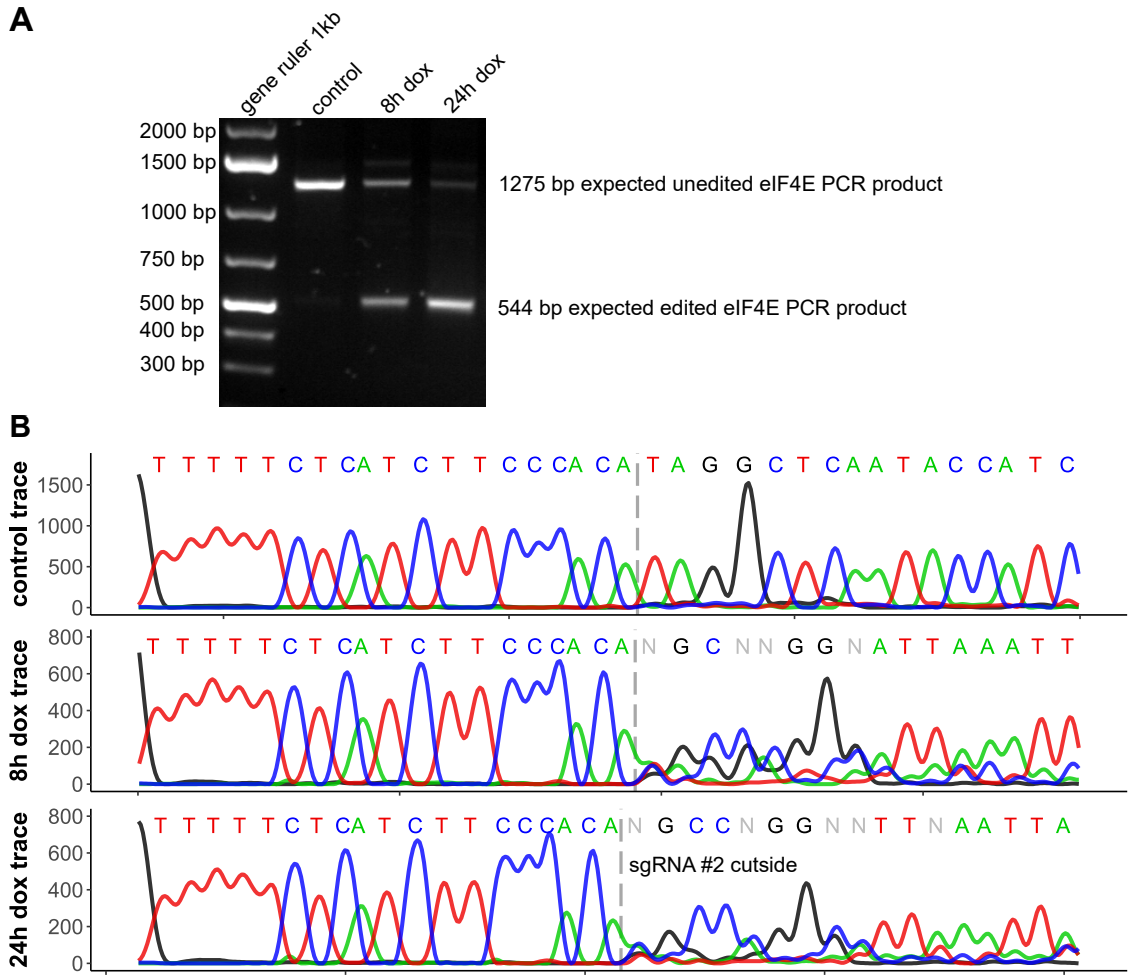


Figure 25: Measuring genetic editing in RKO cells expressing two sgRNAs targeting EIF4E. A) Agarose gel-electrophoresis of PCR amplified EIF4E DNA, covering the two sgRNA binding sites. The time points of doxycycline induced spCas9 expression are indicated above the lanes. B) Sanger sequencing electropherogram of the amplified EIF4E DNA, at the indicated time points of KO induction. The reverse primer (5′-GTCACTTCGTCTCTGCTGTT-3′) was used for sequencing of the PCR products. Cutsides of the sgRNA (5′-GGATGGTATTGAGCCTATGTGGG-3′) are indicated in the electropherograms.

Although the editing of the targeted EIF genes can be measured on the DNA level, for validation of the protein depletion, protein abundance levels need to be determined.

5.3.2 Determining the optimized experimental conditions for the analysis of changes in protein synthesis, caused by eIF depletion

Initial characterisation of the induced eIF knockout (KO) effects on the RKO cells were carried out by measuring cell proliferation. The analysis of cell proliferation of RKO cells with induced eIF KO was carried out by Robert Kalis, via a FACS-based competition assay. For the majority of targeted eIF, reduced proliferation can be observed after 72 h of knockout induction. Not all targeted eIF, however, have an effect on the RKO cells ability to proliferate upon depletion, as was measured for eIF1B, eIF3H, eIF3K, eIF3L, eIF4A2, eIF4E1B, eIF4E3 and eIF4EBP1. Interestingly, even increased proliferation could be observed in response to the induced KO of PDCD4, a known antagonist of eIF4A1 which is major regulator of canonical translation initiation [187]. The measured cell count ratios are not uniform for the panel of targeted eIF at distinct time points. In the case of several eIF, such as eIF3C, eIF4B and eIF4G1, reduced proliferation was measured as early as 48 h after induction of their knockout (fig. 26).

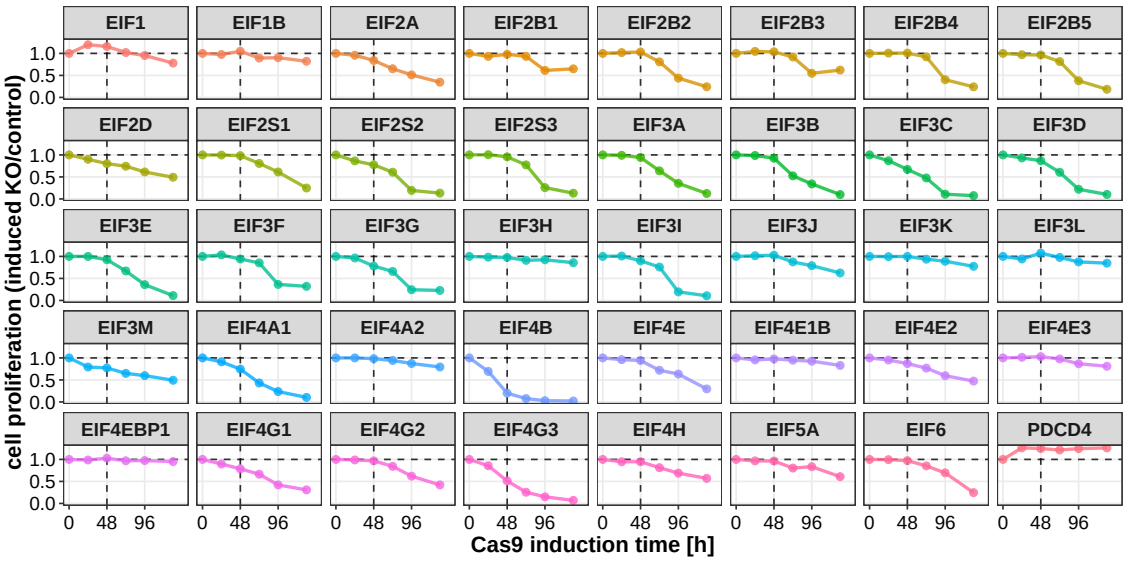


Figure 26: Changes in proliferation of the RKO cell lines, following the induced KO of the indicated eIF, measured over time via a FACS-based competition assay. Ratios of cell counts of RKO cells with induced eIF KO over controls are depicted. The experiment was performed by Robert Kalis in the Institute of Molecular Pathology Vienna.

Ideally, all analyses in the study would be performed with the RKO cells prior to the onset of reduced cell proliferation or viability, with comparable levels of target eIF protein depletion. Due to the vastly different effects on cell proliferation and variation in mRNA and protein

half-lifes of the targeted eIF this would lead hugely complex and impractical experimental design. Instead a uniform time point of 48 h KO induction was selected for the RKO cell line panel. Without the few previously mentioned eIF (3C, 4B and 4G3) no significant changes in cell proliferation were detected at this timepoint (fig. 26). A major concern however, is whether the difference in cell counts, as measured in the competition assay is due to reduced proliferation or induction of apoptotic cell death. To ensure that cell viability is high in the cells with induced KO of eIF, that were shown to strongly reduce cell proliferation, such as eIF4B and 4G3, microscopy images were take and compared to the control (AAV1) KO cells. Reduced numbers of RKO cells with induced KO of eIF4B and eIF4G3, compared to control AAV1 KO cells, could be observed after 48 h induction. However, no accumulation of dead or dying cells could be observed. Furthermore, the average size of cells appeared to increased in cell with the induced KO of eIF4B and eIF4G3 (fig. 27).

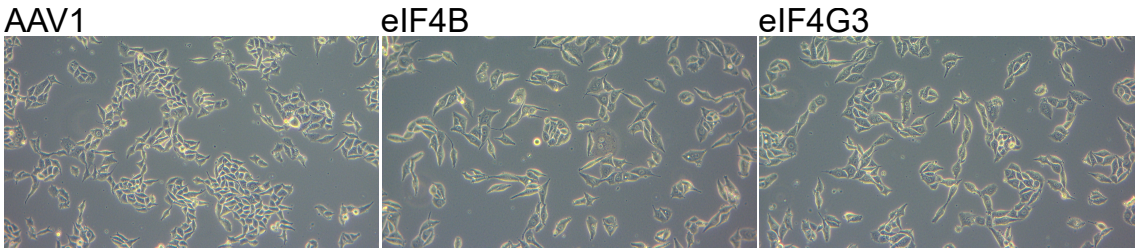


Figure 27: Brightfield microscopy images of RKO cells with induced KO of AAV1, eIF4B and eIF4G3, 48 h after induction with 200 ng/mL doxycycline.

In addition, cell viability measurements were performed for RKO cells with induced KO of eIF4A1 and AAV1. The target eIF4A1 was selected, since it was also shown to reduce cell proliferation upon induced KO (fig. 26). By subtracting the luminescent signal of the cells seeded at the beginning of the experiment, changes in overall ATP content, and cell numbers by proxy, could be monitored over time. The results of the viability assay indicate that indeed the proliferative rate of the RKO cells is reduced, while no significant loss of viability occurs in the timeframe of the planned experiments (fig. S4).

Furthermore, proteomic profiling was performed in RKO cells with expression of two sgRNA targeting AAV1, which serve as controls for each experimental eIF-depletion condition, with either 48 h 200 ng/mL doxycycline induced KO or no KO induction. This experiment serves as additional validation of the induced expression of spCas9, which could be detected with

149 different precursors whereas no precursors could be identified in the sample without doxycycline treatment. The highly confident identification of spCas9 and missing peptide detection in the control sample indicate tightly controlled expression of the inducible system without indications of leaky spCas9 expression (fig. S5).

5.3.3 Analysing changes in protein synthesis, induced by eIF depletion

Changes in protein synthesis, induced by the depletion of the individual eIF, were primarily analysed via quantitative newly synthesized proteome analysis, using the QuaNPA workflow. Additionally, global proteomic profiling of the same samples was carried out. The global proteome analysis served multiple purposes in this study. In the first place, changes in total protein abundances provide important complementary information to NSP expression changes. Secondly, the depletion of the targeted eIF can be validated using protein abundance measurements. Lastly, the intensity ratios of unlabelled proteins, which were synthesized prior to the pulsed SILAC labelling, over labelled NSP can be used to determine the average rate of protein synthesis in the RKO cells following eIF depletion. Using the normalization approach described in section (5.2), normalization of the newly synthesized proteome data can be performed to accurately capture potential down-regulation of a majority of NSP (fig. 21).

In order to avoid potential batch effects and biases in the analysis of the newly synthesized and global proteome of the RKO cells, a block randomization approach was applied for the generation and measurement of the samples. The panel of 40 RKO cell lines were assigned into 4 randomized batches, which represent the cells which were induced, pulse-labelled and harvested on the same day. Since not all samples with 3 replicates fit onto a single 96 well plate, the samples for the newly synthesized proteome analysis and global proteomic analysis were split onto 2 separate plates. Due to the fact that the proteome samples do not contain an internal control, as is the case for the SILAC mixed NSP samples, pooled samples of the AAV1 KO samples from each experimental batch and replicate of the RKO cell line panel were included in the proteomic analysis. For the semi-automated NSP enrichment and automated SP3 sample preparation, of proteome samples, the positions of each sample on the 96 well plates were randomized to avoid additional biases. Measurements were subsequently

performed according in 4 separate blocks, with extensive quality control measurements between each batch, keeping the randomized order of the plate positions (table 1, fig. S6).

Table 1: Targeted eIF and batches for experiments and sample preparation via semi-automated NSP enrichment.

target	experiment batch	NSP enrichment batch	measurement batch
eIF1	1	1	2
eIF1B	2	1	2
eIF2A	3	2	4
eIF2B1	3	2	3
eIF2B2	3	2	4
eIF2B3	1	1	1
eIF2B4	4	2	4
eIF2B5	1	1	1
eIF2D	1	1	2
eIF2S1	1	1	2
eIF2S2	4	2	3
eIF2S3	2	1	2
eIF3A	3	2	4
eIF3B	1	1	1
eIF3C	2	1	1
eIF3D	2	1	1
eIF3E	2	1	2
eIF3F	4	2	3
eIF3G	3	2	4
eIF3H	4	2	3
eIF3I	4	2	4
eIF3J	1	1	1
eIF3K	4	2	3
eIF3L	2	1	2
eIF3M	1	1	2
eIF4A1	1	1	1
eIF4A2	3	2	4
eIF4B	2	1	2
eIF4E	3	2	3
eIF4E1B	2	1	2
eIF4E2	4	2	4
eIF4E3	3	2	3
eIF4EBP1	4	2	3
eIF4G1	4	2	4
eIF4G2	3	2	3
eIF4G3	3	2	3
eIF4H	2	1	1
eIF5A	2	1	1
eIF6	4	2	4
PDCD4	1	1	4

Upon completion of the sample preparation and measurement of the 120 newly synthesized proteome and 132 proteome samples, the quality of the produced data was evaluated in several aspects. The number of identified precursors, with exclusion of precursors which only differ in SILAC labels, was compared for the different eIF KO NSP and proteome samples. Comparable numbers of identifications were obtained for the different sample types, averaging approximately 50000 for NSP samples and approximately 60000 for proteome samples (fig. 28, panel A). These observed differences are most-likely caused by the loss of AHA-containing peptides, which are covalently bound to the MAA beads and not recovered following tryptic digestion. The number of quantified proteins were also comparable for the samples in each data type. For newly synthesized proteome samples, quantified proteins are defined via the detection of heavy- over intermediate SILAC ratios, enabling relative quantification of the eIF KO and AAV1 KO samples. In proteome samples, the number of proteins with label-free quantification normalized (LFQ) intensity is indicated. More proteins were quantified in the NSP samples (on average around 5700) than in the proteome samples (on average around 5300), despite the fact that fewer precursors were identified on average (fig. 28, panel B). This observed difference in the number of identified proteins is consistent with previous reports [142]. Furthermore, the intensity ratios of heavy- and intermediate SILAC labelled precursors, which originate from NSP, and unlabelled (light) precursors, which originate from preexisting proteins, were calculated to asses the efficiency of the click chemistry-based NSP enrichment. In both batches, positive log2 median intensity ratios above 2, indicating more than 75 % signal from NSP derived precursors, were measured on average for all eIF depletion samples (fig. 28, panel C). The median intensity ratios of samples in the second batch are lower than in the average samples of the first batch. However, the range of intensity ratios is very broad in the first batch, indicating higher variability of NSP coupling efficiency (fig. 28, panel D).

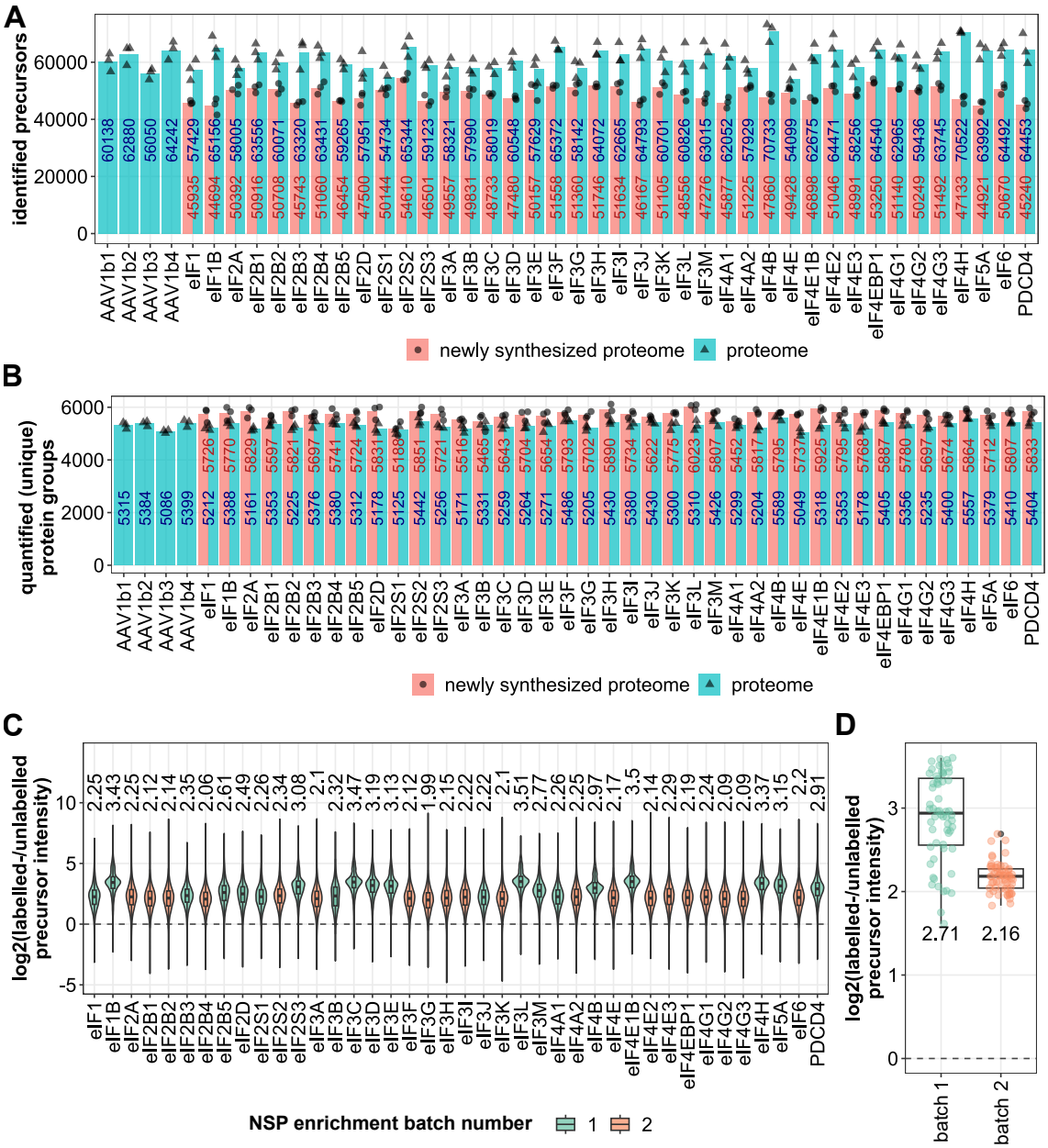


Figure 28: Evaluation of basic sample and data quality of the newly synthesized proteome and global proteome samples, of the RKO cell line panel. A) Number of identified precursors in the different induced eIF- and control KO samples. B) Number of quantified proteins groups. For newly synthesized proteome samples, quantified proteins are defined via the detection of heavy- over intermediate SILAC ratios, enabling relative quantification of the eIF KO and AAV1 KO samples. The different experimental batches of pooled AAV1 control proteome samples are indicated via the added "b(1-4)" label. In proteome samples, the number of proteins with label-free quantification normalized (LFQ) intensity is indicated. Only protein groups with unambiguous assignment to a single protein entry of the SwissProt database (unique proteins) were included to enable direct integration of transcriptomic data in later stages of the analysis. The type of data or more specifically the method by which it was obtained is indicated in separate colors. C) Intensity ratios of heavy- and intermediate SILAC labeled precursors over unlabeled (light) precursors. Data from all 3 experimental replicates in included and the median ratio values are indicated on top of the distributions. D) Median SILAC precursor intensity ratios for NSP samples of both batches. The 2 different semi-automated NSP enrichment batches are indicated in separate colors.

Even with the observed degree of variability in the first batch and slightly lower median observed in the second batch, the precursor SILAC intensity ratios confirm that the primarily NSP were measured and quantified. The quantitative precision, determined via the coefficient of variation (CV) was shown to be high for the proteome (median < 10 %) and newly synthesized proteome data (median < 9 %) across the entire panel (fig. S7).
Additionally, to identify issues in the HPLC-based separation of analysed peptides, the retention times of spCas9 peptides, which were identified in the AAV1 KO control samples (intermediate SILAC label) in the complete set of newly synthesized proteome samples, or complete set of proteome samples, were compared. In the case of the newly synthesized proteome data, no significant retention time deviation for one of the 12 spCas9 peptides was observed in any sample (fig. S8, panel A). For the global proteome samples the larger set of 60 commonly identified spCas9 peptides, indicated severe retention time shifts in the third replicate of the induced eIF4E1B KO sample, which was therefore excluded from subsequent analysis (fig. S8, panel B).

Significant depletion of the targeted eIF protein levels could be measured almost all proteome samples with the exception of eIF2B3 and eIF4A2, whose samples were therefore excluded from subsequent analysis. The fold changes in the levels of the targeted eIF proteins do not correlate with the measured protein abundance, which was calculated using the intensity-based absolute quantification metric (iBAQ) [120]. For the majority of the targeted eIF, protein levels are depleted approximately 2 fold, indicating a 50 % reduction (fig. 29, panel A). Additionally, the fold change values can also be inspected for all identified peptide pairs, to ensure full depletion of the targeted eIFs and exclude truncation of the protein sequence caused via in frame deletions of the 2 sgRNA (fig. S9). For the majority of targeted eIF, significant depletion could only be measured for respective target and not for interaction partners and sub-complex members. In contrast to eIF protein sub-units of the eIF2- and eIF4F complex, a more complex pattern can be seen forthemembers of the eIF3 sub-complex. Depletion of several individual eIF proteins, in turn lead to the depletion of several other eIF3 sub-units, but surprisingly not all members of the sub-complex (fig. S10). Interestingly, the target eIF depletion fold change values do not appear to correlate with changes in cell proliferation (fig. 29, panel B). The similar degree of eIF protein depletion

and diverse effects in cell proliferation indicate that the observed phenotypes are not simply caused by varying degrees of eIF depletion, but due to the specific functions of the individual eIF proteins.

By applying the previously described modified median normalization approach for newly synthesized proteome data (section 5.2), changes in the average protein synthesis rate can be quantified in eIF- and AAV1 KO cells. RKO cells with induced AAV1 KO, serve as reference to which the induced eIF KO data is contrasted to fig. S11. Reduced levels of NSP and thus protein synthesis were measured in a majority of RKO cells lines with induced KO of different eIF. However, minor increases in protein synthesis were also observed in cells with depletions of eIF2D, eIF3J, eIF4E1B, eIF4E2, eIF4EBP1, eIF4H and PDCD4. The most severe reduction in protein synthesis was observed in response to components of the core eIF3 sub-complex such as eIF3A, eIF3B, eIF3C and eIF3E (fig. 29, panel C). Similar patterns of global translation inhibition, induced via knockdown of individual eIF3 sub-units, has been reported before using HeLa cells [188]. Insignificant negative correlation of the measured rate of protein synthesis and the fold change of target eIF protein depletion was observed (Pearson correlation coefficient R=-0.21), further implying that observed cellular responses are due to the specific function of the individual eIF proteins and not simply caused by different degrees of eIF depletion (fig. 29, panel D). Moreover, a moderate correlation between the rate of protein synthesis and the observed changes in cell proliferation was detected (R=0.39). The moderate correlation suggests that the changes in cell proliferation are influenced by changes in the rate of protein synthesis, caused by the induced depletion of eIF proteins. However, two prominent outliers can be observed. Depletion of eIF4B and eIF4G3 lead to strongly reduced cell proliferation, but do not decrease the rate of protein synthesis more strongly than depletion of other eIF proteins. These findings suggest that eIF4B and eIF4G3 play a specific role in the regulation of proliferation in RKO cells (fig. 29, panel E).

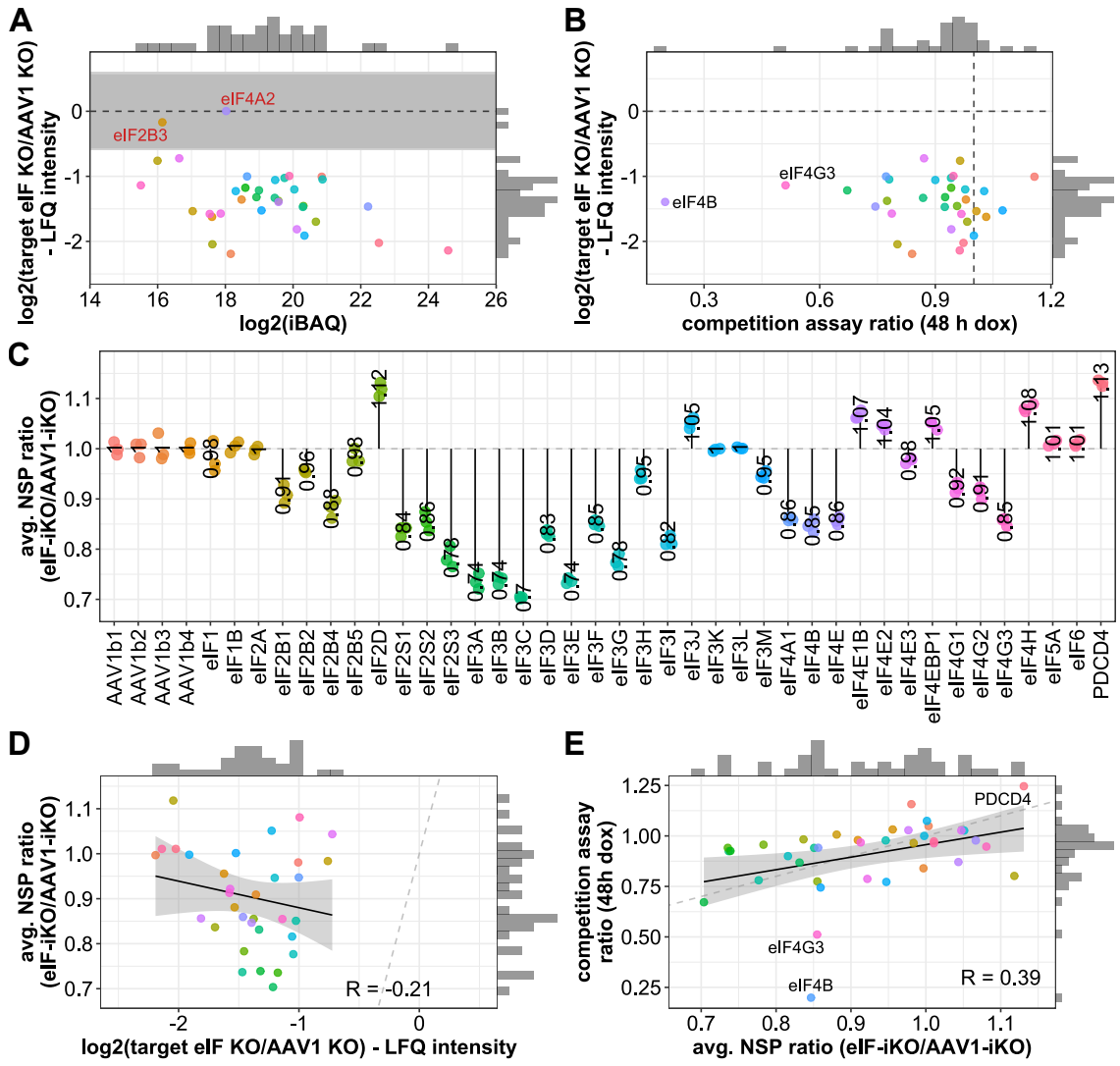


Figure 29: Initial analysis of the newly syntheisized and proteomic data, generated from the induced depletion of different eIF proteins in RKO cell lines. A) Scatter plot of the measured protein LFQ intensity fold change, and iBAQ protein abundance values for the targeted eIF in RKO cells with the respective induced KO. B) Scatter plot of measured protein LFQ intensity fold change values and cell count ratios for the targeted eIF in RKO cells with the respective induced KO. The FACS-based competition assay was performed by Robert Kalis in the Institute of Molecular Pathology Vienna. C) Average intensity ratio of SILAC labeled newly synthesized proteins (NSP) over pre-existing unlabeled proteins, in RKO cells with induced KO of individual eIF compared to induced AAV1 KO controls. The different experimental batches of pooled AAV1 control proteome samples are indicated via the added "b(1-4)" label. D) Scatter plot indicating the correlation of measured protein LFQ intensity fold change and the determined rate of protein synthesis, measured via the average intensity ratio of NSP in AAV1 KO control cells and eIF KO cells. The pearson correlation coefficient (R) is indicated in the graph. E) Scatter plot indicating the correlation of measured change in cell proliferation, measured via a FACS-based competition assay, and the determined rate of protein synthesis, measured via the average intensity ratio of NSP in AAV1 KO control cells and eIF KO cells. The pearson correlation coefficient (R) is indicated in the graph.

The initial analysis of the proteomic data provides clues towards the highly diverse effects

of eIF depletion and their potentially different roles in the regulation of specific mRNA translation initiation.

Principal component analysis (PCA) of the newly synthesized proteome data also showed distinct grouping of individual eIF KO samples with tight association of the experimental replicates, indicating robust quantification and distinct changes in protein expression in response to different eIF depletions (fig. 30). Principal component (PC) 1 aligns with the measured degree of changed protein synthesis, that was observed upon the depletion of targeted eIFs and accounts for highest percentage of observed variance in the data (52.6 %) (fig. 29, panel C). The most extreme values on the PC1 dimension are samples with induced depletion of eIF3C and PDCD4. Interestingly, the eIF4G3 and eIF4B data appear as outliers on the PC3 dimension. Both eIF4B and eIF4G3 stood out in the proliferation assay (fig. 26), which suggests that specific changes in NSP expression could be quantified for both (fig. 30).

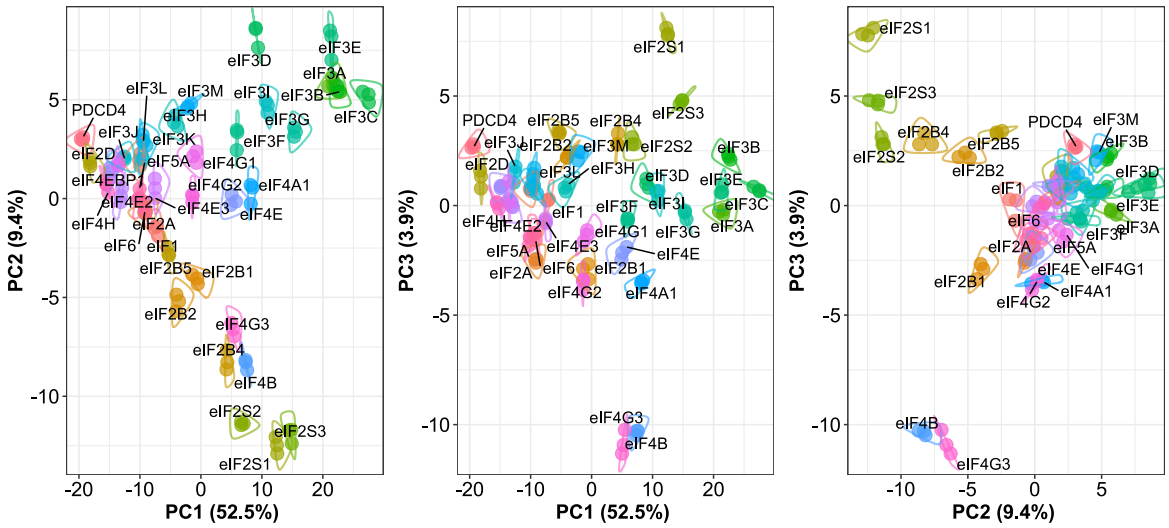


Figure 30: Principal component analysis (PCA) of the newly synthesized proteome data for RKO cell lines with induced depletion of different eIF proteins. The percentage values of explained variance by the respective principal component (PC) are indicated on each axis.

Additionally, PCA was used to asses batch effects in the newly synthesized proteome and proteomic data. For the newly synthesized proteome, no clustering of the distinct samples of the 4 experimental batches or 2 NSP enrichment batches, could be observed (fig. S12, panel A and B). Distinct clustering of the sample preparation batches, or more specifically the samples which were generated 2 separate days (and 96-well plates) via autoSP3, was

detected in the proteomics data. Since each experimental batch contains induced AAV1 KO control samples, no batch correction measures were carried out (fig. S12, panel C and D).

A more detailed look into the proteomic data reveals the drastic differences in the effects of individual eIF depletions. Similarly to the change in protein synthesis, following the induced depletion of different eIF, the effects on the proteome and newly synthesized proteome are highly diverse (fig. 29, panel C). This can be visualized by through comparison of the number of differentially expressed proteins (log2 fold change $> |0.585|$ and adjusted p-value < 0.05) in the proteome and newly synthesized proteome analysis of each induced eIF KO. Interestingly, even among eIF, which belong to the same subcomplex, strong differences can be observed. Importantly, the number of differentially expressed proteins in the proteome anylsis is relatively low in the majority of eIF depletion conditions, which indicates that no extensive proteome remodeling has occurred at the 48 h time point (fig. 31).

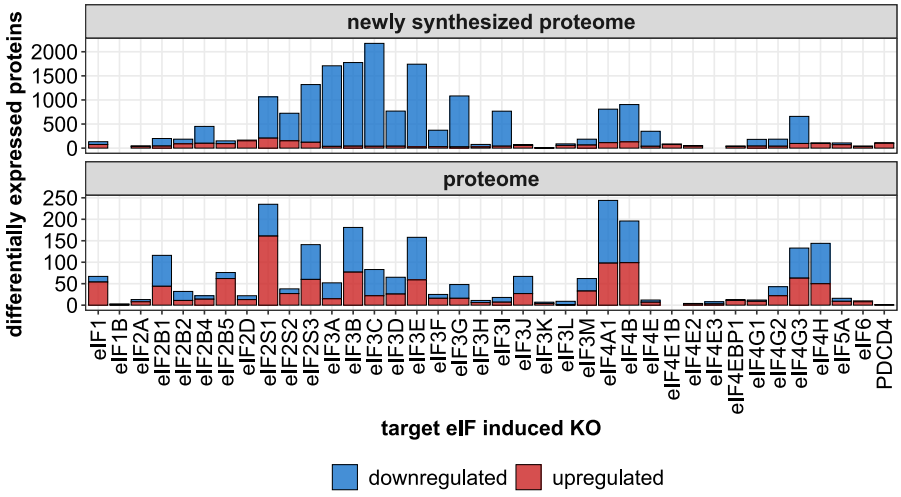


Figure 31: Comparison of the number of differentially expressed proteins (log2 fold change $> |0.585|$ & adj. p-value < 0.05) in the proteome and newly synthesized proteome analysis of RKO cells with different induced eIF KO. The type of differential expression is indicated with different colors. Blue represents down-regulated proteins and red up-regulated proteins.

A strong contrast can for example be seen between eIF3A and eIF3J, which are adjacent members of the eIF3 subcomplex (fig. 5, panel B) [189]. The induced KO of eIF3A and eIF3J results in comparable depletion of the targeted eIF proteins and similar numbers of differentially abundant proteins, as measured in the full proteome analysis fig. 31. However, in contrast to eIF3J depletion which has no major impact on the newly synthesized proteome of RKO cells, depletion of eIF3A leads to the down-regulation of a majority of NSP (fig. 32).

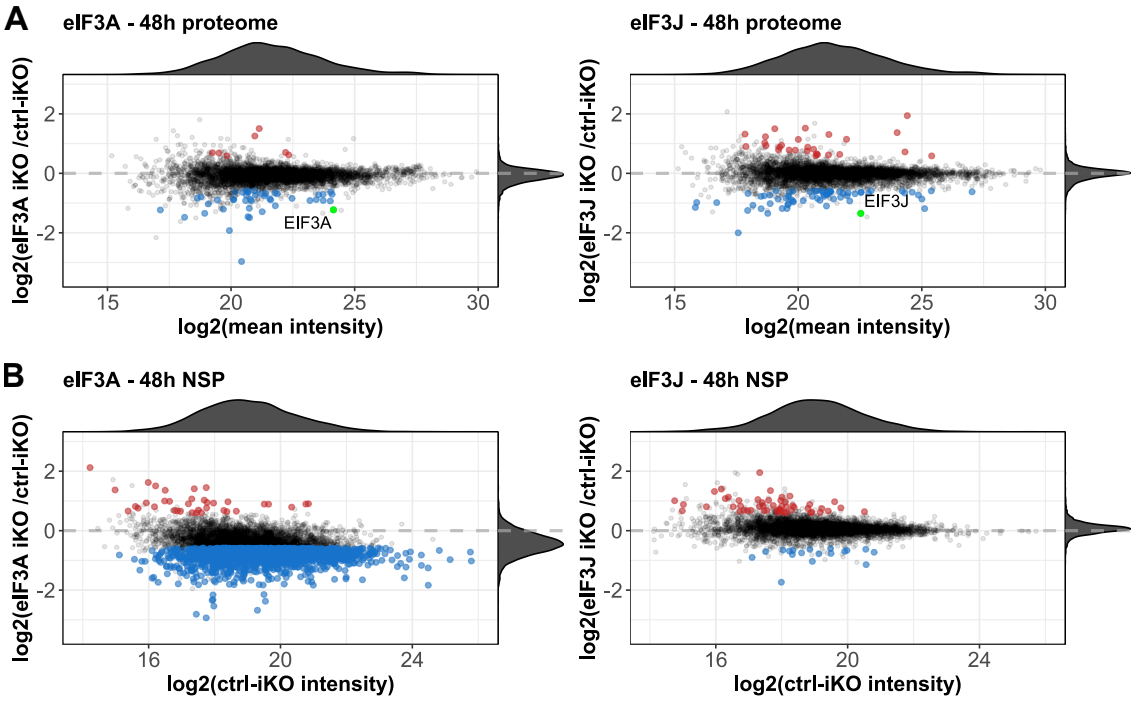


Figure 32: Comparison of changes in newly synthesized and overall protein abundance, induced by the depletion of eIF3A and eIF3J. A) MA-plots of the proteome data, indicating the log2 fold change values of the respective eIF KO over AAV1 KO controls on the y-axis and mean LFQ intensity values on the x-axis. Proteins with significant up-regulation (log2 fold change $>$ 0.585 & adj. p-value $<$ 0.05) in response to 48 h KO induction are indicated in red and down-regulated proteins (log2 fold change $<$ -0.585 & adj. p-value $<$ 0.05) in blue. B) MA-plots of the newly synthesized proteome data, indicating the log2 fold change values of the respective eIF KO over AAV1 KO controls on the y-axis and mean LFQ intensity values of AAV1 controls on the x-axis. NSP with significantly up-regulation (log2 fold change $>$ 0.585 & adj. p-value $<$ 0.05) in response to 48 h KO induction are indicated in red, and down-regulated proteins (log2 fold change $<$ -0.585 & adj. p-value $<$ 0.05) in blue.

However, to better understand the observed differences as to how protein synthesis is impacted in the RKO cell lines following individual eIF depletion, in depth analysis on the level of individual proteins and their respective mRNA is required. Through correlation of mRNA characteristics with protein expression, potential clues towards selective regulatory mechanisms of translation via the individual eIF could be obtained.

5.3.4 Investigating potential mechanisms of selective mRNA translation, regulated by individual eIF

A range of analysis tools and approaches were applied in the analysis of the newly synthesized proteome data, to identify proteins whose synthesis is specifically affected by depletion of individual eIF. The goal of the analysis was to obtain clues towards potential regulatory mechanisms driving selective or preferential mRNA translation via eIF proteins.

To exclude the effects of increased transcription or mRNA degradation, that could contribute to the measured changes in NSP abundance, mRNA abundance was measured using RNA sequencing (QuantSeq) [155]. The obtained data were integrated with the newly synthesized proteome data for subsequent analysis. More specifically, fold changes of the NSP were divided by measured mRNA fold changes, to calculate “translation efficiency“ (TE) as quantitative metric for the comparisons. Translation efficiencies can be compared for individual proteins across the panel of RKO cells with different eIF depletion, pointing towards regulatory roles of different eIF in the translation of specific mRNA subsets.

For the majority of proteins with differential expression, no significant changes in mRNA abundance could be detected, indicating that changes in translation and not transcription or mRNA degradation, are contributing to the measured changes in NSP levels, in response to eIF depletion (fig. 33, panel A). Similarly to the analysis of NSP, reduced TE values can for example be observed for the majority of proteins in response to eIF3A depletion (fig. 33, panel B). Due to relatively moderate effects on mRNA abundance levels in response to eIF3A depletion, the majority of differentially expressed NSP match proteins with differential synthesis. RNA with differential expression can be excluded for subsequent the correlation analysis, even in cases where the calculated TE are still significant (fig. 33, panel C).

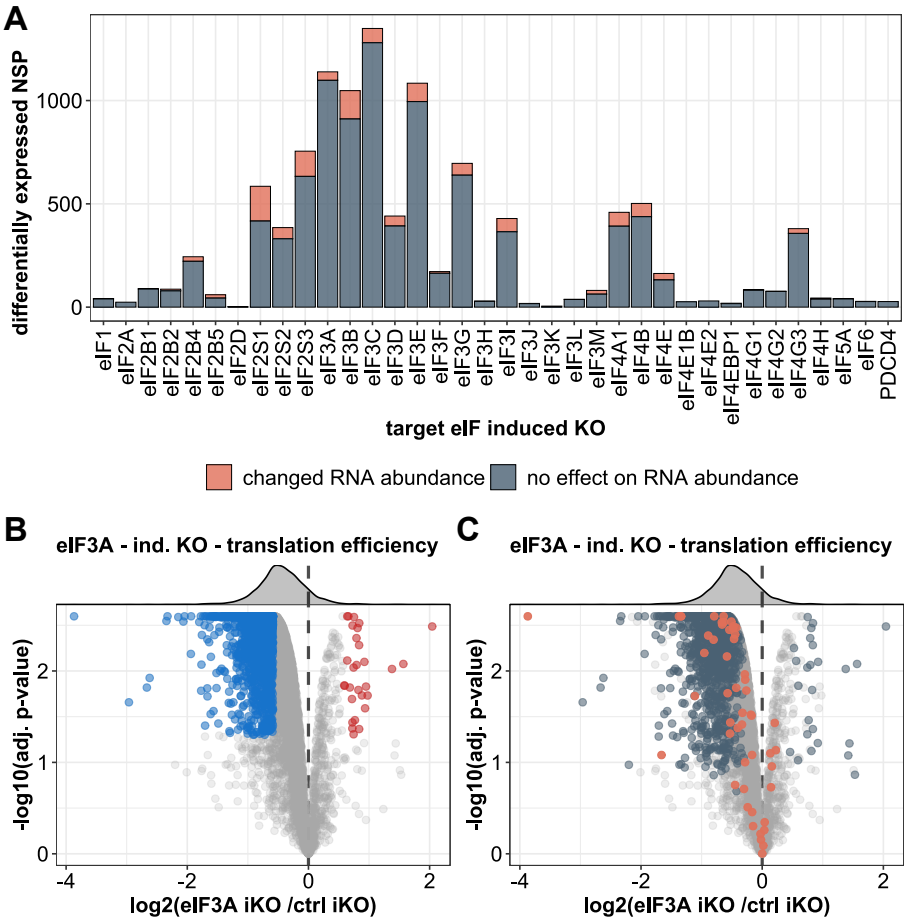


Figure 33: Integration of newly synthesized proteome and RNA sequencing data for the quantitative analysis of mRNA translation. A) Comparison of the number of NSP with differential expression ($|\log2 \text{ fold change}| > 0.585$ & adj. p-value < 0.05 & CV < 20 %). NSP for which no significant changes in mRNA levels were observed are colored in dark gray ($|\log2 \text{ fold change}| < 0.322$ & adj. p-value > 0.05) and NSP with significant changes in mRNA levels are colored in orange ($|\log2 \text{ fold change}| > 0.585$ & adj. p-value < 0.05). B) Volcano plot of TE values from RKO cells with induced KO of eIF3A. Proteins with up-regulated translation (log2 fold change > 0.585 & adj. p-value < 0.05) are indicated in red and down-regulated proteins are indicted in blue (log2 fold change $<$ -0.585 & adj. p-value < 0.05). C) Volcano plot of TE values from RKO cells with induced KO of eIF3A. Subsets of NSP with differential expression, without changed mRNA levels, are indicated in dark gray. NSP with significantly changed mRNA levels are indicated in orange.

Through the normalization of NSP with significantly changed mRNA abundance, the majority did not exhibit significantly changed TE. However, several interesting observations can still be drawn from this subset of proteins. For example, amino acid transporter proteins, SLC1A4, SLC1A4, SLC3A2 and SLC7A5, as well as aminoacyl-tRNA synthetase AARS1, GARS1, SARS1, WARS1 and leucine-sensing protein SESN2 [190] were up-regulated in response to depletion of eIF2 proteins (eIF2S1-3) and eIF2-GEF (eIF2B1,-2,-4,-5). Moreover, increased mRNA and NSP levels of cyclin-dependent kinase inhibitor protein CDKN1A were detected

in cells with depletion of eIF4B and eIF4G3 (fig. S13). Since CDKN1A is a crucial regulator of cell cycle progression and proliferation [191], an increase in its expression could contribute to the strongly reduced proliferation and changed cell morphology in response to induced eIF4B and eIF4G3 depletion (fig. 26 and fig. 27). Interestingly, few significant changes in mRNA abundance were detected for proteins that are associated in the integrated stress response signaling (fig. S14). These results suggest, that the depletion of eIF2 sub-complex proteins and their interaction partners did not trigger strong cellular stress in the RKO cell line panel after 48 h dox treatment.

Several strategies and methods were applied and tested for the analysis of the TE data. Firstly, to detect potential indications of specific protein classes or biological processes, conventional functional enrichment analysis (via GSEA) was performed. However, no significant enrichment of annotated protein sets, whose synthesis was strongly affected by specific eIF depletion were obtained. Instead, the data was analyzed using PCA.
Tight grouping of samples could be observed, indicating strong consistency between experimental replicates and distinct changes in protein synthesis in response to depletion of different eIF proteins.

Heterogenous changes in protein synthesis are also reflected through low- to moderate correlation between the measured TE values (fig. S15). Moreover, overlap between differentially translated mRNA is relatively low, even for neighboring protein sub-units, as can be highlighted for eIF3A-E, indicating that their depletion induces specific changes in translation (fig. S16).

Due to the few significant changes in mRNA levels, the PCA of the TE data produced very similar results to the previous PCA analysis of the newly synthesized proteome data. However, the previous outliers eIF4B and eIF4G3 did not stand out when comparing TE values (fig. 30 and fig. 34, panel A). It is possible that for the majority of NSP that contributed to the principal component 3 in the previous PCA, significant changes in mRNA levels were detected, as was shown for CDKN1A (fig. S13). The 250 proteins with the largest contributions to the variance, which is accounted by the principal components PC1-3, were extracted and subjected to functional enrichment analysis (via over-representation analysis).

Proteins associated with binding to the 3′-UTR of mRNA and to unfolded proteins, and focal adhesion proteins were found to be over-represented in PC1 and PC2 with low significance. Interestingly, significant over-representation of ribosomal proteins and translation-associated proteins was detected in the subset of proteins with the highest variance contribution to PC3 (fig. 34, panel B). However, a closer inspection of the ribosomal protein subset revealed that their average log2 fold change values do not strongly deviate from the median fold change of all quantified proteins in the eIF depletion samples (fig. 34, panel C). Synthesis of several ribosomal proteins, however, strongly deviates from the median of all quantified proteins and exhibits distinct patterns for the individual eIF depletion. Ribosomal proteins RPS20 and RPS27A for example are down-regulated in response to most eIF depletion, whereas synthesis of RPL10, RPL36, RPS15A and RPS19 is increased in response to depletion of specific eIF (fig. 34, panel D).

These results indicate that the tested functional enrichment analysis strategies are unlikely to yield biologically meaningful insights into the data and that other selection criteria for differentially synthesized proteins and their characterization should be applied.

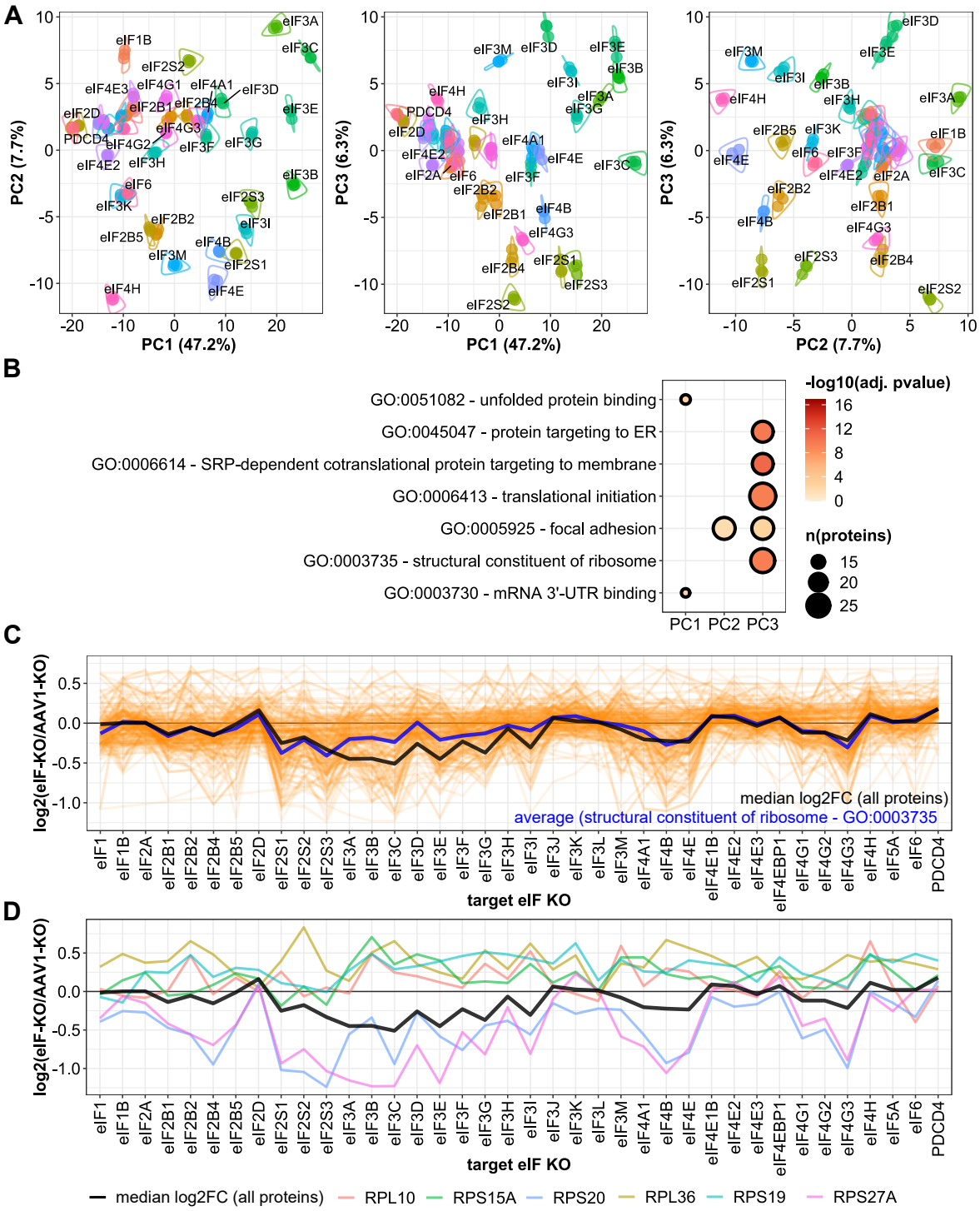


Figure 34: Principal component analysis (PCA) of the translation efficiency (TE) data for RKO cell lines with induced depletion of different eIF proteins. A) PCA plot of the TE data. The percentage values of explained variance by the respective principal component (PC) are indicated on each axis. B) Over-representation analysis of the subsets of 250 proteins with the largest contributions to the principal components PC1-3. The number of proteins in the different gene sets are indicated through point size and statistical significance (adjusted p-value) is indicated via coloration. C) Log2 (TE) fold change of ribosomal proteins, in the RKO cells with induced KO of indicated eIF. Individual proteins of the GO-term “structural constituent of ribosome” (GO:0003735) are displayed as orange lines and their average log2 (TE) fold change is indicated through a blue line. Median log2 (TE) fold change of all quantified proteins in the respective conditions are indicated in black. D) Log2 (TE) fold change of selected ribosomal proteins, in RKO cells with induced KO of the indicated eIF. Median log2 (TE) fold change of all quantified proteins in the respective conditions are indicated in black.

There are no indications in the acquired data and in published reports, that eIF preferentially translate mRNA that encode proteins with specific functions. Therefore, measured changes in protein synthesis (TE metric) were correlated with several characteristics and properties of mRNA templates, instead of focusing on sets of annotated protein classes and proteins involved in specific biological processes. Since untranslated regions (UTR) of mRNA play a crucial role in the regulation of their stability localization and translation [192], basic characteristics of 5′-UTR and 3′-UTR sequences, such as nucleotide composition length etc., were included in the analysis. Moreover, characteristics of the CDS and reported features involved in non-canonical translation initiation [73] were correlated with the measured changes in protein synthesis. For this purpose, the top 5 % most “up-” and “down-regulated” proteins were selected according to their TE log2 fold change and selected mRNA metrics were compared between the protein subsets and the average of all quantified proteins. Observed difference in the properties of mRNA that encode the differentially synthesized proteins, provide initial clues towards potential regulatory mechanisms by which individual eIF proteins could modulate the translation of specific mRNA. However, the high complexity of the data and multitude of regulatory mechanisms of mRNA translation pose a major challenge for the analysis and interpretation.

Since the proteomic data is matched to canonical protein sequences and cannot distinguish between protein isoforms, canonical mRNA transcripts were used for the comparative analysis of basic mRNA metrics. Guanine and cytosine percentages (GC %) in the UTR sequences were calculated to serve as a proxy for the existence of secondary structures, which are more stable with high GC content [193]. The percentage values of pyrimidine bases (uracil and cytosine) in the 5′-UTR was used to include potential 5′-terminal oligo-pyrimidine (TOP) sequences in mRNA, which were shown to function as regulatory sequences in specific protein sets, such as ribosomes [194]. Additionally, the length of the coding sequence (CDS) was compared.

The initial comparative analysis was focused on basic mRNA characteristics, such as 5′-UTR length. Differences in the median length of the 5′-UTR were observed in the top 5 % differentially synthesized protein subsets of several eIF. More specifically, 5′-UTR sequences

of mRNA with strongly down-regulated translation were found to be longer than the median of mRNA with the highest TE. These differences in 5′-UTR length are especially pronounced in RKO cells with depletion of eIF3A, eIF3D, eIF3E, eIF3I and eIF4G2 (fig. 35).

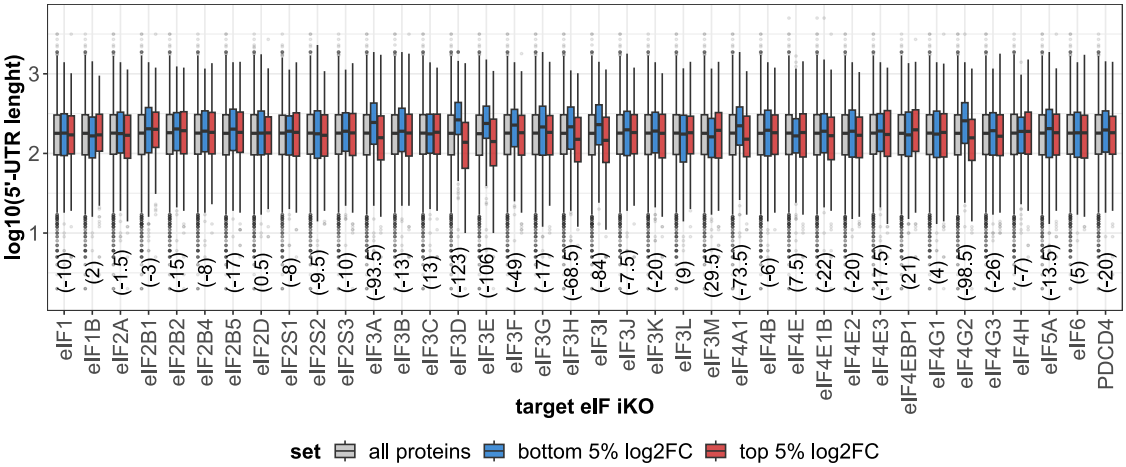


Figure 35: Comparative analysis of 5′-UTR lengths of protein subsets with differential synthesis in response to eIF depletion. Box plots of the log10-transformed 5′-UTR lengths of protein subsets in RKO cells with induced KO of indicated eIF and eIF-associated proteins. The top 5 % of proteins with the largest (TE) log2 fold change are indicated in red and the 5 % most down-regulated proteins are indicated in blue. The median difference in 5′-UTR length (in nucleotides) between the top 5 % protein subsets is indicated as text below the box plots.

Comparisons of additional mRNA characteristics were carried out with the top 5 % differentially synthesized protein subsets, and the percentage deviations of the median values for each set were calculated. No major changes in nucleotide composition as GC- or pyrimidine content could be detected in any of the in the 5′- and 3′-UTR of the subsets (fig. 36, panel A, -C and E). However, strong differences in the median lengths of the 5′- and 3′-UTR, as well as the CDS could be observed between the subsets for several, but not all eIF (fig. 36, panel B, -D and F). As indicated before, these findings indicate that synthesis of proteins with longer 5′-UTR sequences is more strongly decreased than for those with shorter 5′-UTR, upon depletion of eIF3A, eIF3D, eIF3E, eIF3I and eIF4G2 (fig. 35). Interestingly, length of the 3′-UTR was similarly correlated in the cells with depletion of the previously listed eIF (fig. 36, panel B and -D). Moreover, proteins with strongly decreased synthesis contained on average shorter CDS for the majority of depleted eIF. These differences in CDS length were especially pronounced for induced KO of eIF2B4, eIF2B5, eIF2S1, eIF2S3, eIF3B and eIF3C (fig. 36, panel F).

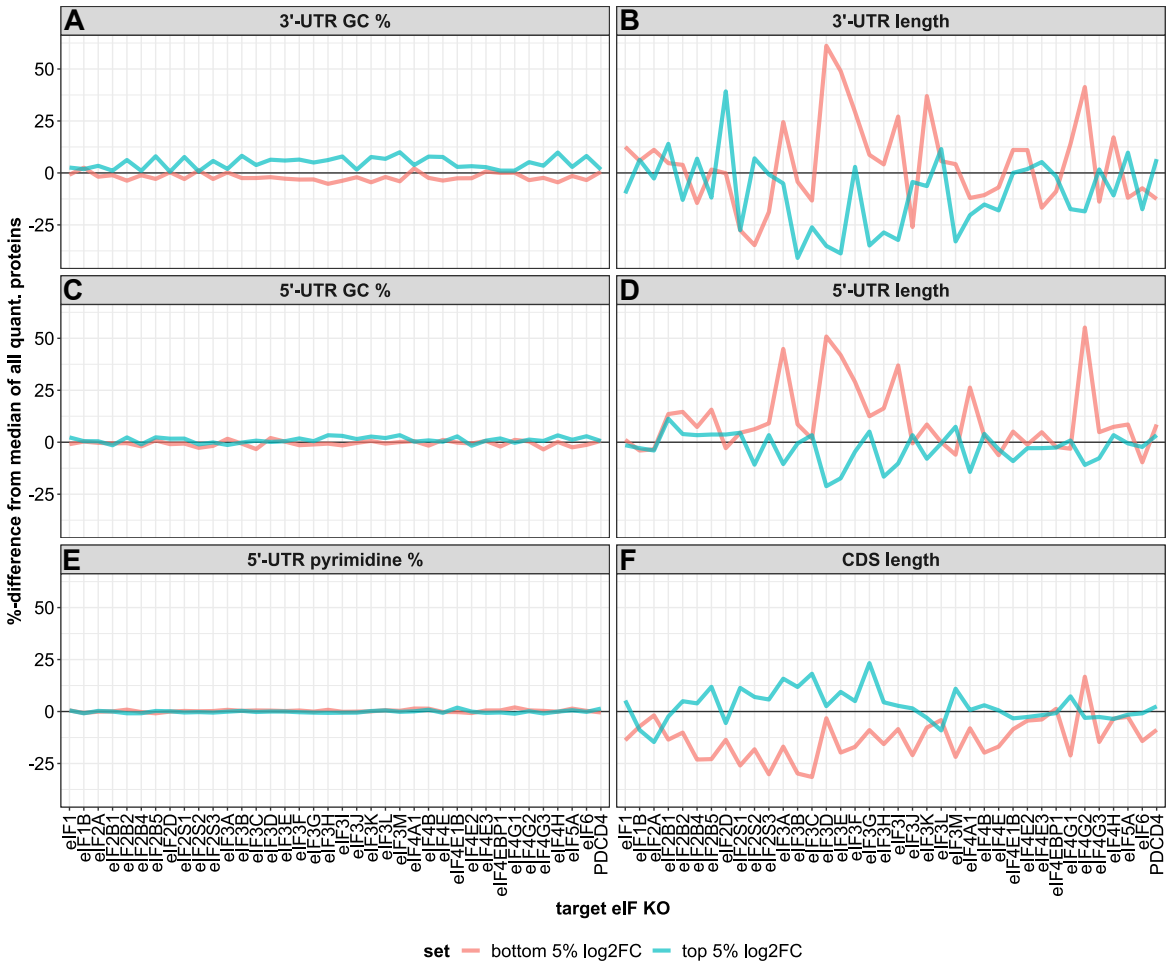


Figure 36: Comparative analysis of changes in translation efficiency, in response to depletion eIF, and basic mRNA characteristics. The average deviation from the median values of all quantified proteins (in %) for the indicated mRNA metrics in the top- and bottom 5 % differentialy synthesized proteins are plotted for each eIF depletion condition. The median deviations of the top 5% down-regulated proteins are indicated in pink and the subset of proteins with top 5 % log2 TE fold change values is indicated in light blue.

In addition to the previously mentioned mRNA characteristics, information with regard to the presence of uORF elements [175] and m6A modifications [176] in the 5′-UTR of mRNA was retrieved and included in the analysis. No clear differences could be observed regarding the numbers of reported uORFs, the predicted secondary structure scores and number of reported m6A modifications in the 5′-UTR of the compared protein subsets (fig. 37, panel A, -D and -E). However, the length and position of reported uORF elements differs widely between the different eIF depletion samples, creating a seemingly chaotic pattern. Proteins, encoded by mRNA with short uORF elements and uORFs in close proximity to the canonical translation start site (TSS) are over-represented among the subset of proteins with the largest TE log2 fold changes, in response to depletion of eIF2B2, eIF2B5, eIF2S1, eIF2S3,

eIF3B, eIF3I, eIF3K, eIF3M, eIF4B and eIF4H. Similarly strong differences can also be observed in the bottom 5 % protein subset in response to eIF3I, eIF3M and eIF4E depletion (fig. 37, panel B and -C).

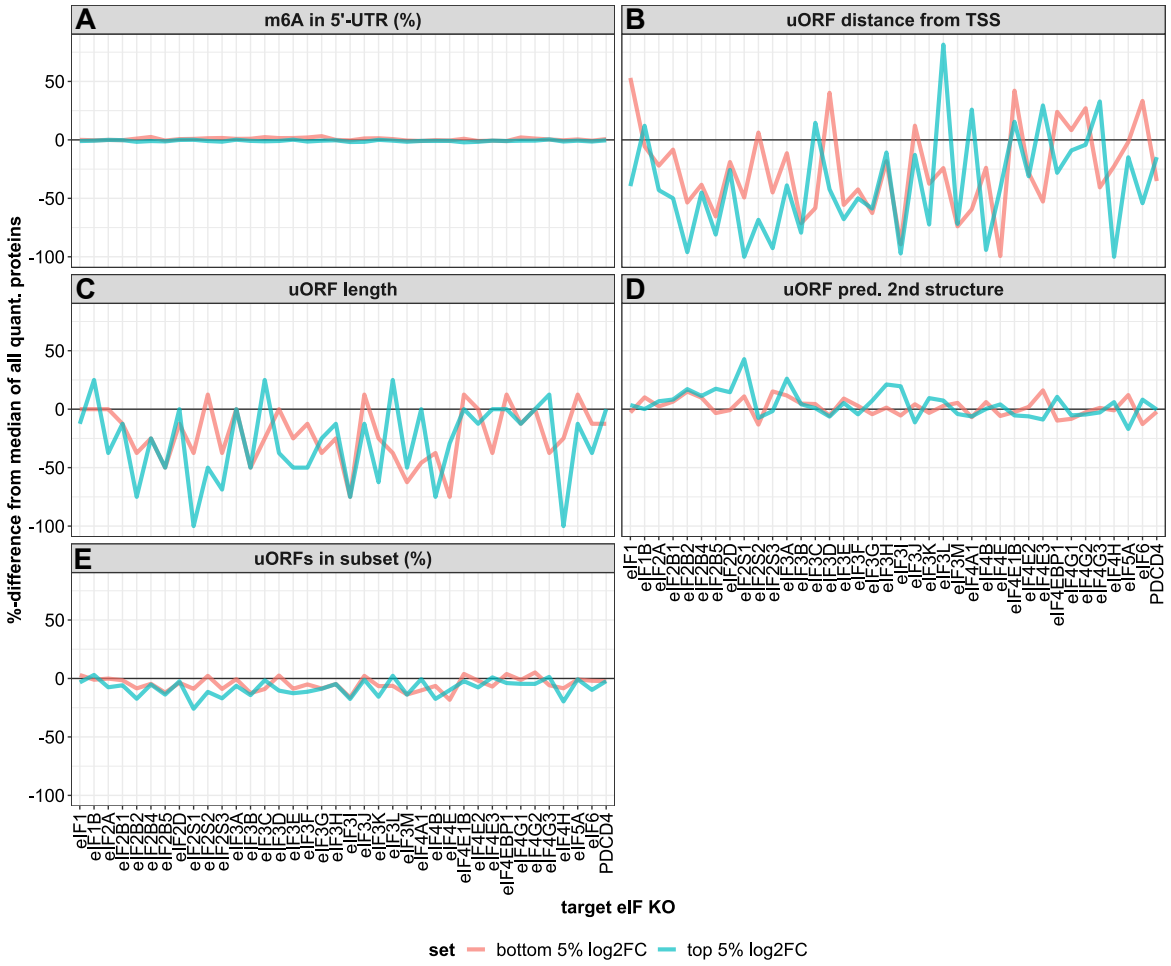


Figure 37: Comparative analysis of changes in translation efficiency in response to eIF depletion, correlated to presence of uORF elements and m6A-modifications in 5′-UTRs. Average deviation from the median values of all quantified proteins (in %) for the indicated mRNA metrics in the top- and bottom 5 % differentialy synthesized proteins are plotted for each eIF depletion condition. The median deviations of the top 5% down-regulated proteins are indicated in pink and the subset of proteins with top 5 % log2 TE fold change values is indicated in light blue.

The comparative analysis of mRNA characteristics provides initial clues towards potential points of selective mRNA translation regulation of eIF. Importantly, patterns could be observed for individual eIF that are distinct from neighboring sub-complex members. However, at its current stage, the results only imply that the highlighted eIF might be involved in the regulation of mRNA translation initiation in an UTR- or uORF-dependent manner. The precise mechanism will need to be investigated more thoroughly in future analyses.

In addition to the correlative analysis of mRNA characteristics and measured protein synthesis, information on shared expression patterns and dependencies on individual eIF can be extracted from the data. For example, similar changes in mRNA translation were measured for EIF4G3, nuclear proteins LMNB1/2 and cell cycle regulatory protein ANP32B [195] (fig. 38).

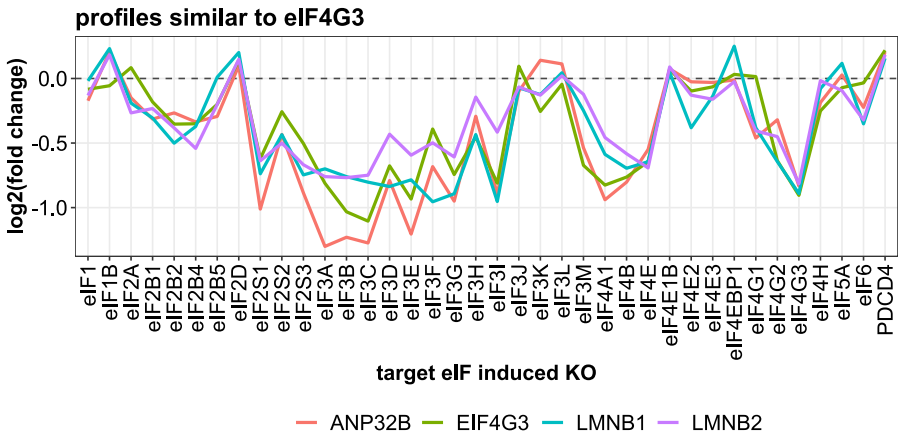


Figure 38: Identification of proteins, with measured changes in translation efficiency, that are similar to those of eIF4G3. Median normalized log2-transformed TE values of the proteins EIF4G3, ANP32B, LMNB1 and LMNB2 are displayed for all 38 eIF depletion conditions.

This finding provides indication of potential translation co-regulation of specific proteins, and could also be explored further in future analysis.

In conclusion, through the integration of data obtained from newly synthesized proteome analysis and RNA sequencing, heterogenous changes in protein synthesis caused by depletion of individual eIF and eIF-associated proteins could be quantified. Correlation of the measured changes in protein synthesis with mRNA characteristics revealed several trends that could point towards additional regulatory functions of individual eIF.

6 Discussion

6.1 Newly synthesized proteome analysis as tool for the study of mRNA translation regulation

6.1.1 Developing an improved proteomic workflow for analysis of NSP

Quantitative newly synthesized proteome analysis is a powerful methodology, which can be used for the measurement of changes in protein synthesis in the context of initial cellular proteome remodeling following various perturbations, or for the investigation of regulatory mechanisms of mRNA translation. However, due to labor-intensive multi-step sample preparation protocols and analytical challenges it remains a rarely applied method. Overcoming these limitations could potentially enable more frequent application of newly synthesized proteome analysis and benefit biological- and medical research.

In this dissertation, the development of a workflow for the semi-automated enrichment and analysis of AHA-and pSILAC labeled NSP is described, which attempts to overcome the limitations of established protocols. More specifically, magnetic alkyne agarose beads were generated and used in a semi-automated protocol on an Agilent Bravo robotic liquid handling platform, which enables the processing of up to 96 samples in parallel. The optimized protocol enables the enrichment of AHA-containing NSP with lower input amounts, simplifying the prior metabolic labeling steps and cell culture work in general. Moreover, through the use of the plexDIA features [69] of the DIA-NN software [52], increased proteome coverage and robust quantification can be achieved in single LC-MS measurements of SILAC-labeled NSP samples, further adding to throughput. Taken together these elements were integrated to create a workflow referred to as “quantitative analysis of the newly synthesized proteome” (QuaNPA) [180].

Initial experiments and comparative analysis of various strategies and magnetic bead types for the enrichment of AHA-contaning NSP, clearly indicated that magnetic beads with high binding capacity are required to create newly synthesized proteome samples of high quality. In addition to their lower binding capacity, compared to agarose beads, coupling with biotin-alkyne and subsequent pull-downs with Streptavidin conjugated beads

require dialysis or precipitation and re-solubilization steps to remove excess biotin-alkyne, adding an additional hurdle to automated sample preparation and increasing throughput [139][136][147]. Due to their high alkyne group density, magnetic alkyne agarose (MAA) beads were shown to be well suited for producing high quality newly synthesized proteome samples (fig. 7). Importantly, using magnetic beads, small scaled (PCR tube scale $\leq$200 μL) NSP enrichments are possible, without major decreases in proteome coverage (fig. 8).

The intensity ratio of SILAC labeled NSP over unlabeled pre-existing proteins, proved to be a highly effective metric for assessing the efficiency of click chemistry-based enrichment of AHA-containing NSP. Comparative analyses of the intensity ratios were especially important in the initial tests with different enrichment strategies and bead types, and for the optimization of the semi-automated enrichment protocol to determine the efficiency of the NSP enrichment (fig. 7 and fig. 10). Furthermore, using the intensity ratio metric it could be demonstrated that even with the inclusion of multiple stringent washing steps, different amounts of unlabeled pre-existing proteins are retained in the samples and cannot be easily excluded according to characteristics, such hydrophobicity [180]. However, it remains unclear whether unlabeled proteins in the enriched samples, originate from pre-existing proteins or from NSP, which incorporate light lysine or arginine due to insufficient amino acid starvation or recycling of degraded proteins. In the context of label-free analysis, the unlabeled proteins cannot be distinguished from NSP and prevent accurate quantification of changes in NSP abundance. The use of SILAC labels is therefore highly beneficial for the quantification and confident identification in NSP, since pre-existing proteins cannot be fully eliminated during washing steps.
However, chemical modifications of AHA-containing NSP also enables their distinction from pre-existing proteins. This has been demonstrated via coupling with biotin alkyne [196] or via alkyne phosphonates [141]. These analyses however, are limited to modified AHA-containing peptides, leading to relatively low sequence coverage.

Slightly reduced peptide identifications, compared to standard proteome samples (approximately 20 % lower), could also be observed in enriched NSP samples (fig. 28) since covalently bound AHA-containing peptides are not recovered from the MAA beads after

tryptic digestion. Interestingly, even with reduced precursor identifications, minor increases in the number of identified proteins are consistently measured in enriched newly synthesized proteome samples (fig. 13, fig. 28) [146]. This increase in protein identification could potentially be caused by a narrower dynamic range of NSP abundance, compared to overall protein abundance.

The optimized semi-automated NSP enrichment protocol has relatively low input requirements and was shown to produce high quality samples with $\leq$25 μg total protein input (fig. 14 and fig. 16). Importantly, the reduced input requirements simplify the handling of cultured cells. Through the use of 6-well plates, cells from the two compared experimental conditions (i.e. treatment and control) can be simultaneously labeled and harvested, reducing potential biases and increasing reproducibility.
Within the tested range of protein input (1-300 μg) the NSP enrichment was shown to be efficient, and deviations in the concentrations of MAA beads did also not severely impact data quality (fig. 13 and fig. 14). These results suggest that the semi-automated enrichment protocol operates in a robust manner across different conditions. Moreover, it can be inferred that moderate deviations in MAA-bead coupling efficiency are unlikely to significantly affects sample quality.

Adopting plexDIA, for the analysis of SILAC-labeled samples proved to be highly effective for newly synthesized proteome analysis. Similar to previously published reports, in which mTRAQ- [69] and di-methyl-labeled samples [68] were analyzed, a major increase in protein and peptide identifications could be achieved. In addition to the increase in identifications, high quantitative accuracy and precision were achieved in benchmark samples (fig. 15 and fig. S2) and subsequent analyses [180].

Previously published plexDIA analyses with chemical labels had to overcome issue with altered chromatograpic retention of peptides, where as SILAC-labeled peptides do not differ in this regard from unlabeled peptides. Therefore, in-silico prediction tools that are optimized for unmodified peptides, can readily be employed for library-free processing in SILAC plexDIA analysis.

Since quantification on both the MS1- and MS2-level are possible in DIA analysis, both modes of quantification were evaluated for plexDIA analysis of SILAC-labeled samples. Similarly to published reports on label-free DIA analysis [64][65], MS2-based quantification was shown to achieve higher accuracy and precision (lower CV values) in the SILAC plexDIA benchmark. The use of an optimized acquisition method, with multiple high resolution MS1 scans and thus shorter effective MS1 cycle time and increased data points per peak, was shown to enhance accuracy and precision for MS1-based quantification (fig. 15 and fig. S2). However, since all of the measurements and analyses, described in this thesis were performed using a QExactive HF orbitrap mass spectrometer, it is unclear whether the higher performance of MS2-based quantification is instrument and setup dependent. Analysis of mTRAQ-labeled samples, using a timsTOF SCP mass spectrometer, was for example carried out using MS1-based quantification [69].

In contrast to conventional DDA search engines, DIA-NN with plexDIA does not include the "re-quantify" function, which in Maxquant [157] and MSFragger [197] enables the calculation of SILAC ratios for peptides with intensities that are close to background levels. Together with the additional channel- and translated q-value filters, the lack of the re-quantify function most-likely leads to observed decrease in identifications in samples with low percentages of labeled proteins. However, even at low overall amounts of labeled proteins, quantification accuracy and linearity are excellent in SILAC plexDIA analysis (fig. 15 and fig. S3).

In contrast, for the analysis of enriched newly synthesized proteome samples, which primarily consist of similar amounts of SILAC labeled peptides, increased identifications was achieved, providing an additional advantage of the enrichment of NSP.

Moreover, the additional translated- and channel q-value filters appear to be highly effective for the confident identification of peptides in distinct labeling states. Strikingly, not a single SILAC-labeled peptide was identified in unlabeled samples that were processed together with samples containing different amounts of SILAC labeled peptides (fig. S3). No extensive optimization of the additional q-value filters were carried out, since high quality data were obtained using relatively stringent thresholds of 0.01. However, the stringent filtering lead to the removal of a large number of labeled precursors with low intensity (fig. S2 and fig. S3).

Optimization of the processing parameters in the DIA-NN software, in regard to q-value threshold and addition of the newly developed QuantUMS algorithm [198] could potentially improve proteomic depth and quantification in SILAC plexDIA analysis.

In addition to the described technical benchmarks, the QuaNPA workflow was used to analyze the effects of IFNg stimulation in a time resolved manner. Since IFNg signaling and its cellular targets are well-studied and described, the measured effects on the newly synthesized proteome serve as additional proof that biologically meaningful data are obtained through the workflow. For this purpose, data from the analysis were compared to published transcriptome [169] and ChIP-seq analyses [170] of IFNg-treated HeLa cells.

Moreover, two recently developed protocols also analyzed changes in newly synthesized- and secreted protein levels via automated enrichment protocols that employ the AssayMap Bravo robotic liquid handling platform. However, both of these protocols require relatively high amounts of protein input ($>$500 μg per condition or cells grown in 15 cm flasks) and include lengthy steps such as overnight de-phosphorylation and dialysis of biotinylated proteins [141][147]. It was demonstrated that the PhosID workflow is capable of detecting significant changes in NSP levels after 4 h and 24 h IFNg treatment. However, in contrast to data generated via QuaNPA, no significant enrichment of IFNg target proteins could be detected at the 4 h time point [180]. The comparison of these two workflows indicates that higher proteomic depth and quantitative accuracy can be achieved through QuaNPA. This assumption is suported by the fact, that up-regulation of several well characterized IFNg-regulated proteins such as ICAM1, STAT1 and SOD2 were detected as early as 2 h after IFNg treatment. Many other IFNg target proteins were also shown to be up-regulated after 4-24 h of IFNg treatment (fig. 18).

The time course analysis enabled the distinction of proteins with rapid changing synthesis rates, which are likely to be direct targets of IFNg signaling, from slower changes in NSP levels that are potentially caused by delayed or secondary effects of IFN stimulation. To determine whether differentially expressed NSP have previously been linked to IFNg, their presence in the “response to IFNg” gene ontology- (GO) and Hallmark gene sets of the Molecular signatures database were checked and the number of evidential datasets were

gathered from the interferome database [172]. Interestingly, several proteins with differential NSP abundance, following IFNg stimulation, had not previously been reported (fig. 19). For two of these potential new IFNg target proteins, additional evidence for changing protein expression could be obtained, via targeted proteomic analysis. Using a label-free PRM assay, moderately increased abundance of SIN3B and decreased abundance of BICD2 could measured after 24 h treatment with IFNg. However, no significant changes in protein abundance could be measured for the remaining candidates SUMO1, S100A6 and CRK (fig. 20). It can be speculated that longer IFNg treatment is required to cause a change in overall abundance of these candidate proteins, or that alternative effects of protein degradation contribute to the observed discrepancy in the newly synthesized proteome data and protein abundance changes that were measured via PRM. However, these alternative mechanisms would need to further investigated using for example measurements of protein degradation in response to IFNg treatment.

6.1.2 Quantifying global inhibition of mRNA translation using newly synthesized proteome analysis

Cellular stress and associated signaling pathways decrease mRNA translation mainly through inhibition of initiation factor complexes eIF2 and eIF4F [199]. Commonly applied normalization methods for label-free and labeling-based quantitative proteomics operate under the assumption that the majority or proteins are unchanged between the tested conditions [181][55]. In the case of newly synthesized proteome analysis of cells with reduced protein synthesis, this assumption is invalid.

Although enriched newly synthesized proteome samples are be capable of quantifying this effect due to unequal fractions of SILAC-labeled NSP, the resulting overall shift in SILAC ratios can not be observed if conventional median normalization is applied. However, the absence of normalization leads to potential mixing inaccuracies, obscuring global changes in protein synthesis.

A new normalization approach was therefore developed to enable the accurate quantification of protein synthesis inhibition. This approach is based on the analysis of samples, generated

from pSILAC-labeled cells, and determination of a scaling factor which is used for the normalization of newly synthesized proteome data (fig. 21).

The recently developed directLFQ algorithm also enables the scaling of data with the use of selected subsets of proteins, such as contaminant proteins and so-called housekeeping genes [56]. However, selecting an appropriate subset for normalization without additional biases is most-likely challenging. The median of commonly quantified NSP across all samples, in comparison likely presents a less biased method for the determination of accurate scaling factors for normalization. Moreover, acquiring data from samples, generated without NSP enrichment, adds valuable and complementary information on protein abundance levels that can be compared to changes in NSP levels (fig. 23).

Through the new normalization approach, a global decrease in NSP levels could be measured in HeLa cells, treated with 0.1 μg/mL. In addition to the quantification of globally decreased protein synthesis, shutdown of mRNA translation could also be measured in HeLa cells treated with high doses of CHX (1-100 μg/mL) (fig. 22).

Taken together these results confirm, that the QuaNPA workflow and proposed normalization method are capable of measuring the loss of protein expression and quantifying protein synthesis inhibition, in newly synthesized protome analysis.

6.1.3 Limitations and potential improvements

Although the described workflow marks an improvement on multiple aspects of newly synthesized proteome analysis, several limitations remain. However, additional optimizations and modifications of the QuaNPA workflow are possible and could expand its utility.

Generally, newly synthesized proteome analysis and the QuaNPA workflow are capable of determining changes in NSP levels on a proteome wide scale, but do not measure and account for changes in protein degradation between the tested conditions. Carrying out pulse-chase experiments with pSILAC and AHA labeling [200] would enable quantitative analysis of protein degradation and could be directly compatible with the QuaNPA workflow.

Due to the metabolic labeling with AHA, the established workflow is currently limited to

cultured cells. However, the established workflow would be compatible with the enrichment of ANL-labeled proteins in genetically modified organisms [130], enabling the analysis of proteins that are synthesized in different cell types. Tissue-derived samples, would most-likely rely on label-free quantification, which requires additional benchmarks for assessing quantification accuracy. Preparation of magnetic azide agarose beads [180], could also enable the enrichment of HPG-, β-ES or OPP-labeled proteins.

Data acquisition with modern mass spectrometers, featuring higher scan speeds and ion-mobility separation, is likely to also be beneficial for increasing proteomic depth in newly synthesized proteome analysis. Moreover, the increased sensitivity of newly developed mass spectrometers could reduce the input requirements of the QuaNPA workflow even further. However, the use of narrow isolation window DIA methods, as demonstrated for the Orbitrap Astral mass spectrometer [201], will increase the number of labeled precursors that will be split into multiple quadrupole isolation windows. The resulting effects on quantification accuracy and numbers of missing values, in the data obtained from multiplexed DIA analyses, would therefore need to evaluated.

AHA-containing peptides that are conjugated to MAA beads cannot be recovered after tryptic digestion, leading to reduced sequence coverage of identified proteins and lower peptide identifications in enriched newly synthesized proteome samples. The presence of cleavable alkyne moieties on magnetic agarose beads could enable the removal of full-length NSP. Cleavage and release has been demonstrated using Diazo- and Dde biotin-alkyne coupled NSP [202][203], but also via proteolytic cleavage of alkyne-containing peptides that are coupled to agarose beads [204]. The creation of magnetic agarose beads, conjugated with synthetic peptides that contain lysine or arginine and HPG or β-ES, would enable the coupling and release of all NSP-derived peptides through tryptic digestion. However, modified AHA-containing peptides are likely to exhibit altered chromatographic retention, adding to the computational challenges of identifying the modified peptides. The DIA-NN software is currently only optimized for the analysis of phosphorylated- and diglycine-conjugated peptides [205][206].

6.2 Systematic analysis of eIF-depletion induced changes in protein synthesis

Increasing evidence indicates that individual eIF and eIF sub-unit proteins are engaged in additional regulatory mechanisms, that facilitate the translation of specific mRNA. [207].

To assess by what extent individual eIF shape the proteome of cancer cells through “selective mRNA translation”, changes in protein synthesis were analyzed in a panel of RKO cells with induced depletion of 40 different eIF. Quantitative analysis of protein synthesis was performed through the calculation of translation efficiency (TE), based on data obtained from newly synthesized proteome analysis and RNA sequencing.

The gentically modified RKO cell lines, with inducible spCas9-mediated KO of the targeted eIF, proofed to be a powerful model system in this study. Induced KO lead to the effective depletion of the majority of targeted eIF (fig. 29), through the deletion of DNA segments, flanked by the two sgRNA (fig. 25). No evidence of leaky spCas9 expression could be detected in proteomic analyses of RKO cells (fig. S5). However, the gene editing-based KO of eIF takes a relatively long time to cause depletion of protein levels. This prolonged time frame required the selection of an optimized KO induction time point, which was uniformly set to 48 h. Despite the onset of reduced proliferation in RKO cells with induced KO of eIF3C, eIF4B and eIF4G3, it could be shown for the latter two that cell viability remains high at this time point (fig. 26 and fig. 27). Moreover, despite different published mRNA and protein half-lives [208][209], similarly decreased target eIF protein abundance could be measured for the majority of eIF, corresponding to an approximately 50 % decrease after 48 h (fig. 29).

Alternative systems for rapid depletion of target proteins include auxin-inducible degron (AID) [210] and dTAG system [211]. Through endogenous tagging of eIF proteins, complete depletion could the targeted eIF could theoretically be achieved, where as slow and gradual degradation of eIF through the Cas9-induced system only enables partial depletion. However, in comparison to the Cas9-induction model, the creation of 40 AID- or dTAG fused eIF proteins would have been a significantly more laborious cloning process for the collaborators in this project. Moreover, additional validation of the eIF fusion proteins would be required to ensure that the inducible degradation-tag does not impact their biological functions.

Quality controls of the newly synthesized- and global proteome data, indicate that proteomic data is of high quality and can be used for the analysis. Peptides derived from spCas9, helped to identify issues in the HPLC-separation for one of the 232 measured samples (fig. S8). No visible batch effects and biases could be observed in the acquired newly synthesized proteome data, where as strong batch effects were detected in proteome samples that were prepared in two batches using the autoSP3 protocol [151]. These observed differences can most likely be explained due to the fact that newly synthesized proteome samples contain an internal control for relative quantification, which is mixed with the lysate of cells in the other experimental condition early in the sample preparation process. Randomized experiment batches and 96-well plate positions are likely to be beneficial in accounting for batch effects as well (table 1 and fig. S6) [212].

Through use of the newly established normalization approach, different degrees of reduced and in several cases increased protein synthesis could be detected in the newly synthesized proteome data. Moderately increased average NSP ratios were detected in cells with depletion of eIF2D, eIF4E1B, eIF4E2, eIF4EBP1, eIF4H and PDCD4 (fig. 29). These results are consistent with the reported functions of eIF4E2, eIF4EBP1, and PDCD4 in inhibiting or competing with components of the eIF4F complex [95][99][213]. Interestingly, reduced protein synthesis has not yet been associated with eIF2D and eIF4H.

Moreover, large differences in the degree of protein synthesis inhibition could be observed for different eIF3 sub-complex proteins (fig. 29). Depletion of core components of eIF3, such as eIF3A, lead to a global decrease in NSP levels, where as depletion of its interaction partner eIF3J did not have similar effects (fig. 32). These findings match previously reported effects of eIF3 sub-unit knockdown and polysome profiling [188]. Interestingly, relatively large overlaps in differentially translated mRNA could be observed in RKO cells with depletion of central eIF3 components eIF3A, -B, -C and -E, where as overlap of these 4 sub-units with eIF3D is significantly lower (fig. S16). Additional cap-binding functions of eIF3D have been described previously, and the acquired data could provide additional evidence towards selective mRNA translation. However, in order to do so additional analysis is required.

The eIF3 sub-complex has been in the focus of numerous studies that identified additional

functions of individual eIF3 sub-units and investigated their roles in pathological conditions [78][118]. Recently published reports, on AID-induced depletion of several eIF sub-units, identified distinct changes in eIF3 sub-complex composition and different impacts on cell proliferation. In the same study, it was reported that eIF3K inhibits translation and selectively represses translation of ribosomal protein RPS15A [214]. However, these findings do not match the data that were acquired using the RKO cell panel via newly synthesized proteome analysis. Changes in RPS15A translation were detected in several conditions, but no significant increase was measured in response to eIF3K depletion (fig. 34).

The heterogenous changes in NSP levels, in response to eIF depletion, could also be highlighted using principal component analysis. Interestingly, eIF4B and eIF4G3, which previously stood out due to their rapid and strong effect on cell proliferation, exhibited strong separation of PC3 (fig. 30 and fig. 27). However, adjusting fold changes of the newly synthesized proteome data with measured changes in mRNA abundance, lead to the removal of these effects, suggesting that changing mRNA abundance and not translation are responsible (fig. 34). Indeed, induced mRNA and NSP levels of cell cycle regulatory protein CDKN1A were measured in eIF4B and eIF4G3 depleted cells (fig. S13). However, significantly changed mRNA abundance could only be measured for a minority of differentially regulated NSP (fig. 33).

A recently published study that analyzed changes in mRNA translation upon depletion of initiation factors in yeast (Saccharomyces cerevisiae), reported a global decrease in mRNA abundance, in response to the majority of depleted eIF. This was inferred from correlation analysis of fold changes from ribosomal occupancy and mRNA abundance data [215]. The importance of cap-binding eIF4F proteins and their inhibition of de-capping enzymes for the stabilization of mRNA have been described before in yeast [216][217]. The contribution of overall increased mRNA degradation can not easily be assessed in the acquired data since no spike-in normalization was carried out [218]. However, since only relatively few significant changes in abundance were measured, for mRNA encoding identified proteins, and since overall correlation between mRNA and NSP log2 fold change values are relatively low (fig. S17, panel A), it is very unlikely that increased mRNA degradation was the predominant

contribution to the measured changes in NSP abundance. Moreover, in the described analyses, NSP fold changes are normalized using changes in mRNA abundance to calculate TE values.

Having measured the eIF depletion specific changes in protein synthesis, the focus was shifted towards the investigation of clues towards potential mechanisms by which individual eIF could modulate specific mRNA translation. For this purpose, correlative analysis of mRNA characteristics and the measured changes in TE was performed.

However, prior to the comparative analysis of mRNA features, protein subsets needed to be selected. Initial analysis attempted to identify annotated protein sets, which are differentially translated in response to eIF depletion. Despite significant enrichment of ribosomal proteins, in the top 250 proteins contributing to the observed variance in the PCA, it could be highlighted that this annotated set of proteins only deviates from the median translation efficiency to a minor degree. Synthesis of several ribosomal proteins, however, exhibited strong differences among the different eIF depletion conditions (fig. 34). The selection of protein subsets was therefore exclusively carried out with the measured TE log2 fold change values. More specifically, the top 5 % and bottom 5% were selected and used for subsequent comparisons.

Several patterns could be observed by comparing mRNA characteristics among the protein subsets in RKO cells with induced depletion of individual eIF. The observed trend points towards strongly reduced translation of mRNA with longer 5′-UTR and 3′-UTR sequences, in response to depletion of several eIF. It can be speculated that these results are due to the presence of regulatory elements, such as secondary structures or sequence motifs, interacting with specific eIF sub-units. Interestingly, the most pronounced differences in UTR length could be observed in RKO cells with depletion of eIF3A, eIF3D and eIF4G2 (fig. 35 and fig. 36). Both eIF3D and eIF4G2 have previously been reported to be involved in the non-canonical translation initiation, through either 5′-cap binding of eIF3D [106][107] or via IRES- and uORF translation regulated by eIF4G2 [80][104]. Additional roles of eIF3A, outside of acting as nucleation core of the eIF3 sub-complex [188] are not as well described.

Moreover, for several eIF depletion conditions, translation of mRNA with short CDS was shown to be more strongly reduced than for mRNA with longer CDS. Differences between the depletion of individual eIF are not as striking as for the UTR lengths of the compared mRNA subsets, but several interesting observations could still be made. Previously mentioned eIF3D and eIF4G2 for example did not exhibit comparable effects to the majority of eIF, in the context of CDS length dependent changes in TE (fig. 36, panel B, -D and -F). Inverse correlation between CDS length and protein abundance has been reported before, and can also be detected in the acquired data sets (fig. S18) [219]. It has been hypothesized that this correlation is caused by different rates of translation re-initiation via ribosome recycling [220].

Several other mRNA characteristics and features have been compared with the measured changes in protein synthesis, following eIF depletion, but no significant differences or clear patterns were obtained (fig. 36 and fig. 37).

No major differences could be observed in the number of m6A modification sites in the 5′-UTR of the compared mRNA subsets. The overall importance and role of m6A modifications for the initiation of translation is still debated. Several eIF3 sub-units have been shown to interact with m6A-associated proteins, such as eIF3A, -B, and -H [105][221]. However, a recently published study, which employed a range of analytical methods, found no evidence for m6A-promoted translation initiation could be identified, in transcripts with m6A modifications in close proximity to the canonical start codon [222].

At the current stage in the analysis of changes in protein synthesis, in response to individual eIF, numerous open question remain and several limitations are present in the study. The following section is focused on discussing these limitations and future plans for the analysis of the acquired data. However, the obtained results highlight the complexity of the regulatory eIF network. The heterogenous effects on protein synthesis, caused by individual eIF depletion, further suggests that certain eIF sub-units modulate the translation of specific mRNA subsets. The precise mechanisms by which preferential mRNA translation occurs need to be further investigated, but the data presented in this thesis could potential aid in this research.

6.2.1 Limitations and outlook

The systematic analysis of eIF-depletion specific changes in mRNA translation, described in this thesis, has created high quality data that can be used for the investigation of selective mRNA translation. However, several limitations exists in regard to the acquired data and type of analysis that was performed in this thesis.

Firstly, mass spectrometry-based proteomics can currently not achieve coverage that is comparable to next-generation sequencing, introducing bias for the identification of highly abundant proteins. Through the use of DIA acquisition and plexDIA analysis, increased coverage could be obtained, but the approximately 5500 quantified (unique) proteins only represent a fraction of the human proteome. The depth of proteomic data should, however, still be sufficient for an investigation of eIF-mediated selective mRNA translation.

Moreover, distinction of protein isoforms is also challenging via bottom-up proteomic analysis [226]. Raw proteomic data was therefore matched to peptides that originate from canonical SwissProt entries [227]. Similarly, only canonical transcripts have been used for the analysis of specific mRNA characteristics. However, future analysis of the data could include different transcript isoforms, since $>$80000 transcripts were identified in the RKO cells using RNA sequencing.

In general, the analysis of the acquired data is exclusively correlative and does generate proof of specific mechanisms of selective mRNA translation. For that purpose, additional validation experiments could be carried out using for example luciferase reporter assays. Cloning different UTR constructs into the reporter could enable quantitative measurements of translation, that can be used to test hypotheses generated from analysis of the newly synthesized proteome analysis.

Moreover, only individual mRNA characteristics were compared at a time. It is possible that an interplay of multiple features is facilitating the selective translation of certain mRNA. Sophisticated bioinformatic analysis is required to to integrate the acquired data and to create a model in which different weights are assigned to the multiple mRNA characteristics. Information on reported or predicted secondary structures and analysis of sequence motifs

in the UTR and CDS of differentially translated mRNA could also provide hints towards preferential translation via eIF. These analyses will be carried out with collaborators in the near future.

Information on m6A modifications in the 5′-UTR of mRNA were retrieved from published datasets, which were acquired through analysis of different cell lines [176]. However, it has been reported that m6A modification patterns are relatively well conserved across human tissues, with the exception of the brain [228].

It also remains unclear if there are any differences in eIF activity in different cell- and tissue types. Evidence points towards different expression levels of eIF proteins throughout the human body and in different cancer sub-types [229]. It is therefore possible that specific interactions and regulatory effects on mRNA translation are more pronounced in one cell type compared to others. Future experiments could be performed in cell line models, other than RKO, to investigate the degree of cell type specificity. EIF proteins with differing expression levels across different tissues or cell types would make excellent candidates for such a study.

Moreover, all experiments were carried out in the presence of high oxygen and nutrient concentrations. Several 5′-cap-independent non-canonical modes of translation initiation have been shown to play crucial roles in protein synthesis under stress conditions. Additional regulatory functions of individual eIF, in regard to selective mRNA translation, could potentially become more prevalent under inhibition of canonical translation inhibition through cellular stress. Few or no significant changes in protein synthesis could be measured in RKO cells with depletion of several eIF, such as for example eIF3K, eIF4E2 and eIF4E3 which have been reported to play important roles in regulation of mRNA translation under stress conditions [214][230].

Future experiments with RKO cells under hypoxic or nutrient deprivation conditions, could provide additional clues towards stress-associated translation driven by individual eIF. The regulation of translation initiation under stress conditions is especially important in the context of cancer research, since tumor cells frequently cope with low oxygen and nutrient availability [73][112]. However, the presented data provides evidence for a high degree

of preferential translation of a large number of mRNA, by individual eIF without stress conditions.

Future analysis of the acquired data could also attempt to investigate known interactions of differentially translated mRNA with RNA-binding proteins (RBP) and non-coding RNA, potentially linking them to eIF driven selective mRNA translation. It has also been reported that ribosomal heterogeneity could add an additional regulatory layer to protein synthesis [231]. Differential synthesis of several ribosomal proteins was detected across the panel of eIF depletion conditions (fig. 34). It would therefore be interesting to investigate if specific eIF proteins facilitate preferential recruitment of specific ribosome sub-types. This could potentially be achieved through affinity purification of individual eIF, coupled with quantitative analysis of ribosomal proteins and their PTM, or ribosomal RNA (rRNA) composition. Moreover, profiling of mRNA bound to sub-populations of ribosomes could be compared with data on differential translation in response to eIF depletion, to investigate clues for this hypothesis.

Lastly, due to the relatively low overlap of differentially synthesized proteins among the individual eIF KO, it can be speculated that specific signatures of eIF sub-units can be identified. If such signatures of preferential mRNA translation of individual eIF proteins could be identified, it should be investigated whether evidence of increased or decreased protein levels are present in human tissues and tumor samples, with high expression levels of the respective eIF. The discovery of individual eIF signatures in the human proteome, could potentially aid in the development of novel therapeutic strategies.

6.2.2 Novel insights and perspective on eIF function

The research described in this thesis, highlights the complexity of mRNA translation initiation and its regulation. Despite the increasing number of reports on selective mRNA translation via individual eIF sub-units, the highly diverse changes in translation that were measured in response to individual eIF depletion are surprising. It will require additional research and analysis to gain a better understanding of the regulatory mechanisms by which certain eIF carry out these processes, but it is my hope that the work described in this thesis can aid in the process.

6.3 Concluding remarks

To summarize, in the course of this project a novel quantitative proteomic workflow for the analysis of newly synthesized proteins has been developed and employed for the analysis of a panel of cell lines with induced depletion of eIF proteins. Through the systematic analysis of changes in protein synthesis, wide-spread evidence for preferential translation of specific mRNA by individual eIF sub-unit proteins could be identified. However, additional analysis is required to determine the mechanisms by which preferential translation is carried out by specific eIF.

7 References

[1] Omenn, G. S. *et al.* The 2022 report on the human proteome from the hupo human proteome project. *Journal of proteome research* **22**, 1024–1042 (2022).

[2] Jiang, L. *et al.* A quantitative proteome map of the human body. *Cell* **183**, 269–283 (2020).

[3] Beck, M. *et al.* The quantitative proteome of a human cell line. *Molecular systems biology* **7**, 549 (2011).

[4] Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics* **1**, 845–867 (2002).

[5] Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nature methods* **10**, 186–187 (2013).

[6] Kelleher, N. L. Peer reviewed: top-down proteomics. *Analytical chemistry* **76**, 196–A (2004).

[7] Aebersold, R. *et al.* How many human proteoforms are there? *Nature chemical biology* **14**, 206–214 (2018).

[8] Chen, B., Brown, K. A., Lin, Z. & Ge, Y. Top-down proteomics: ready for prime time? *Analytical chemistry* **90**, 110–127 (2017).

[9] Kapp, E. A. *et al.* Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Analytical Chemistry* **75**, 6251–6264 (2003).

[10] Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin cleaves exclusively c-terminal to arginine and lysine residues. *Molecular & cellular proteomics* **3**, 608–614 (2004).

[11] Swaney, D. L., Wenger, C. D. & Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *Journal of proteome research* **9**, 1323–1329 (2010).

[12] Tsiatsiani, L. & Heck, A. J. Proteomics beyond trypsin. *The FEBS journal* **282**, 2612–2626 (2015).

[13] Pedder, R. E. Practical quadrupole theory: graphical theory. *Excel Core Mass Spectrometers, Pittsburgh, PA, Extrel Application Note RA_2010 A* (2001).

[14] Guilhaus, M. Special feature: Tutorial. principles and instrumentation in time-of-flight mass spectrometry. physical and instrumental concepts. *Journal of mass spectrometry* **30**, 1519–1532 (1995).

[15] Peters-Clarke, T., Coon, J. & Riley, N. Instrumentation at the leading edge of proteomics (2023).

[16] Hu, Q. *et al.* The orbitrap: a new mass spectrometer. *Journal of mass spectrometry* **40**, 430–443 (2005).

[17] Perry, R. H., Cooks, R. G. & Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass spectrometry reviews* **27**, 661–699 (2008).

[18] Denisov, E., Damoc, E. & Makarov, A. Exploring frontiers of orbitrap performance for long transients. *International Journal of Mass Spectrometry* **466**, 116607 (2021).

[19] Bekker-Jensen, D. B. *et al.* A compact quadrupole-orbitrap mass spectrometer with faims interface improves proteome coverage in short lc gradients. *Molecular & Cellular Proteomics* **19**, 716–729 (2020).

[20] Park, M. A. Apparatus and method for parallel flow ion mobility spectrometry combined with mass spectrometry (2010). US Patent 7,838,826.

[21] Meier, F. *et al.* Parallel accumulation–serial fragmentation (pasef): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device.

Journal of proteome research **14**, 5378–5387 (2015).

[22] Kolakowski, B. M. & Mester, Z. Review of applications of high-field asymmetric waveform ion mobility spectrometry (faims) and differential mobility spectrometry (dms). *Analyst* **132**, 842–864 (2007).

[23] Hebert, A. S. *et al.* Comprehensive single-shot proteomics with faims on a hybrid orbitrap mass spectrometer. *Analytical Chemistry* **90**, 9529–9537 (2018).

[24] Shi, Y., Xiang, R., Horváth, C. & Wilkins, J. A. The role of liquid chromatography in proteomics. *Journal of Chromatography A* **1053**, 27–36 (2004).

[25] Xie, F., Smith, R. D. & Shen, Y. Advanced proteomic liquid chromatography. *Journal of Chromatography A* **1261**, 78–90 (2012).

[26] Lenco, J. *et al.* Reversed-phase liquid chromatography of peptides for bottom-up proteomics: a tutorial. *Journal of Proteome Research* **21**, 2846–2892 (2022).

[27] Åsberg, D. *et al.* The importance of ion-pairing in peptide purification by reversed-phase liquid chromatography. *Journal of Chromatography A* **1496**, 80–91 (2017).

[28] Eeltink, S. *et al.* High-efficiency liquid chromatography–mass spectrometry separations with 50 mm, 250 mm, and 1 m long polymer-based monolithic capillary columns for the characterization of complex proteolytic digests. *Journal of chromatography A* **1217**, 6610–6615 (2010).

[29] Shishkova, E., Hebert, A. & Coon, J. Now, more than ever, proteomics needs better chromatography. cell systems 3: 321–324 (2016).

[30] Neue, U. D. Theory of peak capacity in gradient elution. *Journal of Chromatography A* **1079**, 153–161 (2005).

[31] Ettre, L. S. Nomenclature for chromatography (iupac recommendations 1993). *Pure and Applied Chemistry* **65**, 819–872 (1993).

[32] Tanaka, K. *et al.* Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid communications in mass spectrometry* **2**, 151–153 (1988).

[33] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).

[34] Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization (2013).

[35] Annesley, T. M. Ion suppression in mass spectrometry. *Clinical chemistry* **49**, 1041–1044 (2003).

[36] Chowdhury, S. K. & Chait, B. T. Method for the electrospray ionization of highly conductive aqueous solutions. *Analytical chemistry* **63**, 1660–1664 (1991).

[37] ROEPSTORFE, P. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601–605 (1984).

[38] Wells, J. M. & McLuckey, S. A. Collision-induced dissociation (cid) of peptides and proteins. *Methods in enzymology* **402**, 148–185 (2005).

[39] Olsen, J. V. *et al.* Higher-energy c-trap dissociation for peptide modification analysis. *Nature methods* **4**, 709–712 (2007).

[40] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences* **101**, 9528–9533 (2004).

[41] Gillet, L. C. *et al.* Targeted data extraction of the ms/ms spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics* **11** (2012).

[42] Stahl, D. C., Swiderek, K. M., Davis, M. T. & Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures.

Journal of the American Society for Mass Spectrometry **7**, 532–540 (1996).

[43] Tyanova, S., Temu, T. & Cox, J. The maxquant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols* **11**, 2301–2319 (2016).

[44] Yu, F., Haynes, S. E. & Nesvizhskii, A. I. Ionquant enables accurate and sensitive label-free quantification with fdr-controlled match-between-runs. *Molecular & Cellular Proteomics* **20** (2021).

[45] Kalxdorf, M., Müller, T., Stegle, O. & Krijgsveld, J. Icer improves proteome coverage and data completeness in global and single-cell proteomics. *Nature Communications* **12**, 4787 (2021).

[46] Gallien, S., Kim, S. Y. & Domon, B. Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring (is-prm)*[s]. *Molecular & Cellular Proteomics* **14**, 1630–1644 (2015).

[47] Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics* **11**, 1475–1488 (2012).

[48] Purvine, S., Eppel*, J.-T., Yi, E. C. & Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3**, 847–850 (2003).

[49] Panchaud, A. *et al.* Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Analytical chemistry* **81**, 6481–6488 (2009).

[50] Geiger, T., Cox, J. & Mann, M. Proteomics on an orbitrap benchtop mass spectrometer using all-ion fragmentation. *Molecular & Cellular Proteomics* **9**, 2252–2261 (2010).

[51] Tsou, C.-C. *et al.* Dia-umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature methods* **12**, 258–264 (2015).

[52] Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. Dia-nn: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods* **17**, 41–44 (2020).

[53] Kitata, R. B., Yang, J.-C. & Chen, Y.-J. Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. *Mass spectrometry reviews* **42**, 2324–2348 (2023).

[54] Meier, F. *et al.* diapasef: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature methods* **17**, 1229–1236 (2020).

[55] Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Molecular & cellular proteomics* **13**, 2513–2526 (2014).

[56] Ammar, C., Schessner, J. P., Willems, S., Michaelis, A. C. & Mann, M. Accurate label-free quantification by directlfq to compare unlimited numbers of proteomes. *Molecular & Cellular Proteomics* 100581 (2023).

[57] Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Analytical chemistry* **75**, 1895–1904 (2003).

[58] Aggarwal, K., Choe, L. H. & Lee, K. H. Shotgun proteomics using the itraq isobaric tags. *Briefings in Functional Genomics* **5**, 112–120 (2006).

[59] Li, J. *et al.* Tmtpro-18plex: the expanded and complete set of tmtpro reagents for sample multiplexing. *Journal of proteome research* **20**, 2964–2972 (2021).

[60] Mann, M. Functional and quantitative proteomics using silac. *Nature reviews Molecular cell biology* **7**, 952–958 (2006).

[61] Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed silac. *Proteomics* **9**, 205–209 (2009).

[62] Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature protocols* **4**, 484–494 (2009).

[63] Kang, U.-B., Yeom, J., Kim, H. & Lee, C. Quantitative analysis of mtraq-labeled proteome using full ms scans. *Journal of Proteome research* **9**, 3750–3758 (2010).

[64] Rardin, M. J. *et al.* Ms1 peptide ion intensity chromatograms in ms2 (swath) data independent acquisitions. improving post acquisition analysis of proteomic experiments*[s]. *Molecular & Cellular Proteomics* **14**, 2405–2419 (2015).

[65] Huang, T. *et al.* Combining precursor and fragment information for improved detection of differential abundance in data independent acquisition. *Molecular & Cellular Proteomics* **19**, 421–430 (2020).

[66] Sinitcyn, P. *et al.* Maxdia enables library-based and library-free data-independent acquisition proteomics. *Nature biotechnology* **39**, 1563–1573 (2021).

[67] Pino, L. K., Baeza, J., Lauman, R., Schilling, B. & Garcia, B. A. Improved silac quantification with data-independent acquisition to investigate bortezomib-induced protein degradation. *Journal of proteome research* **20**, 1918–1927 (2021).

[68] Thielert, M. *et al.* Robust dimethyl-based multiplex-dia doubles single-cell proteome depth via a reference channel. *Molecular Systems Biology* **19**, e11503 (2023).

[69] Derks, J. *et al.* Increasing the throughput of sensitive proteomics by plexdia. *Nature biotechnology* **41**, 50–59 (2023).

[70] Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of translation of single mrna molecules in vivo. *Cell* **165**, 976–989 (2016).

[71] Wu, B., Eliscovich, C., Yoon, Y. J. & Singer, R. H. Translation dynamics of single mrnas in live cells and neurons. *Science* **352**, 1430–1435 (2016).

[72] Morisaki, T. *et al.* Real-time quantification of single rna translation dynamics in living cells. *Science* **352**, 1425–1429 (2016).

[73] Sriram, A., Bohlen, J. & Teleman, A. A. Translation acrobatics: how cancer cells exploit alternate modes of translational initiation. *EMBO reports* **19**, e45947 (2018).

[74] Brito Querido, J., Díaz-López, I. & Ramakrishnan, V. The molecular basis of translation initiation and its regulation in eukaryotes. *Nature Reviews Molecular Cell Biology* 1–19 (2023).

[75] Cuchalová, L. *et al.* The rna recognition motif of eukaryotic translation initiation factor 3g (eif3g) is required for resumption of scanning of posttermination ribosomes for reinitiation on gcn4 and together with eif3i stimulates linear scanning. *Molecular and cellular biology* **30**, 4671–4686 (2010).

[76] Brito Querido, J. *et al.* Structure of a human 48 s translational initiation complex. *Science* **369**, 1220–1227 (2020).

[77] Sokabe, M. & Fraser, C. S. Human eukaryotic initiation factor 2 (eif2)-gtp-met-trnai ternary complex and eif3 stabilize the 43 s preinitiation complex. *Journal of Biological Chemistry* **289**, 31827–31836 (2014).

[78] Wolf, D. A., Lin, Y., Duan, H. & Cheng, Y. eif-three to tango: emerging functions of translation initiation factor eif3 in protein synthesis and disease. *Journal of molecular cell biology* **12**, 403–409 (2020).

[79] Robert, F., Cencic, R., Cai, R., Schmeing, T. M. & Pelletier, J. Rna-tethering assay and eif4g: eif4a obligate dimer design uncovers multiple eif4f functional complexes. *Nucleic acids research* **48**, 8562–8575 (2020).

[80] Liberman, N. *et al.* Dap5 associates with eif2β and eif4ai to promote internal ribosome entry site driven translation. *Nucleic acids research* **43**, 3764–3775 (2015).

[81] Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the helicase activity of eif4a by eif4b, eif4h, and eif4f. *Journal of Biological Chemistry* **276**, 30914–

30922 (2001).

[82] Gallie, D. The cap and poly (a) tail function synergistically to regulate mrna translational efficiency. *Genes & development* **5**, 2108–2116 (1991).

[83] Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. Circularization of mrna by eukaryotic translation initiation factors. *Molecular cell* **2**, 135–140 (1998).

[84] Kozak, M. Role of atp in binding and migration of 40s ribosomal subunits. *Cell* **22**, 459–467 (1980).

[85] Kozak, M. Point mutations close to the aug initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308**, 241–246 (1984).

[86] Lapointe, C. P. *et al.* eif5b and eif1a reorient initiator trna to allow ribosomal subunit joining. *Nature* **607**, 185–190 (2022).

[87] Xu, B., Liu, L. & Song, G. Functions and regulation of translation elongation factors. *Frontiers in Molecular Biosciences* **8**, 816398 (2022).

[88] Saini, P., Eyler, D. E., Green, R. & Dever, T. E. Hypusine-containing protein eif5a promotes translation elongation. *Nature* **459**, 118–121 (2009).

[89] Schmidt, C. *et al.* Structure of the hypusinylated eukaryotic translation factor eif-5a bound to the ribosome. *Nucleic acids research* **44**, 1944–1951 (2016).

[90] Wagner, S. *et al.* Selective translation complex profiling reveals staged initiation and co-translational assembly of initiation factor complexes. *Molecular cell* **79**, 546–560 (2020).

[91] Mohammad, M. P., Munzarová Pondělíčková, V., Zeman, J., Gunišová, S. & Valášek, L. S. In vivo evidence that eif3 stays bound to ribosomes elongating and terminating on short upstream orfs to promote reinitiation. *Nucleic acids research* **45**, 2658–2674 (2017).

[92] Proud, C. G. Phosphorylation and signal transduction pathways in translational control. *Cold Spring Harbor perspectives in biology* **11**, a033050 (2019).

[93] Wek, R. C. Role of eif2α kinases in translational control and adaptation to cellular stress. *Cold Spring Harbor perspectives in biology* **10**, a032870 (2018).

[94] Pakos-Zebrucka, K. *et al.* The integrated stress response. *EMBO reports* **17**, 1374–1395 (2016).

[95] Dorrello, N. V. *et al.* S6k1-and ßtrcp-mediated degradation of pdcd4 promotes protein translation and cell growth. *Science* **314**, 467–471 (2006).

[96] Pyronnet, S. *et al.* Human eukaryotic translation initiation factor 4g (eif4g) recruits mnk1 to phosphorylate eif4e. *The EMBO journal* **18**, 270–279 (1999).

[97] Scheper, G. C. *et al.* Phosphorylation of eukaryotic initiation factor 4e markedly reduces its affinity for capped mrna. *Journal of Biological Chemistry* **277**, 3303–3309 (2002).

[98] Pause, A. *et al.* Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* **371**, 762–767 (1994).

[99] Yang, M., Lu, Y., Piao, W. & Jin, H. The translational regulation in mtor pathway. *Biomolecules* **12**, 802 (2022).

[100] Christie, M. & Igreja, C. eif4e-homologous protein (4ehp): a multifarious cap-binding protein. *The FEBS Journal* **290**, 266–285 (2021).

[101] Walters, B. & Thompson, S. R. Cap-independent translational control of carcinogenesis. *Frontiers in oncology* **6**, 128 (2016).

[102] Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mrna directed by a sequence derived from poliovirus rna. *Nature* **334**, 320–325 (1988).

[103] Lee, K.-M., Chen, C.-J. & Shih, S.-R. Regulation mechanisms of viral ires-driven translation. *Trends in microbiology* **25**, 546–561 (2017).

[104] Hanson, P. *et al.* Cleavage of dap5 by coxsackievirus b3 2a protease facilitates viral replication and enhances apoptosis by altering translation of ires-containing genes. *Cell Death & Differentiation* **23**, 828–840 (2016).

[105] Meyer, K. D. *et al.* 5' utr m6a promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).

[106] Lee, A. S., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. eif3d is an mrna cap-binding protein that is required for specialized translation initiation. *Nature* **536**, 96–99 (2016).

[107] Volta, V. *et al.* A dap5/eif3d alternate mrna translation mechanism promotes differentiation and immune suppression by human regulatory t cells. *Nature communications* **12**, 6979 (2021).

[108] Komar, A. A. & Merrick, W. C. A retrospective on eif2a-and not the alpha subunit of eif2. *International Journal of Molecular Sciences* **21**, 2054 (2020).

[109] Dmitriev, S. E. *et al.* Gtp-independent trna delivery to the ribosomal p-site by a novel eukaryotic translation factor. *Journal of Biological Chemistry* **285**, 26779–26787 (2010).

[110] Elfakess, R. *et al.* Unique translation initiation of mrnas-containing tisu element. *Nucleic acids research* **39**, 7598–7609 (2011).

[111] Tahmasebi, S., Khoutorsky, A., Mathews, M. B. & Sonenberg, N. Translation deregulation in human disease. *Nature Reviews Molecular Cell Biology* **19**, 791–807 (2018).

[112] Silvera, D., Formenti, S. C. & Schneider, R. J. Translational control in cancer. *Nature Reviews Cancer* **10**, 254–266 (2010).

[113] Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nature reviews Drug discovery* **14**, 261–278 (2015).

[114] Bartish, M. *et al.* The role of eif4f-driven mrna translation in regulating the tumour microenvironment. *Nature Reviews Cancer* 1–18 (2023).

[115] Chen, Y. *et al.* Loss of pdcd4 expression in human lung cancer correlates with tumour progression and prognosis. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **200**, 640–646 (2003).

[116] Afonja, O., Juste, D., Das, S., Matsuhashi, S. & Samuels, H. H. Induction of pdcd4 tumor suppressor gene expression by rar agonists, antiestrogen and her-2/neu antagonist in breast cancer cells. evidence for a role in apoptosis. *Oncogene* **23**, 8135–8145 (2004).

[117] de la Parra, C., Walters, B. A., Geter, P. & Schneider, R. J. Translation initiation factors and their relevance in cancer. *Current opinion in genetics & development* **48**, 82–88 (2018).

[118] Hershey, J. W. The role of eif3 and its individual subunits in cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1849**, 792–800 (2015).

[119] Iwasaki, S. & Ingolia, N. T. The growing toolbox for protein synthesis studies. *Trends in biochemical sciences* **42**, 612–624 (2017).

[120] Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

[121] Buccitelli, C. & Selbach, M. mrnas, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics* **21**, 630–644 (2020).

[122] Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature reviews genetics* **15**, 205–213 (2014).

[123] Chothani, S. *et al.* deltate: detection of translationally regulated genes by integrative analysis of ribo-seq and rna-seq data. *Current protocols in molecular biology* **129**, e108 (2019).

[124] Chothani, S. P. *et al.* A high-resolution map of human rna translation. *Molecular cell* **82**, 2885–2899 (2022).

[125] Liu, T.-Y. *et al.* Time-resolved proteomics extends ribosome profiling-based measurements of protein synthesis dynamics. *Cell systems* **4**, 636–644 (2017).

[126] Aviner, R. The science of puromycin: From studies of ribosome function to applications in biotechnology. *Computational and Structural Biotechnology Journal* **18**, 1074–1083 (2020).

[127] Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *Proceedings of the National Academy of Sciences* **109**, 413–418 (2012).

[128] van Hest, J. C., Kiick, K. L. & Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *Journal of the American Chemical Society* **122**, 1282–1288 (2000).

[129] Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (boncat). *Proceedings of the National Academy of Sciences* **103**, 9482–9487 (2006).

[130] Ngo, J. T. *et al.* Cell-selective metabolic labeling of proteins. *nature CHeMICaL BIOLOGY* **5**, 715–717 (2009).

[131] Ignacio, B. J. *et al.* Throncat: metabolic labeling of newly synthesized proteins using a bioorthogonal threonine analog. *Nature Communications* **14**, 3367 (2023).

[132] Bagert, J. D. *et al.* Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. *Molecular & Cellular Proteomics* **13**, 1352–1358 (2014).

[133] van Bergen, W., Heck, A. J. & Baggelaar, M. P. Recent advancements in mass spectrometry–based tools to investigate newly synthesized proteins. *Current Opinion*

in Chemical Biology **66**, 102074 (2022).

[134] Aviner, R., Geiger, T. & Elroy-Stein, O. Genome-wide identification and quantification of protein synthesis in cultured cells and whole tissues by puromycin-associated nascent chain proteomics (punch-p). *Nature protocols* **9**, 751–760 (2014).

[135] Forester, C. M. *et al.* Revealing nascent proteomics in signaling pathways and cell differentiation. *Proceedings of the National Academy of Sciences* **115**, 2353–2358 (2018).

[136] Uchiyama, J., Ishihama, Y. & Imami, K. Quantitative nascent proteome profiling by dual-pulse labelling with o-propargyl-puromycin and stable isotope-labelled amino acids. *The Journal of Biochemistry* **169**, 227–236 (2021).

[137] Uchiyama, J. *et al.* psnap: Proteome-wide analysis of elongating nascent polypeptide chains. *Iscience* **25**, 104516 (2022).

[138] Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S. & Krijgsveld, J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nature biotechnology* **30**, 984–990 (2012).

[139] Howden, A. J. *et al.* Quancat: quantitating proteome dynamics in primary cells. *Nature methods* **10**, 343–346 (2013).

[140] Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates III, J. R. Hilaq: A novel strategy for newly synthesized protein quantification. *Journal of proteome research* **16**, 2213–2220 (2017).

[141] Kleinpenning, F., Steigenberger, B., Wu, W. & Heck, A. J. Fishing for newly synthesized proteins with phosphonate-handles. *Nature Communications* **11**, 3244 (2020).

[142] Eichelbaum, K. & Krijgsveld, J. Combining pulsed silac labeling and click-chemistry for quantitative secretome analysis. In *Exocytosis and Endocytosis*, 101–114 (Springer, 2014).

[143] Zhang, G. *et al.* In-depth quantitative proteomic analysis of de novo protein synthesis induced by brain-derived neurotrophic factor. *Journal of proteome research* **13**, 5707–5714 (2014).

[144] Elder, M. K. *et al.* Age-dependent shift in the de novo proteome accompanies pathogenesis in an alzheimer's disease mouse model. *Communications Biology* **4**, 823 (2021).

[145] Ma, Y., McClatchy, D. B., Martínez-Bartolomé, S., Bamberger, C. & Yates III, J. R. Temporal quantitative profiling of newly synthesized proteins during aβ accumulation. *Journal of proteome research* **20**, 763–775 (2020).

[146] Eichelbaum, K. & Krijgsveld, J. Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. *Molecular & Cellular Proteomics* **13**, 792–810 (2014).

[147] Vargas-Diaz, D. & Altelaar, M. Automated high-throughput method for the fast, robust, and reproducible enrichment of newly synthesized proteins. *Journal of Proteome Research* **21**, 189–199 (2021).

[148] Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates, J. R. Quantitative analysis of newly synthesized proteins. *Nature protocols* **13**, 1744–1762 (2018).

[149] Ekkebus, R. *et al.* On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *Journal of the American Chemical Society* **135**, 2867–2870 (2013).

[150] Hong, V., Presolski, S. I., Ma, C. & Finn, M. a. G. Analysis and optimization of copper-catalyzed azide–alkyne cycloaddition for bioconjugation. *Angewandte Chemie* **121**, 10063–10067 (2009).

[151] Müller, T. *et al.* Automated sample preparation with sp 3 for low-input clinical proteomics. *Molecular systems biology* **16**, e9111 (2020).

[152] Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* **10**, 757 (2014).

[153] Saiki, R. K. *et al.* Primer-directed enzymatic amplification of dna with a thermostable dna polymerase. *Science* **239**, 487–491 (1988).

[154] Sanger, F., Nicklen, S. & Coulson, A. R. Dna sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **74**, 5463–5467 (1977).

[155] Moll, P., Ante, M., Seitz, A. & Reda, T. Quantseq 3' mrna sequencing for rna quantification (2014).

[156] Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology* **15**, 1–21 (2014).

[157] Cox, J. *et al.* Andromeda: a peptide search engine integrated into the maxquant environment. *Journal of proteome research* **10**, 1794–1805 (2011).

[158] Frankenfield, A. M., Ni, J., Ahmed, M. & Hao, L. Protein contaminants matter: building universal protein contaminant libraries for dda and dia proteomics. *Journal of proteome research* **21**, 2104–2113 (2022).

[159] MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).

[160] Pino, L. K. *et al.* The skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass spectrometry reviews* **39**, 229–244 (2020).

[161] Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature methods* **16**, 509–518 (2019).

[162] Choi, M. *et al.* Msstats: an r package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).

[163] Kohler, D. *et al.* Msstats version 4.0: statistical analyses of quantitative mass spectrometry-based proteomic experiments with chromatography-based quantifi-

cation at scale. *Journal of Proteome Research* **22**, 1466–1482 (2023).

[164] Pham, T. V., Henneman, A. A. & Jimenez, C. R. iq: an r package to estimate relative protein abundances from ion quantification in dia-ms-based proteomics. *Bioinformatics* **36**, 2611–2613 (2020).

[165] Ritchie, M. E. *et al.* limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research* **43**, e47–e47 (2015).

[166] Zhu, Y. *et al.* Deqms: a method for accurate variance estimation in differential protein expression analysis. *Molecular & Cellular Proteomics* **19**, 1047–1057 (2020).

[167] Wu, T. *et al.* clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation* **2** (2021).

[168] Liberzon, A. *et al.* Molecular signatures database (msigdb) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

[169] Siwek, W., Tehrani, S. S., Mata, J. F. & Jansen, L. E. Activation of clustered ifnγ target genes drives cohesin-controlled transcriptional memory. *Molecular cell* **80**, 396–409 (2020).

[170] Davis, C. A. *et al.* The encyclopedia of dna elements (encode): data portal update. *Nucleic acids research* **46**, D794–D801 (2018).

[171] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).

[172] Rusinova, I. *et al.* Interferome v2. 0: an updated database of annotated interferon-regulated genes. *Nucleic acids research* **41**, D1040–D1046 (2012).

[173] Martin, F. J. *et al.* Ensembl 2023. *Nucleic acids research* **51**, D933–D941 (2023).

[174] Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS computational biology* **9**, e1003118 (2013).

[175] Chew, G.-L., Pauli, A. & Schier, A. F. Conservation of uorf repressiveness and sequence features in mouse, human and zebrafish. *Nature communications* **7**, 11663 (2016).

[176] Tang, Y. *et al.* m6a-atlas: a comprehensive knowledgebase for unraveling the n 6-methyladenosine (m6a) epitranscriptome. *Nucleic Acids Research* **49**, D134–D143 (2021).

[177] Andersson, T., Carlsson, M., Hagel, L., Pernemalm, P.-Å. & Jansson, J.-C. Agarose-based media for high-resolution gel filtration of biopolymers. *Journal of Chromatography A* **326**, 33–44 (1985).

[178] Safarik, I. & Safarikova, M. Magnetic techniques for the isolation and purification of proteins and peptides. *Biomagnetic research and technology* **2**, 1–17 (2004).

[179] Kitata, R. B., Yang, J.-C. & Chen, Y.-J. Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. *Mass spectrometry reviews* e21781 (2022).

[180] Borteçen, T., Müller, T. & Krijgsveld, J. An integrated workflow for quantitative analysis of the newly synthesized proteome. *Nature Communications* **14**, 8237 (2023). URL `https://www.nature.com/articles/s41467-023-43919-3`.

[181] Ting, L. *et al.* Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Molecular & Cellular Proteomics* **8**, 2227–2242 (2009).

[182] Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor perspectives in biology* **4**, a013706 (2012).

[183] Baliga, B., Pronczuk, A. & Munro, H. Mechanism of cycloheximide inhibition of protein synthesis in a cell-free system prepared from rat liver. *Journal of Biological Chemistry* **244**, 4480–4489 (1969).

[184] Brattain, M. *et al.* Initiation and characterization of cultures of human colonic carcinoma with different biological characteristics utilizing feeder layers of confluent fibroblasts. *Oncodevelopmental biology and medicine: the journal of the International Society for Oncodevelopmental Biology and Medicine* **2**, 355–366 (1981).

[185] Michlits, G. *et al.* Multilayered vbc score predicts sgrnas that efficiently generate loss-of-function alleles. *Nature Methods* **17**, 708–716 (2020).

[186] Kotin, R. M., Linden, R. M. & Berns, K. I. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus dna by non-homologous recombination. *The EMBO journal* **11**, 5071–5078 (1992).

[187] Yang, H.-S. *et al.* The transformation suppressor pdcd4 is a novel eukaryotic translation initiation factor 4a binding protein that inhibits translation. *Molecular and cellular biology* **23**, 26–37 (2003).

[188] Wagner, S., Herrmannová, A., Šikrová, D. & Valášek, L. S. Human eif3b and eif3a serve as the nucleation core for the assembly of eif3 into two interconnected modules: the yeast-like core and the octamer. *Nucleic acids research* gkw972 (2016).

[189] Wagner, S., Herrmannová, A., Malík, R., Peclinovská, L. & Valášek, L. S. Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. *Molecular and cellular biology* **34**, 3041–3052 (2014).

[190] Budanov, A. V. & Karin, M. p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mtor signaling. *Cell* **134**, 451–460 (2008).

[191] LaBaer, J. *et al.* New functional activities for the p21 family of cdk inhibitors. *Genes & development* **11**, 847–862 (1997).

[192] Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mrnas. *Genome biology* **3**, 1–10 (2002).

[193] Chan, C. Y. *et al.* A structural interpretation of the effect of gc-content on efficiency of rna interference. *BMC bioinformatics* **10**, 1–7 (2009).

[194] Cockman, E., Anderson, P. & Ivanov, P. Top mrnps: molecular mechanisms and principles of regulation. *Biomolecules* **10**, 969 (2020).

[195] Shen, S.-M. *et al.* Downregulation of anp32b, a novel substrate of caspase-3, enhances caspase-3 activation and apoptosis induction in myeloid leukemic cells. *Carcinogenesis* **31**, 419–426 (2010).

[196] Schiapparelli, L. M. *et al.* Direct detection of biotinylated proteins by mass spectrometry. *Journal of proteome research* **13**, 3966–3978 (2014).

[197] 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. *Nature methods* **14**, 513–520 (2017).

[198] Kistner, F., Grossmann, J. L., Sinn, L. R. & Demichev, V. Quantums: uncertainty minimisation enables confident quantification in proteomics. *bioRxiv* 2023–06 (2023).

[199] Spriggs, K. A., Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Molecular cell* **40**, 228–237 (2010).

[200] McShane, E. *et al.* Kinetic analysis of protein stability reveals age-dependent degradation. *Cell* **167**, 803–815 (2016).

[201] Heil, L. R. *et al.* Evaluating the performance of the astral mass analyzer for quantitative proteomics using data independent acquisition. *bioRxiv* 2023–06 (2023).

[202] Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *Journal of the American Chemical Society* **132**, 18351–18360 (2010).

[203] Phillips, N. J., Vinaithirthan, B. M., Oses-Prieto, J. A., Chalkley, R. J. & Burlingame, A. L. Capture, release, and identification of newly synthesized proteins for improved profiling of functional translatomes. *Molecular & Cellular Proteomics* **22** (2023).

[204] Meng, P. *et al.* Peptide-and protein-level combined strategy for analyzing newly synthesized proteins by integrating tandem orthogonal proteolysis with cleavable bioorthogonal tagging. *Analytical Chemistry* **95**, 628–637 (2022).

[205] Steger, M. *et al.* Time-resolved in vivo ubiquitinome profiling by dia-ms reveals usp7 targets on a proteome-wide scale. *Nature communications* **12**, 5399 (2021).

[206] Lou, R. *et al.* Benchmarking commonly used software suites and analysis workflows for dia proteomics and phosphoproteomics. *Nature Communications* **14**, 94 (2023).

[207] Kovalski, J. R., Kuzuoglu-Ozturk, D. & Ruggero, D. Protein synthesis control in cancer: selectivity and therapeutic targeting. *The EMBO Journal* **41**, e109823 (2022).

[208] Agarwal, V. & Kelley, D. R. The genetic and biochemical determinants of mRNA degradation rates in mammals. *Genome Biology* **23**, 245 (2022). URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-

[209] Zecha, J. *et al.* Peptide level turnover measurements enable the study of proteoform dynamics. *Molecular & Cellular Proteomics* **17**, 974–992 (2018).

[210] Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods* **6**, 917–922 (2009).

[211] Nabet, B. *et al.* The dtag system for immediate and target-specific protein degradation. *Nature chemical biology* **14**, 431–441 (2018).

[212] Burger, B., Vaudel, M. & Barsnes, H. Importance of block randomization when designing proteomics experiments. *Journal of proteome research* **20**, 122–128 (2020).

[213] Rosettani, P., Knapp, S., Vismara, M.-G., Rusconi, L. & Cameron, A. D. Structures of the human eif4e homologous protein, h4ehp, in its m7gtp-bound and unliganded forms. *Journal of molecular biology* **368**, 691–705 (2007).

[214] Duan, H. *et al.* eif3 mrna selectivity profiling reveals eif3k as a cancer-relevant regulator of ribosome content. *The EMBO Journal* e112362 (2023).

[215] Wang, J., Zhang, G., Qian, W. & Li, K. Decoding the heterogeneity and specialized function of translation machinery through ribosome profiling in yeast mutants of initiation factors. *Advanced Biology* 2300494 (2023).

[216] Schwartz, D. C. & Parker, R. Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mrnas in saccharomyces cerevisiae. *Molecular and cellular biology* **19**, 5247–5256 (1999).

[217] Chan, L. Y., Mugler, C. F., Heinrich, S., Vallotton, P. & Weis, K. Non-invasive measurement of mrna decay reveals translation initiation as the major determinant of mrna stability. *Elife* **7**, e32536 (2018).

[218] Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).

[219] Lyu, X., Yang, Q., Zhao, F. & Liu, Y. Codon usage and protein length-dependent feedback from translation elongation regulates translation initiation and elongation speed. *Nucleic acids research* **49**, 9404–9423 (2021).

[220] Rogers, D. W., Böttcher, M. A., Traulsen, A. & Greig, D. Ribosome reinitiation can explain length-dependent translation of messenger rna. *PLoS computational biology* **13**, e1005592 (2017).

[221] Choe, J. *et al.* mrna circularization by mettl3–eif3h enhances translation and promotes oncogenesis. *Nature* **561**, 556–560 (2018).

[222] Guca, E. *et al.* N6-methyladenosine in 5' utr does not promote translation initiation. *Molecular Cell* (2024).

[223] Obermann, W. *et al.* Broad anti-pathogen potential of dead box rna helicase eif4a-targeting rocaglates. *Scientific Reports* **13**, 9297 (2023).

[224] Jiang, S.-L. *et al.* Targeting translation regulators improves cancer therapy. *Genomics* **113**, 1247–1256 (2021).

[225] Ho, J. D. *et al.* Proteomics reveal cap-dependent translation inhibitors remodel the translation machinery and translatome. *Cell reports* **37** (2021).

[226] Sinitcyn, P. *et al.* Global detection of human variants and isoforms by deep proteome sequencing. *Nature Biotechnology* 1–11 (2023).

[227] Uniprot: the universal protein knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).

[228] Li, K. *et al.* Landscape and regulation of m6a and m6am methylome across human and mouse tissues. *Molecular cell* **77**, 426–440 (2020).

[229] Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

[230] Frydryskova, K. *et al.* Distinct recruitment of human eif4e isoforms to processing bodies and stress granules. *BMC molecular biology* **17**, 1–19 (2016).

[231] Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature reviews Molecular cell biology* **13**, 355–369 (2012).

8 Appendix

8.1 Supplementary figures

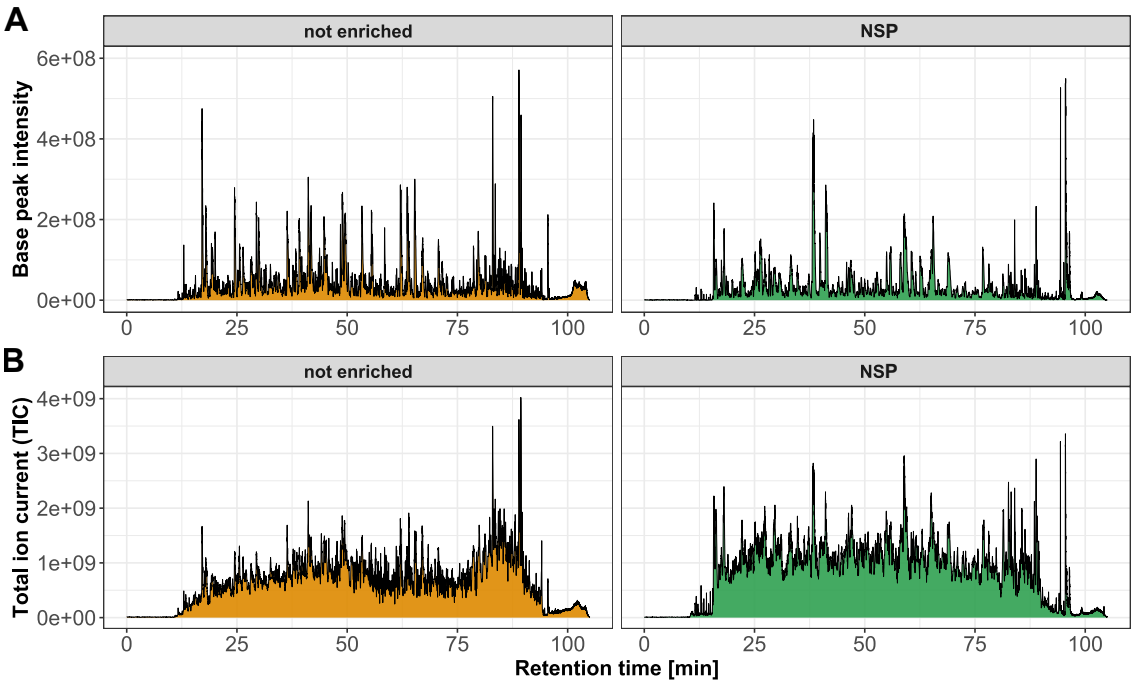


Figure S1: Representative chromatograms of NSP samples, without (left panels) and with (right panels) via click-chemistry based enrichment of NSP. LC-MS analyses was carried out with identical LC gradients and MS acquisition methods, which were optimized for the analysis of enriched samples. A) Base peak- and B) Total ion current (TIC) chromatograms. This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

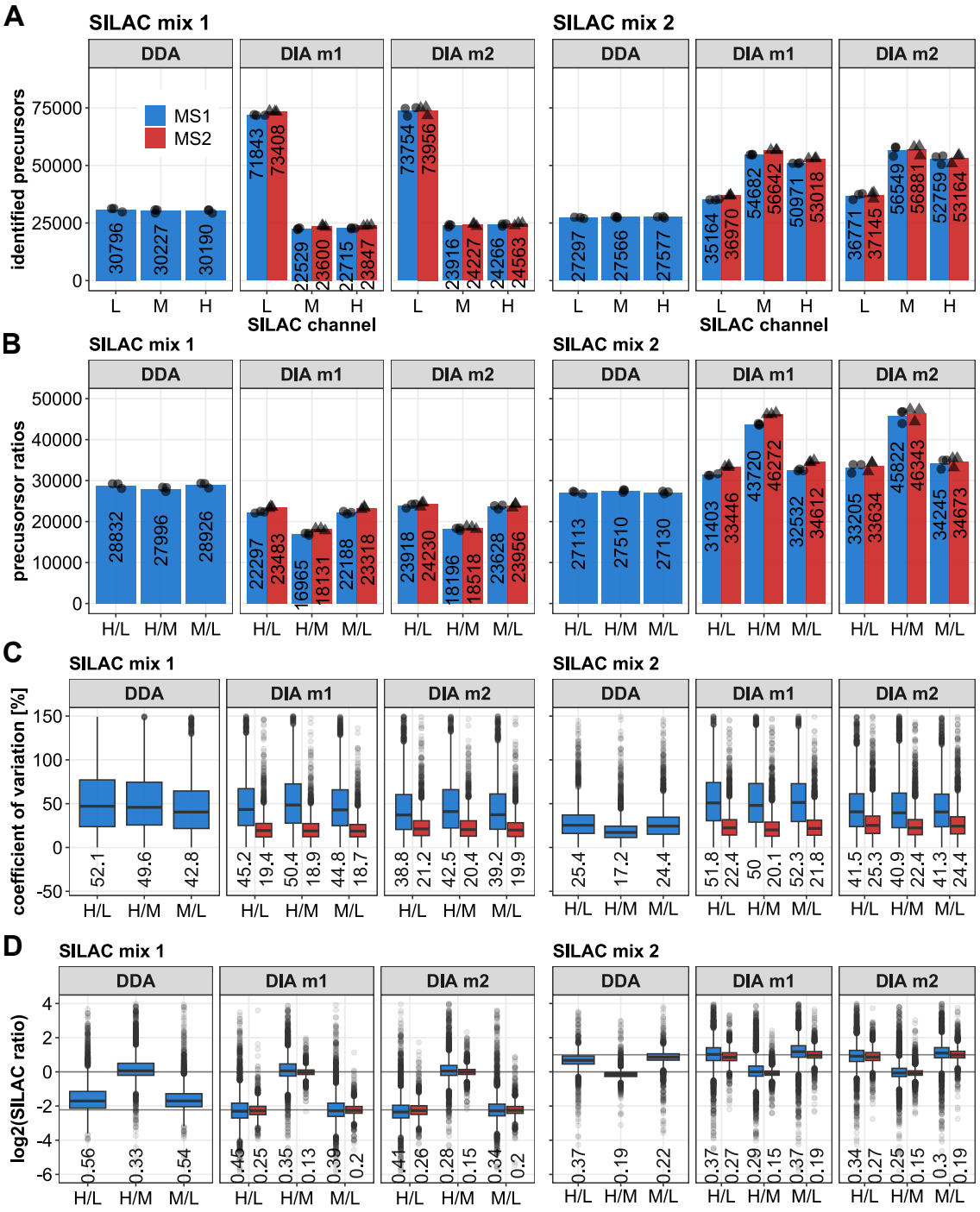


Figure S2: Comparative analysis of SILAC labelled benchmark samples using data-dependent acquisition (DDA) and data-independent acquisition (DIA) mass spectrometry. A) Number of identified precursors in the light- (L), intermediate- (M) and heavy (H) SILAC channels. B) Precursor SILAC ratio numbers. Values based on MS1-based quantification in blue and MS2-based quantification indicated in red. C) Coefficient of variation (CV) values of the precursor SILAC ratios. D) Box plots indicating the distribution of log2-transformed precursor SILAC ratios. This figure was taken from Borteçen et al., An integrated workflow for quantitative analysis of the newly synthesized proteome, Nature Communications, 14, 2286, 2023. [180]

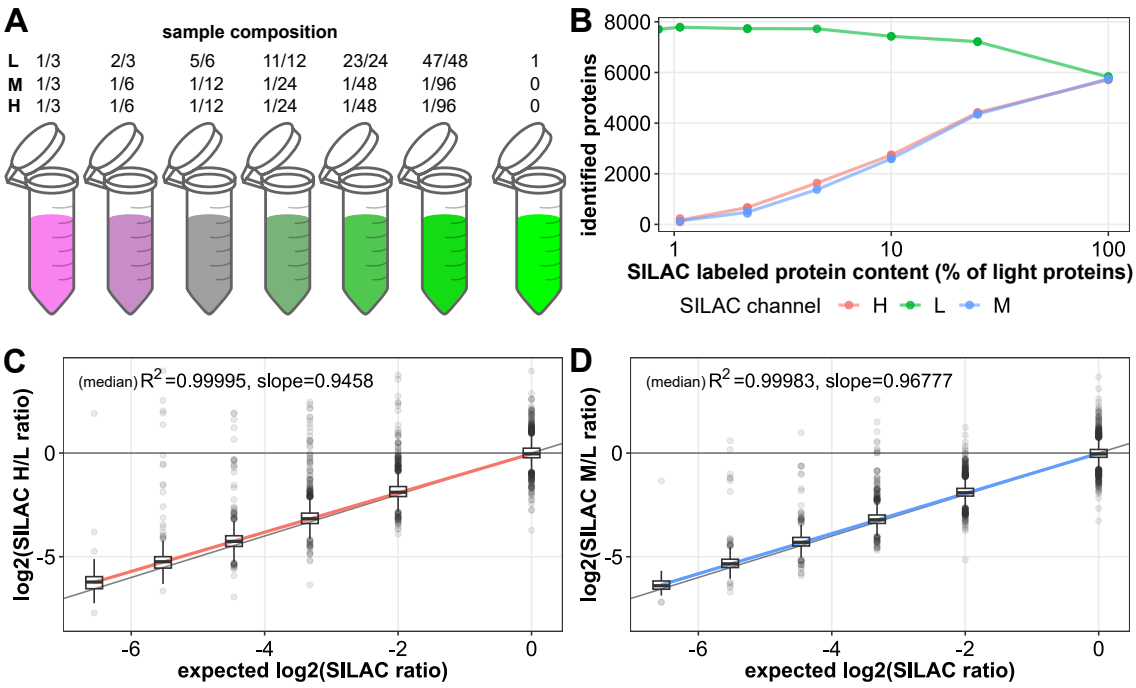


Figure S3: Assessing the linear range of quantification for plexDIA-based SILAC analysis. A) Schematic representation of the dilution of SILAC labeled Hela cell lysates wich were used for the creation of a seven-point calibration curve. B) Number of identified proteins in the light- (L), intermediate- (M) and heavy (H) SILAC channels. C) Linear correlation analysis between the experimentally determined and expected heavy over light SILAC ratios. The coefficient of determination (R^2) and slope of the regression line are indicated in the top left. D) Linear correlation analysis between the experimentally determined and expected intermediate over light SILAC ratios. The coefficient of determination (R^2) and slope of the regression line are indicated in the top left. Box plots indicating the distribution of log2-transformed SILAC ratios. The vector graphic of the conical test tube was created by Nicolàs Palacio-Escat (CC-BY 4.0) and modified for this figure using Inkscape.

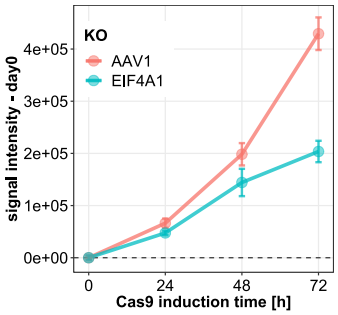


Figure S4: Cell viability of RKO cells lines with induced KO of AAV1 and eIF4A1, measured at multiple time points over 72 h. Cellular ATP content was measured as proxy for cell viability and cell numbers via a the commercial CellTiter-Glo 2.0 assay. The luminescent signal intensity of the cell seeded at the start of the experiment was subtracted from the measured signal at later time points.

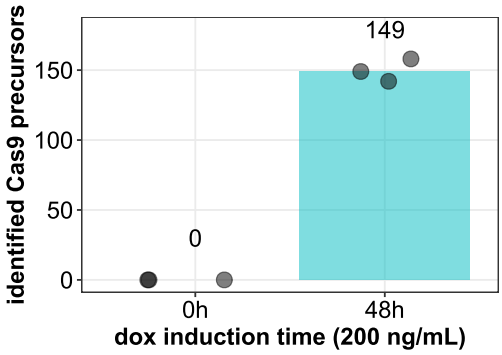


Figure S5: Number of spCas9 precursor identifications in RKO cells, expressing two sgRNAs targeting the AAV1 integration site (AAV1), with 48 h KO induction via 200 ng/mL doxycycline (dox) or no induction.

A

BATCH 1	1	2	3	4	5	6	7	8	9	10	11	12
A	eIF3C_R3	PDCD4_R2	eIF2S3_R2	eIF3C_R2	eIF3D_R1	eIF2D_R2	eIF2D_R3	eIF3B_R2				
B	eIF2B5_R2	eIF3M_R2	eIF4B_R2	eIF3D_R3	eIF3C_R1	eIF1B_R3	eIF3E_R2	eIF2S1_R1				
C	eIF3L_R3	eIF4E1B_R2	eIF1_R1	eIF1B_R1	eIF1_R2	eIF3E_R1	eIF3M_R3	eIF3B_R3				
D	eIF1B_R2	eIF2B5_R3	eIF4H_R1	test	eIF3M_R1	eIF4A1_R1	eIF2S1_R3	eIF3L_R1				
E	eIF4E1B_R1	eIF4B_R3	eIF3J_R3	eIF2B5_R1	eIF3L_R2	PDCD4_R3	eIF2S3_R3	eIF4E1B_R3				
F	eIF4H_R2	test	eIF4B_R1	eIF4H_R3	eIF4A1_R3	eIF2B3_R2	PDCD4_R1	eIF5A_R2				
G	eIF3J_R1	test	eIF3J_R2	test	eIF5A_R1	eIF2D_R1	eIF3B_R1	eIF2B3_R3				
H	eIF2B3_R1	eIF5A_R3	eIF2S3_R1	eIF4A1_R2	eIF3D_R2	eIF1_R3	eIF2S1_R2	eIF3E_R3				

BATCH 2	1	2	3	4	5	6	7	8	9	10	11	12
A	eIF2A_R3	eIF2B4_R3	eIF4G1_R2	eIF2B4_R1	eIF3H_R3	eIF4G1_R3	test	eIF3I_R2				
B	eIF2A_R2	eIF4E3_R2	eIF4E3_R1	eIF4E2_R2	eIF3F_R2	eIF3K_R1	eIF3H_R2	eIF2S2_R3				
C	eIF4E_R3	eIF2B2_R3	eIF4EBP1_R1	test	eIF4EBP1_R2	eIF3F_R3	eIF4A2_R2	eIF4A2_R3				
D	eIF3K_R2	eIF2B2_R1	eIF2B2_R2	eIF4A2_R1	eIF4G3_R2	eIF6_R3	eIF3G_R1	eIF2B4_R2				
E	eIF2B1_R2	eIF2B1_R3	eIF3A_R1	eIF3G_R2	eIF3I_R3	eIF3I_R1	eIF2B1_R1	eIF3A_R2				
F	eIF4G3_R3	eIF4EBP1_R3	eIF4G2_R2	eIF4E_R1	eIF3G_R3	eIF3A_R3	eIF4E_R2	eIF2S2_R1				
G	eIF2S2_R2	eIF3K_R3	eIF4G2_R3	eIF4G2_R1	eIF3H_R1	eIF6_R2	eIF6_R1	test				
H	eIF4G3_R1	eIF4G1_R1	eIF4E2_R3	eIF4E2_R1	test	eIF2A_R1	eIF3F_R1	eIF4E3_R3				

B

BATCH 1	1	2	3	4	5	6	7	8	9	10	11	12
A	eIF3B_R1	eIF3E_R2	eIF3C_R3	eIF2S3_R2	eIF4B_R3	ctrl_set1_R2	empty	ctrl_set1_R1	eIF2D_R3			
B	eIF1_R3	eIF2D_R1	eIF3J_R1	eIF2B3_R1	eIF3D_R3	PDCD4_R3	eIF2S1_R1	eIF2B5_R2	eIF3L_R1			
C	eIF3E_R3	eIF4H_R1	eIF5A_R3	eIF1B_R3	eIF1_R1	ctrl_set2_R2	ctrl_set2_R3	eIF3J_R2	eIF2B3_R2			
D	eIF2B3_R3	eIF1B_R1	eIF2S3_R1	eIF3L_R3	empty	eIF4B_R1	eIF4H_R3	eIF3D_R2	eIF3M_R2			
E	empty	eIF3C_R1	eIF4A1_R1	eIF2S1_R2	eIF4A1_R3	eIF1_R2	eIF4B_R2	eIF1B_R2	eIF5A_R1			
F	eIF4E1B_R1	eIF2D_R2	eIF5A_R2	eIF3D_R1	eIF4E1B_R3	eIF3J_R3	eIF2S1_R3	eIF2B5_R1	eIF4E1B_R2			
G	eIF4H_R2	eIF3B_R2	eIF4A1_R2	empty	empty	ctrl_set2_R1	eIF3M_R1	PDCD4_R2	ctrl_set1_R3			
H	eIF3B_R3	eIF2S3_R3	eIF3E_R1	eIF3M_R3	eIF2B5_R3	eIF3C_R2	eIF3L_R2	PDCD4_R1	empty			

BATCH 2	1	2	3	4	5	6	7	8	9	10	11	12
A	eIF4EBP1_R3	eIF4A2_R2	eIF4G2_R2	eIF4E_R2	eIF3G_R3	eIF4G1_R2	eIF3A_R3	eIF2A_R3	eIF2B2_R3			
B	eIF3G_R1	eIF3H_R3	eIF4G3_R1	eIF3K_R3	eIF3I_R1	eIF3I_R2	eIF2S2_R2	ctrl_set3_R2	eIF3F_R3			
C	eIF4E2_R2	ctrl_set4_R2	eIF2S2_R3	eIF2B4_R2	ctrl_set4_R1	eIF4EBP1_R1	eIF3F_R1	eIF3H_R2	eIF6_R2			
D	eIF2B2_R1	eIF4G3_R2	eIF2B1_R1	eIF6_R3	eIF4E3_R3	eIF4EBP1_R2	eIF2B1_R2	eIF4G1_R1	empty			
E	eIF2B4_R1	empty	eIF4E3_R2	eIF3A_R1	ctrl_set4_R3	eIF3G_R2	eIF3K_R1	eIF2B1_R3	eIF3I_R3			
F	eIF3A_R2	eIF6_R1	eIF4G3_R3	eIF4A2_R1	empty	eIF4G2_R1	eIF3H_R1	eIF2B2_R2	eIF4E2_R3			
G	eIF4G2_R3	empty	eIF2B4_R3	ctrl_set3_R3	eIF2S2_R1	eIF4E_R3	eIF4E3_R1	eIF4E2_R1	eIF2A_R2			
H	eIF2A_R1	eIF3F_R2	eIF4G1_R3	eIF4E_R1	ctrl_set3_R1	empty	eIF3K_R2	empty	eIF4A2_R3			

Figure S6: Schematic representation of the randomized plate positions for the batches of samples for the semi-automated enrichment of NSP (A) and automated SP3 sample preparation for proteomic sample preparation (B). The targeted eIF KO and experimental replicate numbers are indicated in each cell of the grid, which represents a position in the 96 well plate. Pooled AAV1 control KO samples of the 4 experimental batches are indicated as “ctrl_set(1-4)_R(1-3)”

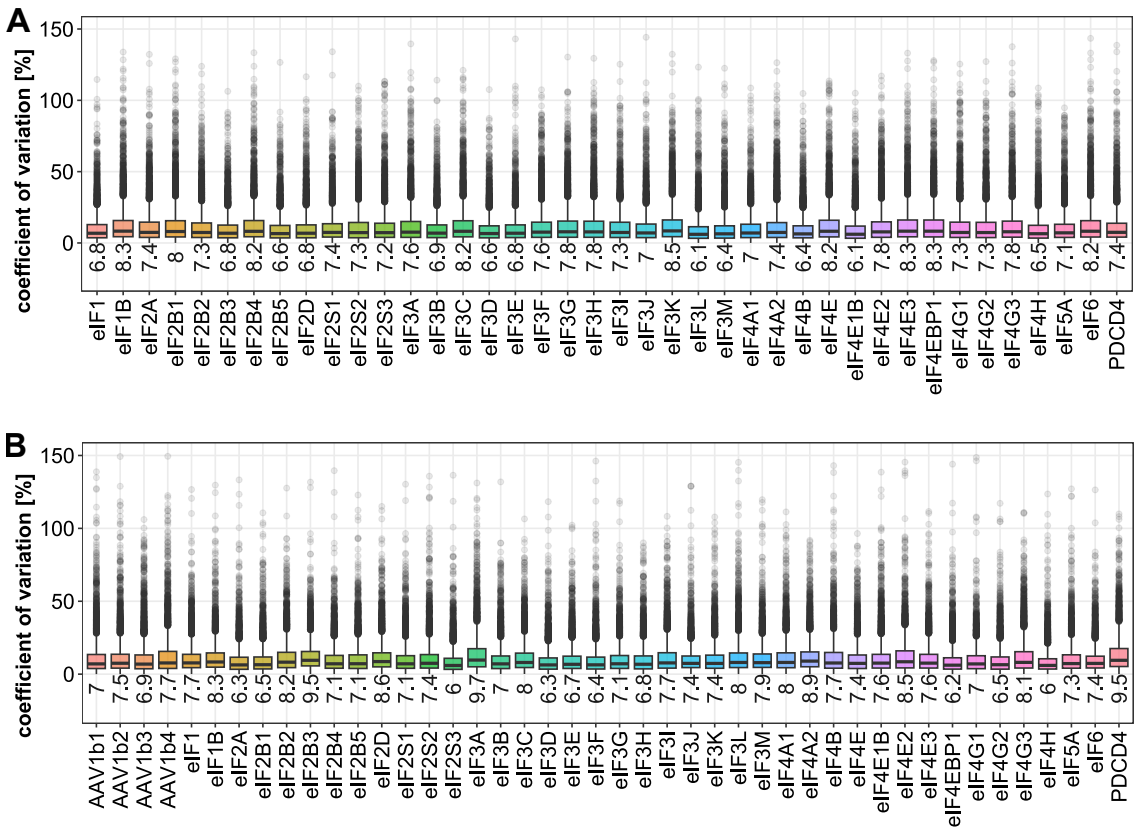


Figure S7: Quantitive precision of the newly snthesized proteomic samples (A) and proteomic samples (B), measured via the coefficient of variation (CV). For the newly synthesized proteome data, CV values were calculated from the heavy over intermediate SILAC ratios, and in the proteome data from LFQ normalized intensities.

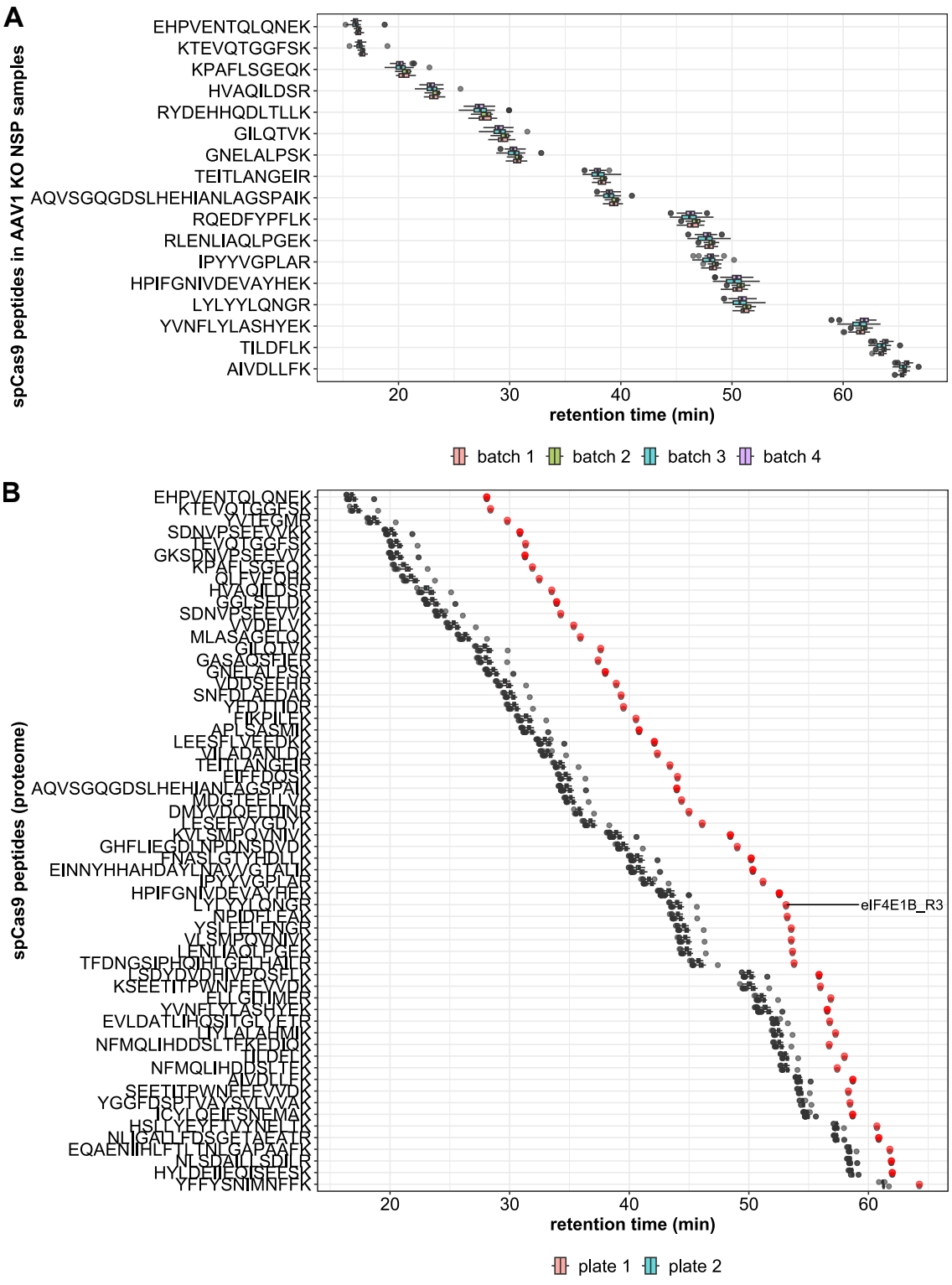


Figure S8: Retention times of spCas9 peptides, which were identified in all newly synthesized proteome or global proteome samples. A) Peptides, derived from spCas9, which were identified in all 120 AAV1 KO control samples (intermediate SILAC label) in the newly synthesized proteome analysis. The respective measurement batches of the samples are indicated with distinct colors. B) Peptides, derived from spCas9, which were identified in all 132 proteome samples. The respective batches of the autoSP3 sample preparaion and LC-MS measurements are indicated with distinct colors. The outlier sample “eIF4E1B_R3” is highlighted in red.

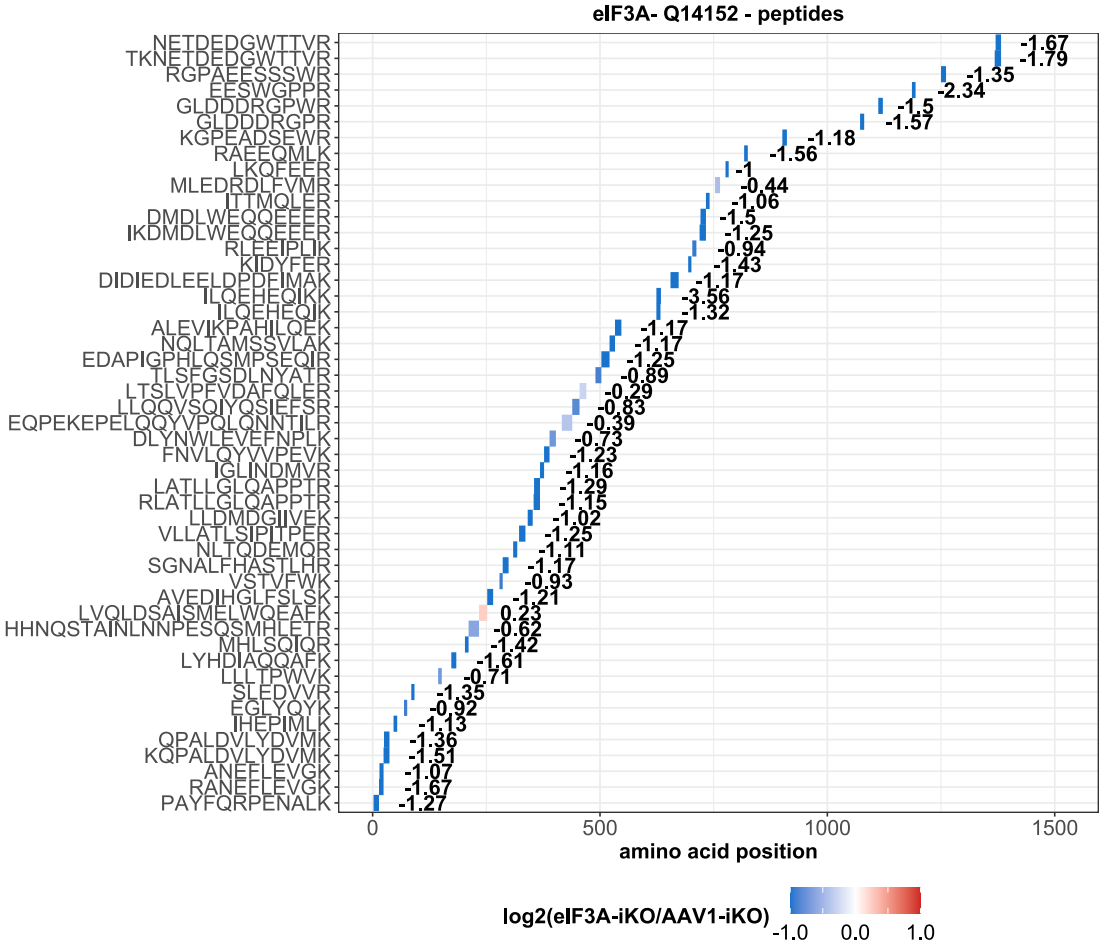


Figure S9: Log2 transformed fold change ratios of all identified eIF3A peptides, in RKO cells with induced KO of eIF3A vs. AAV1 KO controls. The position of the peptides in the full protein sequence is indicated on the x-axis and their color represents the log2 fold change values.

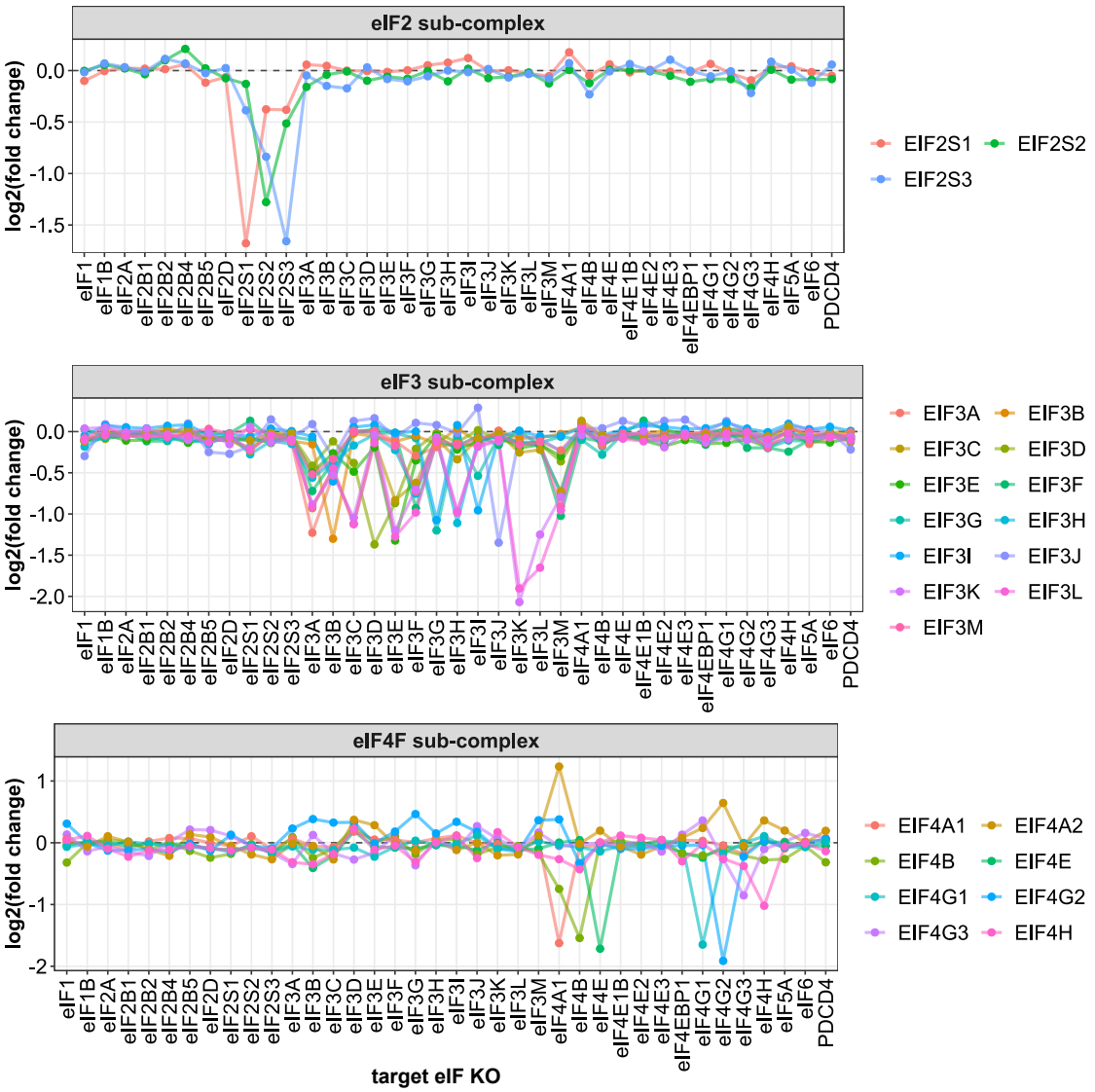


Figure S10: Protein abundance log2 fold change values in RKO cells with induced KO of the indicated eIF, and eIF-associated proteins. KO were induced via treatment with 200 ng/mL dox for 48 h.

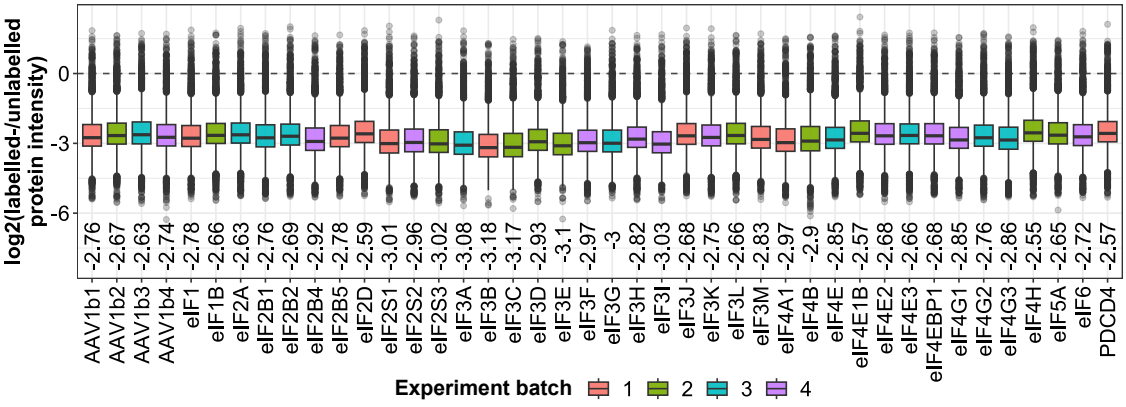


Figure S11: Box plots of intensity ratios of heavy- or intermediate labeled NSP over unlabeled (light) pre-existing proteins, measured in samples of RKO cells with induced KO of different eIF or AAV1 as control. The respective experimental batch number is indicated through different colors.

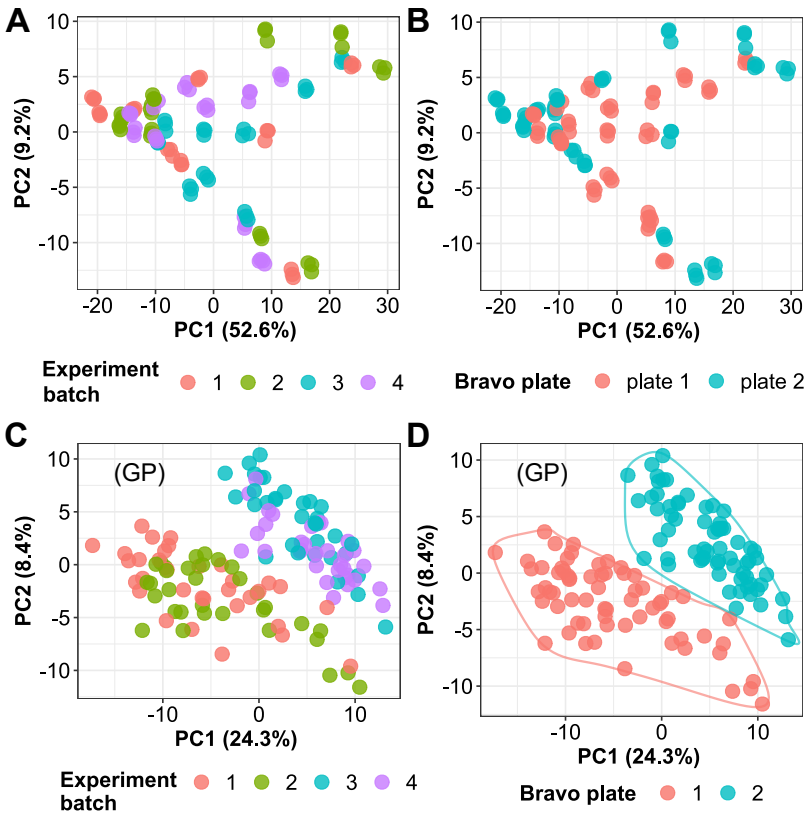


Figure S12: Principal component analysis (PCA) of the proteome and newly synthesized proteome data of RKO cells with induced depletion of distinct eIF. The percentage values of explained variance by the respective principal component (PC) are indicated on each axis. Highlighting of the different experimental batches (A) and semi-automated NSP enrichment batches (B) for the newly synthesized proteome data. Highlighting of the different experimental batches (C) and autoSP3 sample preparation batch (Bravo plate) (D) for the proteome data.

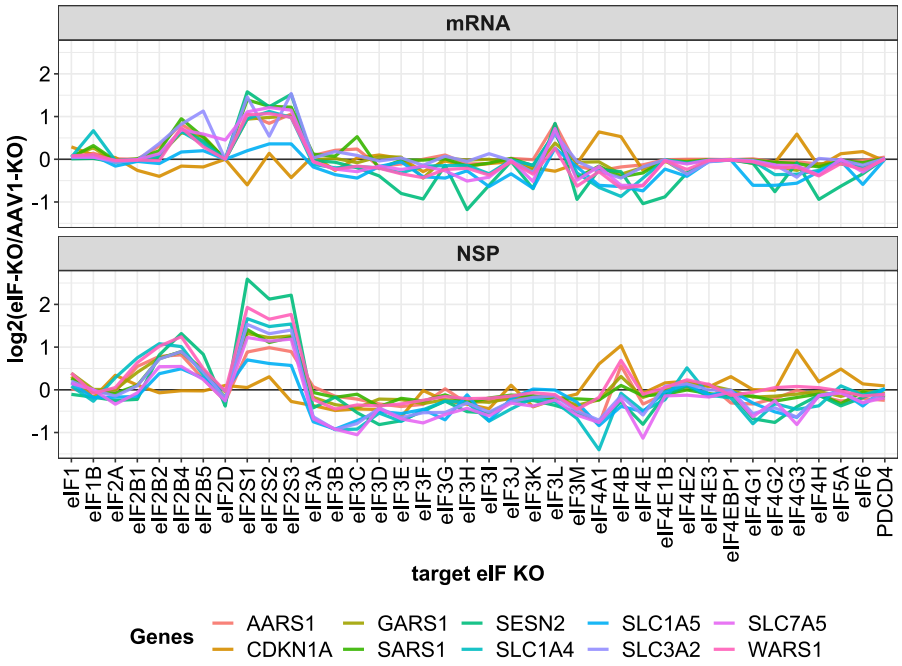


Figure S13: Examples of proteins with differential expression on both mRNA- and NSP levels. Log2 fold change values of selected mRNA and NSP across the panel of RKO cells with induced KO of the indicated eIF and eIF-associated proteins.

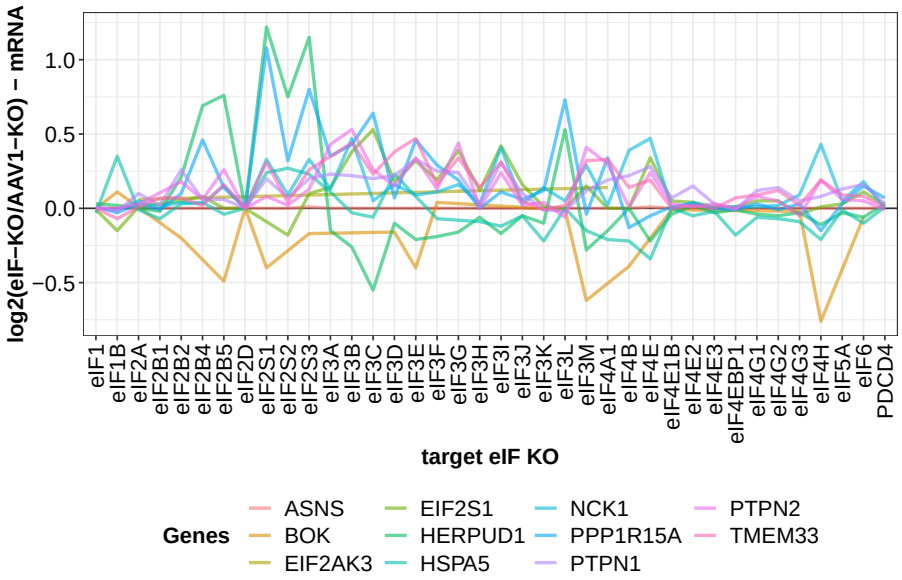


Figure S14: Measured log2 fold change values of mRNA, encoding proteins that are part of the “interated stress response signaling pathway” Go-term (GO:0140467), in RKO cells with induced KO of the indicated eIF and eIF-associated proteins.

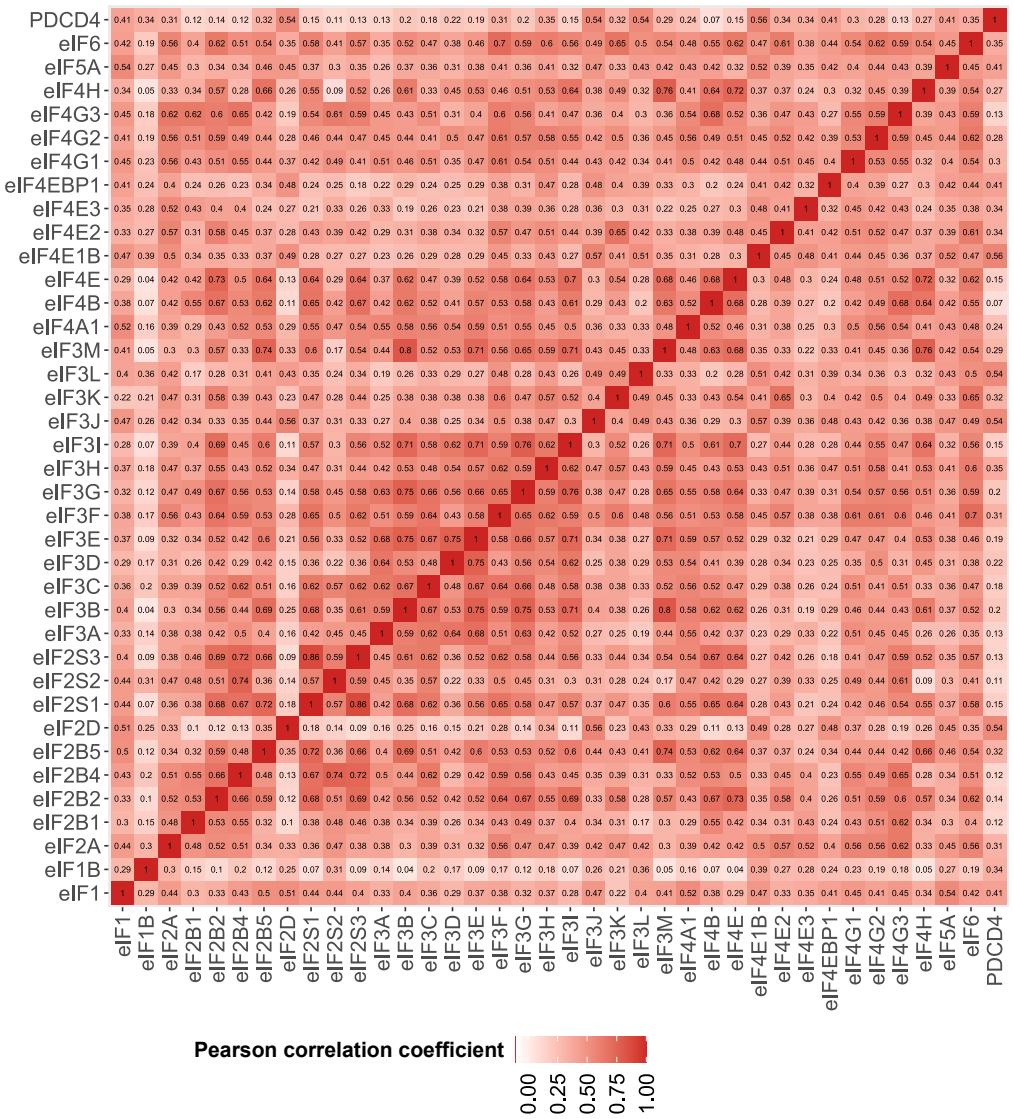


Figure S15: Heatmap of the Pearson correlation coefficient values between the mean log2-transformed translation efficiency (TE) fold changes of RKO cells with induced KO of the indicated eIF and eIF-associated proteins.

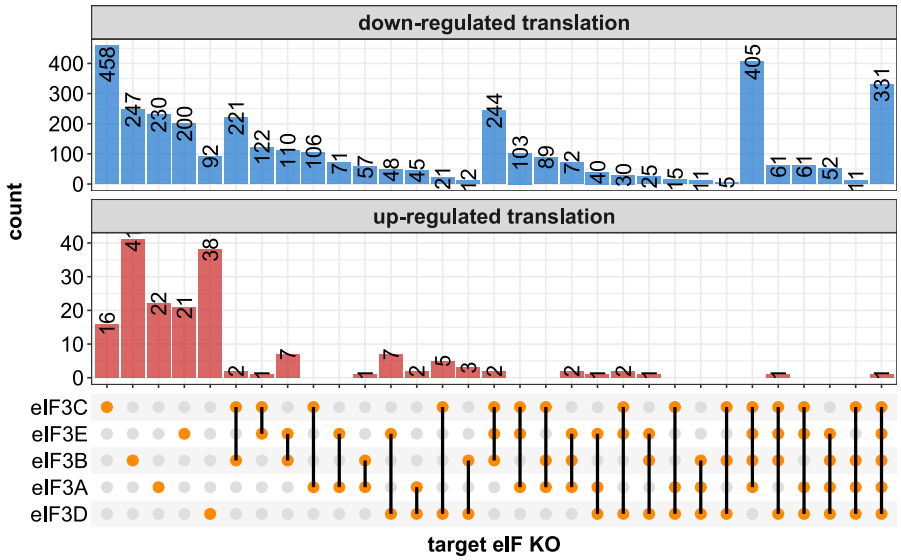


Figure S16: Upset plot of differentially translated mRNA ($|\log2 \text{ fold change}| > 0.585$ and adj. p-value < 0.05), in RKO cells with induced KO of eIF3A,-B,-C,-D or -E.

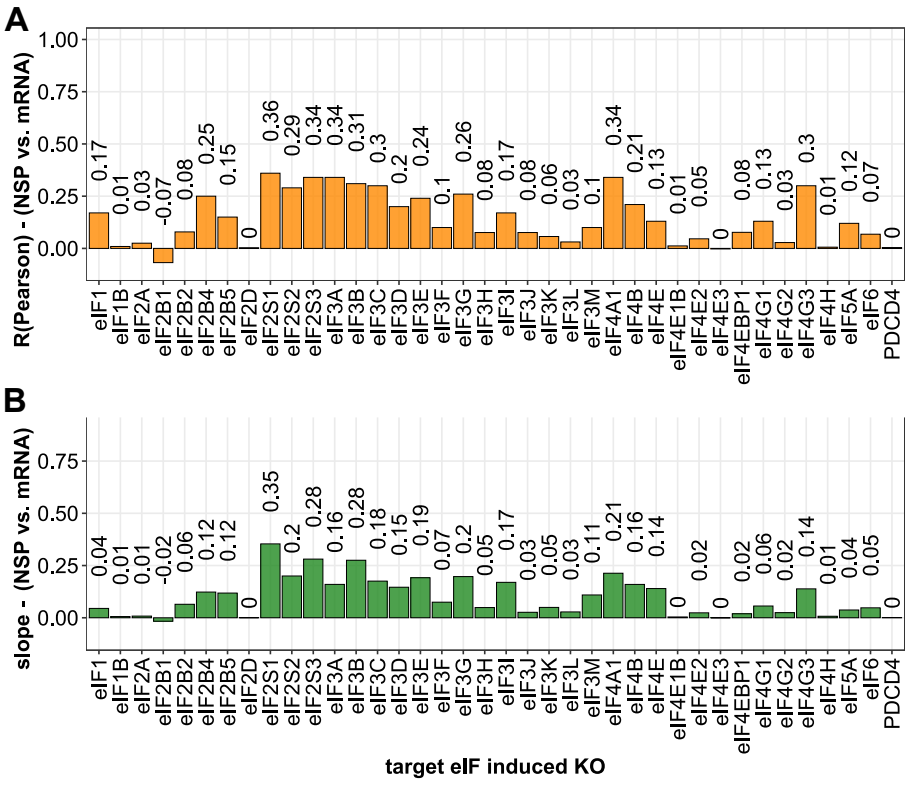


Figure S17: Metrics of linear correlation analysis between the log2 fold change values of mRNA and NSP, in response to depletion of individual eIF. A) Pearson correlation coefficient of mRNA and NSP log2-transformed fold change values. B) Slope of linear regression of mRNA and NSP log2 fold change values.

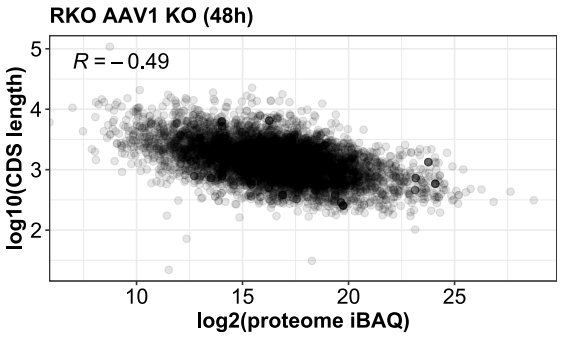


Figure S18: Scatter plot of the canonical mRNA coding sequence length and protein abundance (iBAQ) in RKO cells (expressing sgRNAs targeting the AAV1 integration site). The Pearson correlation coefficient (R) is indicated in the top left corner.

8.2 Supplementary tables

Table S1: Overview of sgRNA that were used for the induced KO of the targeted eIF.

target gene	sgRNA sequence
AAV1	GCTGTGCCCCGATGCACAC
AAV1	GCTTGGCAAACTCACTCTT
EIF1	GATCCCTTGGACAGTAGTAAGGG
EIF1	ATAGCTTACCGAAAGAGTGGAGG
EIF1B	GTAAGGGTGACGACTTACTCC
EIF1B	GGAGAGGTTATTCAGCTTCA
EIF2A	GTCCATGAGGGCTATAGT
EIF2A	GAGGGGACAAATGGAAGTGT
EIF2B1	GACTCACGCCTACTCCAGAG
EIF2B1	GCATCTAGCACCACAGTGACA
EIF2B2	GTGCTCAAGATTATCCGGG
EIF2B2	GGAGCTGATCCGCAGAGA
EIF2B3	GCATGAACTCAGTCACTGTGG
EIF2B3	GATTGGGCCAGAGACACAGAT
EIF2B4	GCCTAATGAAGAACTCTCCA
EIF2B4	GCATGGCTGGGTGGATCACAG
EIF2B5	GAGAGGGTAGACCCATCGG
EIF2B5	GCTAGAGGAAAATGTGCTCCT
EIF2D	GGATGGCAGCCACTGAGTAT
EIF2D	GGAAACCAGGTCCACCACCT
EIF2S1	TGAGTGTGTGGTTGTCATTAGGG
EIF2S1	TCAGATCCATTGCTGAAATGGGG
EIF2S2	GGAAGGGGATACCCAAACAG
EIF2S2	GCTGATGAAGAGGACACT
EIF2S3	GATGGCTACTATGCTGAA
EIF2S3	GTGGCCAGGACAGTCAACAA
EIF3A	GGATCTTCGCAAGAGCCACTTGG
EIF3A	ACTGCCTGTAAGACTCCCACAGG
EIF3B	GTGGACAATGTCCCTCAGGTGGG
EIF3B	TGGGAGATGTACTCAAAGATCGG
EIF3C	GCGCGAGTGCAGAGCATGG
EIF3C	GCTCCTGGGAAAAGCATAT
EIF3D	GCGCTTGATGCTCCGCAG
EIF3D	GATCCCAGGAATACACTGAGC
EIF3E	GCAGAATTGGGATGCAGCCATGG
EIF3E	AGAGAACATGGCTCATTCACTGG
EIF3F	GAGCATCAAAGCCTACGTC
EIF3F	GCCACGAGTACTACAGCCGAG
EIF3G	GACAGAGTACAAGATAGATG
EIF3G	GATCCCCTTGAGGAGCTCGC
EIF3H	GTGAAGCAAGTGCAGATAGATGG
EIF3H	GCTCCACCGCCGGCGCAGCAGGG

target gene	sgRNA sequence
EIF3I	GGTCCTTGGCCACAGTAAAG
EIF3I	GAAGCCGATCCTACTGCA
EIF3J	GGACGAGGACGAGGACGTCA
EIF3J	GCCTTCCCAGCGGTCCCCGC
EIF3K	GGTAAGTGATACCCACAACA
EIF3K	GAGATGCTCGGGGATCTGTC
EIF3L	GTTGGAGGTATACACAAG
EIF3L	GTAAGACTGCCAAGAAGTCAG
EIF3M	GAGTATTCTTATCCATCCCGTGG
EIF3M	TGGTTCCAGGATCAAGAGTAGGG
EIF4A1	TGCTAGGACCAAGGCCTGGGTGG
EIF4A1	GGAGCCCGAAGGCGTCATCGAGG
EIF4A2	GTAAGCATAGATGCCACGA
EIF4A2	GGAGACCCAAGCACTAGTAT
EIF4B	GAGCAGTGGGAAGGATGGAA
EIF4B	GCCTAGAAAAGCAGTGTAGGG
EIF4E	GGATGGTATTGAGCCTATGTGGG
EIF4E	AAAAGTGAGTAGTCACAGCCAGG
EIF4E1B	GGCGGGCGTGCTGCACGT
EIF4E1B	GACACAGCAGAGAGGTATGTG
EIF4E2	GGATTATTCGGCTGCGGAA
EIF4E2	GGTTGGGGAGGAGATCTGT
EIF4E3	GACTGCTTTAAAAGTTATGTG
EIF4E3	GAATGTAAATGCCTCTTTAGT
EIF4EBP1	GGGATCTGCCCACCATTCCGGGG
EIF4EBP1	AGCCCAGAAGATAAGCGGGCGGG
EIF4G1	GATCCAAACCAAGGAGGAA
EIF4G1	GGGGTGCAGCAGTTTCCCAC
EIF4G2	GCCTTGCGAGGAACCCAA
EIF4G2	GAAAATGGATAGGGACCCACT
EIF4G3	GTGGTAGGTGATGGAGGCAGAGG
EIF4G3	AGTCCAGAGAGTCCTAGAGGAGG
EIF4H	GAATCCATGTTGGATCCACGG
EIF4H	GTTCAGGGATGACTTCTTAGG
EIF5A	GGACAGCGGGGAGGTACGAGAGG
EIF5A	TGCATACATACAGGTCCATCTGG
EIF6	CGCTACCAGACAGTAGGTGTTGG
EIF6	GCTCTCCGATACCATCCCCGTGG
PDCD4	CAGTATCCTTAGCATTGGAGGGG
PDCD4	GGATACTCCTAGAGCACCACAGG

8.3 Publications

- Hisaoka M., Schott J., Borteçen T., Lindner D., Krijgsveld J., Stoecklin G. - “Preferential translation of p53 target genes” - **RNA Biology, 19, 437-452, 2022**

- Borteçen T., Müller T., Krijgsveld J. - “An integrated workflow for quantitative analysis of the newly synthesized proteome” - **Nature Communications, 14, 2286, 2023**

- Jiang X., Baig A.H., Palazzo G., Borteçen T., Groessl S., Zaal E.A., Amaya Ramirez C.C., Kowar A., Aviles-Huerta D., Berkers C.R., Palm W., Tschaharganeh D., Krijgsveld J., Loayza-Puch F. - “A p53-dependent program regulates hypusination of eIF5A to promote mitochondrial translation during cellular senescence” - **in revision**

- Hermann J., Borteçen T., Kalis R., Kowar A., Pechincha C., Vogt V., Schneider M., Helm D., Krijgsveld J., Loayza-Puch F., Zuber J., Palm W. - “mTORC1 cooperates with tRNA wobble modification to sustain the protein synthesis machinery” - **submitted**

- Borteçen T., Gaessler L., Kalis R., Zuber J., Krijgsveld J. - “A proteomics approach for the quantification of global protein synthesis inhibition” - **manuscript in preparation**

- Borteçen T., Kalis R., Dieterich C., Zuber J., Krijgsveld J. - “Investigating eIF-mediated preferential translation in a human cancer cell line” - **manuscript in preparation**

8.4 Acknowledgements

I am very privileged to have been given the opportunity to work on a number of exciting and fulfilling projects, and to receive constant support from Prof. Jeroen Krijgsveld and all members of his group. I therefore want to thank Jeroen and all former and current members of the B230 lab.

I am also grateful for the support, love and inspiration that I receive(d) from my wife and family.