DISS. ETH NO. 24044

DYNAMICS OF PROTEIN SYNTHESIS AND DEGRADATION IN FLG22-STIMULATED ARABIDOPSIS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

PHILLIP IHMOR

M.Sc. Biotechnology, Westfälische Wilhelms-Universität Münster

born on 17.06.1986

citizen of Germany

accepted on the recommendation of

Prof. Dr. Wilhelm Gruissem

Priv.-Doz. Dr. Katja Bärenfaller

Prof. Dr. Mark Robinson

Prof. Dr. Jeroen Krijgsveld

2017

TABLE OF CONTENT

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
AK	Aspartate kinase1 , AT5G13280
ANOVA	analysis of variance
bp	base pair
CDS	coding sequence
cpm	count per million mapped reads
DAPL	L,L-diaminopimelate aminotransferase, AT4G33680
DHDPS	Dihydrodipicolinate synthase, AT3G60880
DIA	data-independent acquisition
ESI	electrospray ionisation
ETI	effector-triggered immunity
FC	fold change
h	hours
HCD	higher-energy collisional dissociation
HILIC	Hydrophilic interaction chromatography
HR	hypersensitive response
LC	liquid chromatography
LE	labeling efficiency
LOESS	locally weighted scatterplot smoothing
Lys0	naturally occurring, unlabeled lysine
Lys4	medium-heavy $^2H_4$-labelled lysine
Lys8	heavy $^{13}C_6,^{15}N_2$-labelled lysine
MAD	median absolute deviation

MS	mass spectrometry or mass spectrometer
MS1	survey scan of a MS
MS2	fragmentation spectrum in a MS
nt	nucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen associated molecular pattern
PCD	programmed cell death
PCR	polymerase chain reaction
PTI	PAMP-triggered immunity
RIA	relative isotope abundance
Ribo-Seq	ribosome footprint profiling
RNA-Seq	mRNA sequencing
RPF	ribosome protected fragment
SILAC	stable isotope labeling with amino acids in cell culture
SRM	selected reaction monitoring
TAIR	the Arabidopsis information resource, http://www.arabidopsis.org/
TE	translation efficiency
TTC	2,3,5-triphenyl tetrazolium chloride
uORF	upstream open reading frame
UPS	ubiquitin 26S-proteasome system
UTR	untranslated region

1. ABSTRACT

Gene expression is a multi-step process that shapes the most central aspects of a cell including its physiology and metabolism. As such, it is governed by precise regulatory mechanisms so that the protein levels can respond swiftly and precisely to stimuli. The regulatory cascade includes transcription, translation and protein degradation, yet the relative contributions of these regulation levels are not well studied on a systems-wide level, especially for plants.

In this study, I quantified four aspects of gene expression on a time-resolved, genome-wide scale using RNA sequencing (RNA-Seq, transcription), ribosome footprint profiling (Ribo-Seq, translation), quantitative mass spectrometry with spike-in stable isotopic labelling with amino acids in cell culture (SILAC MS, protein abundance) and dynamic SILAC MS (protein degradation). All datasets were obtained from the same Arabidopsis suspension cell cultures after stimulation with the pathogen-associated molecular pattern (PAMP) flg22, the elicitor-active epitope of bacterial flagellin, or under control conditions.

In the process of obtaining these datasets, I developed an improved SILAC method for plant cell cultures, created an optimized Ribo-Seq protocol for plant samples, identified gene-specific differences in translational efficiency, showed that protein turnover is modelled more accurately with a logistic decay model compared to an exponential decay model, and determined the baseline half-lives for over 2500 Arabidopsis proteins. This is the largest plant protein turnover dataset obtained so far and the first one acquired with dynamic SILAC.

Statistical analyses of the data revealed that transcription showed the largest response to flg22 stimulation both in size of the fold-changes and the number of regulated genes. The translational response reflected the transcriptional response. In fact, I found only few examples with some regulation of translational efficiency upon flg22 treatment. Analysis of the quantitative proteomics data revealed that some protein abundances did not follow the transcriptional and translational changes. For the proteins that maintained constant protein levels despite transcriptional and translational upregulation, I could show that the protein degradation rates increased correspondingly, counteracting the raised protein synthesis. Downregulation of gene expression is determined by a more gradual reduction in transcription, which slowly affects the protein abundance. The speed in which the protein level responds to the transcriptional downregulation presumably depends on the turnover rate of the individual protein.

I conclude that most protein upregulation is determined by transcriptional upregulation and to a lesser degree by regulation of protein degradation. These results highlight the importance of further research to resolve the molecular mechanisms that govern the identified targeted protein degradation in defense response.

1.1 Zusammenfassung

Genexpression ist ein mehrstufiger Prozess, der viele Aspekte einer Zelle reguliert wie deren Physiologie und Metabolismus. Es ist daher für Zellen lebensnotwendig die Genexpression durch präzise und effiziente Regulationsmechanismen zu kontrollieren, damit Protein-Level schnell und präzise auf Stimuli angepasst werden können. Die Regulationsmechanismen der Transkription, Translation und Proteindegradation sind wohl bekannt, aber der relative Einfluss dieser Mechanismen auf das Gesamtsystem ist insbesondere in Pflanzen nicht gut untersucht.

In dieser Studie habe ich vier Aspekte der Genexpression systemweit und zeitsensitiv mittels RNA-Sequenzierung (RNA-Seq, Transkription), ribosome footprint profiling (Ribo-Seq, Translation), quantitativer Massenspektrometrie von Isotopen-markierten Proteinen (SILAC MS, Protein-Abundanz) und dynamischen SILAC MS (Protein-Degradierung) analysiert. Alle Datensätze wurden von der gleichen Arabidopsis Suspensions-Zellkultur unter Kontrollbedingung oder nach Behandlung mit einem Stimulus erhoben. Die Stimulierung der Zellkultur erfolgte durch Zugabe von dem Pathogen-assoziierten molekularen Muster (PAMP) flg22, dem Elicitor-aktiven Epitop bakteriellen Flagellins.

Im Rahmen dieser Untersuchungen habe ich ein robustes plant SILAC Protokoll entwickelt, die Ribo-Seq Methodik für pflanzliche Proben optimiert, habe Gen-spezifische Unterschiede in der Translations-Effizienz identifiziert, habe gezeigt, dass Protein-Degradierungskurven mit einem logistischen Degradationsmodell akkurater beschrieben werden können als mit einem exponentiellen, und habe die Halbwertszeit von mehr als 2500 Arabidopsis Proteinen unter Standardbedingungen bestimmt. Dies stellt den aktuell größten pflanzlichen Proteinumsatz-Datensatz dar und den Ersten, der mit dynamischem SILAC gemessen wurde.

Die statistische Analyse der vier Datensätze zeigt auf, dass die größte Reaktion auf flg22-Stimulation - sowohl in Bezug auf die Anzahl respondierender Gene als auch die Stärke der induzierten Veränderungen - auf Ebene der Transkription erfolgt. Flg22-induzierte Veränderungen der Translation folgen grundsätzlich der Transkriptions-Reaktion, tatsächlich konnten nur wenige Fälle von regulierter Translationseffizienz gefunden werden. Die Analyse der Protein-Level zeigt, dass viele Proteine nicht der transkriptionellen und translationalen Regulation folgen. Für Proteine, die trotz Induktion der Transkription und Translation unveränderte Abundanzen aufzeigen, konnte gezeigt werden, dass die Protein-Degradierung entsprechend verstärkt wurde um der gesteigerten Protein-Produktion entgegenzuwirken. Proteine, deren Abundanz durch flg22-Stimulation reduziert wurde, zeigten eine graduelle Reduktion der Transkription auf. Die Geschwindigkeit, mit der die

veränderten Transkriptions-Level die Protein-Abundanz beeinflussen, scheint von der Halbwertszeit der jeweiligen Proteine abzuhängen.

Zusammengefasst wird die Anpassung der Genexpression nach Stimulation am stärksten durch transkriptionelle Veränderungen bestimmt und zu einem geringen Anteil durch Veränderungen der Protein-Degradierung. Diese Ergebnisse weisen auf die Notwendigkeit zukünftiger Studien hin, um die molekularen Mechanismen der hier identifizierten gezielten Protein-Degradierung aufzudecken.

2. INTRODUCTION

2.1 Gene expression

Gene expression describes the process of transferring genetic information (genotype) to a physiological representation (phenotype), i.e. the production of the respective gene products. Consequently, gene expression regulation refers to regulating the abundance of gene products, which in the majority of cases are proteins. The central dogma of molecular biology describes the information flow from the biomolecules involved in gene expression: The DNA-encoded sequence information is passed on to RNAs and then to proteins (Crick & others, 1970). Yet while the flow of information and the respective encoding are clearly defined, there is no direct relationship determining the concentrations of these molecules, as each of the molecules occurs in different copy numbers, as each of the biosynthesis steps can be separately regulated, and as these molecules are not only produced, but also constantly degraded.

The lack of a clear correlation between RNAs and proteins indicates the high complexity of gene expression regulation, which can take place on the level of DNA (chromatin modification, genetic variation), RNA (transcriptional regulation, degradation, splicing, processing, RNA interference, and sequestration), and protein (degradation, secretion, subcellular localization and post-translational modifications (PTMs)) (Payne, 2015). The relative extent by which these regulatory mechanisms influence protein expression is not well understood and demands a new type of research, where the quantities of RNAs and proteins and their changes over time are compared on a systems-wide level (Payne, 2015). This defines one hallmark of the modern field of science termed systems biology.

With the advent of high-throughput technologies for transcriptomic and proteomic analyses, information on the influence of the different gene expression regulatory mechanisms became accessible. Absolute mRNA and protein concentrations across various species, tissues and time points grown at steady state correlate only moderately with correlation coefficients between 0.4 and 0.6 (Baerenfaller *et al*, 2008; de Sousa Abreu *et al*, 2009; Koch *et al*, 2014). This led to the conclusion that protein abundances are only partly determined by transcriptional regulation and that regulation of translation and protein degradation may be equally important in determining protein levels (Vogel *et al*, 2010; Schwanhäusser *et al*, 2011). However, some of the interpretations in these studies seem to be arguable and it was suggested that transcriptional regulation might account for up to 81% of the overall variation in protein expression (Li *et al*, 2014b).

These analyses and interpretations were based on biological systems at steady-state. Alternatively, the contributions of the different gene expression regulation mechanisms can be studied by stimulating a steady-state biological system and determining how much of the measured proteomic variation can be observed on the different regulatory levels (Liu *et al*, 2016). Using this approach, transcription was deemed the most prevalent regulatory mechanism, followed by translational regulation and only to a small extent by protein degradation (Kristensen *et al*, 2013; Jovanovic *et al*, 2015). However, studies that acquire such genome-wide datasets from the same biological system are still rare (Kuersten *et al*, 2013; McManus *et al*, 2015), especially for plant systems.

Recently, these considerations and more were excellently reviewed by Liu *et al*, (2016).

2.2   Protein synthesis

Regulation of gene expression manifests as changes in protein synthesis or protein degradation. Protein synthesis is determined by the number of mRNAs per gene and their rate of translation.

2.2.1   Transcription

The first step of gene expression is transcription, where the genetic information on the DNA is copied to mRNAs by RNA polymerases. After splicing, capping and polyadenylation, the mature mRNAs are transported from the nucleus into the cytosol where they are translated by ribosomes.

Transcriptional analyses have made major progress in recent years with the advent of next-generation sequencing technologies, resulting in unprecedented precision and coverage of the transcriptional landscape (Metzker, 2010). In an mRNA sequencing approach, usually abbreviated as RNA-Seq, the mRNA is extracted from the biological sample, depleted from rRNAs either by enrichment of poly-adenylated transcripts by oligo-dT capture or by negative hybridization with anti-rRNA oligomers. rRNA-depletion allows also the detection of regulatory, non-protein coding RNA species that are not polyadenylated. The enriched RNAs are then fragmented, transcribed to cDNA, ligated to barcoded adapters, amplified by polymerase chain reaction (PCR) and sequenced by various chemical approaches. Third generation sequencers like the Oxford nanopore minION (Eisenstein, 2012), are characterized by detecting single molecules and therefore can omit PCR amplification (Mikheyev & Tin, 2014). Combined with new approaches of separating individual cells to perform cell-specific cDNA reverse-transcription and barcoding (Masoko et al., 2015), transcriptional profiles can now be obtained at single-cell resolution, which allows deciphering transcriptional networks (Shalek *et al*, 2013, 2014).

For transcriptomics, the main benefits of using sequencing-based technologies compared to earlier microarray methodologies are that they can 1) resolve RNA-isoforms resulting from differential splicing, 2) allow absolute quantitation if standards are mixed into the sample and 3) massively increases the coverage and the accuracy of quantitation (Liu *et al*, 2016).

2.2.2 Translation

Translation describes the process in which ribosomes produce proteins and polypeptides based on the information encoded on mRNAs.

The translation process can be separated into three steps. 1) In the initiation step, one of the ribosome subunits with its associated proteins binds to the mRNA and scans along it to find a START codon, where the translationally competent ribosome is fully assembled. 2) During elongation, the ribosome translocates along the mRNA in steps of 3 nucleotides, translating the genetic information from the mRNA into an elongating polypeptide chain. 3) Translation is terminated when the translation machinery reaches a STOP codon on the mRNA (usually UAA, UAG or UGA), where the ribosome-mRNA complex is disassembled. Therefore, only the mRNA sequence between the START and STOP codon is actually translated and is defined as open reading frame (ORF), but not the 5' or 3' untranslated regions (UTRs) of the mRNA sequence. In eukaryotes, 5' UTRs can contain upstream open reading frames (uORFS) that contain a START codon, where translation can start, which usually affects translational efficiency of the main ORF (Morris & Geballe, 2000; Gaba *et al*, 2001; Meijer & Thomas, 2002; Kim *et al*, 2007; Von Arnim *et al*, 2014). For Arabidopsis, it was found that around 30% of all mRNAs contain uORFS in their 5'UTRs (Kim *et al*, 2007).

In a process called leaky scanning, it is also possible for ribosomes to bypass an AUG start codon and to start translation at a later start codon (Herzog *et al*, 1995). This can give rise to a multitude of differentially translated proteins.

2.2.2.1 *Quantifying translation*

Three techniques are primarily employed to measure the gene-specific translational activity: Polysomal profiling, translating ribosome affinity purification (TRAP), and ribosome footprint profiling (reviewed in King & Gerber, 2016). Polysomal profiling separates mRNA-ribosome complexes by ultracentrifugation on a sucrose density gradient. Highly translated mRNAs have multiple ribosomes attached to them, which forms a complex called polyribosomes or polysomes. The complex has a high density so that it sediments during ultracentrifugation in a sucrose density gradient. This separates the highly translated from the lowly translated mRNAs with less than two ribosomes that have a lower density. The lowly and highly translated parts of the transcriptome can then be compared by high throughput sequencing or microarray hybridization. The enrichment ratio in the polysomal fraction can be compared between treated and untreated biological systems to determine translational regulation of expression. This method is well established, also for plants (Loraine, 2009), can be applied to multiple organisms and relies on standardized sequencing methods.

In plants, various stimuli were analyzed for translational regulation effects, including dehydration stress (Kawaguchi *et al*, 2004), cold stress (Juntawong *et al*, 2013), light regulation (Juntawong & Bailey-Serres, 2012), oxygen deprivation / hypoxia (Branco-Price *et al*, 2005), or sucrose levels (Gamm *et al*, 2014). Many of these studies detected large global changes in the actively translated part of the transcriptome and determined gene-specific changes in translation rates of up to 100 fold. The mechanism of this regulation could partially be described by the activity of oligouridylate binding protein 1 (UBP1) which selectively sequesters mRNAs when stimulated by hypoxia (Sorenson & Bailey-Serres, 2014).

Alternatively, building upon an approach in yeast (Inada *et al*, 2002), Zanetti *et al*, (2005) developed an approach called TRAP (translating ribosome affinity purification) in which the ribosomal protein RPL18 is fused with a His6-FLAG dual epitope tag and overexpressed in Arabidopsis. The FLAG-tag is used to precipitate the ribosomes and all bound RNAs, which are then quantified by microarray hybridization. Mustroph *et al*, (2009b) expanded on that approach by using a tissue specific promoter for the expression of the FLAG-tagged RPL18 to determine the translation profiles on a tissue-specific level. The enrichment of the TRAP method is comparable to the polysome profiling (Mustroph *et al*, 2009), and is also compatible with modern sequencing methodologies (Reynoso *et al*, 2015).

Both ribosome profiling and TRAP will give a quantitative measure of the amount of mRNA present in polysomes or associated with ribosomes. However, it will not reveal information on translation efficiency, i.e. the number of ribosomes associated with a given mRNA, as polysomes can have varying numbers of more than two ribosomes attached. Furthermore, no information on the localization of the translating ribosomes on the respective mRNAs will be obtained.

2.2.2.2 Ribo-Seq

In 2009, Ingolia *et al*, reported a new method to query translational activity, which they named ribosome footprint profiling, now often termed Ribo-Seq. Similar in concept to DNase footprinting, this method uses endonucleases to digest all nucleic acids that are not bound by a protein. In the case of Ribo-Seq, an RNase is used to digest unbound mRNAs while the translating ribosomes block the RNase, resulting in ribosome "footprints" on the mRNA (Figure 2.1). In detail, cycloheximide is added to the cells beforehand to stall the translating ribosomes on the mRNAs. Then, the cells are lysed and RNase is added to digest all mRNA sequences that are not bound by a ribosome. The ribosome-mRNA-complexes are then enriched by sucrose cushion ultracentrifugation. The mRNA fragments are purified from the proteins and depleted of rRNAs by subtractive hybridization (e.g. by using the Ribo-Zero kit, Illumina). Oligo-dT enrichments cannot be used as these fragments do not contain poly-A tails. The rRNA-depleted mRNA fragments are then separated by denaturing polyacrylamide gel electrophoresis (PAGE) and the ribosome-protected fragments (RPFs) in the area of 25-35 nt are excised. Then adapters are ligated to the RNA fragments, the ligation products are circularized, amplified by PCR and finally sequenced. The resulting reads are then mapped onto the organism's transcriptome, where successfully mapped, true RPFs display a length of 25 to 32 nt. The amount of reads that map to the ORF of a gene is indicative for the abundance of the corresponding mRNA and the number of bound ribosomes, and can thus be used as a quantitative measure of the translation rate. True Ribo-Seq reads should map only to translated areas of the mRNA, i.e. not the 5' and 3' UTRs.

The strength of this approach is that the position of a ribosome on the mRNA can be determined with single-codon resolution. This allows in-depth analyses of translational regulation events, such as the effect of untranslated regions (UTRs), sequence-specific translation speeds and identification of alternatively spliced transcripts, to predict open reading frames (ORFs) and alternative translation start sites, or events of ribosome stalling. Hu et al, (2016) used Ribo-Seq data for instance to identify 5360 potentially translated uORFs in 2051 transcripts of the Arabidopsis genome. Using harringtonin, lactimidomycin or puromycin, which stall initiating ribosomes instead of elongating ribosomes, exact translation start sites can be determined (Ingolia, 2016). A key metric for the

annotation of ORFs is the periodicity of the mapped RPFs, as true ORFs should follow the 3 nucleotide periodicity of the ribosomal movement following the triplet code (Ingolia *et al*, 2014). Ribo-Seq data can also be used to identify actively translated short or alternative open reading frames (sORFs and altORFs), which could have regulatory effects on gene expression. However, experimental evidence and functional characterizations of the resulting polypeptides are still rare (Mackowiak *et al*, 2015; Landry *et al*, 2015; Tavormina *et al*, 2015; Lauressergues *et al*, 2015). Similarly, it could be shown that the number of mass spectrometrically identified proteins can be increased by 2.5% when the Ribo-Seq derived translatome is used instead of the transcriptome as the reference database to assign the MS spectra to peptide sequences (Koch *et al*, 2014). The ribosome profiling approach can also be applied on TRAP-isolated ribosome-mRNA complexes (Juntawong *et al*, 2014), which allows to resolve tissue-specific translational changes.

For our purposes, however, the most interesting feature is the dynamic range of Ribo-Seq in quantifying translational activity, as each ribosome bound to an mRNA is quantified independently. It can therefore be distinguished whether an mRNA is bound by two or by many more ribosomes and thereby allows to assess the translational rates more accurately than the polysome fractionation approach mentioned before. To calculate translation rates or efficiencies from Ribo-Seq data, the Ribo-Seq counts have to be compared to the mRNA-Seq counts derived from the same sample (Ingolia *et al*, 2009; Larsson *et al*, 2010; Xiao *et al*, 2016).

Ribo-Seq was first developed in yeast (Ingolia *et al*, 2009) and was then employed to various other biological systems, including mice (Guo *et al*, 2010), mammalian embryonic stem cells (Ingolia *et al*, 2011), and plants, which are discussed in more detail below.

Juntawong *et al,* (2013a) enriched ribosome-mRNA complexes by ultracentrifugation in a 60 % (w/v) sucrose cushion. The pellet was taken up in an RNase digestion compatible buffer and the mRNAs were digested. Then, a second, differential ultracentrifugation in a 0 to 40 % (w/v) sucrose gradient was performed to extract the monosomal ribosome-mRNA complexes. This is an adaptation from their frequently applied polysome profiling protocol (Loraine, 2009). They compared these quantifications with Ribo-Seq data from TRAP-isolated ribosome-mRNA complexes. Biologically, they analyzed the effect of hypoxia on Arabidopsis seedlings and saw a global decline in translation initiation, consistent with earlier polysome profiling and TRAP studies (Branco-Price *et al*, 2005; Mustroph *et al*, 2009).

Liu *et al,* (2013) analyzed translational regulation in etiolated Arabidopsis seedlings after exposure to light for 4 h. By using a fold-change cut-off of greater than 2 on the translation efficiencies they identified 294 downregulated and 519 upregulated genes. They report that uORFs with an alternative CUG START codon have a stronger translation repression effect than AUG initiated uORFs. Furthermore, they reported that light exposure reduced the number of RPFs that bind to the 5' and 3' UTRs of the mRNAs. As translating ribosomes are not expected to bind in 3'UTR, the identified 3'UTR-binding RPFs can either be indicative for alternative splicing events that are currently not covered in the genome annotation or represent sequencing or mapping artifacts. Experimentally, Liu *et al,* (2013) followed the protocol for mouse embryonic stem cells (Ingolia *et al*, 2011) but scaled it up from 100 µL cell lysate in the original protocol to 14 mL.

Merchante et al, (2015) studied the effect of ethylene on the translational regulation in Arabidopsis seedlings and identified several genes with differential translational efficiencies. These genes included EBF1 and EBF2, which encode F-box proteins involved in the degradation of the key transcriptional regulators EIN3/EIL1. Experimentally, they used the lysis buffer developed for plant polysomal profiling followed by the Ribo-Seq protocol developed for yeast cells (Ingolia *et al*, 2009). They report only a low correlation of $R^2 = 0.22$ between Ribo-Seq and RNA-Seq fold-changes.

Lei *et al*, (2015) assessed the global translational response in maize seedlings upon drought stress. They found that photosynthesis genes have an increased translational efficiency and detected global changes in translational efficiency induced by the drought stress of over 3 orders of magnitude. Experimentally, they used the same method as Juntawong et al, (2013a).

Chotewutmontri *et al*, (2016) analyzed the translational changes of maize leaves over four stages of development, focusing especially on chloroplast-encoded genes. They determined 2- to 4fold changes in translational efficiency over the developmental stages. Experimentally, they did not enrich the ribosomes-mRNA complexes before RNase digestion and therefore used high amounts of RNase I (3500 U) in a high amount of lysate (2.5 mL).

Chung *et al*, (2015) performed Ribo-Seq on Chlamydomonas and mouse samples. The focus of the paper is on the implementation of sequence-unspecific rRNA depletion approaches using DNA-duplex-specific nucleases. Experimentally, they employed the approach developed for yeast cells (Ingolia *et al*, 2009), but had to take 40 times more RNase I for the Chlamydomonas preparations than for the murine samples.
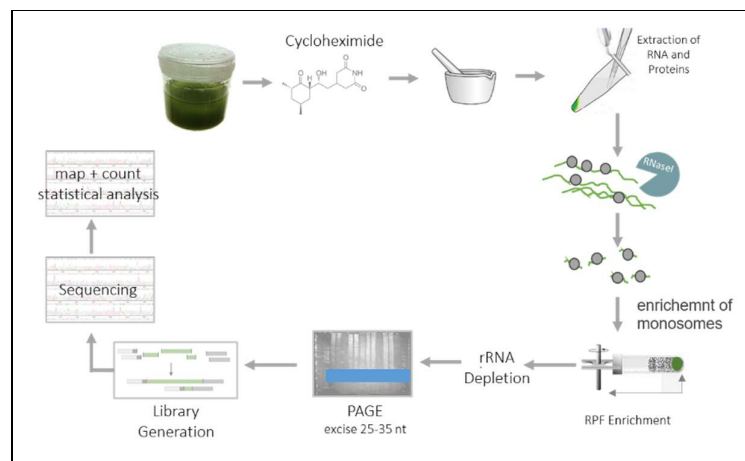

Figure 2.1: Schema of the Ribo-Seq procedure.

2.3 Proteomics

The centerpiece of the gene expression cascade is the protein level. Quantifying proteins however, requires substantial technology prowess, especially if the analysis is to be performed on a system-wide scale. Various approaches have been developed depending on organism and sample types, all with their own benefits and shortcomings. In the following, I will give an overview of the current state of quantitative high-throughput proteomics and explain the analytical approach we have chosen.

2.3.1 Mass spectrometric proteomics

There are various approaches to quantify proteins biochemically, including microscopy, cell cytometry, and immunoblotting. But only mass spectrometric (MS) approaches can quantify proteins in high throughput at a near genome-wide scale. MS ionizes the analytes and separates the ions according to their mass to charge value (m/z). As intact proteins don't ionize easily, typical bottom-up MS-based proteomics approaches digest the protein extraction by an endoprotease such as trypsin. For a typical eukaryote, this results in a mixtures that can exceed 1 million unique peptides, so that a separation by liquid chromatography before ionization is advised. This approach is called LC-MS, where the liquid eluting from the LC is ionized constantly, typically by means of electrospray ionization (ESI). The ionized particles including the peptide ions are then transported with ion optics into the MS and their mass to charge values are determined.

Typically, an MS frequently obtains MS1 survey scans that detect all inflowing ions. Ion species identified in these survey scans can then be isolated and fragmented in a subsequent fragmentation scan (MS2). The fragment spectra give information on the chemical structure of the precursor ions, which means in the case of peptide precursor ions that the peptide's amino acid sequences can be deduced from the MS2 spectra.

This is the most widely used proteomics approach, called data-dependent acquisition (MS1 scans trigger MS2 scans), bottom-up (based on peptides, rather than on whole proteins) LC-MS proteomics.

2.3.2 Label-free proteomics

Mass spectrometry approaches are great for their throughput and depth of analysis, but by itself, MS is highly variable and not quantitative. Each analyte has a unique ionization efficiency and signal response, which cannot be predicted from the chemical structure of the analyte. Over the last decades, mass spectrometrists have developed various technologies to normalize MS proteomics so that the technology can be used to not just identify analytes but also to quantify their abundances.

Early quantitation approaches were based on spectral counting, where the number of MS2 spectra associated to a protein is used as a proxy for its abundance. This is based on the idea that sampling for MS2 fragmentation events occurs equally over the LC-MS run, and that peptides from a more abundant protein will be picked more frequently for MS2 fragmentation. This spectral counting value has to be normalized for several factors like the number of observable peptides per protein, which is included in the normalization methods such as emPAI (Ishihama *et al*, 2005) and APEX (Lu et al, 2007).

Alternatively, the MS1 intensity of the peptide ions can be used to quantify protein abundance. Examples of this approach are the Top3 method (Silva, 2005), MaxLFQ (MaxQuant's label-free quantification (Cox *et al*, 2014)) or Progenesis QI from Nonlinear Dynamics. Comparisons of label-free quantification approaches showed that MS2-based spectral counting approaches have a lower dynamic range than the MS1-based quantitations (Grossmann *et al*, 2010; Arike *et al*, 2012; Ahrné *et al*, 2013; Fabre *et al*, 2014).

Another label-free approach that quantifies peptides based on their MS1 intensities is SWATH-MS (Gillet *et al*, 2012). In contrast to the before mentioned approaches, SWATH-MS does not work in a data-dependent fashion, where the mass spectrometer selects the MS2 precursors from MS1 data. Instead, SWATH-MS lets the mass spectrometer scan over the MS1 range with small m/z windows (swaths), fragmenting all ions within these ranges. The resulting fragment ions are then mapped back onto a precursor ion according to their shared LC-elution behavior. SWATH-MS approaches are discussed to hold a high potential to increase the reproducibility of MS-proteomics (Navarro *et al*, 2016).

The mentioned schemes give relative quantifications of the same protein in different samples measured in the same experiment. Absolute quantities of different proteins within a given sample are not readily obtained. In absolute quantification approaches such as iBAQ (intensity-based absolute quantification, (Schwanhäusser *et al*, 2011)), proteins with known concentrations are spiked as references into the samples. The MS1 signal intensity of histones can also be used as a reference for absolute quantification without additional spiked-in proteins (Wiśniewski *et al*, 2014). The benefit of such absolute quantification is that it allows to create large repositories, such as the drafts for the tissue-specific human proteome (Kim *et al*, 2014; Wilhelm *et al*, 2014). However, the reliability of these large-scale repositories are being debated (Li *et al*, 2014b; Ezkurdia *et al*, 2015).

2.3.3 Labeling proteomics

The downside of label-free MS studies is that all the quantitative comparisons are done in silico after the actual measurements (Karp & Lilley, 2007). Alternatively, samples can be labeled isotopically and can be combined earlier during sample processing. They are then measured as a mixture and de-multiplexed in the mass spectrometer according to their isotopic label. Technical variations during sample processing, LC or MS affect both samples equally, leading generally to more stable and less variable quantitation.

Depending on the type of labeling, the samples will be mixed at various levels of the proteomic sample processing workflow – be it at the peptide , protein, or tissue level (Russell & Lilley, 2012). The earlier the integration, the more technical variation can be avoided.
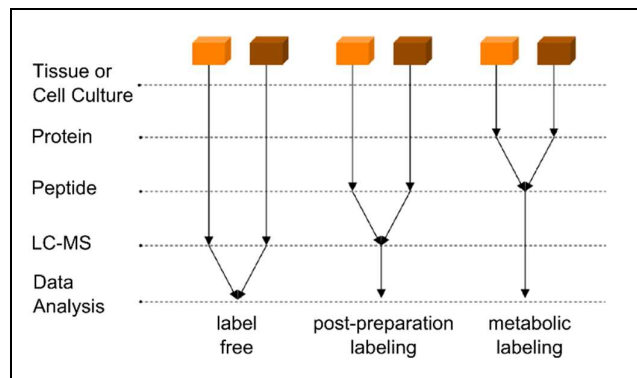
Figure 2.2: The key difference in label-free, post-preparation labeling and metabolic labeling proteomics is the time point when the samples are compared or mixed together (modified from Russell & Lilley (2012)).

2.3.4   Post-preparation labelling

Labelling approaches can be separated into metabolic labeling and post-preparation labeling. For the latter, the samples are mixed as peptides, after separate protein extraction and proteolysis. There are various chemical strategies available to label the peptides, such as ICAT (Gygi *et al*, 1999), iTRAQ (Ross *et al*, 2004b), Tandem Mass Tag (TMT, Thermo Scientific), and dimethyl labeling (Boersema *et al*, 2008). All approaches use different isotope combinations to label the peptides, but differ in their quantitation strategy. Dimethyl and ICAT labeling, for instance, result in peptides that differ in their absolute masses, leading to m/z shifts on the MS1 level. iTRAQ and TMT tags, on the other hand, add isobaric masses to the peptides so that the different label states can be quantified on the MS2 level, allowing for higher quantitative accuracy and higher sample multiplexing (Mertins *et al*, 2012). MS3 quantitation schemes are also possible, increasing the multiplexing capabilities to 10 samples within the same LC-MS run (Viner *et al*, 2013). Increasing the number of samples that can be combined massively reduces the required MS runtime, one of the most expensive factors for proteomics experiments.

The key benefits of these post-preparation labeling schemes are their high quantification accuracy, their multiplexing potential, and a straightforward application to different organisms or sample types. The downside of such approaches is that the labeling efficiencies could vary between the labeling reactions and that the labeling agents are comparatively expensive, with the exception of dimethyl labeling.

2.3.5   Metabolic labelling

Metabolic labeling approaches allow mixing samples one step earlier in the sample preparation procedure than the before mentioned post-preparation labeling schemes. Metabolic labeling utilizes the protein biosynthesis machinery of the biological system itself to create differentially labeled proteins so that samples can be combined as tissues or protein extracts.

Generally, metabolic labeling approaches supply the organism with an isotopically labeled compound that is introduced into the analyte in an ideally unbiased fashion. The most fundamental metabolic labeling strategies change the isobaric status of a nutrient source, for instance using ^{15}N-labelled nitrate as the sole nitrogen source (Engelsberger *et al*, 2006; Bindschedler *et al*, 2008; Kierszniowska *et al*, 2009). Theoretically, all biochemically incorporated elements - C, H, O, S, P, and

N - can be used for metabolic labeling, though with varying degrees of practicality (Nelson *et al*, 2014b). P, for instance, cannot be used for metabolic labeling, as it has no stable isotope. S has a stable isotope, but not all peptides contain S atoms. ^{2}H is also a stable isotope but can have detrimental effects on the biological system during the labeling (Yang *et al*, 2010). The stable isotope ^{18}O was effectively applied to label mice without apparent side effects (Zhou *et al*, 2012). However, the employed $H_2^{18}O$ is rather expensive. As O represents about 10% of all atoms of a peptide, ^{18}O labeling creates complex mass spectra, yet the isotope envelopes of ^{18}O-labeled peptides are uniquely shaped due to the 2 Da mass shift of the label. Most naturally occurring isotope variants induce only a 1 Da mass shift.

About 1% of all natural C atoms are in the form of the stable ^{13}C isotope. Using it as a labeling agent holds great promise as it has no detrimental side effects. Yet full labeling would require exchanging all medium components with their ^{13}C counterparts, which is either very costly or outright impossible. But also partial labeling with a ^{13}C-labelled carbon source can be informative, for instance for tracking metabolic fluxes through organisms (Hiller *et al*, 2011) or for the interrelation of subcellular amino acid pools (Allen *et al*, 2012). Auxotrophic organisms can be labeled with $^{13}CO_2$, but working with gaseous labeling agents is experimentally challenging (Kölling *et al*, 2013) and results in non-complete labeling rates (Young *et al*, 2011; Ishihara *et al*, 2015).

In proteomics, the most frequently employed elemental labeling uses the stable isotope ^{15}N. Peptides contain at least one N per amino acid so that peptides are effectively labeled by ^{15}N. Yet the labeling creates a sequence-specific mass shift so that label pairs are more difficult to match. ^{15}N-labeling results in mass spectra with broad isotope envelopes that can be challenging to interpret and quantify.

Yet, ^{15}N labelling is rising in popularity in part due to its low cost and straightforward application for autotroph organisms, where simply the N-source in the form of inorganic salts has to be exchanged (Engelsberger *et al*, 2006; Bindschedler *et al*, 2008; Kierszniowska *et al*, 2009; Höhner *et al*, 2013). For non-autotrophic organisms, the ^{15}N-label has to be introduced into all compounds of the medium or feed, so that the medium has to be prepared for instance from ^{15}N-labelled algae. Still, ^{15}N-labelling was applied to various biological systems, including mammalian cells (Conrads *et al*, 2001), *E. coli* (Ross *et al*, 2004a), yeast (Kolkman *et al*, 2006), algae (Höhner *et al*, 2013), Arabidopsis cell culture (Engelsberger *et al*, 2006), Arabidopsis hydroponic plants (Bindschedler *et al*, 2001), and even whole mice (Price *et al*, 2010). The low cost of ^{15}N-labelling makes it the method of choice for studies where bulk amounts of biological material are needed such as quantitative enrichments of low abundant phosphorylated peptides (Kline *et al*, 2010).

All mentioned elemental metabolic labeling strategies compare between the labeled and unlabeled protein populations, allowing for 2-fold multiplexing.

2.3.6 Metabolic labelling with amino acids

For prototrophic organisms, however, a different labeling technology gained favor in the field. Prototrophs have to take up their essential amino acids from the growth medium. These can easily be exchanged to an isotopically labeled version, resulting in high labeling rates without any known side effects. Labeling the proteins themselves allows mixing the samples either as complete cells or

after protein extraction. All subsequent biochemical procedures including the trypsination and the LC-MS measurements are then performed on the mixture of the samples. This minimizes technical variabilities and allows for reliable quantitations (Russell & Lilley, 2012).

This approach was called stable isotope labeling with amino acids in cell culture, coining the acronym SILAC (Ong, 2002). It leads quickly to high labeling efficiencies of >97%, close to the theoretical limit of the isotopic purity of the labeling agent (Ong & Mann, 2006). If supplied in excess, non-essential amino acids such as valine, tyrosine or methionine can also be used for SILAC (Ong, 2002). However, such approaches have lower labeling efficiencies.

A prime benefit of SILAC over other metabolic labeling strategies is the simplicity of the resulting spectra. SILAC was first designed with leucine as the labeled essential amino acid (Ong, 2002), but was quickly adapted to lysine (K) and arginine (R) labeling. The frequently used endoprotease trypsin cleaves after exactly these two amino acids, creating ideally peptides with exactly one differentially labeled amino acid. This massively simplifies the search space that peptide search algorithms have to analyze. Given that there is only one differentially labeled amino acid per peptide, the delta mass between a labeled and unlabelled peptide is identical for all peptides of an experiment. This allows to identify the labeling partner of a peptide even if the amino acid sequence can only be assigned to one peptide of the pair.

SILAC peptide pairs can be quantified on the MS1 level with a variety of software solutions with MaxQuant (Cox & Mann, 2011) as the most commonly used tool.

As expressed by its name, SILAC works best in a cell culture, where the amino acids are fully dispersed and equally available to all cells. It was originally developed for mammalian cell culture (Ong, 2002), but is equally applicable to bacteria and archaea (Veenstra *et al*, 2000), yeast (Gruhler *et al*, 2005a),or unicellular algae (Naumann *et al*, 2005). However, also multicellular organisms can be labeled by feeding SILAC-labelled yeasts or algae over an extended period. This leads to a slowly increasing but complete, organism-wide labeling. This approach opened up SILAC labeling to most model organisms such as *Drosophila melanogaster* (Sury *et al*, 2010), *Caenorhabditis elegans* (Fredens *et al*, 2011)*, Mus musculus* (Krüger *et al*, 2008; Zanivan *et al*, 2012)*, Danio reio* (Nolte et al, 2014), chicken (Doherty *et al*, 2005)*,* and even the newt *Notophthalmus viridescens* (Looso *et al*, 2012)*.*

As the labeling has to be passed on through the food chain, labeling a full organism is costly. It is thus more economical for such approaches to purchase pre-labeled tissue samples and use them as a common reference to the non-labelled samples (called spike-in SILAC if derived from a single reference (Geiger *et al*, 2011) or Super-SILAC if derived from a mix of references (Geiger *et al*, 2010)). Such spike-in references can even be employed across organismal borders. SILAC-labelled mouse tissue was used for instance as a reference for human serum samples (Zhao *et al*, 2013). Partially labeled organisms can also be used as spike-in references, substantially reducing the cost for the SILAC labeling (Looso *et al*, 2012).

It is important to remember that all SILAC peptides are produced by the protein biosynthesis machinery of the biological system itself and that the absolute concentration of each peptide is unknown. In comparison to SRM, where a synthetic peptide with a known concentration can be

spiked into the analyte mixture to obtain absolute quantifications, SILAC by itself does only allow for relative quantifications between the sample and its reference.

SILAC labeling schemes with delta masses of up to 16 Da were designed, which allows to combine up to 5 labeling states, without creating a strong overlay of the isotope envelopes (Chen *et al*, 2015). Alternatively, NeuCode was proposed to expand SILAC's multiplexing capacity (Hebert et al, 2013). It suggests separating SILAC populations in the MS based on the neutron binding energy variations of stable isotopes. The ~1 Da mass shift of a ^{15}N atom differs slightly from the mass shift of a ^{13}C or ^{2}H atom. Consequently, a lysine molecule with a ~8 Da delta mass can be labeled with up to 39 different isotopologue combinations, creating unique mass shifts just within the minuscule delta mass range of 8.12 to 8.16 Da. Current high-performance mass spectrometers with a resolution higher than 200'000 are able to resolve some of these differences, but resolutions of up to 1'000'000 are needed to fully resolve all 39 proposed NeuCode label states.

2.3.7 Calculating labeling efficiencies

On a side note, there are different ways of calculating the labeling efficiency, which is a quantitative value indicative of the amount of labeled proteins compared to the amount of unlabeled proteins. Most protein abundance studies do not have to assess the labeling efficiency, as they only compare the labeled peptide pools. However, when using metabolic labeling including SILAC, the labeling efficiency is a crucial factor to quantify as a non-complete labeling efficiency would create a bias in the common case of quantifying a labeled vs an unlabeled culture.

Incomplete labeling can occur, when the labeling phase was not long enough or when a non-essential amino acid is used for labeling and the endogenously produced unlabeled amino acid can be incorporated instead. On top of that, labeling compounds such as isotopically labelled amino acids are only enriched for a chemical isotope, which means that they always contain trace amounts of different labels. SILAC amino acids can be purchased with isotopic purities in the range of 97 to 99 %. The commonly employed SILAC quantification software MaxQuant uses a median centralization on the labelled-to-unlabeled ratio to balance out small differences in labeling efficiencies (Cox & Mann, 2008).

In the SILAC field, the following formula is used frequently to determine the labeling efficiency (LE, website of Prof. Mathias Mann's group, Max Planck Institute of Biochemistry, Martinsried http://www.biochem.mpg.de/221777/SILAC , and e.g. (Schütz *et al*, 2011; Visscher *et al*, 2016)):

$$LE = \frac{Heavy - Ligh}{Heavy} = 1 - \frac{1}{ratio\left(\frac{Heavy}{Ligh}\right)} = 1 - ratio\left(\frac{Light}{Heavy}\right)$$

In the protein turnover field, the relative isotope distribution (RIA) is rather used to make statements on the labeling efficiency and the progression of a label swap.

$$RIA = \frac{Heavy}{Heavy + Light} = \frac{1}{1 + \frac{1}{ratio\left(\frac{Heavy}{Light}\right)}}$$

The RIA is limited to a range of 0 to 1, while the LE can range from 1 to negative infinite (Figure 2.3). The difference is negligible at high labeling rates, but a 1:3 mix of heavy to unlabeled proteins

computes a RIA of 25 %, while the LE leads to an illogically negative value of -200 %. Therefore, I consider the RIA as a more appropriate measure of the labeling efficiency and employ the RIA calculation within this thesis if not noted otherwise. Relative isotope fraction (RIF), as proposed by Rahman *et al*, (2016), is actually a more accurate term than RIA, but it is used less frequently in the scientific literature as of yet. We therefore remain with the term RIA.

For triple labeling, these formulas translate into:

$$LE = \frac{Heavy + Medium\_Heavy - Light}{Heavy + Medium\_Heavy}$$

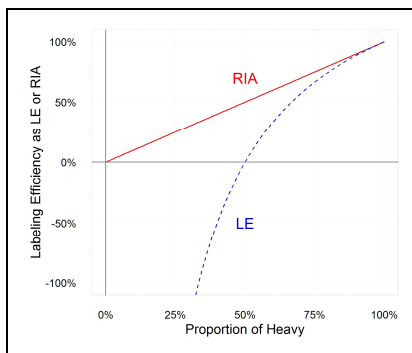$$RIA = \frac{Heavy + Medium\_Heavy}{Heavy + Medium\_Heavy + Light}$$

Figure 2.3: Relative isotope distribution (RIA, solid line) and Labelling Efficiency (LE, dashed line) as a function of the amount of heavy-labeled compounds.

2.3.8 Metabolic labeling in plant proteomics

Plants, like all prototrophic organisms, lend themselves ideally to metabolic labeling with inorganic compounds such as ^{15}N-labelled nitrate and ammonium or $^{13}CO_2$ (reviewed in e.g. Matthes *et al*, (2014); Nelson *et al*, (2014)). Labeling with complex organic molecules like amino acids in SILAC is employed only infrequently, as the labeling efficiencies are inconsistently low and as even physiological side-effects of high amino acid supplementation were reported (Gruhler *et al*, 2005b).

In the algae Chlamydomonas reinhardtii, the strain CC-424 is used frequently for SILAC studies (Naumann *et al*, 2005; Terashima *et al*, 2010; Mastrobuoni *et al*, 2012). CC-424 is prototrophic for arginine, as it has a dysfunctional argininosuccinate lyase (*arg2*).

In Arabidopsis, I found three SILAC studies:

1. (Gruhler *et al*, 2005b): This first plant SILAC study tested the labeling efficiencies when using isotopically labelled lysine, arginine or leucine in Arabidopsis suspension cell cultures. By adding one dose of 760 µM arginine or 800 µM leucine, they reaching incorporation rates of 70-80% after 8 days of growth. Adding 330 µM lysine led to a comparatively low incorporation rate of 45%. All three labeling rates were deemed too low for thorough proteomics approaches. Gruhler and colleagues did not detect any decrease in culture viability from the labeling procedure but discuss potential negative side-effects regarding lysine labeling as lysine affects the biosynthesis of several amino acids.

Lysine inhibits Dihydrodipicolinate synthase (DHDPS or DapA, IC_{50} of 15 μM, (Ghislain *et al*, 1990)), which catalyzes the first unique reaction of the lysine biosynthesis in plants, which is the condensation of aspartate-semialdehyde (ASA) and pyruvate. Lysine also inhibits Aspartate kinase (AK, IC_{50} of 750 μM (Heremans & Jacobs, 1997)), an enzyme preceding DapA in the lysine biosynthesis pathway. AK is essential for the biosynthesis of lysine, methionine, and threonine, which is also a direct precursor for isoleucine. As threonine also inhibits AK, the addition of 1 mM of both lysine and threonine leads to growth inhibition of maize callus cultures, which can be recovered by supplementation with 1 mM of methionine or homoserine, a methionine precursor (Green & Phillips, 1974). Adding only 1 mM of lysine to the growth medium did not have a negative effect on the maize callus cultures. In wild-type Arabidopsis seedlings, 2 mM of lysine is sufficient to inhibit growth (Heremans & Jacobs, 1997).
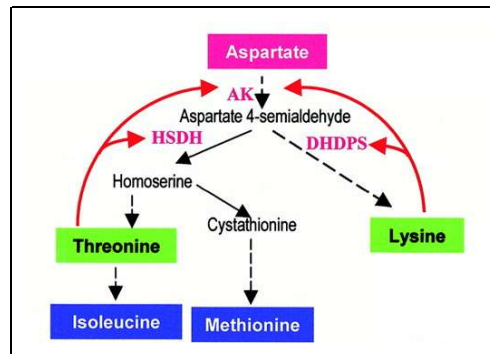

Figure 2.4: Overview of the biosynthesis pathways of threonine, lysine, methionine, and isoleucine together with the known inhibitory activities (upward red arrows) (Galili, 1995; Wang & Larkins, 2001).

2. (Schütz *et al*, 2011): This study substantially improved the general SILAC strategy in Arabidopsis suspension cell cultures by comparing medium-heavy-labeled $^4H_2$-Lys4 cultures with heavy-labeled $^{13}C_6{}^{15}N_2$-Lys8 cultures (Figure 2.5). This overcomes elegantly the problem of incomplete labeling efficiencies, as the unlabeled Lys0 peptides should amount to the same quantity in the two compared cultures. The Lys0 peptides can therefore be excluded from the quantitation and just the Lys4- and Lys8-labelled subproteome are compared.

Additionally, Schütz *et al* increased the labeling efficiency to 91%, coming close to the range of mammalian systems, by 1) adding substantially more labeled amino acids to the culture (daily supplementation of 350 μM lysine, ~5x more than in the standard mammalian protocol (Ong & Mann, 2006), and 7x more than in (Gruhler *et al*, 2005b)), 2) reducing the nitrogen content of the growth medium by 66% (20.6 mM nitrogen from 10.3mM NH_3NO_4 instead of 60.2 mM nitrogen from 20.6mM NH_3NO_4 and 19mM KNO_3), 3) cultivating the cells in the dark, which decreases the number of chloroplasts, the compartment of endogenous lysine biosynthesis, and 4) extending the labelling period to 12 days. When they employed their approach on the same cell culture grown in the light, the labeling rate was decreased to only 58% and the variability of the labeling was increased (Schütz et al, 2011). So, although effective, the prerequisite of growing the cells in the dark was deemed by (Holper *et al*, 2014) to "not represent a particularly meaningful condition for most plant physiological studies". The study by Schütz and colleagues was purely methodological and did not address a biological question.
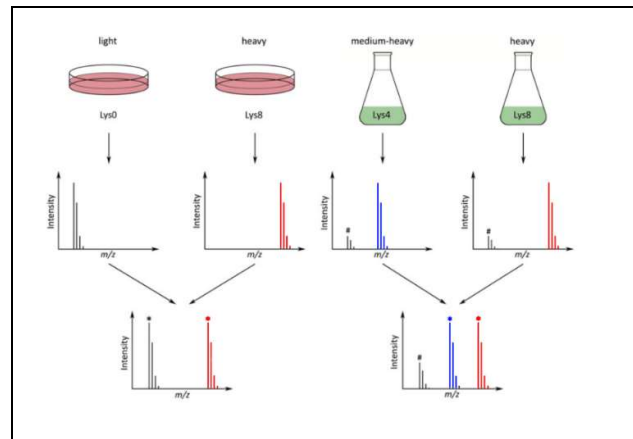
Figure 2.5: Improved plant SILAC strategy (Schütz et al, 2011).
In contrast to classical SILAC (left), the improved plant SILAC can cope with incomplete labeling. The plant cell culture's proteomes are compared by quantifying the red heavy Lys8-labelled proteins with the blue medium-heavy Lys4-labelled proteins. The black, unlabeled part of the proteome arises from the endogenous lysine biosynthesis and is excluded from the quantitation as an unchanging contaminant.

3. (Lewandowska *et al*, 2013): In this study, Arabidopsis seedlings were grown in liquid culture supplemented with both arginine and lysine. The seedlings were germinated in the liquid medium itself and after 19 days of labeling, labeling efficiencies of >95% could be determined for representative peptides. The seedlings were supplied every 2 days with fresh Gamborg B5 medium supplemented with 1 mM of labeled lysine and arginine. The labeled seedlings were used to quantify the response to salt stress after treating the seedlings with 80 mM NaCl.

Although these studies gained high attention from the community and are mentioned in numerous reviews, I could only find these three published proof-of-concept studies and no reproduction of the approaches neither from the same nor independent laboratories.

2.4   Protein degradation

Protein abundances are regulated either by changes in protein synthesis or in protein degradation. The main route of protein degradation is through proteasomal degradation, but it can also occur through authophagy or through the action of proteases. Protein degradation through proteases is especially important in plastids and mitochondria, as these organelles do not possess proteasomes. Instead, they contain prokaryote-related ATP-dependent proteolytic machineries, such as the carboxy-terminal processing protease CtpA, the stromal Clp, the lumenal serine-type protease DegP and the ATP-dependent thylakoidal metalloprotease FtsH (Sakamoto, 2006; Chi *et al*, 2012; Nishimura *et al*, 2016).

Autophagy describes the process of orderly degradation of cellular material by engulfing large cytoplasmic volumes, potentially containing whole organelles, in a double-membrane structure called autophagosome (Levine *et al*, 2011). In mammalian cells, these structures fuse with lysosomes, while in plants and fungi these fuse with the vacuole for degradation of the formerly

cytoplasmatic cargo via acidic lysosomal hydrolases. This process is also described as macroautophagy and can be differentiated from microautophagy, where cytoplasmatic material gets directly introduced into the lysosome or vacuole by membrane invagination (Li *et al*, 2012b). The third mode is called chaperone-mediated autophagy (CMA) where proteins are selectively bound by chaperones such as Hsc70 (Heat shock 70 kDa protein 8) or Hsp90 (heat shock protein 90) and are threaded in an active process through the double membrane into the lysosome or vacuole (Bandyopadhyay *et al*, 2008). Compared to the other two non-selective modes of autophagy, CMA describes a protein-specific, targeted process. In yeast, mammals and plants autophagy is primarily regulated via TOR signaling (Díaz-Troya *et al*, 2008), which integrates nutrition signaling and affects biosynthesis regulation (Hay & Sonenberg, 2004).

2.4.1   Proteasomal degradation

In comparison to autophagy and protein degradation through proteases, protein degradation by the ubiquitin 26S-proteasome system (UPS) is a more targeted method, in which proteins are degraded in the 26S proteasome. This is one of the largest protein complexes in a cell, consisting of the cylindrical 20S core particle and two cap like structures. The 20S core is composed of four stacked rings, each consisting of 7 β subunits in the case of the two central two rings and 7 α subunits in the case of the outer two rings (Smith *et al*, 2007). These rings form a hollow cylinder in which multiple proteolytic sites are located. These degrade entering proteins by chymotrypsin-like, trypsin-like and peptidylglutamyl-peptide hydrolyzing mechanisms (Wilk & Orlowski, 1983). On both sides of the 20S particle, cap-like 19S regulatory particles are situated, which are large multi-protein complexes that regulate and support all steps required for protein degradation (Zwickl *et al*, 1999). In place of the 19S particle, a protein complex called 11S regulatory particle can also bind to the ends of the 20S core particle, which also regulates and supports proteolysis (Förster *et al*, 2005).

The signal that marks a protein for proteasomal degradation is the attachment of a polyubiquitin tag. Ubiquitin is a small 8.5 kDa regulatory protein that is bound to a lysine residue of the target protein in a cascade consisting of three steps: activation of the ubiquitin by ubiquitin-activating enzymes (E1s), conjugation by ubiquitin-conjugating enzymes (E2s), and ligation to the target protein by ubiquitin ligases (E3s) (Pickart & Eddins, 2004). In each step of the enzymatic cascade an increasing number of unqiue proteins is involved. The genome of Arabidopsis encodes 2 E1s, at least 37 E2s and potentially over 1400 E3s, which sum up to 10-times more genes involved in the UPS than in the same pathway in yeast (Vierstra, 2003, 2009).

Once a single ubiquitin is attached to the target protein (monoubiquitylation), more ubiquitin moieties can be bound to one of the lysine residues of that ubiquitin. The linkages of these polyubiquitylations define different outcomes for the protein they are attached to, while polyubiquitylation at K48 or K11 is in general associated with protein degradation (Ikeda *et al*, 2010; Komander & Rape, 2012). Polyubiquitin chains can also be removed from the target protein by a class of enzymes called deubiquitinating enzymes (DUBs, Reyes-Turcu *et al*, (2009)).

With the ubiquitin proteins, the E1s, E2s and E3s, the proteasome complex subunits and the DUBSs, about 6% of the proteome (> 1'600 loci) are involved in the UPS (Vierstra, 2003, 2009). With this genomic complexity, protein degradation by the UPS was suggested to be equally important for plant cell regulation as transcriptional regulation or protein kinase cascades (Vierstra, 2003).

2.4.2 Theory of measuring protein degradation

The degradation of any molecule at constant speed can be described by a first order exponential decay reaction: $Molecule \xrightarrow{\Delta time} \emptyset$. This is used in physics for instance to describe the decay of radioactive material, where the degradation rate is a property of the molecule itself and is unaffected by any factors. In biology, we employ this framework to describe various processes including the degradation of proteins.

Mathematically, such a degradation reaction can be described by the following formula for exponential decay:

$$[molecule_{Tx}] = [molecule_{T0}] * exp^{(- k_{deg} * time_x)}$$

It is based on an initial concentration at time point zero $[Molecule_{T0}]$ multiplied by an exponential function of the negative degradation rate constant k_{deg} multiplied by time x. At an initial concentration of 1 for the molecule, this creates the following trajectories for a k_{deg} of 10 (red), 1 (blue) and 0.1 (green). The half-lives are indicated by dashed vertical lines (the green trajectory does not reach the half-life in the indicated time.
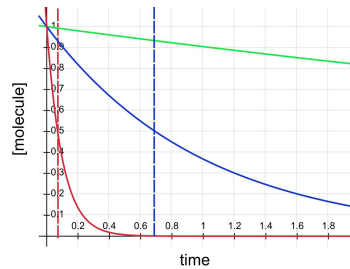
The k_{deg} parameter is an abstract term and can be unwieldy to work with, so the notion of half-life is frequently used to describe decay processes in a more graspable manner. A molecule's half-life describes the time until the molecule's population is being reduced to half of its initial value. By transformation, this results in the following formula:

At half-life: $\quad [molecule_{Tx}] = ½ * [molecule_{T0}] \xrightarrow{yields} time_x = \dfrac{log\left(\dfrac{½* [molecule_{T0}]}{[molecule_{T0}]}\right)}{- k_{deg}} = \dfrac{log(½)}{- k_{deg}}$

For exponentially decaying molecules, the degradation rate and the half-life are constant and can be determined at any time point during the degradation.

In contrast to radioactively decaying material, proteins can also be synthesized anew. Under steady-state conditions, a protein's abundance does not change. The proteins are constantly degraded to the same extent as they are replenished, so the degradation and synthesis rates are equal. In that context, the notion of a protein's turnover time was put forth to describe the average lifetime of a protein before it is replaced by newly synthesized protein. The turnover time is a crucial term for cellular protein homeostasis and is calculated by the concentration of a molecule divided by its flow-through rate (Dettmann, 2008).

Only under steady-state conditions, one can truly speak of turnover times where k_{syn} and k_{deg} are the same and don't have to be delineated. Under these conditions, the turnover time can therefore be

calculated by dividing the molecule concentration with either the degradation or synthesis rate. However, true steady-state conditions can only rarely be upheld in biological experiments. Optimally, one would need a combined turbidostat and chemostat fermenter for such an experiment. Still, obtaining only the synthesis rate can be an informative parameter for protein homeostasis (Schwanhäusser *et al*, 2009). However, when quantifying newly synthesized proteins, it has to be taken into consideration that these proteins also undergo degradation processes.

If a biological system is not in steady state, the synthesis rate (k_{syn}) and the degradation rate (k_{deg}) can be non-equal and have to be delineated by measuring at least two of the three components of the formula: $[Protein] = k_{syn} - k_{deg}$.

The degradation rate k_{deg} can be determined experimentally by labeling the molecule of interest and tracking the loss of this label over time. When acquiring this label-loss rate k_{loss} in a biological system, one has to consider that these rates are obtained in growing and dividing cells. Even under steady-state conditions with constant molecule levels, the labeled molecules are diluted by cell expansion. This process can be factored in by subtracting k_{dil}, the rate of dilution by cell growth, from the label-loss rate to obtain the true rate of degradation: $k_{deg} = k_{loss} - k_{dil}$ (Pratt *et al*, 2002).

2.4.3 Experimental assessment of protein degradation

Several techniques have been developed to experimentally determine the degradation rate k_{deg}. I will describe in the following paragraphs the four main approaches, depicted schematically in Figure 2.6.
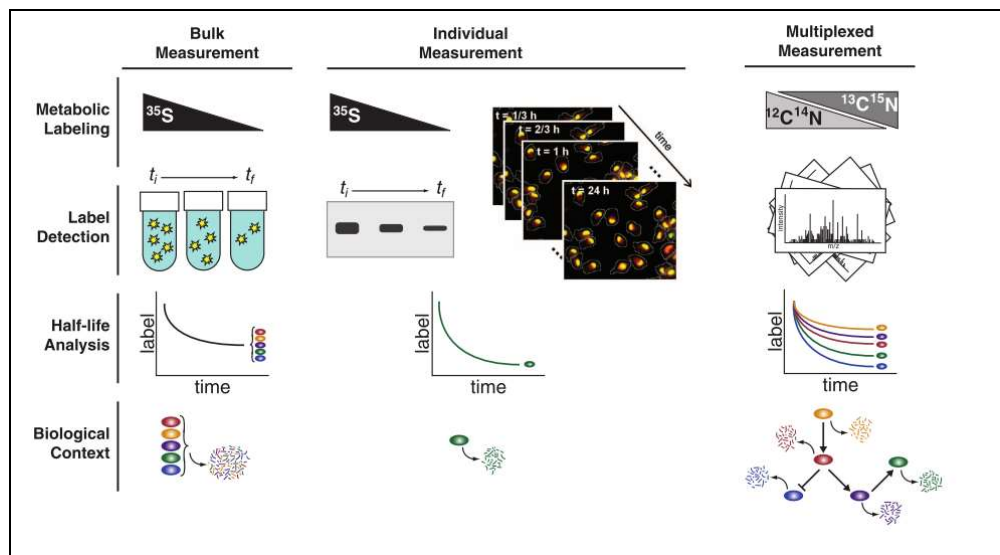
Figure 2.6: Schematic depiction of the main experimental approaches to obtain protein degradation rates adapted from Hinkson & Elias, (2011) and Eden et al (2011).

The first developed approach to obtain protein degradation rates is to quantify protein abundances over time by e.g. quantitative Western Blot while inhibiting de novo protein synthesis. Chemicals such as cycloheximide arrest mRNA-bound ribosomes by inhibiting translocation elongation so that no de novo protein synthesis can occur. This will lead to a decrease in protein abundance over time, as protein degradation proceeds normally. From these decreasing abundance values one can then

extrapolate the protein degradation rate. This approach is still employed today (e.g. Ferretti *et al*, (2016)) due to its straightforward methodology, but it has severe limitations. The cycloheximide treatment affects all cellular protein biosynthesis and therefore disrupts the overall protein homeostasis, which will eventually kill any cell. Therefore, the reliability of these degradation rates has to be considered carefully. Additionally, this approach requires specific and reliable antibodies to quantify the protein of interest and is generally labor intensive as it does not allow multiplexing.

Secondly, the degradation of proteins can be quantified microscopically when they are tagged with fluorophores. Protein-bound fluorophores can be inactivated with a strong light impulse without affecting protein abundance. The reemergence of the fluorescence signal is then informative for the protein synthesis time. This approach is called bleach-chase strategy (Eden *et al*, 2011) and gives *in vivo* kinetics of high temporal resolution on single cell level. The downside of this approach is that it requires substantial genetic engineering work and that tagged proteins might not behave equally to native proteins. Also, the intensity of the bleaching light pulse has to be balanced carefully so that it does not harm the cells. This issue can be avoided by fusing two fluorescence tags with unique maturation kinetics to the protein of interest. The ratio of the fluorescence from the two tags is then informative of the average protein age, which can be used to determine differences in the age of proteins in subcellular compartments (Khmelinskii *et al*, 2012).

The third approach allows to obtain protein turnover rates without genetic engineering or affecting protein abundances. These so-called pulse-chase metabolic labeling experiments can be split into two parts: During the pulse phase, a labeling agent such as a radiolabeled amino acid is added to proliferating cultured cells, which gets taken up by the cells and is incorporated into the proteome. In the following chase phase, the labeling agent is removed so that any newly synthesized protein does not incorporate the label. The time it takes for the cells to replace the preexisting labeled proteins by new unlabeled ones is the turnover time. Given the exponential decay characteristics, the labeling does not have to be complete. Also partial labeling can result in meaningful data (Doherty *et al*, 2005; Looso *et al*, 2012), but the higher the labeling rate, the better the dynamic range of the turnover assessment.

^{35}S- radiolabeled methionine is the most frequently employed labeling agent for such experiments. Without biochemical fractionation, the decay of the radioactivity in the protein extract is informative of the turnover rate of the whole proteome. To obtain protein-specific turnover rates, (immuno-) affinity enrichments have to be used. This approach, therefore, requires protein-specific antibodies or protein tagging and does not allow for multiplexed analyses. To improve the throughput, the ^{35}S-methionine labeled proteins can also be separated by 2D polyacrylamide gel electrophoresis (2D-PAGE), which allows assessing the turnover rates of more than 100 proteins in parallel (Galland *et al*, 2014).

As an interesting alternative imaging-based approaches were presented recently by Shen *et al*, (2014). Labeling phenylalanine with ^{13}C atoms results in a shift in the Raman scattering spectrum of 38 cm^{-1} compared to the ^{12}C-labelled isoform. This physical property was used by Shen and colleagues to determine the *in vivo* subcellular turnover of the full proteome of a mammalian cell line using a pulse-chase strategy.

2.4.4 Mass spectrometric assessments of protein degradation

The fourth and currently most frequently employed approach to assess protein degradation utilizes the ability of MS to differentiate between stable, non-radioactive, isotopically-labelled subpopulations of proteins. Employing the same pulse-chase strategy as described before, protein populations can be labeled with an isotopic marker and the decrease of the labeled population over time during the chase period is informative of the degradation rate. MS-based proteomics resolves the kinetics of individual proteins, giving this approach the highest throughput compared to the approaches described above. Nowadays, MS approaches have reached a stage of maturity that allows to reliably assess thousands of proteins in parallel, reaching near-proteome coverage for bacteria and small eukaryotes (Aebersold & Mann, 2016). In our view, dynamic mass spectrometry-based proteomics, therefore, represents the most powerful strategy for obtaining proteome-wide, protein-specific degradation rates. It is also important to note that a metabolic labeling approach does not affect proteome homeostasis, compared to the first discussed translation-arrest approach. Yet, compared to the imaging-based approaches, it can only obtain population-wide turnover data and cannot assess the variations on a single cell level.

Protein subpopulations can be labeled with different approaches. In theory, any compound that is taken up and integrated into proteins can be used for such protein turnover studies. This includes stable isotopes from all biological elements C, H, O, S, or N (Nelson *et al*, 2014b). P cannot be included on this list, as it does not have a stable non-radioactive isotope.

For autotrophic organisms, it is possible to apply non-organic substances such as CO_2, H_2O or NH_4NO_3 for metabolic labeling schemes. By supplying culture medium with differentially labeled inorganic salts complete labelling can be achieved. For plants the method to compare protein abundances of completely ^{15}N-labelled proteins is well established. The plants used in such experiments are grown hydroponically and all nitrate and ammonium in the hydroponic medium are exchanged by their ^{15}N-labelled counterparts (Engelsberger *et al*, 2006; Nelson *et al*, 2007; Bindschedler *et al*, 2008; Kierszniowska *et al*, 2009; Arsova *et al*, 2012). The advantage here is that ^{15}N labeling has no physiological side effect on growth rates or yield (Li et al, 2012).

A challenge of this approach is that the mass shift between a 100% ^{14}N labeled peptide and a 100% ^{15}N labeled peptide depends on the peptide's individual amino-acid sequence. To identify labeling pairs, the amino acid sequence of the peptide has first to be determined to identify the number of N atoms, which results in complex data analysis schemes. ^{15}N quantifications are consequently only well established for 100%-labelled proteomes, where next to every N-atom is isotopically labeled exclusively in the ^{15}N configuration.

However, protein turnover studies require that intermediately labeled peptides can also be quantified, which represents an analytically much more difficult task. A typical 10 amino acid tryptic peptide contains 11 to 20 nitrogen atoms, all of which can be labeled in either ^{14}N or ^{15}N configuration. This creates a broad isotope pattern over a range of 20 Da (Figure 2.7, discussed in (Li *et al*, 2012a; Nelson *et al*, 2014b)). Combined with the naturally occurring isotope envelope from ^{13}C isotopes, this creates highly convoluted mass spectra both at MS1 and MS2 level that are difficult to identify and even more difficult to quantify. On top of this comes the fact that every amino acid has a unique half-life ranging from 3 to 70 h, resulting in a non-stable ^{15}N labeling rate for each amino

acid (Li et al, 2012). Consequently, the labeling efficiency of every amino acid would have to be taken into consideration for the overall protein turnover. Additionally, photorespiration in photosynthetic tissues produces unlabeled ammonium that can exceed the primary nitrogen uptake of a cell, which consequently affects the nitrogen labeling efficiency of that particular tissue (Keys et al, 1978; Nelson et al, 2014b)

Nowadays, different algorithms are available that quantify such dynamic ^{15}N labeling changes (Li *et al*, 2012a; Nelson *et al*, 2013; Lyon *et al*, 2014; Trouillard *et al*, 2015; Fan *et al*, 2016), but no consensus has yet been established in the field. But even with modern algorithms, such a convoluted labeling means that many spectra have to be discarded as they are not interpretable (88.7% of all mass spectra in (Nelson *et al*, 2014a)).

To overcome the limitations mentioned above, various approaches have been developed. To name a few: Helbig *et al*, (2011) quantified the label change by only quantifying the decrease of the pure ^{14}N envelope without quantifying the emerging ^{15}N envelope. Li *et al*, (2012a) recommended to use MALDI-MS, which produces preferentially single charged peptides that are easier to interpret and quantify. They also established an approach to separate the protein sample at first by 2D-gel-electrophoresis to further reduce the protein complexity for the MS analysis. Naturally, this decreased the throughput and coverage of that study. Trouillard *et al*, (2015) presented a novel algorithm that quantifies intermediary labeled ^{15}N peptides by matching their spectra to a multimodal m/z distribution. Fan *et al*, (2016) recommended to spike a known concentration of unlabeled ^{14}N proteins into the ^{15}N sample so that there is always an ion species with known N composition to anchor the isotope envelope to. It also helps with the identification of the peptides, as intermediary labeled fragmentation spectra are difficult to assign to peptide sequences due to increased complexity of the spectrum.
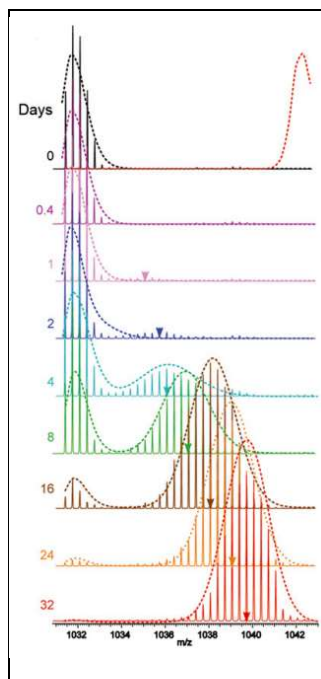

Figure 2.7: ^{15}N-labelled peptides produce complex mass spectra (Price et al, 2010).
A Cofilin-1 tryptic peptide is shown during ^{15}N-pulse labeling of mice. Both the intensity and the centroid mass (indicated by arrowheads) of the newly synthesized peptides change over the course of the labeling.

In summary, ^{15}N labeling has been successfully applied to determine protein turnover rates after adoption of special experimental workflows or algorithms. In part because of the cheap and easy labeling procedure, ^{15}N labeling is rising in popularity and is being applied to quantify protein turnover levels in e.g. mice (Price *et al*, 2010; Guan *et al*, 2011; Turck *et al*, 2016), Arabidopsis (Li *et al*, 2012a; Nelson *et al*, 2013; Huang *et al*, 2015; Fan *et al*, 2016), micrococcus algae (Martin *et al*, 2012), *Medicago truncatula* (Lyon *et al*, 2014), barley (Nelson *et al*, 2014a), and *Caenorhabditis elegans* (Dhondt *et al*, 2016). Dynamic ^{15}N labeling is a powerful methodology that was used for instance to identify very long lived proteins such as nucleoporins, histones, and proteins of the cornea that hardly degrade over the lifetime of a cell (Toyama *et al*, 2013).

2.4.5 SILAC for protein degradation assessments

As metabolic labeling with isotopic elements creates mass spectra that are convoluted and difficult to interpret and quantitate, the method termed stable isotope labeling with amino acids in cell culture, SILAC (Ong, 2002)), has gained large popularity for protein turnover studies. This is due to a rather straightforward labeling and quantitation methodology. SILAC can be used with any amino acid, but especially essential amino acids allow quick incorporation and high labeling efficiencies of ~97%, close to the theoretical limit of the isotopic purity of the supplied amino acid (Ong & Mann, 2006). Classically, SILAC studies use both lysine and arginine, as they account for all terminal amino acids of fully trypsinized peptides. This creates peptide pairs that are differentially labeled by exactly one amino acid per trypsinized peptide and therefore only have two labeling states (native light label vs artificial heavy label). This results in mass spectra with only moderately increased complexity, which often allows for the peptide assignment to fragmentation spectra from both labeled and unlabeled peptides. Each peptide pair has a clear fixed mass shift (e.g. a shift of +8 Da for Lys0 vs Lys8 labeled peptides) and can be identified with a variety of software solutions such as MaxQuant (Cox & Mann, 2008) or Mascot (Perkins et al, 1999). Quantification of SILAC is also rather straightforward, as the monoisotopic MS1 intensity of the labeled and unlabeled peptide pair can be compared directly in good approximation. In contrast to labeling with isotopic elements, where protein turnover leads to peptides with partial labelling states and therefore shifts in isotope envelopes, in SILAC it leads to an exchange of labelled peptide with unlabeled peptide or vice versa. SILAC is also compatible with selected reaction monitoring (SRM) data-independent MS that gives superior quantitation accuracy and sensitivity for dynamic SILAC turnover studies (Holman *et al*, 2016). Furthermore, supplementation with stable-isotope labeled amino acids does not have detrimental effects compared to supplementation with their natural occurring counterparts (Ong, 2002) - in comparison to for instance ^{2}H-labelling (Yang *et al*, 2010).

2.4.6 Noteworthy dynamic SILAC studies

The earliest MS-based protein turnover study used a SILAC labeling approach with ^{2}H$_{10}$-leucine and could determine protein turnover rates for 50 proteins in a chemostat yeast culture (Pratt *et al*, 2002). In a proof-of-concept study, (Doherty *et al*, 2005) showed that protein degradation rates can also be obtained from living higher organisms. They labeled live chicken by feeding them with a diet of 50% ^{2}H$_8$-valine and obtained protein turnover rates for 8 proteins. In 2009, Doherty and colleagues followed up on this study with a ground-breaking, truly large-scale study, coining the term dynamic SILAC for pulse-chase SILAC approaches (Doherty *et al*, 2009). They could obtain turnover rates for 600 proteins in cultured human cells using ^{13}C$_6$-arginine labeling. This chase period was performed

on a comparatively short timescale of 8 h, so proteins with long half-lives were not well resolved. Still, with this wealth of data, Doherty and colleagues could challenge several protein-degradation theories that are based on lower-throughput data, e.g. the N-end rule, the influence of disorder of a protein's structure, or the effect of PEST sequences (protein stretches enriched in proline (P), glutamic acid (D), aspartic acid (E) and serine (S) or threonine (T) residues).

Claydon *et al*, (2012) measured protein turnover in higher throughput in whole organisms. They labeled mice with 50% $^8H_2$-valin feed and determined the protein turnover with 2D-PAGE MALDI-ToF MS/MS. This approach allowed to determine tissue-specific turnover rates, showing a correlation between tissue-specific growth rate and turnover time. Also, they showed circadian fluctuations of labeling rates and assessed protein recycling rates.

In 2009, Schwanhäusser *et al*, showed that protein synthesis rates quantified with pulsed SILAC are equally informative for the protein turnover rate at steady-state as pulse-chase designs. This reduces the experimental complexity of most designs.

Holman *et al*, (2016) designed SRM assays witch which they could determine protein turnover with unprecedented accuracy and precision. Data-independent SRM is several times more accurate than data-dependent shotgun proteomics, but with substantially decreased throughput.

Nowadays, the high speed and mass accuracy of modern mass spectrometers allow to create more multiplexed and convoluted labeling schemes that allow to disentangle synthesis, degradation and protein abundance in a single labeling scheme (Jayapal *et al*, 2010; Dephoure & Gygi, 2012; Boisvert *et al*, 2012; Fierro-Monti *et al*, 2013). This also allows to obtain degradation rates under non-steady-state conditions.

The turnover studies that currently have the highest coverage include reports by Schwanhäusser *et al*, (2011) who used pulsed SILAC to quantify proteins with differential synthesis rates, and by Jovanovic *et al*, (2015), who combined protein turnover assessments with transcriptomics to study the impact of transcription, translation and protein degradation on protein homeostasis.

2.4.7   Protein turnover in Plants
All before mentioned SILAC studies were conducted in mammalian or bacterial systems, because these organisms are prototrophic and incorporate supplemented amino acids quickly, which results in next to complete labeling efficiencies. Autotrophic systems will still incorporate SILAC amino acids, yet get labeled at lower rates (Gruhler *et al*, 2005b). This complicates dynamic SILAC approaches for autotrophic organisms but does not make them generally inadmissible.

2.4.8   Implications of non-steady-state conditions
A major challenge of MS-based metabolic labeling studies is that they obtain only relative quantifications between the protein subpopulations with the old or new label. Changes in these labeling rates can thus be caused by 1) the decrease of the old labeled population by protein degradation or 2) the increase of the newly-labeled population by increased protein synthesis.

To overcome this problem, most studies mentioned before are conducted under the steady-state assumption. This requires that all physiological conditions stay constant over the course of the experiment including - most importantly - protein abundances. An assumption that is most likely

not valid for experimental conditions in which developmental processes (Doherty *et al*, 2005; Nelson *et al*, 2014a) or reaction to stimuli (Trotschel *et al*, 2012; Jovanovic *et al*, 2015) are analyzed.

Consequently, it is crucial in such non-steady-state metabolic labeling studies to disentangle protein synthesis and protein degradation by quantifying at least two of the three factors: abundance, synthesis, and degradation. This has been done for instance by quantifying protein abundances separately by additional MS assays (Piques *et al*, 2009), difference 2D gel electrophoresis (DIGE) (Li *et al*, 2012a) or multiplexed labeling schemes (Boisvert *et al*, 2012; Jovanovic *et al*, 2015). The latter two studies reached the highest proteomic coverage by spiking an unlabeled Lys0 protein population into Lys4- to Lys8-labeled protein turnover samples. The Lys0 population was then used as a reference for protein level changes of the combined Lys4 and Lys8 subpopulations. With this approach, protein turnover and abundance could be determined in one MS assay.

2.5   Plant defence mechanisms

In the present study, we stimulated the biological system with one of the best studied and one of the most important processes for plant systems, the defense mechanism against bacteria and microbes.

In the course of the never-ending molecular arms race of plants and microbial pathogens, a two layered plant innate immune defense system evolved. The first layer of defense response is induced by transmembrane pattern recognition receptors (PRRs, mostly receptor-like kinases (RLKs) or receptor like proteins (RLPs)) (Zipfel, 2014). They bind and respond to common classes of biomolecules from invasive species, such as bacterial cold shock proteins, elongation factors (e.g. EF-Tu), lipopolysaccharides, peptidoglycans, chitin and chitosan as well as the here employed flg22 peptide from bacterial flagellin (Felix *et al*, 1999; Zipfel *et al*, 2004; Das *et al*, 2015). These molecules are classified as pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) which, when sensed by PRRs, induce the so called PAMP-triggered immune response (PTI). PTI can also be induced by endogenous damage-associated molecular patterns (DAMPs) (Seong & Matzinger, 2004; Lotze *et al*, 2007)

Microbial pathogens can overcome detection of the PTI system, e.g. by translocating suppressive cytoplasmic effectors into the plant cell. To oppose this mechanism, Plants have a second layer of defense mechanisms that do not act at the cell membrane but inside of the cell (Schneider & Collmer, 2010). There, plant disease resistance (R) proteins sense the pathogen transduced virulence effectors and induce a secondary, usually stronger effector-triggered immune (ETI) response, which often culminates in a hypersensitive response and cell death. ETI can be induced by nucleotide-binding domain leucine-rich repeat-containing receptors (NLRs) (Zebell & Dong, 2015).

2.5.1 Response to flg22

To induce a strong PTI signal, we used one of the most potent stimulants known for plant systems, flg22. This 22 amino acid long peptide is the elicitor-active epitope of bacterial flagellin (Felix *et al*, 1999; Navarro *et al*, 2004). It is derived from the conserved N-terminal part of flagellin and induces a multilevel PTI response in plants. Flg22 is bound by the PRR FLS2 (flaggelin sensing 2, Chinchilla *et al*, (2006)), which responds to picomolar concentrations of flg22. After binding of flg22, FLS2 dimerizes with BAK1 (BRI1-associated kinase1), which leads to reciprocal phosphorylation of their kinase domains (Chinchilla *et al*, 2007). This initiates an intracellular calcium/calmodulin-mediated signaling cascade (Cheval *et al*, 2013) as well as a rapidly and transient activation of three major mitogen-activated protein kinases (MAPKs), MPK3, MPK4, and MPK6 (Pitzschke *et al*, 2009). The result of these cascades is a multilevel defense response including the production of reactive oxygen species (ROS), stomatal closure, ethylene and jasmonic acid production, salicylic acid production, callose deposition, and glucosinolate activation (Clay *et al*, 2009; Li *et al*, 2016). WRKY transcription factors are upregulated upon PTI signaling (Eulgem *et al*, 2000), including WRKY33, which modulates the biosynthesis of camalexin (Mao *et al*, 2011), the main phytoalexin of Arabidopsis (Schlaeppi *et al*, 2010). Camalexin is synthesized from tryptophan, which is also strongly upregulated upon PTI stimulation (Ishihara *et al*, 2008; Consonni *et al*, 2010).

2.6 Aim of the study and experimental design

In the presented study, we aimed to increase our understanding of gene expression regulation by quantifying in a systems biology approach the impact of transcription, translation, and protein degradation on the regulation of protein abundance in response to a strong stimulus. We therefore wanted to acquire genome-wide datasets for transcription, translation, protein abundance and protein turnover after elicitation of a defense response, and to compare them statistically to assess the size of the individual contributions of transcription, translation and protein degradation on the induced changes in protein abundances.

We aimed to conduct the experiment in a coherent and as uniform as possible plant system where we can induce a system-wide stimulus without tissue-specific variations. As we wanted to take the biological material for establishing all four datasets from exactly the same biological samples we would need rather large amounts of biomass. We therefore chose to conduct the experiment in undifferentiated cultured Arabidopsis thaliana suspension cells that can easily be cultivated in a uniform fashion at large scale.

The proteomic part of the experiment requires that the proteins of the cell cultures are labeled by heavy isotopologues of lysine. The cultures are therefore cultivated for 12 days in growth medium supplemented with either Lys4 or Lys8. After the labeling, the densely grown cell culture is split into two parts. One part of the samples is used for acquiring the data on transcription, translation and protein abundance, and the other part for the protein turnover study, where the previously established lysine-labeling is switched, Lys4 labelled cultures are labelled with Lys8 and vice versa. After a brief adjustment to the new cultivation conditions of 2 h, half of the cultures are treated with flg22, while the other half are mock-treated. The treatment was split between the two differentially labelled cultures, treating two Lys4-labelled cultures (L41 and L43) and one Lys8-labelled culture (L82), while one Lys4-culture (L42) and two Lys8-lableld cultures (L81 and L83) were mock-treated. The samples for the protein turnover part are collected over a timeline of 72 h, while the samples for establishing the transcription, translation and protein abundance data are collected over a timeline of 8 h.

The experimental design is outlined in Figure 2.1 together with the employed color scheme for the samples. All flg22-treated samples are colored in blue, while untreated control samples are colored in red. Cultures labeled with Lys4 are colored in a darker tone, while Lys8 cultures have a lighter tone. Individual samples are encoded by the initial labeling condition (L4 or L8), followed by the replicate number (1 to 3) and the time after treatment in hours (T0 to T72). Samples taken 0.5 h after treatment are encoded by T1 to facilitate computational analyses. We strive to employ this naming and color scheme throughout the thesis.
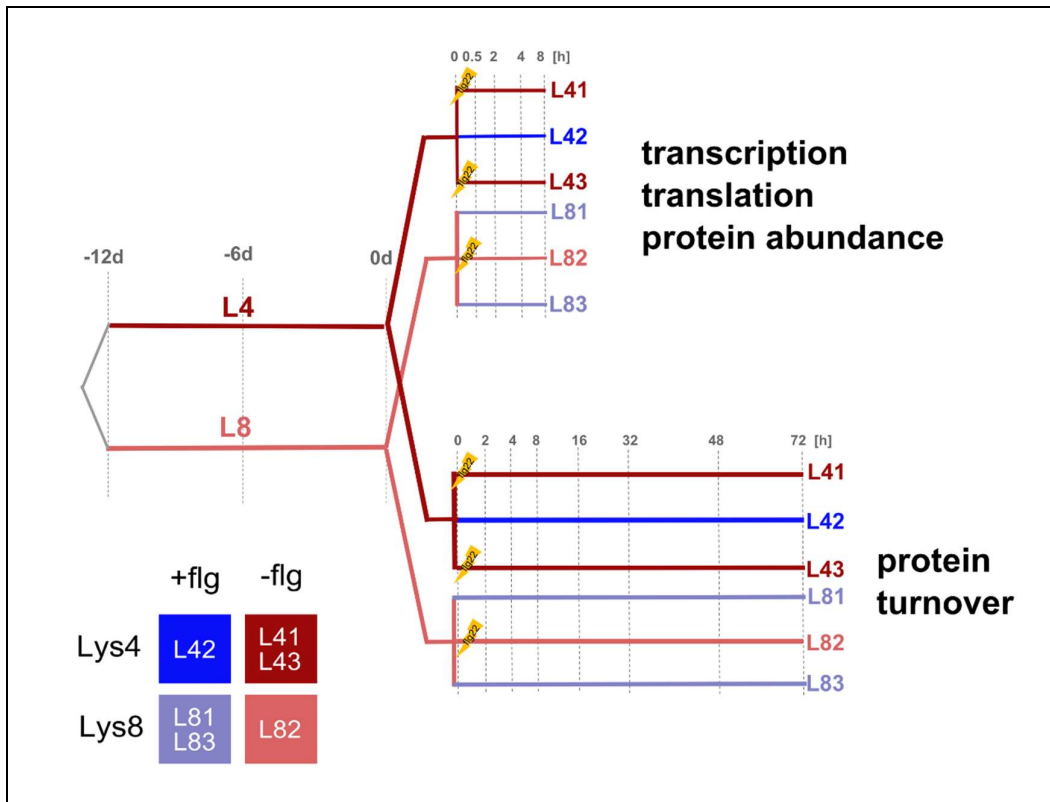
Figure 2.8: Experimental design and color scheme.

2.7 Contributions of the author

All experimental, statistical and computational analyses were conducted by the author with exception to the library generation and sequencing of the RNA-Seq samples (chapter 4) and the sequencing of the Ribo-Seq libraries (chapter 5), which were performed by Catharine Aquino and Dr. Maria Domenica Moccia from the Functional Genomics Center Zurich (FGCZ). All mass spectrometric measurements were performed by the author at the Functional Genomics Center Zurich under the excellent supervision of Dr. Bernd Roschitzki.

Chantal Ernst supported the analysis of differential translational efficiencies (chapter 5.6) during a semester project in our group.

3. MATERIAL AND METHODS

3.1 Suspension cell culture

Leaf derived Arabidopsis suspension cell cultures ecotype Columbia were cultured similarly to (Schütz *et al*, 2011). In brief, the cultures grew in liquid Murashige and Skoog (MS) medium with 3% (w/v) sucrose, and supplemented with 2mg/L Indole-3-acetic acid, 0.2 mg/L kinetin, 0.1mg/mL 2,4-dichlorophenoxyacetic acid and MS vitamins (2.0 µg/ml glycine, 100ug/ml myo-inositol, 0.50 µg/ml nicotinic acid, 0.50 µg/ml pyridoxine hydrochloride, 0.10 µg/ml thiamine hydrochloride). The cultures were grown under shaking (120 rpm) in Erlenmeyer flasks (filled to 30% of the flask's volume) at 22° C in the dark. Cultures were passaged every 7 days, transferring 10% of the growing culture to fresh medium, inoculating at >30 mg/mL.

During the SILAC experiments, the cultures grew in medium depleted to 70% of the original MS nitrogen that is 14 mM NH_4NO_3 an 7.5 mM KNO_3. Additionally, the cultures were supplemented daily with 350 µM isotopically labeled L-lysine, either Lys4 (4,4,5,5-H_2) or Lys8 ($^{13}C_6{}^{15}N_2$) (Cambridge Isotope Laboratories). After 12 days of labeling, including one passage at day 6, pooled Lys4 or Lys8 cultures were split into three replicates each and treated either with 100 nM flg22 (RP19986, GeneScript) or with water.

The same pool of Lys4 or Lys8 labeled cultures was used to inoculate the cultures used for protein turnover assessments. These cultures were washed two times with 1 volume of growth medium without any lysine, were transferred to fresh nitrogen depleted medium at a density of 100 mg/mL and got supplemented with 350 µM of the reciprocally labeled lysine. After an acclimation period of 2 h, the cultures were treated with either 100 nM flg22 or water. The supplementation of 350 µM heavy lysine was repeated every 24 h.

3.1.1 Assessing culture densities

There are numerous ways to determine the growth of a suspension cell culture (Mustafa *et al*, 2011), with even some non-invasive ones. As it provides good sensitivity in low concentrations independent of the shape of the culturing flask, we decided to determine the culture density by determining the culture's fresh weight using the following method: 3 times 300 or 500 µL cell culture were taken from the suspension cell culture and put into a pre-weighed tube, using a cut-off, wide-bore 1 mL pipet tip. The aliquots were spun for 1 min at >16'000 x g and all supernatant was taken off using a 200 µL pipet tip. The tube with the leftover cell pellet was weighed.

This method's accuracy can vary with the granularity of the cell line.

3.1.2   TTC reduction assay

The suspension cell culture's viability was determined by measuring the mitochondrial oxidoreductase activity by reducing TTC (2,3,5-triphenyl tetrazolium chloride) to insoluble formazan (Castro-Concha *et al*, 2006), as it has a low tendency for false positive results. In brief, culture aliquots of equal mass (50 mg usually), were spun 1 min at >2'000 x g (higher speed harms the cultures and gives false negative readouts), the supernatant was taken off, and the cells were washed once in 1 mL 50 mM phosphate buffer pH 7.5. Then, 125 mM TCC dissolved in 1 mL phosphate buffer were added. The cultures are shaken for 30 min at room temperature in the dark. The cells are pelleted for 1 min at >16'000 x g and all supernatant was discarded. The resulting formazan crystals were dissolved in 300 µL DMSO during 10 min incubation at 37° C. The formazan amount was quantified at 485 nm wavelength using a Nanodrop spectrophotometer. A sample without cells provides a negative control and 2mM DTT acted as a positive control.

3.2   Transformation of cultured Arabidopsis cells by Agrobacterium

The transformations were performed according to the protocol by Shimamoto et al. (2005). In brief, plasmid-containing agrobacterium cultures were grown at 27° C in 20 mL LB medium to an OD_{600} of 1.0, peletted, washed with 10 mL MS medium and added to a 50 mL, 3 day old Arabidopsis culture grown in MS medium supplemented with 1 mM lysine. After 48 h, 25 µg/mL cefotaxime is added to kill the agrobacterium. 48 h later, 15 mL of the transformed Arabidopsis culture is transferred to 35 mL of fresh MS medium together with 1 mM lysine, 50 µg/mL cefotaxime and the plasmid-compatible antibiotics. After 48 h of cultivation, the transformed culture are spread on petri dishes with jellified, ½-concentrated MS medium containing 1 mM lysine, 50 µg/mL cefotaxime and the plasmid-compatible antibiotics. After 1 week of growth, individual cell clumps were transferred to liquid MS medium with 1 mM lysine but without antibiotics for further propagation.

3.3   RNA-Seq

RNA was extracted using the Qiagen RNeasy Plant Mini Kit using the proprietary RLT buffer and an on-column DNase digestion following the manufacturer's instructions. The purified RNA was sequenced on an Illumina HiSeq 2000 (12 samples per lane, 125bp single-end reads) using the Illumina TruSeq stranded mRNA library preparation kit including an oligo-dT enrichment, RNA fragmentation, and random hexamer-primed cDNA synthesis.

The reads were processed with the *sushi* pipeline framework of the FGCZ Zürich (http://fgcz-sushi.uzh.ch/, Hatakeyama *et al*, 2016). The reads were de-multiplexed and trimmed by *Trimmomatic* (Bolger *et al*, 2014), resulting in 7 to 30 million reads per sample, and mapped in a stranded fashion with *STAR* (Dobin *et al*, 2013) to the TAIR10 genome (downloaded 07.09.2012). Reads aligned to TAIR10 genes and natural antisense transcripts (NATs, annotations created from (Matsui *et al*, 2008; Okamoto *et al*, 2010)) were counted with *featureCounts* of the R package *Subread* (Liao *et al*, 2014). Genes with less than 1 count per million mapped reads in more than 3 samples were discarded. Differentially expressed genes were identified with a generalized linear model tool from the *edgeR* package (Robinson *et al*, 2009; Chen *et al*, 2014). We set the lysine background as the

blocking factor and testing for changes at any time point between all flg22 and control cultures. We used TMM normalization (Robinson & Oshlack, 2010) and a tag-wise estimation of the dispersion.

3.4 Ribo-Seq

The ribosome-mRNA complexes were enriched similarly to Juntawong *et al*, (2013). In brief, ~5 mL Arabidopsis suspension cell culture were incubated for 1 min at room temperature with 100 µg/mL cycloheximide. The culture was centrifuged for 1 min at 2'000 x g, the medium was discarded and the cells (>1 g) were snap frozen. The cells were ground to a powder in liquid nitrogen and were thawed in 1 volume of polysomal extraction Buffer (PEB, 0.2 M Tris pH 9.0, 0.2 M KCl, 25 mM EGTA, 35 mM $MgCl_2$, 1 % (w/v) polyoxyethylene(23)lauryl ether (Brij-35), 1 % (v/v) Triton X-100, 1 % (v/v) octylphenyl-polyethylene glycol (Igepal CA 630), 1% (v/v) polyoxyethylene sorbitan monolaurate 20 (Tween 20), 1 % Sodium Deoxycholate (DOC), 5 mM DTT, 2x Roche cOmplete Protease Inhibitor, 100 µg/mL cycloheximide). The extraction was incubated for 20 min at RT. Cell debris was precipitated by centrifuging twice for 15 min at >16'000 x g at 4°C, transferring the supernatant to a clean tube each time. The spun supernatant was loaded onto a 1 mL 60 % sucrose cushion (1.75 M sucrose, 0.4 M Tris pH 9.0, 0.2 M KCl, 5 mM EGTA, 35 mM $MgCl_2$, 100 µg/mL cycloheximide. The sucrose was added directly to a 10x stock solution of the salts, only few additional water was needed to bring the solution to the full volume. The sucrose was dissolved by heating it up in a <70° C water bath). The sucrose cushion was overlaid with the extract and were spun for 150 min at 80'000 rpm in a MLA-80 rotor (Beckman Coulter). The cushion and supernatant were discarded and the clear ribosome-mRNA pellet was further processed with the ARTseq ribo profile mammalian kit (Illumina Epicentre). In brief, the pellet (~ 20 µg RNA) was dissolved in 200 µL 1x ARTseq mammalian polysome extraction buffer and digested for 2 h at room temperature by addition of 50 U TruSeq nuclease (ARTseq kit, Illumina/Epicenter). The RNA digestion was stopped with 50 U SUPERase•In RNase Inhibitor (Life Technologies), and purified by RNA Clean & Concentrators-25 (Zymo Research). 5 µg of the purified ribosome protected fragments (RPFs) were depleted of rRNAs using the Ribo-Zero plant leaf kit (Illumina), repeating the magnetic separation step of the rRNA bound beads.

From the purified RPFs, a sequencing library was prepared following the instructions of the ARTseq ribo profile kit (Illumina/Epicenter), continuing the protocol with the PAGE purification of the RPFs. The cut-out PAGE gel pieces were extracted by the crush-and-soak method (Buratowski & Chodosh, 2003), incubating the crushed gel pieces for 2 h at 30° C and separating the gel pieces from the eluate by spinning for 2 min at 2'000 rpm in a QiaShredder spin column. It was deemed useful in our approaches to clean up all PCR amplified libraries by denaturing PAGE to separate adapter amplicons, as discussed in the ARTseq protocol.

The final RPF libraries were sequenced with 125 bp single-end reads on an Illumina HighSeq 2000 (12 samples per lane).

The Illumina reads were processed equally to the RNA-Seq reads (chapter 3.2). The only difference being that we used Bowtie2 (Langmead *et al*, 2009) to align the reads to the TAIR 10 genome using the default settings (minimum read length of 20 nt, and a minimum average alignment quality of 10). As discussed in chapter 5.3.4, we excluded all reads that map to the following genes containing non-reproducible artifact reads: AT2G01021, AT3G55850, AT1G70185, AT3G52700, AT2G23880, AT3G55850, AT1G70185, ATMG00030. We excluded all lowly transcribed genes that did not have more than 5 counts per million mapped reads in more than 8 samples.

3.5 Proteomics

3.5.1 Protein extraction and FASP

All cell culture samples used in proteomics analyses were centrifuged for 1 min at 4'000 x g, spent supernatant was discarded, and the pellet was washed once in 1 volume of 50mM phosphate buffer pH 7.5, before being snap frozen in liquid nitrogen. The cell pellet was ground in liquid nitrogen and taken up in 1.5 volumes of 40 mM Tris base, 4 % SDS, 2x cOmplete Protease Inhibitor Cocktail (Roche). The extracts were incubated for 30 min at room temperature and were cleared of cell debris by spinning twice for 10 min at >16'000 x g. After determining the protein concentration by a Pierce BCA assay (Thermo Fischer Scientific), the extract was reduced by adding 50 mM DTT. The protein extracts from the protein abundance experiment were mixed with the same amount of protein from a reference protein mixture, as described in chapter 6.

Up to 200 µg protein extract were digested each with 4 µg (1:50, 2 % of the protein input) of trypsin (Promega) following the FASP procedure (Wisniewski *et al*, 2009) using Microspin YM-30 concentrators (Millipore). The eluted peptides were dried in a vacuum concentrator and purified with Finisterre C18 SPE columns (Teknokroma). In brief, the SPE columns were wetted with 1 mL MeOH, equilibrated with 1 mL 0.1 % trifluoroacetic acid (TFA) in 60 % ACN and 2 ml 0.1 % TFA. The dried peptides were dissolved in 0.6 mL 0.1 % TFA, loaded onto the column, washed two times with 1 mL 0.1 % TFA and eluted with 0.8 mL 0.1% TFA in 60 % ACN. The purified peptides were dried again in a vacuum concentrator at 37°C.

3.5.2 Peptide separation by HILIC

The trypsinized peptides were separated into 5 fractions using an Agilent 1200 series HPLC system with a YMC-Pack Polyamine II, 250 x 3.0 mm, 5 µm, 120 Å HILIC column. The sample were dissolved in 100 µL Buffer A (75% ACN, 8 mM KH_2PO_4, pH 4.5) and separated with Buffer B (5% ACN, 100mM KH_2PO_4, pH 4.5). The samples got injected in 100% A at a flow rate of 300 µL/min. From 7.5 min to 37.5 min, Buffer A decreased linearly to 50%. Till 42.5 min B increased linearly to 100%; during 1 min the buffer were switched to 100% A to re-equilibrate the column till 60 min. The eluate of these HILIC runs were separated into 24 fractions of 750 µL each. These fractions were pooled to 5 samples so that each sample contained an equal amount of peptides according to the UV absorption. We pooled the HILIC fractions 6-10, 11-12, 14-15, and 16-22. Fraction 13 contained the apex of the HILIC elution curve and was not pooled further. These combined fractions were dried at 37 °C in a vacuum concentrator and desalted by Finisterre C18 SPE columns using 4 washes of 1 mL 0.1% TFA each.

3.5.3 Mass spectrometry

The fractionated and desalted peptide pools were analyzed by LC-MS using an Eksigent nanoLC-Ultra and a Thermo Orbitrap Fusion mass spectrometer. The LC was performed with a 150 x 0.075 mm column heated to 50° C, packed in-house with 1.9 µm ReproSil-Pur C18 AQ particles with 120 Å pores (Dr. Maisch GmbH, Ammerbuch, Germany).

The samples (1 µg peptides) were separated with the following linear gradients of Buffer A (0.1% FA in water) and Buffer B (0.1% FA in Acetonitrile) using a conserved flow of 300 nl/min: B increases linearly from 3% to 25% till 60 min; then B increases over 10 min to 35%; then to 97% over 5 min; then the column was washed for 5 min at 97% B and was re-equilibrated at 3% B during another 5

min. For the most complex samples (the middle fractions of the HILIC: fraction 11-12, 13, 14-15), the first linear elution phase to 35% B was extended from 60 min to 90 min.

The Orbitrap mass analyzer was set to a resolution of 240'000 FWHM at a scan range of 300-1500 m/z. The injection time was capped to maximally 50 ms. Precursors for MS2 scans were selected using the Monoisotopic Precursor Selection filter, including charge states 2 to 7, and using an intensity threshold of minimally 5E3. Starting with the most intense precursors, a maximum of 250 ms injection time was allowed for the ion trap. The precursors were isolated using a 1.6 m/z window and got fragmented at 30% HCD collision energy. The ion trap ran in "top speed" setting at the "rapid" scan rate. After a single fragmentation, precursors were excluded from fragmentation for the next 25 sec.

3.5.4 MaxQuant

The proteomic data were analyzed with *MaxQuant* Version 1.5.3.3 (Cox & Mann, 2008), with disabled re-quantification and 1 min "match between runs" using a 20 min run alignment window. The spectra were compared against a trypsin-digested Arabidopsis TAIR10 database (downloaded on 17.10.2011) including common contaminants. A reverted decoy database of these sequences was added. The comparison algorithm allowed for variable acetylation of the protein N-terminus and methionine oxidation, while carbamidomethylation of cysteines was set as a fixed modification. The mass accuracy was set for 20 ppm for the orbitrap, and 0.5 Da for the ion trap. A maximal false discovery rate (FDR) of 1% was allowed on the peptide level and 1% on the protein level. Quantification events were limited to unmodified, unique peptides, requiring at least 2 quantification events. The "re-quantify" option was not enabled, as each sample was mixed with a reference sample that should contain all peptides and every peptide should be paired to the unlabeled Lys0 isoform, reaching 15% of the labeled peak's intensity. MaxQuant was used with a multiplicity of 2 for the protein abundance data set (Lys4 or Lys8 as the labeling state) and with a multiplicity of 3 for the protein turnover dataset (no-label, Lys4, or Lys8). For quantifying a protein, a minimum of 2 peptides in that samples was required (called "Min. Ratio Count" setting in MaxQuant).

3.6 Bioinformatics

All analyses were performed with R (version 3.2.5, (Environment & for statistical computing. Vienna, 2015)) with the libraries given in the results chapters. All turnover models were fitted with the nls() function of the base R package. All R scripts are available upon request.

3.7 Gene set enrichment

Gene ontology biological process (GO) enrichments were performed with *PANTHER* (version 10.0) (Mi *et al*, 2016), using the statistical overrepresentation test with Bonferroni correction for multiple testing and a 5% adjusted p-value cutoff (Mi *et al*, 2013). The reference gene list to test for enrichment were taken from all quantified genes of the particular dataset. The GO enrichments were visualized with VirtualPlant (Katari *et al*, 2010).

4. TRANSCRIPTION

In the following, I will present the results of the experiment divided into the four chapters, one for each genome-wide dataset including its interpretation and comparison to the other obtained datasets.

4.1 Experimental design and quality control

We used RNA-Seq to assess the transcriptional response of Arabidopsis suspension cell culture AtB to flg22-treatment. We took samples from three flg-22 treated and three control cultures before treatment and 0.5, 2, 4 and 8 hours after treatment (Figure 4.1A,B). We extracted the RNAs, enriched for mRNAs by oligo-dT hybridization, fragmented the mRNAs with divalent cations under elevated temperature, transcribed the RNA to cDNA, and sequenced 125 nt long strand-specific reads with an Illumina HiSeq sequencer. Library preparation and sequencing were performed by Catharine Aquino from the Functional Genomics Center Zürich (FGCZ). The quality of the sequencing was verified by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All samples showed a good mean quality per read and per base of >36 points on the Phred score (negative logarithm of base 10 of the base-calling error probability, Figure 4.1C,D).

The reads were trimmed and mapped strand-specifically to the TAIR10 genome. On average, 14.8 million reads were mapped for each sample, and each sample had at least 5 million mapped reads (Figure 4.1E). Reads mapping to open reading frames (ORFs) were counted and normalized to the total number of mapped reads, resulting in *counts per million mapped reads* (cpm) values for each ORF. ORFs with less than 1 cpm in at least 3 samples were discarded. This resulted in transcript expression values for 18'892 ORFs with on average of 51 cpm over 26 samples (Figure 4.1F). This data matrix was used to assess the transcriptional response of Arabidopsis to flg22 treatment.
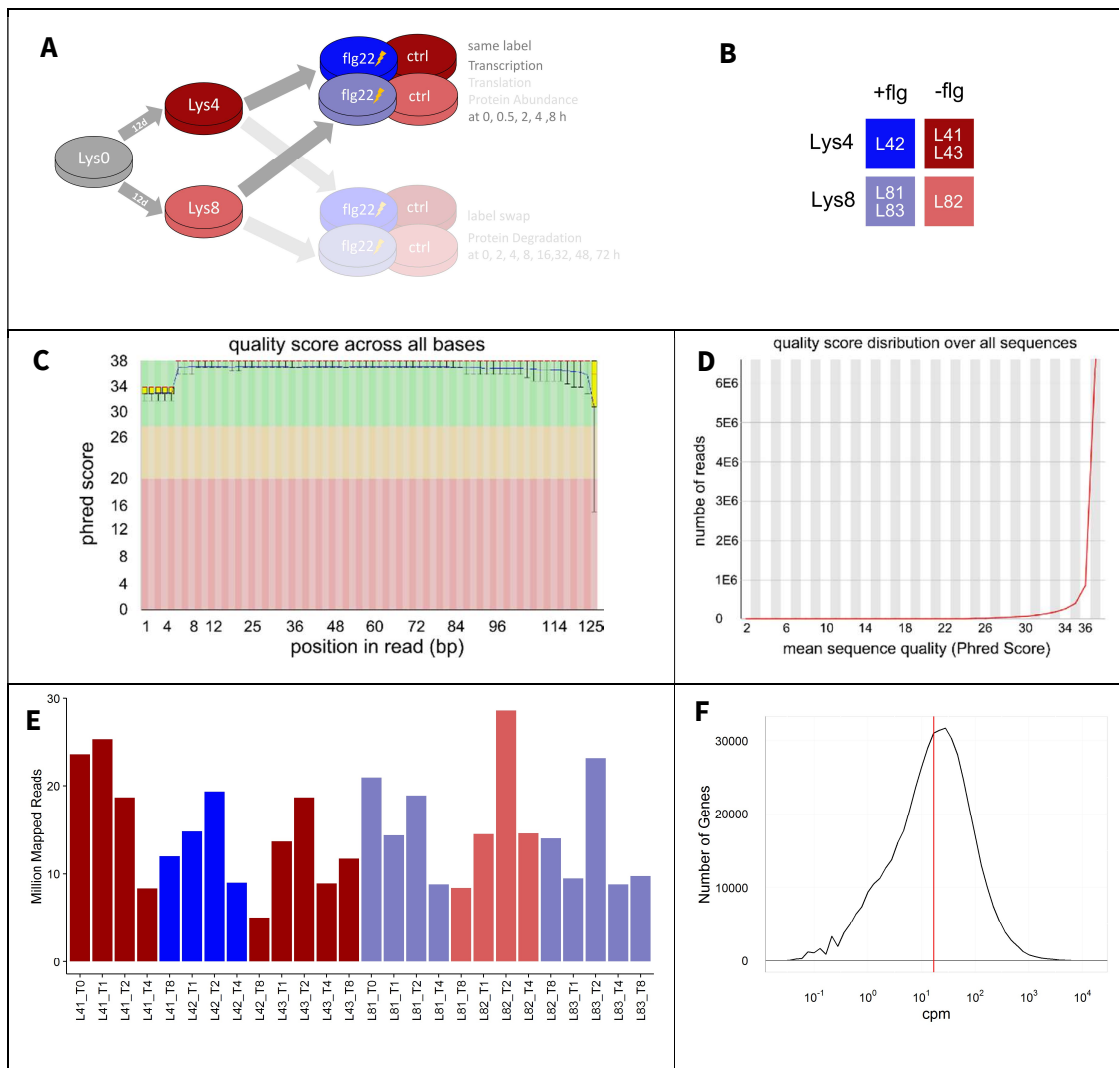
Figure 4.1: RNA-Seq Quality Control.
 A: Experimental design. B: Naming and color scheme of the samples. All flg22-treated samples are colored in blue, while untreated control samples are colored in red. Cultures labeled with Lys4 are colored with darker colors, while Lys8 cultures are colored lighter. We strive to employ this color scheme throughout all plots of the thesis. C: Quality score per base of a representative RNA-Seq sample. It is typical for an Illumina sequencer to have a slightly lowered quality score for the very first bases. D: Quality score per read of a representative sequencing sample. E: Number of reads per sample that mapped to TAIR10 open reading frames. F: Frequency plot of the average cpm per gene. The median expression value is indicated by the red line.

4.2 Dimensionality reduction visualizations

To assess such a large quantitative transcriptional dataset in a nonparametric way, dimensionality reduction methods can be very helpful. Both the principal coordinate analysis (PCoA, Figure 4.2A) and the multidimensional scaling plot (MDS, Figure 4.2B) display the same pattern. Replicates of the same treatment and time point cluster closely together, while both time and treatment separate the clusters. In the PCoA, the first principal coordinate (33.24 % of the variance) mostly

indicates changes in time, while the second coordinate (18.87 % of the variance) indicates changes induced by the treatment.

Samples treated with flg22 show especially at the first time point (0.5 h after treatment, named T1 here) large differences to their corresponding control samples. The MDS displays these differences particularly striking as it restricts the analysis to the 500 transcripts with the largest standard deviation over all samples. The flg22-treated samples collected at 4 h and 8 h after treatment (T4 and T8) group considerably closer to their respective untreated clusters, indicating an attenuation of the treatment effect.

A consistent difference between the Lys4- and Lys8-labelled samples can be seen. Especially the MDS indicates a small but consistent shift between these two labeling backgrounds (Figure 4.2C), which is discussed in more detail in chapter 4.7.
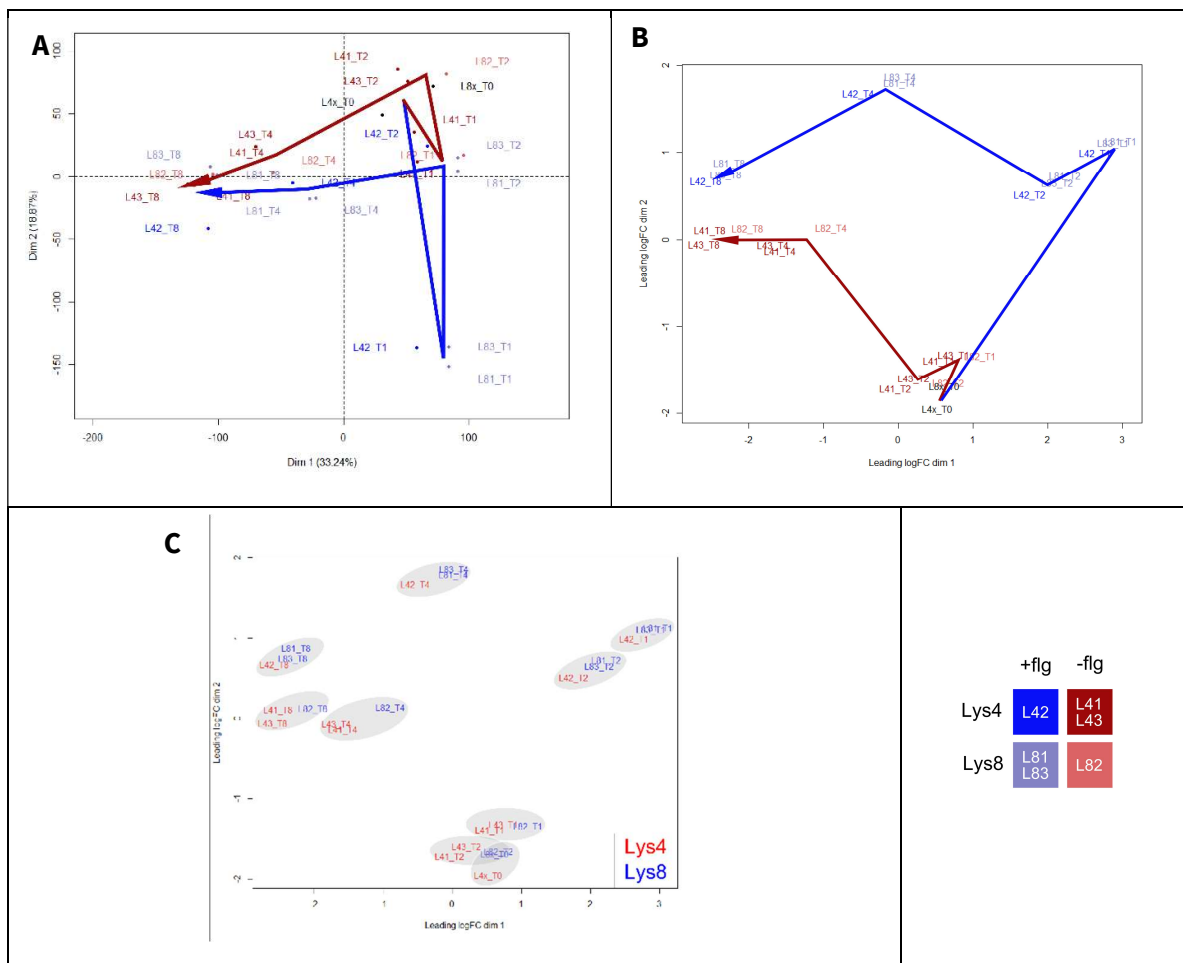
Figure 4.2: Dimensionality reduction plots of the transcription dataset.
A: Principal coordinate analysis (PCoA). The position of the samples is indicated by the dot. B: multidimensional scaling plot (MDS) calculated by using the version available in *limma* (Smith, 2005). The experimental timeline is indicated by connecting with an arrow the centers of each replicate cluster. The samples are colored according to the legend to the bottom right.

4.3 Flg22-induced differential transcription

We tested for differential expression of genes with the *edgeR* framework (Robinson *et al*, 2009). To estimate the dispersion, we used the tag-wise, that is gene-wise, dispersion (Figure 4.3A), which is recommended for a multi-parametric experimental design (Chen *et al*, 2014). We included the lysine labeling background of the cultures as a regression coefficient of the statistical test (Figure 4.3B) and the total number of reads of each sample (LibrarySize) so that the gene-specific number of reads (counts) depends on the lysine labeling, treatment and sampling time:

$$counts \sim LibrarySize + LysineBackground + Time + Treatment$$

We tested for differential expression over all time points (Figure 4.3C), or each time point separately (Figure 4.3D-G). We observed the highest number of differentially regulated transcripts at the first time point, 0.5 h after treatment. The more time passed after the treatment, the fewer differentially regulated genes were detected. In the following text, we will discuss the results from the test for differential expression at any time point.
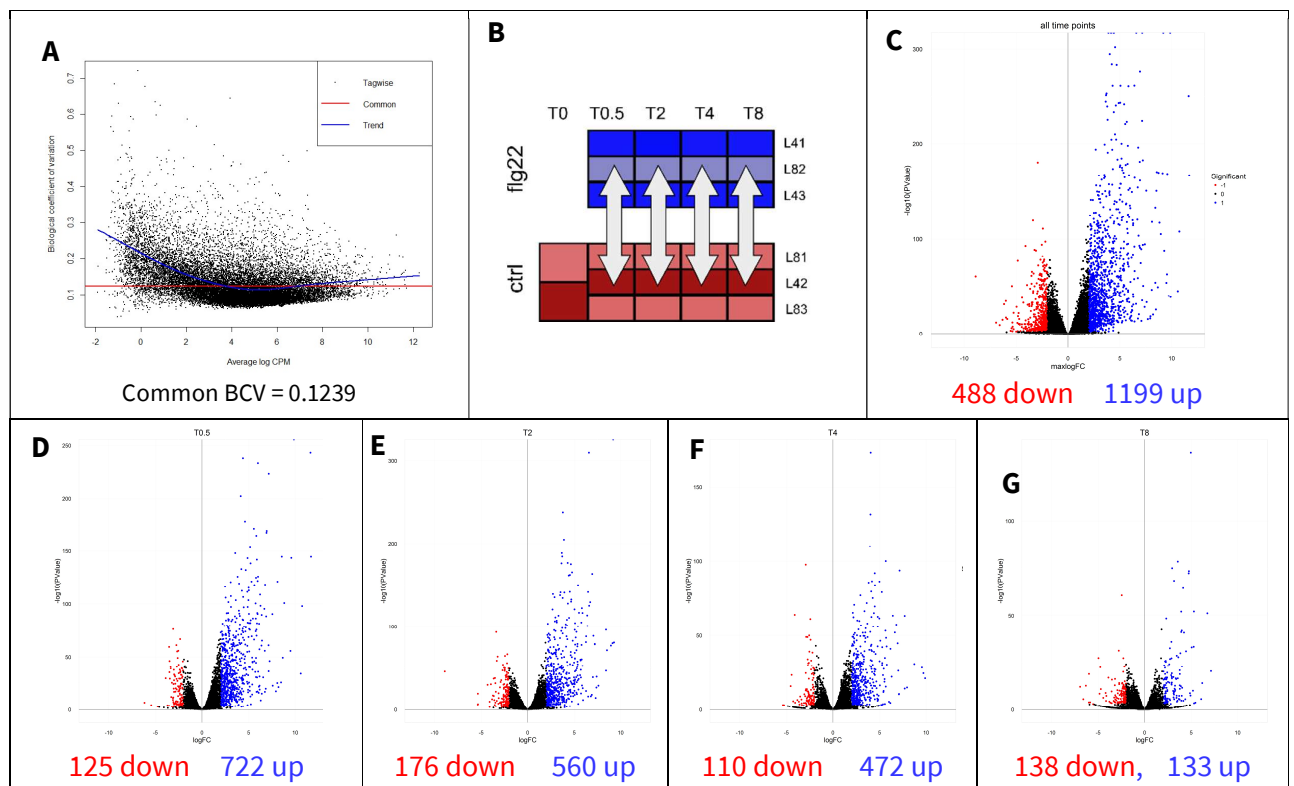
Figure 4.3: *EdgeR* tests for flg22-induced differentially transcribed genes.
A: Plot of the biological coefficient of variation (BCV) against the average expression in cpm to estimate the dispersion. We chose the tag-wise dispersion estimate for our statistical analysis. B: Schema of the statistical design. C-G: Volcano plots of the differentially expressed genes over all time points (C) or at the individual time points: 0.5 h (D), 2 h (E), 4 h (F), 8 h (G) after treatment. In C, the genes are plotted on the x-axis according to the highest log2-transformed fold-change at any time point (maxlogFC). The numbers of differentially expressed genes are given below the volcano plots. The threshold for differential transcription was set at Bonferroni-Holm-adjusted p-value < 0.01 and |log2(fold change)| > 2. In total, 18'894 genes were tested statistically.

The up- and downregulated response differ in their consistency and / or length of the response: While only 4.3 % (38 transcripts) of all downregulated transcripts are found significantly downregulated in more than one time point (Figure 4.4A), 60.2 % of all significantly upregulated transcripts (580 transcripts) are determined to be upregulated in more than a single time point (Figure 4.4B), 48 of which were upregulated over all four time points.

Figure 4.4: Venn diagram of the significantly downregulated (A) or upregulated (B) genes.

The 1199 genes with increased transcription under flg22 treatment were enriched in a plethora of GO categories (Table 4.1, Figure 4.5), of which nearly all can be related to plant defense responses. The 488 downregulated genes were associated with only one GO term (*water homeostasis*, 5 genes, adjusted p-value 0.011).

Table 4.1: Gene ontology biological process enrichment for the 1199 genes with increased transcription upon flg22 treatment (adjusted p-value <5 %).
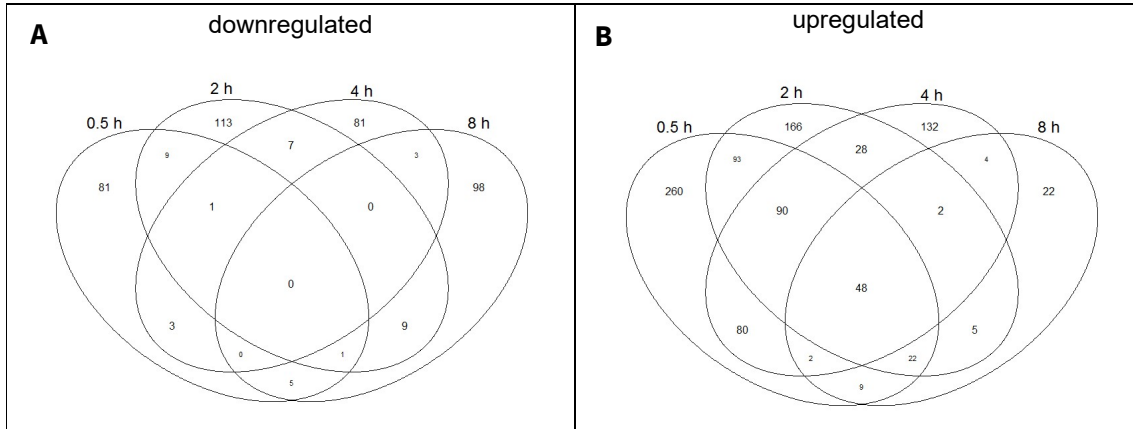Underrepresented GO categories, where the number of transcripts were less frequently observed than expected by pure chance, are listed at the bottom of the table, separated by a black bar. In total, 134 GO terms were significantly enriched and only the child nodes of the GO categories are listed. The enrichment analysis was performed with Panther (Mi *et al*, 2013) and all expressed genes were used as the reference list.

| GO BIOLOGICAL PROCESS	# IN REFERENCE	# IN QUERY	# EXPECTED IN QUERY	FOLD ENRICHMENT	ADJUSTED P-VALUE
indole-containing compound catabolic process	6	5	0.26	19.4	1.62E-02
defense response by callose deposition in cell wall	15	10	0.64	15.52	3.91E-06
camalexin biosynthetic process	9	6	0.39	15.52	7.03E-03
chorismate metabolic process	11	6	0.47	12.7	2.18E-02
response to chitin	108	54	4.64	11.64	3.38E-35
tryptophan biosynthetic process	20	8	0.86	9.31	7.21E-03
induced systemic resistance	28	10	1.2	8.31	1.22E-03
defense response to bacterium	281	52	12.07	4.31	1.77E-14
cellular response to hypoxia	23	8	0.99	8.1	1.97E-02
cellular response to decreased oxygen levels	26	9	1.12	8.06	5.71E-03
sulfur compound catabolic process	40	11	1.72	6.4	4.16E-03
response to karrikin	104	28	4.47	6.27	1.20E-10
toxin catabolic process	43	11	1.85	5.95	8.22E-03
glutathione metabolic process	56	13	2.41	5.4	3.22E-03
cellular modified amino acid metabolic process	103	17	4.42	3.84	8.49E-03
hydrogen peroxide catabolic process	70	16	3.01	5.32	2.55E-04
response to wounding	179	39	7.69	5.07	1.39E-12
glucosinolate metabolic process	82	17	3.52	4.83	4.04E-04
cellular amino acid biosynthetic process	152	31	6.53	4.75	6.75E-09
phenylpropanoid biosynthetic process	104	21	4.47	4.7	2.40E-05
phenylpropanoid metabolic process	130	27	5.58	4.83	1.12E-07
lignin metabolic process	69	13	2.96	4.39	2.93E-02
response to jasmonic acid	173	31	7.43	4.17	1.60E-07
response to salicylic acid	164	28	7.05	3.97	3.68E-06
regulation of defense response	180	30	7.73	3.88	1.70E-06
flavonoid biosynthetic process	110	18	4.73	3.81	5.03E-03
ethylene-activated signaling pathway	135	22	5.8	3.79	4.28E-04
response to oxidative stress	360	55	15.47	3.56	5.91E-12
response to salt stress	415	48	17.83	2.69	3.71E-06
cellular response to acid chemical	342	34	14.69	2.31	2.09E-02
defense response to fungus	425	39	18.26	2.14	2.91E-02
response to inorganic substance	687	61	29.51	2.07	3.66E-04
protein phosphorylation	828	72	35.57	2.02	6.35E-05
oxidation-reduction process	1293	93	55.55	1.67	3.14E-03
peptide biosynthetic process	537	6	23.07	0.26	5.00E-02
amide biosynthetic process	554	6	23.8	0.25	2.84E-02
DNA metabolic process	403	2	17.31	< 0.2	9.82E-03
RNA processing	412	1	17.7	< 0.2	7.28E-04
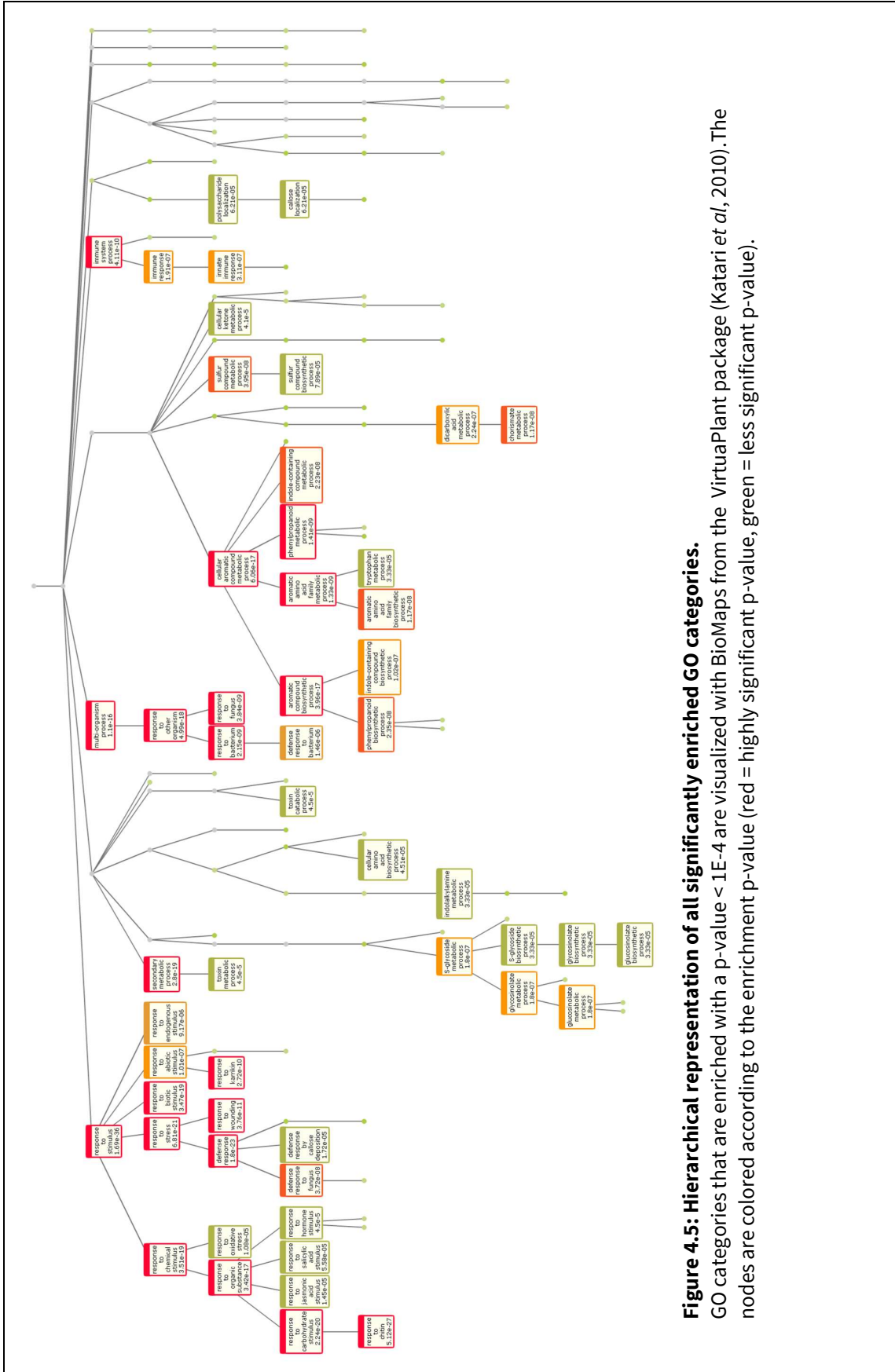
Figure 4.5: Hierarchical representation of all significantly enriched GO categories.
GO categories that are enriched with a p-value < 1E-4 are visualized with BioMaps from the VirtuaPlant package (Katari *et al*, 2010).The nodes are colored according to the enrichment p-value (red = highly significant p-value, green = less significant p-value).

4.4   Clustering transcriptional response to flg22-treatment

To disentangle the complex response to flg22-treatment, we performed a hierarchical clustering of the flg22-treatment over control fold-changes of all significantly changing genes. The 1'199 upregulated genes were clustered into 3 groups (Figure 4.6A): Cluster *1up* comprises transcripts that are upregulated at 0.5 h and return quickly to basal levels, while clusters *2up* and *3up* showed more long lasting, later-stage responses.

All three clusters are enriched for defense response genes. The GO categories *plant-type hypersensitive response* and *regulation of cellular defense response* were only enriched in the early response cluster *1up*. Cluster *2up* is enriched in *hydrogen peroxide catabolic process*, indicating compensatory mechanisms to the ROS produced by the defense mechanism. Metabolic responses such as tryptophan, phenylpropanoid, and glutathione biosynthetic processes are found predominantly in the mid- to late-stage clusters.

The downregulated genes were clustered into 4 groups (Figure 4.6B): cluster 3down contains genes that respond only at the first time point, cluster *2down* contains mostly late stage responding genes, and cluster *1down* contains a large group of variably responding genes. Cluster *4down* contains two genes (AT1G49200 and AT4G20270) with 0 cpm in one control group, leading to atypical high fold-changes. The separate clustering of these genes is therefore not biologically relevant.

The downregulated genes are enriched for only few GO terms. Cluster *1down* contains the 5 genes associated with *water homeostasis*. Interestingly, the early-stage downregulated genes from cluster 3down are enriched for developmental and transcriptional regulation function. The same enrichment did emerge when analyzing all genes downregulated 0.5 h after treatment, while the other time points were not significantly enriched for any GO term.
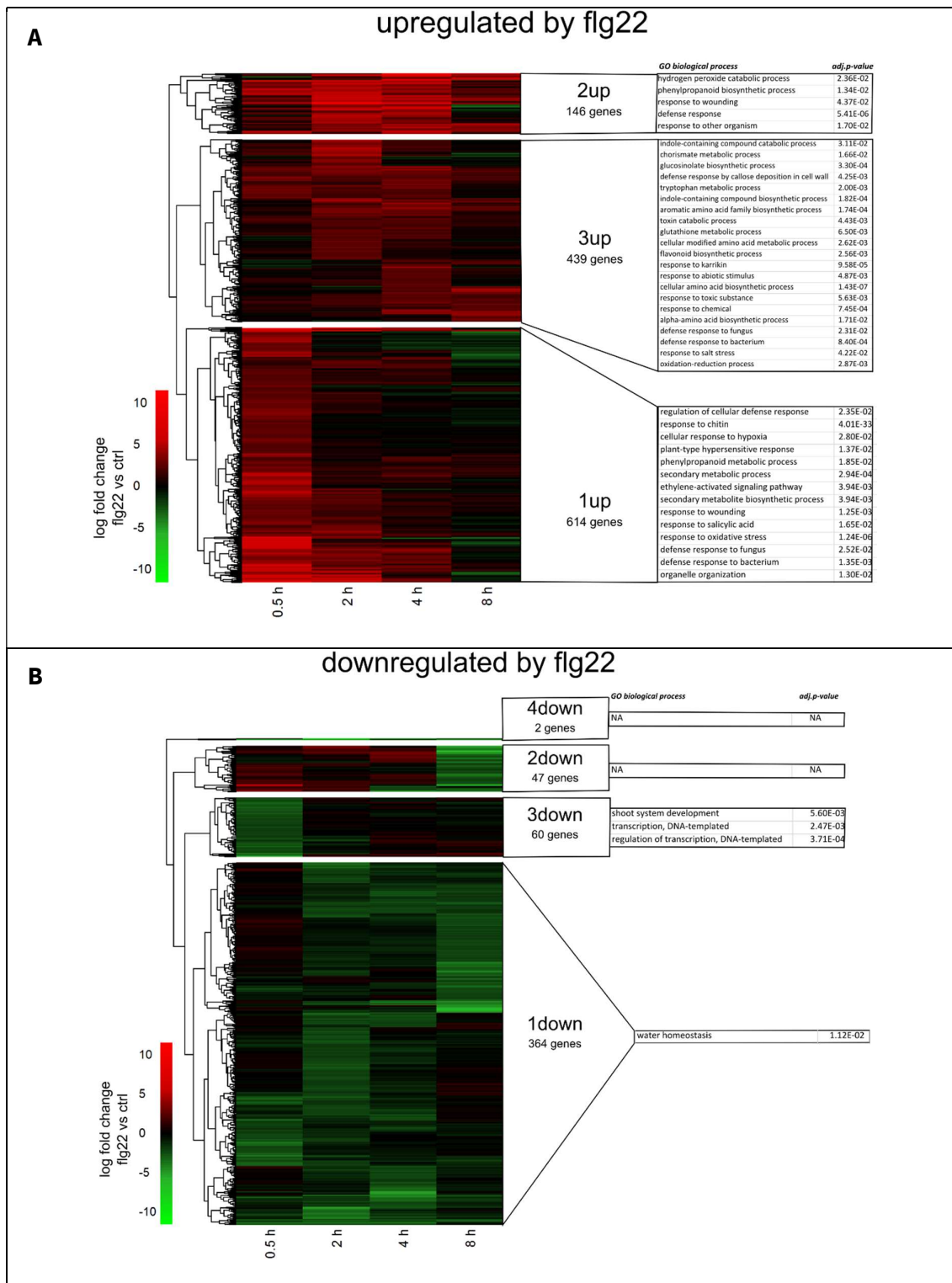
upregulated by flg22

A

2up 146 genes

GO biological process	adj.p-value
hydrogen peroxide catabolic process	2.36E-02
phenylpropanoid biosynthetic process	1.34E-02
response to wounding	4.37E-02
defense response	5.41E-06
response to other organism	1.70E-02

3up 439 genes

indole-containing compound catabolic process	3.11E-02
chorismate metabolic process	1.66E-02
glucosinolate biosynthetic process	3.30E-04
defense response by callose deposition in cell wall	4.25E-03
tryptophan metabolic process	2.00E-03
indole-containing compound biosynthetic process	1.82E-04
aromatic amino acid family biosynthetic process	1.74E-04
toxin catabolic process	4.43E-03
glutathione metabolic process	6.50E-03
cellular modified amino acid metabolic process	2.62E-03
flavonoid biosynthetic process	2.56E-03
response to karrikin	9.58E-05
response to abiotic stimulus	4.87E-03
cellular amino acid biosynthetic process	1.43E-07
response to toxic substance	5.63E-03
response to chemical	7.45E-04
alpha-amino acid biosynthetic process	1.71E-02
defense response to fungus	2.31E-02
defense response to bacterium	8.40E-04
response to salt stress	4.22E-02
oxidation-reduction process	2.87E-03

1up 614 genes

regulation of cellular defense response	2.35E-02
response to chitin	4.01E-33
cellular response to hypoxia	2.80E-02
plant-type hypersensitive response	1.37E-02
phenylpropanoid metabolic process	1.85E-02
secondary metabolic process	2.94E-04
ethylene-activated signaling pathway	3.94E-03
secondary metabolite biosynthetic process	3.94E-03
response to wounding	1.25E-03
response to salicylic acid	1.65E-02
response to oxidative stress	1.24E-06
defense response to fungus	2.52E-02
defense response to bacterium	1.35E-03
organelle organization	1.30E-02

log fold change flg22 vs ctrl: 10, 5, 0, -5, -10

0.5 h | 2 h | 4 h | 8 h

downregulated by flg22

B

4down 2 genes

GO biological process	adj.p-value
NA	NA

2down 47 genes

NA	NA

3down 60 genes

shoot system development	5.60E-03
transcription, DNA-templated	2.47E-03
regulation of transcription, DNA-templated	3.71E-04

1down 364 genes

water homeostasis	1.12E-02

log fold change flg22 vs ctrl: 10, 5, 0, -5, -10

0.5 h | 2 h | 4 h | 8 h

Figure 4.6: Gene-wise hierarchical clustering of the log2-transformed fold changes of all differentially changing transcripts.
Clustering of the 1'199 upregulated genes (A) and 488 downregulated genes (B). The number of clusters was chosen manually to split the genes into groups specific for temporal responses. Significantly enriched biological process GO terms for each cluster are listed with their corresponding adjusted p-values.

4.5   Response of bacterial defence genes to flg22 treatment

Flg22 is a well-studied PAMP, so we verified our biochemical and statistical analysis with defense response markers, such as the flagellin receptor-like kinase 1 (FRK1), and the transcription factors WRKY29, WRKY22, WRKY30, MYB15, MYB51 and PEP3 (Asai *et al*, 2002; Lee *et al*, 2011). All of these factors are highly significantly upregulated upon flg22 treatment (Figure 4.7A-G). WRKY30 was expressed too low to be included in the statistical analysis, yet the expression plot also shows a clear transcriptional induction by flg22 (Figure 4.8D).
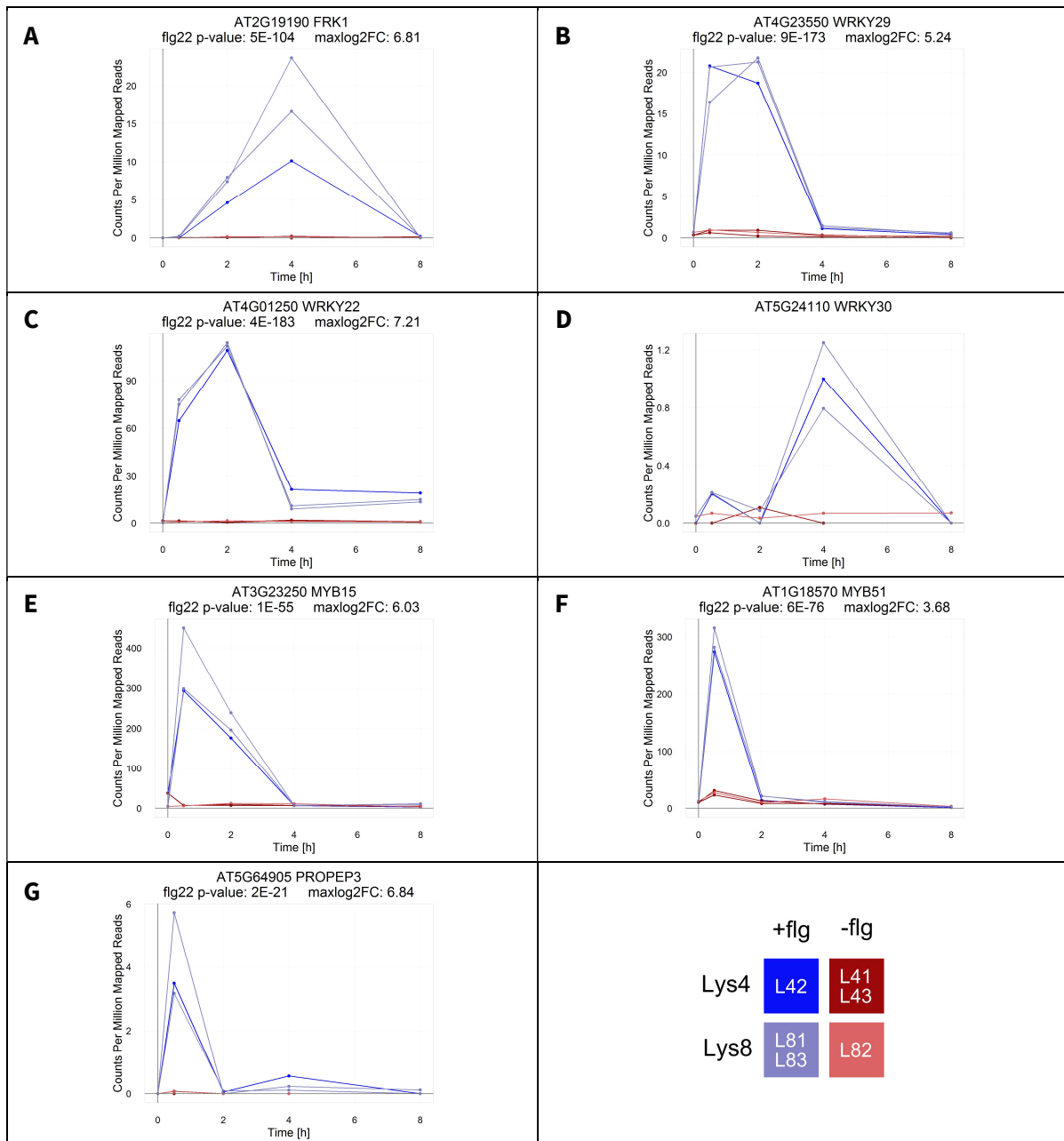
Figure 4.7: Transcriptional expression of well–characterized defense response markers. maxlog2FC represents the highest flg22-induced log2-tranformed fold-change at any measured time point.

Flg22 treatment induces only small to no changes for infection or systemic acquired resistance markers such as PR1 and NPR1 (Cao *et al*, 1994) and disease markers or markers such as PAD4 and EDS1 (Feys et al, 2001). EDS1 is stronger upregulated as it is also involved in the plant hypersensitive response (HR).
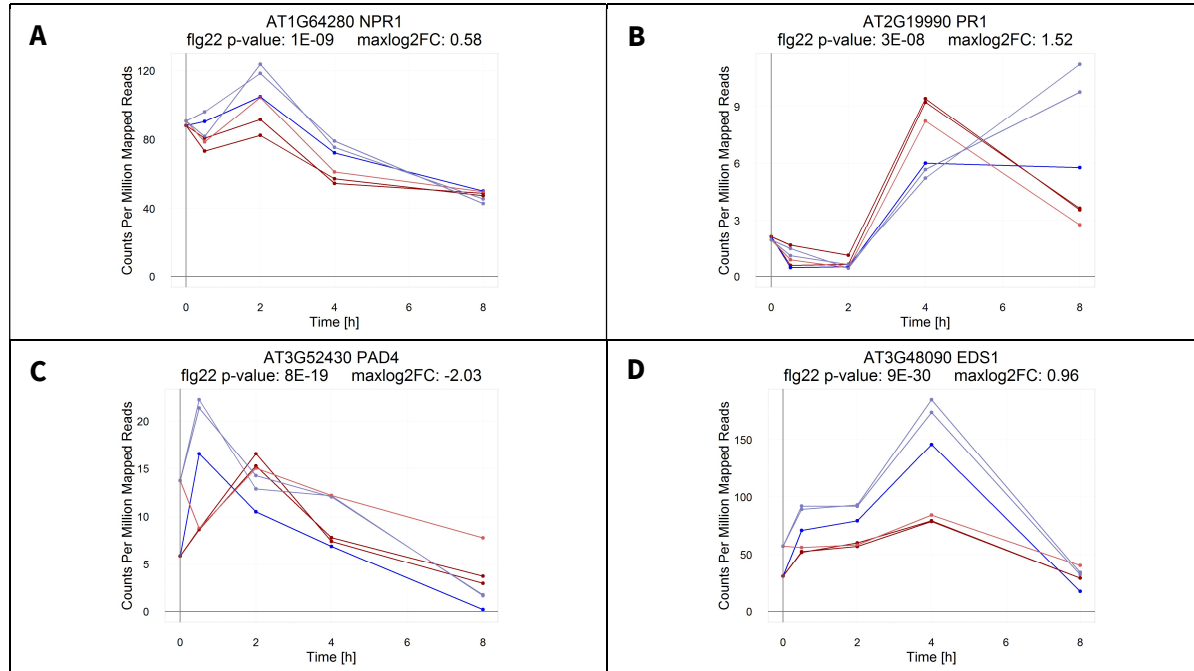

Figure 4.8: Exemplary infection and disease markers.
maxlog2FC represents the highest flg22-induced log2-tranformed fold-change at any measured time point.

To assess this more systematically, we compared all genes that are associated in TAIR to *defense against bacteria* (236 genes), *defense against fungus* (161 genes), *defense against virus* (35 genes) and *systemic acquitted resistance* (53 genes). There is a partial overlap of the genes in these categories (Figure 4.9A).

We determined highly significant differences in the log2-transformed fold changes induces by the flg22-treatment for *defense against bacteria* transcripts (p-value = 3.474e-16, Wilcoxon rank sum test, Figure 4.9D) and *defense against fungus* transcripts (p-value = 1.172e-09). *Systemic acquitted resistance* transcripts were only slightly differentially expressed (p-value = 0.034)*,* while the fold-changes of the *defense against virus* group were not significantly affected by flg22 (p-value = 0.456). The highly flg22-responsive genes of the *defense against virus* category are overlapping with defense genes against fungal or bacterial infection such as WRKY8 (discussed in more detail in 4.6.1) and the lipid acyl hydrolase patatin-like protein 2 (PLP2, Figure 4.9C). PLP2 is also associated with cell death execution upon fungal infection (La Camera *et al*, 2009).
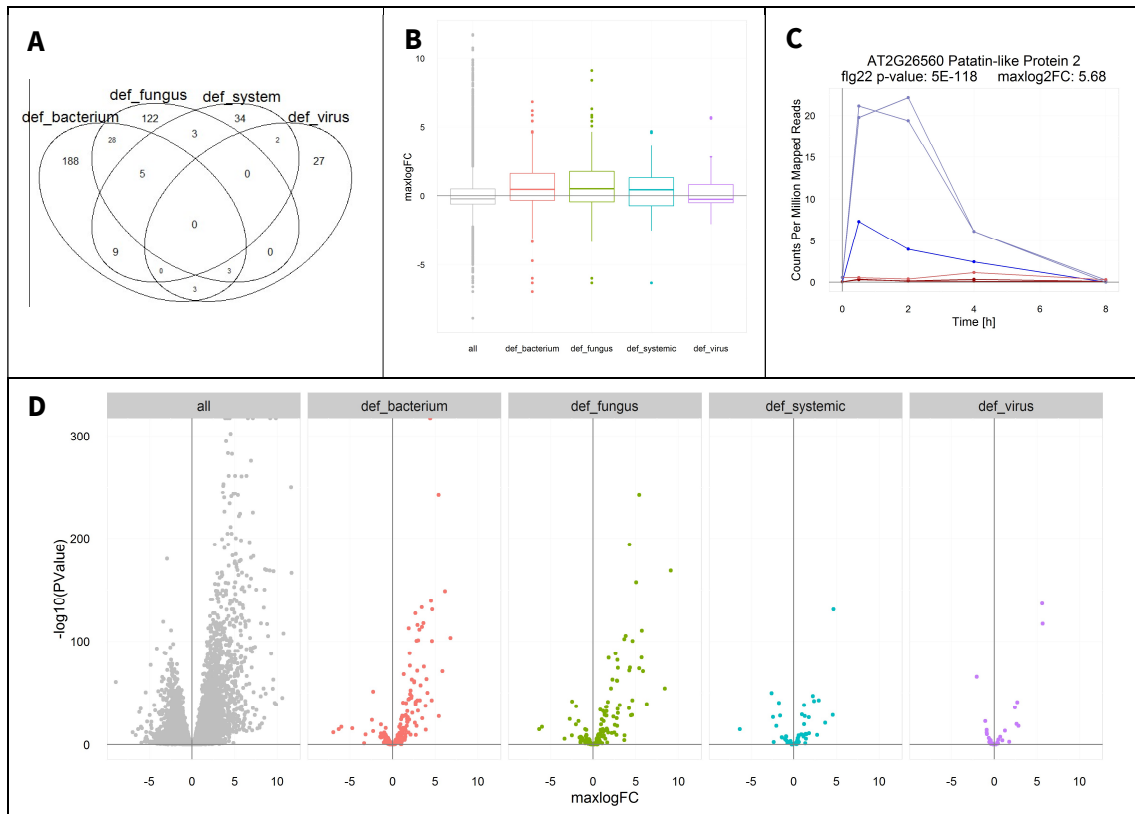
Figure 4.9: Analysis of genes associated with defense related GO categories.
A: Venn-diagram of all genes associated with defence–associated categories. B: Boxplots of the highest flg22-induced fold-change at any time point for all the genes assigned to the different categories. C: Transcript expression profile of PLP2. D: Volcano plots of the genes associated with the different categories.
def_bacterium = 236 *defence against bacteria* genes, def_fungus = 161 defence *against fungus* genes*,* def_virus = 35 *defence against virus* genes, def_systemic = 53 *systemic acquitted resistance,* all = all 18'894 genes assessed in this study.

4.6 Comparison to other transcriptional flg22-studies

The transcriptional response of Arabidopsis to bacterial infection or flg22 treatment has been analyzed in three previous genome-wide studies (Thilmony *et al*, 2006; Howard *et al*, 2013; Li *et al*, 2014a). Surprisingly, the overlap of the significantly regulated genes is only moderate (Figure 4.10A-D). There are a multitude of possible reasons for the low overlap: differences in the quantitative method (microarrays in Thilmony et al. (2006), RNA-Seq in the other three), in the analyzed tissue (seedlings for Li et al (2014), fully grown plants in Howard et al. (2013) and Thilmony et al. (2006), and tissue culture in our case), the time points investigated, stimulant (flg22 for Li et al. (2014) and here, *Pseudomonas syringae* infection for Howard et al (2013), and *Pseudomonas syringae* and *Escherichia coli* treatment for Thilmony et al. (2006)) or the statistical framework and significance thresholds.

The highest overlap was seen with the methodologically most similar study by Li et al. (2014), where RNA-Seq was used to study the transcriptional response of Arabidopsis seedlings after 30 min of

flg22 treatment. The flg22-induced fold-change and adjusted p-values correlate highly with our data 0.5 h after stimulation (Figure 4.10E,F).
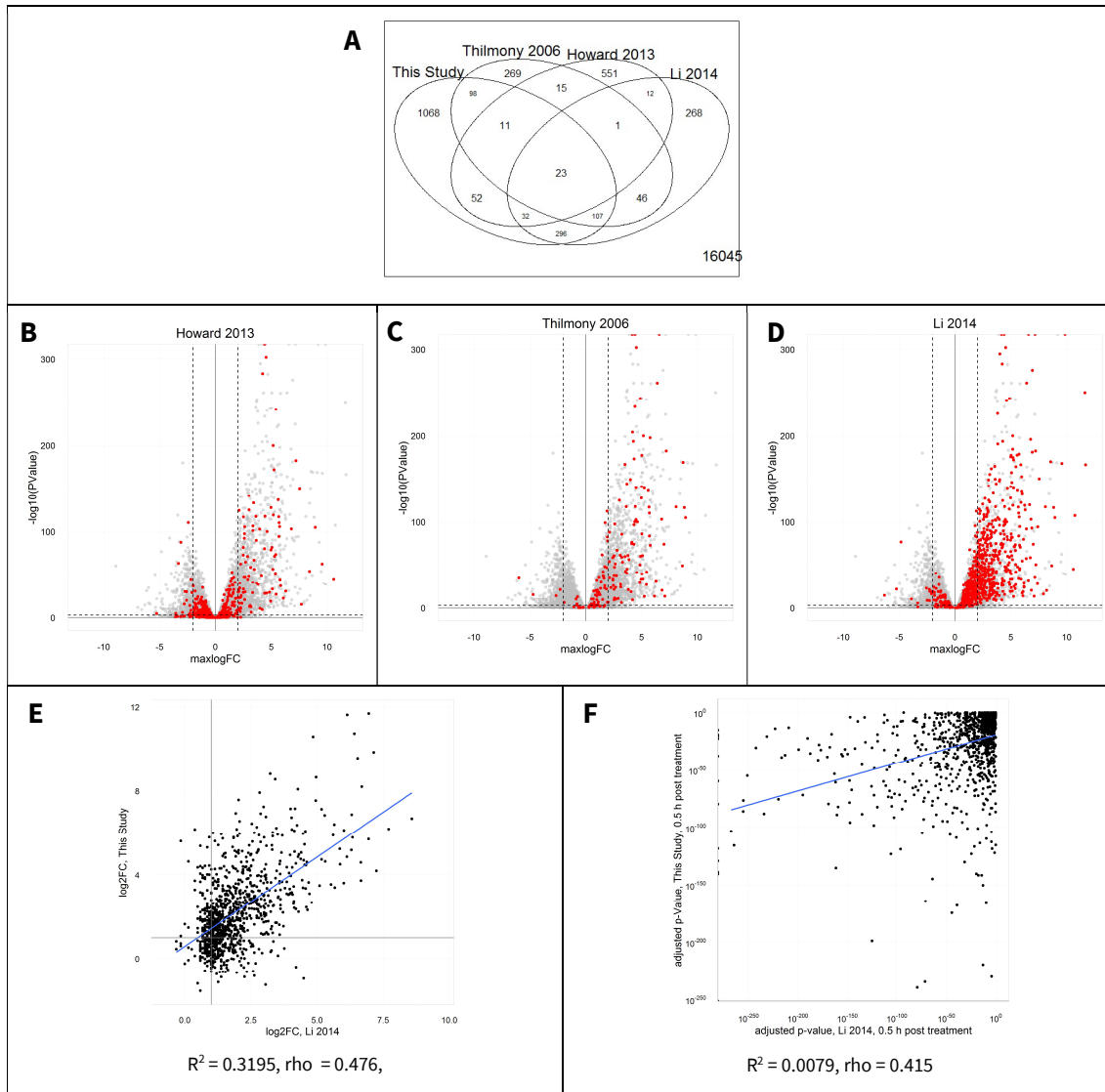

Figure 4.10: Comparison with other genome-wide bacterial defense response studies.
A: Venn diagram of the genes assessed as significantly changing in the indicated studies. B - D: The genes identified as differentially expressed in the indicated study are highlighted in the Volcano plot displaying the transcriptional response in our data set. Our threshold for assessing significant changes are indicate by dashed vertical ($|maxlogFC| > 2$) and horizontal (FDR-adjusted p-value < 0.01) lines E: Correlation of the flg22-induced fold-change our study and from Li et al, (2014). A regression line is indicated in blue with its correlation statistics listed below the plot. F: Correlation of the gene-wise p-values from our study and Thilmony et al. (2006). The regression line is plotted in blue, adjusted $R^2 = 0.0029$

4.6.1   WRKY transcription factor superfamily

WRKY transcription factors are defined by two WRKY domains, which are about 60 amino acids long with the WRKYGQK sequence followed by a C2H2 or C2HC zinc finger motif. Arabidopsis encodes 72 WRKY genes, 52 of which are expressed under our experimental conditions. Of these, 7 genes are significantly down- and 17 are upregulated upon flg22 stimulation, making it a highly responsive class of genes with a significantly different average fold-change (p-value = 0.000179, Wilcoxon rank sum test). The WRKY transcription factor family is highly redundant for their function (Eulgem *et al*, 2000).

In our data set we identify four WRKY transcription factors (WRKY18, WRKY33, WRKY40, and WRKY46) that respond actually more sharply to the flg22 treatment than the classically reported WRKY factors. These transcription factors are reported to be regulated also by other stimuli such as abscisic acid application (Geilen & Böhmer, 2015), drought (Luo *et al*, 2013), osmotic, salt or metal stress (Van Aken *et al*, 2013; Carrio-Segui *et al*, 2016). To disentangle their exact molecular function in the flg22-signaling cascade would be of high interest.

The expression of some members of the WRKY family also peaks at later time points. These late-response transcription factors are actually described to be immune modulators or even antagonistic regulators of the defense mechanism, as WRKY17 (Ali et al, 2014), WRKY8 (Chen *et al*, 2013; Hu *et al*, 2013), and WRKY63 overexpression downregulates the expression of flg22-responsive genes (Van Aken *et al*, 2013). WRKY48 is a plant defense repressor (Xing *et al*, 2008), whose transcription peaks early on and elevated expression rates are maintained over the later time points unlike any other WRKY factor. The transcriptional increase of these WRKY transcription factors could initiate the downregulation of defense response as a measure to return to steady-state.

Interestingly, FRK1, the flagellin receptor-like kinase 1 also shows a more delayed transcriptional response similar to the group of WRKY immune modulator genes (Figure 4.7A). After stimulation, the membrane-bound FRK1 gets activated and internalized. To return to steady state, the amount of non-activated FRK1 has to be replenished on the cellular membrane. A transcriptional increase of FRK1 therefore does not necessarily mean an increase in signaling, like for the WRKY-transcription factors. It could rather be an indication that the flg22 signal has already been passed on.

In our understanding, this is indicative for differences in signal propagation. While FRK1 is a kinase that passes on the signal by phosphorylation, transcription factors have to increase in abundance to "pass on the signal". An increase in FRK1 thus means a return to non-signaling steady state.

This highlights that our transcriptional dataset allows studying the kinetics of flg22 signal propagation in Arabidopsis cells. We can identify early response genes, and genes that are working antagonistically to a pathogen response – all within the same superfamily of genes.
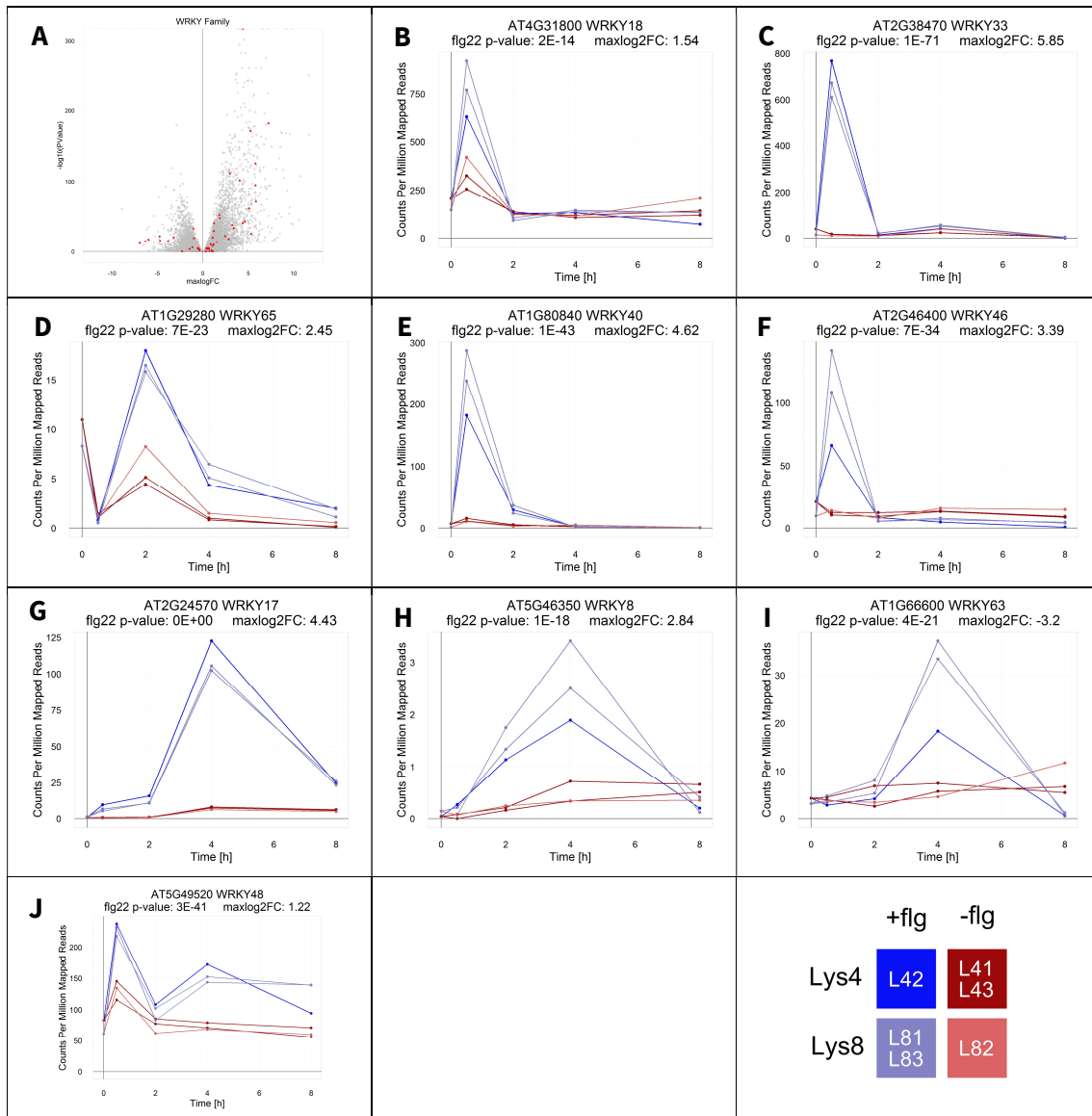
Figure 4.11: Transcriptional expression of flg22-responsive WRKY transcription factors.
A: Volcano plot of flg22-induced transcriptional changes over all time points with the 52 expressed WRKY genes indicated in red. B-F: Examples of early-peaking WRYK genes. G-J: WRKY genes that could regulate the plant defense response in an antagonistic manner. maxlog2FC represents the highest flg22-induced log2-tranformed fold-change at any measured time point.

4.6.2 Programmed cell death

We found that transcripts associated with programmed cell death (PCD) regulation are also regulated by flg22-treatment (Figure 4.12A). The strongest upregulation could be seen for transcripts associated in TAIR to the keyword 'PCD downregulation'. For example Cyclase1 (AT4G34180, Figure 4.12B) is strongly upregulated, which is a negative regulator of cell death and regulates pathogen-induced symptom development (Smith *et al*, 2015). Only one transcript of the genes associated in TAIR to the keyword 'PCD associated' shows a strong trasncriptional response upon flg22 treatment. This gene, cysteine-rich receptor-like kinase 20 (CRK20, AT4G23280, (Figure 4.12C), is described to be a modulator of the plant defence response and promotes conditions that are favorable for infection (Ederli *et al*, 2011).
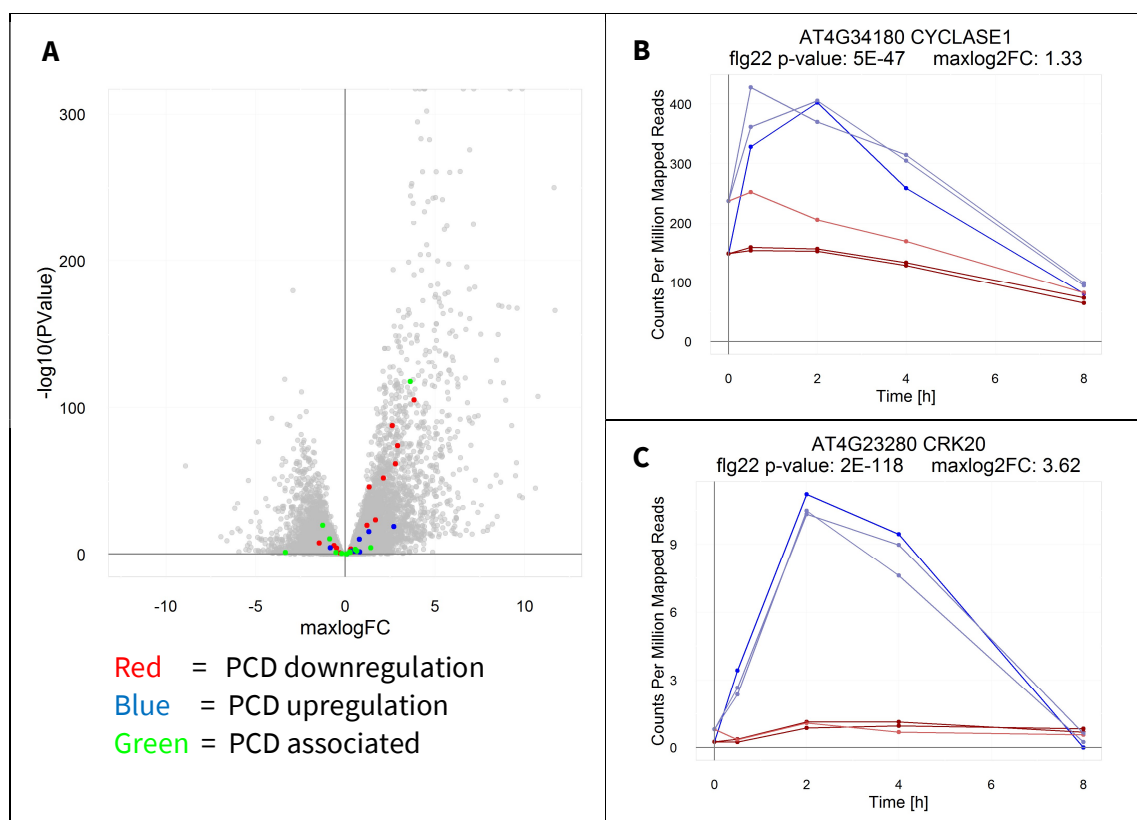

Figure 4.12: Response of transcripts associated to programmed cell death.
A: Volcano plot highlighting all transcripts associated to with three indicated PCD categories. The genes were selected by a keyword search in TAIR. B-C: Exemplary transcriptional responses of genes associated with PCD.

4.6.3 Regulation of Cyclins

Cyclins affect the progression of cells through the cell cycle by activating cyclin-dependent kinases (Galderisi *et al*, 2003). They contain a structural motif called a cyclin box of about 100 amino acids length. Of the 66 genes that were associated in TAIR to the keyword *cyclin-dependent protein serine/threonine kinase regulator activity*, 56 were detected in our transcriptional dataset. These genes have a significantly lower average log2-transformed fold change of -0.6 compared to all quantified genes (average of 0.0494, p-value = 4.811e-05, Wilcoxon rank sum test, Figure 4.13A). The majority of cyclins show a steady and stable transcriptional downregulation upon flg22-stimulation (Figure 4.13C-T), while one undescribed cyclin-associated protein shows a spiked increase at 2 h post treatment (AT5G48640, Figure 4.13Q). The transcription of retinoblastoma-related 1 (RBR1) remains unaffected by flg22-treatment (Figure 4.13B).

Interestingly, many cyclins show a shift in transcript expression between Lys4 and Lys8-labeleld cultures at the beginning of the experiment. The cyclins from Lys4-labelled cultures have a lower expression compared to Lys8 cultures, except for cyclin P1;1 (AT3G63120, Figure 4.13T), which is described to be a negative regulator of CDKA1 (Torres Acosta et al, 2004). Again, the differences between the Lys4 and Lys8 cultures reduces with time. At the latest two hours after the treatment, the trajectories align according to the flg22-treatment rather than the labeling background.

In summary, the transcriptional status of the cyclins hint at a slowed cell cycle progression when treated with flg22, which would reduce the growth rate of the treated cultures. In chapter 7.2, the actual growth rates over of the treated cultures are discussed in more detail.
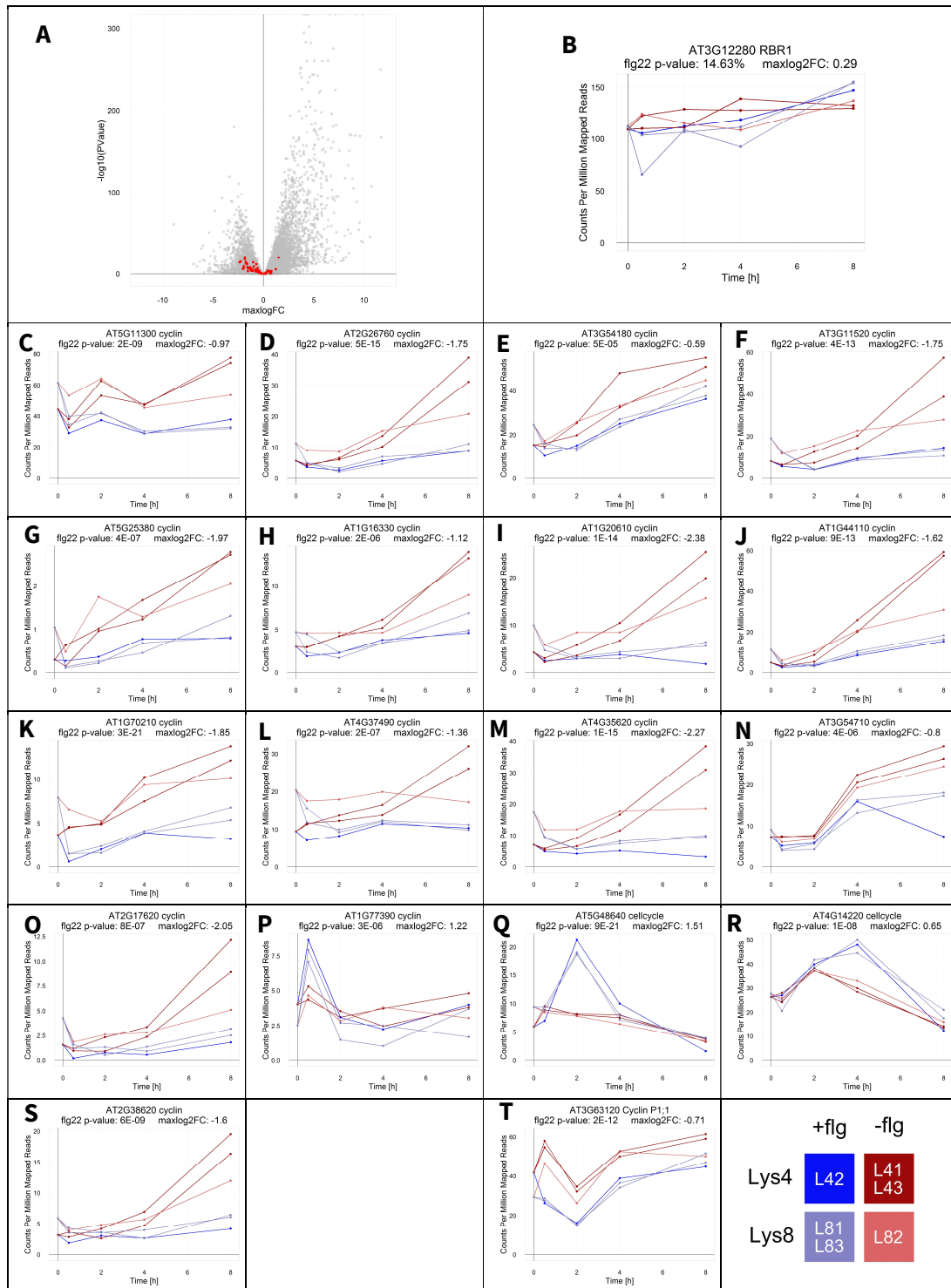
Figure 4.13: Effect of flg22 on transcript expression of cyclins.
A: Flg22-induced expression changes of all detected cyclins. B-T: Exemplary flg22-affected cyclins and cyclin-related proteins. T: Note the transcriptional increase of cyclin P1;1 (AT3G63120) in Lys4 cultures.

4.7 Differences between lysine-labelled cultures

Schweizer et al. (2013) identified 41 transcription factors that were responsive to herbivory by the leafworm *Spodoptera littoralis,* including AZF2, BT1, ATRD26, ANAC019, ZAT10, and ZAT12 (Figure 4.14). Most of these transcription factors are only moderately regulated upon flg22-treatment (FC < 2.0). Interestingly, though, many of these genes indicate an upregulation in the Lys4-labelled culture before any flg22-treatment. This might hint towards an explanation for the differences between Lys4- and Lys8-labelled cultures indicated in the MDS plot (Figure 4.2). Herbivory can be excluded, yet some signaling process might have been triggered for the Lys4 cultures – potentially to wounding during the passaging of the cultures or to senescence as the Lys4 culture grew quicker. For most of these genes, the expression trajectories of the Lys4- and Lys8-labelled cultures align again at the first measurement after the treatment (0.5 h). Despite of the Lys-labeling effect, the flg22-treatment affects gene expression of these transcription factors more prominently.
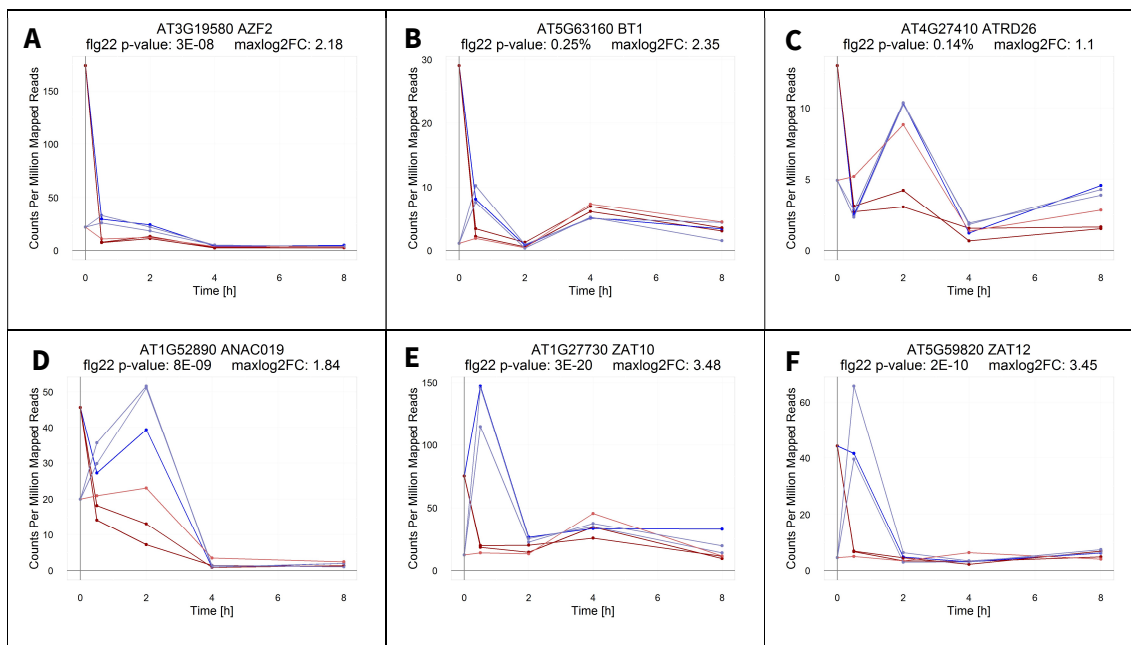

Figure 4.14: Exemplary expression profiles of herbivory-induced genes (Schweizer et al. 2013).

When testing systematically for differences between Lys4 and Lys8 cultures over all time-points, we identify 36 differentially expressed genes when applying the significance thresholds defined before (Figure 4.15A). These genes are not enriched for a stringent set of GO categories. Only the GO category *cellular response to ozone* was enriched at an adjusted p-value cutoff of 0.05, yet by containing only two superoxide dismutase genes (Figure 4.15B-C).

However, testing explicitly for differences between the two untreated Lys4- and Lys8-labelled cultures at the onset of the experiment, we identify 104 differentially expressed genes (log2(FC) >2 & cpm > 15) that are significantly enriched for the GO categories *response to chitin* (9 genes, adj. p-value 1.34E-05) and *ethylene-activated signaling pathway* (6 genes, adj. p-value 4.49E-02). Genes from these categories show strong expression differences between Lys4- and Lys8-labelled cultures

at the onset of the experiment, but their expression trajectories align again 0.5 h after treatment (Figure 4.15D-I).

We conclude that the Lys4- and Lys8-labelled cultures are differently affected by a signaling process, which is most likely caused by slight differences in growth rate or handling of the cultures. It is important to note though that the effects disappear quickly after the treatment and are substantially smaller than the response to flg22.
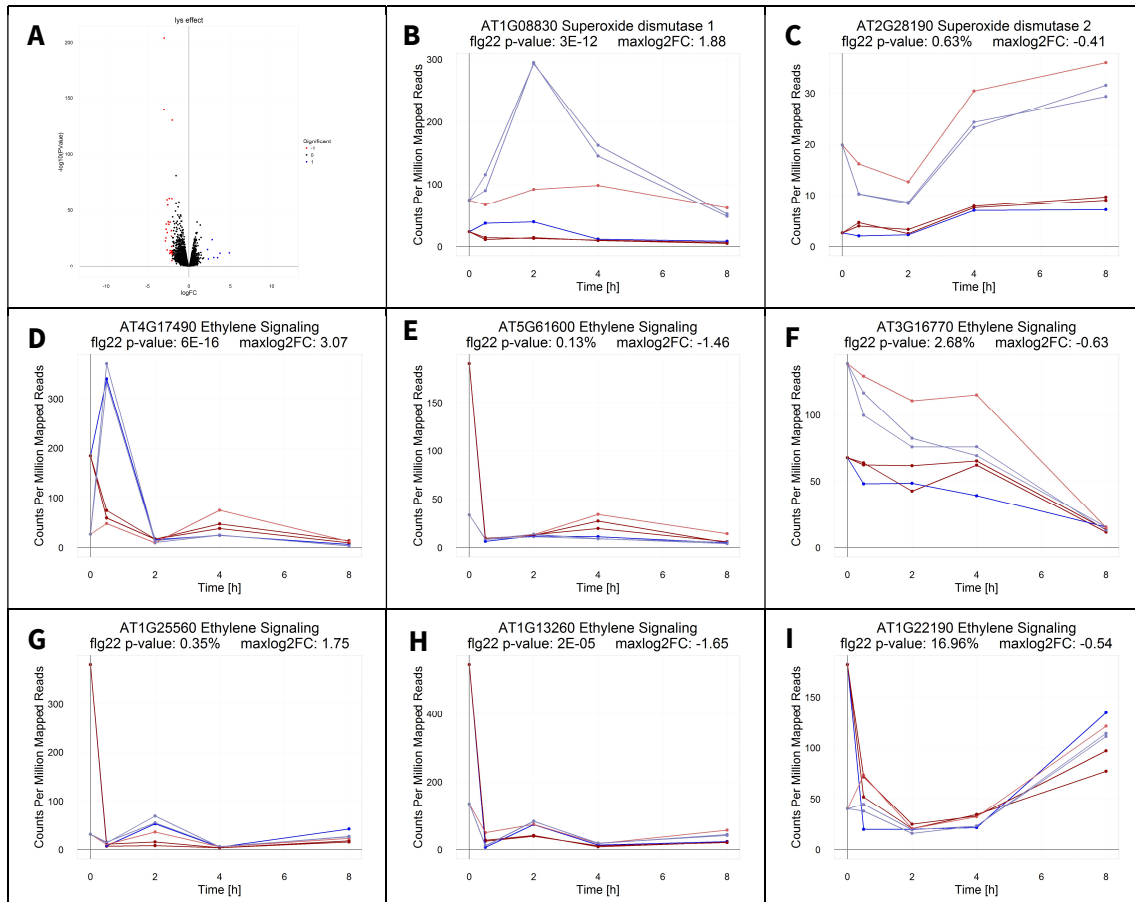

Figure 4.15: Differences in transcription expression between Lys4- and Lys8-labeled cultures.
A: Volcano plot for genes affected by the type of lysine labeling at any time point. B-I: Exemplary genes that are differentially expressed in Lys4- and Lys8-labelled cultures.

5.

TRANSLATION

5.1 Development of a plant Ribo-Seq protocol

As we wanted to perform the Ribo-Seq study on the same isotopically labelled culture material as the RNA-Seq and proteomics studies, we aimed to minimize input of cell material for successful RFP enrichment. In the studies by Juntawong et al. (2013a) and Li et al. (2013), extensive amounts of plant material was used to enrich for RFPs. One reasons for this is that previous Ribo-Seq protocols and commercial kits, notably the ARTseq kits by Epicenter / Illumina, were developed for yeast and mammalian cell cultures and are designed to lyse the cells in high concentrations and in buffers that are compatible with RNase degradation (Figure 5.1A). Plant extracts in comparison are more diluted due to the large vacuole (>90% of the cell's volume) and contain secondary metabolites that hamper effective Ribo-Seq RNase digestion.

In our understanding, another bottleneck in the protocol published by Juntawong et al. (2013a) is that the first ultracentrifugation is performed in a Type70Ti rotor. The tubes for this rotor must be filled to 36 mL, thus requiring extensive centrifugation (3 - 18 h at 170'00 rpm) and high amounts of cell material (about 5 g (Loraine, 2009)). We adapted that approach by scaling down the volume during ultracentrifugation using a MLA-80 rotors (2 h 30 min at 80'000 rpm, 1 mL 60 % sucrose cushion and 1.5 mL cell extract) or a TLA 120.1 rotor (3 h at 80'000 rpm, 1 mL 60 % sucrose cushion and 0.8 mL extract). These ultracentrifugations have the same clearing factor (k factor) as the one in the Type70Ti rotor, yet require less volume and less cell material. With this we obtained frequently >20 μg enriched mRNA-ribosome complexes from 1 g cell material. 10 μg RNA are sufficient for the subsequent steps of the Ribo-Seq protocol.

The pelleted mRNA-ribosomes complexes were resuspended in RNase-compatible buffer and digested with RNase. The resulting ribosome-protected fragments (RPFs) were then purified by denaturing PAGE, where we cut out the 20-35 nt area. This circumvents the additional differential centrifugation step performed by Juntawong et al. (2013a), which further decreases the extraction efficiency. We had more success digesting the mRNA-ribosome complexes with the ARTseq RNase (Epicenter / Illumina) than with RNase I (New England Biolabs, Figure 5.1B). Using the approach employed for yeast and mammalian cell culture (Ingolia *et al*, (2013), ARTseq kit Epicenter / Illumina)

without prior polysome enrichment with ultracentrifugation in a sucrose cushion did not lead to efficient degradation with either RNase.

We also tested enriching the ribosome-mRNA complexes by size exclusion chromatographic spin columns (S-400 HR columns, GE Healthcare) as suggested by Freeberg *et al,* (2013). The spin columns can be loaded with maximally 100 μL sample volume, making them well suited for the yeast and mammalian extracts. Yet, the cell extracts from plant material usually exceed 1 mL in volume, making it expensive and cumbersome compared to ultracentrifugation enrichment.

We conclude that the enrichment and buffer exchange achieved by pelleting the mRNA-ribosome complexes is highly recommended for efficient RNase digestion in plant Ribo-Seq studies as it removes metabolites and other cellular components that might interfere with the subsequent steps. The proposed small-scale ultracentrifugation is effective and should also be applicable for sucrose gradient polysome profiling studies.
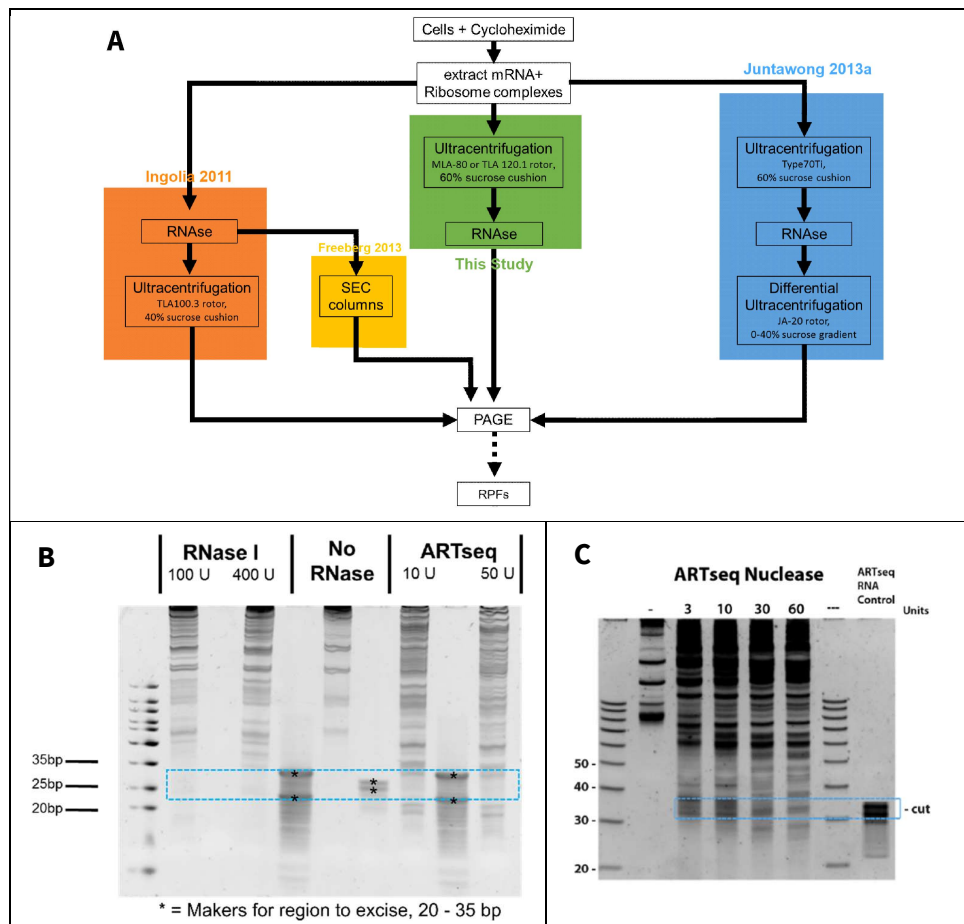

Figure 5.1: Plant Ribo-Seq optimization.
A: Enrichment strategies for RPFs as employed by Ingolia et al. (2011), ARTseq kit (Epicenter / Illumina, Freeberg *et al*, (2013)), Junatwong et al. (2013a) and as implemented here. B: Optimization of RNase digestion in mRNA-ribosome enrichments by RNase I (NEB) or ARTseq nuclease (Illumina / Epicenter), analyzed by denaturing PAGE. We obtained more efficient degradation using at least 50 U ARTseq nuclease. The RPFs are excised from the blue dashed area of 25-35 nt. C: Digestion efficiency reference from the ARTseq protocol (Illumina/Epicenter). They assessed that digestion with 60 U ARTseq nuclease was optimal for digesting the mRNA-ribosome complexes of that sample.

5.2 Ribo-Seq results

To assess the translational response of flg22 in Arabidopsis suspension cell culture AtB, we performed Ribo-Seq on the same cell culture samples as the RNA-Seq (chapter 4) and protein abundance samples (chapter 6). To reiterate, these cultures were supplemented for 12 days with daily doses of 350 µM Lys4 or Lys8 and were passaged to fresh medium on day 6 of the experiment. On day 12, the two labeled cultures were split into three cultures each and one-half of the cultures were treated with 100 µM flg22 and the other half remained untreated. The treatment was added to two Lys4-labelled cultures (L41 and L43) and one Lys8-labelled culture (L82), while one Lys4-culture (L42) and two Lys8-lableld cultures (L81 and L83) remained untreated. We took samples from the cultures before the treatment and 0.5 h, 2 h, 4 h, and 8 h after the treatment and analyzed their translational profile with the plant Ribo-Seq method established before (chapter 5.1).
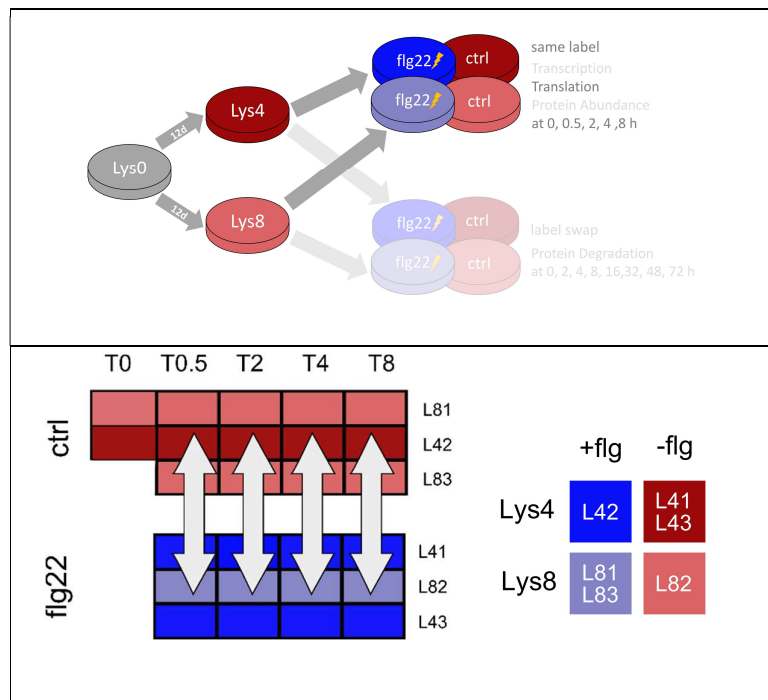

Figure 1.3: Experimental design schema of the Ribo-Seq part of the study, where we analyzed the translational response of Arabidopsis cell culture AtB on flg22. We use the color code in the legend to the bottom right throughout this thesis.

5.3 Ribo-Seq quality control

5.3.1 Mapping rate

We sequenced on average 21 million reads per sample with a good per-base sequencing quality (>36 phred score, Figure 5.2A, C). After trimming of the adapters, the sequences displayed a length of 14 - 38 nt and a mode of 28 nt (Figure 5.2D).

When mapping the reads to the Arabidopsis genome with Bowtie2 ((Langmead *et al*, 2009), 20 nt minimum read length and minimum average alignment quality of 10), we could map on average only

37.7 % of all reads (Figure 5.2B). For comparison, > 95 % of the RNA-Seq reads of chapter 4 could get mapped. The mapping rates vary substantially between the samples (standard deviation of 15.62 percentage points), yet no batch effect or pattern could be observed that would link it to an experimental procedure such as RNase digestion, library preparation or PCR amplification.

We also tested other alignment algorithms such as STAR (Dobin *et al*, 2013) and CLC Genomics Workbench (Qiagen Bioinformatics), but obtained worse mapping rates (> 50% fewer mapped reads per sample). Lowering the Bowtie2 minimal alignment cutoff increases the mapping rate, yet with unclear consequences for the subsequent biological analysis. We therefore decided for the strict but reliable average alignment quality cutoff of 10. It remains unclear what caused this variation in mapping rates, yet mapping short reads such as RPFs will always be a challenging task for any algorithm.

Juntawong et al. (2014) report mapping rates of 66.9%, while Chotewutmontri & Barkan (2016) report mapping rates of >90% using Bowtie2 using not specified parameters. Their main experimental difference seems to be the high amount of RNase added to each sample (3'500 U RNase I, while we tested maximally 400 U RNase I). However, it seems unlikely that this can cause the differences in mapping efficiency. In any case, in future studies it might be advisable to assess successful RNase degradation not just by denaturing PAGE analysis, but by sequencing a range of samples with different RNase concentrations to assure a good mapping rate.
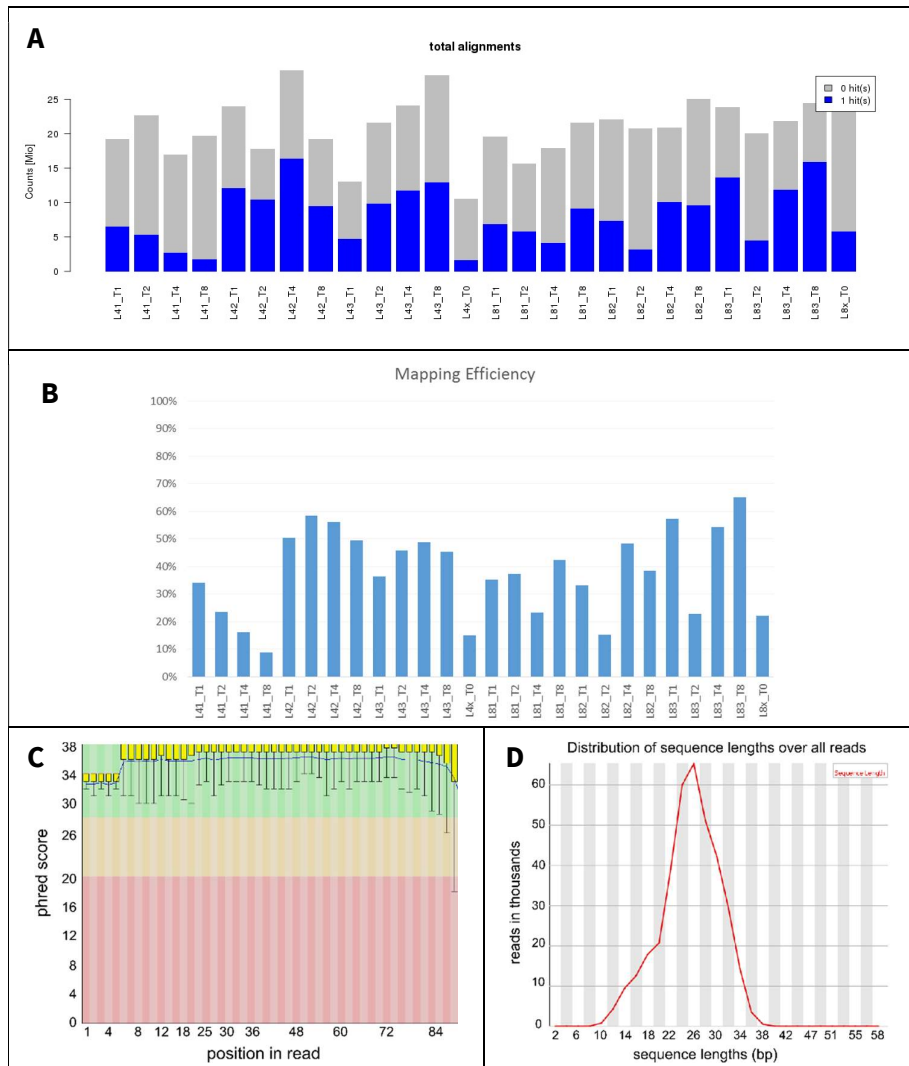
Figure 5.2: Ribo-Seq quality control plots.
A: Reads per sample. Total number of reads are indicated in grey and number of reads mapped with Bowtie2 are indicated in blue. B: Per sample mapping rate of Bowtie2. C: Representative sequencing quality score. The sequenced part of the RPFs libraries have an average length of <80 nt, therefore the quality score drops at higher sequence lengths. D: Representative read length distribution after trimming of the adapters.

5.3.2 Ribosome-protected fragment length

Similarly to the findings of Chotewutmontri *et al* (2016), we see differences in the distribution of the read lengths of the RPFs regarding their subcellular localization. Nuclear-encoded genes are translated in the cytosol by eukaryotic ribosomes, while the chloroplast- and mitochondrion-encoded genes are translated by smaller prokaryotic-like ribosomes.

In our dataset, RPFs from nuclear-encoded mRNAs have a read length mode of 31 nt (Figure 5.3A), while the chloroplast- (Figure 5.3B) and mitochondrion-encoded (Figure 5.3C) mRNAs have a shorter read length mode of 24 and 28 nt, respectively. When only genes with a continous RPF read length distribution are taken into account, both mitochondrial and plastidic RPFs peak at 27 nt.

A substantial amount of RPFs show a non-optimal read length of < 30 nt for nuclear-encoded mRNA and < 25 nt for mitochondrion- and chloroplast-encoded mRNAs. This might be indicative of problems in the preparation of the sequencing libraries, which could explain the low mapping rate.
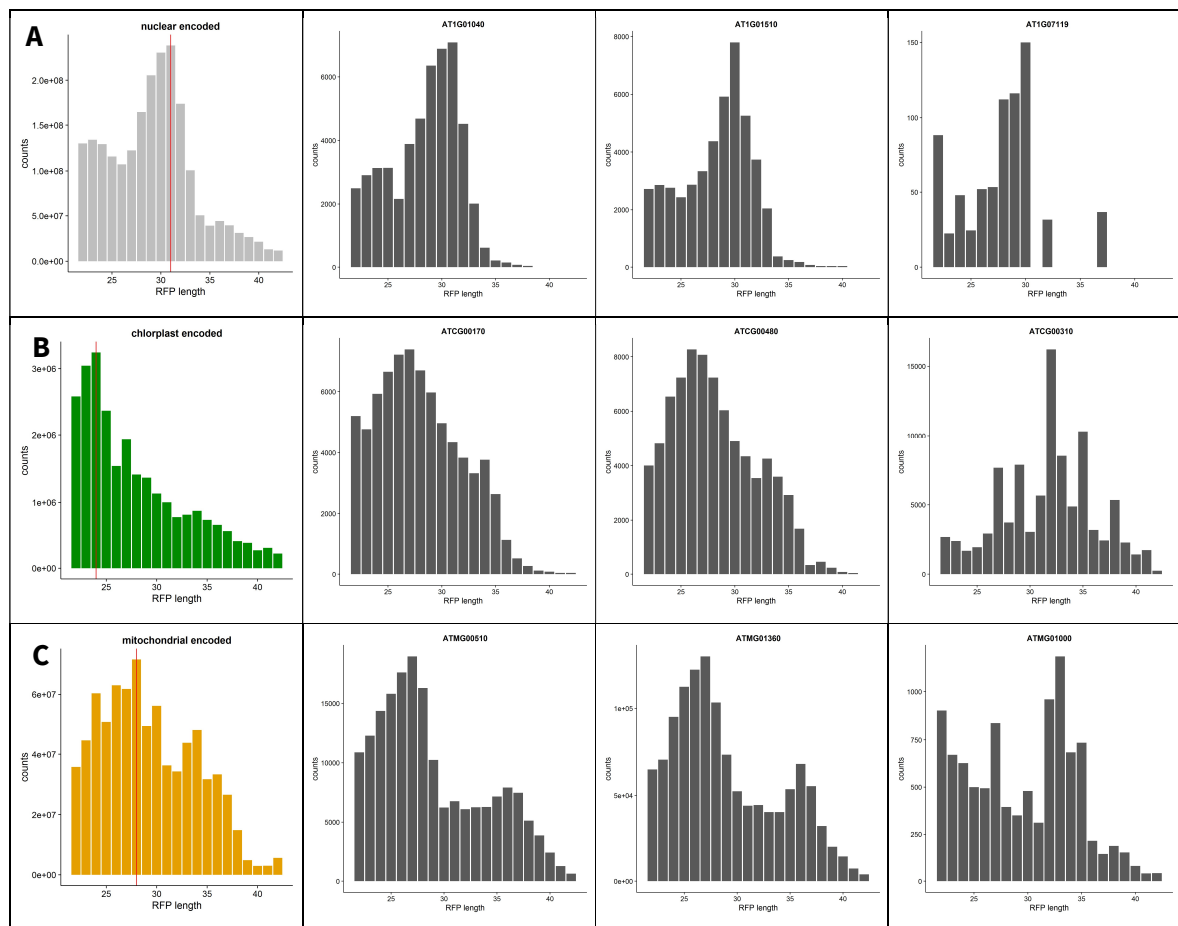

Figure 5.3: RFP length distribution of the ribosome protected fragments (RFPs) that map to nuclear (A), plastid (B) or mitochondrial (C) genes.
For each distribution, the sum of all genes is presented where the modus RFP length is indicated by a vertical line. Additionally, three exemplary genes are shown, two of which showed a continuous read length distribution and one that showed a more erratic pattern.

5.3.3 rRNA depletion and repetitive elements

Another important quality check for Ribo-Seq samples is to determine the number of reads that map to rRNA genes. Each RNase-digested ribosome-mRNA complex contains several kb of rRNA but only ~30 bases of mRNA. tRNAs are also frequent contaminants as they are highly abundant in the cell and are short enough (~71 nt) to reach into the size selection window of the denaturing PAGE.

Ribo-Seq cannot rely on oligo-dT enrichment approaches as mRNA-Seq does because the RPFs by definition do not contain poly-A tails. rRNAs are therefore depleted from the samples by e.g. subtractive hybridization, which has a lower efficiency than oligo-dT enrichment. rRNAs thus remain a steady contamination in many Ribo-Seq experiments (Ingolia *et al*, 2013).

In our samples, on average 36.17 ±19.7% of all reads map to rRNA and tRNA genes (Figure 5.4), which lies in the expected range for a Ribo-Seq experiment. In Chotewutmontri *et al*, (2016) 53 - 81% rRNA and tRNA reads are reported and in Juntawong *et al*, (2013a) 52.13 ±7.56 %. An alternative sequence-independent rRNA depletion approach using duplex-specific nucleases after PCR amplification of the library resulted in a lower rRNA depletion rate (50-80% rRNA reads, Chung *et al*, 2015)

In our samples, we observed that the efficiency of the subtraction hybridization was highest for samples processed early on and decreased over time. Furthermore, we achieved improved rRNA depletion when we repeated the magnetic separation of the rRNA-bound beads. Based on these observations, we conclude that repeated freeze-thaw cycles of the hybridization solution should be avoided and that the facultative additional separation of the rRNA-bound beads should always be performed.
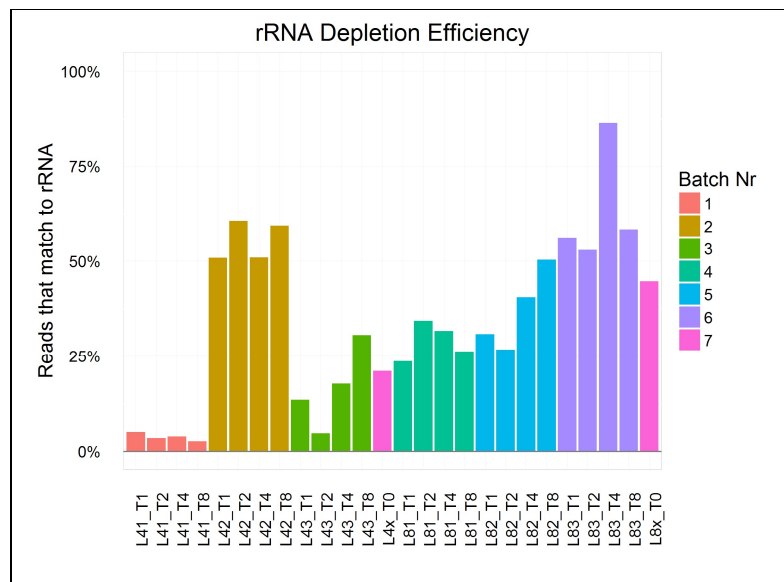
Figure 5.4: rRNA depletion efficiency per sample.
RNA samples were processed in batches using Ribo-Zero. The batch number is indicated. In batch 2, the magnetic separation of the rRNA-bound beads was performed only once, as recommended by the Ribo-Zero protocol. In the other batches, the separation was repeated for a second time.

After exclusion of all rRNAs and tRNA reads, we identified 8 ORFs with atypically high read count (>1000 counts per million mapped reads) that correlated poorly over the biological replicates. These reads map to genes with repetitive sequences such as a non-LTR retrotransposon family gene (LINE, AT2G23880, 2.7% of all reads) or to completely unannotated ORFs (AT2G01021, AT3G55850, AT1G70185, AT3G52700, AT3G55850, AT1G70185, ATMG00030). Ingolia et al, (2013) observed similar highly abundant, non-reproducible reads and presume that these originate from contaminating sequences including ncRNAs. The reads aligned to these 8 ORFs were excluded from further analysis.

5.3.4   Reproducibility

After exclusion of all contaminating reads, the number of reads per samples ranged from 300'000 to 4.5 million (Figure 5.5A). Despite the differences in library size, mapping rate, rRNA depletion, or amount of contaminating reads, we obtain good correlations between the gene-wise counts per million mapped reads (cpm) of the replicate samples (mean rho = 0.9332 ±0.027, Figure 5.5B-E).
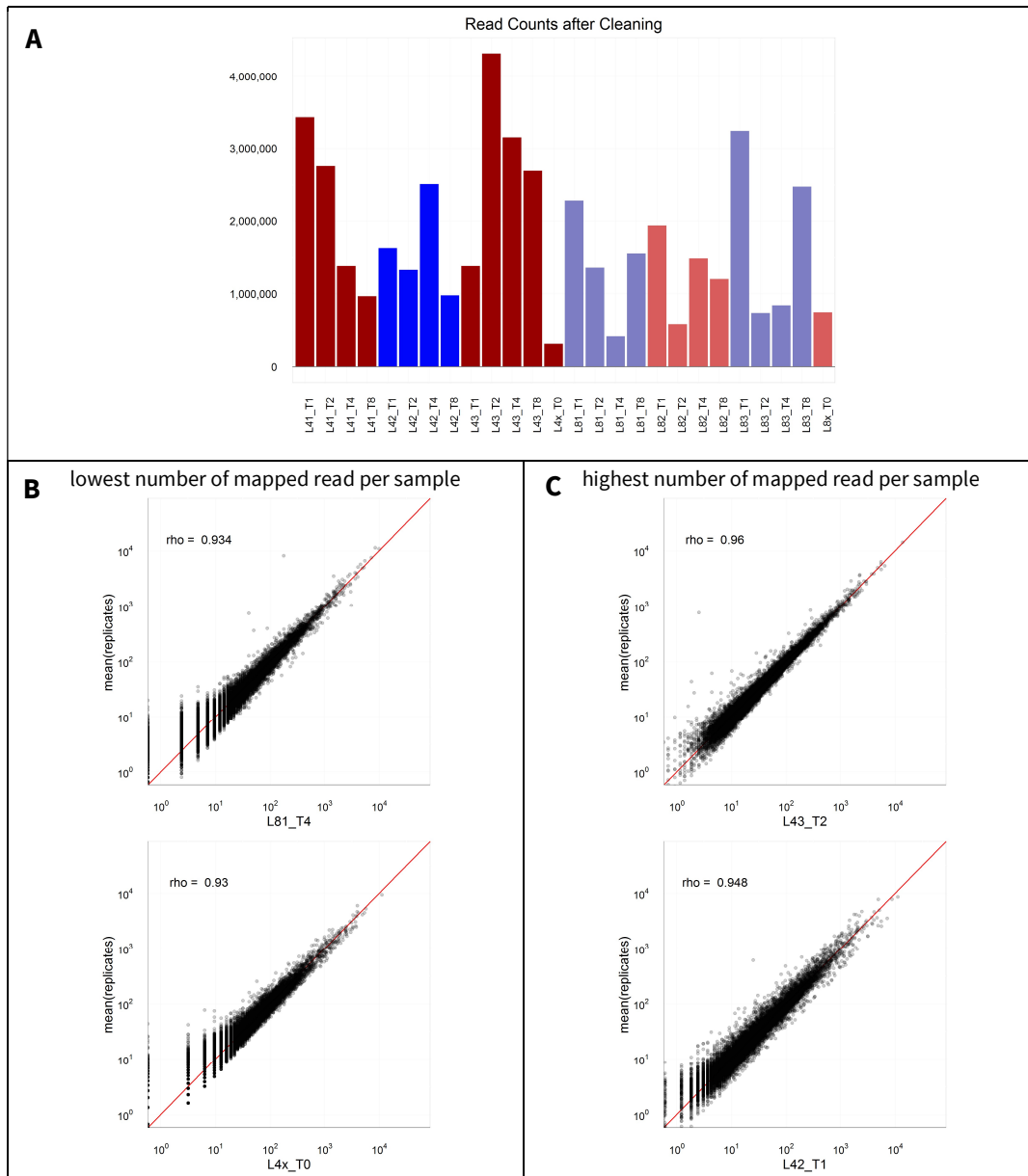

Figure 5.5. Reproducibility analysis.
A: Number of mapped reads per sample after removal of all contaminating reads. B-E: Correlation plots of the two samples with the lowest (B, D) and highest number of reads (C, E) each compared to the mean of their two biological replicate samples. Each point represents a gene and they are plotted by their counts per million mapped reads (cpm).

5.3.5   Periodicity of mapped RPFs

Ribo-Seq data sets are informative for various analyses, such as discovery and annotation of open reading frames, translation initiation sites and uORFs, or identification of sequence-specific ribosomal stalling (Ingolia, 2016). Figure 5.6A-B indicates the capacity of the presented data set for such analyses. Aligning all reads with respect to the translation start of the corresponding ORF lets a 3 nt periodicity pattern emerge over the translated gene body (Figure 5.6A). This pattern results from the movement of the ribosome in intervals of the triplet code. The 3 nt pattern is particularly enriched in 30 - 32 nt long reads. Reads with a different length show a more stochastic pattern (Figure 5.6B).

While the 3 nt periodicity is primarily a quality control measure, we use the Ribo-Seq data here primarily as a measure of translational activity. The Ribo-Seq counts represent the ribosome occupancy over all mRNA copies of a gene, which we take as a proxy for the translation rate of that gene.
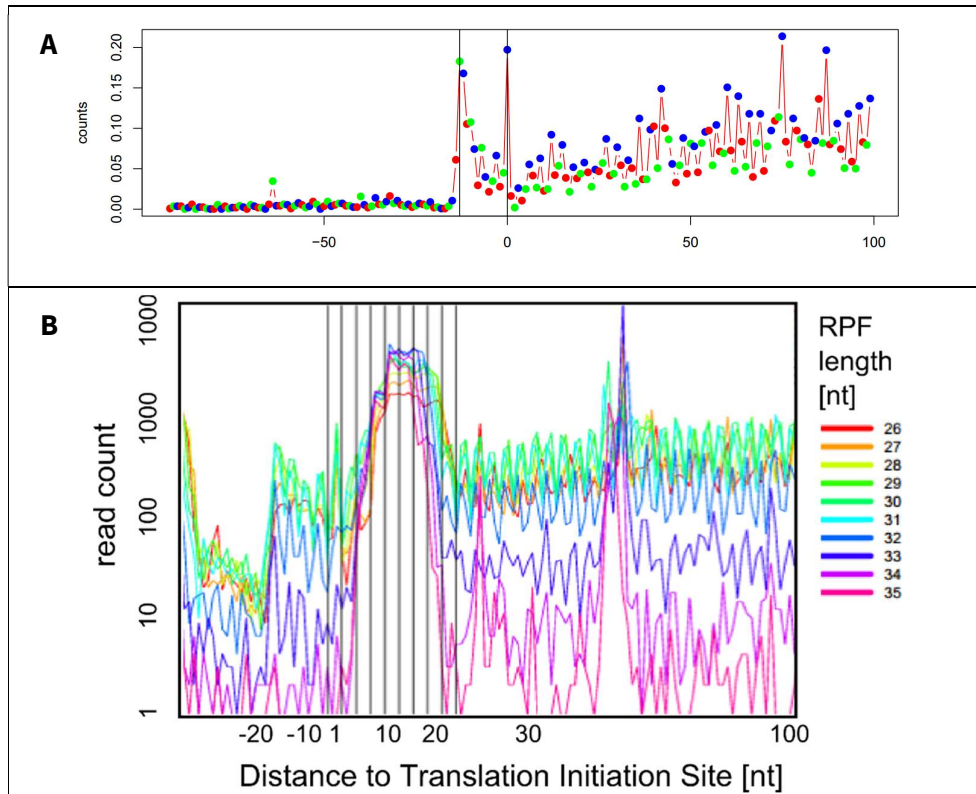
Figure 5.6: Periodicity plots of sample L41_T8
A: Alignment of all reads to the translation initiation sites of their respective ORF. Reads that align to the +0 frame are colored in blue, the +1 frame in red, and +2 frame in green. Only the RPF reads with a length of 30 nt are plotted. B: Similarly to A, the read counts per base aligned to the translation start site are plotted. The lines are colored with respect to the read length of the RPFs. The green (30 nt), turquois (31 nt), and blue (32 nt) read lengths show the highest number of reads and show the 3 nt pattern most consistently.

5.4 Dimensionality reduction plots

We used principal coordinate analysis (PCoA, Figure 5.7A) and multi-dimension scaling (MDS, Figure 5.7B) to non-parametrically assess the variance within the obtained data set. Similarly to the RNA-Seq dataset, we see a clearer pattern in the MDS plot, which represents the variance of the 500 most differentially expressed genes.

Both plots show a clustering of the samples according to treatment and time points. Either factor leads to a discrimination of the samples. In the MDS plot, the effect of the flg22 treatment is more pronounced. The outlier in the MDS, L83_T4, aligns well with its biological replicates in the PCoA, indicating a primarily statistical reason for the strong difference. Comparable to the RNA-Seq dataset we see the strongest flg22-induced differences at 2 h and 4 h post treatment. The distance between treatment and control at 0.5 h post treatment is clearly visible yet less pronounced than in the RNA-Seq dataset and the treated and untreated samples align again 8 h after treatment.

The type of lysine labeling is again a consistent effector on the samples' variances, as it separates the samples within their sampling time and treatment cluster. In the MDS, the Lys8-labelled samples are shifted in most cases to the bottom right of their respective replicate cluster (Figure 5.7C), indicating a small but consistent difference in translation of the differentially labeled cultures.

Overall, the plots assure that both time and treatment affect the sample composition in a biologically coherent way.
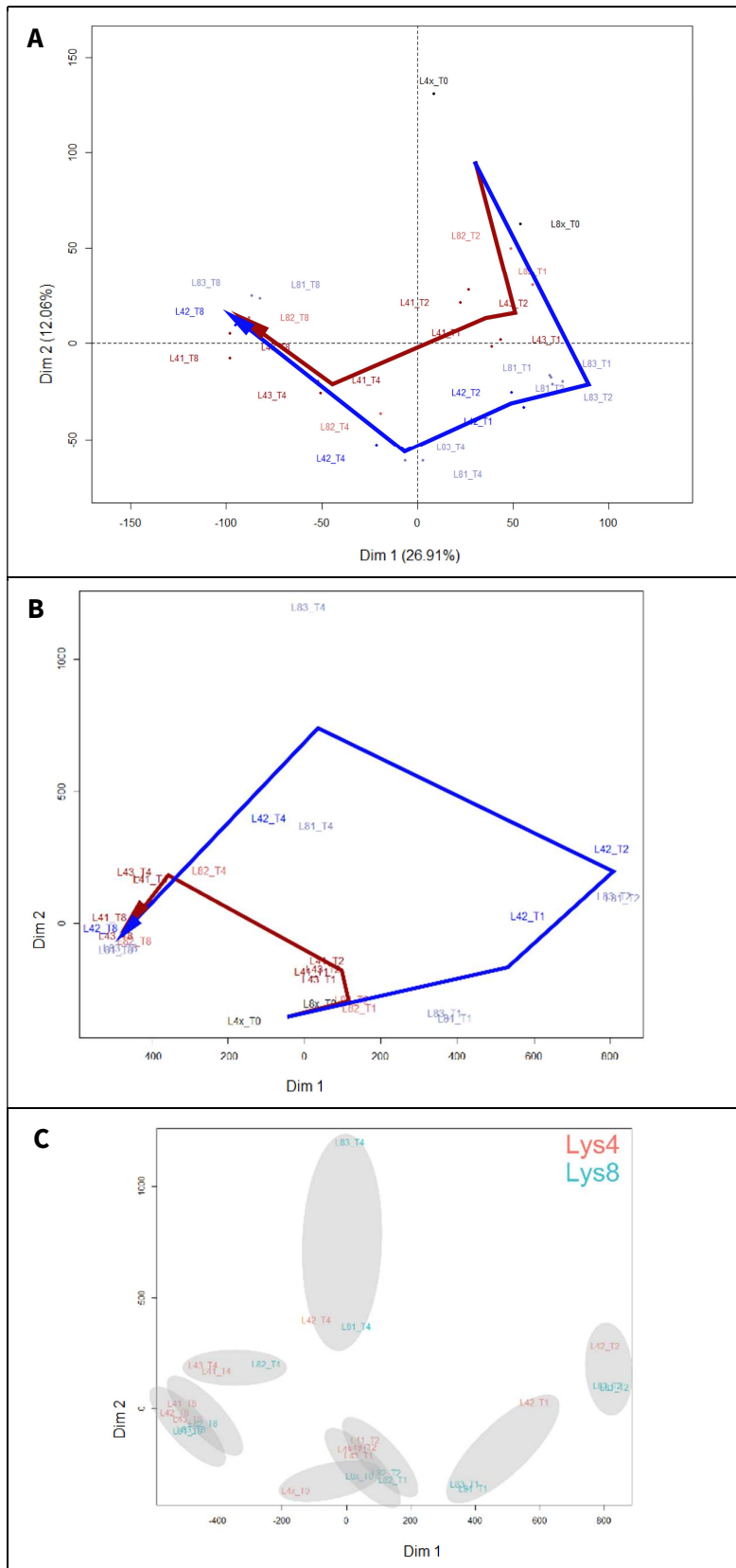
Figure 5.7: Dimensionality reduction plot of the Ribo-Seq data set.
A: Dimensions 1 and 2 of the principal coordinate analysis. B: Multi-dimension scaling (MDS) plot calculated with the R package *limma* (Smith, 2005). In both plots, the averages of the replicate clusters (same time and treatment) are connected by a blue arrow for the flg22-treated cultures and by a red arrow for the control cultures. C: MDS where the samples are colored by the lysine-labeling background of the cultures. Time and replicate clusters are indicated by gray shadings.

5.5 Flg22-induced differential translation

We used *edgeR* to test for significant changes in the translational rates between flg22-treated and untreated cultures (Figure 5.8A) after exclusion of genes with low coverage. Compared to the RNA-Seq dataset we used a more conservative minimal expression threshold of at least 5 cpm in more than 8 samples to account for the higher variability caused by the lower read count per sample (common biological coefficient of variation (BCV) of 0.196, Figure 5.8B). After this final filter, the Ribo-Seq data set contained a measure for translational activities for 13'730 ORFs in 26 samples, which is 5'000 genes less than the RNA-Seq data set, but lies within the expected range of published Ribo-Seq studies (11'000 genes in Toyama *et al*, (2013), 12'487 genes in Juntawong *et al*, (2013a), 15'000 genes in Liu et al, (2013)).

We used the same statistical design and significance cutoff as for the RNA-Seq data (chapter 4, FDR-adjusted p-value < 0.01 & |log2(fold change)| > 2). The lysine background was again included as a blocking factor:

$$count \sim libary\ size + \ lysine\ background + time\ point + flg22\ treatement$$

We compare the effect of flg22 treatment over all four time points in one analysis (Figure 5.8C) or separately for each time point (Figure 5.8D-G). Overall, the differentially translated mRNAs display the same pattern as the differentially transcribed genes. The highest number of differentially translated mRNAs is detected 0.5 h after treatment and every successive time point contains fewer differentially translated genes. The number of genes with increased translation was 3 to 10 times higher than the number of genes with reduced translation.

The number of differentially translated genes in the Ribo-Seq dataset is lower than the number of differentially transcribed genes in the RNA-Seq dataset. This is presumably a consequence of the lower coverage and a higher variation between replicates of the Ribo-Seq dataset (illustrated by the elevated biological coefficient of variation (BCV), Figure 5.8B), caused by the more complex biochemical preparation. This means that differences in the absolute number of significantly regulated genes is not indicative for translational regulation.
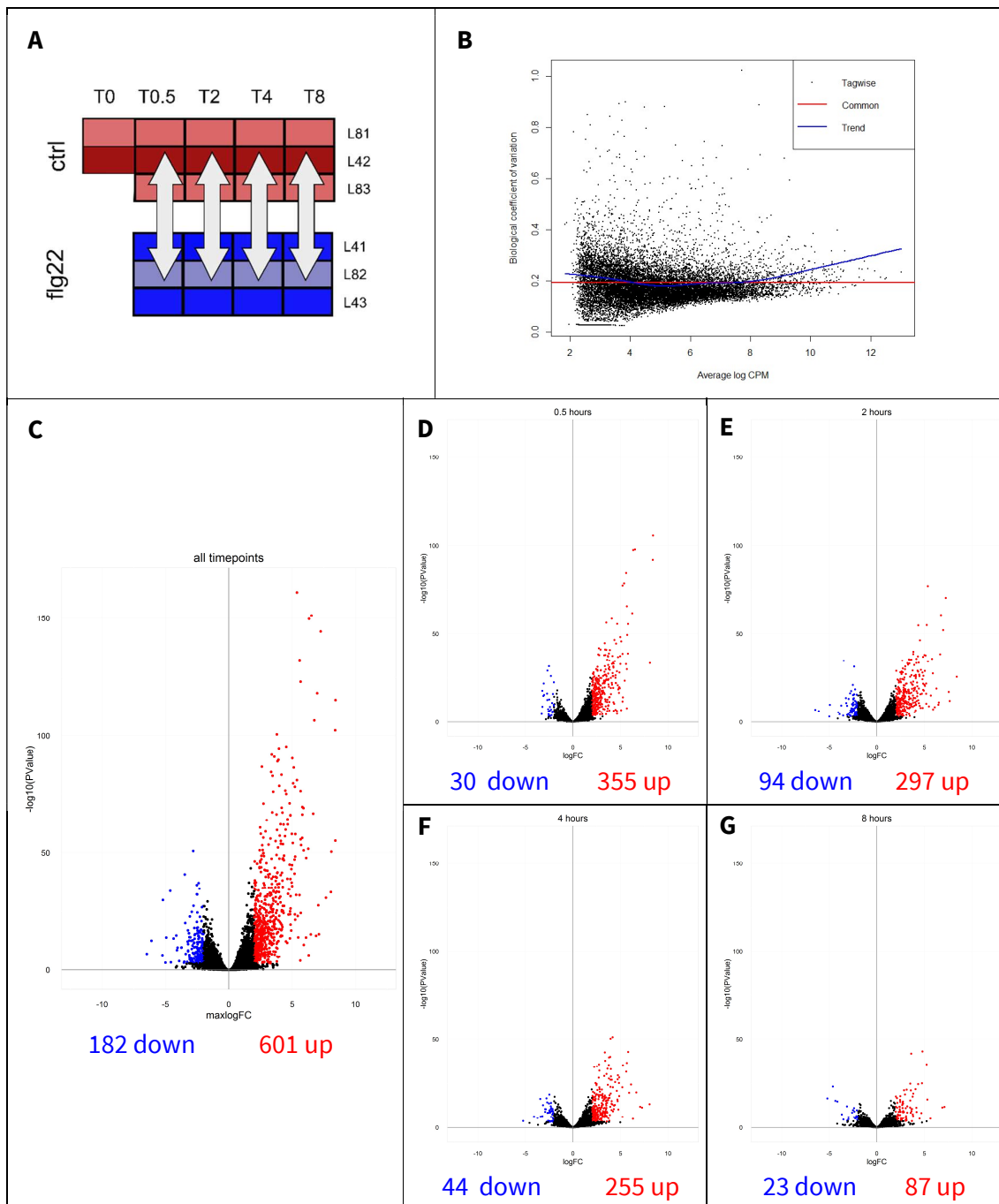
Figure 5.8: Statistical test for differential translation with *edgeR*.
A: Schema of the statistical test. B: Plot of the biological coefficient of variation (BCV) from *edgeR*, displaying common, trended and tagwise dispersions. C: Volcano plot of the statistical test for differences in translational activity after flg22 treatment over all time points. The number of genes with significantly different translation is indicated below the plots. In total 13'730 genes were assessed. D-G: Same analysis as in D, but limited to the sample from each individual time point.

The GO term enrichment analysis gave a similar picture as for the RNA-Seq dataset. The set of translationally upregulated genes shows the full scale of plant defense responses (Table 5.1, Figure **5.9**). The translationally downregulated genes were only enriched for one GO term (*transmembrane receptor protein tyrosine kinase signaling pathway*, 6 genes, adj. p-value 5.88E-03), which does not overlap with the GO categories enriched in the downregulated genes of the RNA-Seq dataset (there, the term water homeostasis is significantly enriched).

Table 5.1: GO term enrichment for the 601 genes with a translational increase after flg22 stimulation during any of the four measured time points.
The enrichment was performed with PANTHER (Mi *et al*, 2016). The set of translationally upregulated genes were compared against all 13'730 statistically assessed genes and a Bonferroni-adjusted p-value cutoff of < 0.05 was used. Only the lowest nodes of the GO hierarchy are listed, parent nodes are left out.

GO BIOLOGICAL PROCESS	# IN REFERENCE	# IN UPREGULATED	# EXPECTED IN UPREGULATED	FOLD ENRICHMENT	ADJ. P-VALUE
chorismate biosynthetic process	8	6	0.35	17.23	3.26E-03
defense response by callose deposition in cell wall	14	8	0.61	13.13	4.79E-04
tryptophan metabolic process	15	8	0.65	12.25	8.01E-04
indole-containing compound biosynthetic process	24	12	1.04	11.49	2.23E-06
response to chitin	73	35	3.18	11.02	1.27E-21
aromatic amino acid family biosynthetic process	32	15	1.39	10.77	4.68E-08
drug transmembrane transport	18	7	0.78	8.93	3.21E-02
hydrogen peroxide catabolic process	26	10	1.13	8.84	5.82E-04
toxin catabolic process	24	9	1.04	8.62	2.77E-03
phenylpropanoid biosynthetic process	37	12	1.61	7.45	2.41E-04
response to karrikin	65	19	2.83	6.72	3.15E-07
glutathione metabolic process	38	11	1.65	6.65	2.38E-03
glucosinolate metabolic process	39	11	1.7	6.48	3.05E-03
flavonoid biosynthetic process	46	12	2	5.99	2.31E-03
response to wounding	121	27	5.27	5.13	2.15E-08
sulfur compound biosynthetic process	56	12	2.44	4.92	1.66E-02
defense response to fungus	131	28	5.7	4.91	2.39E-08
response to toxic substance	60	12	2.61	4.59	3.25E-02
cellular amino acid biosynthetic process	121	23	5.27	4.37	1.40E-05
defense response to bacterium	206	33	8.97	3.68	6.01E-07
response to jasmonic acid	102	16	4.44	3.6	2.81E-02
response to oxidative stress	244	34	10.62	3.2	1.04E-05
innate immune response	181	24	7.88	3.05	4.22E-03
peptide biosynthetic process	443	4	19.28	0.21	4.46E-02
protein localization	403	3	17.54	< 0.2	3.89E-02
DNA metabolic process	308	1	13.41	< 0.2	3.44E-02
RNA processing	335	1	14.58	< 0.2	1.12E-02

Figure 5.9: Visualization of the significantly enriched GO terms of the translationally upregulated genes using VirtualPlant (Katari et al, 2010). Only nodes with a p-value less than 1E-4 are depicted. The nodes are colored according to the enrichment p-value (red = highly significant p-value, green = less significant p-value).

Hierarchical clustering of the flg22-treatment over control fold-changes of the Ribo-Seq counts gave the expected results for the upregulated genes, as cluster *1up* consisted of early responsive genes involved in for instance hydrogen peroxide processes. Cluster *2up* genes were enriched for more long-term responses and genes in this cluster encode ion transporters and metabolic enzymes for tryptophan and chorismate biosynthesis.

Clustering the translationally downregulated genes did not result in significant GO over-representations. Cluster *1down* showing strongest downregulation 2 h post-stimulus was enriched for *transmembrane receptor protein tyrosine kinase signaling pathway* genes.
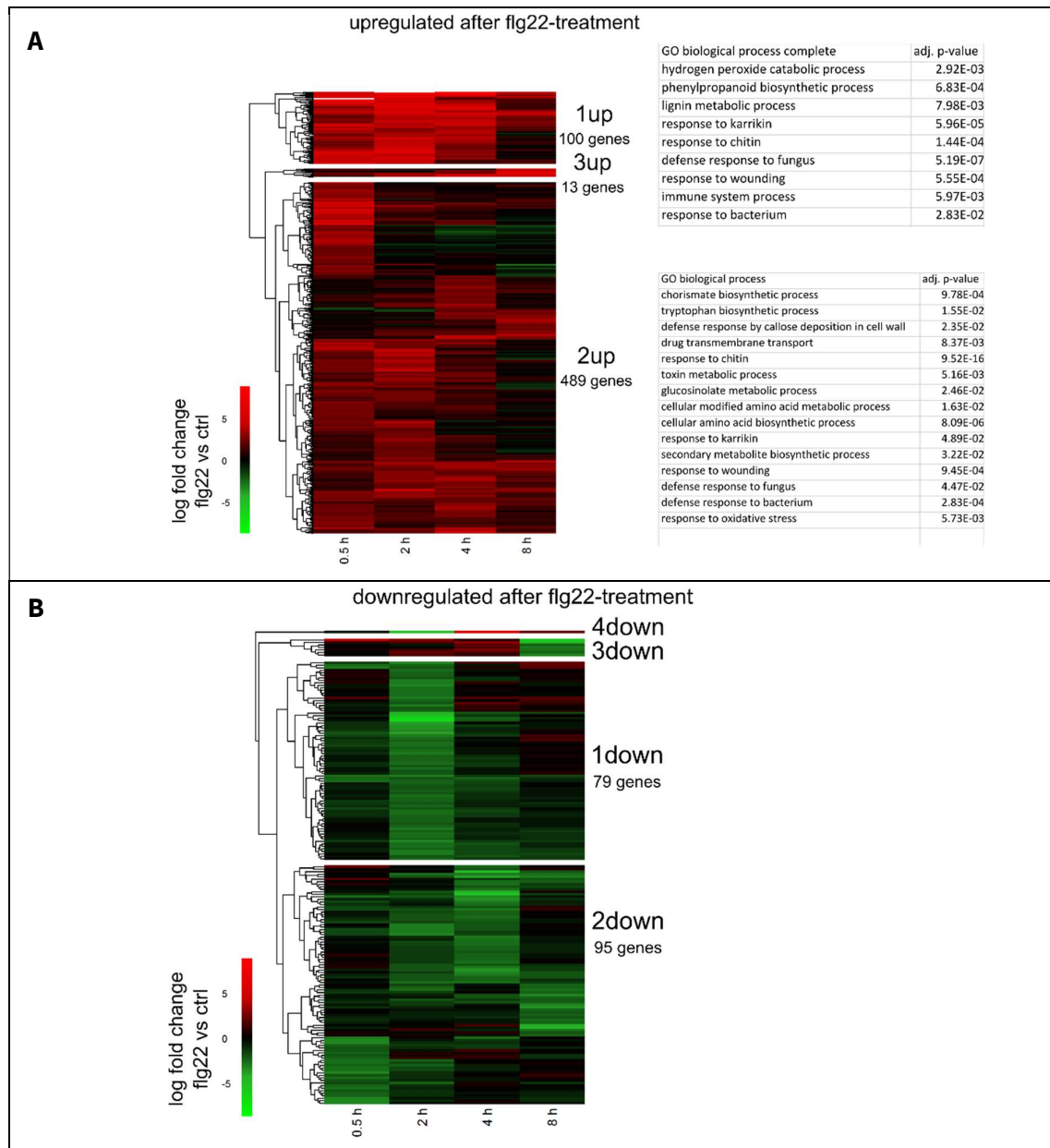

Figure 5.10: Hierarchical clustering of all genes with significantly upregulated (A) or downregulated (B) translation after flg22-stimulation. The significantly enriched GO categories for each cluster are shown next to the cluster.

Similarly, as in the hierarchical clustering, the flg22-induced response leading to a translational increase is more consistent over the four time points than the translational downregulation. Of the 601 significantly upregulated genes, 259 were determined as upregulated in more than one time point (43.1%, Figure 5.11A), while only 9 out of the 182 downregulated genes (4.9 %, Figure 5.11B) were found to be downregulated in more than one time point.
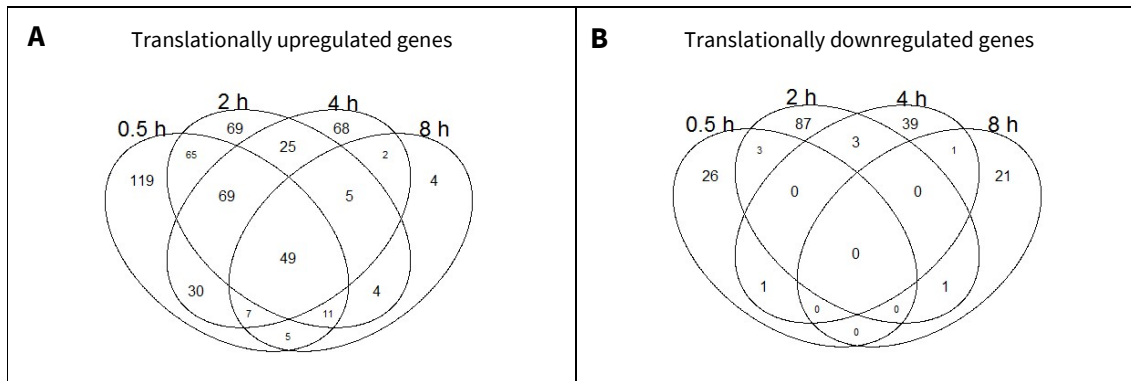

Figure 5.11: Overlap of the genes assessed as upregulated (A) or downregulated (B) at the indicated time points.

5.6   Differential translational efficiencies

Naturally, the transcriptional and translational responses to a stimulus are highly interconnected. The number of ribosome protected fragments (RPFs) depends strongly on the initial number of mRNAs that are present at a given time point. Consequently, RNA-Seq and Ribo-Seq datasets by default correlate strongly with each other, yet the differences between the two data sets are of particular interest, as they point towards translational regulation of gene expression. In the following, we will analyze the relationship between the two sequencing data sets in detail to identify such translational regulation.

I analyzed here the 13'568 genes detected in both datasets, which comprises 99% of all translationally assessed genes and 72% of all transcriptionally assessed genes.

5.6.1   Gene-specific translational efficiencies
As mentioned earlier, the RNA-Seq and Ribo-Seq datasets correlate strongly. Already the count per million (cpm) values averaged over all samples correlate highly across both data sets (rho = 0.838, Figure 5.12A). This distribution is highly similar to the distribution observed in yeast (Figure 1.A in Albert et al. (2014).

The ratio of Ribo-Seq counts to RNA-Seq counts describes the number of ribosomes per transcript, which means the average translational efficiency (TE) for that transcript. In our data set, 95% of the gene-specific translational efficiencies fall within a range of ¼ to 2, differing by a factor of 8 (Figure 5.12B). Genes that deviate strongly from this range (|TE | > 2) were not significantly enriched for any GO category.

It is assumed that ORF length has a strong effect on the number of bound ribosomes per mRNA (Ingolia *et al*, 2013). The longer a mRNA, the more ribosomes are able to bind to it and the longer it takes for each ribosome to translate the ORF fully. A longer ORF is therefore expected to produce more RFPs given the same translational activity. To our surprise, adding ORF length as a linear factor in the correlation analysis increased the correlation only by a small margin (R^2 increases by 0.005, p-value = 1.96e-06). Testing the effect of the gene length directly on the log-transformed TEs did not show a significant correlation using either the spliced mature mRNA length (cDNA length, $R^2 = 0.0001$, p-value = 0.2428, Figure 5.12C) or the ORF length (coding sequence (CDS), $R^2 = 0.00025$, p-value = 0.092, Figure 5.12D).

The length of the coding sequence, therefore, cannot explain the nature of this strong gene-specific effect on TE. For the future, it would be interesting to test other mRNA sequence-specific effects such as the tRNA availability and codon-usage efficiency. In a similar analysis, Lei et al, (2015) detected effects on TE by CDS length, GC content and normalized minimal free energy of the mRNA secondary structure. Juntawong et al, (2014a) propose that gene-specific differences in TE can be explained by translation initiation processes such as start codon recognition and ribosome subunit joining rates.
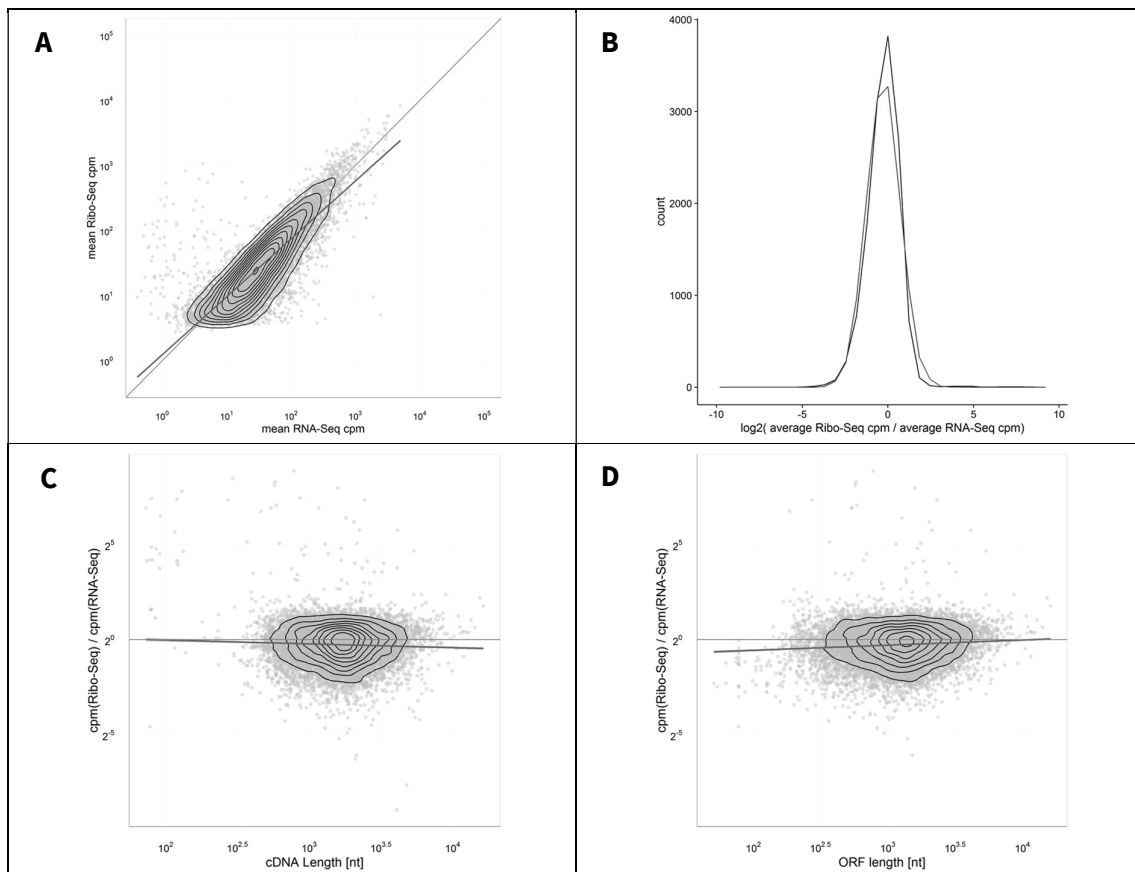
Figure 5.12: Gene-wise translational efficiencies.
A: mean count per million mapped reads (cpm) of the transcriptional RNA-Seq dataset plotted against the cpm of the Ribo-Seq dataset. A linear correlation is indicated in blue ($R^2 = 0.6986$). B: Distribution of the log-transformed, gene-wise translational efficiencies. These efficiencies are not exactly normal distributed (p-value < 2.2e-16, Shapiro-Wilk normality test), but can be roughly described by the fitted normal distribution in red

(mean -0.2603, SD = 0.9727). C: Plotting the gene-wise translational efficiency against the length of the cDNA. In blue, the linear regression is plotted. D: Same as C, but plotted against the length of the open reading frame (ORF).

5.6.2   Correlation of transcriptional and translational flg22-induced fold-changes

The translational and transcriptional data sets correlate strongly when assessing the flg22-induced fold-changes (Figure 5.13A). Following the classification from Albert et al. (2014, Figure 5.13B), genes whose FC ratios follow the identity line imply unchanged translational efficiencies. Deviations from the identity line indicate either reinforced translation (steeper line) or translational buffering (shallower line).

The correlation coefficient is highest at the earliest time point and decreases with the progression of the experiment time (Spearman rho 0.5 h = 0.746, 2 h = 0.674, 4 h = 0.607, 8h = 0.337), as fewer and fewer genes are differentially transcribed and translated. The non-regulated genes make up a larger proportion of the analyzed genes at the later time points, resulting in a mostly stochastic distribution of log2-transformed FC values close to 0. Still, the correlation at any time point is highly significant, resulting in the smallest p-value that the Spearman's rank correlation test computes in R (p-value < 2.2e-16). These values are comparable to the data on drought-stress treated maize seedlings (Lei et al, 2015), where positive correlations between the fold changes at transcriptional and translational level were reported (R^2 = 0.69).

The data also highlight that the upregulation reaches higher fold-changes than the downregulation response on both transcriptional and translational level (Figure 5.13A).

Additionally, I observed a slight decrease in translational efficiency for the transcriptionally highly responsive genes (Figure 5.13B), indicative for translational buffering of the transcriptional response. Transcriptionally highly upregulated genes (on the right side of the plot) show a translational efficiency decrease of up to 2^{-1} = 50 %, exemplified in Figure 5.13C and D. The translational buffering observed here is less pronounced than in hypoxia–treated Arabidopsis seedlings (Figure 5.13C, Juntawong et al. (2013a)).
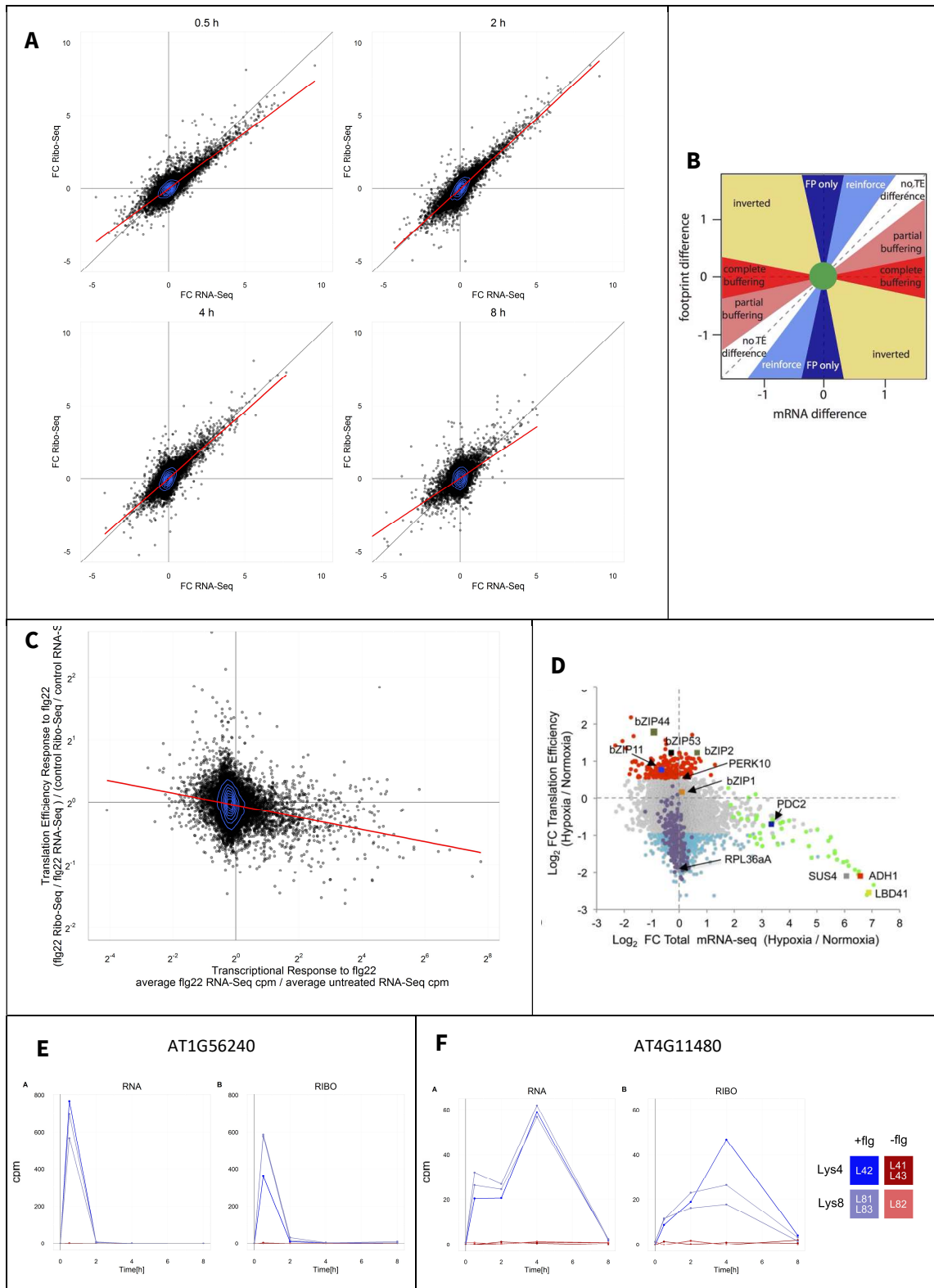
Figure 5.13: Genome-wide translational efficiency responses.
A: Comparisons of the log2-transformed fold-changes in the RNA-Seq and Ribo-Seq data sets at the four time points. The density of the points is indicated by blue contour lines, to give an estimate for the over-plotted area. Linear regression lines are plotted in red

(R^2 T1 = 0.7385, T2 = 0.7524, T4 = 0.6288, T8 = 0.3481. B: Classification of translational efficiency (TE) regulation (Albert *et al,* 2014). FP = footprints = Ribo-Seq counts. B: Correlation of the transcriptional response to the response in translational efficiency. Genes that are strongly transcriptionally upregulated (right side of the plot) have a decreased translational efficiency in flg22-treated cultures. The correlation indicated by the red line representing a linear regression is highly significant (p-value < 1E-16) at a low correlation efficiency (R^2 = 0.00839). The distribution of the points is indicated by the blue contour lines. D: For comparison, the translational response of hypoxia-treated Arabidopsis seedlings (Juntawong *et al,* 2013). E and F: Transcriptional and translational data of two exemplary genes that show a translational buffering effect.

5.6.3 Overlap of flg22-induced differentially transcribed or translated genes

Comparing the genes with significantly different transcription or translation we observed a large overlap (Figure 5.14A). Again, the overlap is higher for the upregulated genes (69 %, Figure 5.14B) than for the downregulated genes (54 %, Figure 5.14C) due to the larger flg22-induced FC of the upregulated genes.

Over the time course, the overlap slightly decreases from 27% of all regulated genes to 13% 8 h post treatment (Figure 5.14D). It again seems that the datasets align better when the biological signal is the "strongest". The proportion of shared significantly different transcripts are slightly lower than what was reported for drought-stressed maize seedlings, where 26% of the upregulated and 39% of the downregulated genes were assessed as differentially regulated in both the transcriptional and translational data set (Lei *et al*, 2015).

We inspected every gene that was significantly different in in one but not the other dataset and could not detect any clear example of a transcript that diverges in the translational response from the transcriptional response. Figure 5.14E-F exemplifies such cases, where the significance assesments do vary between the transcriptional and translational dataset, yet the overall trend is consistent in both datasets. In my assessment the main reason for the decreased number of significantly changing transcripts in the Ribo-Seq dataset is primarily caused by the increased technical variation of that dataset and not by a biological reason.
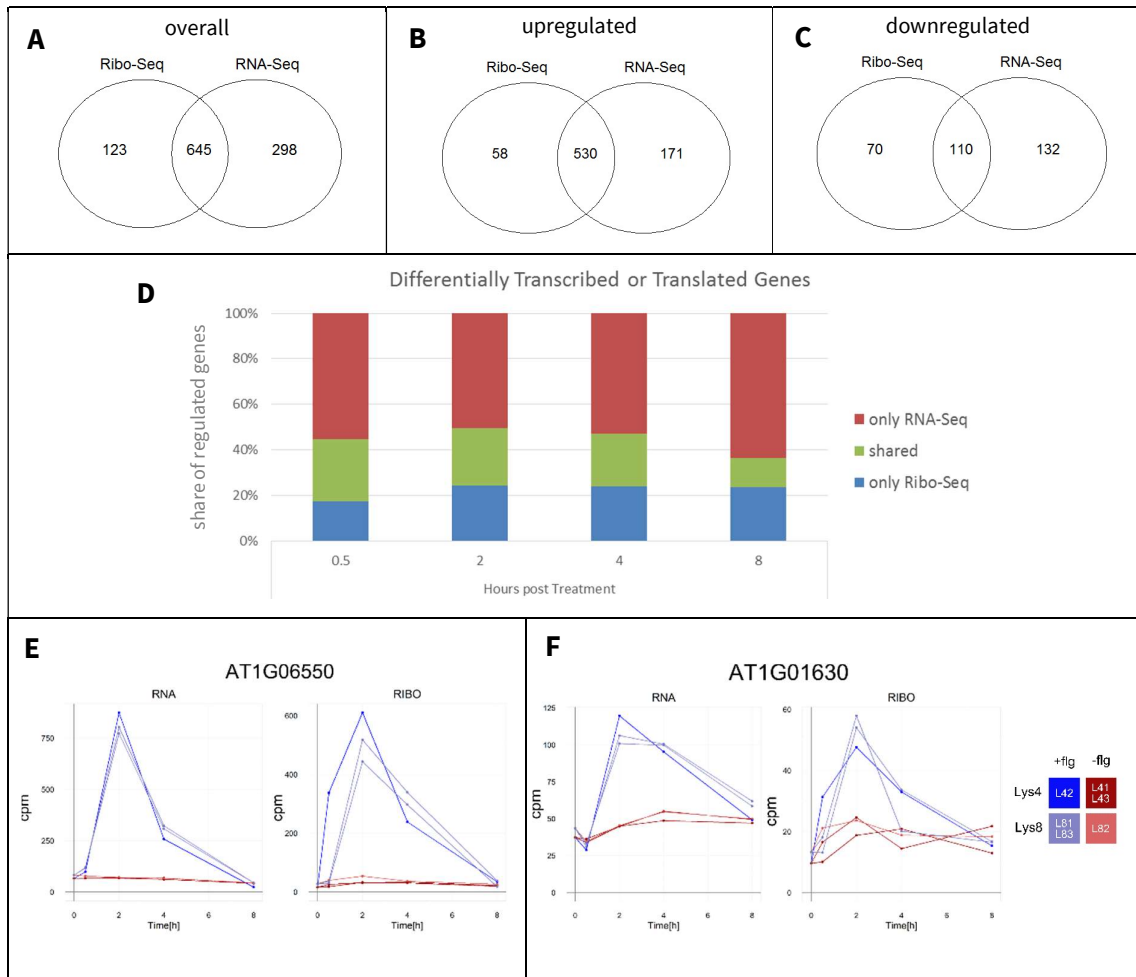
Figure 5.14: Comparison of significantly differentially transcribed or translated genes upon flg22-stimulation.
A: Overlap of all genes assessed as significantly differentially transcribed (RNA-Seq) or translated (Ribo-Seq) at any time point. The genes in plot A were divided into upregulated (B) or downregulated (C) genes. D: Ratios of the number of shared or uniquely significantly regulated genes at the individual time points. E-F: Two example genes that were assessed as significantly upregulated 0.5 h after flg22 treatment in the Ribo-Seq dataset but not in the RNA-Seq dataset.

5.6.4 Gene expression regulation through Flg22-induced changes in translational efficiency

5.6.4.1 Interactional model

To test more rigorously for differences in translational efficiency we used the same statistical approach as was used for yeast samples by Albert et al. (2014). We created the null hypothesis that the number of reads in cpm depends on lysine-labeling, treatment, time point and the type of sequencing library (Ribo-Seq or RNA-Seq). We then tested with an ANOVA if the alternative hypothesis H1 with an interaction term between the library-type and the flg22-treatment describes the cpm in a better way.

$$H0: \quad counts \sim library\ size + type\ +\ lysine\ +\ time\ +\ flg22$$

$$H1: \quad counts \sim library\ size + type\ +\ lysine\ +\ time\ +\ flg22 + type : flg22$$

In other words, this test asks for each gene whether the ratio of RNA-Seq to Ribo-Seq cpm is significantly different for the flg22-treated samples at any time point regardless of a potential lysine-labeling effect.

I adjusted the resulting p-values for multiple testing by FDR and obtained 29 genes with an adjusted p-value < 0.05 (Figure 5.15A). These genes were not significantly enriched for any GO category (BH adjusted p-value < 0.05). Lowering the FDR-adjusted p-value threshold to < 0.1 or taking the 100 most significantly regulated genes did also not result in a GO term enrichment. However, this was not necessarily expected as a translational regulation response does not have to be described by GO categories.

Visual inspection of the highly significant cases of translational regulation resulted only in few convincing cases where the Ribo-Seq data showed different trajectories than the RNA-Seq data. To our surprise, all of the strongly differing cases were from antisense transcripts, which are presumed not to be translated. When checking the Ribo-Seq counts from their overlapping sense transcripts, it became obvious that they showed the same trajectories. The mapping of the RPFs was done in a non-stranded fashion so that the Ribo-Seq reads were counted both for the sense and antisense strand transcripts ("multimapping").

While these results are unsatisfying for identifying translational regulation, it shows that these cases can be used as a positive control for the statistical test. I can identify genes where the RNA-Seq and Ribo-Seq trajectories differ significantly, yet we don't seem to have clear cases in our data set.

When inspecting the significantly translationally regulated cases, we had the impression that they were enriched for transcriptionally downregulated genes. Yet that could not be verified statistically, as I do not see a significant difference in the p-value distribution for the transcriptionally downregulated or upregulated genes (p-value 0.4366, Wilcoxon rank sum test).
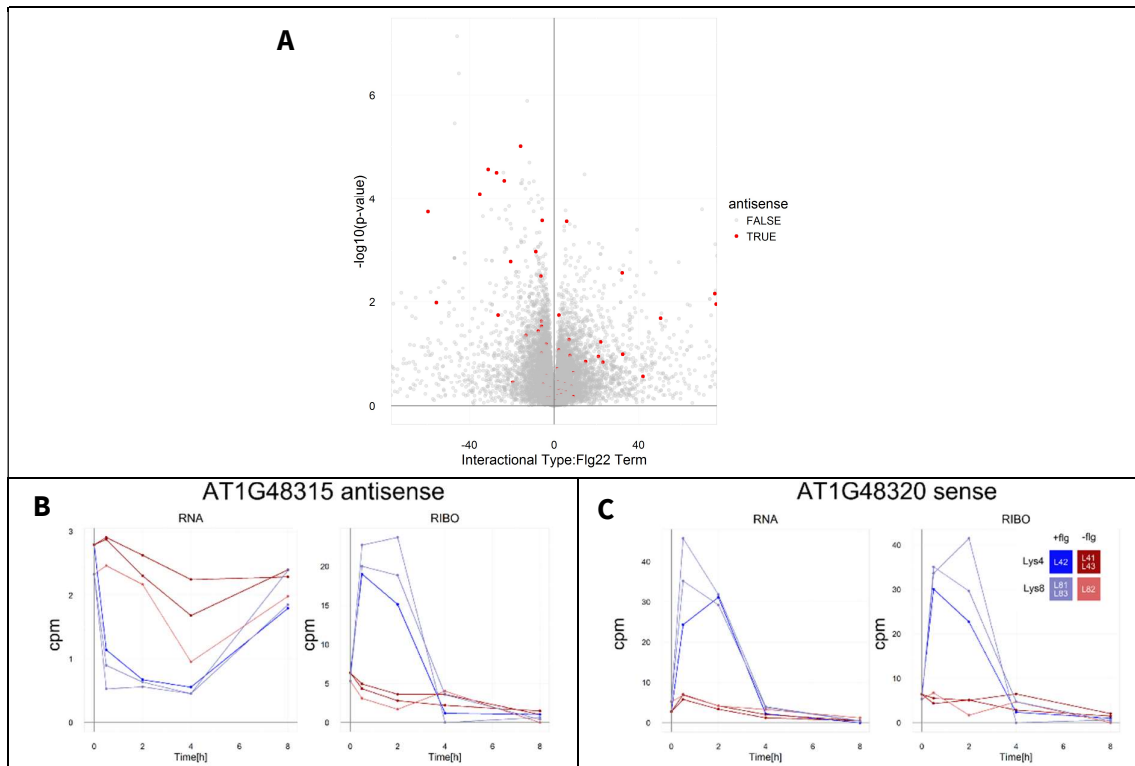
Figure 5.15: Results of the interactional model.
A: Pseudo-volcano plot of the results with the interactional model. The size of the interactional term is plotted on the x-axis. Genes annotated as antisense are highlighted in red. B-C: Transcriptional and translational profile of a sense-antisense gene pair. The antisense gene (B) is assessed as significantly differentially translated, yet the Ribo-Seq reads actually originate from the sense transcript (C), which does not indicate any differential translation.

5.6.4.2 ANOTA

We also tested additional statistical approaches to identify cases of translational regulation. Larsson et al. (2010) published an approach called analysis of translational activity (ANOTA) based on an analysis of partial variance, which produces estimates of translational activity that are independent of cytosolic mRNA levels. This approach should be superior to calculating the ratio of RNA-Seq counts over Ribo-Seq counts as this can lead to spurious correlations. Inspection of the genes assessed as differentially translated by ANOTA resulted only in the previously observed antisense genes and non-convincing results.

5.6.4.3 Fold-change model

We also used an ANOVA to compare the flg22-induced fold-changes from both data sets. The test should identify genes where the sequencing library type had a significant effect on the flg22-induced fold-change.

$$H0: \quad fold\ change \sim time$$

$$H1: \quad fold\ change \sim time + type$$

Using this approach, we could not even identify the antisense genes. Most genes identified as significantly regulated were spuriously transcribed or translated and were considered to be false positives.

5.6.4.4 Xtail

Generally, there are two parameter pairs to compare when identifying stimulation-induced translational efficiency changes. A: RNA-Seq vs Ribo-Seq and B: treatment vs control. As statistical tests rely on probability distributions, the order of both calculations can affect the outcome of the test. The third employed statistical approach, Xtail (Xiao *et al*, 2016), therefore uses both orders separately (A followed by B (Figure 5.16A) and B followed by A (Figure 5.16B)) to compute for each gene a probability distribution. It then takes the more conservative of the two p-values. It also assumes that the distribution of the Ribo-Seq and RNA-Seq counts follow a negative binomial distribution, similarly to *edgeR*.

Xtail compares maximally one factor - treatment against control. We, therefore, omitted the lysine-labeling factor as it has only a small effect on the sequencing data sets and performed the Xtail assessment on the samples of each time point separately. In a volcano plot displaying the minimal p-value at any time point and the corresponding fold change, we again identified the antisense genes (Figure 5.16C). The next most significantly regulated gene is RAP (AT2G31890), which overlaps with the flg22-responsive SRRLK (AT2G31880). The translational data of RAP mirror the SRRLK data, indicating again a technical artifact of the mapping.

Xtail assessed 82 genes as significantly differently translated at a p-value cutoff of 1E-4 and 252 genes at a p-value cutoff of 1E-3. Both gene lists were not significantly enriched for any GO term (FDR-adjusted p-value threshold of 0.05).
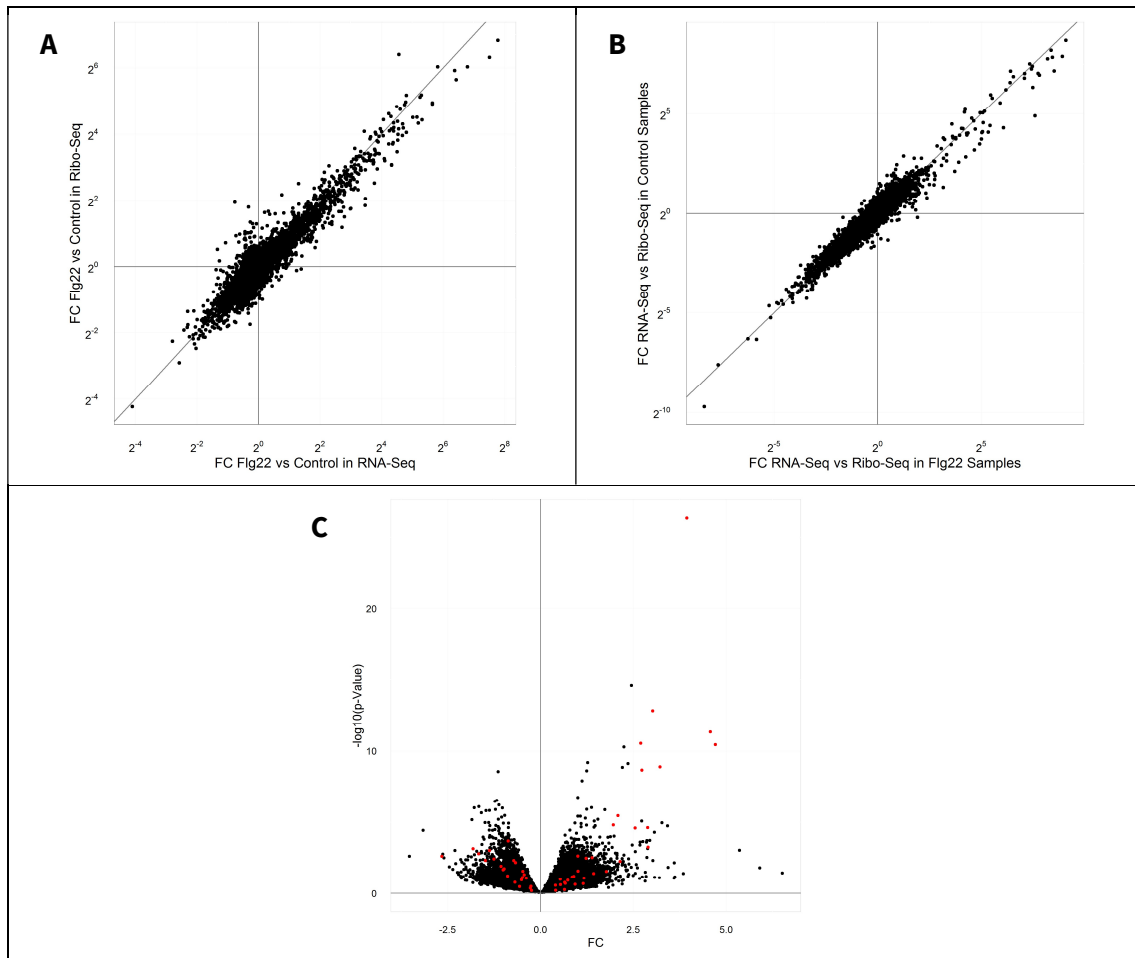
Figure 5.16: Xtail test.
A, B: Explanatory plots for the two ways of assessing translational regulation. Either, the treatment to control fold-changes are computed first (A), or the Ribo-Seq to RNA-Seq fold-changes (B). C: Volcano plot from the Xtail test, printing the lowest p-value at any time point and its corresponding log2-transformed fold-change. Antisense transcripts are highlighted in red.

5.6.5 Differentially translated genes according to Xtail

In the next step, we verified the genes, which were significantly regulated on the translational level according to Xtail. The genes that only showed spurious translational changes when all time points were considered were excluded as well as those that overlapped with another gene's ORF. In the end, eleven putative translationally regulated genes remained. None of these genes contain a uORFs according to the analysis by Hu *et al,* (2016).

In the following section, we will discuss the eleven genes with differential translational efficiencies, addressing them with the single-letter-codes as depicted in Figure 5.17.

Five genes (A, B, C, D, E) depict a pattern of enhanced translation after the transcriptional burst has passed and the number of transcripts has decreased again. The translational efficiency subpanels show a decreased efficiency at the early time points that then gets strongly compensated at the later

time points. The five genes do seem to share a common characteristic. The gene of A encodes for a phenylalanine lyase (AT3G53260), B an unknown protein associated with Geminivirus infection defense (Ascencio-Ibáñez *et al*, 2008), C the membrane-bound ATP-binding cassette transporter G36 (AT1G59870) that acts as a Cd^{2+} pump and is activated upon infection with Pseudomonas (Xin *et al*, 2013), D a largely undescribed leucine-rich repeat transmembrane protein kinase (AT1G56140), that is also responsive to Geminivirus infection (Ascencio-Ibáñez *et al*, 2008), and E an undescribed plasma-membrane-bound long coiled-coil protein with unknown molecular function (AT2G32240).

F and G display a decreased translational efficiency upon flg22 treatment. In F, the translation of the kelch repeat superfamily protein with unknown function (AT4G39570) does not respond to the initial transcriptional burst. The trajectories for G show the constantly reduced translation efficiency of the tubulin alpha-2 chain (AT1G50010) when treated with flg22. The translational regulation becomes particularly clear in the translational efficiency subpanel. In contrast, transcription, translation, and TE of the tubulin alpha-1 chain (AT1G64740) remains unaffected by flg22 treatment.

The genes shown in H and I encode the two EIN3-binding F-box proteins 1 (EBF1, AT2G25490) and 2 (EBF2, AT5G25350) that are described to be regulated post-transcriptionally in the ethylene signaling pathway (Merchante *et al*, 2015). In both genes, the flg22-induced transcriptional response does not result in a corresponding translational response at 2 h post treatment.

In J, the gene encoding the lectin protein kinase (AT1G61370) shows a strong translational response at T1 without a transcriptional response. K shows a strong transcriptional downregulation response of the oxidation-related zinc finger 1 (AT2G19810), which is dampened on the translational level. Interestingly, both genes have been identified as two of the four strongest differentially expressed bacterial defense response regulators when comparing the wild-type response with the *constitutive defense without defect in growth and development 1* (cdd1) mutant to bacterial infection (Swain *et al*, 2015). The genetic mutation of cdd1 is currently undetermined. The translationally upregulated AT1G61370 and dampened AT2G19810 identified here were identified by Swain at al, (2015) as negative and positive regulators of bacterial defense response, respectively.
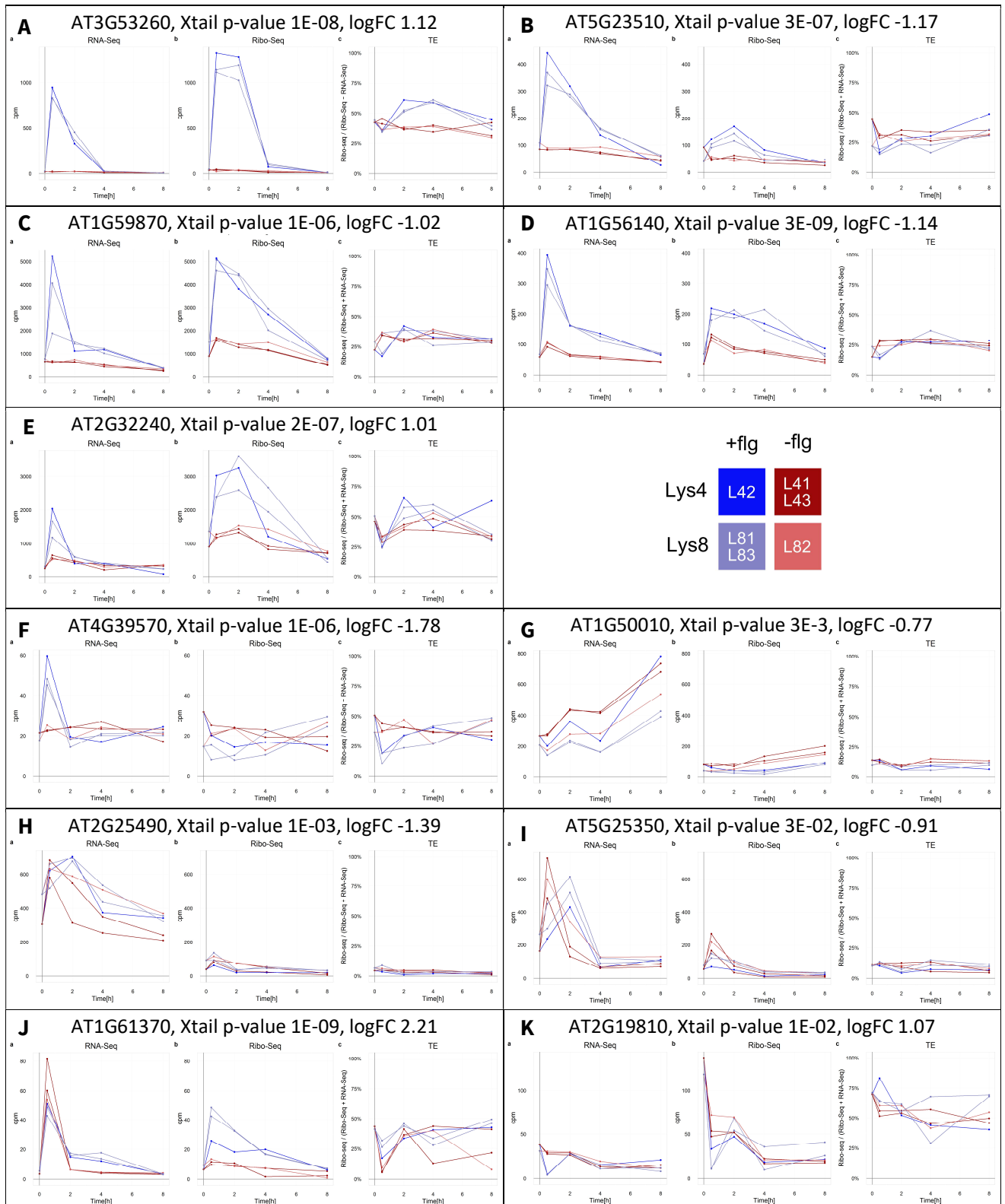
Figure 5.17: Verified genes with differential translational efficiencies upon flg22 treatment at any time point as determined by Xtail.
A-K: For each of the eleven genes, the RNA-Seq (a, left subpanel), Ribo-Seq (b, middle subpanel) and translational efficiency (c, right subpanel) data are displayed.

6.

PROTEIN ABUNDANCE

6.1 Development of plant SILAC

Motivated by the plant SILAC study by Schütz and colleagues (2011), I set out to adopt their labeling strategy in our laboratory by 1) using a medium with low nitrogen content, 2) supplementing the culture with a high daily dosage of 350 µM lysine and 3) cultivating the cells in the dark. However, I encountered serious challenges when establishing this approach. Our in-house cell line (*Arabidopsis thaliana* A, abbreviated as AtA) showed decreased growth or even full growth arrest under the published conditions, especially when passaging the culture to fresh medium at day 6 of the labeling regimen. To improve the culture's viability, we modified various culturing conditions.

I noticed that inoculation density is a crucial factor for high viability when starting the culture. Especially in the first days of the cultures, the supplemented lysine does not get taken up fully by the low-density cultures (Mustafa et al, 2011), leading presumably to a too high accumulation of lysine. In my experiments, it turned out to be crucial to inoculate fresh cultures at >30 mg fresh weight / mL medium.

Also, the culturing vessel is an important factor, as it affects the dispersal of the cells in the medium. Cells accumulating at the bottom of the vessel are not aerated properly and grow only marginally. Therefore, cell dispersion has to be assured by optimizing the rotation speed of the orbital shaker, the diameter of the culture's growth vessel and the maximal volume of medium filled into the vessel. For our culture, I found that Erlenmeyer flasks with their steep walls provide efficient dispersal and aeration when they are filled less than 30% of the maximal volume and the flasks are shaken at 120 rpm. In my hands, Erlenmeyers of different sizes performed equally well, which allows for easy scaling of the culture volume, as long as the medium does not exceed 30% of the Erlenmeyer's total volume.

6.1.1 Lysine supplementation and culture growth in plant SILAC

After initial problems using our in-house AtA cell line, we obtained the cell line used in Schütz *et al.* (2011), named AtB, from the co-first author Dr. Niklas Hausmann (Laboratory of Prof. Rüdiger Hampp, University of Tübingen). Comparing this cell culture to our in-house AtA line, we observed substantial physiological differences. AtB cultures grow to higher culture densities (>400 mg fresh

weight / mL medium after 7 days of growth in complete Murashige & Skoog medium, compared to <190 mg fresh weight / mL medium for the AtA line, Figure 6.1A). AtB forms fewer macroscopically observed cell aggregates (Figure 6.1B) and the paler color of the culture suggests a reduced chloroplast count. AtB also does not turn fully green when grown in the light, although it is described to be derived from leaf tissue.
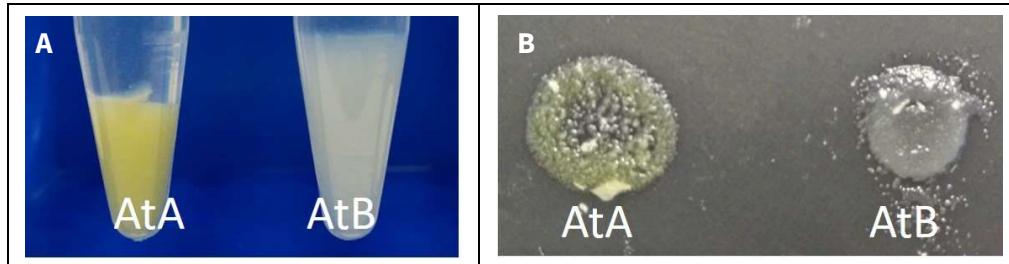

Fig. 6.1: Phenotypic comparison of the in-house Arabidopsis suspension cell culture AtA and the AtB culture used in the study of Schütz *et al.* (2011).
Under the same culturing conditions (complete Murashige & Skoog medium, 7 days of growth), AtB grows to nearly twice the density as AtA (A) and forms fewer cell aggregates during cultivation (B).

Employing the published labeling scheme on the AtB culture did lower the propensity of growth arrest during the labeling regimen, however, it did not prevent it completely. Infrequent culture collapses during the labeling procedure were also experienced by the authors of the Schütz et al. (2011) study (Dr. Niklas Hausmann, personal communications). They did not employ the labeling for more than 12 days.

As discussed in chapter 2.3.8, lysine inhibits also aspartate kinase (AK, Heremans & Jacobs, 1997), whose product 4-phospho-L-aspartate is required for the biosynthesis of lysine, threonine, methionine, and isoleucine. We speculate that these feedback regulations lead to imbalances of the amino acid pools, which could impact culture viability. However, when adding the amino acids produced by AK (5 mg/L of each methionine, isoleucine, and threonine) to lysine-supplemented cell cultures, we could not detect any benefits on growth or viability.

Lowering the daily dosage of lysine, however, did increase the culture's viability, yet led to low labeling efficiencies that are unsuitable for in-depth SILAC studies. After 12 days of labeling, we obtained 64% labeling efficiency when supplementing with 175 µM lysine and 72% labeling efficiency when supplementing with 262 µM lysine. Extending the labeling period to 19 days did not further increase the labeling efficiency (73% labeling after 19 days of daily supplementation with 262 µM lysine).

We conclude that the amount of supplemented lysine is a crucial factor for a plant SILAC experiment. A high rate of supplementation has detrimental effects on the culture's viability, yet it is also needed to obtain high labeling rates by downregulation of the endogenous lysine biosynthesis through inhibition of DHDPS and unfortunately also AK, and by outcompeting endogenous lysine during protein biosynthesis. Consequently, in design of the experiment, the lysine supplementation has to be finely adjusted to the growth rate of the employed cell culture.

6.1.2   Chloride in plant SILAC

While optimizing the culturing conditions, I also looked into the effect of chloride ions on cell culture growth, as we repeatedly detected lower viability in AtA cultures supplemented with heavy lysine compared to cultures supplemented with Lys0 (Figure 6.1A). This detrimental effect was equally strong for Lys4- and Lys8-labelled cultures. Lys4 and Lys8 are provided as crystals with 2 mole Cl^- per mole lysine, while Lys0 is provided in crystals with 1 mole Cl^- per mole Lys0. I therefore analyzed whether increased Cl^- concentrations can impact culture viability.

I determined that salt stress of less than 20 mM NaCl showed no significant effect on the growth of Arabidopsis suspension cultures and that the K_{50} was at 50 mM NaCl (Figure 6.1B). Adding 350 µM lysine daily for 6 days increases the Cl^- concentration in the medium from 6 mM (baseline MS medium) to 11 mM when supplementing with Lys4 and Lys8, compared to 8.5 mM for Lys0. These Cl^- concentrations are therefore too low to explain the differences in culture viability, which thus remain elusive.

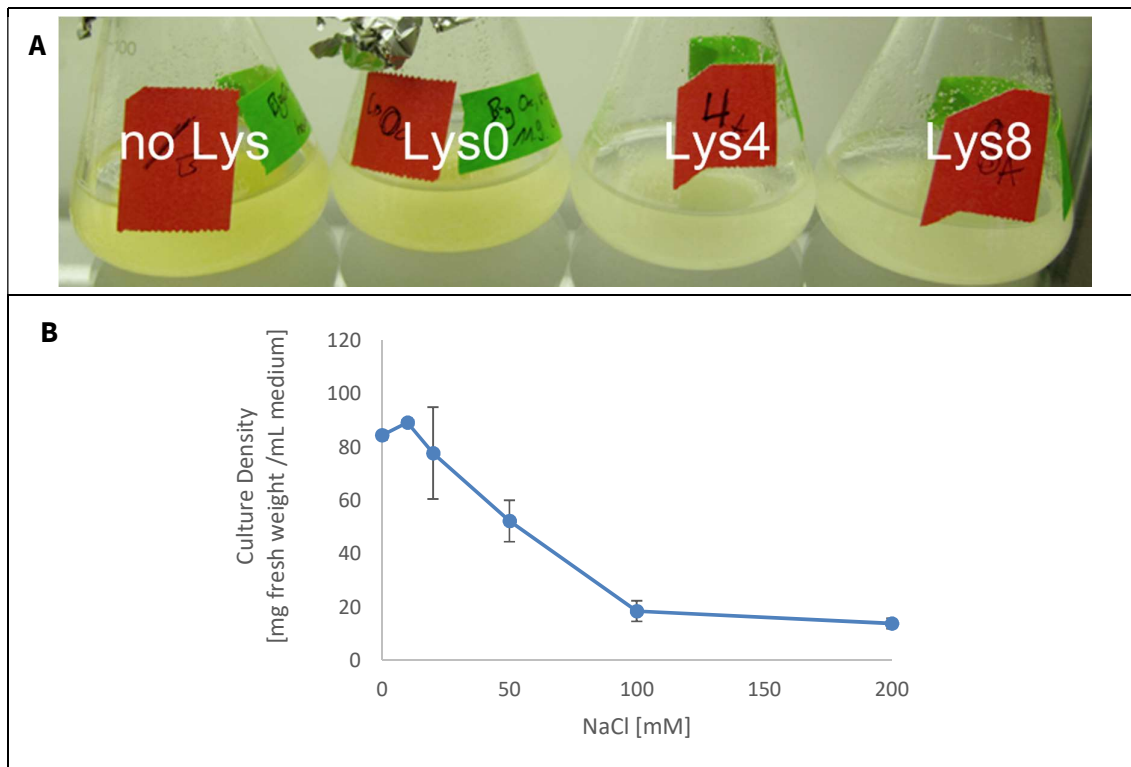**Figure 6.1. Effect of chloride ions on the cultivation of AtA.**
A: Lys4- and Lys8-supplemented AtA cultures show severe growth deficiencies when they are supplemented for 12 days with daily doses of 350 µM lysine. The decolored, whiter cultures indicate partially dead, non-proliferating cells.
B: NaCl inhibits the growth of 8 day old AtA cultures. The cultures were inoculated with 20 mg fresh weight cells per mL medium.

6.1.3   Nitrogen starvation in plant SILAC

Even with all settings as close as possible to the original study (cell line, phytohormone composition, culturing vessels, etc.), the cultures frequently did not survive 12 days of labeling - also in the hands of an independent experimenter (Doris Russenberger, ETH Zurich). Often, they stopped growing after the passage on day 6. Therefore, I tested a crucial setting of the plant SILAC method, the extent of nitrogen starvation. We grew AtB in Murashige-Skoog media with different amounts of supplied KNO_3 and NH_4NO_3, ranging from 20 mM (Schütz *et al*, 2011) to 56 mM nitrogen atoms as in the full Murashige-Skoog medium (Table 6.1). The cultures were grown for more than 12 days and were passaged on day 6. The extent of nitrogen starvation affected the culture density, labeling efficiency and viability of AtB cultures (Figure 6.2).

Table 6.1: Tested nitrogen starvation conditions

Total N-Atoms	20 mM	29.5mM	39 mM	56 mM
KNO_3	0 mM	4.5 mM	9 mM	18 mM
NH_4NO_3	10 mM	12.5 mM	15 mM	19 mM
Nitrogen content relative to full Murashige & Skroog medium	35%	53%	70%	100%
	used by Schütz et al. (2011)			full Murashige & Skoog medium

The culture density increased with the amount of supplied nitrogen. When more than 29.5 mM nitrogen were supplied, the culture did reach a stationary phase of > 400 mg fresh weight / mL medium on day 13 of the cultivation (Figure 6.2A). Reducing the nitrogen supply to 20 mM lowers the culture density substantially. In this experiment, we saw no significant effect of the lysine supplementation on the culture density. The nitrogen supply therefore seems to have a stronger effect on culture growth than the lysine supplementation.

On day 12 of the labeling, we determined the labeling efficiency. It is negatively correlated with the culture density: The more nitrogen is available, the lower the labeling of the cultures (Figure 6.2B). Increasing the nitrogen supply from 20 mM to 56 mM lowered the labeling by 10 percent points from 86.4% to 76.4%. On average, the labeling reduces by 0.25 percentage points per 1 mM additional nitrogen.

The viability of the cultures was determined by the reduction of TTC (2,3,5-triphenyl tetrazolium chloride) to formazan as a measure of the activity of mitochondrial oxidoreductase (Towill & Mazur, 1975; Castro-Concha *et al*, 2006). The reduction activity depends on the amount of viable cells and their metabolic activity.

At day 9 of the labeling, all cultures – treated with lysine or not - showed a strong reductase activity (Figure 6.2C). Cultures with a low nitrogen supply showed a trend for reduced reductase activity. This could result either from the lowered culture density and growth rate or from a reduction of the culture's viability. This effect was more pronounced on day 13 of the labeling. Cultures with 20 mM nitrogen supply showed reductase activities close to the detection limit, indicating full growth arrest and collapse of the cultures. Cultures with more nitrogen supply showed linearly more reductase activity.

Supplementing the cultures with lysine also had an effect on the viability of the cultures. Cultures without any lysine supplementation showed the highest reductase activity, suggesting a negative effect of lysine supplementation on culture viability. Supplementation with Lys0 or Lys4 showed a similar reduction of reductase activity.

Cultures treated with full Murashige-Skoog medium (56 mM nitrogen) were not influenced by the addition of lysine, but showed a generally low reductase activity. This could indicate that the cultures reached the stationary phase of the growth curve, where the metabolic activity is reduced. Yet, the 56 mM nitrogen cultures still had a higher reductase activity than the 20 mM nitrogen cultures that had collapsed.

This interpretation was corroborated by the macroscopic physiology of the cultures (Figure 6.2D). Cultures with a low nitrogen supply are whiter in color. Healthy AtB cultures grow denser and show a yellow-brown color, similar to the cultures with elevated nitrogen supply.
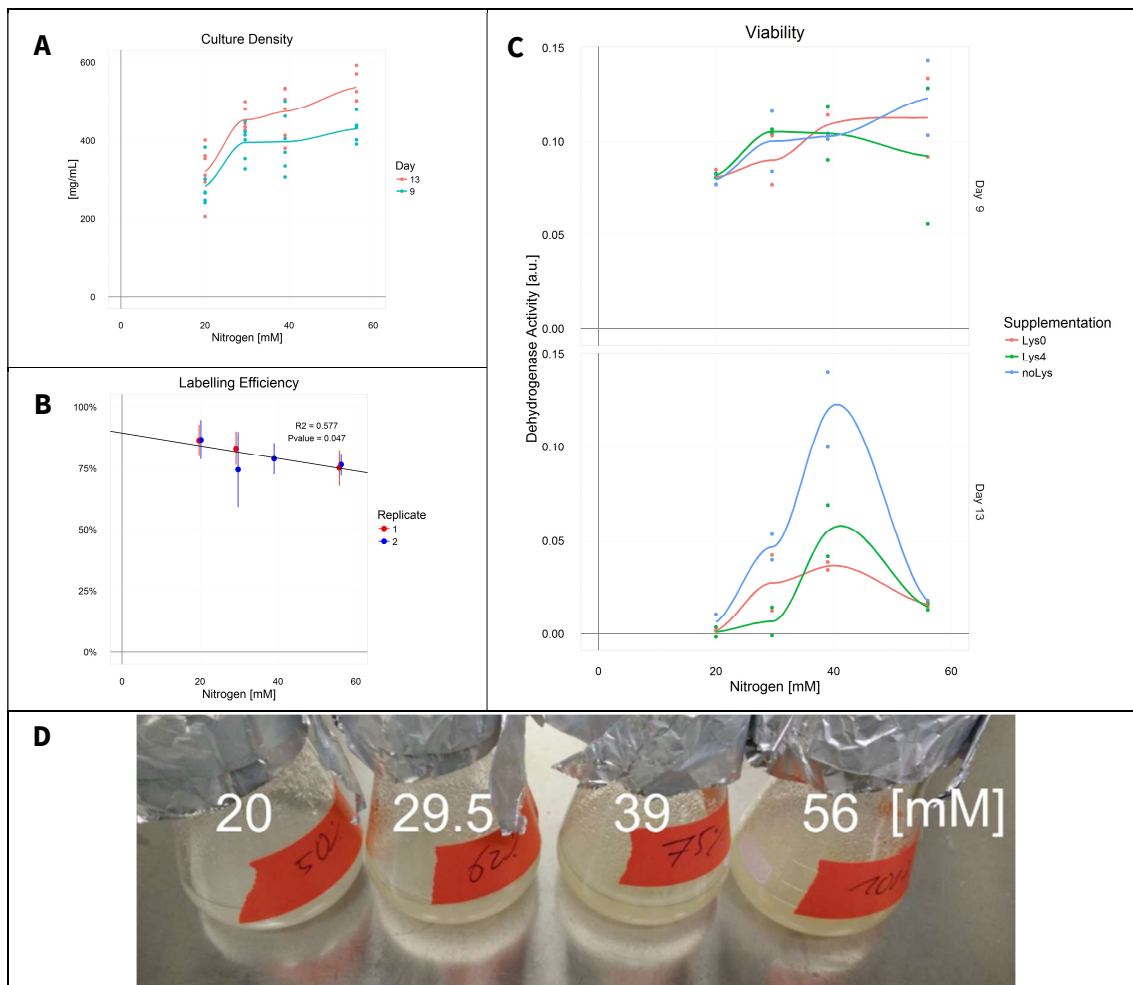
Figure 6.2: Effect of nitrogen starvation on culture density (A), labeling efficiency (B), reductase activity (C), and physiology (D) of 9 to 13 day old AtB cultures.
The cultures were grown for up to 13 days in medium with different amounts of KNO_3 and NH_4NO_3 (**Table** 6**.**1) while being supplemented daily with 350 µM Lys0, Lys4 or no lysine. The cultures were inoculated at 30 mg fresh weight / mL medium and were passaged at day 6. The cultures depicted in D are the control cultures not supplemented with lysine.

We conclude from this experiment that reducing the nitrogen supply to 20 mM affects the cell cultures negatively and ultimately leads to full growth arrest. Increasing the nitrogen supply increases the viability of the culture substantially, leading to higher culture densities and viability while reducing the labeling efficiencies only gradually.

For all subsequent experiments, we optimized the labeling efficiency and culture viability by increasing the nitrogen supply to 14 mM NH_4NO_3 and 7.5 mM KNO_3, which sums up to 35.5 mM nitrogen atoms or 63% of the full Murashige-Skoog medium. These particular settings are optimized for the cell line AtB and might not be directly transferable to other cell lines.

6.1.4 Alternative SILAC approaches

In parallel to the labeling strategy of Schütz and coworkers (2011), we explored the following alternative SILAC labeling approaches:

6.1.4.1 SILAC in seedlings

Lewandowska *et al*. (2013) reported a SILAC labeling strategy for light-grown Arabidopsis seedlings reaching >95% labeling. They grew seedlings from about 15 mg of seeds for 19 days in unmodified liquid Gamborg B5 growth medium containing 25 mM KNO_3, without any reported viability decrease. The medium was supplemented with 1 mM of labeled lysine and arginine and was exchanged every 2 days with fresh, supplemented medium.

When we reproduced this study, the germinating seedlings grew with highly variable rates and impaired viability. Because lysine supplementation also inhibits AK, 2 mM lysine are enough to inhibit the growth of Arabidopsis seedlings (Heremans & Jacobs, 1997). Although in this protocol the growth medium is exchanged every two days, the supplemented lysine will still be taken up by the seedlings and could build up intracellularly to concentrations that affect the amino acid homeostasis.

At day 19 after germination, I quantified the labeling efficiencies of the seedlings. It ranged from 32% to 72%, presumably depending on the stratification efficiency and overall growth rate. The higher the growth rate, the stronger the dilution of the supplemented amino acids within the seedlings, decreasing the overall labeling efficiency.

I checked the labeling efficiency of the original MS data submitted to PRIDE by Lewandowska *et al*. (2013, www.ebi.ac.uk/pride/archive/projects/PXD002069) and found that the median global labeling efficiency was only 30.8% (Figure 6.3). In the original publication, they determined the labeling efficiency by analyzing single spectra of representative peptides.

I concluded that this approach was not suitable for our intended study.
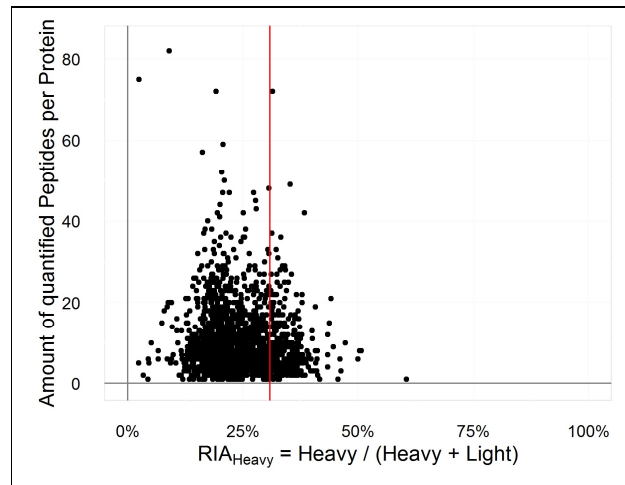
Figure 6.3: Reanalysis of the labeling efficiency from Lewandowska *et al.* (2013).
The median labeling efficiency is indicated by a vertical line.

6.1.4.2   DapL *knockout*

We tested strategies to increase the labeling efficiency by inhibiting the endogenous amino acid synthesis by genetic knock-down. Similarly to the SILAC strategy employed for the algae *Chlamydomonas reinhardtii*, whose strain CC-424 is arginine-autotroph as it has a dysfunctional argininosuccinate lyase (Naumann *et al*, 2005), I aimed to knock-out the LL-diaminopimelate aminotransferase (DapL or Aberrant Growth And Death2, AT4G33680), as it is essential for lysine biosynthesis in plants (Hudson *et al*, 2006; Dobson *et al*, 2011) (Figure 2.4). Homozygous *DapL* knock-outs are embryo-lethal (Song *et al*, 2004; Dobson *et al*, 2011) and no knock-out lines are available. We therefore aimed to knock-down *DapL* in suspension cell cultures, while supplementing the culture simultaneously with lysine. We successfully transformed AtA suspension cell cultures with both DapL-targeted amiRNA constructs (Schwab et al, 2006) and pHellsgate hairpin-RNAi constructs. The clones contained the inserted constructs as determined by PCR on genomic DNA, however, I could not identify any transcriptional downregulation of DapL by qRT-PCR.

Over the course of the transformation, the cells have to be singularized to create a monoclonal line afterwards. This selection takes place over a period of 7 days on solidified medium. During the whole transformation protocol including the singularization, lysine is supplied to support any transformed, lysine-deficient DapL knockout cells. Yet the amount of lysine to add to the medium is difficult to estimate – especially on solidified medium. Ideally the lysine concentration has to be high enough to allow growth of successfully transformed DapL knockouts, but low enough to not affect amino acid homeostasis and create the negative effects discussed in chapter 6.1.1.

6.1.4.3 *Chemical inhibition of* DapL

Besides genetic manipulation, we tested whether *DapL* can be inhibited chemically. The group of Prof. John C. Vederas (University of Alberta, Canada) designed *DapL* inhibitors from in vitro assays (Fan *et al*, 2010; Fan & Vederas, 2012) and generously provided us with the *DapL* inhibitor compound 1a (Figure 6.4A). When I tested it in our Arabidopsis suspension cell culture AtA, I identified a dose-dependent inhibition of culture growth with an IC_{50} at <20 µM (Figure 6.4B), which is comparable to the reported in vitro IC_{50} of 5 µM. Daily addition of 350 µM lysine partially reverts this growth impairment when the inhibitor was used in a range of 15 to 25 µM, while 50 µM of the inhibitor stopped any culture growth, presumably due to side reactions of the free hydrazide moiety of the inhibitor (Figure 6.4A).

Despite these robust effects on the culture's growth, we did not see a significant change in the labeling efficiencies of AtA cultures when treated with 15 µM compound 1a (Figure 6.4C). From this we concluded that compound 1a can induce detrimental side-effects and does not improve labeling efficiency under the tested conditions.
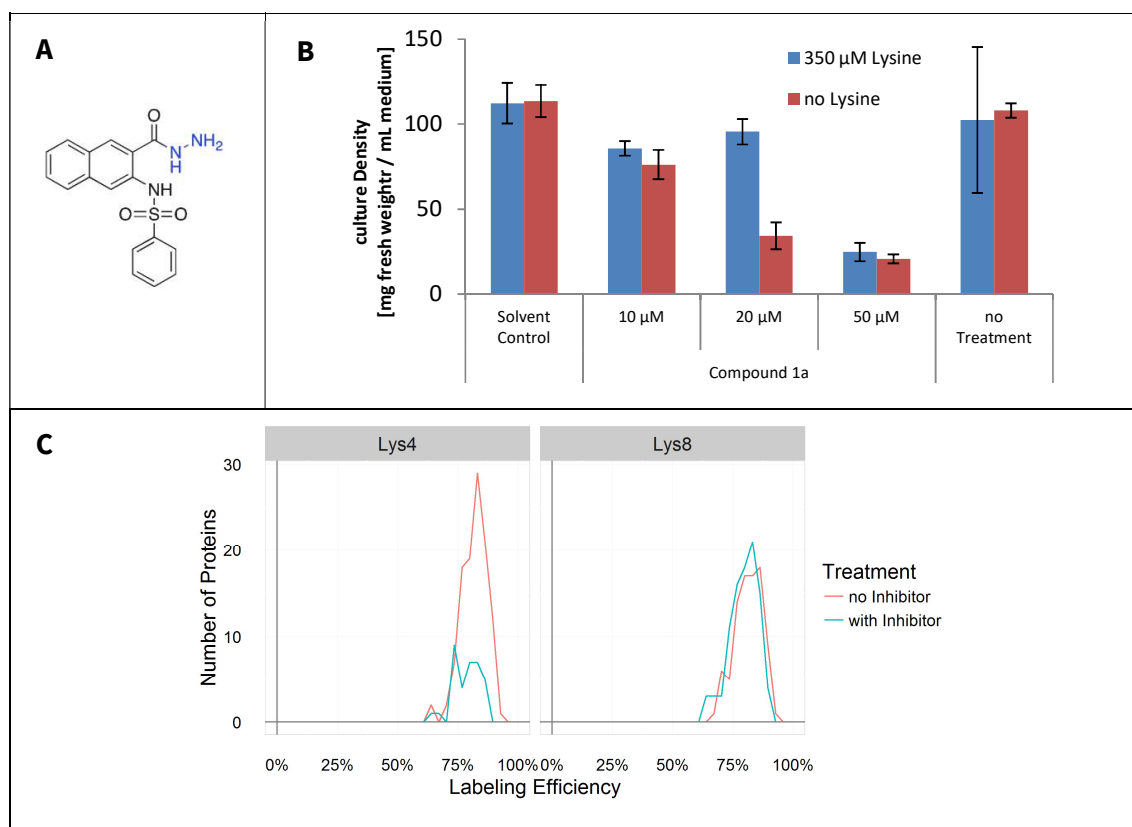
Figure 6.4: Chemical Inhibition of *DapL* by compound 1a.
A: Chemical structure of the *DapL* inhibitor compound 1a (Fan & Vederas, 2012). The free hydrazide moiety is colored in blue. B: Effect of compound 1a on the growth of 7 day old AtA cell cultures. The blue labeled cultures were supplemented each day with 350 µM lysine. When administering compound 1a, it was dissolved in DMSO. The effect of the maximally added volume of DMSO was tested in a separate solvent control. C: Labeling efficiency of proteins of 7-day old AtA cultures supplemented daily with 350 µM Lys4 or Lys8. In blue are the labeling efficiencies of cultures, which were treated with 15 µM compound 1a at inoculation.

6.2 Experimental design

In conclusion, the optimized plant SILAC labeling approach based on (Schütz *et al*, 2011) in Arabidopsis suspension cell culture AtB was deemed most successful and was employed for the protein quantification part of this study.

To reiterate, two AtB suspension cell cultures were grown over 12 days with daily supplementation of 350 µM medium-heavy Lys4 or heavy Lys8, respectively. At day 6 the cultures were passaged to fresh medium so that they maintained continuous growth till day 12. During the two 6 day passages, the cultures grew from 30 to 450 mg/mL, doubling roughly 4 times. At day 12, the Lys4- and the Lys8-labeled cultures were split into three replicate cultures each (Figure 6.5A). After a brief acclimatization of 2 h, three of the six cultures were treated with 100 nM flg22, while the other three cultures were mock treated with water. To control for labeling specific effects, we split the treatment group randomly between Lys4- and Lys8-labelled cultures: two Lys8 and one Lys4 culture were treated with flg22, and one Lys8 and two Lys4 cultures were mock treated (Figure 6.5B).

In order to compare the protein abundance data to the other datasets, we extracted the protein samples at the same time points: before the treatment, and 0.5 h, 2 h, 4 h, and 8 h post treatment. In the following part, I will discuss the results from the protein abundance data set, which aims to identify the effects of the bacterial defense elicitor flg22 on the protein levels of the Arabidopsis suspensions cell culture AtB.
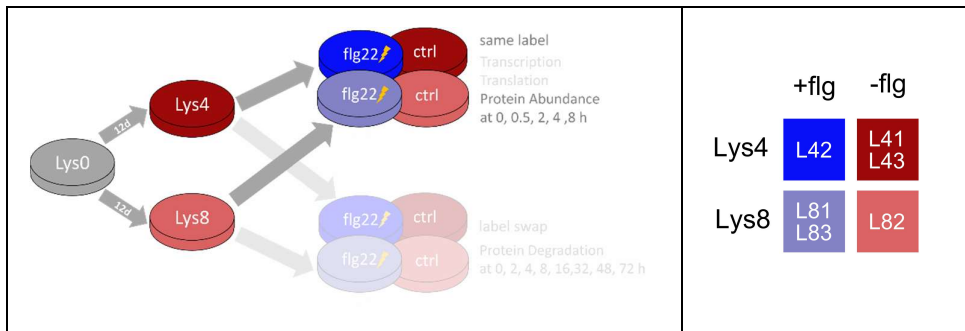
Figure 6.5: Overview of the experimental design for the protein abundance. A: Schema of the overall experimental design. This part of the thesis will discuss the results from the protein abundance dataset, which is not grayed out. B: Schema for the sample naming convention. The first two letters indicate the labeling background of the culture. Samples from Lys4-labelled cultures will be colored slightly darker than Lys8-labelled cultures. Flg22-treated cultures will be colored in blue, while untreated, control samples will be colored in red. The third letter of the sample code indicates the replicate number. We strive to follow the naming conventions and color scheme throughout this thesis.

6.3 Labelling efficiency and viability

After 12 days of labelling, the culture's proteins had a median labelling efficiency of RIA = 85.0% (LE = 85.9%) for the Lys4 culture and RIA = 85.3% (LE = 86.5%) for the Lys8 culture (Figure 6.6A). The labeling had an average standard deviation of 9.1% and 9.8%, respectively, and a median absolute

deviation of 2.87% and 2.83%, respectively. At the end of the time course 8 hours later, the labeling efficiency was practically unchanged at RIA = 85.8% for the Lys4 cultures and RIA = 85.30% for the Lys8 cultures. The Lys0 incorporation remained stable over the course of the experiment (Figure 6.6B). Consequently, all protein abundances were quantified between the heavy Lys8 and the medium-heavy Lys4 label state, excluding Lys0-labelled proteins as it was done by Schütz *et al.* (2011).

After the time course, the viabilities of the cultures were tested by TTC assay (Figure 6.6C). Treatment or labeling did not affect the viability of the cultures.
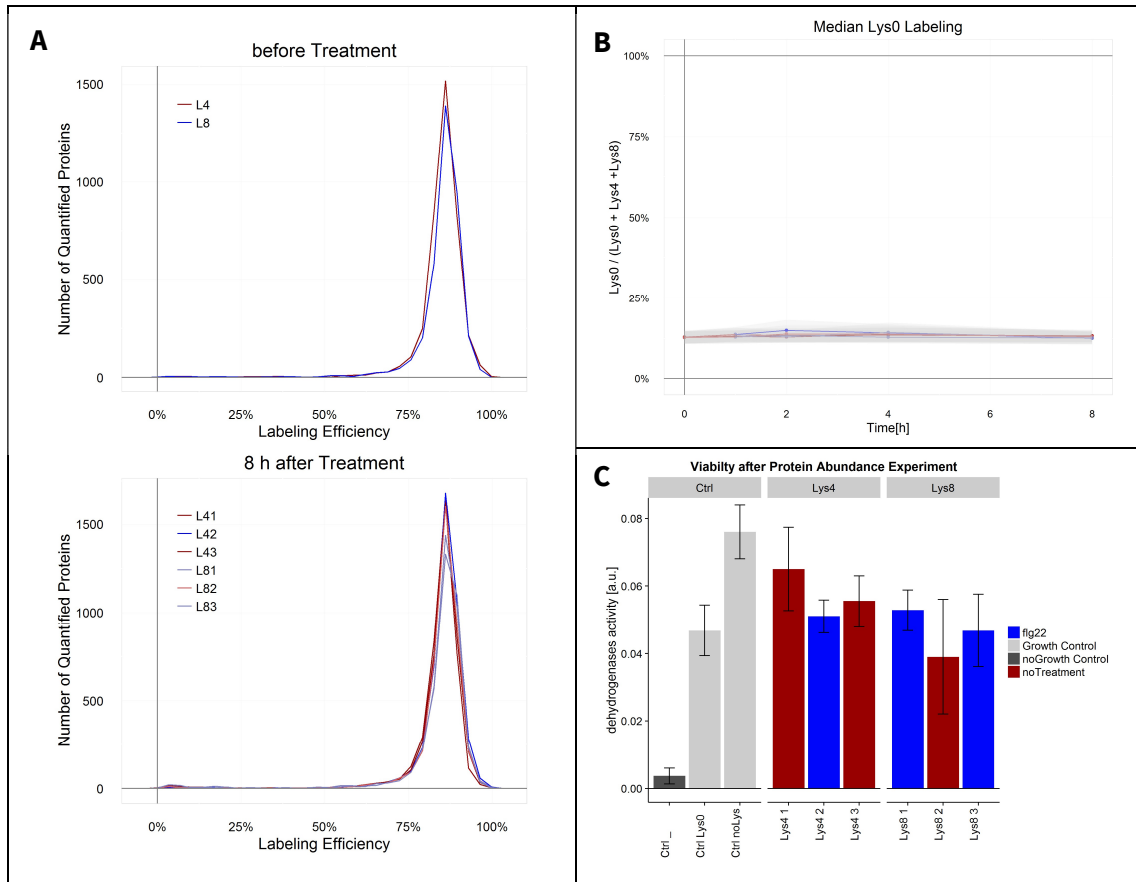
 Figure 6.6: Labelling efficiencies and viabilities during the protein abundance experiment.
A: Labeling efficiencies of the cultures before treatment and at the last time point. There was no change in labeling over the course of the experiment. B: The ratio of the unlabeled Lys0 proteins remained stable during the experiment. C: The viability of the cultures after the experiment was determined by TTC assay. There are no significant changes between the cultures used in the experiment and the control culture labeled with Lys0. The culture without Lysine supplementation showed a stronger dehydrogenase activity. The negative control ("noGrowth") was obtained from a non-proliferating 4-week old culture.

6.4   Coverage

In total, 7397 proteins were identified in all samples combined, of which 4110 proteins were quantified. On average, 2611 proteins were quantified in a single sample.

The overlap of quantified proteins over all samples is rather low, as it is typical for a large-scale MS experiments (Bruderer *et al*, 2015). Only 1134 or 27.6% of all quantified proteins were quantified in all 30 samples. While 2585 or 62.8% of all quantified proteins were quantified in more than 50% of all samples.
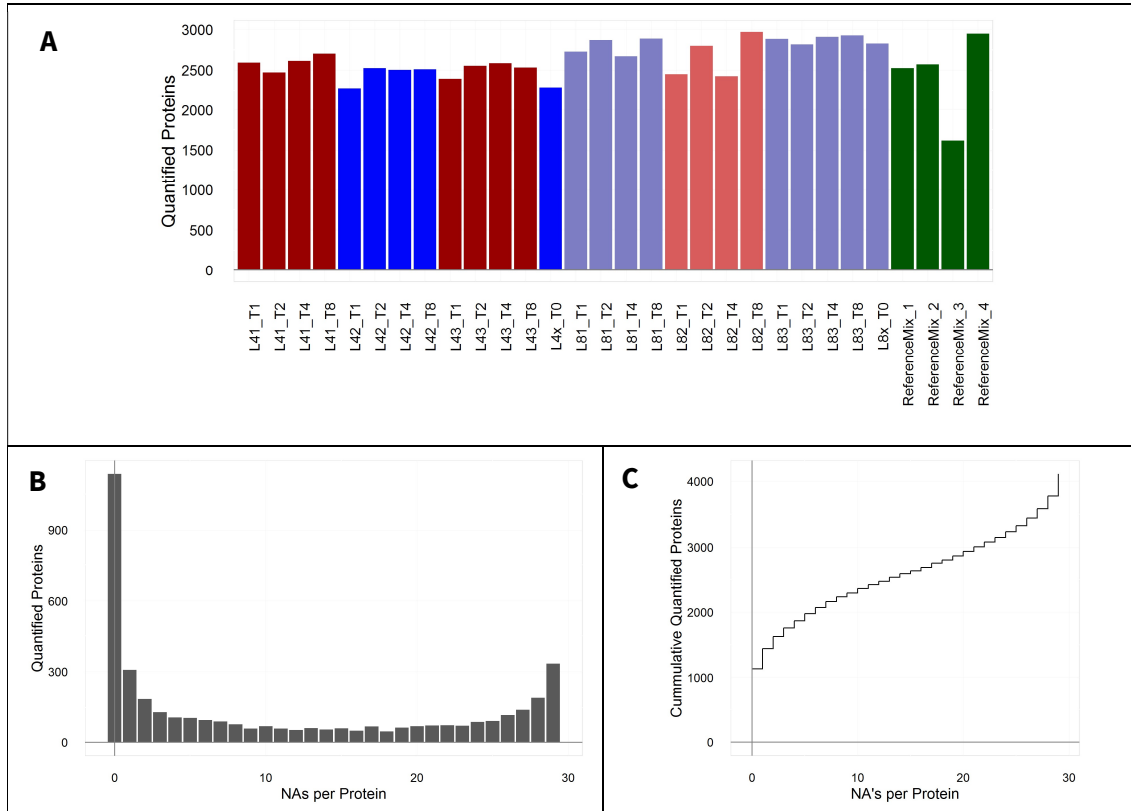
Figure 6.7: Coverage QC plots.
A: Amount or proteins quantified in each sample. B: Histogram of the number of quantified proteins per number of not-assigned (NA) values over all 30 samples. C: Cumulative plot of the number of proteins with a set number of NAs per protein.

6.5 Quality control

We checked the quality of the MS analysis with the tool PTXQC (Bielow et al, 2016). It analyzes a multitude of different quality control (QC) parameters from the MaxQuant output files and combines them to 20 QC metrics. Figure 6.8A shows the overview of these metrics on all 150 LC-MS runs (30 samples x 5 fractions).
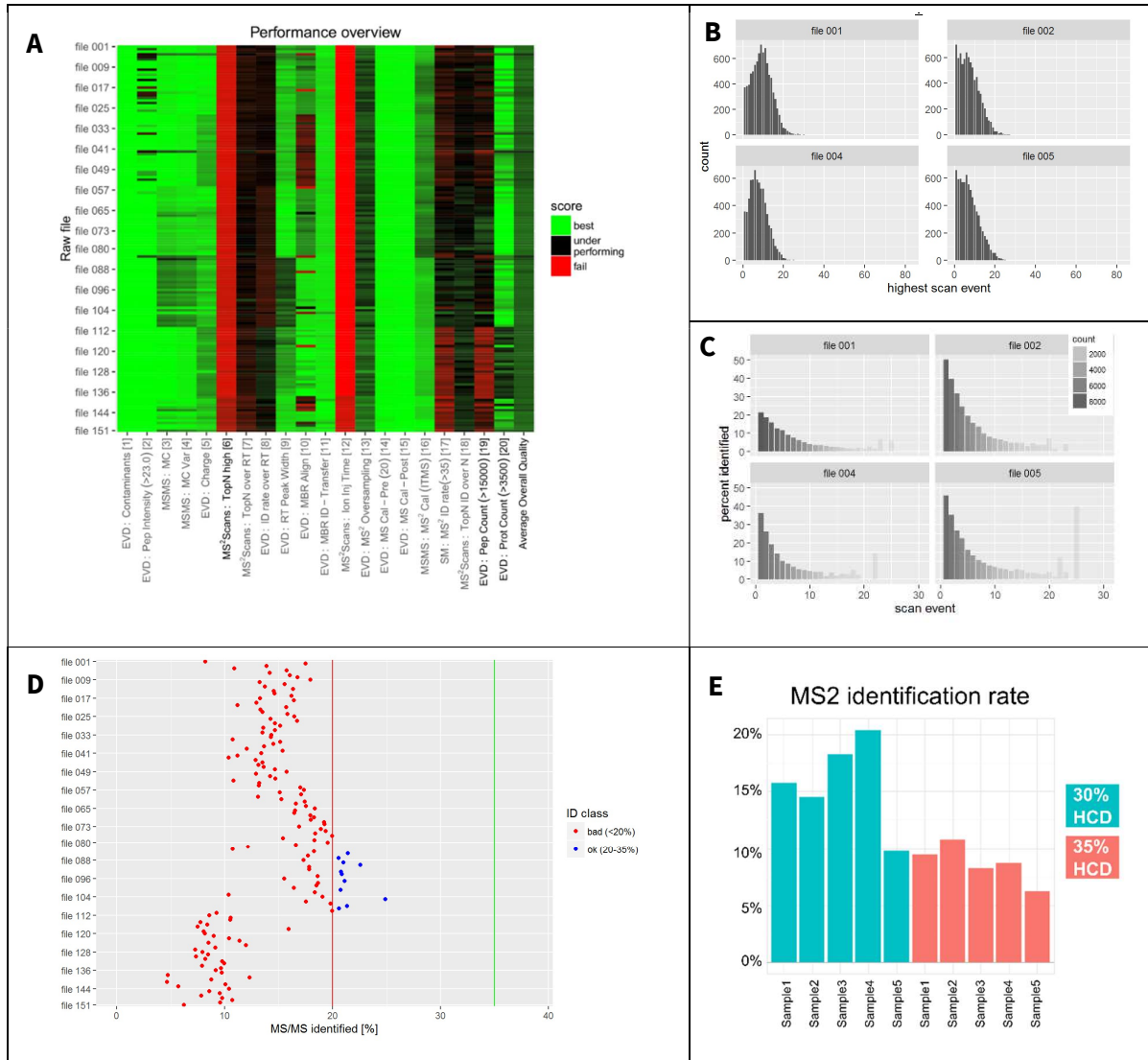
Figure 6.8: Quality control of the protein abundance samples by PTXQC.
A: Overview of all samples across all 20 quality metrics of PTXQC. B: Metric [6] TopN high, showing the number of MS2 scans per duty cycle. C: Metric [18] TopN ID over N, showing the MS2 identification rate relative to the number of MS2 scan events per duty cycle. D: metric [17] MS2 ID rate of all 150 samples. E: Comparison of HCD collision energy on MS2 spectra identification rate. The samples colored in red were fragmented with the higher 35% HCD energy, while the blue samples were fragmented at 30% HCD energy.
EVD = MaxQuant Evidence file, ID = identification, MBR = Match between runs, Cal = mass calibration, MC = missed cleavage.

The overview heat map shows that most samples perform uniformly and generally well. The QC metric [6], TopN high, describes the number of MS2 scans that are triggered by the machine, before it initiates a new MS1 scan. All samples underperform here, showing that on average only ~10 MS2 scans are initiated before a new MS1 scan starts (Figure 6.8B), while the Orbitrap Fusion mass spectrometer was set up to initiate up to 20 MS2 before finishing the duty cycle. Consequently, the underperforming TopN also shows a non-optimal value for the metric [7], TopN over retention time. This also comes into play for metric [18], which rewards a uniform identification rate independent of the TopN cycle. This means, that a peptide from a Top20 cycle should be as well identified as one of a Top5 cycle. However, there is a clear decrease in the identification rate with every additionally fragmented precursor per duty cycle (Figure 6.8C). There, optimization of the LC conditions could help to assure reliable loading of the MS with enough precursor ions. Metric [12], ion injection time, though flagged as failed, is only affected at the end of the LC gradient. The average ion injection time of ~40 ms is acceptable.

6.5.1 MS2 identification rate

The most revealing metric is number 17, the MS2 ID rate (Figure 6.8D). In only a few samples more than 20% of the MS2 scans could get identified as a peptide. An excellent MS run would identify more than 35% of all MS2 spectra. As a reference, samples deposited in PRIDE have on average an MS2 ID rate of 25% (Griss *et al*, 2016).

This low identification rate is also explained by the TopN plots. The difference between Figure 6.8B and Figure 6.8C shows that a majority of MS2 scans are obtained from higher TopN cycles, yet the identification rate drops substantially for every additional MS2 scan per duty cycle. Clearly, the LC-MS method has to be optimized to accommodate for these differences. Either, one could set the maximal TopN to a lower value of ~7, so that the MS is forced to start more frequently an MS1 scan. This would assure that the time difference between MS1 and MS2 scans is smaller and the machine can isolate MS2 precursors more effectively. The more time passes between an MS1 and the consecutive MS2 scans, the more likely it is that a non-optimal time of the peptide elution peak is chosen for fragmentation, and for instance, the peptide did already elute from the LC column.

Another idea to consider is to increase the ion inflow by optimizing the spray conditions or by increasing the concentration of peptides loaded onto the column. However, the overall peptide intensity is deemed sufficiently high for the majority of samples (metric [2]).

We tested the effect of an increased HCD fragmentation energy on the MS2 identification rate. During the processing of the MS2 scans, we identified a large proportion of spectra with non-optimal fragmentation, where the non-fragmented precursor ion remained the most prominent peak of the fragment spectrum. We consequently tested if increasing the HCD collision energy from 30% to 35% would increase the spectra identification rate. However, as shown in Figure 6.8E, an increased collision energy lowered the average MS2 spectra identification rate.

We have tested alternatively if the MS2 identification rate could be improved by adding more peptide modifications to the database, against which the spectra are compared to. We analyzed some exemplary MS runs with the search engine ProteinPilot and the "error tolerant search" from the Mascot search engine. Both analyses allow for all peptide modifications reported in UNIMOD

(www.unimod.org) and report if any modification was frequently identified in the queried LC-MS run. Both searches returned the Lys4 and Lys8 labeling and the already added peptide modification (fixed carbamidomethylated cysteines, and variable acetylation of the protein N-terminus and oxidation of methionines). The next most frequently found modification was a deamination of asparagine or glutamine. Adding this modification to the MaxQuant search, however, lowered the overall identification rate by 3.2% on average. Adding the modification to the search space lets us identify the truly deaminated peptides, but requires a stricter FDR threshold that excludes more unmodified true peptides than adds true-positive modified peptides. Consequently, we did not include these additional modifications in the MaxQuant analysis.

Figure 6.9: Testing for additional peptide-modifications with ProteinPilot (A) and the Mascot error-tolerant search (B).
Both algorithms match spectra to peptides with deaminated N or Q in addition to the already included peptide modifications (carbamidomethylated cysteines, acetylation of the protein N-terminus and oxidation of methionines).

6.5.2 Arginine-terminating peptides

Another important factor to consider are the identified but not quantified peptides. 97.5% of these peptides contain an arginine, but no lysine (Figure 6.10). Only 1.3% of the non-quantified peptides contain a lysine. In total, 45% of all detected peptides could not be quantified, as they did not contain a differentially labeled lysine. To remedy this, one would either have to include arginine in the labeling strategy, which would require another round of cell culture optimization, or to exchange the Arg- and Lys-specific trypsin with a protease that only cleaves after lysine residues. The endoproteinase Lys-C from *Lysobacter enzymogenes* cleaves only at the carboxyl side of lysine residues, even when that lysine is followed by a proline (Jekel *et al*, 1983). Its digestion efficiency can be lower than that of trypsin (Gershon, 2014), but it can be improved under alternating high-pressure conditions (Cheng *et al*, 2016).

Lys-C cleaves less frequently than trypsin, which results in longer but fewer peptides. All arginine-terminating peptides, which make up 45% of all true tryptic peptides in Arabidopsis, would not be generated. This lowered complexity of the produced peptide mixtures could improve the MS analysis and increase the overall identification rate – given an adaptation of the MS2 fragmentation settings to longer peptides with higher charge states. Especially electron-transfer dissociation (ETD) is discussed to be beneficial for sequencing such peptides (Giansanti *et al*, 2016). Schütz *et al.* (2011)

used trypsin for analyzing their lysine-labelled Arabidopsis cell culture, while Christiano *et al.* (2014) used Lys-C to analyze their lysine-labelled yeast culture.
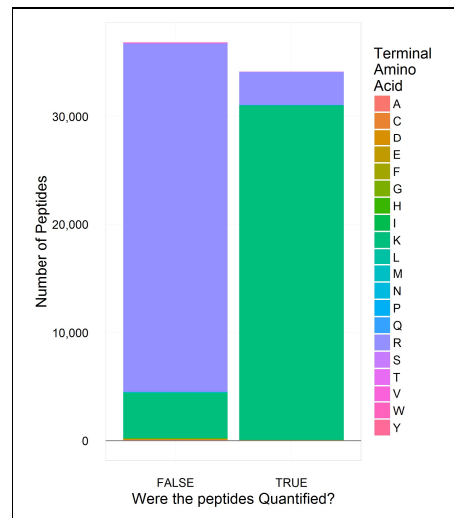
Figure 6.10: 94.3% of all non-quantified peptides terminate on arginine (R).

Another strategy to improve the coverage of an MS study is to reduce the protein database size to the actively transcribed or even only the translated part of the genome. This transcriptome or translatome can be collated from all open reading frames (ORFs) identified in the RNA-Seq or Ribo-Seq experiment. With this, the database to compare the spectra against becomes smaller, and the chance to match with a good score to a decoy peptide gets reduced, thus lowering the chance of a false identification at a given FDR threshold. This follows the same logic as the lowered identification rate when the database was enlarged by another peptide modification. In metaproteomics studies, such approaches are fundamental for reliable peptide identification (Muth *et al*, 2016).

In summary, the overall quality of the samples is good (as shown in the summarizing column of Figure 6.8A). However, the quality control assessment has raised issues in sample processing and MS settings that can be optimized in future studies.

6.5.3 Data normalization

The plant SILAC approach allows to run two samples mixed together (medium-heavy Lys4 and heavy Lys8 channel) and to draw direct comparisons between these two samples. But we wanted to compare all samples with each other so that we can make statements about changes over time as well as changes between flg22-treated and control cultures. Therefore, we decided for a common reference design, which is commonplace in microarray experiments (Smith, 2005) or the spike-in SILAC design (Geiger *et al*, 2011), where each unlabeled sample is mixed with the same labeled reference tissue culture (Figure 6.11A).

In our case, the samples were labeled both as Lys4 and Lys8, so that we needed to create two reference samples: one to contrast each label state. We took equal amounts of cells from all Lys4- or Lys8-labelled cultures at the onset of the experiment (T0) and the conclusion of it (T8), and combined them to a Lys4- and Lys8-reference sample, respectively. These samples contain both flg22-treated

and non-treated material. The proteins extracted from these reference samples were mixed 1:1 with the protein samples of the complementary label state and were analyzed by MS (Figure 6.11B). This two-reference spike-in SILAC design allows to compare all samples along a time-series when only relative quantifications can be performed.
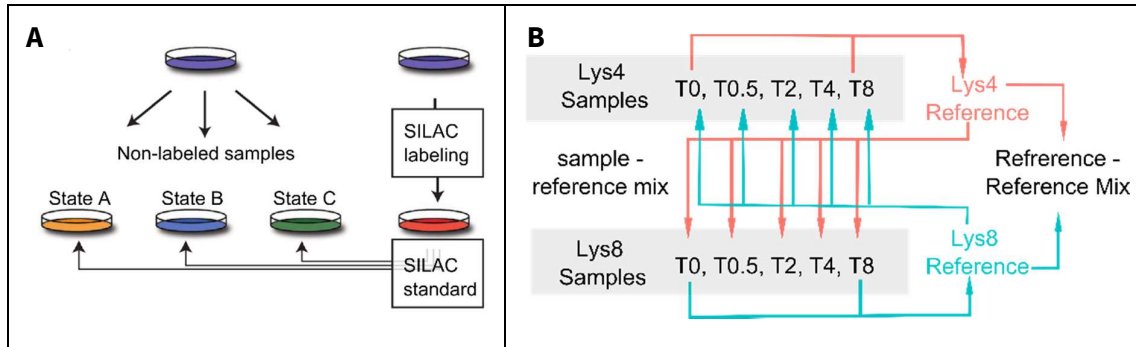
Figure 6.11: A: general spike-in SILAC approach (Geiger et al, 2011), B: our implemented two-reference spike-in SILAC design.

The design includes a label-swap that helps to exclude labeling-induced artifacts. The downside of this approach is that the two references are not perfectly identical as they are derived from cultures, which were cultivated separately for more than 12 days. Also, the reference mixes contain an unequal amount of treated or untreated cultures due to the label-swap design. Reference_Lys4 contains one treated and two untreated cultures (L41_control, L42 _flg22, L43_control), while ReferenceLys4 contains two treated and one untreated culture (L81 _flg22, L82 _control, L83 _flg22). We measured four technical replicates of 1:1 mixtures of the Lys4- and Lys8-reference, which allows us to normalize for changes between the references.

In the following paragraphs, we describe the data processing and normalization of the MS data, which we applied to identify protein level changes induced by flg22 treatment. In Figure 6.12, the process is illustrated for the data set as a whole and by the specific example of cinnamate-4-hydrolase (C4H, AT2G30490). The values in Figure 6.12 are log2-transformed for visualization purposes only, all calculations are performed on the untransformed ratios.

As the Lys0-labelling stayed constant during the experiment, we excluded it from the search space of the conducted MaxQuant analysis. In this case, $\frac{Heavy}{Light}$ ratios describe $\frac{Lys8}{Lys4}$ ratios.

1. Centralization (Figure 6.12B&H)

I extracted the non-normalized protein-level $\frac{Heavy}{Light} = \frac{Lys8}{Lys4}$ ratios from the MaxQuant output file *proteinGroups.txt*. We performed a median centralization ($values_{centralized} = \frac{values}{median(values)}$) so that experimentally introduced deviations from the 1:1 mixing are balanced out (Cox & Mann, 2008). We did not use the directly reported MaxQuant-normalized ratios as both the median and the mean of these ratios differs slightly from 1.0. It is unclear how exactly that commonly used normalization is being computed.

The distributions of the $\frac{Lys8}{Lys4}$ ratios before the centralization are depicted in (Figure 6.12A&G).

2.   Inversion (Figure 6.12C&I)

The experimental design contains a label-swap. $\frac{Lys8}{Lys4}$ ratios of Lys8 samples contain the reference value in the denominator, while for Lys4 samples it is in the numerator. We inverted all $\frac{Lys8}{Lys4}$ ratios of Lys4-labelled samples, so that the reference value was always in the denominator: $\frac{Ly_{Sample}}{Lys8_{Reference}}$ or $\frac{Lys_{Sample}}{Lys4_{Reference}}$.

3.   Reference Normalization (Figure 6.12D&J)

The ratios refer to two different references in the denominator. We multiplied the ratios of Lys4-samples with the MaxQuant-centralized $\frac{Lys8_{Reference}}{Lys4_{Reference}}$ ratio, so that all ratios refer to the same $Lys4_{Reference}$ value:

$$\frac{Lys4_{Sample}}{\cancel{Lys8_{Reference}}} * \frac{\cancel{Lys8_{Reference}}}{Lys4_{Reference}} = \frac{Lys4_{Sample}}{Lys4_{Reference}}$$

This is a protein-specific normalization. If the protein levels of $Lys4_{Reference}$ and $Lys8_{Reference}$ did not differ, the $\frac{Sample}{Reference}$ ratios remained unchanged.

4.   Labeling Minimization (Figure 6.12E&K)

Despite the reference normalization, there remained an offset between Lys4-lableled and Lys8-labelled samples. The size of this offset varies for each protein but is constant over all time points. We could not associate the offset to any other cause than biological differences between the reference samples. We measured the $\frac{Lys8_{Reference}}{Lys4_{Reference}}$ mix in four technical replicates with independent protein mixing, trypsination, fractionation and LC-MS analysis, and there seems to be a systematic difference between the Lys4-labelled and Lys8-labelled references that is not technical in nature. We decided to add a normalization that minimizes these differences between Lys4- and Lys8-samples. For that, we divided all ratios from Lys4 samples with a Lys4-labeling constant, and all Lys8 samples with Lys8-labeling constant.

$$LysX \in Lys4, Lys8$$

$$\frac{LysX_{ProteinY,normalized}}{Lys4_{Reference,ProteinY}} = \frac{LysX_{ProteinY}}{Lys4_{Reference,ProteinY}} \div constant_{LysX,ProteinY}$$

The labeling constants are calculated for each time point by taking the median $\frac{Sample}{Reference}$ ratio over all treated and untreated samples, taking into effect the unequal amount of treated and untreated samples. We then averaged that protein-specific constant over all time points.

$$constant_{LysX,ProteinY} = median\left(median\left(\frac{LysX_{flg22,ProteinY}}{Lys4_{Reference,ProteinY}} \right), median\left(\frac{LysX_{control,ProteinY}}{Lys4_{Reference,ProteinY}} \right) \right)$$

We then averaged that protein-specific constant over all time points.

$$constant_{LysX,ProteinY} = \frac{\sum_{T=1}^{4} constant_{LysX,ProteinY,TimeT}}{4}$$

The logic of this normalization becomes clear with a specific example. In Figure 6.12K, an offset between the darker colored Lys4-labelled and the lighter colored Lys8-labelled samples is evident. With the labeling minimization normalization (Figure 6.12K), the differences between these lines are reduced and the samples align closer according to the flg22-treatment groups.

Interestingly, the constant shows an unprecedented systematic bias regarding the employed label. Over all proteins, the median Lys4-labeling constant is 0.950, while it is 1.001 for Lys8. We speculate

that this difference is caused by the MS quantitation, maybe due to an overlap of the Lys0 isotope envelopes with the Lys4-labelled peaks.
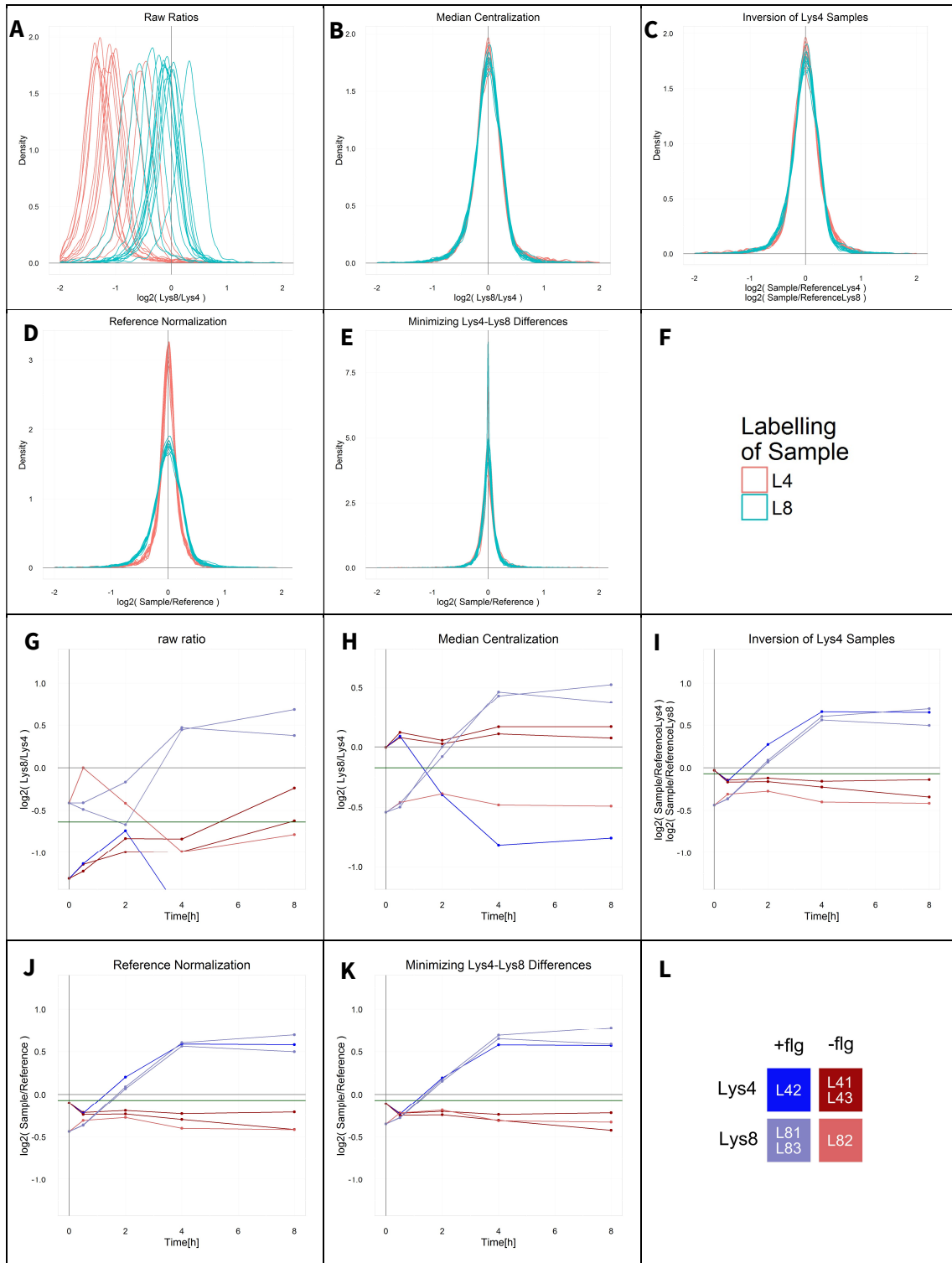
Figure 6.12: Protein-level ratios during the normalization procedure.
The ratios are log2-transformed for visualization purposes but remained untransformed for the calculations. On the top (A-F) are the density plots of all samples

throughout the normalization depicted to visualize the spread of the data. The legend for these plots are depicted in panel F. In the lower half (G-L), the normalization procedure is described with the example of cinnamate-4-hydrolase (C4H, AT2G30490). The legend for these plots is in panel L. The C4H-specific $\frac{Lys8_{Reference}}{Lys4_{Reference}}$ ratio is depicted as a green horizontal line throughout the normalization. Note the remaining labeling-specific offset after the reference normalization (J), which is balanced out by the labeling minimization (K).

The variance of the ratios varies heavily over the course of the data normalization (Figure 6.13). The standard deviation is however strongly affected by outliers (Figure 6.13A), as the distance to the mean is inserted into the standard deviation by the power of 2. Therefore, I propose to monitor the normalization procedure with the median absolute deviation (MAD, Figure 6.13B). It shows that the wide-spreading raw ratios are effectively reduced to a constant variation by the MaxQuant centralization. Every step after the inversion reduced the variation and especially the labeling minimization reduced it strongly.
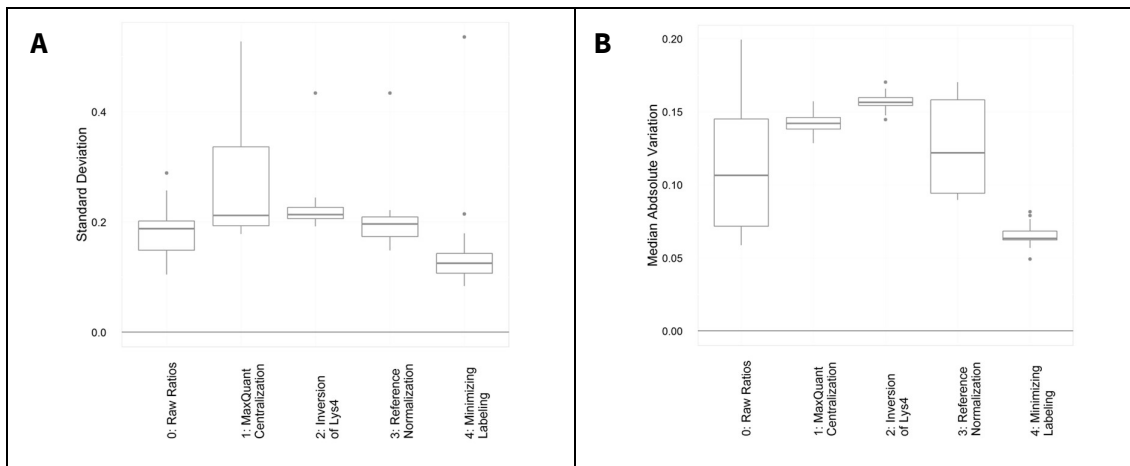
Figure 6.13: Variance during the normalization procedure, measured as the standard deviation (SD, A) or the median absolute deviation (MAD, B).

After concluding the data processing, I see the highest correlation between sample replicates (Figure 6.14). Samples with the same treatment show a high correlation. Interestingly, treated and untreated samples of the early time points (T1 = 0.5 h after treatment and T2 = 2 h after treatment) correlate highly as well. These trends are upheld regardless of the labeling background of the samples.
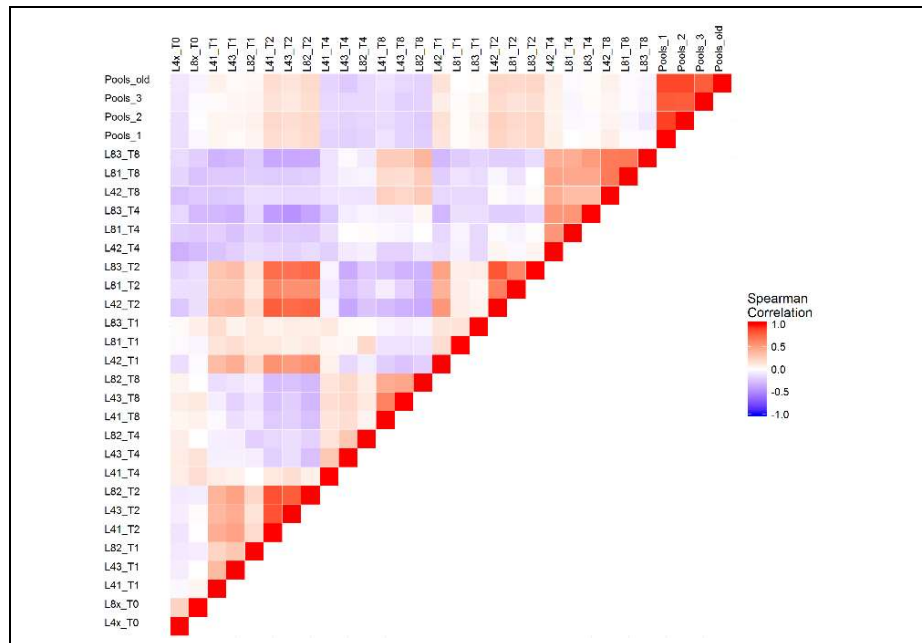
Figure 6.14: **Correlation analysis after normalization.**
The highest correlation of the data set can be observed for the technical replicates of the $\frac{\text{Lys8}_{\text{Reference}}}{\text{Lys4}_{\text{Reference}}}$ mix (named Pools in this plot).

6.6 Dimension reduction plots

To get an unbiased understanding of the obtained data matrix, it is useful to describe it in a 2-dimensional plot by some dimension reduction technique. We employed both principle component analyses (PCA, Figure 6.15A) and multi-dimension scaling plots (MDS, Figure 6.15B) for this task. In both plots, the sample replicates cluster together according to both time point and treatment – regardless of the Lys8 or Lys4 labeling state of the sample.

The PCA scales according to the variation within the dataset. Principle coordinate 1 (Dim1, 12.89% of the variation) describes the progression of time, while principle coordinate 2 (Dim2, 7.93% of the variation) describes the variation induced by the treatment. The progression of time thus induced more variation in the data set than the treatment itself. Yet it has to be taken into consideration that the two first principle components only sum up to 20.82% of the variation of the data set.

Furthermore, both plots show that the strongest variation between the flg22 and control samples occurs at the end of the time course. The more time passes, the more disparate the flg22 and control samples get – as it is to be expected for protein data. The variation in the data set describes a biologically meaningful trajectory.

This clustering according to treatment and time only becomes apparent when all normalization steps described above are performed. Any step earlier in the data normalization procedure creates dimension reduction plots, where the samples are not separated according to biologically meaningful parameters and separate primarily by the sample's Lys4 and Lys8 labeling state.
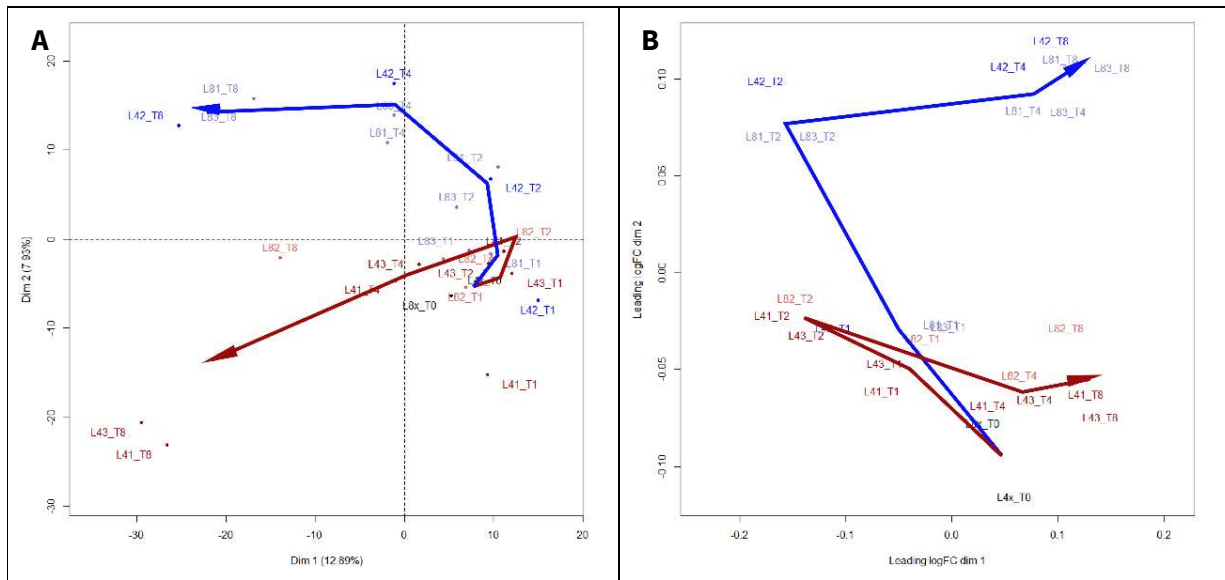
Figure 6.15: Dimension Reduction Plots of the log2-transformed normalized protein abundance data set.
A: Principle Component Analysis (PCA). B: Multi-Dimension Scaling Plot (MDS). The averages of all flg22-treated (blue) and untreated samples (red) over the time course are indicated with arrows.

6.7 Statistical analysis for flg22 effect

After the data processing and normalization, I tested which proteins change their abundance following the flg22 treatment. I employed the linear modeling framework of the R package limma (Smith, 2005). Limma was designed for microarray experiments but has also been applied on relative quantification proteomics data sets (Schwämmle et al, 2013). The benefit of limma in comparison to a classical gene-specific ANOVA is that it uses global information of all proteins to estimate the variance. Limma can be applied to log-transformed ratios or expression values.

We designed the statistical test so that we identified differential expression at any time points (T0.5, T2, T4, T8) between all replicates treated with flg22 or the control solution - regardless of the Lys4 or Lys8 labeling background (Figure 6.16A). The labeling background was included as a blocking factor.

$$\log_2(ratio) \sim lysine + time + treatment$$

With this statistical test, we identified 142 upregulated proteins and 84 downregulated proteins (p-value < 0.05, no fold change threshold), while for 2'235 proteins no significant change could be determined. In total, 2'451 proteins could be assessed by the *limma* test.
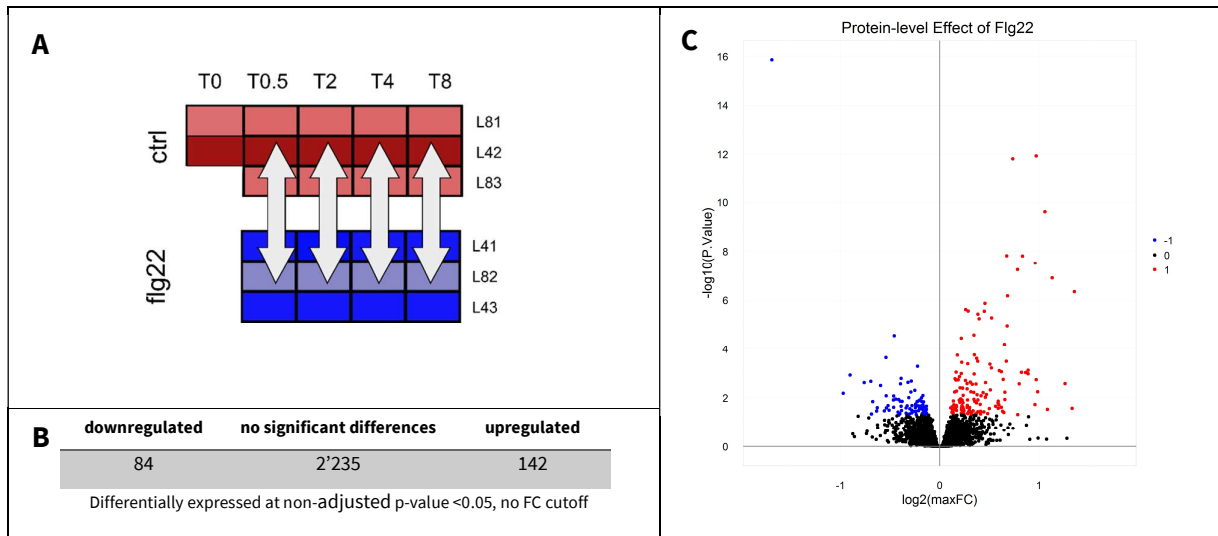
Figure 6.16: Results of the limma test for flg22 on protein level changes
A) Schema of the limma test design. Each block represents one sample. The differently treated samples at the same time point were compared, with the lysine labeling as a blocking factor. B) Amount of proteins regulated according to the given cutoff thresholds. C) Volcano plot comparing the statistical strength (p-value) and the effect size (maximal fold change at any time point, maxFC).

The upregulated proteins are enriched for various biological processes associated with bacterial defense reactions (Table 6.2 and Figure 1.20, PantherGO (Mi *et al*, 2013), Bonferroni adjusted p-Value < 0.05). The downregulated proteins are not enriched for any GO biological process term at Bonferroni multiple testing adjusted p-value threshold of 0.05. Using a non-adjusted p-value leads to an incoherent set of GO terms (data not shown). Similarly, the downregulated proteins change more irregularly at individual time points, while the upregulated proteins show a more consistent pattern over all four time points (Figure 6.18).

Table 6.2: Enrichment of biological process GO terms of all upregulated proteins.
Only the lowest daughter nodes are listed. Parent nodes are excluded for the sake of brevity, refer to Figure 6.17 for a full graphical presentation. The p-value is adjusted for multiple testing by the Bonferroni method. The list of all 2451 limma-tested proteins was used as reference. # = number of proteins.

GO biological process complete	# in reference	# in upregulated	# expected in upregulated	Fold Enrichment	Adjusted p-Value
tryptophan biosynthetic process	14	5	0.28	18.17	1.39E-02
toxin catabolic process	19	6	0.37	16.06	3.52E-03
glutathione metabolic process	30	7	0.59	11.87	3.67E-03
response to karrikin	43	7	0.85	8.28	3.69E-02
cellular amino acid biosynthetic process	112	12	2.2	5.45	3.69E-03
response to bacterium	159	13	3.13	4.16	2.49E-02

Figure 6.17: Hierarchical presentation of all enriched GO categories in the set of upregulated proteins.
Visualized with BioMaps from the VirtuaPlant package (Katari et al, 2010).The nodes are colored according to the enrichment p-value (orange = highly significant p-value, green = less significant p-value). For the sake of clarity, only nodes with a p-value <0.001 are presented. The list of all 2451 limma-tested proteins was used as the background list to compare against.

Figure 6.18: Heat map of the flg22-induced fold changes (FC) for the 126 significantly upregulated proteins (A) and the 91 downregulated proteins (B).

We compared the flg22-induced fold changes of the *limma* test with the *edgeR* test results from the RNA-Seq (Figure 6.19A) and Ribo-Seq data sets (Figure 6.19B). We identified a small but highly significant correlation to both data sets (rho = 0.2444 and 0.2033, respectively, both p-values < 2.2e-16). The fold-changes from the RNA-Seq dataset correlate better to the protein data, presumably due to the smaller variance within that dataset. The correlational analyses show the same trend of a higher correlation for genes with increased protein abundances. When limiting the analysis to the proteins that change significantly according to the limma test, the Spearman correlation increased to 0.634 for the upregulated and 0.515 for the downregulated proteins.
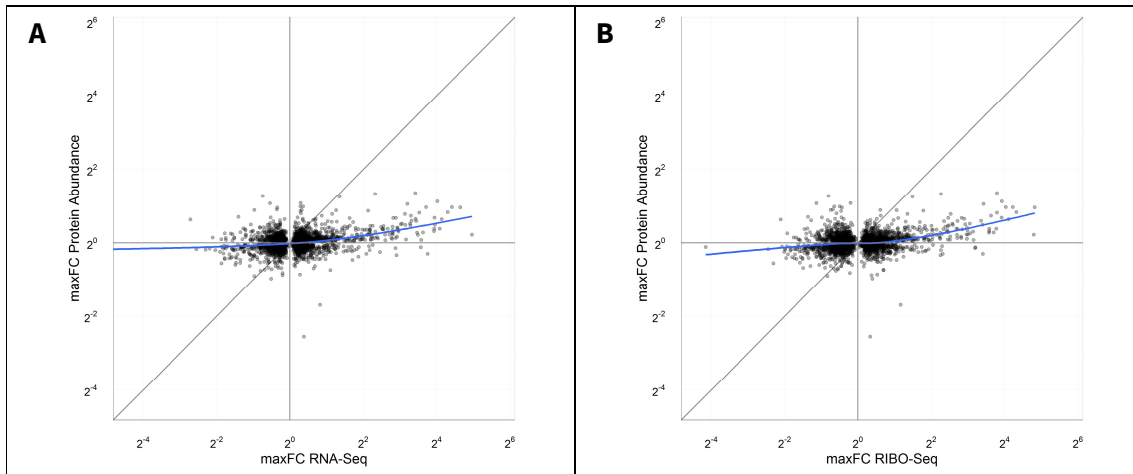
Figure 6.19: Correlation analysis of the protein and the transcriptional (A) and translational (B) data set.
A locally weighted scatterplot smoothing (LOESS) trend line is indicated. maxFC = highest fold-change between flg22-treated and untreated samples at any of the four tested time points.

To assure the correct application of the *limma* framework, we employed an ANOVA test with the same formula onto the same data set ($\log_2(ratio) \sim lysine + time + treatment$). The ANOVA p-value for the flg22-treatment is generally lower than the p-value from *limma*, but it correlates highly (p-value < 2.2e-16, Pearson's product-moment correlation, Figure 6.20A). It identified 120 downregulated and 159 upregulated proteins at an unadjusted p-value cutoff of 5% (Figure 6.20B). The upregulated proteins are enriched for similar GO categories (Table 6.) as for the *limma* test and the downregulated genes are also not significantly enriched for any GO category. The overlap of the regulated proteins between the ANOVA and limma test is higher for the upregulated proteins (Figure 6.20C) than for the downregulated proteins (Figure 6.20D). A complete list of the proteins with changed abundances according to both tests can be found in the appendix chapter 10.1.
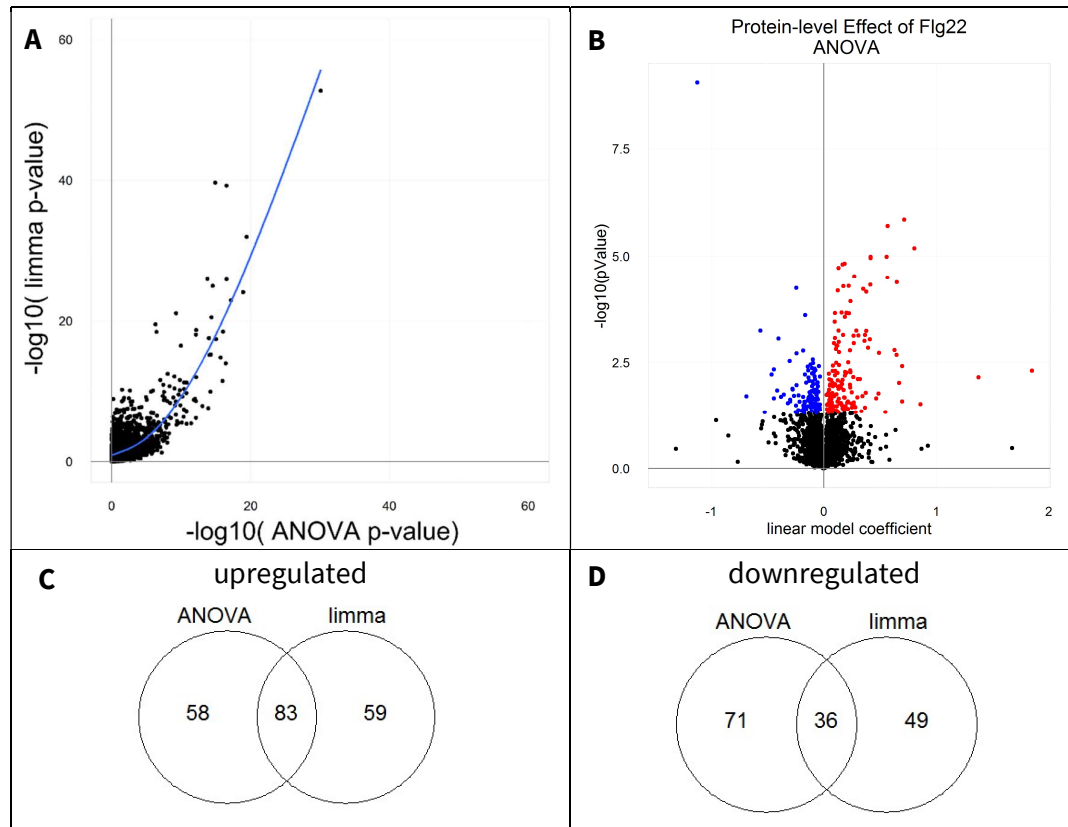
Figure 6.20: Comparison of ANOVA and limma test.
A: Both tests compute highly correlating p-values for the effect of flg22. A LOESS regression line is indicated in blue. B: volcano plot of the regulated genes identified by ANOVA. C-D: Venn diagram of the up- (C) and downregulated genes as assessed by the two tests.

Table 6.3: GO category enrichment for upregulated proteins with a p-value < 0.05 according to the ANOVA test. The enrichment was performed with Panther (Mi *et al*, 2016) and the list of all ANOVA assessed genes was used as the reference list.

GO BIOLOGICAL PROCESS COMPLETE	PROTEIN IN REFERENCE	PROTEIN IN UPREGULATED	# EXPECTED IN UPREGULATED	FOLD ENRICHMENT	ADJ. P-VALUE
tryptophan biosynthetic process	14	6	0.3	19.73	1.10E-03
auxin biosynthetic process	14	5	0.3	16.44	2.24E-02
secondary metabolic process	106	13	2.3	5.65	9.77E-04
sulfur compound metabolic process	145	16	3.15	5.08	1.98E-04
response to cadmium ion	191	21	4.15	5.06	2.04E-06
response to bacterium	159	14	3.45	4.05	1.61E-02
carboxylic acid biosynthetic process	221	17	4.8	3.54	1.05E-02
cellular amino acid metabolic process	211	16	4.58	3.49	2.36E-02
response to stress	963	41	20.92	1.96	2.03E-02

7. PROTEIN DEGRADATION

7.1 Experimental approach

High-throughput turnover studies require MS-based proteomics with a metabolic pulse(-chase) labeling design. As mentioned previously the two most established approaches are ^{15}N and SILAC labeling. Here we chose to use the plant SILAC labeling approach primarily due to the superior mass spectrometric and bioinformatic quantitation framework.

We used the same cell cultures as in the RNA-Seq, Ribo-Seq, and protein abundance data. To reiterate briefly, we grew two Arabidopsis suspension cell cultures over 12 days in low-nitrogen medium and supplemented them every 24 h with 350 µM heavy lysine (either Lys4 or Lys8) with a passage to fresh medium at day 6 to keep the cultures in steady growth. During this pulse-phase of the experiment, the culture's proteomes got labeled with the heavy lysine to a rate of ~85% (chapter 6).

After a labeling phase of 12 days, one part of the cultures was taken for the RNA-Seq, Ribo-Seq, and protein abundance analyses. The second part was used for the protein turnover experiment, where the cultures were washed twice with non-supplemented medium and were split into three subcultures each. The 15 mL cultures were inoculated with densities of 100 mg fresh weight / mL medium. The cultures were then supplemented with 350 µM of the reciprocal isotopically labelled lysine: formerly Lys4-labelled cultures were supplemented with Lys8 and formerly Lys8-labelled cultures were supplemented with Lys4. After an acclimation period of 2 h, three of these cultures were treated with 100 nM flg22, while the remaining three cultures were mock treated. The treatment and control groups were split among the Lys4- and Lys8-labelled cultures so that a potential effect of the labeling could be distinguished from the effect of treatment with flg22.

During the chase period, samples were taken before the flg22 treatment and 2, 4, 8, 16, 32, 48, and 72 h post-treatment. The supplementation with daily doses of 350 µM lysine was maintained during the chase phase. Proteins were extracted from the samples and analyzed mass spectrometrically to quantify the progression of the label-switch. The high number of sampling time points should allow us to accurately describe the labeling trajectories of both quickly and slowly degrading proteins.

7.2 Growth and viability of the cultures

During the 72 h of the chase period, I monitored the growth of the cultures by taking aliquots and determining the fresh weight of the cell pellet after centrifugation. While all cultures grew steadily during the experiment, the flg22 treatment had a substantial effect on the growth rates (Figure 7.1A).

To describe the effect of flg22 on the growth rates in a quantitative manner, I fitted an exponential growth curve to the culture densities using the following formula (Figure 7.1C):

$$Densi_{Tx} = Densi_{T0} * exp(k_{dil} * time)$$

The fitted growth rate constants k_{dil} indicate a decrease of the growth rate upon flg22 treatment by 37 % on average (Figure 7.1B, average flg22-treated k_{dil} = 0.0107, average untreated k_{dil} = 0.0171). The labelling background also had an effect on the growth rates, as Lys8-labelled cultures grew on average 15.9% slower than the Lys4-labelled cultures (average Lys4-labelled k_{dil} = 0.0151, average Lys8-labelled k_{dil} = 0.0127). As a point of reference, the untreated ^{15}N-labeled Arabidopsis suspension cell culture employed by Li *et al*, (2012) grew at a k_{dil} of 0.007, significantly slower than the growth rates determined for the cell culture employed in this study.

Irrespective of the differences in growth rates, all cultures remained viable during the experiment (Figure 7.1D). Neither the flg22 treatment nor the lysine labeling led to a significant difference in dehydrogenase activity (lysine p-value = 0.35, flg22 p-value = 0.33, t-test). I did not detect any macroscopic, physiological differences between the cultures.
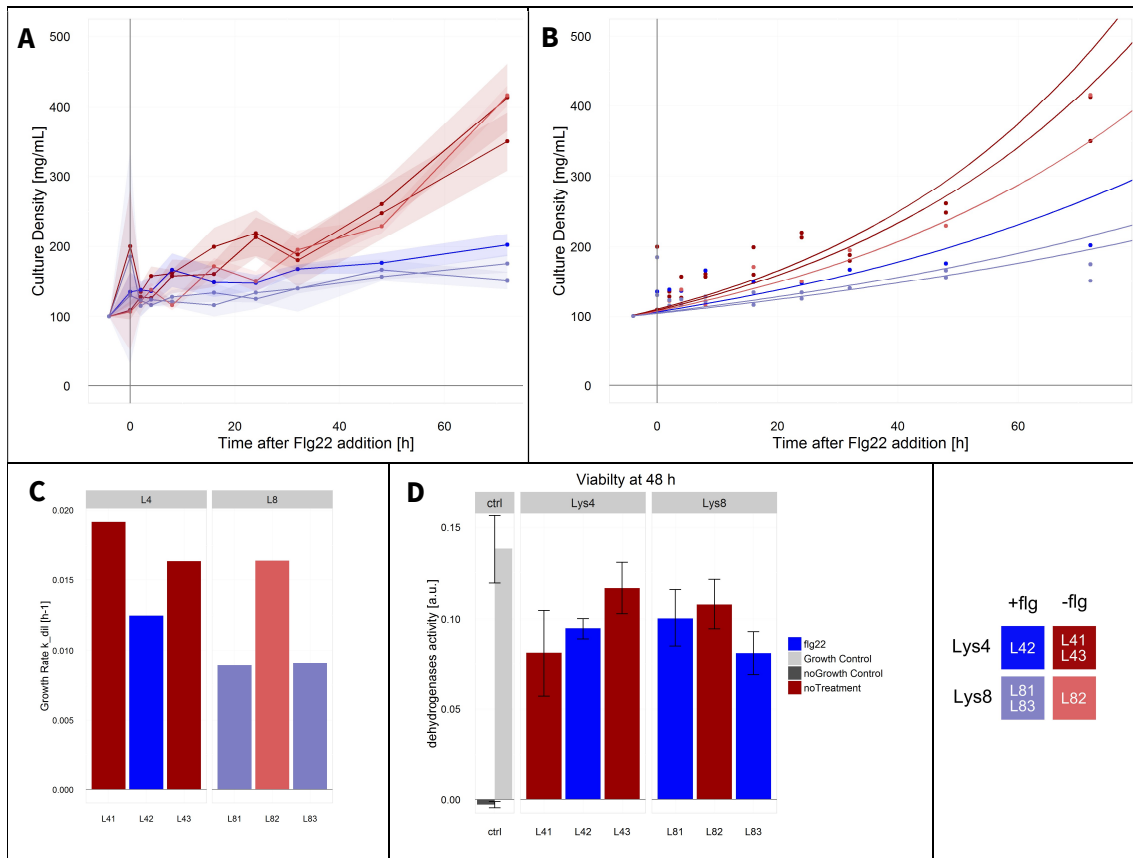
Figure 7.1: Growth rate and viability during the turnover experiment.
A: Measured culture densities of the flg22-treated (blue) and untreated (red) cultures. The standard deviations of the culture density measurements are indicated by colored ribbons. B: Exponential growth curves (lines) fitted to the measured culture densities (dots). C: Growth rate constants k_{dil} from the fitted exponential curves. D: Cell culture viability 48 h after the treatment as determined by the reduction of TTC to formazan. The coloring scheme at the bottom right is representative for all plots.

7.3 Mass spectrometric coverage and quality control

In the MS analyses of the protein turnover samples, I identified in total 73'098 peptides that map to 8'669 proteins at a false discovery rate of 1% at both the peptide and protein level. As these samples were processed identically to the protein abundance samples, the same issues are valid regarding non-quantifiable arginine-terminating peptides and a limited overlap of quantified proteins in all samples as discussed in chapter 6.5. On average, 2128 proteins were quantified in each sample (Figure 7.2A), 1966 proteins were quantified in more than 50% of all samples and 481 proteins were quantified in all 44 samples (Figure 7.2B,C).
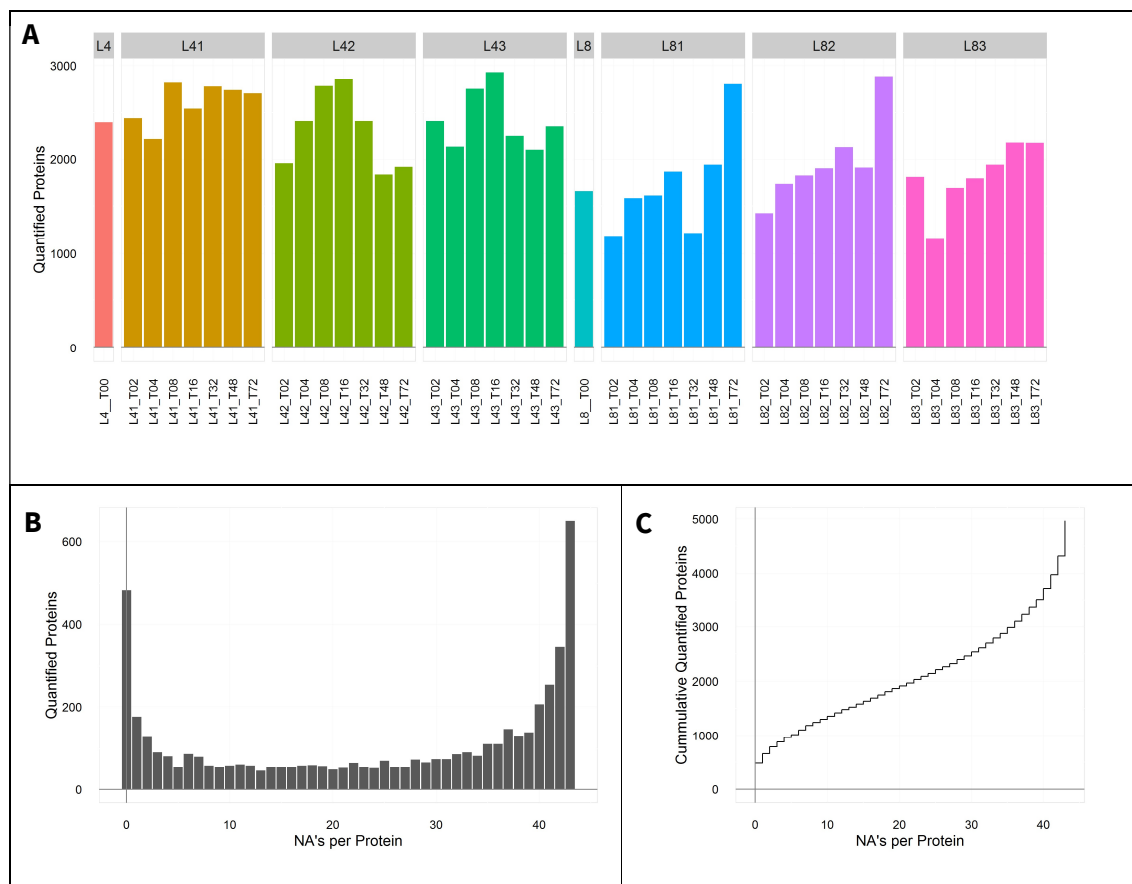
Figure 7.2: Mass spectrometric coverage.
A: Number of proteins quantified in each sample. The names of the samples describe the labeling background with the first two letters, the replicate number with the third letter and the sampling time in hours after the underscore. B: Number of missing quantitations per protein (NA = not acquired). C: Cumulative curve of plot B. Note that the number of quantified proteins is lower than the number of identified proteins due to non-labelled arginine-terminating peptides.

I verified the quality of the mass spectrometric analysis with PTXQC (proteomics quality control, Bielow *et al*, 2016), which analyzes quality control (QC) parameters from the MaxQuant output. PTXQC identified 6 LC-MS runs where every quality metrics failed (Figure 7.5A). Manual inspection of the total ion current verified that these samples were of low quality, presumably due to problems with the electrospray ionization. MaxQuant relies on stable LC performance in all samples to match

MS1 features between runs. I therefore excluded the problematic samples to assure the quality of our dataset and reanalyzed the cleaned data set with MaxQuant and PTXQC (Figure 7.5B).

Now, the MS measurements are of constant good quality, as indicated by the average overall quality column of the heat map (Figure 7.5B). The MS2 identification rate over the LC retention time (RT) was assessed as underperforming for all samples, but this means only that the number of MS2 identifications was not uniformly spread over the LC elution time. That is to be expected when the peptide mixtures are fractionated by HILIC before the LC-MS analysis. As exemplified in Figure 7.5C, each fraction produces a unique, non-constant peptide elution profile, which cannot be optimized easily and is per se not problematic for identification or quantification – it just shows that the MS sample processing pipeline could be optimized here. More interesting is metric 14, indicating an overall low MS2 identification rate over all samples (Figure 7.5D). This metric behaves similarly to the protein abundance data set, where the effect of low MS2 ID rates for high TopN cycles (Figure 7.5E, F) is discussed in more detail (chapter 6.5). Additionally, the ID rate decreases steadily with the number of analyzed samples, which might have been caused by a contaminated orifice or a deteriorating LC column. The mass spectrometer's mass calibration seems to be unproblematic according to metric 11 and 12.

In summary, the mass spectrometric measurements were of good quality, but improvements in the MS setup could increase the analysis depth in the future.

Figure 7.5: Quality control plots of the mass spectrometric analyses from PTXQC.
A: Overview of all PTXQC metrics over all samples and B: after exclusion of the six problematic samples. C: Exemplary non-uniform ID rate over retention time (RT) of the 1st and 5th HILIC fraction. During the LC, the peptides did not elute uniformly. D: MS2 identification rate over all samples. E: Exemplary TopN plot, which is a histogram of duty cycles with n-number of MS2 scans per duty cycle. F: Exemplary MS2 ID rate per

TopN. The MS2 scan identification rate decreases, the more MS2 scans per duty cycle are performed. EVD = Evidence file from MaxQuant, ID = identification, MBR = Match between runs feature of MaxQuant, Cal = mass calibration, MC = missed cleavages.

7.4   RIA calculations

I used MaxQuant to obtain the ratios of the MS-quantified Lys0-, Lys4- and Lys8-subproteomes. MaxQuant reports protein-level quantitations in the file "protein_groups.txt" as the median of the non-normalized MS1 intensity ratios of all peptides mapping uniquely to that protein. It computes within each individual sample the ratios of all three label states (Heavy = H, Medium-heavy = M and Light = L): $\frac{H}{L} = \frac{Lys8}{Lys0}, \frac{H}{M} = \frac{Lys8}{Lys4}$ and $\frac{M}{L} = \frac{Lys4}{Lys0}$. Lys0 represents the unlabeled protein subpopulation containing the naturally occurring light lysine isotope. Lys4 and Lys8 represent the protein subpopulations containing the artificially heavy- and mid-heavy-labelled lysines.

Starting from these ratios, I calculated the relative isotope abundances (RIA) for each label state. RIAs are defined as the ratio of one label state against the sum of all other states. In our triple-labeling case, it is defined as

$$RIA_{LysX} = \frac{LysX}{Lys0+Lys4+Lys}$$

I calculated the RIA of each label state from the non-normalized protein-level ratios using the following formulae:

$$RIA_{Lys0} = \frac{Lys0}{Lys0+Lys4+Lys8} = \frac{1}{\frac{Lys0+Lys4+Lsy}{Lys0}} = \frac{1}{\frac{Lys0}{Lys0}+\frac{Lys4}{Lys0}+\frac{Lys8}{Lys0}} = \frac{1}{1+\frac{Lys4}{Lys0}+\frac{Lys8}{Lys0}} = \frac{1}{1+\frac{M}{L}+\frac{H}{L}}$$

$$RI_{Lys4} = \frac{Lys4}{Lys0+Lys4+Lys8} = \frac{1}{\frac{Lys0+Lys4+Lsy8}{Lys4}} = \frac{1}{\frac{Lys0}{Lys4}+\frac{Lys4}{Lys4}+\frac{Lys8}{Lys4}} = \frac{1}{1/\frac{Lys4}{Lys0}+1+\frac{Lys8}{Lys4}} = \frac{1}{1/\frac{M}{L}+1+\frac{H}{M}}$$

$$RIA_{Lys8} = \frac{Lys8}{Lys0+Lys4+Lys8} = \frac{1}{\frac{Lys0+Lys4+Lsy}{Lys8}} = \frac{1}{\frac{Lys0}{Lys8}+\frac{Lys4}{Lys8}+\frac{Lys8}{Lys8}} = \frac{1}{1/\frac{Lys8}{Lys0}+1/\frac{Lys8}{Lys4}+1} = \frac{1}{1/\frac{H}{L}+1/\frac{H}{M}+1}$$

Taking into account the label-swap design, I calculated the RIAs for the old and the newly-added label from the lysine-specific RIAs:

$$RIA_{newLys} = \begin{cases} RI_{Lys4} \ for\ initially\ Lys8\ labeled\ samples \\ RI_{Lys8} \ for\ initially\ Lys4\ labeled\ samples \end{cases}$$

$$RIA_{oldLys} = \begin{cases} RI_{Lys8} \ for\ initially\ Ly\ 8\ labeled\ samples \\ RIA_{Ly} \ for\ initially\ Lys4\ labeled\ samples \end{cases}$$

Equally, I will speak in following paragraphs of oldLys- and newLys-labelled proteins to take the label-swap design into account.

7.5   Median label switch

From the calculated gene-wise RIA values we determined the average progression of the label-swap. Figure 7.3A shows the RIA trajectories of the three label states, indicating a constant exchange of the old to the new label while RIA_{Lys0} stays rather low. At the end of the experiment, the majority of the old label has been exchanged.

When examining the trajectories for each individual culture, pronounced differences become apparent. Figure 7.3B-D depict the culture-specific median RIA trajectories for RIA_{oldLys}, RIA_{newLys}, and RIA_{Lys0}. In all three figures those trajectories of the direct replicate pairs, which underwent the identical combination of treatment and labeling, align best (cultures L41 and L43 and cultures L81 and L83, respectively). Both treatment and labeling have an effect on RIA trajectories, but the differences between flg22-treated (blue) and untreated (red) trajectories are larger than between Lys8-labelled (lighter colors) and Lys4-labelled (darker colors) cultures. Lys8-labelled cultures exchange their label substantially quicker than their Lys4-labelled counterparts.

Importantly, the RIA_{Lys0} trajectories, which are indicative for the labeling efficiencies, do not stay constant over the course of the experiment. For the Lys4-labelled untreated cultures, the unlabeled proteins reached up to 25 % of the total. These cultures also showed the highest growth rates during the experiment (Figure 7.1A). We suppose that these differences in the labeling efficiencies are caused by differences in the growth rates. A higher growth rate leads to increased protein biosynthesis, which leads to a higher demand for all amino acids. This lowers the intracellular lysine concentration, which releases the negative-feedback down-regulation of aspartate kinase (Heremans & Jacobs, 1997) and LL-diaminopimelate aminotransferase *DapL* (Ghislain *et al*, 1990) . Consequently, the cells increase their endogenous biosynthesis of unlabeled lysine. Newly synthesized proteins will, therefore, incorporate either newLys or increasingly also Lys0.

This connection between growth rate and lysine labeling efficiency can be well observed when comparing Lys8-labelled cultures with Lys4-labelled ones. Lys8-cultures grew on average 15.9 % slower than Lys4-cultures, leading to a lower RIA_{Lys0} (-37.4 %) and a proportionally higher RIA_{newLys} (+25.1 %) at 72 h post-treatment irrespective of the flg22 treatment.

However, assessments of RIA_{oldLys} as a proxy for protein degradation are not affected by a non-constant labeling rate, as the denominator in the RIA_{oldLys} formula contains both the Lys0- and newLys-labelled proteins (chapter 7.4). Newly synthesized proteins, regardless of their labeling state, are summed up with the initial Lys0-labelled proteins to a constant factor. Consequently, the differences between Lys4 and Lys8 cultures are less striking for RIA_{oldLys} than for $RIA_{newLys.}$ Importantly, RIA_{oldLys} is more strongly affected by the flg22 treatment, than by the differences between Lys4- and Lys8-labeling.
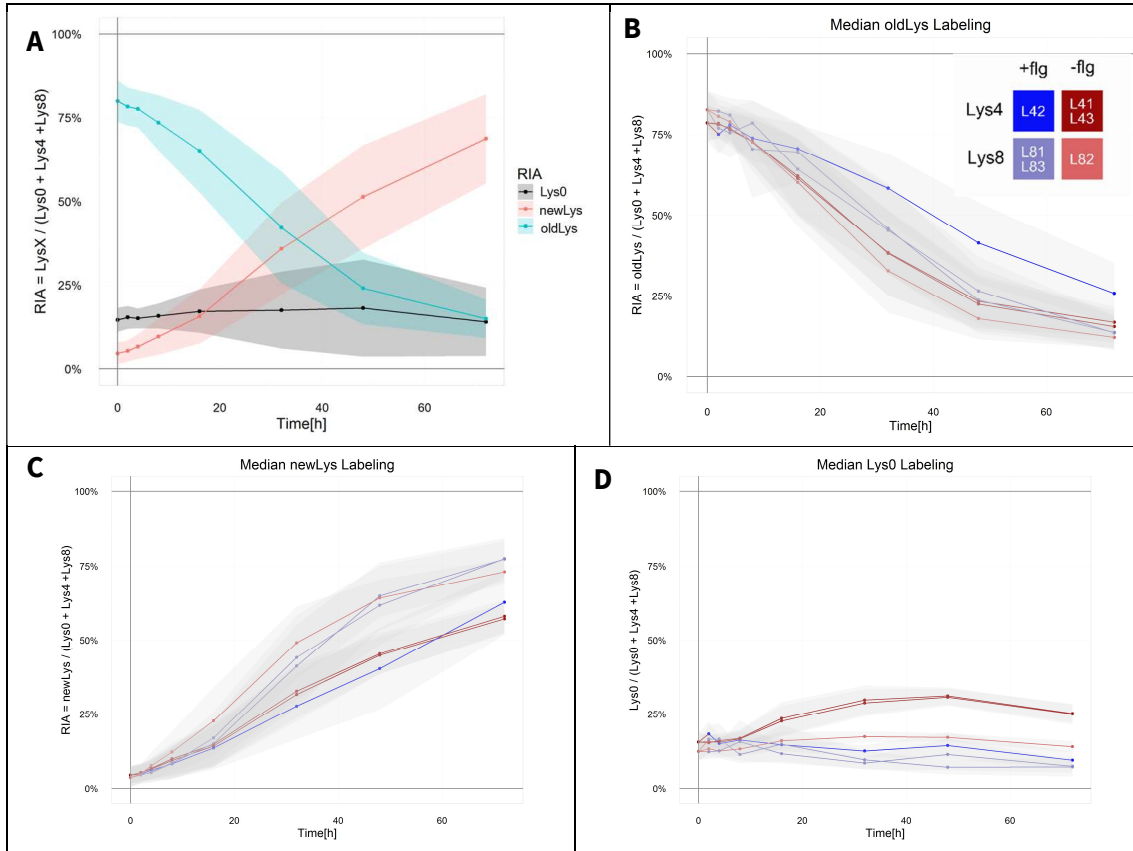
Figure 7.3: RIA trajectories during the protein turnover experiment.
A: Trajectories of the median RIA_{oldLys}, RIA_{newLys}, and RIA_{Lys0} averaged over all six cultures.
B-D: Culture-specific median RIA_{oldLys} (B), RIA_{newLys} (C), and RIA_{Lys0} (D) trajectories. As a measure of variance, the MAD (median absolute deviation) is indicated as ribbons. The color scheme for all culture-specific trajectories is indicated in B.

7.6   Assessing protein degradation

Protein degradation is often described with exponential decay functions (chapter 2.4):

$$Protein_{Time_X} = Protein_{Ti\quad 0} * exp^{(- k_{deg} * Time_X)}$$

We refactored the formula such that the curve always starts at 1 at time point zero:

$$\frac{Protein_{Time_X}}{Protein_{Time_0}} = exp^{(- k_{deg} * Tim\ x)}$$

And we use RIA_{oldLys} as the metric for protein abundance. RIA_{oldLys} relative to time point zero was termed relative RIA_{oldLys}.

$$relative\ RIA_{oldLys} = \frac{RIA_{oldLys_{Time_X}}}{RIA_{oldLys_{Time_0}}} = exp^{(- k_{deg} * Time_x)}$$

Fitting the relative RIA$_{oldLys}$ trajectories to this exponential decay model revealed systemic discrepancies (exemplified by the median relative RIA$_{oldLys}$, Figure 7.4A-B), as especially the early time points show a clear difference between the measured data and the fitted curve.

An alternative description of protein decay curves was proposed by Martin-Perez & Villen (2015). They describe that in yeast 10-13% of the protein degradation curves can be described more accurately with a logistic decay model. Such a model assumes a sigmoidal or "s"-shaped curve with a non-steady decay rate. At the beginning, the decay rate is slow and increases until it reaches the maximal decay rate at the inflection point and decreases steadily thereafter. This trajectory was formalized by Martin-Perez & Villen (2015) with the following logistic decay function:

$$Protein_{TimeX} = Protein_{Time0} - \frac{Protein_{Time0} - Protein_{Time}}{1 + exp^{(-k_{deg} * (Time_X - Time_{hal\ -life}))}}$$

Time$_{half-life}$ signifies the inflection point of the sigmoidal curve or, in other words, the y-axis or time-offset. A value larger than 0 leads to the above-mentioned slow decay or lag phase at the beginning of the degradation curve.

I adapted this logistic decay model for our data, using the relative RIA$_{oldLys}$ value as the protein degradation metric ($Protein_{TimeX} = \frac{RIA_{TimeX}}{RI\ _{Time}} =$ relative RIA$_{oldLys}$) and determined that the labelled proteins are completely degraded at the infinite time point ($Protein_{Time\infty} = 0$). This simplifies the logistic decay formula to:

$$\text{RIA}_{\text{oldLysT0}} = 1 - \frac{1}{1 + exp^{(-k_{deg} * (Time_X - Time_{half-life}))}}$$

I also allow Time$_{half-life}$ to be smaller than zero, which is biologically not meaningful, but allows to shift the inflection point of the curve to the left of time point zero. In such a case, the logistic curve becomes more similar to an exponential decay curve with an initially quick and constantly slowing decay rate. Consequently, this version of a logistic decay model is rather a combination of an exponential and logistic decay model.

When fitting my logistic decay model to the median relative RIA$_{oldLys}$ (Figure 7.4C), I see a more accurate fit between the model and the data compared to the classical exponential decay model (Figure 7.4B).
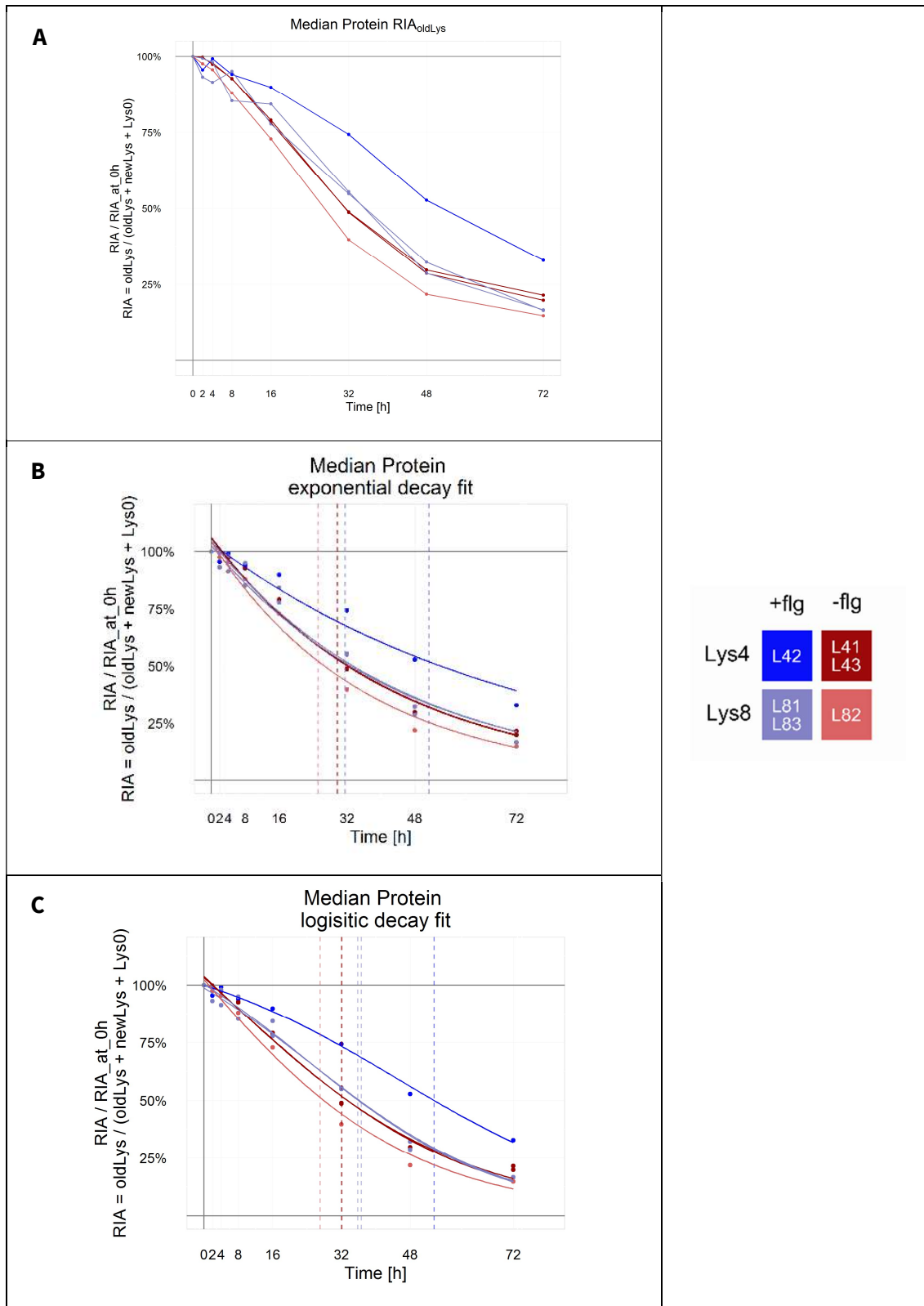
Figure 7.4: Modeling the culture-specific median protein turnover trajectories (relative RIA$_{oldLys}$, A) by exponential (B) or logistic decay (C).
The half-life times calculated from the fitted degradations are indicated by dashed vertical lines.

7.7 Comparing exponential and logistic decay models

I compared the quality of the fitted curves from both models by calculating the weighted residual sum-of-squares of the fits. Nearly all fitted exponential decay models have larger weighted residual sum-of-squares than the corresponding fits with the logistic decay model (Figure 7.6A, B), indicating that the logistic decay models generally align better to the relative RIA$_{oldLys}$ values than the exponential decay models. However, it is important to note that the logistic model is more complex as it is based on two parameters, k_{deg} and Time$_{half-life}$, while the exponential model is based only on one, k_{deg}. I therefore compared each logistic and exponential decay model pair with an ANOVA to determine if the logistic decay fit was significantly better despite the fewer available degrees of freedom. According to the ANOVA comparison, 29.7 % of all logistic decay models describe the trajectories more accurately than the exponential decay model (p-value < 0.05). This ratio varies with the flg22-treatment. The slower-growing flg22-treated samples showed more trajectories that can be fitted better with a logistic model than the non-treated cultures (Figure 7.6C).

I verified the assessment of the ANOVA with comparisons of the Akaike's information criterion (AIC) of both models (Figure 7.6 D). The AIC gives the log-likelihood of the fit and 58% of all fittings had a lower AIC with the logistic decay model. The AIC comparison also indicated that trajectories of flg22-treated cultures were better described by the logistic decay function (Figure 7.6E).

It is important to note that goodness-of-fit comparisons of non-linear models are difficult to perform robustly. To our knowledge, there is no general-purpose comparison available such as the R^2 statistic for linear models. Still, the presented comparisons of the two models are useful to identify trends about the two models. The logistic decay model depicts the RIA$_{oldLys}$ trajectory in our opinion at least as accurately as the exponential decay model and in a number of cases significantly better (Figure 7.6A-D). Several of the protein decay trajectories simply cannot be described faithfully with an exponential decay model. In a conventional approach, these proteins would have been excluded based on a low quality-of-fit statistic to the exponential decay model as applied in Schwanhäusser et al. (2011).

Regardless of the differences in the fitting quality, I observed for the majority of proteins only small differences in the calculated half-lives using either model (indicated by vertical dashed lines in the graphs of Figure 7.4B-C and Figure 7.6H-K). Figure 7.6F depicts these differences systematically and shows only for long-lived proteins stronger deviations between the models. Their half-life times seem to be overestimated by the exponential decay model, as it cannot take initial lag phases into consideration (e.g. Figure 7.6J). These long-lived proteins also show an enrichment for higher AIC differences, indicating significantly better fits for the logistic decay models (Figure 7.6G).

Proteins for which degradation trajectories were significantly better modeled by a logistic decay model were enriched for several GO categories of the response class and two large metabolic categories (Table 7.1). There is no enrichment for ribosomal processes as it was reported in yeast (Martin-Perez & Villén, 2015).

In the following text I will focus on the logistic decay model, as I consider it superior in describing the true degradation processes.
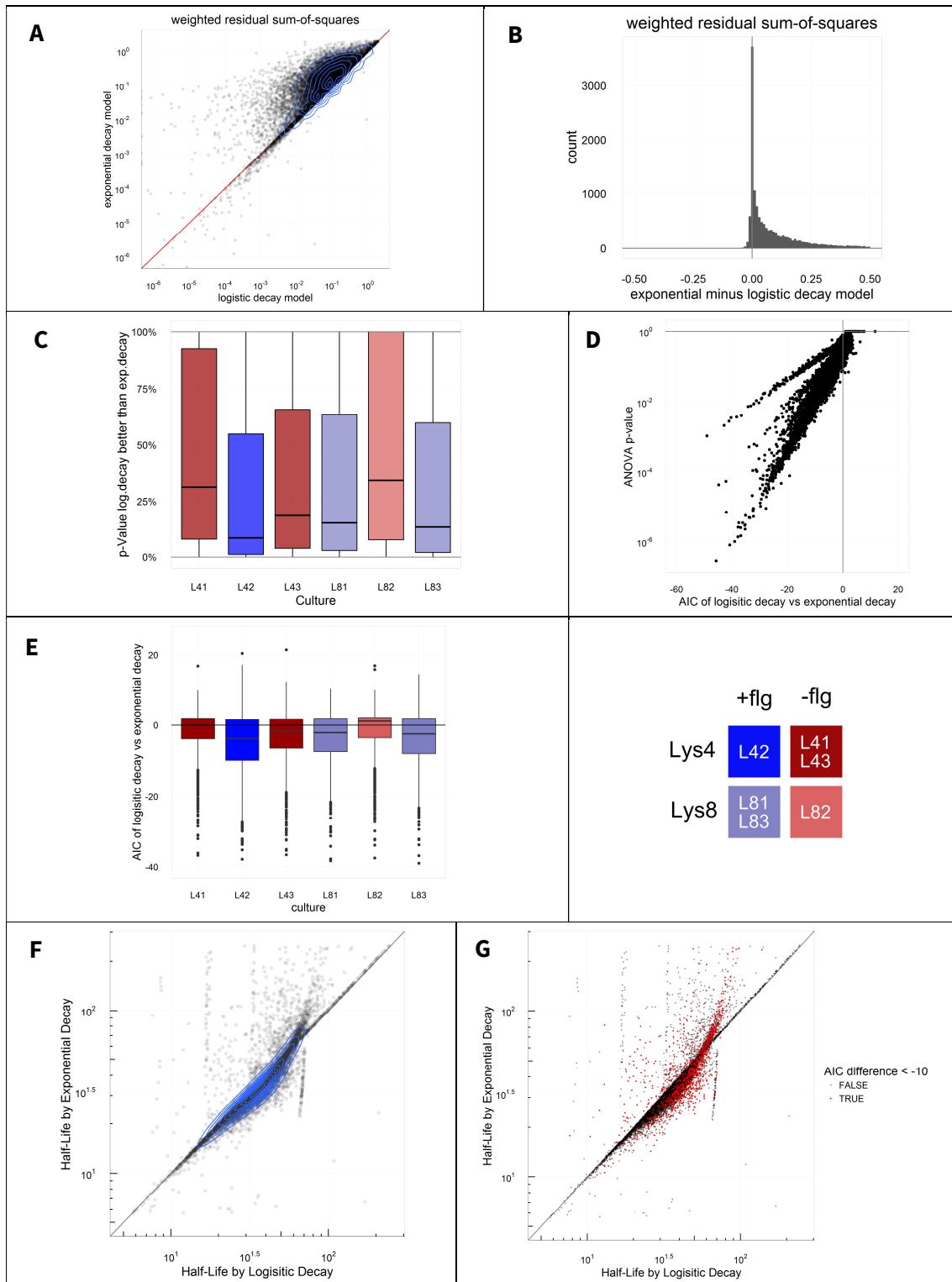
Figure 7.5: Comparison of the exponential and logistic decay models fitted to the protein- and culture-specific relative RIA$_{oldLys}$ trajectories.

A: Comparison of the weighted residual sum-of-squares of both models. The majority of points are located above the red identity line, indicating a lower deviation for most logistic decay models. The density of the points is indicated by blue topographic-like contours. B: Histogram of the differences between the weighted residual sum-of-

squares of both models. For the majority of models, the difference is positive, indicating a larger weighted residual sum-of-squares for the exponential decay model. C: P-values of the ANOVA comparison of both models along the six cultures. D: Comparison of the ANOVA p-values and differences in AICs comparing the two models. E: Culture-specific AIC differences between the two models. A more negative value is indicative for a better fit by the logistic decay model. F: Half-life times calculated by the exponential or logistic decay models. The density of the points is indicated as blue topographic-like contours.
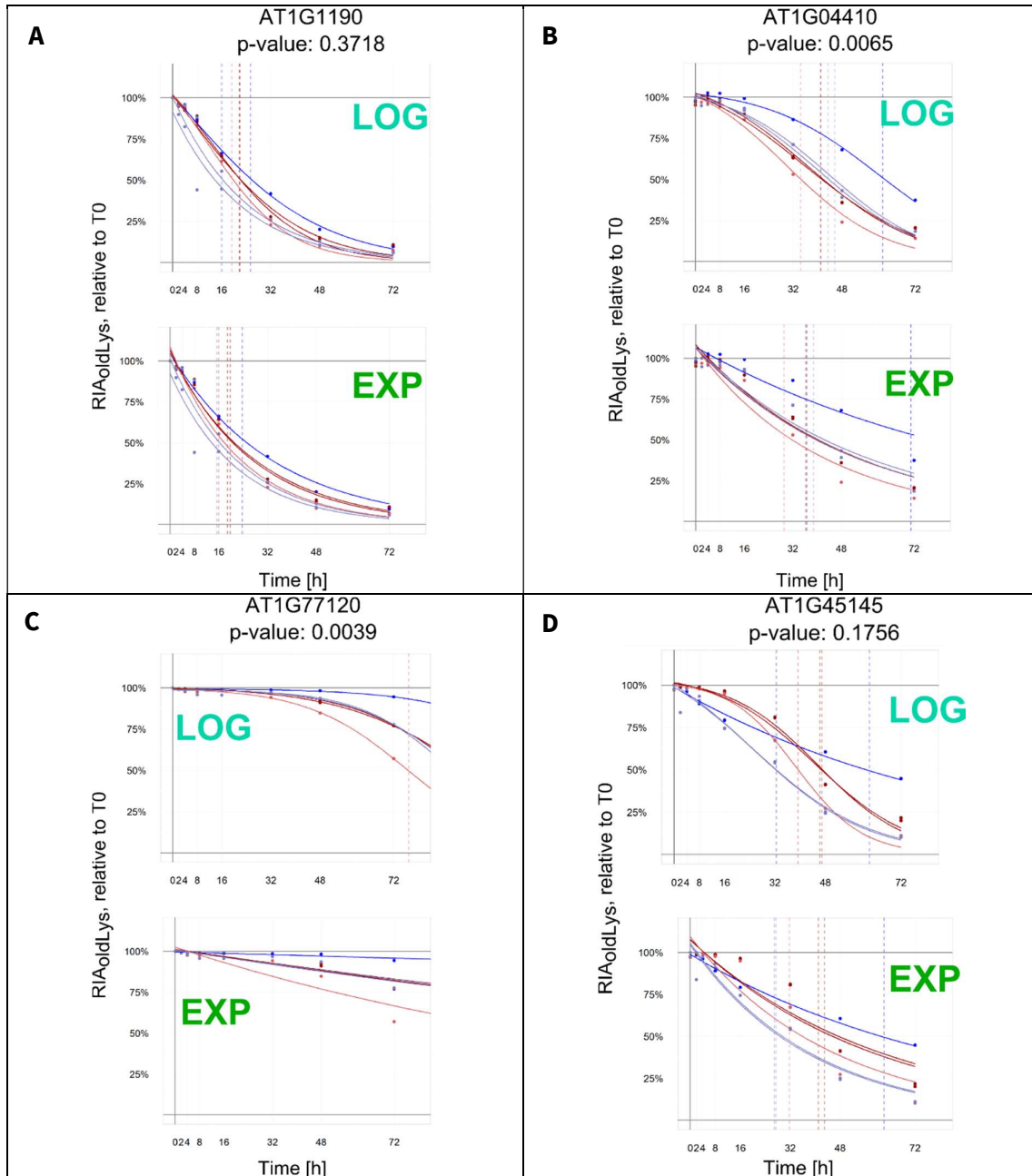
Figure 7.6: Logistic (top) and exponential (bottom) decay fits to the relative RIA$_{oldLys}$ trajectories of exemplary genes (A-D). The half-lives of the proteins calculated based on the fitted models are indicated as dashed vertical lines in the plots. The average p-values from the ANOVA comparisons of the two models are listed under the plots.

Table 7.1: GO term enrichment for the 282 proteins with decay trajectories that were fit significantly better by logistic decay models (ANOVA p-value < 0.05 in ≤ 3 cultures).

The enrichment was performed with Panther (Mi *et al*, 2016) with a Bonferroni-adjusted p-value cutoff of < 0.05. The reference list consisted of all 2'155 modeled proteins. Here only the lowest nodes of a hierarchy are listed, parent nodes are left out. # = number of proteins.

GO BIOLOGICAL PROCESS	# IN REFERENCE	# IN QUERY	# EXPECTED IN QUERY	FOLD ENRICHMENT	ADJ. P-VALUE
response to oxidative stress	70	43	20.39	2.11	3.68E-03
response to acid chemical	97	52	28.25	1.84	1.62E-02
response to cadmium ion	145	77	42.23	1.82	2.33E-04
response to oxygen-containing compound	129	68	37.57	1.81	1.53E-03
single-organism catabolic process	106	55	30.87	1.78	2.27E-02
oxidation-reduction process	237	109	69.03	1.58	8.27E-04
carboxylic acid metabolic process	227	98	66.12	1.48	3.78E-02
RNA processing	60	2	17.48	< 0.2	2.37E-03

7.8   Protein complex associations

McShane et al. (2016) reported that proteins that are better described by a non-exponential decay model are twice as likely associated with protein complexes. I tested whether the proteins identified here with decay trajectories that are better described by a logistic decay model, are similarly more likely associated with a protein complex. When using the protein complex associations from the SUBA database (http://suba3.plantenergy.uwa.edu.au,(Tanz *et al*, 2013)), we detected a small but significant change in the ANOVA p-value distribution for the proteins that are associated with a protein complex (p-value = 0.003074, Wilcoxon rank sum test, Figure 7.7A). This means that proteins with a protein complex annotation are more likely better described by a logistic decay model.

Additionally, I tested the protein complex associations from the plant interactome consortium (http://interactome.dfci.harvard.edu/, (Braun P, Carvunis AR, Charloteaux B, Dreze M, Ecker JR, Hill DE, Roth FP, Vidal M, Galli M, Balumuri P, Bautista V, Chesnut JD, Kim RC, de los Reyes C, Gilles P, Kim CJ, Matrubutham U, Mirchandani J, Olivares E, Patnaik S, Quan R, Ramaswamy G, Shinn P, Sw, 2011)). Also with these associations I found that the proteins associated with a protein complex are significantly more likely better described by a logistic decay model than an exponential decay model (p-value < 2.2e-16, Wilcoxon rank sum test, Figure 7.7B).

These differences are much smaller than the two-fold enrichment reported by McShane et al. (2016). However, the quality of the protein complex associations in Arabidopsis are presumably not as high as those in the manually curated, mammalian database (CORUM, Ruepp *et al*, 2010), which might explain the differences. It remains that the proteins that are better described by a logistic decay function are more likely associated with a protein complex. As mentioned above, the decay trajectories of especially long-lived proteins are generally modeled better by logistic decay models. The finding that proteins with decay trajectories that are better described with a logistic decay

function are enriched for complex formation might therefore reflect the fact that complex formation leads to protein stabilization (Goldberg, 2003; McShane *et al*, 2016).
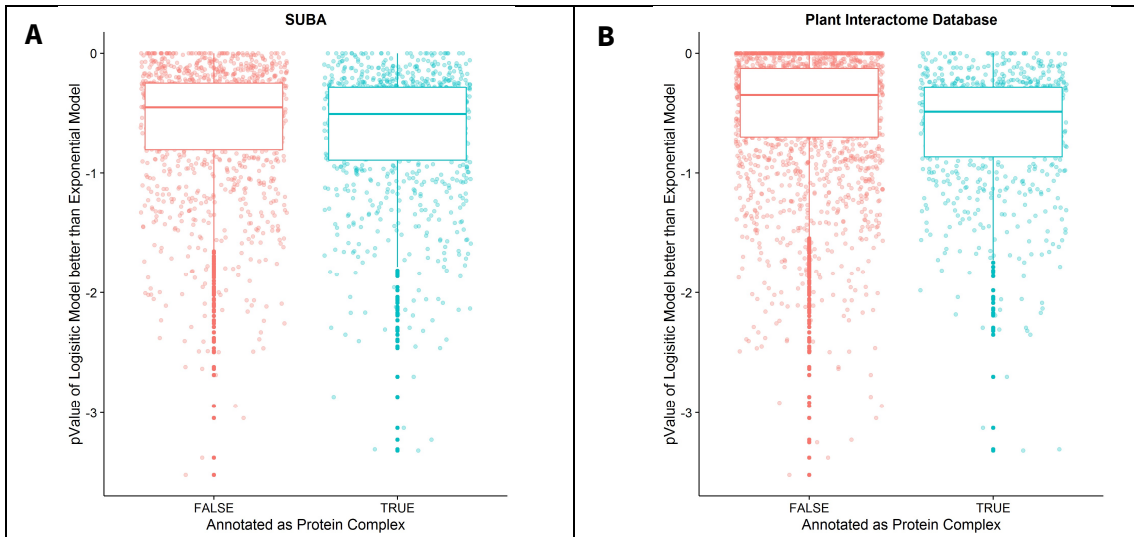
Figure 7.7: Effect of protein complex associations on the distribution of the ANOVA p-values comparing the exponential and logistic decay models.

The information on protein complex associations was either taken from SUBA (A, Tanz *et al*, (2013) or from the plant interactome database (B, Dreze *et al*, (2011)).

7.9 Bona fide protein half-lives

Under the assumption that the cell cultures are in steady-state with constant protein abundance and protein synthesis, we can calculate the protein turnover from the fitted relative RIA_{oldLys} trajectories. The turnover of proteins can be expressed in half-lives, which is the time until 50% of a protein population is being renewed. The term was originally defined for exponential decay models, but can also be calculated for non-exponential decay models such as the logistic decay model by calculating the time until the fitted curve reaches 50% of the RIA_{oldLys} value at the beginning of the experiment.

I determined the half-lives of 2802 proteins, 1509 of which were characterized in all six cultures (Figure 7.8A). The culture-specific density distributions of these half-lives are depicted in Figure 7.8B. Again, cultures with the same treatment and labeling background correlate better, as was observed before for the median RIA trajectories (Figure 7.3) and the growth curves (Figure 7.1). As discussed previously, this is most probably due to the differences in growth rates caused by the treatment and the initial labeling.

Consequently, I normalize the half-lives by the growth rate, following the approach published by Martin-Perez & Villen, (2015), who compared the half-lives of prototroph- and auxotroph-grown yeast cultures. Accordingly, the half-lives were divided by the culture's doubling time, which is the time until an exponentially growing culture doubles in density, giving the dimension-less normalized half-life:

$$normalized\ half\text{-}life = \frac{Half\text{-}Life}{doubling\ time}$$

A normalized half-live smaller than 1 means that the protein has a higher turnover rate than the culture's growth rate. Following the interpretation of Eden et al, (2011), this means that the protein's half-life is dominated by targeted degradation processes. A relative half-life larger than 1 means in contrast that the protein's half-life is dominated by general dilution processes caused by the growth of the cells. Normalized half-lives are dimensionless values.

The doubling time of a culture is calculated similarly to the half-life by solving the exponential growth formula for time at $Density_{Tx} = 2 * Densi_{T0}$:

$$2 * Densi_{T0} = Density_{T0} * exp(k_{dil} * Doubling\ Time) \xrightarrow{yields} Doubling\ Time = \frac{log(2)}{k_{dil}}$$

The calculated doubling times range from 36.2 h for the L41 culture to 77.9 h for the L81 culture.

The distributions of the growth-normalized relative half-lives are depicted in Figure 7.8C. The untreated cultures show a higher correlation (average rho = 0.806) than the flg22-treated cultures (average rho = 0.664). By limiting the analysis to the untreated cultures where the proteome is unaffected from stimuli or developmental reorganization, I can make statements on the half-lives of proteins in Arabidopsis cell cultures. The calculated median half-life in our data set was 30.4 h ± 2.9 and the relative median half-life is 0.76 ± 0.12. This growth-normalized value is substantially smaller than 1, indicating that the majority of Arabidopsis cell culture proteins have half-lives smaller than the culture's growth rate and are subject to targeted protein degradation processes.

The relative half-lives are in the same order of magnitude as the relative half-lives reported for auxotrophic and prototrophic yeast cultures (Martin-Perez & Villen, 2015, Figure 7.8D), while the absolute non-growth normalized half-lives are un-comparable (yeast median half-life = 2.12 h at 2.04 h culture doubling time, Arabidopsis median half-life determined here = 30.4 h). Martin-Perez and Villen (2015) report a median relative half-life of 1.026 using a dynamic SILAC approach. Using a microscopy based bleach-chase approach, Eden et al, (2011) report a median relative half-life of 0.659 for yeast cells, which is closer to our determined median relative half-life of 0.762.
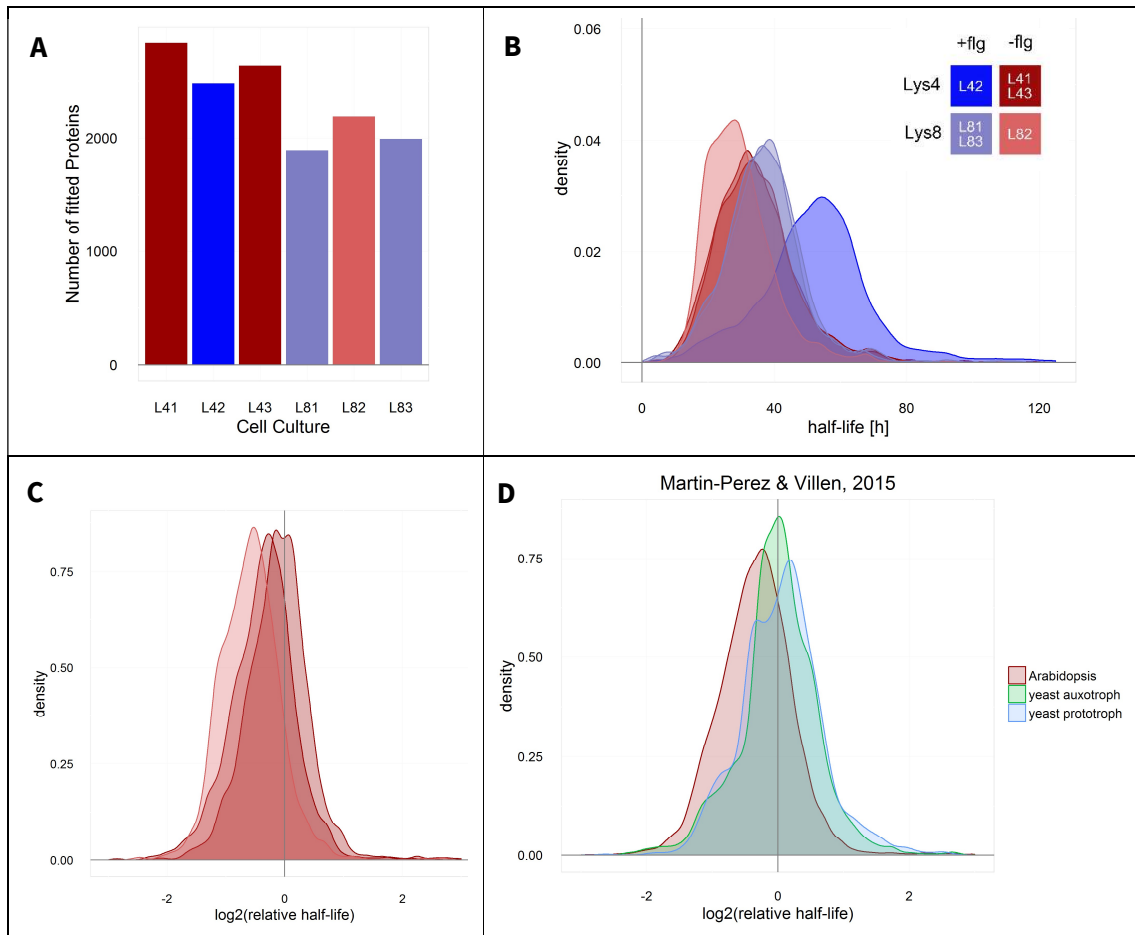
Figure 7.8: Protein half-lives.
A: Number of calculated protein half-lives in this study. B: Density plots of the half-lives calculated from the logistic decay fitting. C: log2-transformed protein half-lives normalized by the cell culture's doubling time used for the bona fide protein turnover rate calculation. D: Comparison of the relative half-lives of subplot C with the relative half-lives reported for auxotrophic and prototrophic yeast cultures (Martin-Perez & Villén, 2015).

7.9.1 Comparison to literature data

It is difficult to draw direct comparisons between protein turnover studies as they differ largely in both biochemical and mathematical methodology. There are also only few turnover studies published in the plant sciences field. I focused the comparisons on studies where both the protein-specific half-lives and growth rates were reported.

The laboratory of Harvey Millar (University of Western Australia) reported several studies on plant protein turnover by fitting ^{15}N pulse-labeling trajectories to exponential decay models. In Nelson *et al*, (2013), they determined the half-lives for 258 mitochondrial proteins from an Arabidopsis cell culture and in Nelson *et al*, (2014a) they reported the half-lives for 521 proteins in maturing barley leaves, 231 of which have homologs in our study.

Comparing the half-lives without normalizing for the growth rate creates large differences for individual proteins of more than a factor of 10. Yet the growth rate normalization brought the half-lives into the same order of magnitude (Figure 7.9A, B). While there was no correlation to the barley proteins (rho = -0.04876, Figure 7.9B), it was substantially higher for the mitochondrial proteins from the Arabidopsis cell culture (rho = 0.2463, Figure 7.9A).

In the most recent protein turnover study in Arabidopsis reported by Fan et al, (2016), they determined the turnover for 250 protein in Arabidopsis seedlings by fitting the ^{15}N pulse-labelled trajectories to an exponential decay model. They did not publish the growth rate of the seedlings so that we could only compare the non-normalized half-lives (Figure 7.9C, rho = 0.373). Interestingly, the half-lives of both studies do not show a clear offset, hinting towards a similar growth rate of the seedlings and our quickly growing cell line.

In conclusion, the relative half-lives we obtained are in a comparable range to those in other published studies, where different biochemical and mathematical frameworks were adopted or even different species. A systematic normalization that includes the growth rate is needed to draw reliable comparisons between studies. Without this, the results from the protein degradation field will remain largely incomparable.
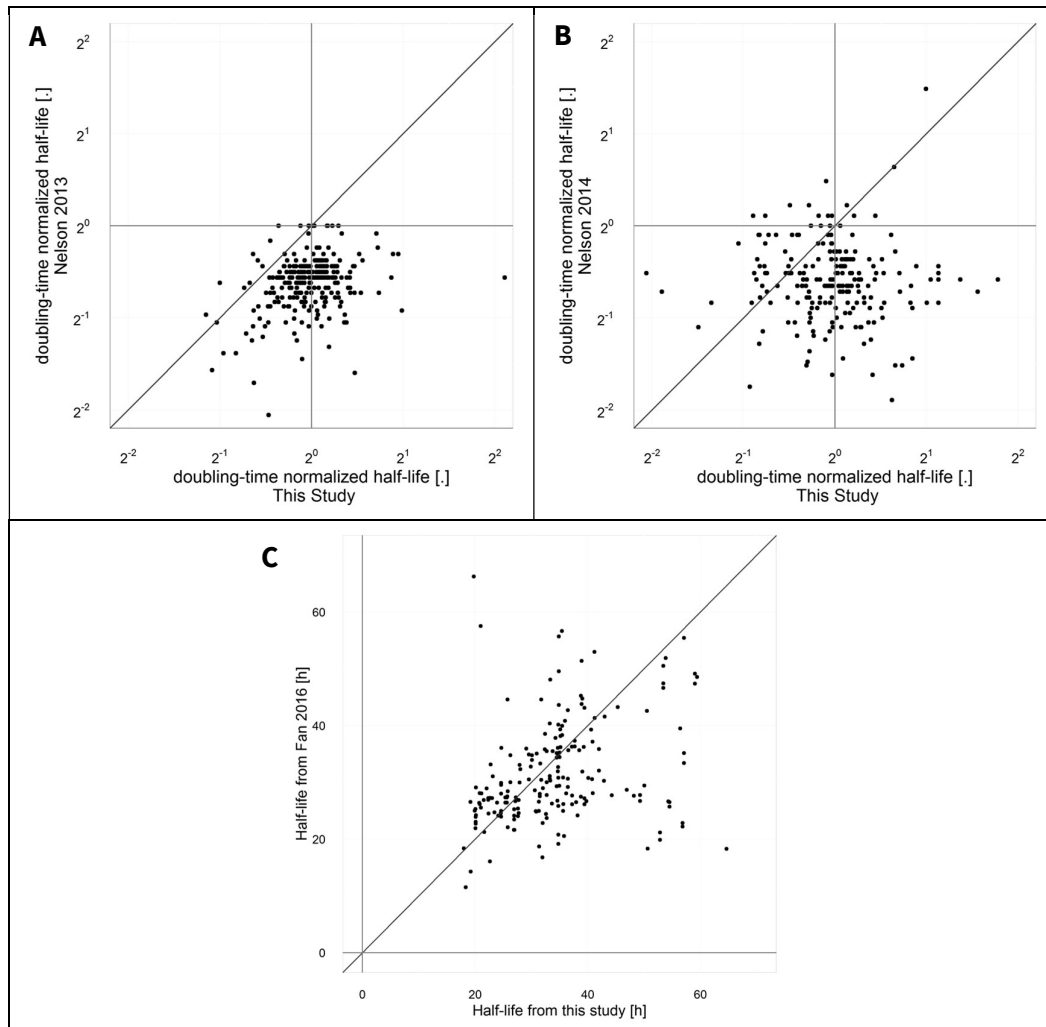
Figure 7.9 Comparison of the determined protein half-lives to those in published data sets. A: Comparison to half-lives of mitochondrial proteins from an Arabidopsis cell culture (Nelson *et al*, 2013). B: Comparison to half-lives of proteins from barley leaves (Nelson *et al*, 2014a). C: Comparison to non-normalized half-lives of Arabidopsis seedlings.

7.9.2   Labelling of plastid-localized proteins

During the comparison to the literature data it became apparent that a subgroup of proteins depicts a peculiar labeling pattern. Proteins encoded in the plastid genome contained a particularly high amounts of unlabeled lysine, in other words, they have a low labeling efficiency (Figure 7.10A). When analyzing individual plastid encoded proteins, a particularly stochastic labeling efficiency becomes apparent (Figure 7.10B-D). In a more systematic analysis, I observed that the plastid-encoded proteins are labeled on average to 85% with endogenous Lys0 (Figure 7.10E). Proteins that are encoded in the nuclear and mitochondrial genomes have in contrast a constant median Lys0 labeling rate of ~15%.

Chloroplasts are the place of lysine biosynthesis (Galili, 1995). It is therefore to be expected that the concentration of endogenous unlabeled lysine is higher in the chloroplasts compared to the cytosol.

This fact was used by Allen *et al*, (2012) to study compartment-specific metabolism by determining the labeling rate of the small, nuclear-encoded and large plastid-encoded subunits of RUBISCO when adding ^{13}C-labelled carbon sources to *Brassica napus* embryos. They determined surprisingly small differences between the different RUBISCO subunits and interpreted this as an indication for a quick exchange rate of amino acids between the two compartments.
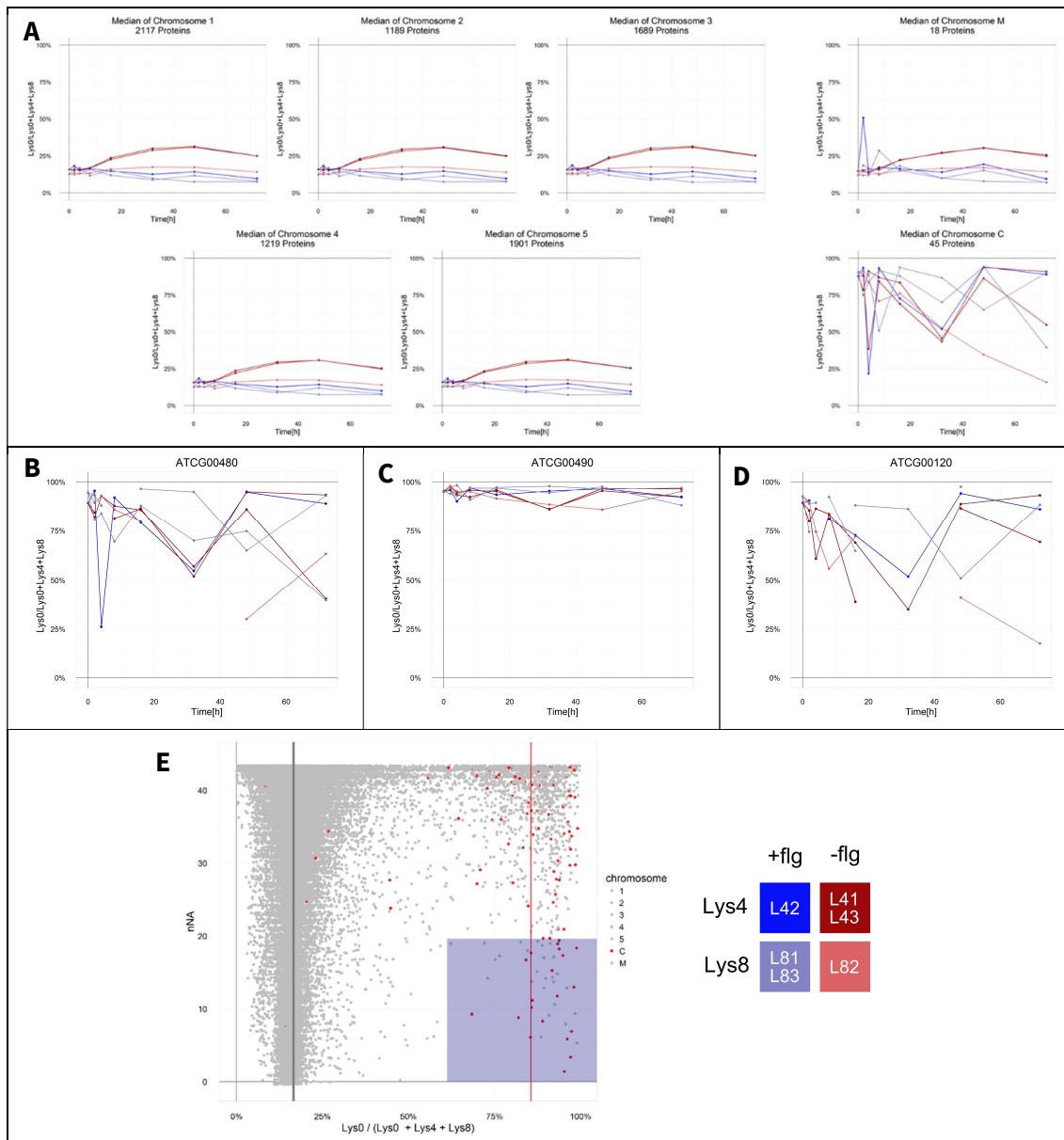
Figure 7.10: Compartment-specific protein labeling efficiencies
A: Chromosome-specific median Lys0 labeling rate over the course of the turnover experiment. B-D: Lys0 labeling rate of exemplary chloroplast-encoded proteins. E: Lys0 labeling rate of all proteins detected in the turnover experiment in any sample. As a measure for the certainty of the quantification, the number of samples where the protein was not quantified is listed on the y-axis (nNA = number of "not assessed" values. nNA of zero indicates that the protein was identified in all samples.). The more often the protein is quantified, the more certain is the quantification. Plastid-encoded proteins are highlighted in red. The chromosome-specific medians are indicated as

vertical lines. The blue shaded box comprises all proteins listed in Table 7.2 that are identified in more than 20 samples with a labeling efficiency of less than 40%.

Curiously, the gene-specific labeling-efficiency plot also revealed a subset of nuclear-encoded proteins with a high Lys0 labeling rate (Figure 7.10E). Proteins, which were quantified in only few samples (high nNA = at the top of the plot in Figure 7.10E) have a wide spread of labeling efficiencies, indicating quantification errors. When limiting the analysis to proteins that are quantified in more than 20 samples, still 11 nuclear-encoded proteins remained that show extremely low labeling efficiencies of less than 40% (Table 7.2). Each of these proteins plays a central role in chloroplast metabolism, including photosynthesis and Calvin cycle.

When expanding the analysis to all quantified proteins with a low labeling rate of less than 40 % regardless how many times these proteins were quantified, I could determine a strong enrichment for chloroplast-related GO categories (Table 7.3). A list of these 101 low-labelled proteins can be found in the appendix 10.3.

Table 7.2: High-confidence proteins with a low labeling efficiency that are not plastid-encoded.

AGI	GENE DESCRIPTION
AT1G31330	PHOTOSYSTEM I SUBUNIT F, PSAF
AT1G42970	GAPB, GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE B SUBUNIT
AT1G61520	LHCA3, PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 3
AT1G67090	RBCS1A, RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1A
AT2G39730	RCA, RUBISCO ACTIVASE
AT3G08940	LHCB4.2, LIGHT HARVESTING COMPLEX PHOTOSYSTEM II
AT3G27850	RIBOSOMAL PROTEIN L12-C, RPL12-C
AT4G10340	LHCB5, LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5
AT5G01530	LHCB4.1, LIGHT HARVESTING COMPLEX PHOTOSYSTEM II
AT5G26000	ATTGG1, BETA GLUCOSIDASE 38, TGG1, THIOGLUCOSIDE GLUCOHYDROLASE 1
AT5G38420	RBCS2B, RUBISCO SMALL SUBUNIT 2B

Table 7.3: Cellular component GO category enrichment for the 101 proteins with a low labeling efficiency. The enrichment was performed with PANTHER (Mi *et al*, 2016). The reference list contains all quantified proteins.

GO CELLULAR COMPONENT	# IN REFERENCE	# IN QUERY	EXPECTED IN QUERY	FOLD ENRICHMENT	ADJUSTED P-VALUE
chloroplast photosystem II	4	4	0.18	21.88	1.19E-02
photosystem I	15	12	0.69	17.5	2.76E-09
light-harvesting complex	10	7	0.46	15.31	1.54E-04
plastoglobule	32	19	1.46	12.99	4.30E-13
chloroplast thylakoid lumen	20	11	0.91	12.03	9.87E-07
stromule	20	11	0.91	12.03	9.87E-07
apoplast	161	36	7.36	4.89	1.45E-12
extracellular region	374	52	17.1	3.04	1.16E-10
chloroplast membrane	59	11	2.7	4.08	3.10E-02

I could verify this labeling pattern in the protein abundance SILAC dataset (chapter 6). However, the effect was less pronounced there, presumably as the experiment has a more constant average labeling efficiency (Figure 7.11A). In the pilot studies that I have performed to set up the plant SILAC protocol, I also observed an equally low labeling efficiency for chloroplast-encoded and some nuclear-encoded proteins (Figure 7.11B). These nuclear-encoded genes overlap (20 % overlap) with the proteins that showed a low labeling rate in the protein turnover study (Figure 7.11C).

We tested whether this labeling pattern was caused by peptide sequence mismatches or peptide ambiguities. The isobaric amino acids isoleucine and leucine for instance cannot be distinguished in MS, which can lead to potentially erroneous matching of peptides to proteins. We, therefore, limited the analysis to peptides, which map uniquely to a protein even when one amino acid mismatch is allowed (Figure 7.11D). This analysis showed the same labeling pattern, including the low labeling rate for some peptides that map to nuclear-encoded proteins.
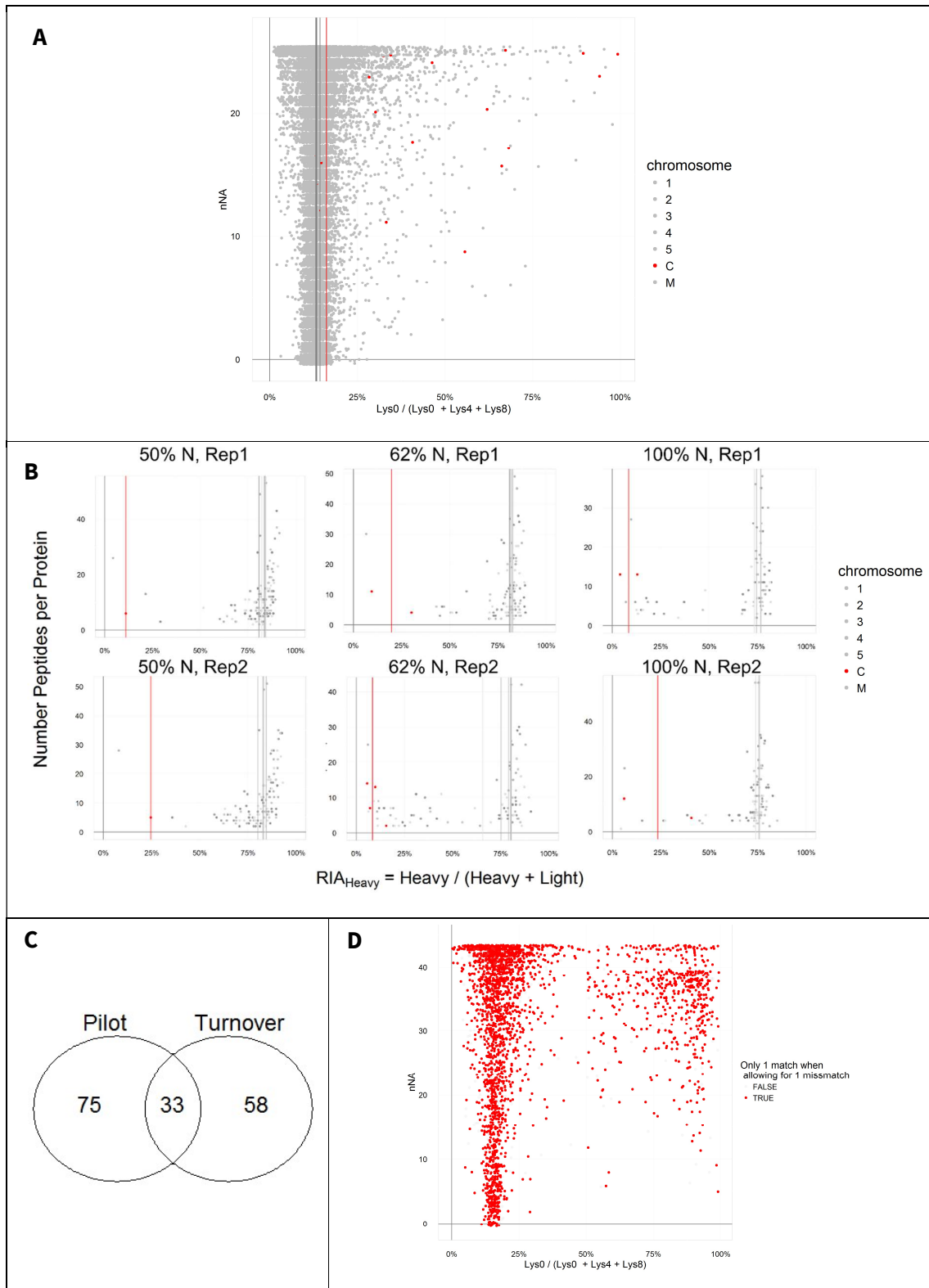
Figure 7.11: Verification of the chloroplast-specific low labeling rate.
A: Labeling efficiencies for each protein detected in the protein abundance data set.
nNA = number of samples in which the protein was not identified. Red dots indicate
plastid-encoded proteins and the median labeling rate for each chromosome is

indicated by vertical lines. B: Protein-specific labeling efficiencies plotted against the number of unique peptides with which the proteins were identified in the pilot experiments. Vertical lines indicate the chromosome-specific median labelling efficiencies. C: Venn plot of nuclear-encoded genes with at low labeling efficiency as determined in the pilot studies of B or the protein turnover data set. D: Labeling efficiency plot of the peptides quantified in the protein turnover data set. Peptides that map uniquely to one protein even when one amino acid mismatch is allowed are colored in red. The analysis is limited to peptides that do not map to plastid-encoded proteins.

A naïve and biologically arguable interpretation of this phenomenon would be that some transcripts of nuclear-encoded genes are somehow translated in the chloroplasts. We therefore analyzed the lengths of the RPFs that map to the transcripts for the nuclear-encoded proteins with a low labeling efficiency. Yet their observed footprint length of ~31 nt is typical for eukaryotic translation of cytoplasmic 80S ribosomes. RPFs mapping to transcripts of plastid-encoded proteins depict a footprint length of only 26 nt, which is typical for transcripts translated by prokaryotic ribosomes.

One RFP length pattern of the low label-efficiency proteins deviates from this analysis: The RFPs of the ATP synthase AT2G07698 have an average footprint length of 26 nt. When comparing the coding sequence of AT2G07698 to the Arabidopsis genome, we identified a next-to-perfect copy of the gene in the mitochondrial genome (ATMG01200, 98% DNA identity). It is therefore more likely that the prokaryotic pattern of the ribosomal footprints derives from the mitochondrial copy than from the nuclear-encoded version.
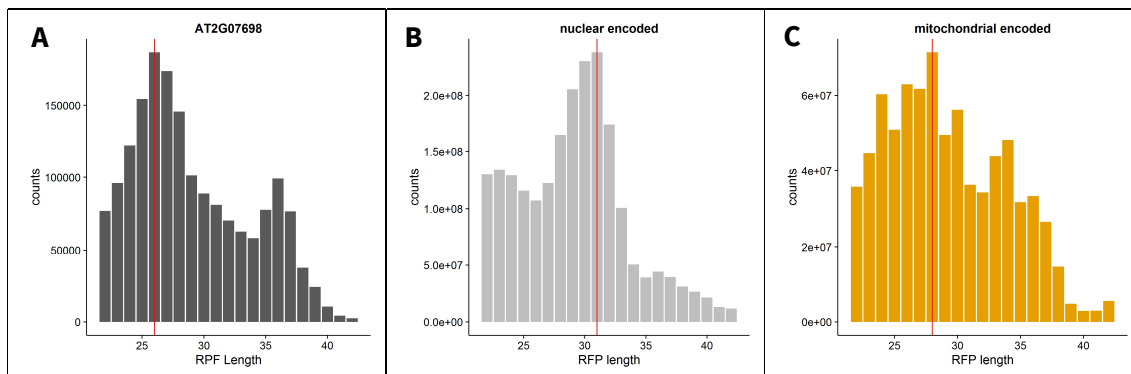
Figure 7.12. A: Length distribution of the ribosome-protected fragments that map to AT2G07698. For comparison, the footprint length distributions for nuclear- (B) and mitochondrial-encoded (C) genes with the respective footprint length modus indicated by vertical lines.

Zhelyazkova *et al,* (2012) published an RNA-Seq data set from isolated barley chloroplasts. We remapped these plastid reads to the whole barley genome to find out whether some of them might map to nuclear-encoded genes in barley. Yet only the nuclear-encoded cytochrome C oxidase gene MLOC_370 was mapped by a high number of reads, which has a perfect copy on the mitochondrial genome. It is therefore more likely that these reads originate from the mitochondrial copy, which could be co-isolated in the chloroplast enrichment.

In conclusion, the RFP lengths and chloroplast RNA-Seq analyses give no indication that the low-labeling proteins would be translated in the chloroplast. The reason for the peculiar difference in labeling thus remains unclear. What remains to be stated is that due to the labeling effect, we cannot assess the protein turnover of chloroplast-encoded genes and some chloroplast-associated proteins. Unfortunately, this also entails proteins that would be important to be included in protein turnover studies such as the RUBISCO subunits, the D1 and D2 proteins of photosystem II, or THIC and THI1, which were described to have a particularly high turnover rates (Chatterjee *et al*, 2011; Li *et al*, 2012a; Fitzpatrick & Thore, 2014).

On a side note, when analyzing the plastid-encoded genes I detected one outlier that showed a regular labeling pattern (acetyl-CoA carboxylase carboxyltransferase, ACCD, ATCG00500, Figure 7.13A). The reason for this difference in labeling is equally unclear, as this gene shows a prokaryotic RPF length distribution (Figure 7.13B) indicating plastid translation. Comparing the nucleotide and amino acid sequence of ATCG00500 by BLAST did not reveal any highly similar copy of the gene (Figure 7.13C-D). ATCG00500 is detected reliably in all samples with 3 to 8 unique peptides.
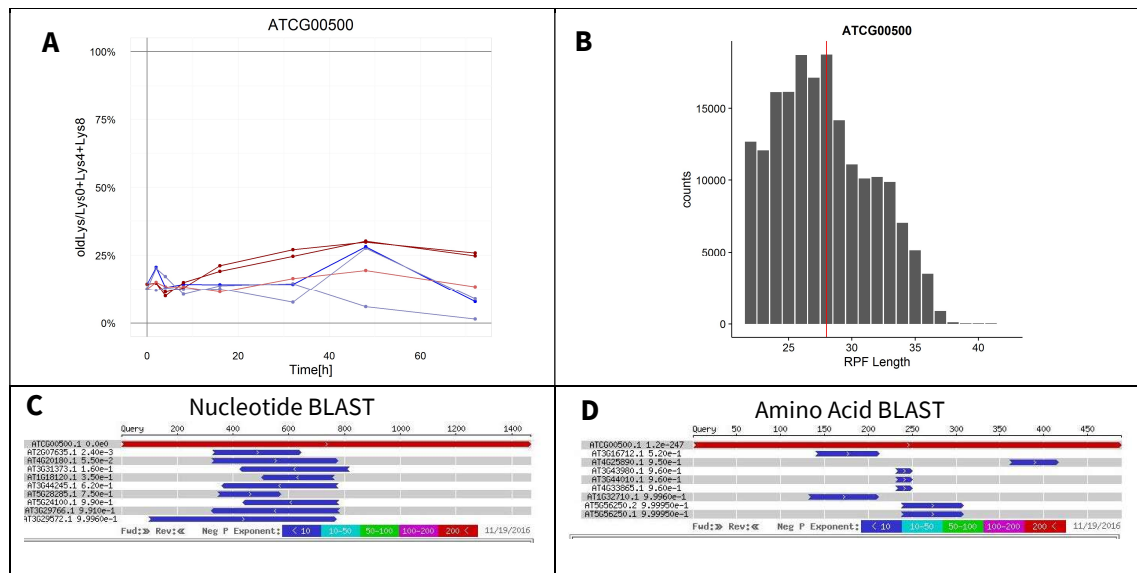
Figure 7.13: ATCG00500, a plastid-encoded gene with a cytosolic Lys0 labeling pattern.
A: Labeling efficiency of ATCG00500. B: RFP length distribution with the modus indicated by a red line. C: Nucleotide BLAST search of ATCG00500 against the Arabidopsis genome. D: Amino acid BLAST search against Arabidopsis proteome.

7.9.3 Example proteins

Nucleoporins and histones are reported to have very long protein half-lives (Toyama & Hetzer, 2013; Martin *et al*, 2012). However, in our dataset, I don't see a significantly different turnover for these protein classes (nucleoporins p-value = 0.687, histones p-value = 0.789, Wilcoxon rank sum test). The only protein annotated in TAIR to the nuclear pore complex that shows a distinctly high half-life is Sucrose synthase 1 (ATSUS1, AT5G20830, relative half-life = 1.522), whose role is clearly not limited to the nuclear pore.
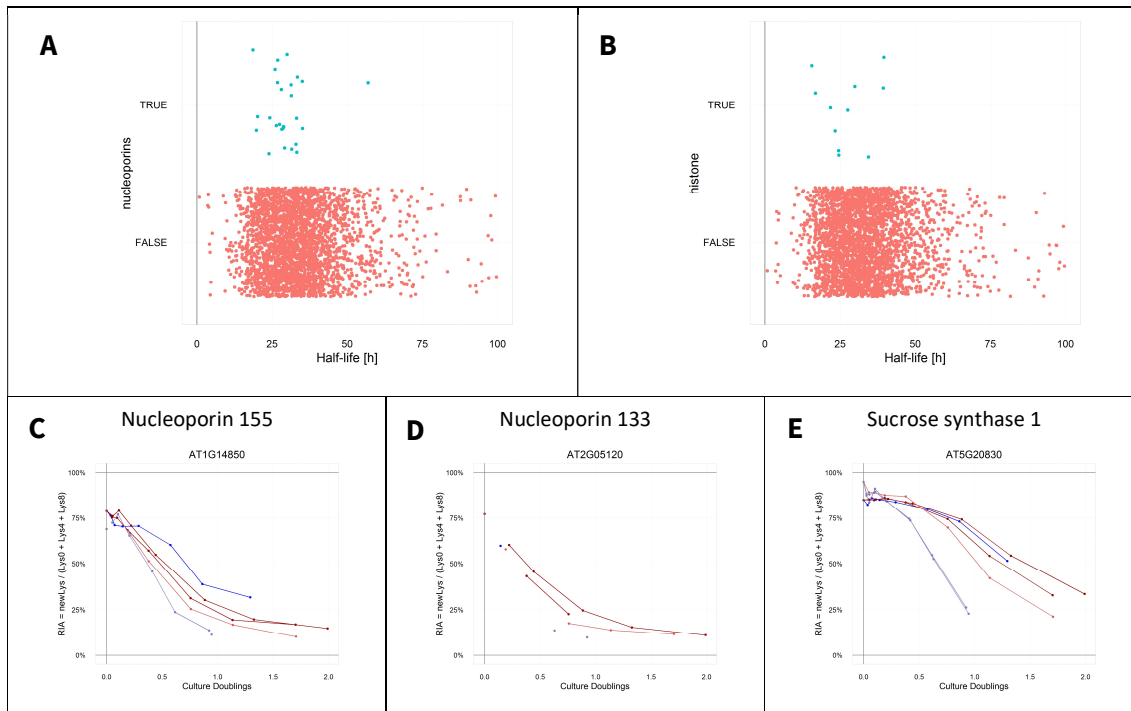
Figure 7.14: Protein half-lives of nucleoporins (A) and histones (B). C-E: Protein turnover plots of example proteins associated with the nuclear pore.

Several cases show a high turnover rate where the protein and/or transcript is reported to follow a circadian rhythm:

Li et al, (2012) determined that glycine-rich RNA-binding family proteins (GRP) depict a particularly high turnover rate. In our data set we can identify three GRPs. GRP2 (AT4G13850, relative half-life = 0.51440, Figure 7.15A) and GRP7 (AT2G21660, relative half-life = 0.4909, Figure 7.15B) have relatively short half-lives, while GRP8 (AT4G39260, relative half-life = 0.8250, Figure 7.15C) shows a regular turnover. GRP7 is the most studied GRP as it is associated with the regulation of circadian oscillations (Staiger *et al*, 2003) and flowering time (Li *et al*, 2012a).

D-myo-inositol-3-phosphatase synthase 1 (ATIPS1, AT4G39800, Figure 7.15D) that shows diurnal protein level fluctuations (Sira Echevarría-Zomeño, ETH Zurich, personal communication) also has a short relative half-life of 0.5431.

Nitrate reductases are key enzymes in nitrogen homeostasis. These proteins catalyze also a side reaction with nitrite to form toxic nitric oxide (NO), whose scavenging is limited in the dark by the availablilty of reduced ferredoxin (Rockel *et al*, 2002). The abundance of these proteins is therefore

down-regulated throughout the day so that the proteins are not active during the dark (Piques *et al*, 2009). In our data set, Nitrate Reductase 1 (AT1G77760, Figure 7.15E) and Nitrate Reductase 2 (AT1G37130, Figure 7.15F) show a high relative turnover of 0.4516, and 0.5516, respectively.

Retinoblastoma-related 1 (RBR1, AT3G12280, Figure 7.15G), a key regulator of the cell cycle (Gutzat *et al*, 2012), also shows a shortened relative half-life of 0.4961.
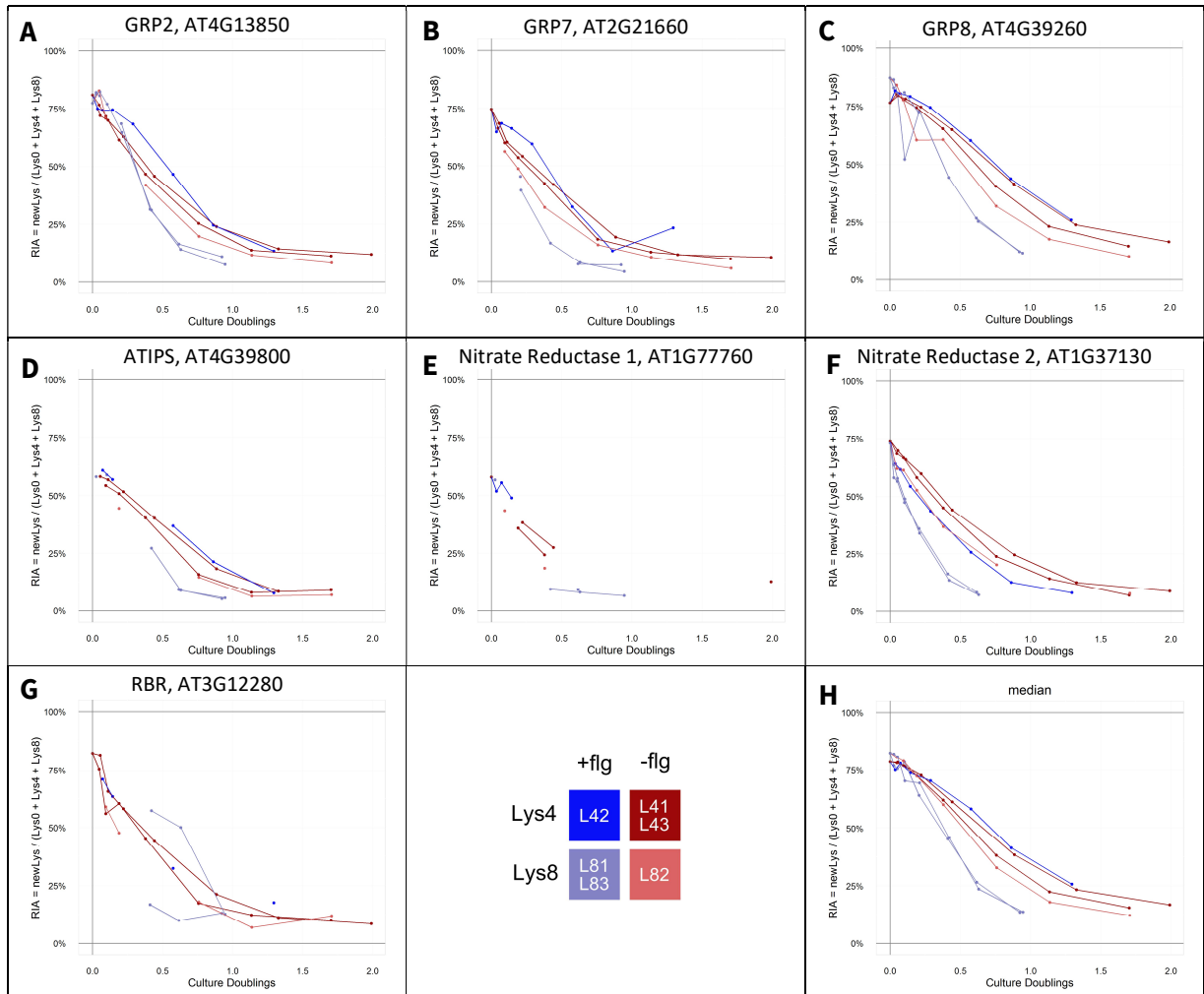
Figure 7.15: Protein turnover plots of proteins associated with circadian regulation (A-D), nitrate metabolism (E&F) and cell cycle regulation (G). As a reference, the median turnover is plotted in the bottom right (H).

7.10 Effect of flg22-treatment on protein turnover

In the next step, I aimed to determine gene-specific effects on protein turnover upon flg22 treatment. In chapter 6, I have observed that flg22 induces substantial changes in the proteome composition of the treated cultures. An increase in protein abundance leads to an increase of newLys–labeled protein in the turnover dataset. RIA values are determined in relation to the other two labeled states so that an increase in RIA_{newLys} leads to a proportional decrease of RIA_{oldLys}, which I used to determine the protein half-lives. Consequently, changes in protein abundance should be reflected in the here determined protein half-lives.

This relationship can be exemplified for Cinnamate-4-hydroxylase (C4H, AT2G30490). This protein is among the proteins that increase the most in response to flg22 treatment (Figure 7.16A). It also shows a drastic reduction in RIA_{oldLys} when treated with flg22. Consequently, when assessing statistically the effect of flg22 on our protein turnover data, genes with increased protein abundance such as C4H should be flagged as significantly changing.

I compared the doubling-time-normalized relative half-lives of all proteins with an ANOVA on the flg22-treatment that includes the labeling background as a blocking factor. I found 149 proteins with a differential relative half-life at a p-value cutoff of 0.05. The positive control C4H was determined to be significantly different at a p-value of 0.0054. As a negative control, I used the median relative half-lives, assuming that the majority of proteins remain unaffected by the flg22-treatment. The ANOVA computes a non-significant p-value of 0.8903 for this control.
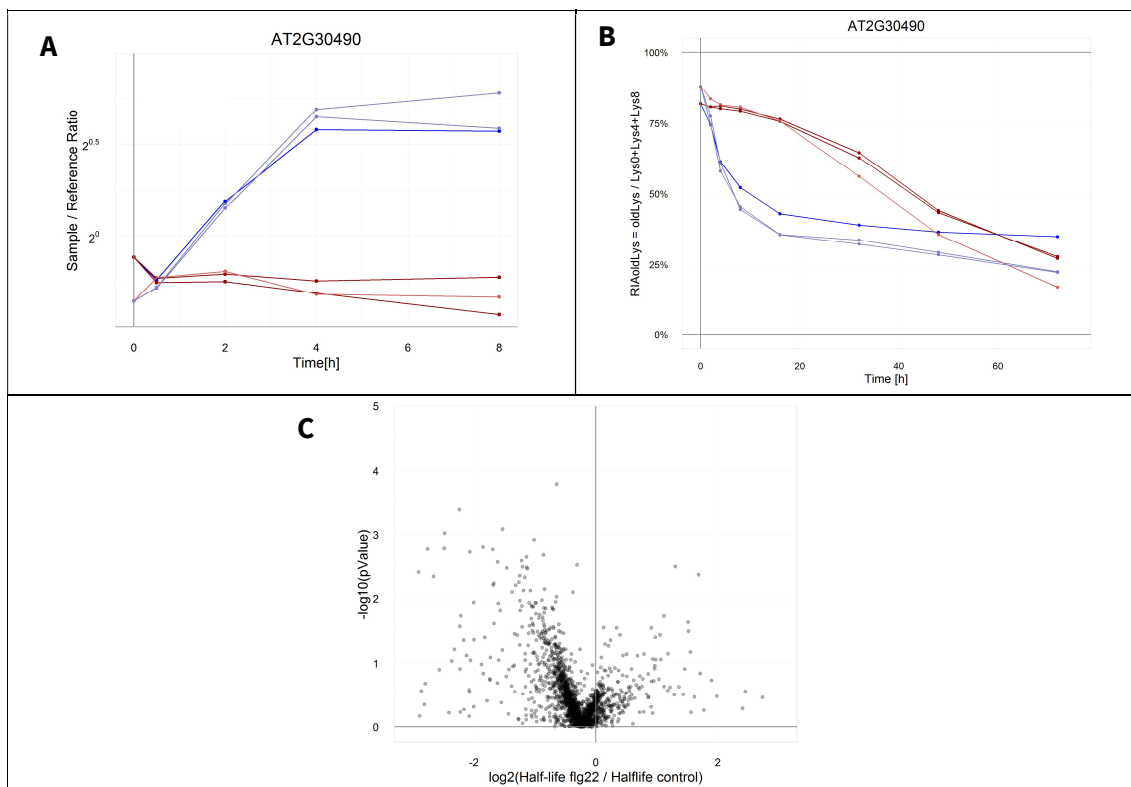
Figure 7.16: Testing for flg22-effect on the relative protein half-lives with an ANOVA.
A: Protein abundances of C4H (AT2G30490), which is used as a positive control for flg22-

induced changes in the turnover rate. B: Protein degradation curves of C4H (AT2G30490). C: Volcano plot of the ANOVA test for the effect of flg22-treatement on the relative protein half-lives.

I conclude that the ANOVA can assess the effect of flg22-treatement correctly, however, by taking the computed half-lives instead of the RIA_{oldLys} values statistical power is lost, leading to reduced sensitivity.

I therefore designed a statistical test within the limma framework (Smith, 2005) for the effect of flg22-treatment based on the RIA_{oldLys} values. First, I tested the effect of flg22 treatment on RIA_{oldLys} over all sampling time points. The lysine labeling background was included as a blocking factor.

$$RIA_{oldLys} \sim Labeling\ Background\ +\ Sampling\ Time\ +\ flg22\ Treatment$$

This test would assess unrealistically that nearly all proteins are affected by flg22 (Figure 7.17A). The negative control median RIA_{oldLys} values were also determined to be highly significantly regulated by flg22 (p-value = 0.00021).

In a second test, I normalized for the different growth rates of the cultures by dividing the absolute sampling time by the culture-specific doubling time. This results in the number of culture doublings at the time of sampling, which could also be called the average number of cell divisions at that time point.

$$\frac{Sampling\ Time}{Doubling\ Time} = Culture\ Doublings$$

$$RI_{oldLys} \sim Labeling\ Background\ +\ Culture\ Doublings\ +\ flg22\ Treatment$$

This transformation is visualized in the plots of Figure 7.17, where the median RIA_{oldLys} is plotted either against the sampling time (A) or against the culture doublings (B). By including the culture-specific growth rates, the differences between the treated and untreated median RIA_{oldLys} trajectories get smaller, which results in a non-significant assessment of the flg22-treatment (p-value = 0.12401). The positive control C4H is determined to be affected by flg22 with a highly significant p-value of 1.86E-13 (Figure 7.17E). The proteins with higher turnover are enriched for GO categories associated with bacterial defense reactions (Table 7.4).

Manual inspection of the two proteins determined to have a slowed degradation upon flg22 treatment showed that these proteins were only determined in few samples and have scattered RIA_{oldLys} trajectories. This is reflected in relatively high p-values (2.7E-4 for AT1G09310 and 2.3E-4 for AT2G24270) compared to the proteins with increased turnover that reach p-values of up to 1E-16 (Figure 7.17D).

I conclude that normalizing for the different growth rates is essential for testing the effect of flg22 on protein turnover. But as mentioned before, changes in protein abundance also have an effect on the protein turnover assessments. In the next step, protein turnover changes that are caused by protein abundance changes have to be differentiated from the ones caused by changes in protein degradation rates.
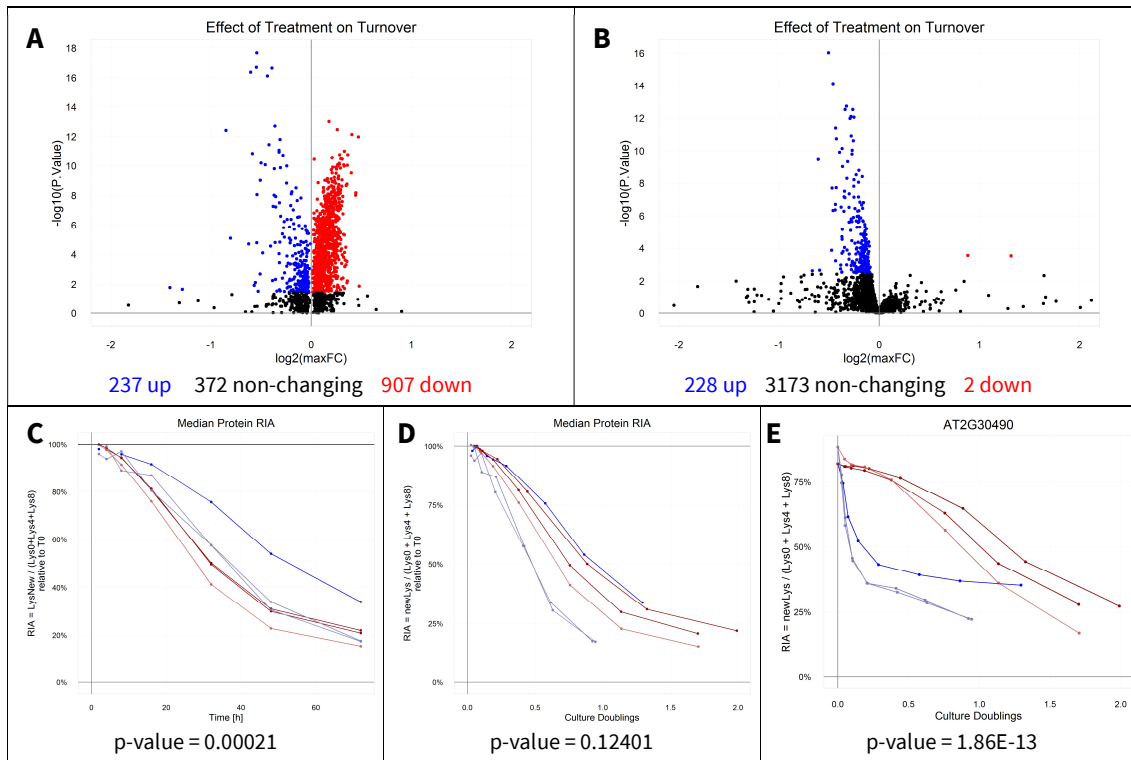
Figure 7.17: Statistical tests on the effect of flg22-treatment on protein turnover.
A: Volcano plot of the first limma test based on the actual sampling times. The number of proteins that are significantly affected by flg22-treatment at an FDR-adjusted p-value threshold of 0.05 are given underneath the plots. B: Volcano plot of the second limma test based on the number of culture doublings. C: Median relative RIA_{oldLys} plotted against the sampling time. The p-value for the effect of flg22-treatment is given underneath the plot. D: Median relative RIA_{oldLys} plotted against the number of culture doublings. E: Relative RIA_{oldLys} values of C4H plotted against the number of culture doublings.

Table 7.4: GO category enrichment for the 228 proteins with increased turnover upon flg22-treatment.

The enrichment was performed with Panther (Mi *et al*, 2016) using the Bonferroni correction for multiple testing and an adjusted p-value cutoff of 0.05. Here only the daughter nodes of a hierarchy are listed. The 3403 proteins for which the limma test was performed were used as reference list.

GO BIOLOGICAL PROCESS	PROTEINS IN REFERENCE	PROTEINS IN QUERY	EXPECTED IN QUERY	FOLD ENRICHMENT	ADJUSTED P-VALUE
flavonoid glucuronidation	9	6	0.39	15.3	2.89E-03
polysaccharide localization	8	5	0.35	14.34	2.65E-02
tryptophan biosynthetic process	9	5	0.39	12.75	4.60E-02
jasmonic acid biosynthetic process	12	6	0.52	11.47	1.46E-02
flavonoid biosynthetic process	14	6	0.61	9.83	3.42E-02
secondary metabolite biosynthetic process	23	8	1	7.98	7.85E-03
glutathione metabolic process	26	8	1.13	7.06	1.88E-02
response to wounding	35	10	1.53	6.56	3.30E-03
sulfur compound biosynthetic process	37	9	1.61	5.58	3.55E-02
defense response to fungus	37	9	1.61	5.58	3.55E-02
cellular amino acid biosynthetic process	84	15	3.66	4.1	4.11E-03
defense response to bacterium	80	13	3.49	3.73	4.75E-02
response to bacterium	95	16	4.14	3.86	4.06E-03
response to cadmium ion	171	22	7.45	2.95	4.85E-03
cellular protein metabolic process	457	4	19.92	< 0.2	6.05E-03
nucleobase-containing compound metabolic process	414	1	18.04	< 0.2	8.29E-05

7.11 Effect of flg22-treatment on protein degradation

As discussed in chapter 7.10, changes in protein abundances affect the turnover estimations. An increase in protein synthesis leads to a reduction of the proportional RIA_{oldLys} value, regardless of the actual degradation rate of the protein. Consequently, there is a high overlap of genes that were assessed as differentially changing at the protein level and in protein turnover (Figure 7.18A). Yet, when limiting the analysis to the 153 proteins that change in their turnover yet remain unchanged in protein abundance, I can study the flg22-induced effect on protein degradation.

I analyzed these genes manually to verify that the two statistical tests describe a reliable effect over multiple timepoints and not just a spurious change. Of the 153 assessed proteins, I determined 120 proteins to show a clear increase in protein degradation. Upon flg22-stimulation, these proteins show a faster turnover, constant protein levels and for the most cases an increase in RNA-Seq and Ribo-Seq counts. Only 8 cases showed non-significant changes in the transcription and translation data. 33 cases depict a brief, burst-like response where the degradation rate changes quickly after the stimulus and seemingly returns afterwards to a degradation rate comparable to that in untreated control samples. The whole list of proteins is found in the appendix (chapter 10.2) and representative proteins are shown in Figure 7.18B-G.

The list of proteins with increased degradation rates contains a diverse list of enzymes. According to SUBA3, these proteins can be found in all subcellular compartments and 13 proteins are secreted (Figure 7.18B). The proteins are enriched for several GO terms associated with bacterial defense mechanisms (Table 7.5, Figure 7.19), indicating that regulation of protein degradation is a biologically important part of the flg22-response.

Table 7.5: GO category enrichment of the 120 proteins with increased degradation upon flg22-stimulation. The enrichment was performed with PANTHER (Mi *et al*, 2016) and all proteins detected in the turnover study were used as the reference list. Parent nodes of the GO term hierarchies were excluded.

GO BIOLOGICAL PROCESS	PROTEINS IN REFERENCE	PROTEINS IN QUERY	EXPECTED IN QUERY	FOLD ENRICHMENT	ADJ. P-VALUE
flavonoid glucuronidation	15	5	0.34	14.8	3.57E-02
jasmonic acid biosynthetic process	19	6	0.43	14.02	7.42E-03
regulation of biological quality	258	18	5.81	3.1	3.17E-02
oxidation-reduction process	541	35	12.18	2.87	1.50E-05
response to chemical	784	38	17.66	2.15	4.94E-03
response to stress	921	41	20.74	1.98	1.49E-02
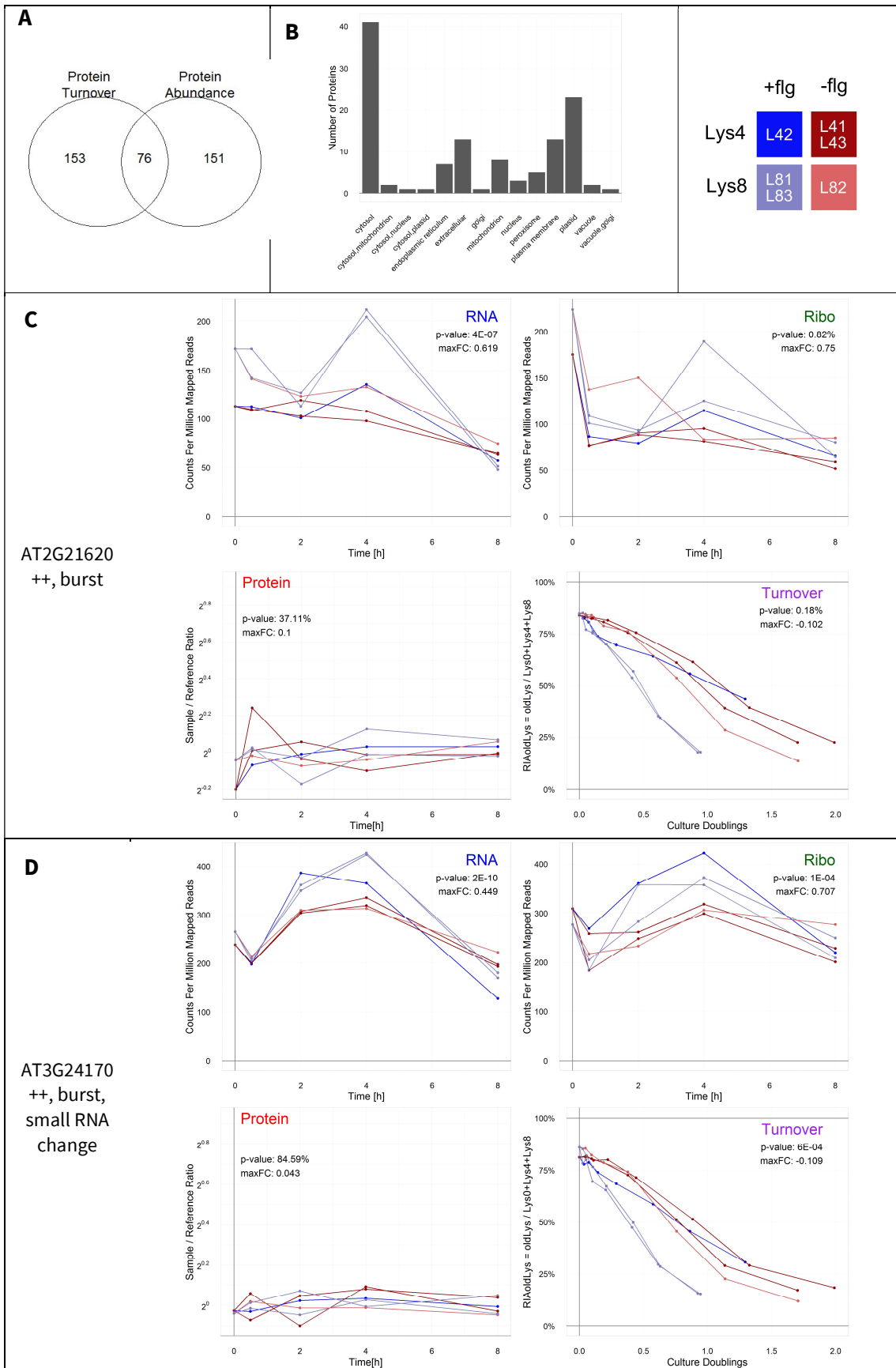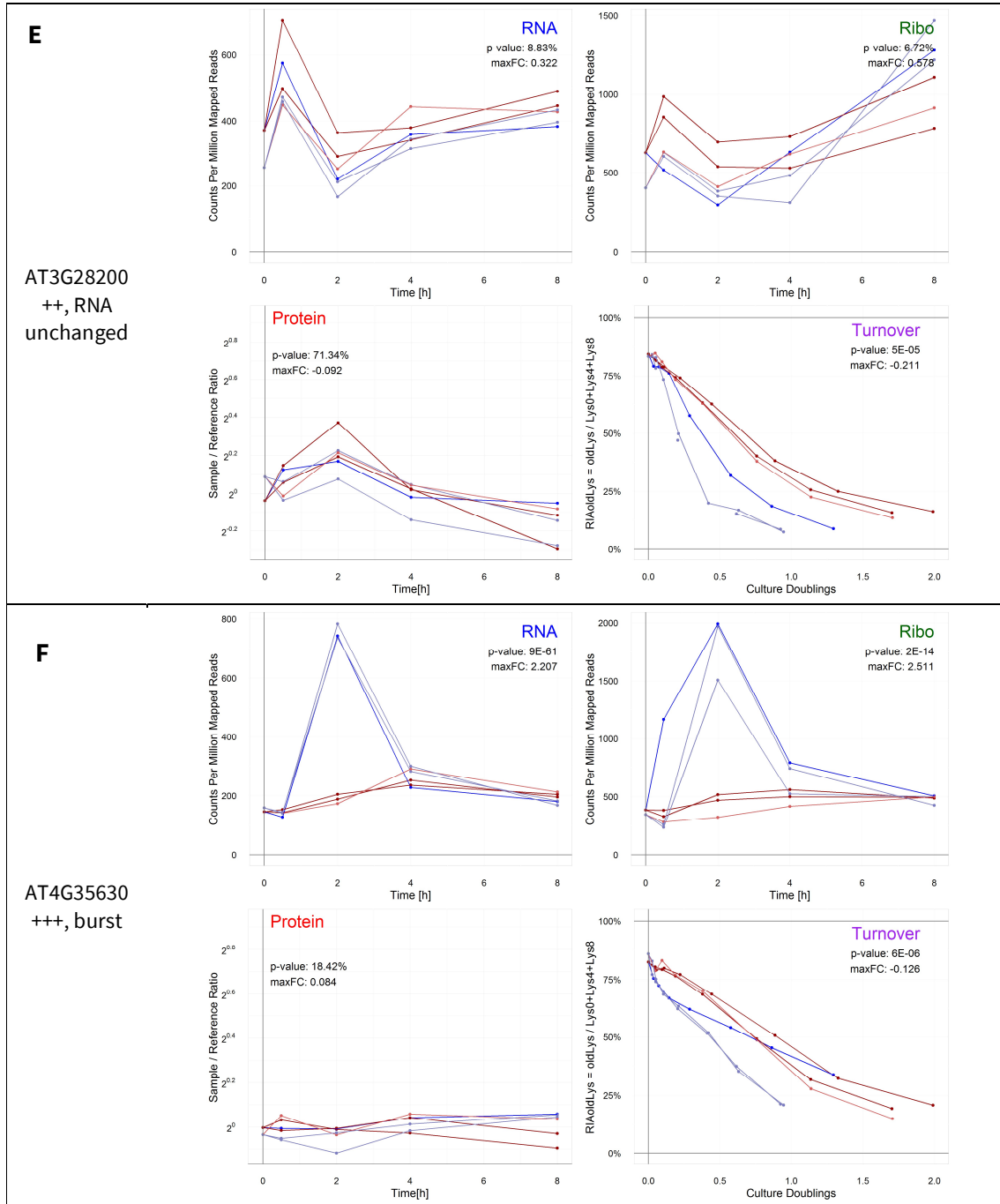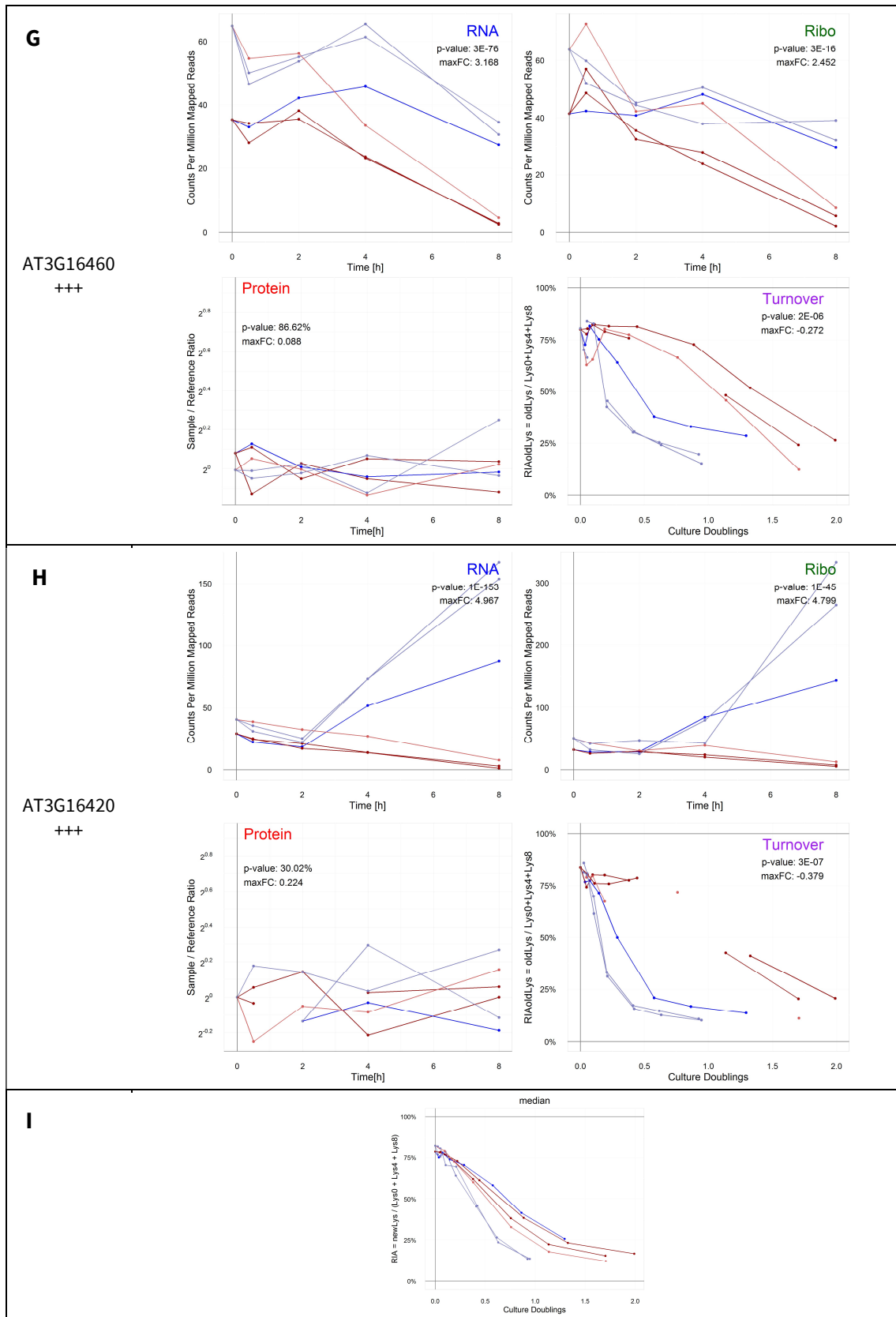
Figure 7.18: Examples of genes with increased protein degradation upon flg22-stimulation.

A: Venn diagram of the proteins that change significantly in protein abundance or protein turnover upon flg22-treatment. B: Subcellular localization according to SUBA3 consensus. C-H: Examples of genes with increased protein degradation upon flg22-stimulation. The RNA-Seq, Ribo-Seq, protein abundance and protein turnover assessments are presented for each protein together with the p-values and fold change assessments for a flg22-induced effect in each individual dataset. The manual classifications of the degradation regulation size are given above each plot. The number of + indicate the size of the degradation regulation. I: As a reference, the protein turnover of the median RIA_{oldLys} values. The color scale at the top right is representative for all plots.
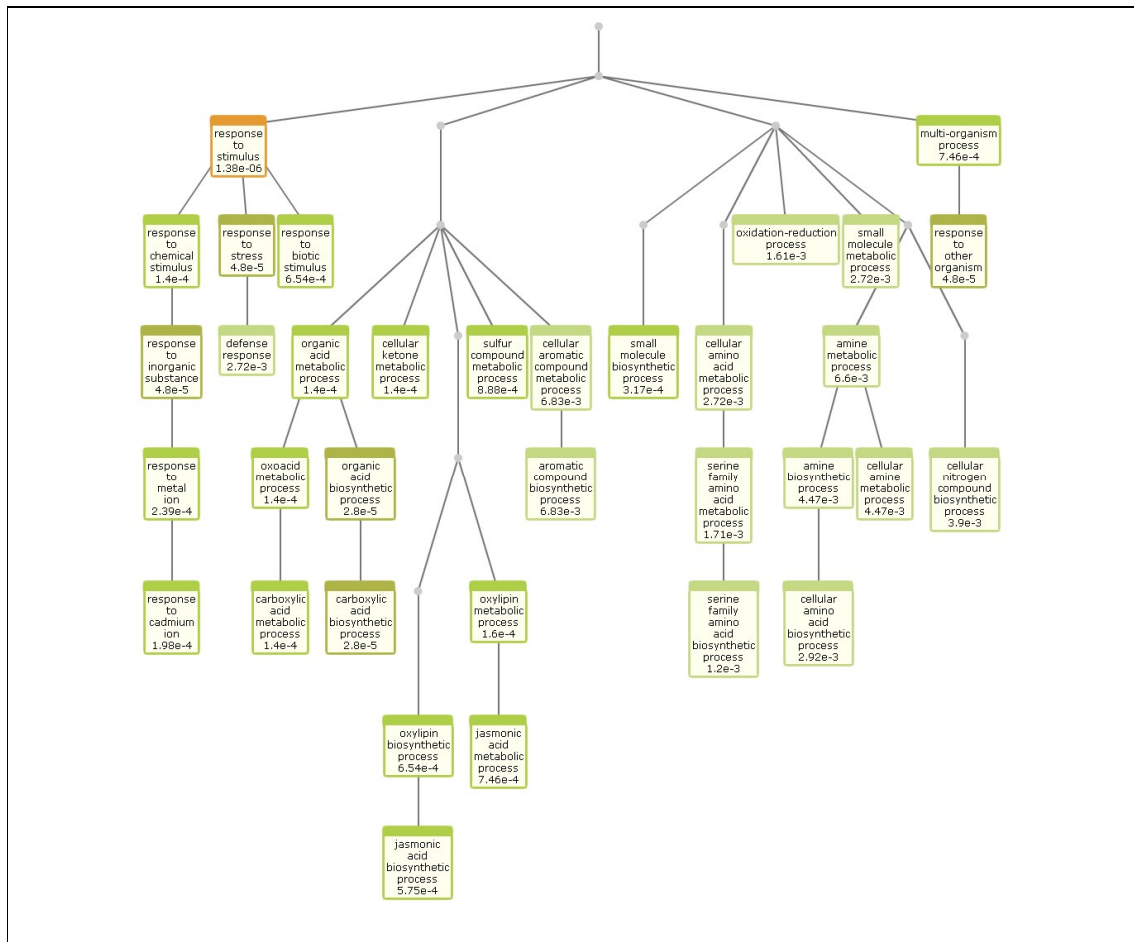
Figure 7.19: Visualization of the GO enrichment of the proteins with increased degradation upon flg22-stimulation by Virtualplant (Katari *et al*, 2010).
The nodes are colored according to the enrichment p-value (orange = highly significant p-value, green = less significant p-value).

8. DISCUSSION

8.1 Gene expression regulation

The aim of this study was to assess in a systems biology approach the relative contributions of transcription, translation and protein degradation on the global regulation of protein abundances under the influence of an external stimulus. We quantified four genome-wide gene expression datasets (transcription, translation, protein abundance and protein turnover) in cultured Arabidopsis suspension cells with or without elicitation of defense response through treatment with flg22. Flg22 is a short peptide derived from the amino acid sequence of the bacterial flagellum and it was chosen as the stimulant because it induces a strong PAMP-triggered immune response on multiple physiological levels (Felix *et al*, 1999).

In summary, I found for genes whose expression increases by flg22 stimulation that 1) the strongest changes after flg22 stimulation are detected at the transcriptional level, 2) transcriptional changes are passed on to the translational level without major regulation and 3) in cases where protein levels do not follow the transcriptional and translational upregulation, protein degradation rates increase, counteracting the increased protein synthesis.

I did not detect clear examples where these gene expression levels did anti-correlate, which would require more complex gene expression regulation mechanisms. Visual inspection of the cases where the observed changes were not significant in all data sets revealed the same trend in the different datasets, indicating that the non-significance was merely an issue of the employed significance thresholds. Many of such non-corresponding cases could also be explained by spurious correlations in one dataset at one time point caused by measurement noise. This highlights the importance of including multiple biological replicates and adopting a time-resolved experimental design. With that wealth of data, spurious cases that would otherwise be deemed significant, could be assessed accurately.

When flg22-treatment induces transcriptional downregulation of a gene, the interpretation gets less clear. Downregulation does not seem to have such a strong effect at transcriptional and translational level, as the fold-changes for downregulated genes are much smaller than for upregulated genes. We could also not identify genes with strongly and quickly reduced protein

abundances, which would be indicative for substantially increased protein degradation. It seems that the primary downregulation response is triggered by a transient decrease in transcription, which slowly affects the protein abundances depending on the individual protein turnover rates. That means that proteins with a slow turnover need a long time to respond to a transcriptional downregulation. A slow protein turnover buffers against quick and brief stimuli. This mechanism could primarily be detected for constitutively expressed proteins with high abundances, while transcription factors and stress response genes are expected to show a more burst-like protein abundance behavior and a more strongly regulated protein degradation. Unfortunately, the nature of the MS-based protein quantitation makes the assessment of these low abundant proteins difficult. Assessing the protein degradation regulation of these low abundant proteins would be very interesting. This could be analyzed for instance by using a click-chemistry-based enrichment of newly synthesized proteins using AHA-methionine supplementation briefly after the stimulus (Eichelbaum *et al*, 2012).

The relative contributions of transcription, translation and protein degradation on gene expression regulation also depend on the type and timing of the stimulus. It will, therefore, be interesting for future studies to assess whether the assessments made here can be transferred to other stimuli including hormonal treatment or exposure to biotic or abiotic stressors.

These assessments so far are based on my personal interpretation of the data. A systemic mathematical assessment of the impact of transcription, translation and protein degradation on the protein levels is currently being conducted.

8.2   Comparison to other reported studies

Only few studies have systematically quantified the gene expression cascade on the four levels discussed here. The following studies are noteworthy to discuss in this context:

1. Schwanhäusser *et al*, (2011) determined the absolute abundances and turnover rates of mRNAs and proteins in mouse fibroblasts without further stimulation. They conclude that variation in transcription can explain up to 40% of the protein level variation and that protein levels are equally strongly affected by translational regulation, while regulation of protein degradation has only a minor effect.
2. In response to Schwanhäusser *et al*, (2011), Li *et al*, (2014) criticized the employed MS-based absolute quantifications to estimate the impact of translational regulation. When comparing the MS-based estimation of the translation rate with more immediately determined Ribo-Seq data, only 12% of the protein level variation can be explained by translational regulation. Li et al, (2014) thus conclude that the effect of transcription was underestimated in the study of Schwanhäusser et al, (2011) and can rather explain ~ 81 % of the protein level variation.
3. Maier *et al*, (2011) studied transcription, protein abundance, and protein turnover of the genome-reduced (690 ORFs) bacterium *Mycoplasma pneumoniae*. They determined absolute levels of mRNAs and proteins and found that the dynamics of mRNAs and of their corresponding proteins are largely decoupled and that abundance ratios of mRNAs and proteins are affected by regulation of translation rather than protein degradation. Yet, how well a genome-reduced bacterium can be compared to higher eukaryotes is debatable.

4. Battle *et al*, (2014) obtained and correlated expression quantitative trait loci (eQTL) across transcription, translation and protein abundance datasets in 62 mammalian cell lines. They saw a strong association of the eQTL to transcriptional expression and by an equal extent to translation. The eQTLs exhibit attenuated effects on protein abundance, indicating a buffering mechanism. They did not assess protein degradation, which is the main remaining regulatory mechanism, which could explain the reduced effect size of the protein abundances on the eQTLs.

5. Jovanovic *et al*, (2015) analyzed primary mouse dendritic cells treated with lipopolysaccharides (LPS). They determined protein synthesis and degradation with a multiplexed SILAC assay and determined the transcriptional response by RNA-Seq. It is very similar to our experimental design, as it also compares the response of a biological system to a stimulus, coincidentally also a bacterial stimulus that elicits a defence response. They identified 63 proteins with increased abundance and 4 with decreased abundance – a similar number compared to our dataset. They obtained each data set in two replicates and determined that transcriptional regulation contributes by far the most to protein level regulation (66 %), followed by translational regulation (26 %). Protein degradation regulation was determined to have only a marginal effect on protein level variation (8 %). Yet their translation rates are also determined from MS data and could therefore equally overestimate the effect of translational regulation analogously to the discussion by Li *et al*, (2014) of the data in Schwanhäusser *et al*, (2011). In support of this assumption, Jovanovic et al, (2015) report that their MS-based translation rate estimations correlate only marginally well with the translation rates obtained from Ribo-Seq data (Pearson correlation = 0.5).

Compared to these landmark studies, our study explicitly quantifies all four levels of gene expression. Yet our data set has issues that arise from using a stimulus that affects the growth rate and from using a plant-based system, which complicates the SILAC labeling efficiency and the Ribo-Seq protocols. Still, with the presented normalizations the data allow us to draw biologically meaningful conclusions.

My finding that gene expression is largely determined by transcription corroborates the assessment of Li et al, (2014) and in part of Jovanovic et al, (2015). In contrast to the latter, I determined the effect of translational regulation to be of lower significance. As mentioned, they did not determine the translational response explicitly but inferred it from MS-determined protein turnover data. As these are more variable compared to sequencing-based quantifications such as Ribo-Seq, it can be that the regulatory effect of translational regulation is overestimated. In our data set, we cannot detect clear regulatory effects on the level of translation but rather on the level of protein degradation.

In conclusion, based on the data presented here, I found that the primary causal agent for protein level regulation is transcriptional change. On the second level, protein degradation seems to be a crucial regulatory mechanism, especially for proteins with constant abundance levels despite of increasing transcript levels. Translational regulation has only a minor effect on gene expression regulation and is limited to few individual cases. Most importantly, though, this assessment has to be backed up by mathematical analyses that are currently being performed in collaboration with the group of Prof. Jörg Stelling (ETH Zürich).

8.3 Experimental design

In the following, I will discuss in more detail the benefits and challenges of the experimental design and the presented data sets. With this, I aim to extend the interpretation of the datasets, especially with regard to their limitations, and to support future research in building upon the employed experimental approaches.

One challenge with the employed experimental design is the different time scale. The nature of the biological response demands a shorter timeframe for transcription and translation data sets than for the protein abundance data set. The protein turnover analysis requires an even longer timeframe to produce a large enough label change. Analyzing the response to a stimulus at different levels in one experiment will inevitably lead to compromises.

One of the compromises is that the protein turnover assessments span a longer timeframe, which is partially not covered by the protein abundance data set. Consequently, it cannot be excluded that significant protein level changes occur at a later time point, which would affect the protein turnover assessment. Future studies should expand the timeline for the protein abundance analysis.

8.4 Replicates

It was instrumental to have three replicates in all four data sets so that we could assess the technical variation and biological that supported me in identifying biological signal from technical noise.

Following the same line of thought, the Lys4- and Lys8-labeling background did complicate the analysis, especially for the protein abundance data set. Yet, it also introduced biological variation that helped to identify true biological responses to the flg22-treatment. I arrived at the conclusion that label swaps are powerful to hedge against spurious correlations (Kierszniowska *et al*, 2009). The substantial variation in Lys4- and Lys8-labeling trajectories in the protein datasets came as a surprise to us, and in hindsight, a higher number of replicates would have been advisable.

Equally, to have a time-resolved data set enabled me to inspect noisy genes that were passing the statistical tests due to spurious correlations. Manual assessment is not a statistically robust approach, but invaluable to exclude spurious correlations that are problematic in any high-throughput quantification (Larsson *et al*, 2010).

8.5 RNA-Seq

In the following paragraphs, I want to address important points for each individual data set.

Only a few things have to be said about the RNA-Seq data set. It is a robust, reliable, sensitive and accurate methodology, thanks to a high degree of standardization in the workflow. It indicates where the other genome-wide technologies still have to improve. I would like to highlight the importance of rigorous standardized QC measures that verified each step along the data acquisition. In MS analysis, for instance, such measures are still the exception.

The data set contains highly valuable information for modeling approaches to tease apart the kinetics of the signaling cascade. One could identify upstream and downstream targets of the signaling cascade due to the time-sensitive nature of the analysis. A systematic kinetic model of the

data set could improve our understanding of the flg22-transcriptional response network similar to the analysis of the Toll-like receptor response of immune dendritic cells (Chevrier *et al*, 2011).

8.6 Ribo-Seq

Although Ribo-Seq is also based on a sequencing methodology, the data are substantially more complex to obtain due to the many experimental processing steps and the short read lengths. The Ribo-Seq protocol had first to be adjusted for plant samples. The biochemical approaches for plant Ribo-Seq samples differ substantially in their ribosome-mRNA enrichment and RNase digestion procedures.

My strategy of pelleting the mRNA-ribosome complexes by ultracentrifugation represents a more economical approach to obtain RPFs than to start with large amounts of plant material and adding high concentrations of RNases. Yet we cannot assess whether this leads to any differences in the quality of the enrichment or the RNase digestion. Using a smaller rotor for the ultracentrifugation is a minor improvement to reduce the required amount of biological material and time. In conclusion, I hope that my adaptations will help in creating more reliable plant Ribo-Seq data sets in the future.

The analysis of the RPF lengths revealed a non-negligible number of incomplete short RPFs. These could be the reason for the low mapping rate and indicate issues in the library generation. Yet, where exactly to improve the protocol can currently not be determined.

Regardless of this issue, the data set contains enough reads for each sample to describe the translational regulation response induced by flg22. Using several statistical tools, we could not detect cases where the translational efficiency was strongly altered upon flg22 treatment. It was unexpected to see so few changes on the translational level, as flg22-stimulation seems to affect nearly all physiological parameters. Maybe a different stimulus could cause a more pronounced translational response. Hypoxia was described to lead to a global decline in the initiation of translation, yet also there the number of ribosomes per transcript stayed stable across the tested conditions (Juntawong *et al*, 2014).

In contrast, Liu et al. (2013) describe a clear impact of translational control on thousands of genes after exposure of etiolated seedlings to light. As photomorphogenesis is a process that requires reshaping the whole cell including the organelles, translational regulation might be much more pronounced. Clearly, more data on different stimuli will be required to assess the relative contribution of translational control on gene expression.

8.7 Protein abundance

8.7.1 Spike-in SILAC

MS-based proteomics remains a powerful, yet equally challenging technology to wield. We present here a comprehensive data set of thousands of quantified proteins over many time points, where each sample can be compared directly to any other sample. Yet especially the moderate overlap of quantifiable peptides in all the samples raises an important issue. Without reproducible detection of peptides in a complex sample mixture, the quantification will always be more stochastic and therefore variable. Future studies could there therefore focus on the maturing data-independent

acquisition (DIA) method SWATH-MS (Gillet *et al*, 2012), which holds the promise of a more robust identification of peptides over multiple samples (Bruderer *et al*, 2015) and a more robust quantification independently of the employed algorithm (Navarro *et al*, 2016).

SWATH-MS does also not depend on labeling, which would overcome the issue that arose from using two spike-in reference samples. The employed approach to normalize the spike-in samples is not a standard technique but results in the only way to reliably quantify these samples. McShane et al, (2016) developed a similarly complex approach to normalize between triple pulsed SILAC samples, where one label state was used as a common reference. Battle *et al*,(2014) detected 4'000 proteins over all samples in a spike-in SILAC approach when increasing the FDR to an atypically high value of 10 %. It seems to remain a non-trivial issue to compare large amounts of samples by SILAC MS, thus giving motivation to invest in setting up reliable DIA SWATH-MS spectral libraries and assays. A direct mixing of treated and untreated samples from the same time point might have been a more robust approach to quantifying protein abundance changes upon flg22 treatment, yet with such an approach we would not have been able to compare abundance changes between the different time points, which would have impacted successive kinetic modeling approaches.

The adopted spike-in SILAC approach resulted in a protein abundance data set of good quality, yet improvements regarding peptide identification rate could be done in future studies, for instance by more frequent tuning and exchange of the orifice or LC column. Robust use of online (Bittremieux *et al*, 2016) and offline QC tools such as PTXQC (Bielow *et al*, 2016) should help with improving the reliability of MS studies. Furthermore, when using only lysine as the labeling agent, like in the employed plant SILAC approach, the usage of Lys-C as the sole protease should be tested, as it holds great potential of increasing the amount of quantifiable information per MS run because no unlabeled arginine-terminating peptides will be created. This will reduce peptide complexity, which could further improve the identification rate and increase the chance of identifying a particular peptide in all samples.

8.7.2   Plant SILAC

Our results show that SILAC studies in plants are possible yet remain challenging. Balancing labeling efficiency and vitality of cultured cells under nitrogen and lysine stress is a nontrivial task, and probably needs to be established for each culture separately. Our study reveals the importance of keeping the nitrogen supply high enough to not overly stress the cell cultures. Murashige-Skoog medium might be overly saturated in nitrogen, yet reducing it to the amount employed by Schütz et al, (2011) holds the risk of performing the experiment in a substantially stressed culture.

It would be interesting to explore in more detail whether the identified culture viability issues could be resolved by supplementation with methionine, as it was indicated by Green & Phillips (1974). I could not find a coherent explanation for the detrimental effects of lysine supplementation other than the inhibition of aspartate kinase, which potentially leads to a depletion of methionine, threonine, and isoleucine. When I added the three amino acids to lysine-supplemented cultures, I could not detect a clear improvement of the cultures' condition. However, increasing the amount of supplemented methionine to non-physiological levels might show an effect in the more sensitive formazan-based viability assay.

Improving the labeling rate by genetic engineering still represents an interesting approach. The arginine-deficient Chlamydomonas strain CC-424 clearly shows the benefit of such a culture for SILAC studies (Naumann et al, 2005). The generation of an inducible lysine deficient cell culture definitely holds great promise to further facilitate plant SILAC studies. Possible genetic targets are all genes unique to the lysine biosynthesis pathway, including the here tested *DapL*.

Without such a lysine-deficient line, the plant SILAC approach used here can only be recommended, when the timeline of the experiment is so short that the seemingly inevitable fluctuating levels of unlabeled protein do not affect the quantitation substantially. Also, the differences between the Lys4- and Lys8-labelled cultures show that this approach should only be applied with a label-swap included in the experimental design (Kierszniowska *et al*, 2009).

8.8 Protein turnover

The protein turnover data set presented here is, to our knowledge, the most comprehensive plant turnover data set. Global degradation processes, especially in plants, are seldomly studied, presumably due to the challenges regarding plant SILAC (Gruhler *et al*, 2005b; Schütz *et al*, 2011) and dynamic ^{15}N labeling (Nelson *et al*, 2014b, 2014a). Our data set highlights the importance of studying both baseline protein degradation and in response to stimuli. Interestingly, I detected a higher number of proteins with changed protein turnover after flg22 stimulation than with changed protein abundances. Clearly, plant cells respond to stimuli with a complex protein degradation machinery, which would deserve more attention of the research community.

With the current approach, I could only identify cases of increased or decreased protein degradation if the protein levels remain stable. A more sophisticated analysis using mathematical modeling could take protein level changes into consideration and analyze, if the protein turnover only reflects the protein level change or whether it is further regulated by changes in protein degradation. This is clearly an important task to follow up in the future.

8.8.1 Transcriptional and translational upregulation can be counteracted by increased protein degradation

Interesting biological questions emerge when focusing on the subgroup of proteins that respond to a transcriptional increase by raising protein degradation rates, resulting in constant protein levels. To our knowledge, these cases show for the first time a coherent reconciliation of the gene expression regulation model. In most studies not all four levels of the gene expression model were quantified and therefore had to explain discrepancies by referring to the non-measured level (e.g. Battle *et al*, 2014). The phenomenon of constant protein levels despite transcription level changes is described frequently both at genome-scale and individual gene level (e.g. Dutilleul *et al*, 2003; Yang *et al*, 2004; Baerenfaller *et al*, 2012), yet it is not well resolved which post-transcriptional processes ultimately cause this discrepancy. Our data indicate that – at least for the flg22-induced PTI response - regulation by protein degradation plays an important role and the finding that the effect size of the transcriptional increase is coherent with the effect size of the protein degradation validates this assessment. Equally so for the observed burst-like changes in the turnover trajectories, which could indicate a stimulus-induced increase in degradation rate and a subsequent return to a non-stimulated degradation rate after some time.

Seeing the high prevalence of targeted protein degradation processes upon stimulation brings up the question why cells invest in increased protein synthesis and degradation when the protein abundance finally remains constant. I could think of the following possible explanations:

1. Protein complex homeostasis.

   It has been proposed that proteins become stabilized when they are integrated into protein complexes and that protein complex subunits produced in excess are rapidly degraded (Goldberg, 2003; McShane *et al*, 2016). This could also be the case for the proteins discussed here. Increased synthesis of protein complex subunits could lead to increased degradation if the synthesis of the protein complex partners does not increase equally. This would explain how this phenomenon comes to be, but not what the benefit for the cell might be.

2. Protein rejuvenation.

   An increase in protein synthesis and a concurrent increase in protein degradation results in a higher turnover of the protein, reducing the average molecular age of that particular protein. Such a rejuvenation could have two benefits for the cell. First, assuming that the metabolic activity of a protein decreases with age e.g. by accumulating oxidation damage or stochastically occurring non-reversible misfolding, a rejuvenation could increase the average metabolic activity for that protein. This might be particularly important as flg22-treatment leads to an oxidative burst (Wojtaszek, 1997), which also damages proteins. Secondly, the rejuvenation of a protein would reset the accumulated post-translational modifications (PTMs). Newly synthesized proteins could be modified differently, resulting in a different proteoform (Smith & Kelleher, 2013), which could modulate the average enzymatic activity of that protein. For signaling proteins such as receptor kinases, such a PTM reset could also result in a return to the non-induced steady-state.

3. Changed protein pool composition by alternative splicing.

   In comparison to prokaryotes or lower eukaryotes, higher eukaryotes do not necessarily contain a higher number of genes, yet the number of different proteoforms resulting from differential splicing and PTMs is substantially higher (Barbosa-Morais *et al*, 2012). The protein abundance analysis described here does not resolve splicing isoforms, however, it could be that the protein pools actually do change in composition due to differential splicing. Resolving such events on the protein level by MS is challenging, mainly due to the fact that peptides that are unique for only one splice variant and readily identifiable with MS are very rare. Yet the RNA-Seq and Ribo-Seq data would hold the information on differentially spliced isoforms and their systematic analysis could reveal interesting findings, especially for the here discussed gene set.

4. Two signal stimulation.

   It might be that the signaling system for some genes did evolve to be more robust and require stronger stimulation to actually change in its protein abundance, potentially requiring a secondary stimulus. Our experimental setting is limited to PTI (PAMP-triggered immunity) signaling, the first part of plant defence mechanism (Li *et al*, 2016). The second part of the defence mechanism is initiated by a bacteria-secreted, intracellularly-detected effector molecule, called ETI (effector-triggered immunity). It could be that the here described proteins are part of that secondary ETI response and are initiated by PTI signalling in anticipation of that ETI signal. Upon that second stimulus, the targeted protein degradation could quickly be stopped so that the more slow reacting protein synthesis machinery is in place to quickly raise

the protein abundance. This would describe a form of plant defence priming (Martinez-Medina *et al*, 2016)

5. Secretion of newly produced protein.
    A plausible explanation for increased turnover rates at constant protein levels is that a fraction of these proteins is secreted from the cell. Secretion would keep the intracellular protein abundance constant despite of increased protein synthesis. The turnover signal RIA$_{oldLys}$ cannot differentiate between protein degradation and secretion, as in both cases, the intracellular labeling composition of such a protein would change.

    Yet according to SUBA3 analysis (Tanz *et al*, 2013), only 13 of the 121 proteins are predicted to be secreted. One of these proteins is Cyclase1 (AT4G34180), an important regulator of the programmed cell death response upon pathogen infection. It can be found in increased concentrations in the growth medium of Arabidopsis suspension cell cultures when stimulated by salicylic acid (Smith *et al*, 2015).

The comparably high prevalence of proteins with increased degradation that balances increased transcription points to an important phenomenon that should be investigated in further detail.

8.8.2   Growth rates affect turnover trajectories

Clearly, the growth rate is a crucial factor to consider when determining protein degradation processes. Several studies publish turnover rates without assessing the growth rates of the biological system (Lyon *et al*, 2014; Fan *et al*, 2016), which makes the comparisons between studies difficult even within the same organisms.

The interconnection of growth rate and protein turnover labeling was shown in my view most clearly by a pulsed-SILAC study in salt-stressed Chlamydomonas cultures (Mastrobuoni *et al*, 2012). The salt stress affects the growth rate (Figure 8.1A) and the label incorporation rate (Figure 8.1B). Yet, the label incorporation becomes comparable between the different treatments when compared to the cell numbers instead of the sampling time (Figure 8.1C), thus by taking the differences in growth rate into consideration.
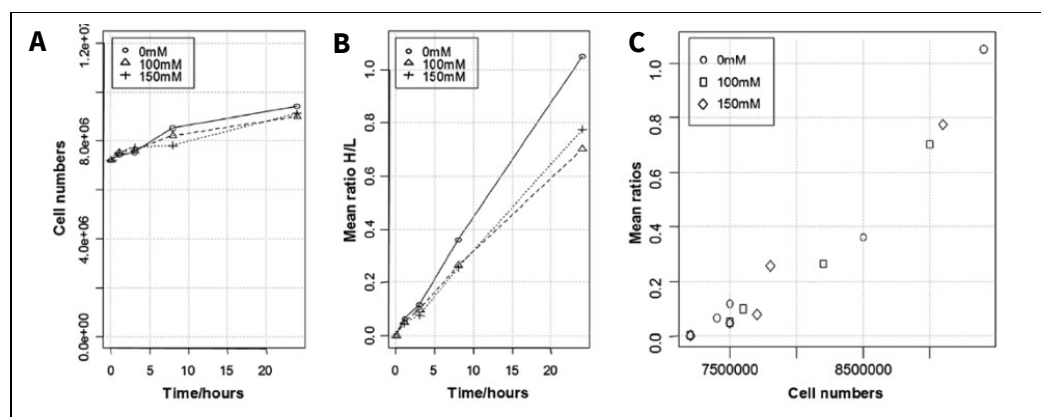
Figure 8.1: Pulse-SILAC labeled Chlamydomonas cultures affected by 0 mM, 100 mM or 150 mM of NaCl (Figure 1, Mastrobuoni et al, 2012).
Culture density (A) and label incorporation (B) over time. By plotting the label against culture density in cell numbers, the labeling becomes comparable across the treatments (C).

Price *et al*, (2010) identified different turnover rates for the same proteins within three mammalian tissues - liver, brain, and blood. I think that these findings might be explained by differences in the general growth rate of these tissues, as growth rate and turnover rate are negatively correlated. The quicker the cells of a tissue divide, the shorter the determined half-lives. Normalizing to the tissue-specific growth rates could increase the value of these findings.

Turck *et al*, (2016) detected two pools of proteins with different turnover rates for the same protein of the DPYSL family when analyzing brain tissue of ^{15}N pulse-labelled mice. This could indicate either a differentially regulated turnover within the same cell or, in my view more likely, a celltype-specific turnover rate within the brain (e.g. different growth rates of neurons and glia cells).

From the first large-scale degradation study (Pratt *et al*, 2002) until today (Nelson *et al*, 2014a), the growth rate has been taken into consideration by subtracting the dilution rate from the label-loss rate to obtain the degradation rate ($k_{deg} = k_{loss} - k_{dil}$, Pratt *et al*, (2002)). This approach has major drawbacks. First, it can only be applied when the protein degradation curves are interpreted strictly as an exponential decay process. Second, it results in negative protein decay values for up to 5% of all proteins (Price *et al*, 2010; Nelson *et al*, 2014a), making them uninterpretable.

The alternative approach is to divide the label-loss rate by the growth rate (relative $k_{deg} = k_{loss} / k_{dil}$; Eden *et al*, 2011) or by dividing the half-life by the doubling time (relative half-life = half-life / doubling time; Martin-Perez & Villen, 2015). These values are more biologically relevant and make the interpretation across species and studies more meaningful. It also does not exclude proteins with a relative decay rate of > 1, which would be a negative value in the subtraction approach. Instead, it incorporates them in the biological interpretation as dominated by dilution processes (Eden *et al*, 2011). Focused research on these cases will help to increase our fundamental understanding of untargeted protein degradation and general protein homeostasis.

In a review on plant protein turnover, Nelson *et al.* (2014b) warn explicitly against comparing half-lives determined in whole plants and in undifferentiated tissue cultures, and call for more baseline studies. In my view, the comparisons are possible as long as the growth rates are included more systematically. Most turnover experiments compare cultures with unaffected growth rates, making comparisons between treatment and control straightforward. Yet, our more difficult to interpret data on a treatment that affects the growth rate of the culture might help to understand more fundamentally how degradation rates are governed across organisms and cell lines.

8.8.3 Non-exponential decay is a widespread phenomenon

Many protein decay trajectories showed a peculiar pattern with a slowly decaying, initial lag phase that was followed by a phase with increased decay. I could show that many of these trajectories can be described more accurately by a non-exponential decay model such as the logistic decay model. When I became aware of this particular degradation pattern, I also detected it in several other studies in different organisms (Figure 8.2A-F). This decay pattern therefore seems to be a widespread biological phenomenon for which the scientific community does not have a full explanation.

I observed that the duration of the lag phase differs substantially between proteins. Therefore, I presume that the effect is protein-specific and cannot be explained by effects on the general amino acid pool, e.g. an incomplete exchange of the labeling amino acids or a high recycling rate of labeled amino acids. Therefore, adding a fixed offset in the decay model of all proteins (Li *et al*, 2012a) does not describe the phenomenon correctly and a more complex mathematical model of the decay process is needed.

It should also be asked if this initial slowed decay rate is merely a technical issue. Potentially, MS-based analyses could have problems to differentiate in a mass spectrum the initial small peaks at the beginning of a protein turnover experiment from background noise. Hughes & Krijgsveld, (2012) therefore recommend to use a 1:1 mix of the labeled amino acids in the pulse phase of a turnover experiment and to chase with just one of the two labeled amino acids. Despite of this concern, this initial lag phase pattern can be observed even when the protein turnover is quantified by SRM (Figure 8.2E, Holman *et al*, (2016)). Curiously, in the published figure, only one peptide shows this diverging pattern. It would be interesting to find out if that peptide originates from a differently spliced protein isoform or whether that peptide might be ambiguously mapping to another protein. In any case, a systematic analysis of this dataset for protein isoforms would be interesting.

In conclusion, the observed non-exponential decay pattern seems to be a wide-spread phenomenon that is worth to be analyzed systematically.
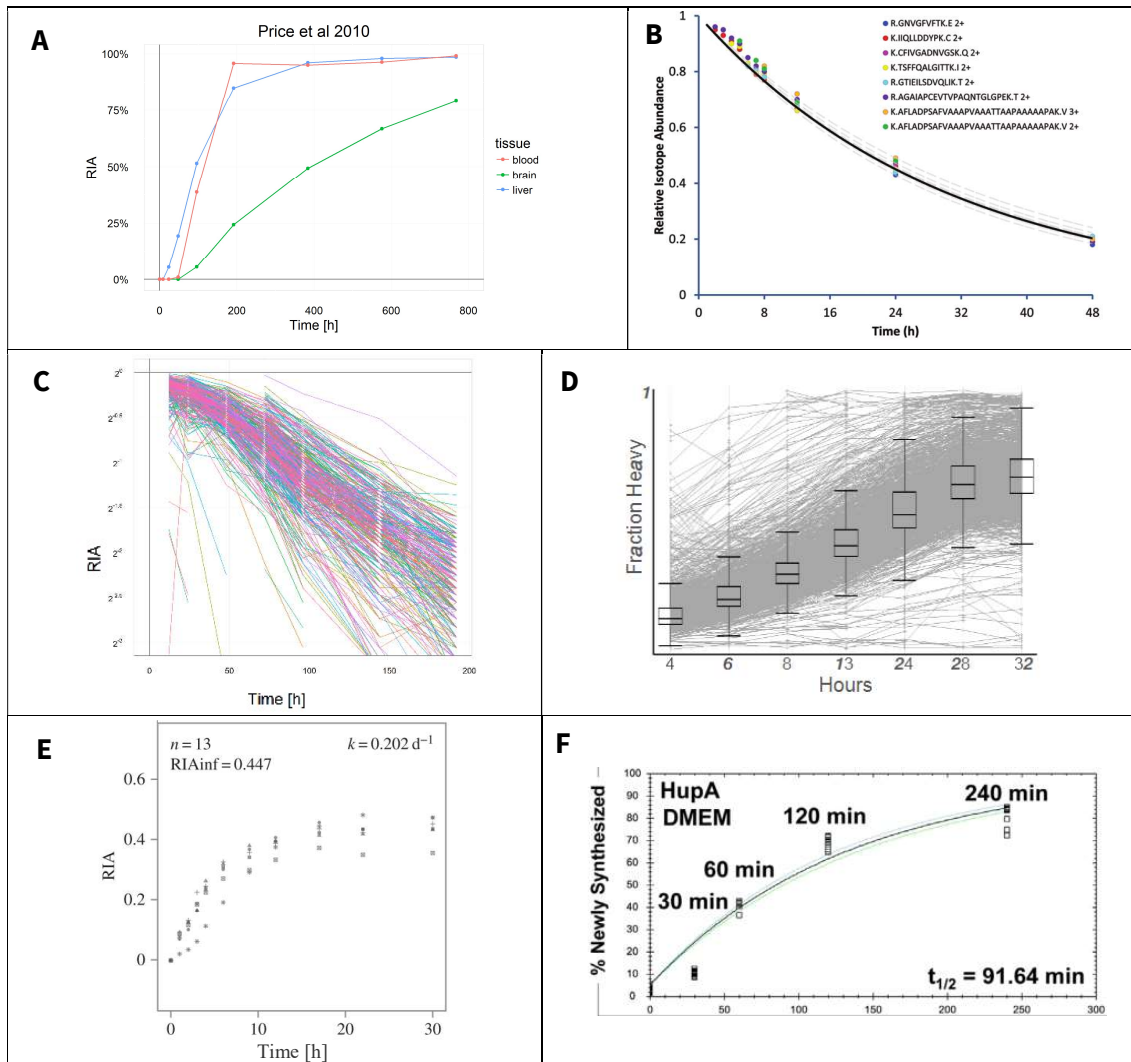
Figure 8.2: Exemplary figures from various publications that indicate a lag phase at the beginning of the protein degradation.

A: [15]N pulse-labeled mice, medians of all proteins of each tissue (Price et al, 2010). B: SILAC labeled HeLa cells (Figure 2, Hoopmann *et al*, 2011). C: ^{15}N-labelled barley leaves (Nelson *et al*, 2014a). Each line represents one quantified protein. An exponential decay pattern would result in a straight line in this semi-log plot. D: pulse SILAC labeling of *C. elegans* (Visscher *et al*, 2016). Each line represents one protein E: lysine SILAC labeling of mice, where turnover was quantified by SRM (Figure 4B, Holman *et al*, 2016). Each symbol represents a different peptide mapping to 2-oxoglutarate dehydrogenase (Q60597). F: pulse SILAC of *Salmonella typhimurium* (Figure 1, Wang *et al*, 2016).

8.8.4 Recycling of labeled peptides

When labeled proteins are degraded, their labeled amino acids can get recycled intact and incorporated into newly synthesized proteins. A high recycling rate would influence the absolute turnover kinetics, yet it is to be expected that the recycling rate would affect the total amino acid pool and as such all proteins equally. Also, it would affect all cultures equally, because the overall protein recycling rate would not be affected substantially by the treatment. In my view, the recycling rate does therefore not need to be taken into account for comparative analyses.

Yet, to model the turnover rate in a consistent mathematical framework together with the other three genome-wide datasets would require an estimation of the amino acid recycling rate to describe the turnover kinetics correctly. Cambridge *et al*, (2011) and Jovanovic *et al*, (2015) have determined the recycling rate from the labeling rates of misscleaved peptides with two labeled amino acids. Peptides with two differently labeled lysines can only occur, when 1) the amino acid pool is not fully exchanged yet or 2) amino acids are recycled from previously labeled proteins. This is a convincing approach, yet the number of such identified misscleaved peptides is typically very low. Furthermore, it is unclear how these misscleaved peptides with mixed labels were quantified, as to my knowledge, MaxQuant only reports quantitative values for uniformly labeled peptides.

In conclusion, while the recycling rate is not critical for comparative analyses, it is an important factor to consider for the absolute kinetics of protein turnover and consequently also affects the here presented half-life calculations. However, we expect especially for the comparison between different datasets that the recycling rate is a minor factor compared to differences in the employed kinetic model and the consideration of the growth rate.

8.8.5 Non-exponential decay models describe protein degradation trajectories more accurately

I used logistic decay models to describe the measured turnover trajectories, as they fit our data substantially better, yet there are no biological reasons for this particular mathematical model. The $Time_{half-life}$ parameter has no biological or physical representation in the protein's characteristic, such as e.g. its age. We chose this model primarily to indicate that the exponential decay model is too simple to explain the detected degradation trajectories.

To our knowledge, three studies have so far modeled these non-exponential decay patterns explicitly:

1. Similar to our interpretation, (Martin-Perez & Villén, 2015) modelled the non-exponential decaying proteins with a logistic decay model without further biological interpretation.
2. Rahman *et al*, (2016) presented a Gaussian process model to describe protein turnover data. The model describes protein turnover trajectories more accurately than an exponential decay model (Figure 8.3). Yet their model is not interpreted biologically and primarily highlights the need for a more complex description of protein turnover trajectories.
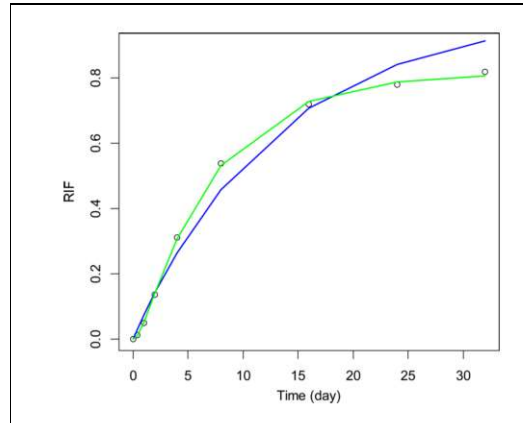
Figure 8.3: Modeling of the ¹⁵N pulse-labelled mouse protein Q8BMS1 (Figure 3, Rahman et al, 2016).
The experimental data (points) are fitted by exponential decay (blue line) and a Gaussian process model (green line). RIF = relative isotope frequency, an alternative term for the more frequently used term relative isotope abundance (RIA).

3. A very recent study by McShane et al, (2016) employed a more complex two-step exponential decay model to describe protein degradation trajectories, where the proteins have different decay rates according to their molecular age. The model was originally developed for non-exponential decay patterns of mRNA (Deneke *et al*, 2013), where the degradation rate depends on the age of the molecule (Figure 8.4A, B). The model is based on a Markov chain (Figure 8.4C), where each state of the chain degrades exponentially at an individual rate μ, and the progression from one state to the next is described by a linear rate λ. In other words, the molecules pass through the chain with increasing age. The mRNA decay model is based on a five-state Markov chain (Deneke *et al*, 2013), while the protein decay data model consists of two states (McShane *et al*, 2016). This means that protein decay is described by two different exponential decay rates and a transition rate. McShane et al, (2016) find that more than 10 % of all proteins are better described by such a non-exponential decay (NED) model. For all NED proteins was the later decay rate lower than the first one, indicating a decrease in protein decay with increasing age of the protein (Figure 8.4D). Proteins that increase in degradation rate with age could not be detected. For the degradation of mRNAs, both decay rates that increase and decrease by age were found (Figure 8.4D).

    This model interprets the non-exponential decay pattern in a biologically meaningful way by including the age of a molecule in the description.

With a small exploration, we could find degradation trajectories in our data set that depict such NED patterns, e.g. for disulfide isomerase-like protein 1 (AT1G21750, Figure 8.4F). The abundance of this protein does not change. The decay rate in the untreated sample potentially slows down with the age of the molecule, while in the flg22-treated sample the degradation rate seems to increases with age. Clearly, this phenomenon has to be analyzed systematically in future studies.
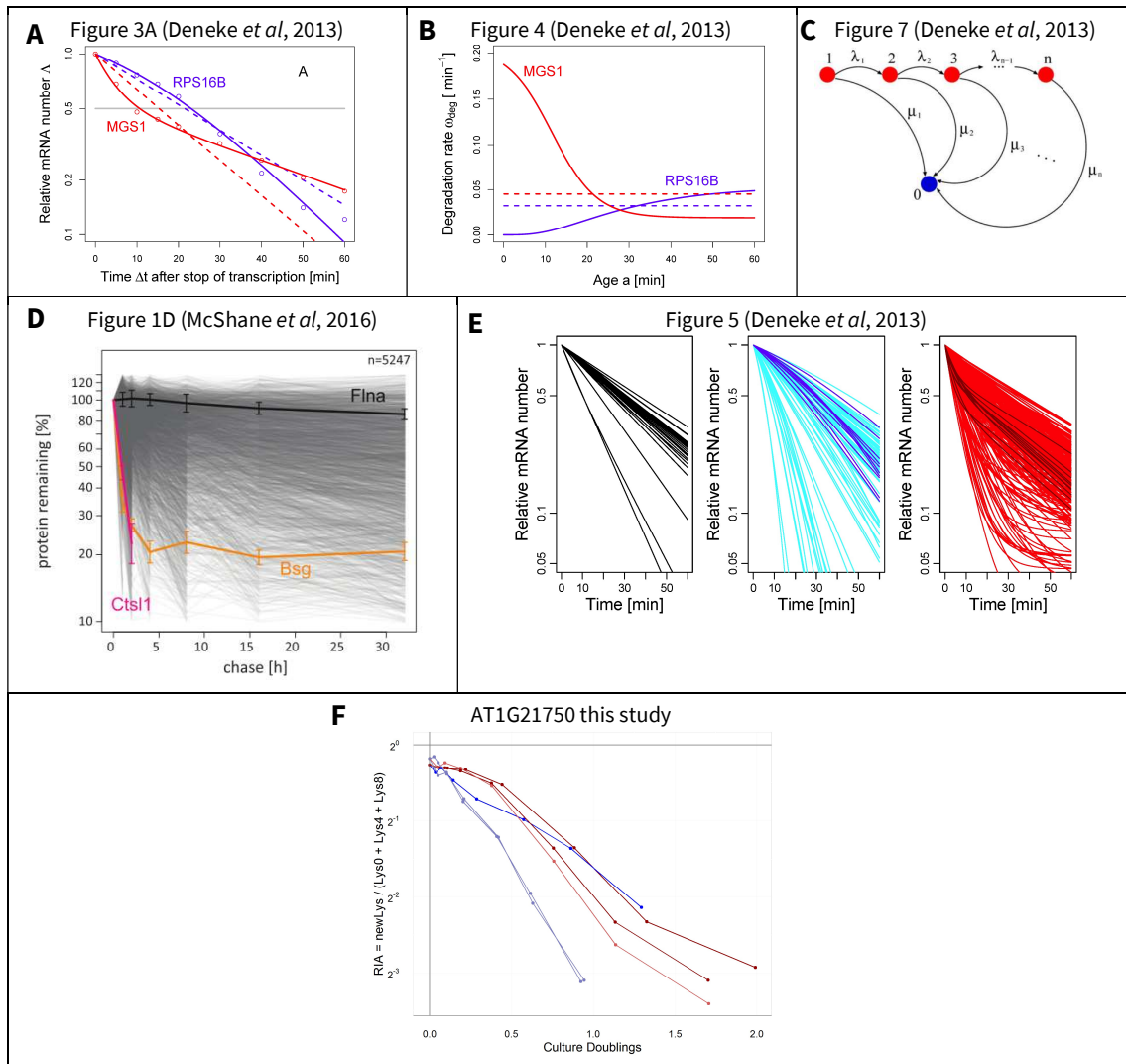
Figure 8.4: Modeling non-exponential decay patterns by considering the age of the molecules.

A: Degradation rates of two mRNAs. The data (points) were modeled by a multistep exponential decay process (drawn through lines) (Deneke *et al*, 2013). A classical one-step exponential decay pattern is indicated by a dashed line. B: Distribution of the degradation rate by age of the mRNAs from A (Deneke *et al*, 2013). C: Markov chain model of the degradation (Deneke *et al*, 2013). A molecule progresses from each state $\mathbb{N}$ at the rate λ and at each state, it decays exponentially with a rate μ. D: Decay rates of pulse-chase labeled proteins in a murine cell line (McShane *et al*, 2016). Flna and Ctsl1 show classical exponential decay patterns, while Bsg depicts a non-exponential decay pattern where the degradation of the protein decreases with time. E: mRNA decay rates modeled by a multistep exponential decay (Deneke *et al*, 2013). mRNAs whose degradation increases with age are indicated in blue, and mRNAs that decrease their degradation with age are indicated in red. F: Decay pattern of disulfide isomerase-like protein 1. Note that the degradation data are plotted in a semi-logarithmic fashion, which helps to distinguish the decay pattern from true exponential decay that would show a linear trajectory.

8.9 Future research directions

Even though the results described here provide a comprehensive view on gene expression regulation in response to a stimulus, more questions are being raised than answered. According to my view, the following avenues should be focused on:

1. Quantitative model of gene expression.
   The ultimate goal of acquiring the four datasets was to incorporate them into a quantitative kinetic model of gene expression. Similarly to Jovanovic et al (2015), this could give exact values on the contributions of transcription, translation and protein degradation on protein level regulation. Equally, such an analysis could expand the interpretation of the protein turnover analysis to the proteins with changing protein abundances, which had to be excluded in our current analysis. A kinetic model could evaluate whether the degradation of proteins that increase in abundance are also modulated, which would make them also relevant for the rejuvenation hypothesis discussed in chapter 8.8.1. We have established a collaboration with the Computational Systems Biology group of Prof. Jörg Stelling (ETH Zürich) to develop such a model.

2. Model degradation by an age-dependent multi-step exponential decay process.
   Following up on the discussion in chapter 8.8.3 it is of high interest, if our dataset can be modeled better with an age-dependent model as indicated by McShane et al, (2016) than with the logistic decay model. If the age-dependent degradation model holds true for our data, a fascinating question is raised: how can a cell discriminate proteins by their molecular age to degrade them accordingly?

3. Functional description of rejuvenated protein pools.
   The discussion in chapter 8.8.1 outlines several hypotheses for the functionality of a rejuvenated protein pool. All of these should be considered and tested experimentally. Most intriguing to me is the question if the metabolic activity of a rejuvenated protein pool does change. One would need to set up enzymatic assays for these proteins and test the activity before and after rejuvenation. Equally, the proteins' status regarding post-translational modifications before and after rejuvenation could be determined mass spectrometrically. Yet, before starting these analyses, Occam's razor recommends to start the test with the simplest explanation, which is the explanation by protein secretion (number 5 in chapter 8.8.1). Therefore, the list of rejuvenating proteins should be analyzed for localization and ER-targeting signals.

4. Subcellular localization of the nuclear-encoded, unlabelled proteins.
   The lack of lysine labeling of certain chloroplast-related proteins remains unexplained. They are highly enriched for chloroplast-localized proteins. The RPF length distribution verifies that these proteins are translated by 80S ribosomes, yet the low labeling rates indicate some form of compartmentalization. Microscopy localization studies with fluorescence tags require that the protein of interest is fully translated, folded and not bound by chaperones. In contrast, a LOPIT (Localization of Organelle Proteins by Isotope Tagging, (Dunkley *et al*, 2004)) study could reveal the subcellular localization of these proteins when they are not fully matured. In that approach, the subcellular compartments are fractionated by ultracentrifugation in a sucrose gradient and the proteins in the different compartment fractions are labeled by chemical post-preparation labels. The labeled peptides are then combined and analyzed mass spectrometrically. The latest development of the protocol (hyperLOPIT) can resolve up to 10

different subcellular compartments and was used to describe intracellular trafficking (Christoforou *et al*, 2016). Newly translated proteins will be low in abundance compared to the pre-existing proteins, so it is advised to enrich the newly synthesized proteins by AHA-methionine click-chemistry enrichment (Eichelbaum *et al*, 2012).

5. Disentangle the E3 ligase network.

   I see a tightly regulated degradation for at least 123 proteins (8.8.1). Currently, it is unknown, which E3 ligases are responsible for mediating this targeted degradation. The proteins identified here would be interesting and promising candidates for identifying the responsible E3 ligases that target them for proteasomal degradation, as these might play important roles in plant defense mechanisms.

6. Interplay between autophagy and proteasomal degradation.

   Equally, the proteins with longer half-lives than doubling times are interesting candidates to study the interplay of proteasomal degradation and degradation by autophagy. It is unknown to us, whether proteins can be degraded by the proteasome in an untargeted fashion by a non-discriminating E3 ligase or whether these proteins are rather degraded by autophagy.

9.

BIBLIOGRAPHY

Aebersold R & Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**:**7620** 347–355

Ahrné E, Molzahn L, Glatter T & Schmidt A (2013) Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **13**:**17** 2567–2578

Van Aken O, Zhang B, Law S, Narsai R & Whelan J (2013) AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. *Plant Physiol.* **162**:**1** 254–71

Albert FW, Muzzey D, Weissman JS & Kruglyak L (2014) Genetic Influences on Translation in Yeast. *PLoS Genet.* **10**:**10** e1004692

Ali Amjad M, Wieczorek K, Kreil DP & Bohlmann H (2014) The beet cyst nematode Heterodera schachtii modulates the expression of WRKY transcription factors in syncytia to favour its development in Arabidopsis roots. *PLoS One* **9**:**7** e102360

Allen DK, Laclair RW, Ohlrogge JB & Shachar-Hill Y (2012) Isotope labelling of Rubisco subunits provides in vivo information on subcellular biosynthesis and exchange of amino acids between compartments. *Plant, Cell Environ.* **35**:**7** 1232–1244

Arike L, Valgepea K, Peil L, Nahku R, Adamberg K & Vilu R (2012) Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *J. Proteomics* **75**:**17** 5437–5448

Von Arnim AG, Jia Q & Vaughn JN (2014) Regulation of plant translation by upstream open reading frames. *Plant Sci.* **214** 1–12

Arsova B, Kierszniowska S & Schulze WX (2012) The use of heavy nitrogen in quantitative proteomics experiments in plants. *Trends Plant Sci.* **17**:**2** 102–112

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W-L, Gomez-Gomez L, Boller T, Ausubel FM & Sheen J (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* **415**:**6875** 977–83

Ascencio-Ibáñez JT, Sozzani R, Lee T-J, Chu T-M, Wolfinger RD, Cella R & Hanley-Bowdoin L (2008) Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* **148**:**1** 436–454

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W & Baginsky S (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science* **320**:**5878** 938–41

Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Bühlmann P, Hennig L, Hirsch-Hoffmann M, Howell K a, Kahlau S, Radziejwoski A, Russenberger D, Rutishauser D, Small I, Stekhoven D, Sulpice R, Svozil J, Wuyts N, Stitt M, Hilson P, … Gruissem W (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Mol. Syst. Biol.* **8**:**606** 606

Bandyopadhyay U, Kaushik S, Varticovski L & Cuervo AM (2008) The chaperone-mediated autophagy receptor organizes in dynamic protein complexes at the lysosomal membrane. *Mol. Cell. Biol.* **28**:**18** 5747–63

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Çolak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ & Blencowe BJ (2012) The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science* **338**:**6114**

Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK & Gilad Y (2014) Impact of regulatory variation from RNA to protein. *Science* **347:6222** 664–667

Bielow C, Mastrobuoni G & Kempa S (2016) Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J. Proteome Res.* **15:3** 777–787

Bindschedler L V., Minibayeva F, Gardner SL, Gerrish C, Davies DR & Bolwell GP (2001) Early signalling events in the apoplastic oxidative burst in suspension cultured French bean cells involve cAMP and Ca2+. *New Phytol.* **151:1** 185–194

Bindschedler L V., Palmblad M & Cramer R (2008) Hydroponic isotope labelling of entire plants (HILEP) for quantitative plant proteomics; an oxidative stress case study. *Phytochemistry* **69:10** 1962–1972

Bittremieux W, Valkenborg D, Martens L & Laukens K (2016) Computational quality control tools for mass spectrometry proteomics. *Proteomics*

Boersema PJ, Aye TT, Van Veen TAB, Heck AJR & Mohammed S (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **8:22** 4624–4632

Boisvert F-M, Ahmad Y, Gierliński M, Charrière F, Lamont D, Scott M, Barton G & Lamond AI (2012) A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11:3** M111.011429

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30:15** 2114–2120

Branco-Price C, Kawaguchi R, Ferreira RB & Bailey-Serres J (2005) Genome-wide analysis of transcript abundance and translation in arabidopsis seedlings subjected to oxygen deprivation. *Ann. Bot.* **96:4** 647–660

Braun P, Carvunis AR, Charloteaux B, Dreze M, Ecker JR, Hill DE, Roth FP, Vidal M, Galli M, Balumuri P, Bautista V, Chesnut JD, Kim RC, de los Reyes C, Gilles P, Kim CJ, Matrubutham U, Mirchandani J, Olivares E, Patnaik S, Quan R, Ramaswamy G, Shinn P, Sw VM (2011) Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science* **333**:**July** 601–607

Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng L-Y, Messner S, Ehrenberger T, Zanotelli V, Butscheid Y, Escher C, Vitek O, Rinner O & Reiter L (2015) Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **14:5** 1400–1410

Buratowski S & Chodosh LA (2003) Unit 12.2 Mobility Shift DNA-Binding Assay Using Gel Electrophoresis. *Curr. Protoc. Mol. Biol.* **12:1996** Unit 12.2

Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Krueger M & Mann M (2011) Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10:12** 5275–5284

La Camera S, Balagué C, Göbel C, Geoffroy P, Legrand M, Feussner I, Roby D & Heitz T (2009) The Arabidopsis patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. *Mol. Plant. Microbe. Interact.* **22:4** 469–481

Cao H, Bowling S, Gordon A & Dong X (1994) Characterization of an Arabidopsis Mutant That Is Nonresponsive to Inducers of Systemic Acquired Resistance. *Plant Cell* **6:11** 1583–1592

Carrio-Segui A, Romero P, Sanz A & Penarrubia L (2016) Interaction between ABA signaling and copper homeostasis in Arabidopsis thaliana. *Plant Cell Physiol.* **57:7** 1568–1582

Castro-Concha LA, Escobedo RM & de Lourdes Miranda-Ham M (2006) Measurement of Cell Viability in In Vitro Cultures. *Plant Cell Cult. Protoc.* **318:6** 071–076

Chatterjee A, Abeydeera ND, Bale S, Pai P-J, Dorrestein PC, Russell DH, Ealick SE & Begley TP (2011) Saccharomyces cerevisiae THI4p is a suicide thiamine thiazole synthase. *Nature* **478:7370** 542–546

Chen L, Zhang L, Li D, Wang F & Yu D (2013) WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **110:21** E1963-71

Chen X, Wei S, Ji Y, Guo X & Yang F (2015) Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics* **15:18** 3175–3192

Chen Y, Mccarthy D, Robinson M & Smyth GK (2014) edgeR: differential expression analysis of digital gene expression data User ' s Guide.

Cheng Y, Chen Y & Yu C (2016) Fast and Efficient non-reduced Lys-C digest using pressure cycling technology for antibody disulfide mapping by LC-MS. *J. Pharm. Biomed. Anal.* **129** 203–209

Cheval C, Aldon D, Galaud J-P & Ranty B (2013) Calcium/calmodulin-mediated regulation of plant immunity. *Biochim. Biophys. Acta* **1833:7** 1766–71

Chevrier N, Mertins P, Artyomov MN, Shalek AK, Iannacone M, Ciaccio MF, Gat-Viks I, Tonti E, Degrace MM, Clauser KR, Garber M, Eisenhaure TM, Yosef N, Robinson J, Sutton A, Andersen MS, Root DE, Von Andrian U, Jones RB, … Hacohen N (2011) Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147:4** 853–867

Chi W, Sun X & Zhang L (2012) The roles of chloroplast proteases in the biogenesis and maintenance of photosystem II. *Biochim. Biophys. Acta - Bioenerg.* **1817:1** 239–246

Chinchilla D (2006) The Arabidopsis Receptor Kinase FLS2 Binds flg22 and Determines the Specificity of Flagellin Perception. *Plant Cell Online* **18:2** 465–476

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, Jones JDG, Felix G & Boller T (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448:7152** 497–500

Chotewutmontri P & Barkan A (2016) Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLoS Genet.* **12:7** e1006106

Christoforou A, Mulvey CM, Breckels LM, Geladaki A, Hurrell T, Hayward PC, Naake T, Gatto L, Viner R, Martinez Arias A & Lilley KS (2016) A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7** 8992

Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC & Brierley I (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *Rna* **21:10** 1731–1745

Clay NK, Adio AM, Denoux C, Jander G & Ausubel FM (2009) Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* **323:5910** 95–101

Claydon AJ, Thom MD, Hurst JL & Beynon RJ (2012) Protein turnover: Measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. *Proteomics* **12:8** 1194–1206

Conrads TP, Alving K, Veenstra TD, Belov ME, Anderson GA, Anderson DJ, Lipton MS, Paša-Tolić L, Udseth HR, Chrisler WB, Thrall BD & Smith RD (2001) Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. *Anal. Chem.* **73:9** 2132–2139

Consonni C, Bednarek P, Humphry M, Francocci F, Ferrari S, Harzen A, Ver Loren van Themaat E & Panstruga R (2010) Tryptophan-derived metabolites are required for antifungal defense in the Arabidopsis mlo2 mutant. *Plant Physiol.* **152:3** 1544–1561

Cox J, Hein MY, Luber C a & Paron I (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13:9** 2513–2526

Cox J & Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26:12** 1367–72

Crick F & others (1970) Central dogma of molecular biology. *Nature* **227:5258** 561–563

Das SN, Madhuprakash J, Sarma PVSRN, Purushotham P, Suma K, Manjeet K, Rambabu S, Gueddari NE El, Moerschbacher BM & Podile AR (2015) Biotechnological approaches for field applications of chitooligosaccharides (COS) to induce innate immunity in plants. *Crit. Rev. Biotechnol.* **35:1** 29–43

Deneke C, Lipowsky R & Valleriani A (2013) Complex Degradation Processes Lead to Non-Exponential Decay Patterns and Age-Dependent Decay Rates of Messenger RNA. *PLoS One* **8:2** e55442

Dephoure N & Gygi SP (2012) Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Sci. Signal.* **5:217** rs2

Dettmann EH (2008) Turnover Time. In *Encyclopedia of Ecology* pp 3639–3644.

Dhondt I, Petyuk VA, Cai H, Vandemeulebroucke L, Vierstraete A, Smith RD, Depuydt G & Braeckman BP (2016) FOXO/DAF-16 Activation Slows Down Turnover of the Majority of Proteins in C.elegans. *Cell Rep.* **16:11** 3028–3040

Díaz-Troya S, Pérez-Pérez ME, Florencio FJ & Crespo JL (2008) The role of TOR in autophagy regulation from yeast to plants and mammals. *Autophagy* **4:7** 851–865

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA -seq aligner. *Bioinformatics* 1–7

Dobson RCJ, Girón I & Hudson AO (2011) L,L-diaminopimelate aminotransferase from chlamydomonas reinhardtii: A target for algaecide development. *PLoS One* **6:5** e20439

Doherty MK, Hammond DE, Clague MJ, Gaskell SJ & Beynon RJ (2009) Turnover of the human proteome: Determination of protein intracellular stability by dynamic SILAC. *J. Proteome Res.* **8:1** 104–112

Doherty MK, Whitehead C, McCormack H, Gaskell SJ & Beynon RJ (2005) Proteome dynamics in complex organisms: Using stable isotopes to monitor individual protein turnover rates. *Proteomics* **5:2** 522–533

Dunkley TPJ, Watson R, Griffin JL, Dupree P & Lilley KS (2004) Localization of Organelle Proteins by Isotope Tagging (LOPIT). *Mol. Cell. Proteomics* **3:11** 1128–1134

Dutilleul C, Garmier M, Noctor G, Mathieu C, Chétrit P, Foyer CH & Paepe R De (2003) Leaf Mitochondria Modulate Whole Cell Redox Homeostasis , Set Antioxidant Capacity , and Determine Stress Resistance through Altered Signaling and Diurnal Regulation. *Plant Cell* **15:May** 1212–1226

Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A & Alon U (2011) Proteome half-life dynamics in living human cells. *Science* **331:6018** 764–8

Ederli L, Madeo L, Calderini O, Gehring C, Moretti C, Buonaurio R, Paolocci F & Pasqualini S (2011) The Arabidopsis thaliana cysteine-rich receptor-like kinase CRK20 modulates host responses to Pseudomonas syringae pv. tomato DC3000 infection. *J. Plant Physiol.* **168:15** 1784–1794

Eichelbaum K, Winter M, Diaz MB, Herzig S & Krijgsveld J (2012) Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* **30:10** 984–990

Eisenstein M (2012) Oxford Nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.* **30:4** 295–296

Engelsberger WR, Erban A, Kopka J & Schulze WX (2006) Metabolic labeling of plant cell cultures with K(15)NO3 as a tool for quantitative analysis of proteins and metabolites. *Plant Methods* **2:3** 14

Environment RCT 2015 R a language and & for statistical computing. Vienna ARF for SC (2015) R: a language and for statistical computing. (http://www. Rproject.org)

Eulgem T, Rushton PJ, Robatzek S & Somssich IE (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5:5** 199–206

Ezkurdia I, Calvo E, Del Pozo A, Vázquez J, Valencia A & Tress ML (2015) The potential clinical impact of the release of two drafts of the human proteome. *Expert Rev. Proteomics* **12:6** 579–93

Fabre B, Lambour T, Bouyssié D, Menneteau T, Monsarrat B, Burlet-Schiltz O & Bousquet-Dubouch MP (2014) Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteomics* **4** 82–86

Fan C, Clay MD, Deyholos MK & Vederas JC (2010) Exploration of inhibitors for diaminopimelate aminotransferase. *Bioorganic Med. Chem.* **18:6** 2141–2151

Fan C & Vederas JC (2012) Synthesis and structure-activity relationships of o-sulfonamido-arylhydrazides as inhibitors of LL-diaminopimelate aminotransferase (LL-DAP-AT). *Org. Biomol. Chem.* **10:30** 5815–9

Fan KT, Rendahl AK, Chen WP, Freund DM, Gray WM, Cohen JD & Hegeman AD (2016) Proteome Scale-Protein Turnover Analysis Using High Resolution Mass Spectrometric Data from Stable-Isotope Labeled Plants. *J. Proteome Res.* **15:3** 851–867

Felix G, Duran JD, Volko S & Boller T (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18:3** 265–276

Feys BJ, Moisan LJ, Newman MA & Parker JE (2001) Direct interaction between the Arabidopsis disease resistance signaling proteins, EDS1 and PAD4. *EMBO J.* **20:19** 5400–5411

Fierro-Monti I, Racle J, Hernandez C, Waridel P, Hatzimanikatis V & Quadroni M (2013) A novel pulse-chase SILAC strategy measures changes in protein decay and synthesis rates induced by perturbation of proteostasis with an Hsp90 inhibitor. *PLoS One* **8:11**

Fitzpatrick TB & Thore S (2014) Complex behavior: From cannibalism to suicide in the vitamin B1 biosynthesis world. *Curr. Opin. Struct. Biol.* **29** 34–43

Förster A, Masters EI, Whitby FG, Robinson H & Hill CP (2005) The 1.9 A structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol. Cell* **18:5** 589–599

Fredens J, Engholm-Keller K, Giessing A, Pultz D, Larsen MR, Højrup P, Møller-Jensen J & Færgeman NJ (2011) Quantitative proteomics by amino acid labeling in C. elegans. *Nat. Methods* **8:10** 845–847

Freeberg L, Kuersten S & Syed F (2013) Isolate and sequence ribosome-protected mRNA fragments using size-exclusion chromatography. *Nat. Methods* **10:5** i–ii

Gaba A, Wang Z, Krishnamoorthy T, Hinnebusch AG & Sachs MS (2001) Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *EMBO J.* **20:22** 6453–63

Galderisi U, Jori FP & Giordano A (2003) Cell cycle regulation and neural differentiation. *Oncogene* **22:33** 5208–5219

Galili G (1995) Regulation of Lysine and Threonine Synthesis. *Plant Cell* **7:7** 899–906

Galland M, Huguet R, Arc E, Cueff G, Job D & Rajjou L (2014) Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. *Mol. Cell. proteomics* **13:1** 252–68

Gamm M, Peviani A, Honsel A, Snel B, Smeekens S & Hanson J (2014) Increased sucrose levels mediate selective mRNA translation in Arabidopsis. *BMC Plant Biol.* **14:1** 306

Geiger T, Cox J, Ostasiewicz P, Wisniewski JR & Mann M (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7:5** 383–385

Geiger T, Wisniewski JR, Cox J, Zanivan S, Kruger M, Ishihama Y & Mann M (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* **6:2** 147–157

Geilen K & Böhmer M (2015) Dynamic subnuclear relocalisation of WRKY40 in response to Abscisic acid in Arabidopsis thaliana. *Sci. Rep.* **5** 13369

Gershon PD (2014) Cleaved and missed sites for trypsin, Lys-C, and Lys-N can be predicted with high confidence on the basis of sequence context. *J. Proteome Res.* **13:2** 702–709

Ghislain M, Frankard V & Jacobs M (1990) Dihydrodipicolinate synthase of nicotiana sylvestris, a chloroplast-localized enzyme of the lysine pathway. *Planta* **180:4** 480–486

Giansanti P, Tsiatsiani L, Low TY & Heck AJR (2016) Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* **11:5** 993–1006

Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R & Aebersold R (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **11:6** O111.016717-O111.016717

Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* **426:6968** 895–899

Green CE & Phillips RL (1974) Potential Selection System for Mutants with Increased Lysine, Threonine, and Methionine in Cereal Crops1. *Crop Sci.* **14:6** 827

Griss J, Perez-Riverol Y, Lewis S, Tabb DL, Dianes JA, del-Toro N, Rurik M, Walzer M, Kohlbacher O, Hermjakob H, Wang R & Vizcaíno JA (2016) Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat. Methods* **13:8** 651–656

Grossmann J, Roschitzki B, Panse C, Fortes C, Barkow-Oesterreicher S, Rutishauser D & Schlapbach R (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics* **73:9** 1740–1746

Gruhler A, Olsen J V, Mohammed S, Mortensen P, Faergeman NJ, Mann M & Jensen ON (2005a) Quantitative Phosphoproteomics Applied to the Yeast Pheromone Signaling Pathway. *Mol Cell Proteomics* **4:3** 310–327

Gruhler A, Schulze WX, Matthiesen R, Mann M & Jensen ON (2005b) Stable Isotope Labeling of Arabidopsis thaliana Cells and Quantitative Proteomics by Mass Spectrometry. *Mol. Cell. Proteomics* **4:11** 1697–1709

Guan S, Price JC, Prusiner SB, Ghaemmaghami S & Burlingame a. L (2011) A Data Processing Pipeline for Mammalian Proteome Dynamics Studies Using Stable Isotope Metabolic Labeling. *Mol. Cell. Proteomics* **10:12** M111.010728-M111.010728

Guo H, Ingolia NT, Weissman JS & Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:7308** 835–840

Gutzat R, Borghi L & Gruissem W (2012) Emerging roles of RETINOBLASTOMA-RELATED proteins in evolution and plant development. *Trends Plant Sci.* **17:3** 139–148

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH & Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17:10** 994–9

Hatakeyama M, Opitz L, Russo G, Qi W, Schlapbach R, Rehrauer H, Noble W, Goecks J, Nekrutenko A, Taylor J, Fisch K, Meißner T, Gioia L, Ducom J-C, Carland T, Loguercio S, Su A, Halbritter F, Vaidya H, … Dijkstra E (2016) SUSHI: an exquisite recipe for fully documented, reproducible and reusable NGS data analysis. *BMC Bioinformatics* **17:1** 228

Hay N & Sonenberg N (2004) Upstream and downstream of mTOR. *Genes Dev.* **18:16** 1926–1945

Helbig AO, Daran-Lapujade P, van Maris AJ a, de Hulster E a F, de Ridder D, Pronk JT, Heck AJR & Slijper M (2011) The diversity of protein turnover and abundance under nitrogen-limited steady-state conditions in Saccharomyces cerevisiae. *Mol. Biosyst.* **7:March** 3316–3326

Heremans B & Jacobs M (1997) A nutant of Arabidopsis thaliana (L.) heynh. with modified control of aspartate kinase by threonine. *Biochem. Genet.* **35:3**–**4** 139–153

Herzog E, Guilley H & Fritsch C (1995) Translation of the Second Gene of Peanut Clump Virus RNA 2 Occurs by Leaky Scanning in Vitro. *Virology* **208:1** 215–225

Hiller K, Metallo C & Stephanopoulos G (2011) Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics. *Curr. Pharm. Biotechnol.* **12:7** 1075–1086

Hinkson I V. & Elias JE (2011) The dynamic state of protein turnover: It's about time. *Trends Cell Biol.* **21:5** 293–303

Höhner R, Barth J, Magneschi L, Jaeger D, Niehues A, Bald T, Grossman A, Fufezan C & Hippler M (2013) The metabolic status drives acclimation of iron deficiency responses in Chlamydomonas reinhardtii as revealed by proteomics based hierarchical clustering and reverse genetics. *Mol Cell Proteomics* **12:10** 2774–2790

Holman SW, Hammond DE, Simpson DM, Waters J, Hurst JL & Beynon RJ (2016) Protein turnover measurement using selected reaction monitoring-mass spectrometry (SRM-MS). *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.* **374:2079**

Holper S, Ruhs A & Kruger M (2014) Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *Methods Mol. Biol.* **1188** 191–205

Hoopmann MR, Chavez JD & Bruce JE (2011) SILACtor: Software to enable dynamic SILAC studies. *Anal. Chem.* **83:22** 8403–8410

Howard BE, Hu Q, Babaoglu AC, Chandra M, Borghi M, Tan X, He L, Winter-Sederoff H, Gassmann W, Veronese P & Heber S (2013) High-Throughput RNA Sequencing of Pseudomonas-Infected Arabidopsis Reveals Hidden Transcriptome Complexity and Novel Splice Variants. *PLoS One* **8:10** e74183

Hu Y, Chen L, Wang H, Zhang L, Wang F & Yu D (2013) Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J.* **74:5** 730–745

Huang S, Nelson CJ, Li L, Taylor NL, Ströher E, Peteriet J & Millar AH (2015) INTERMEDIATE CLEAVAGE PEPTIDASE55 Modifies Enzyme Amino Termini and Alters Protein Stability in Arabidopsis Mitochondria. *Plant Physiol.* **168:2** 415–27

Hudson O, Singh BK, Leustek T & Gilvarg C (2006) An LL -Diaminopimelate Aminotransferase Defines a Novel Variant of the Lysine Biosynthesis Pathway in Plants. *Plant Physiol.* **140**:**January** 292–301

Hughes C & Krijgsveld J (2012) Developments in quantitative mass spectrometry for the analysis of proteome dynamics. *Trends Biotechnol.* **30:12** 668–676

Ikeda F, Crosetto N & Dikic I (2010) What determines the specificity and outcomes of Ubiquitin signaling? *Cell* **143:5** 677–681

Inada T, Winstall E, Tarun SZ, Yates JR, Schieltz D & Sachs AB (2002) One-step affinity purification of the yeast ribosome and its associated proteins and mRNAs. *RNA* **8:7** 948–958

Ingolia NT (2016) Ribosome Footprint Profiling of Translation throughout the Genome. *Cell* **165:1** 22–33

Ingolia NT, Brar GA, Rouskin S, McGeachy AM & Weissman JS (2013) Genome-wide annotation and quantitation of translation by ribosome profiling. *Curr. Protoc. Mol. Biol.* Chapter 4:SUPPL.103 Unit 4.18

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJS, Jackson SE, Wills MR & Weissman JS (2014) Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* **8:5** 1365–1379

Ingolia NT, Ghaemmaghami S, Newman JRS & Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:5924** 218–23

Ingolia NT, Lareau LF & Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:4** 789–802

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J & Mann M (2005) Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* **4:9** 1265–1272

Ishihara A, Hashimoto Y, Tanaka C, Dubouzet JG, Nakao T, Matsuda F, Nishioka T, Miyagawa H & Wakasa K (2008) The tryptophan pathway is involved in the defense responses of rice against pathogenic infection via serotonin production. *Plant J.* **54:3** 481–495

Ishihara H, Obata T, Sulpice R, Fernie AR & Stitt M (2015) Quantifying protein synthesis and degradation in arabidopsis by dynamic 13CO2 labeling and analysis of enrichment in individual amino acids in their free pools and in protein. *Plant Physiol.* **168:1** 74–93

Jayapal KP, Sui S, Philp RJ, Kok YJ, Yap MGS, Griffin TJ & Hu WS (2010) Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. *J. Proteome Res.* **9:5** 2087–2097

Jekel PA, Weijer WJ & Beintema JJ (1983) Use of endoproteinase Lys-C from Lysobacter enzymogenes in protein sequence analysis. *Anal. Biochem.* **134:2** 347–354

Jovanovic M, Rooney MS, Mertins P, Przybylski D, Chevrier N, Satija R, Rodriguez EH, Fields AP, Schwartz S, Raychowdhury R, Mumbach MR, Eisenhaure T, Rabani M, Gennert D, Lu D, Delorey T, Weissman JS, Carr SA, Hacohen N & Regev A (2015) Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347:6226** 1259038

Juntawong P & Bailey-Serres J (2012) Dynamic Light Regulation of Translation Status in Arabidopsis thaliana. *Front. Plant Sci.* **3:April** 1–16

Juntawong P, Girke T, Bazin J & Bailey-Serres J (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **111:1** E203-12

Juntawong P, Sorenson R & Bailey-Serres J (2013) Cold shock protein 1 chaperones mRNAs during translation in Arabidopsis thaliana. *Plant J.* **74:6** 1016–1028

Karp NA & Lilley KS (2007) Design and analysis issues in quantitative proteomics studies. *Pract. Proteomics* **7** 42–50

Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE, Coruzzi GM & Gutierrez RA (2010) VirtualPlant: A Software Platform to Support Systems Biology Research. *Plant Physiol* **152:2** 500–515

Kawaguchi R, Girke T, Bray EA & Bailey-Serres J (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *Plant J.* **38:5** 823–839

Keys AJ, Bird IF, Cornelius MJ, Lea PJ, Wallsgrove RM & Miflin BJ (1978) Photorespiratory nitrogen cycle. *Nature* **275:5682** 741–743

Khmelinskii A, Keller PJ, Bartosik AA, Meurer M, Barry JD, Mardin BR, Kaufmann A, Trautmann S, Wachsmuth M, Pereira G, Huber W, Schiebel E & Knop M (2012) Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat. Biotechnol.* **30:7** 708–714

Kierszniowska S, Walther D & Schulze WX (2009) Ratio-dependent significance thresholds in reciprocal 15N-labeling experiments as a robust tool in detection of candidate proteins responding to biological treatment. *Proteomics* **9:7** 1916–1924

Kim B-H, Cai X, Vaughn JN & von Arnim AG (2007) On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biol.* **8:4** R60

Kim M-S, Pinto S, Getnet D, Nirujogi R, Manda S, Chaerkady R, Madugundu A, Kelkar D, Isserlin R, Jain S, Thomas J, Muthusamy B, Pamela L-R, Kumar P, Sahasrabuddhe N, Balakrishnan L, Advani J, George B, Renuse S, … Pandey A (2014) A draft map of the human proteome. *Nature* **509:7502** 575–581

King HA & Gerber AP (2016) Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Brief. Funct. Genomics* **15:1** 22–31

Kline KG, Barrett-Wilt GA & Sussman MR (2010) In planta changes in protein phosphorylation induced by the plant hormone abscisic acid. *Proc. Natl. Acad. Sci. U. S. A.* **107:36** 15986–15991

Koch A, Gawron D, Steyaert S, Ndah E, Crappd J, De Keulenaer S, De Meester E, Ma M, Shen B, Gevaert K, Van Criekinge W, Van Damme P & Menschaert G (2014) A proteogenomics approach integrating proteomics and ribosome profiling increases the efficiency of protein identification and enables the discovery of alternative translation start sites. *Proteomics* **14:23–24** 2688–2698

Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM a, Pronk JT, Slijper M & Heck AJR (2006) Proteome analysis of yeast response to various nutrient limitations. *Mol. Syst. Biol.* **2:1** 2006.0026

Kölling K, Müller A, Flütsch P & Zeeman SC (2013) A device for single leaf labelling with CO2 isotopes to study carbon allocation and partitioning in Arabidopsis thaliana. *Plant Methods* **9:1** 45

Komander D & Rape M (2012) The Ubiquitin Code. *Annu. Rev. Biochem.* **81:1** 203–229

Kristensen AR, Gsponer J & Foster LJ (2013) Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* **9:689** 689

Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fässler R & Mann M (2008) SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134:2** 353–364

Kuersten S, Radek A, Vogel C & Penalva LOF (2013) Translation regulation gets its 'omics' moment. *Wiley Interdiscip. Rev. RNA* **4:6** 617–630

Landry CR, Zhong X, Nielly-Thibault L & Roucou X (2015) Found in translation: Functions and evolution of a recently discovered alternative proteome. *Curr. Opin. Struct. Biol.* **32** 74–80

Langmead B, Trapnell C, Pop M & Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10:3** R25

Larsson O, Sonenberg N & Nadon R (2010) Identification of differential translation in genome wide studies. *Proc. Natl. Acad. Sci. U. S. A.* **107:50** 21487–92

Lauressergues D, Couzigou J-M, Clemente HS, Martinez Y, Dunand C, Bécard G & Combier J-P (2015) Primary transcripts of microRNAs encode regulatory peptides. *Nature* **520:7545** 90–93

Lee H, Chah O-K & Sheen J (2011) Stem-cell-triggered immunity through CLV3p-FLS2 signalling. *Nature* **473:7347** 376–379

Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, Zhao H, Song W, Fan Z & Lai J (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* **84:6** 1206–1208

Levine B, Mizushima N & Virgin HW (2011) Autophagy in immunity and inflammation. *Nature* **469:7330** 323–335

Lewandowska D, ten Have S, Hodge K, Tillemans V, Lamond AI & Brown JWS (2013) Plant SILAC: Stable-Isotope Labelling with Amino Acids of Arabidopsis Seedlings for Quantitative Proteomics. *PLoS One* **8:8** e72207

Li B, Meng X, Shan L & He P (2016) Transcriptional Regulation of Pattern-Triggered Immunity in Plants. *Cell Host and Microbe* **19:5** 641-650

Li F, Cheng C, Cui F, De Oliveira MV V, Yu X, Meng X, Intorne AC, Babilonia K, Li M, Li B, Chen S, Ma X, Xiao S, Zheng Y, Fei Z, Metz RP, Johnson CD, Koiwa H, Sun W, … He P (2014a) Modulation of RNA polymerase II phosphorylation downstream of pathogen perception orchestrates plant immunity. *Cell Host Microbe* **16:6** 748–758

Li JJ, Bickel PJ & Biggin MD (2014b) System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2** e270

Li L, Nelson CJ, Solheim C, Whelan J & Millar a. H (2012a) Determining Degradation and Synthesis Rates of Arabidopsis Proteins Using the Kinetics of Progressive 15N Labeling of Two-dimensional Gel-separated Protein Spots. *Mol. Cell. Proteomics* **11:6** M111.010025-M111.010025

Li WW, Li J & Bao JK (2012b) Microautophagy: Lesser-known self-eating. *Cell. Mol. Life Sci.* **69:7** 1125–1136

Liao Y, Smyth GK & Shi W (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:7** 923–930

Liu Y, Beyer A & Aebersold R (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165:3** 535–550

Looso M, Michel CS, Konzer A, Bruckskotten M, Borchardt T, Krüger M & Braun T (2012) Spiked-in pulsed in vivo labeling identifies a new member of the CCN family in regenerating newt hearts. *J. Proteome Res.* **11:9** 4693–4704

Loraine A (2009) Plant Systems Biology. *Dmitry A. Belostotsky (ed.), 1* **553:4** 247–265

Lotze MT, Zeh HJ, Rubartelli A, Sparvero LJ, Amoscato AA, Washburn NR, DeVera ME, Liang X, Tor M & Billiar T (2007) The grateful dead: Damage-associated molecular pattern molecules and reduction/oxidation regulate immunity. *Immunol. Rev.* **220:1** 60–81

Luo X, Xi Bai, Sun X, Zhu D, Liu B, Ji W, Cai H, Cao L, Wu J, Hu M, Liu X, Tang L & Zhu Y (2013) Expression of wild soybean WRKY20 in Arabidopsis enhances drought toleranc. *J. Exp. Bot.* **64:8** 2155–2169

Lyon D, Castillejo MA, Staudinger C, Weckwerth W, Wienkoop S & Egelhofer V (2014) Automated protein turnover calculations from 15n partial metabolic labeling lc/ms shotgun proteomics data. *PLoS One* **9:4** e94692

Mackowiak SD, Zauber H, Bielow C, Thiel D, Kutz K, Calviello L, Mastrobuoni G, Rajewsky N, Kempa S, Selbach M & Obermayer B (2015) Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol* **16:1** 179

Maier T, Schmidt A, Güell M, Kühner S, Gavin A-C, Aebersold R & Serrano L (2011) Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* **7:1** 511

Mao G, Meng X, Liu Y, Zheng Z, Chen Z & Zhang S (2011) Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *Plant Cell* **23:4** 1639–53

Martin-Perez M & Villén J (2015) Feasibility of Protein Turnover Studies in Prototroph Saccharomyces cerevisiae Strains. *Anal. Chem.* **87:7** 4008–4014

Martin SF, Munagapati VS, Salvo-Chirnside E, Kerr LE & Le Bihan T (2012) Proteome turnover in the green alga Ostreococcus tauri by time course 15N metabolic labeling mass spectrometry. *J. Proteome Res.* **11:1** 476–486

Martinez-Medina A, Flors V, Heil M, Mauch-Mani B, Pieterse CM., Pozo MJ, Ton J, van Dam NM, Conrath U, Conrath U, al. et, Hilker M, al. et, Conrath U, al. et, Luna E, al. et, Pieterse CMJ, al. et, … al. et (2016) Recognizing Plant Defense Priming. *Trends Plant Sci.* **21:10** 818–822

Mastrobuoni G, Irgang S, Pietzke M, Aßmus HE, Wenzel M, Schulze WX & Kempa S (2012) Proteome dynamics and early salt stress response of the photosynthetic organism Chlamydomonas reinhardtii. *BMC Genomics* **13:1** 215

Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, Satou M, Kim JM, Kobayashi N, Toyoda T, Shinozaki K & Seki M (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol.* **49:8** 1135–1149

McManus J, Cheng Z & Vogel C (2015) Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation. *Mol. Biosyst.* **11:10** 2680–9

McShane E, Sin C, Zauber H, Wells JN, Donnelly N, Wang X, Hou J, Chen W, Storchova Z, Marsh JA, Valleriani A & Selbach M (2016) Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **167:3** 803–815.e21

Meijer HA & Thomas AAM (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem. J.* **367:Pt 1** 1–11

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN & Alonso JM (2015) Gene-Specific Translation Regulation Mediated by the Hormone-Signaling Molecule EIN2. *Cell* **163:3** 684–697

Mertins P, Udeshi ND, Clauser KR, Mani D, Patel J, Ong S -e., Jaffe JD & Carr S a. (2012) iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics. *Mol. Cell. Proteomics* **11:6** M111.014423-M111.014423

Metzker ML (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11:1** 31–46

Mi H, Muruganujan A, Casagrande JT & Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8:8** 1551–1566

Mi H, Poudel S, Muruganujan A, Casagrande JT & Thomas PD (2016) PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* **44:D1** D336–D342

Mikheyev AS & Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* **14:6** 1097–1102

Morris DR & Geballe AP (2000) Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol. Cell. Biol.* **20:23** 8635–42

Mustafa NR, de Winter W, van Iren F & Verpoorte R (2011) Initiation, growth and cryopreservation of plant cell suspension cultures. *Nat. Protoc.* **6:6** 715–742

Mustroph A, Zanetti ME, Jang CJH, Holtan HE, Repetti PP, Galbraith DW, Girke T & Bailey-Serres J (2009) Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **106:44** 18843–8

Muth T, Renard BY & Martens L (2016) Metaproteomic data analysis at a glance: advances in computational microbial community proteomics. *Expert Rev. Proteomics* **13:8** 757–69

Naumann B, Stauber EJ, Busch A, Sommer F & Hippler M (2005) N-terminal processing of Lhca3 is a key step in remodeling of the photosystem I-light-harvesting complex under iron deficiency in Chlamydomonas reinhardtii. *J. Biol. Chem.* **280:21** 20431–20441

Navarro L, Zipfel C, Rowland O, Keller I, Robatzek S, Boller T & Jones JDG (2004) The Transcriptional Innate Immune Response to flg22. Interplay and Overlap with Avr Gene-Dependent Defense Responses and Bacterial Pathogenesis. *Plant Physiol.* **135:2** 1113–1128

Navarro P, Kuharev J, Gillet LC, Bernhardt OM, MacLean B, Röst HL, Tate SA, Tsou C-C, Reiter L, Distler U, Rosenberger G, Perez-Riverol Y, Nesvizhskii AI, Aebersold R & Tenzer S (2016) A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **34:11** 1130–1136

Nelson CJ, Alexova R, Jacoby RP & Millar AH (2014a) Proteins with high turnover rate in barley leaves estimated by proteome analysis combined with in planta isotope labeling. *Plant Physiol.* **166:1** 91–108

Nelson CJ, Huttlin EL, Hegeman AD, Harms AC & Sussman MR (2007) Implications of 15N-metabolic labeling for automated peptide identification in Arabidopsis thaliana. *Proteomics* **7:8** 1279–1292

Nelson CJ, Li L, Jacoby RP & Millar AH (2013) Degradation rate of mitochondrial proteins in Arabidopsis thaliana cells. *J. Proteome Res.* **12:7** 3449–3459

Nelson CJ, Li L & Millar AH (2014b) Quantitative analysis of protein turnover in plants. *Proteomics* **14:4–5** 579–592

Nishimura K, Kato Y & Sakamoto W (2016) Chloroplast Proteases: Updates on Proteolysis within and across Suborganellar Compartments. *Plant Physiol.* **171:4** 2280–2293

Nolte H, Konzer A, Ruhs A, Jungblut B, Braun T & Krüger M (2014) Global protein expression profiling of zebrafish organs based on in vivo incorporation of stable isotopes. *J. Proteome Res.* **13:4** 2162–2174

Okamoto M, Tatematsu K, Matsui A, Morosawa T, Ishida J, Tanaka M, Endo TA, Mochizuki Y, Toyoda T, Kamiya Y, Shinozaki K, Nambara E & Seki M (2010) Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays. *Plant J.* **62:1** 39–51

Ong S-E (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1:5** 376–386

Ong S-E & Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1:6** 2650–60

Payne SH (2015) The utility of protein and mRNA correlation. *Trends Biochem. Sci.* **40:1** 1–3

Petersen TN, Brunak S, von Heijne G & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8:10** 785–786

Pickart CM & Eddins MJ (2004) Ubiquitin: Structures, functions, mechanisms. *Biochim. Biophys. Acta - Mol. Cell Res.* **1695:1–3** 55–72

Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, Rohwer J & Stitt M (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Mol. Syst. Biol.* **5:314** 314

Pitzschke A, Schikora A & Hirt H (2009) MAPK cascade signalling networks in plant defence. *Curr. Opin. Plant Biol.* **12:4** 421–426

Pratt JM, Petty J, Riba-Garcia I, Robertson DHL, Gaskell SJ, Oliver SG & Beynon RJ (2002) Dynamics of Protein Turnover, a Missing Dimension in Proteomics. *Mol. Cell. Proteomics* **1:8** 579–591

Price JC, Guan S, Burlingame A, Prusiner SB & Ghaemmaghami S (2010) Analysis of proteome dynamics in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **107:32** 14508–14513

Rahman M, Previs SF, Kasumov T & Sadygov RG (2016) Gaussian Process Modeling of Protein Turnover. *J. Proteome Res.* **15:7** 2115–2122

Reyes-Turcu FE, Ventii KH & Wilkinson KD (2009) Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu Rev Biochem* **78** 363–397

Reynoso MA, Juntawong P, Lancia M, Blanco FA, Bailey-Serres J & Zanetti ME (2015) Translating ribosome affinity purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translatomes. In *Plant Functional Genomics: Methods and Protocols: Second Edition* pp 185–207.

Robinson M & Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11:3** 1–9

Robinson MD, McCarthy DJ & Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:1** 139–140

Rockel P, Strube F, Rockel A, Wildt J & Kaiser WM (2002) Regulation of nitric oxide (NO) production by plant nitrate reductase in vivo and in vitro. *J. Exp. Bot.* **53:366** 103–10

Ross A, Kessler W, Krumme D, Menge U, Wissing J, Van Den Heuvel J & Flohé L (2004a) Optimised fermentation strategy for 13C/15N recombinant protein labelling in Escherichia coli for NMR-structure analysis. *J. Biotechnol.* **108:1** 31–39

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A & Pappin DJ (2004b) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **3.12:12** 1154–1169

Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C & Mewes HW (2009) CORUM: The comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* **38:1** D497–D501

Russell MR & Lilley KS (2012) Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *J. Proteomics* **77** 441–454

Sakamoto W (2006) Protein Degradation Machineries in Plastids. *Annu. Rev. Plant Biol.* **57:1** 599–621

Schlaeppi K, Abou-Mansour E, Buchala A & Mauch F (2010) Disease resistance of Arabidopsis to Phytophthora brassicae is established by the sequential action of indole glucosinolates and camalexin. *Plant J.* **62:5** 840–851

Schneider DJ & Collmer A (2010) Studying plant-pathogen interactions in the genomics era: beyond molecular Koch's postulates to systems biology. *Annu. Rev. Phytopathol.* **48:1** 457–479

Schütz W, Hausmann N, Krug K, Hampp R & Macek B (2011) Extending SILAC to proteomics of plant cell lines. *Plant Cell* **23:5** 1701–5

Schwab R, Ossowski S, Riester M, Warthmann N & Weigel D (2006) Highly Specific Gene Silencing by Artificial MicroRNAs in *Arabidopsis*. *Plant Cell* **18:5** 1121–1133

Schwämmle V, León IR & Jensen ON (2013) Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates. *J. Proteome Res.* **12:9** 3874–3883

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W & Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:7347** 337–342

Schwanhäusser B, Gossen M, Dittmar G & Selbach M (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9:1** 205–209

Schweizer F, Bodenhausen N, Lassueur S, Masclaux FG & Reymond P (2013) Differential Contribution of Transcription Factors to Arabidopsis thaliana Defense Against Spodoptera littoralis. *Front. Plant Sci.* **4:February** 13

Seong S-Y & Matzinger P (2004) Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nat. Rev. Immunol.* **4:6** 469–478

Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H & Regev A (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498:7453** 236–240

Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, Schwartz S, Fowler B, Weaver S, Wang J, Wang X, Ding R, Raychowdhury R, Friedman N, Hacohen N, … Regev A (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**:**7505** 363–9

Shimamoto K, Okada K & Tabata S (2005) Rice Arabidopsis japonicus Hen - Experimental protocol of model plant 3rd ed. Shujunsha

Silva JC (2005) Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition. *Mol. Cell. Proteomics* **5**:**1** 144–156

Smith DM, Chang SC, Park S, Finley D, Cheng Y & Goldberg AL (2007) Docking of the Proteasomal ATPases' Carboxyl Termini in the 20S Proteasome's α Ring Opens the Gate for Substrate Entry. *Mol. Cell* **27**:**5** 731–744

Smith GK (2005) limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* pp 397–420.

Smith LM & Kelleher NL (2013) Proteoform: a single term describing protein complexity Lloyd. *Nat Methods* **10**:**3** 186–187

Smith SJ, Kroon JTM, Simon WJ, Slabas AR & Chivasa S (2015) A Novel Function for Arabidopsis CYCLASE1 in Programmed Cell Death Revealed by Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) Analysis of Extracellular Matrix Proteins. *Mol. Cell. Proteomics* **14**:**6** 1556–68

Song JT, Lu H & Greenberg JT (2004) Divergent Roles in Arabidopsis thaliana Development and Defense of Two Homologous Genes, ABERRANT GROWTH AND DEATH2 and AGD2-LIKE DEFENSE RESPONSE PROTEIN1, Encoding Novel Aminotransferases. *Plant Cell Online* **16**:**2** 353–366

Sonnhammer EL, von Heijne G & Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings* **6** 175–182

Sorenson R & Bailey-Serres J (2014) Selective mRNA sequestration by OLIGOURIDYLATE-BINDING PROTEIN 1 contributes to translational control during hypoxia in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **111**:**6** 2373–8

de Sousa Abreu R, Penalva LO, Marcotte EM & Vogel C (2009) Global signatures of protein and mRNA expression levels. *Mol BioSyst* **5**:**12** 1512–1526

Staiger D, Zecca L, Wieczorek Kirk DA, Apel K & Eckstein L (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.* **33**:**2** 361–371

Sury MD, Chen JX & Selbach M (2010) The SILAC fly allows for accurate protein quantification in vivo. *Mol Cell Proteomics* **9**:**10** 2173–2183

Swain S, Singh N & Nandi AK (2015) Identification of plant defence regulators through transcriptional profiling of Arabidopsis thaliana cdd1 mutant. *J. Biosci.* **40**:**1** 137–146

Tanz SK, Castleden I, Hooper CM, Vacher M, Small I & Millar HA (2013) SUBA3: A database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. *Nucleic Acids Res.* **41**:**D1** D1185-91

Tavormina P, De Coninck B, Nikonorova N, De Smet I & Cammue BPA (2015) The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. *Plant Cell* **27**:**8** 2095–118

Terashima M, Specht M, Naumann B & Hippler M (2010) Characterizing the anaerobic response of Chlamydomonas reinhardtii by quantitative proteomics. *Mol. Cell. Proteomics* **9**:**7** 1514–32

Thilmony R, Underwood W & He SY (2006) Genome-wide transcriptional analysis of the Arabidopsis thaliana interaction with the plant pathogen Pseudomonas syringae pv. tomato DC3000 and the human pathogen Escherichia coli O157:H7. *Plant J.* **46**:**1** 34–53

Torres Acosta JA, De Almeida Engler J, Raes J, Magyar Z, De Groodt R, Inzé D & De Veylder L (2004) Molecular characterization of Arabidopsis PHO80-like proteins, a novel class of CDKA;1-interacting cyclins. *Cell. Mol. Life Sci.* **61**:**12** 1485–1497

Towill LE & Mazur P (1975) Studies on the reduction of 2,3,5-triphenyltetrazolium chloride as a viability assay for plant tissue cultures. *Can. J. Bot.* **53**:**11** 1097–1102

Toyama BH & Hetzer MW (2013) Protein homeostasis: live long, won't prosper. *Nat. Rev. Mol. Cell Biol.* **14**:**1** 55–61

Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR & Hetzer MW (2013) Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154:5** 971–982

Trotschel C, Albaum SP, Wolff D, Schroder S, Goesmann A, Nattkemper TW & Poetsch A (2012) Protein Turnover Quantification in a Multilabeling Approach: From Data Calculation to Evaluation. *Mol. Cell. Proteomics* **11:8** 512–526

Trouillard R, Hubert-Roux M, Tognetti V, Guilhaudis L, Plasson C, Menu-Bouaouiche L, Coquet L, Guerineau F, Hardouin J, Ele Ekouna JP, Cosette P, Lerouge P, Boitel-Conti M, Afonso C & Ségalas-Milazzo I (2015) Determination of Multimodal Isotopic Distributions: The Case of a ^{15}N Labeled Protein Produced into Hairy Roots. *Anal. Chem.* **87:12** 5938–5946

Turck CW, Webhofer C, Nussbaumer M, Teplytska L, Chen A, Maccarrone G & Filiou MD (2016) Stable isotope metabolic labeling suggests differential turnover of the DPYSL protein family. *Proteomics - Clin. Appl.* **10:12** 1269–1272

Veenstra TD, Martinović S, Anderson GA, Paša-Tolić L & Smith RD (2000) Proteome analysis using selective incorporation of isotopically labeled amino acids

Vierstra RD (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.* **8:3** 135–142

Vierstra RD (2009) The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10:6** 385–97

Viner R, Bomgarden R, Blank M & Rogers J (2013) Increasing the Multiplexing of Protein Quantitation from 6- to 10-Plex with Reporter Ion Isotopologues. *ASMS13 Poster* 1–7

Visscher M, De Henau S, Wildschut MHE, van Es RM, Dhondt I, Michels H, Kemmeren P, Nollen EA, Braeckman BP, Burgering BMT, Vos HR & Dansen TB (2016) Proteome-wide Changes in Protein Turnover Rates in C. elegans Models of Longevity and Age-Related Disease. *Cell Rep.* **16:11** 3041–3051

Vogel C, Abreu R de S, Ko D, Le S-YY, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM & Penalva LO (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. TL - 6. *Mol. Syst. Biol.* **6:400** 400

Wang X & Larkins B (2001) Genetic Analysis of Amino Acid Accumulation in opaque-2 Maize Endosperm. *Plant Physiol.* **125:4** 1766–1777

Wang Z, Han QQ, Zhou MT, Chen X & Guo L (2016) Protein turnover analysis in Salmonella Typhimurium during infection by dynamic SILAC, Topograph, and quantitative proteomics. *J. Basic Microbiol.* **56:7** 801–811

Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J-H, Bantscheff M, … Kuster B (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509:7502** 582–7

Wilk S & Orlowski M (1983) Evidence that Pituitary Cation-Sensitive Neutral Endopeptidase Is a Multicatalytic Protease Complex. *J. Neurochem.* **40:3** 842–849

Wiśniewski JR, Hein MY, Cox J & Mann M (2014) A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13:12** 3497–506

Wisniewski JR, Zougman A, Nagaraj N & Mann M (2009) Universal sample preparation method for proteome analysis. *Nat. Meth.* **6:5** 359–362

Wojtaszek P (1997) Oxidative burst: an early plant response to pathogen infection. *Biochem. J.* **322 (Pt 3:3** 681–92

Xiao Z, Zou Q, Liu Y & Yang X (2016) Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* **7** 11194

Xin X-F, Nomura K, Underwood W & He SY (2013) Induction and suppression of PEN3 focal accumulation during Pseudomonas syringae pv. tomato DC3000 infection of Arabidopsis. *Mol. plant-microbe Interact.* **26:8** 861–7

Xing DH, Lai ZB, Zheng ZY, Vinod KM, Fan BF & Chen ZX (2008) Stress- and pathogen-induced Arabidopsis WRKY48 is a transcriptional activator that represses plant basal defense. *Mol. Plant* **1:3** 459–470

Yang M, Wardzala E, Johal GS & Gray J (2004) The wound-inducible Lls1 gene from maize is an orthologue of the Arabidopsis Acd1 gene, and the LLS1 protein is present in non-photosynthetic tissues. *Plant Mol. Biol.* **54:2** 175–191

Yang XY, Chen WP, Rendahl AK, Hegeman AD, Gray WM & Cohen JD (2010) Measuring the turnover rates of Arabidopsis proteins using deuterium oxide: An auxin signaling case study. *Plant J.* **63:4** 680–695

Young JD, Shastri AA, Stephanopoulos G & Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary 13C flux analysis. *Metab. Eng.* **13:6** 656–665

Zanetti ME (2005) Immunopurification of Polyribosomal Complexes of Arabidopsis for Global Analysis of Gene Expression. *Plant Physiol.* **138:2** 624–635

Zanivan S, Krueger M & Mann M (2012) In Vivo Quantitative Proteomics: The SILAC Mouse. *Methods Mol. Biol.* **757:4** 1–25

Zebell SG & Dong X (2015) Cell-Cycle Regulators and Cell Death in Immunity. *Cell Host Microbe* **18:4** 402–407

Zhao S, Li R, Cai X, Chen W, Li Q, Xing T, Zhu W, Chen YE, Zeng R & Deng Y (2013) The Application of SILAC Mouse in Human Body Fluid Proteomics Analysis Reveals Protein Patterns Associated with IgA Nephropathy. *Evidence-Based Complement. Altern.* **2013:1** 1–10

Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J & Borner T (2012) The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase. *Plant Cell* **24:1** 123–136

Zhou H, Li W, Wang S-P, Mendoza V, Rosa R, Hubert J, Herath K, McLaughlin T, Rohm RJ, Lassman ME, Wong KK, Johns DG, Previs SF, Hubbard BK & Roddy TP (2012) Quantifying apoprotein synthesis in rodents: coupling LC-MS/MS analyses with the administration of labeled water. *J. Lipid Res.* **53:6** 1223–31

Zipfel C (2014) Plant pattern-recognition receptors. *Trends Immunol.* **35:7** 345–351

Zipfel C, Robatzek S, Navarro L, Oakeley E & al et (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **428**:**April** 15–18

Zwickl P, Ng D, Woo KM, Klenk HP & Goldberg AL (1999) An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. *J. Biol. Chem.* **274:37** 26008–26014

10. APPENDIX

10.1 Proteins that change in protein abundances under flg22 treatment

MaxFC describes the highest fold change at any time point between flg22-treated and untreated samples. The p-values are derived from the limma or ANOVA test from chapter 6.7 and are adjusted by FDR. If no p-value is given, then it is was determined to be not significant (> 0.05).

10.1.1 Downregulated Proteins

AGI	maxFC	ANOVA adj. p-value	Limma adj. p-value	Description
AT4G34230	-1.692	2.58E-06	3.41E-13	cinnamyl alcohol dehydrogenase 5
AT1G51630	-0.458	9.94E-03	3.29E-03	O-fucosyltransferase family protein
AT5G61740	-0.541	1.07E-01	2.15E-02	ABC2 homolog 14
AT2G06850	-0.225	2.11E-01	3.86E-02	xyloglucan endotransglucosylase/hydrolase 4
AT3G27400	-0.905	5.28E-02	6.67E-02	Pectin lyase-like superfamily protein
AT3G51670	-0.390	2.17E-01	8.61E-02	SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein
AT2G38530	-0.695	-	9.65E-02	lipid transfer protein 2
AT2G16430	-0.288	2.76E-02	9.65E-02	purple acid phosphatase 10
AT4G31120	-0.763	-	1.01E-01	SHK1 binding protein 1
AT3G15660	-0.320	4.34E-01	1.01E-01	glutaredoxin 4
AT5G03540	-0.396	-	1.06E-01	exocyst subunit exo70 family protein A1
AT5G22740	-0.597	1.54E-01	1.20E-01	cellulose synthase-like A02
AT3G28580	-0.251	-	1.70E-01	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G43160	-0.292	-	1.87E-01	ENTH/VHS family protein
AT1G65150	-0.973	-	2.00E-01	TRAF-like family protein
AT1G09430	-0.169	-	2.36E-01	ATP-citrate lyase A-3
AT3G07540	-0.540	3.78E-01	2.37E-01	Actin-binding FH2 (formin homology 2) family protein
AT4G18030	-0.464	-	2.38E-01	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT4G39320	-0.312	9.82E-02	2.72E-01	microtubule-associated protein-related
AT1G62790	-0.184	-	2.74E-01	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G28050	-0.443	-	2.98E-01	Cytidine/deoxycytidylate deaminase family protein
AT4G25370	-0.466	-	3.00E-01	Double Clp-N motif protein
AT5G51550	-0.415	-	3.00E-01	EXORDIUM like 3
AT5G27670	-0.204	-	3.00E-01	histone H2A 7
AT5G27850	-0.397	-	3.18E-01	Ribosomal protein L18e/L15 superfamily protein
AT1G59610	-0.379	-	3.18E-01	dynamin-like 3

AT4G13200	-0.229	3.82E-01	3.18E-01	unkown protein
AT4G34870	-0.221	-	3.18E-01	rotamase cyclophilin 5
AT3G25150	-0.182	1.45E-01	3.18E-01	Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain
AT5G65760	-0.676	4.26E-01	3.20E-01	Serine carboxypeptidase S28 family protein
AT2G34970	-0.163	4.26E-01	3.20E-01	Trimeric LpxA-like enzyme
AT3G13330	-0.243	-	3.78E-01	proteasome activating protein 200
AT2G16600	-0.245	3.35E-01	3.86E-01	rotamase CYP 3
AT1G26110	-0.170	-	3.86E-01	decapping 5
AT1G53840	-0.521	-	4.03E-01	pectin methylesterase 1
AT2G19570	-0.389	-	4.03E-01	cytidine deaminase 1
AT5G42790	-0.329	2.73E-01	4.03E-01	proteasome alpha subunit F1
AT1G51760	-0.282	-	4.03E-01	peptidase M20/M25/M40 family protein
AT3G49910	-0.219	3.89E-01	4.03E-01	Translation protein SH3-like family protein
AT1G67560	-0.196	-	4.03E-01	PLAT/LH2 domain-containing lipoxygenase family protein
AT1G22530	-0.135	1.99E-01	4.03E-01	PATELLIN 2
AT1G13440	-0.631	-	4.14E-01	glyceraldehyde-3-phosphate dehydrogenase C2
AT2G16570	-0.539	3.14E-01	4.14E-01	GLN phosphoribosyl pyrophosphate amidotransferase 1
AT4G12720	-0.465	-	4.14E-01	MutT/nudix family protein
AT3G49430	-0.339	-	4.14E-01	SER/ARG-rich protein 34A
AT5G17980	-0.244	4.04E-01	4.14E-01	C2 calcium/lipid-binding plant phosphoribosyltransferase family protein
AT2G25270	-0.204	1.73E-01	4.14E-01	unkown protein
AT3G63130	-0.191	-	4.14E-01	RAN GTPase activating protein 1
AT3G16110	-0.177	2.48E-01	4.14E-01	PDI-like 1-6
AT3G08580	-0.153	-	4.14E-01	ADP/ATP carrier 1
AT5G20950	-0.198	-	4.16E-01	Glycosyl hydrolase family protein
AT3G15180	-0.191	-	4.16E-01	ARM repeat superfamily protein
AT5G13050	-0.183	-	4.16E-01	5-formyltetrahydrofolate cycloligase
AT5G11710	-0.302	-	4.26E-01	ENTH/VHS family protein
AT1G80270	-0.187	3.35E-01	4.26E-01	PENTATRICOPEPTIDE REPEAT 596
AT1G72150	-0.135	3.94E-01	4.40E-01	PATELLIN 1
AT4G14300	-0.510	-	4.53E-01	RNA-binding (RRM/RBD/RNP motifs) family protein
AT3G15590	-0.283	-	4.53E-01	Tetratricopeptide repeat (TPR)-like superfamily protein
AT5G61510	-0.167	1.94E-01	4.53E-01	GroES-like zinc-binding alcohol dehydrogenase family protein
AT5G14250	-0.385	-	4.59E-01	Proteasome component (PCI) domain protein
AT1G55150	-0.214	2.58E-01	4.59E-01	DEA(D/H)-box RNA helicase family protein
AT3G08030	-0.559	-	4.61E-01	Protein of unknown function, DUF642
AT1G27500	-0.636	-	4.67E-01	Tetratricopeptide repeat (TPR)-like superfamily protein
AT3G56090	-0.260	-	4.71E-01	ferritin 3
AT1G12920	-0.241	4.26E-01	4.71E-01	eukaryotic release factor 1-2
AT4G39260	-0.159	-	4.71E-01	cold, circadian rhythm, and RNA binding 1
AT5G16070	-0.157	1.92E-01	4.71E-01	TCP-1/cpn60 chaperonin family protein
AT5G65700	-0.420	5.12E-01	4.87E-01	Leucine-rich receptor-like protein kinase family protein
AT5G16970	-0.136	-	5.07E-01	alkenal reductase
AT1G15120	-0.284	3.78E-01	5.08E-01	Ubiquinol-cytochrome C reductase hinge protein
AT3G07410	-0.153	-	5.08E-01	RAB GTPase homolog A5B
AT3G28715	-0.338	-	5.10E-01	ATPase, V0/A0 complex, subunit C/D
AT1G28290	-0.175	-	5.10E-01	arabinogalactan protein 31
AT2G03800	-0.401	4.20E-01	5.12E-01	D-aminoacyl-tRNA deacylases
AT1G71840	-0.306	-	5.12E-01	transducin family protein / WD-40 repeat family protein
AT2G37690	-0.135	-	5.12E-01	phosphoribosylaminoimidazole carboxylase, putative / AIR carboxylase, putative
AT3G61260	-0.133	-	5.12E-01	Remorin family protein
AT2G31660	-0.689	3.28E-01	5.17E-01	ARM repeat superfamily protein

AT4G26900	-0.165	-	5.19E-01	HIS HF
AT4G31460	-0.323	4.53E-01	5.25E-01	Ribosomal L28 family
AT2G36070	-0.237	2.36E-01	5.25E-01	translocase inner membrane subunit 44-2
AT5G47760	-0.181	-	5.25E-01	2-phosphoglycolate phosphatase 2
AT3G01640	-0.147	-	5.36E-01	glucuronokinase G
AT1G67360	-0.413	4.74E-01	5.39E-01	Rubber elongation factor protein (REF)
AT3G44300	-0.318	-	5.39E-01	nitrilase 2
AT1G44790	-0.556	3.66E-01	-	ChaC-like family protein
AT1G64490	-0.526	3.94E-01	-	DEK, chromatin associated protein
AT1G22200	-0.479	3.22E-01	-	Endoplasmic reticulum vesicle transporter protein
AT3G07170	-0.432	3.76E-01	-	Sterile alpha motif (SAM) domain-containing protein
AT4G09150	-0.381	5.26E-01	-	T-complex protein 11
AT5G56760	-0.342	5.10E-01	-	serine acetyltransferase 1
AT5G43980	-0.320	4.19E-01	-	plasmodesmata-located protein 1
AT5G11240	-0.316	2.84E-01	-	transducin family protein / WD-40 repeat family protein
AT2G27880	-0.309	3.79E-01	-	Argonaute family protein
AT3G49870	-0.299	4.63E-01	-	ADP-ribosylation factor-like A1C
AT2G37130	-0.297	4.53E-01	-	Peroxidase superfamily protein
AT2G22450	-0.281	4.26E-01	-	riboflavin biosynthesis protein, putative
AT5G52520	-0.277	4.83E-01	-	Class II aaRS and biotin synthetases superfamily protein
AT1G51570	-0.263	1.79E-01	-	Calcium-dependent lipid-binding (CaLB domain) plant phosphoribosyltransferase family protein
AT1G36280	-0.258	4.52E-01	-	L-Aspartase-like family protein
AT4G34700	-0.228	3.82E-01	-	LYR family of Fe/S cluster biogenesis protein
AT2G01600	-0.228	3.78E-01	-	ENTH/ANTH/VHS superfamily protein
AT2G24940	-0.224	2.65E-01	-	membrane-associated progesterone binding protein 2
AT3G10410	-0.217	4.63E-01	-	SERINE CARBOXYPEPTIDASE-LIKE 49
AT5G15400	-0.216	4.63E-01	-	U-box domain-containing protein
AT1G16870	-0.213	2.80E-01	-	mitochondrial 28S ribosomal protein S29-related
AT2G18110	-0.206	3.46E-01	-	Translation elongation factor EF1B/ribosomal protein S6 family protein
AT4G05420	-0.205	3.57E-01	-	damaged DNA binding protein 1A
AT4G15640	-0.186	1.99E-01	-	unkown protein
AT5G64270	-0.182	2.86E-01	-	splicing factor, putative
AT2G25080	-0.181	2.29E-01	-	glutathione peroxidase 1
AT5G42020	-0.179	4.94E-01	-	Heat shock protein 70 (Hsp 70) family protein
AT1G73430	-0.178	4.52E-01	-	sec34-like family protein
AT5G11340	-0.177	5.12E-01	-	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT1G12410	-0.170	3.22E-01	-	CLP protease proteolytic subunit 2
AT1G50940	-0.165	4.63E-01	-	electron transfer flavoprotein alpha
AT4G25630	-0.161	3.82E-01	-	fibrillarin 2
AT2G04280	-0.156	4.20E-01	-	unkown protein
AT4G09720	-0.156	3.78E-01	-	RAB GTPase homolog G3A
AT3G10850	-0.153	3.94E-01	-	Metallo-hydrolase/oxidoreductase superfamily protein
AT1G11680	-0.152	3.87E-01	-	CYTOCHROME P450 51G1
AT5G43280	-0.143	5.04E-01	-	delta(3,5),delta(2,4)-dienoyl-CoA isomerase 1
AT3G01740	-0.139	4.26E-01	-	Mitochondrial ribosomal protein L37
AT4G34740	-0.133	3.46E-01	-	GLN phosphoribosyl pyrophosphate amidotransferase 2
AT4G35470	-0.132	1.63E-01	-	plant intracellular ras group-related LRR 4
AT1G70710	-0.128	4.74E-01	-	glycosyl hydrolase 9B1
AT2G40290	-0.121	5.22E-01	-	Eukaryotic translation initiation factor 2 subunit 1
AT1G08680	-0.121	4.34E-01	-	ARF GAP-like zinc finger-containing protein ZIGA4
AT5G52310	-0.118	2.05E-01	-	low-temperature-responsive protein 78 (LTI78) / desiccation-responsive protein 29A (RD29A)
AT2G45240	-0.117	4.34E-01	-	methionine aminopeptidase 1A

AT5G47030	-0.117	3.66E-01	-	ATPase, F1 complex, delta/epsilon subunit
AT3G26080	-0.117	4.74E-01	-	plastid-lipid associated protein PAP / fibrillin family protein
AT4G13850	-0.115	3.66E-01	-	glycine-rich RNA-binding protein 2
AT3G60980	-0.112	5.26E-01	-	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G62390	-0.110	3.98E-01	-	Octicosapeptide/Phox/Bem1p (PB1) domain-containing protein / tetratricopeptide repeat (TPR)-containing protein
AT1G25490	-0.109	4.53E-01	-	ARM repeat superfamily protein
AT3G60960	-0.106	2.80E-01	-	Tetratricopeptide repeat (TPR)-like superfamily protein
AT5G52470	-0.105	5.01E-01	-	fibrillarin 1
AT4G20360	-0.100	5.26E-01	-	RAB GTPase homolog E1B
AT1G04810	-0.097	5.12E-01	-	26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit
AT4G02580	-0.097	4.94E-01	-	NADH-ubiquinone oxidoreductase 24 kDa subunit, putative
AT4G36680	-0.095	3.82E-01	-	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G20440	-0.095	4.98E-01	-	cold-regulated 47
AT3G12260	-0.094	4.83E-01	-	LYR family of Fe/S cluster biogenesis protein
AT5G48480	-0.091	4.53E-01	-	Lactoylglutathione lyase / glyoxalase I family protein
AT4G29840	-0.091	3.87E-01	-	Pyridoxal-5'-phosphate-dependent enzyme family protein
AT3G12780	-0.090	2.86E-01	-	phosphoglycerate kinase 1
AT1G42960	-0.084	3.22E-01	-	;expressed protein localized to the inner membrane of the chloroplast.
AT5G16990	-0.078	2.73E-01	-	Zinc-binding dehydrogenase family protein
AT1G09760	-0.076	1.78E-01	-	U2 small nuclear ribonucleoprotein A
ATCG00500	-0.075	4.34E-01	-	acetyl-CoA carboxylase carboxyl transferase subunit beta
AT3G21110	-0.069	3.94E-01	-	purin 7
AT1G18070	-0.059	3.44E-01	-	Translation elongation factor EF1A/initiation factor IF2gamma family protein
AT2G21870	-0.056	4.63E-01	-	copper ion binding
AT3G18190	-0.054	4.35E-01	-	TCP-1/cpn60 chaperonin family protein
AT3G23990	-0.052	2.30E-01	-	heat shock protein 60

10.1.2 Upregulated Proteins

AGI	maxFC	ANOVA adj. p-value	Limma adj. p-value	Description
AT2G30490	0.971	7.97E-03	1.28E-09	cinnamate-4-hydroxylase
AT4G39980	0.736	4.77E-03	1.28E-09	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1
AT4G31500	1.060	2.00E-03	1.50E-07	cytochrome P450, family 83, subfamily B, polypeptide 1
AT1G22410	0.835	4.77E-03	6.35E-06	Class-II DAHP synthetase family protein
AT1G02930	0.676	1.04E-02	6.35E-06	glutathione S-transferase 6
AT2G38860	0.962	9.55E-03	1.07E-05	Class I glutamine amidotransferase-like superfamily protein
AT4G08850	0.785	2.00E-03	1.76E-05	Leucine-rich repeat receptor-like protein kinase family protein
AT2G39518	1.133	4.77E-03	3.44E-05	Uncharacterised protein family (UPF0497)
AT1G02920	1.358	9.82E-02	1.12E-04	glutathione S-transferase 7
AT1G52760	0.683	9.72E-03	1.50E-04	lysophospholipase 2
AT2G30860	0.455	3.03E-01	2.77E-04	glutathione S-transferase PHI 9
AT2G04400	0.260	2.58E-02	4.60E-04	Aldolase-type TIM barrel family protein
AT2G15620	0.448	2.86E-01	4.65E-04	nitrite reductase 1
AT3G02360	0.287	5.22E-03	4.65E-04	6-phosphogluconate dehydrogenase family protein
AT3G03780	0.385	2.58E-02	5.79E-04	methionine synthase 2
AT2G36880	0.522	9.94E-03	7.65E-04	methionine adenosyltransferase 3
AT1G59870	0.396	7.97E-03	7.91E-04	ABC-2 and Plant PDR ABC-type transporter family protein
AT4G30530	0.679	7.25E-02	1.46E-03	Class I glutamine amidotransferase-like superfamily protein
AT5G11670	0.344	9.72E-03	3.29E-03	NADP-malic enzyme 2
AT2G32240	0.217	5.70E-03	3.97E-03	unkown protein

AT4G20830	0.653	4.77E-03	6.80E-03	FAD-binding Berberine family protein
AT4G30190	0.348	1.76E-01	1.78E-02	H(+)-ATPase 2
AT1G80460	0.177	1.78E-01	1.78E-02	Actin-like ATPase superfamily protein
AT2G20610	0.371	2.58E-02	2.23E-02	Tyrosine transaminase family protein
AT1G78850	0.671	2.48E-01	2.77E-02	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain
AT3G16470	0.382	7.62E-02	2.77E-02	Mannose-binding lectin superfamily protein
AT4G39230	0.222	5.22E-03	2.94E-02	NmrA-like negative transcriptional regulator family protein
AT3G54640	0.505	5.28E-02	3.26E-02	tryptophan synthase alpha chain
AT1G48850	0.281	6.12E-02	3.26E-02	chorismate synthase, putative / 5-enolpyruvylshikimate-3-phosphate phospholyase, putative
AT1G33590	0.521	1.61E-01	4.42E-02	Leucine-rich repeat (LRR) family protein
AT3G14840	0.890	1.06E-01	5.31E-02	Leucine-rich repeat transmembrane protein kinase
AT5G54810	0.600	6.12E-02	5.38E-02	tryptophan synthase beta-subunit 1
AT3G27960	0.620	-	5.70E-02	Tetratricopeptide repeat (TPR)-like superfamily protein
AT3G27310	0.865	-	5.72E-02	plant UBX domain-containing protein 1
AT1G10370	0.823	3.53E-01	5.72E-02	Glutathione S-transferase family protein
AT2G45290	0.165	7.62E-02	5.72E-02	Transketolase
AT5G63620	0.891	3.78E-01	6.12E-02	GroES-like zinc-binding alcohol dehydrogenase family protein
AT3G45960	0.217	9.72E-03	6.12E-02	expansin-like A3
AT1G74010	0.346	2.05E-01	6.15E-02	Calcium-dependent phosphotriesterase superfamily protein
AT3G23570	0.199	8.34E-02	7.54E-02	alpha/beta-Hydrolases superfamily protein
AT5G07440	0.155	6.68E-02	8.81E-02	glutamate dehydrogenase 2
AT1G06550	0.972	1.14E-01	9.07E-02	ATP-dependent caseinolytic (Clp) protease/crotonase family protein
AT4G08770	0.639	4.52E-01	9.07E-02	Peroxidase superfamily protein
AT1G07750	0.187	-	9.07E-02	RmlC-like cupins superfamily protein
AT2G30870	0.162	1.99E-01	9.07E-02	glutathione S-transferase PHI 10
AT1G30510	0.258	-	9.36E-02	root FNR 2
AT3G61820	0.462	2.09E-01	1.01E-01	Eukaryotic aspartyl protease family protein
AT3G12700	0.309	2.92E-02	1.01E-01	Eukaryotic aspartyl protease family protein
AT1G15950	1.262	1.78E-01	1.06E-01	cinnamoyl coa reductase 1
AT1G70980	0.801	-	1.06E-01	Class II aminoacyl-tRNA and biotin synthetases superfamily protein
AT1G74100	0.366	9.82E-02	1.06E-01	sulfotransferase 16
AT5G17530	0.275	2.58E-02	1.06E-01	phosphoglucosamine mutase family protein
ATMG00090	0.329	-	1.09E-01	structural constituent of ribosome
AT3G01290	0.229	7.49E-02	1.46E-01	SPFH/Band 7/PHB domain-containing membrane-associated protein family
AT3G22890	0.229	1.04E-01	1.46E-01	ATP sulfurylase 1
AT2G47960	0.204	-	1.46E-01	unkown protein
AT4G27070	0.275	1.65E-02	1.55E-01	tryptophan synthase beta-subunit 2
AT3G10220	0.510	-	1.66E-01	tubulin folding cofactor B
AT4G27500	0.168	1.02E-02	1.84E-01	proton pump interactor 1
AT2G35790	0.987	-	1.87E-01	unkown protein
AT3G23710	0.312	-	1.87E-01	Tic22-like family protein
AT5G17380	0.151	-	1.87E-01	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein
AT5G47700	0.656	-	1.88E-01	60S acidic ribosomal protein family
AT3G13150	0.342	-	1.88E-01	Tetratricopeptide repeat (TPR)-like superfamily protein
AT3G52400	0.438	6.12E-02	1.91E-01	syntaxin of plants 122
AT1G37130	0.469	5.28E-02	2.21E-01	nitrate reductase 2
AT1G69410	0.375	-	2.21E-01	eukaryotic elongation factor 5A-3
AT2G20940	0.235	-	2.36E-01	Protein of unknown function (DUF1279)
AT5G01750	0.415	3.46E-01	2.37E-01	Protein of unknown function (DUF567)
AT4G02520	0.460	-	2.62E-01	glutathione S-transferase PHI 2

AT5G07340	0.210	2.84E-01	2.96E-01	Calreticulin family protein
AT3G51430	0.656	-	3.00E-01	Calcium-dependent phosphotriesterase superfamily protein
AT1G25220	0.384	6.12E-02	3.00E-01	anthranilate synthase beta subunit 1
AT4G39640	0.345	-	3.00E-01	gamma-glutamyl transpeptidase 1
AT5G24300	0.279	-	3.00E-01	Glycogen/starch synthases, ADP-glucose type
AT3G06350	0.226	1.63E-01	3.00E-01	dehydroquinate dehydratase, putative / shikimate dehydrogenase, putative
AT2G18040	0.195	-	3.00E-01	peptidylprolyl cis/trans isomerase, NIMA-interacting 1
AT2G13272	0.589	-	3.18E-01	unkown protein
AT5G45280	0.352	4.26E-01	3.18E-01	Pectinacetylesterase family protein
AT3G59970	0.137	2.58E-02	3.18E-01	methylenetetrahydrofolate reductase 1
AT1G25570	0.202	-	3.28E-01	unkown protein
AT3G29635	0.234	-	3.38E-01	HXXXD-type acyl-transferase family protein
AT3G46460	0.587	-	3.43E-01	ubiquitin-conjugating enzyme 13
AT3G54400	0.370	2.05E-01	3.44E-01	Eukaryotic aspartyl protease family protein
AT4G17260	0.960	-	3.86E-01	Lactate/malate dehydrogenase family protein
AT3G50930	0.595	9.19E-02	3.86E-01	cytochrome BC1 synthesis
AT2G44450	0.407	1.99E-01	3.86E-01	beta glucosidase 15
AT1G74060	0.141	-	3.86E-01	Ribosomal protein L6 family protein
AT5G06870	0.167	-	4.02E-01	polygalacturonase inhibiting protein 2
AT1G22840	0.278	4.18E-01	4.02E-01	CYTOCHROME C-1
AT1G08640	0.562	6.81E-02	4.03E-01	Chloroplast J-like domain 1
AT1G70320	0.224	-	4.03E-01	ubiquitin-protein ligase 2
AT4G35090	0.621	-	4.14E-01	catalase 2
AT4G37640	0.300	2.51E-01	4.14E-01	calcium ATPase 2
AT1G27130	0.276	2.80E-01	4.14E-01	glutathione S-transferase tau 13
AT5G64350	0.234	-	4.14E-01	FK506-binding protein 12
AT1G53900	0.228	-	4.14E-01	Eukaryotic translation initiation factor 2B (eIF-2B) family protein
AT1G14650	0.219	3.66E-01	4.14E-01	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein / ubiquitin family protein
AT4G01850	0.170	8.34E-02	4.14E-01	S-adenosylmethionine synthetase 2
AT5G65020	0.160	5.28E-02	4.14E-01	annexin 2
AT4G26690	0.124	3.66E-02	4.14E-01	PLC-like phosphodiesterase family protein
AT4G34200	0.111	-	4.14E-01	D-3-phosphoglycerate dehydrogenase
AT4G30280	1.334	2.73E-01	4.16E-01	xyloglucan endotransglucosylase/hydrolase 18
AT5G37740	0.588	-	4.18E-01	Calcium-dependent lipid-binding (CaLB domain) family protein
AT4G17770	0.500	-	4.18E-01	trehalose phosphatase/synthase 5
AT3G28940	0.151	9.57E-02	4.18E-01	AIG2-like (avirulence induced gene) family protein
AT3G56070	0.161	-	4.40E-01	rotamase cyclophilin 2
AT5G17770	0.240	-	4.40E-01	NADH:cytochrome B5 reductase 1
AT3G05280	1.084	-	4.41E-01	Integral membrane Yip1 family protein
AT2G05920	0.298	3.87E-01	4.53E-01	Subtilase family protein
AT4G05400	0.220	-	4.53E-01	copper ion binding
AT1G70410	0.158	2.86E-01	4.53E-01	beta carbonic anhydrase 4
AT2G22250	0.130	2.80E-01	4.57E-01	aspartate aminotransferase
AT5G48020	0.169	-	4.59E-01	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT1G74020	0.245	-	4.59E-01	strictosidine synthase 2
AT5G58110	0.336	-	4.61E-01	chaperone binding
AT1G53070	0.633	-	4.71E-01	Legume lectin family protein
AT4G23710	0.411	4.99E-01	4.71E-01	vacuolar ATP synthase subunit G2
AT2G22125	0.292	-	4.71E-01	CELLULOSE SYNTHASE-INTERACTIVE PROTEIN 1
AT2G21250	0.238	2.53E-01	4.71E-01	NAD(P)-linked oxidoreductase superfamily protein
AT3G59350	0.385	-	4.76E-01	Protein kinase superfamily protein

AT1G44130	0.446	-	4.78E-01	Eukaryotic aspartyl protease family protein
AT2G46880	0.317	-	4.92E-01	purple acid phosphatase 14
AT3G44100	0.229	-	4.94E-01	MD-2-related lipid recognition domain-containing protein
AT1G79530	0.123	-	4.94E-01	glyceraldehyde-3-phosphate dehydrogenase of plastid 1
AT5G56360	0.646	3.57E-01	5.00E-01	calmodulin-binding protein
AT1G12470	0.411	-	5.00E-01	zinc ion binding
AT3G09260	0.319	-	5.08E-01	Glycosyl hydrolase superfamily protein
AT1G60710	0.136	-	5.09E-01	NAD(P)-linked oxidoreductase superfamily protein
AT1G78830	0.372	-	5.10E-01	Curculin-like (mannose-binding) lectin family protein
AT3G57530	0.483	-	5.12E-01	calcium-dependent protein kinase 32
AT1G24100	0.412	4.62E-01	5.12E-01	UDP-glucosyl transferase 74B1
AT5G18400	0.363	4.19E-01	5.12E-01	Cytokine-induced anti-apoptosis inhibitor 1, Fe-S biogenesis
AT3G12145	0.317	5.22E-01	5.12E-01	Leucine-rich repeat (LRR) family protein
AT1G08190	0.281	-	5.12E-01	vacuolar protein sorting 41
AT5G35700	0.281	-	5.12E-01	fimbrin-like protein 2
AT4G19210	0.244	-	5.12E-01	RNAse l inhibitor protein 2
AT3G28930	0.217	-	5.12E-01	AIG2-like (avirulence induced gene) family protein
AT4G12470	0.785	-	5.25E-01	azelaic acid induced 1
AT1G02500	0.358	4.52E-01	5.25E-01	S-adenosylmethionine synthetase 1
AT4G13930	0.134	2.73E-01	5.25E-01	serine hydroxymethyltransferase 4
AT1G55490	0.325	-	5.36E-01	chaperonin 60 beta

10.2 Proteins with quicker turnover and steady protein abundance

The relative size of the increased turnover is described by the "Effect" category. The protein were characterized with the following prediction tools: Subcellular localization with SUBA3 (Tanz et al, 2013), presence of a signal peptide with signalP4.1 (Petersen *et al*, 2011) and the number of transmembrane helices (TMH) with TMHMM (Sonnhammer *et al*, 1998).

AGI	Effect	Comment	SUBA consensus	Signal peptide	TMH	Description
AT1G03220	+++		extracellular	yes	0	Eukaryotic aspartyl protease family protein
AT1G35140	+++		extracellular	yes	0	Phosphate-responsive 1 family protein
AT1G45145	+++		cytosol	no	0	thioredoxin H-type 5
AT1G55450	+++	sparse protein data	cytosol	no	0	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G24200	+++		cytosol	no	0	Cytosol aminopeptidase family protein
AT2G39310	+++	effect at later time point	cytosol	no	0	jacalin-related lectin 22
AT3G09940	+++		cytosol	no	0	monodehydroascorbate reductase
AT3G16420	+++		cytosol	no	0	PYK10-binding protein 1
AT3G16460	+++		cytosol	no	0	Mannose-binding lectin superfamily protein
AT3G25780	+++		plastid	no	0	allene oxide cyclase 3
AT4G11850	+++		cytosol	no	0	phospholipase D gamma 1
AT4G16760	+++		peroxisome	no	0	acyl-CoA oxidase 1
AT4G18950	+++		cytosol, nucleus	no	0	Integrin-linked protein kinase family
AT4G30210	+++		endoplasmic reticulum	no	1	P450 reductase 2
AT4G34150	+++		cytosol	no	0	Calcium-dependent lipid-binding (CaLB domain) family protein
AT4G35630	+++		plastid	no	0	phosphoserine aminotransferase
AT4G35830	+++		cytosol, mitochondrion	no	0	aconitase 1

AT5G05730	+++		plastid	no	0	anthranilate synthase alpha subunit 1
AT5G11520	+++	RNA small change	plastid	no	0	aspartate aminotransferase 3
AT5G44400	+++		extracellular	yes	0	FAD-binding Berberine family protein
AT1G05620	++	burst	cytosol	no	0	uridine-ribohydrolase 2
AT1G06840	++	sparse protein data	plasma membrane	yes	2	Leucine-rich repeat protein kinase family protein
AT1G07890	++	burst	cytosol	no	0	ascorbate peroxidase 1
AT1G17745	++	RNA small change	plastid	no	0	D-3-phosphoglycerate dehydrogenase
AT1G18210	++	burst	nucleus	no	0	Calcium-binding EF-hand family protein
AT1G19130	++	burst	cytosol	no	0	unkown protein
AT1G20510	++	burst	plastid	no	2	OPC-8:0 CoA ligase1
AT1G21750	++	burst	endoplasmic reticulum	yes	0	PDI-like 1-1
AT1G22360	++	sparse protein data	cytosol	no	0	UDP-glucosyl transferase 85A2
AT1G78380	++	burst at late time point	cytosol	no	0	glutathione S-transferase TAU 19
AT1G80300	++	sparse protein data	plastid	no	9	nucleotide transporter 1
AT1G80360	++		mitochondrion	no	0	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
AT2G15490	++	protein level rise	plasma membrane	no	0	UDP-glycosyltransferase 73B4
AT2G17720	++	sparse protein data	mitochondrion	no	1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT2G18690	++	sparse protein data	plasma membrane	no	5	unkown protein
AT2G21620	++	burst	cytosol	no	0	Adenine nucleotide alpha hydrolases-like superfamily protein
AT2G30140	++		cytosol	no	0	UDP-Glycosyltransferase superfamily protein
AT2G38940	++		plasma membrane	no	11	phosphate transporter 1
AT2G41380	++		cytosol	no	0	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G45300	++		plastid	no	0	RNA 3'-terminal phosphate cyclase/enolpyruvate transferase, alpha/beta
AT3G06650	++		mitochondrion	no	0	ATP-citrate lyase B-1
AT3G08640	++		plastid	no	2	Protein of unknown function (DUF3411)
AT3G11820	++		plasma membrane	no	1	syntaxin of plants 121
AT3G14990	++	RNA unchanged	cytosol	no	0	Class I glutamine amidotransferase-like superfamily protein
AT3G16400	++	sparse protein data	cytosol	no	0	nitrile specifier protein 1
AT3G19930	++		plasma membrane	no	11	sugar transporter 4
AT3G22850	++		cytosol	no	0	Aluminium induced protein with YGL and LRDR motifs
AT3G28200	++	RNA unchanged	extracellular	yes	0	Peroxidase superfamily protein
AT3G28450	++		plasma membrane	yes	2	Leucine-rich repeat protein kinase family protein
AT3G48000	++		mitochondrion	no	0	aldehyde dehydrogenase 2B4
AT3G49120	++		extracellular	yes	0	peroxidase CB
AT3G52850	++	sparse protein data	golgi	yes	1	vacuolar sorting receptor homolog 1
AT3G53180	++		cytosol	no	0	glutamate-ammonia ligases
AT3G57330	++		endoplasmic reticulum	no	10	autoinhibited Ca2+-ATPase 11

AT3G58730	++	RNA unchanged	vacuole, golgi	no	0 vacuolar ATP synthase subunit D (VATD) / V-ATPase D subunit / vacuolar proton pump D subunit (VATPD)
AT3G59760	++		plastid	no	0 O-acetylserine (thiol) lyase isoform C
AT3G61440	++		mitochondrion	no	0 cysteine synthase C1
AT4G14880	++	burst	cytosol	no	0 O-acetylserine (thiol) lyase (OAS-TL) isoform A1
AT4G19880	++	RNA small change	plastid	no	0 Glutathione S-transferase family protein
AT4G23100	++	burst	plastid	no	0 glutamate-cysteine ligase
AT4G23850	++	burst	peroxisome	no	0 AMP-dependent synthetase and ligase family protein
AT4G25900	++	changing protein levels	extracellular	yes	1 Galactose mutarotase-like superfamily protein
AT4G29900	++		plasma membrane	no	8 autoinhibited Ca(2+)-ATPase 10
AT4G34050	++	changing protein levels	cytosol	no	0 S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT4G34131	++	sparse protein data	plasma membrane	no	0 UDP-glucosyl transferase 73B3
AT4G34180	++	burst	extracellular	yes	0 Cyclase family protein
AT4G37370	++	burst	endoplasmic reticulum	no	1 cytochrome P450, family 81, subfamily D, polypeptide 8
AT4G38420	++	burst	extracellular	yes	1 SKU5 similar 9
AT5G01500	++	burst	plastid	no	0 thylakoid ATP/ADP carrier
AT5G03160	++	burst	endoplasmic reticulum	no	1 homolog of mamallian P58IPK
AT5G03630	++		cytosol	no	0 Pyridine nucleotide-disulphide oxidoreductase family protein
AT5G04590	++	burst	plastid	no	0 sulfite reductase
AT5G04740	++	burst	plastid	no	0 ACT domain-containing protein
AT5G06320	++		plasma membrane	no	1 NDR1/HIN1-like 3
AT5G06860	++	changing protein levels	extracellular	yes	0 polygalacturonase inhibiting protein 1
AT5G13200	++	burst	cytosol	no	0 GRAM domain family protein
AT5G13420	++	burst	plastid	no	0 Aldolase-type TIM barrel family protein
AT5G17330	++	burst	cytosol	no	0 glutamate decarboxylase
AT5G17990	++	sparse protein data	plastid	no	0 tryptophan biosynthesis 1
AT5G19440	++	burst	cytosol	no	0 NAD(P)-binding Rossmann-fold superfamily protein
AT5G21105	++		extracellular	yes	1 Plant L-ascorbate oxidase
AT5G24760	++	sparse protein data	cytosol	no	0 GroES-like zinc-binding dehydrogenase family protein
AT5G26030	++	burst	plastid	no	1 ferrochelatase 1
AT5G27380	++	burst	cytosol, plastid	no	0 glutathione synthetase 2
AT5G39950	++	burst	cytosol	no	0 thioredoxin 2
AT5G40370	++	burst	cytosol	no	1 Glutaredoxin family protein
AT5G40760	++	burst	cytosol	no	0 glucose-6-phosphate dehydrogenase 6
AT5G44720	++	burst	cytosol	no	0 Molybdenum cofactor sulfurase family protein
AT5G45510	++	sparse protein data	cytosol	no	0 Leucine-rich repeat (LRR) family protein
AT5G48810	++	burst	endoplasmic reticulum	no	1 cytochrome B5 isoform D
AT5G48930	++	sparse protein data	cytosol	no	0 hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
AT5G54500	++		cytosol	no	1 flavodoxin-like quinone reductase 1
AT5G58590	++	sparse protein data	nucleus	no	0 RAN binding protein 1

AGI						Description
AT5G63570	++	burst	plastid	no	0	glutamate-1-semialdehyde-2,1-aminomutase
AT5G66120	++	burst	cytosol	no	0	3-dehydroquinate synthase, putative
AT1G06290	+	RNA unchanged	mitochondrion	no	0	acyl-CoA oxidase 3
AT1G07510	+	sparse protein data	mitochondrion	no	0	FTSH protease 10
AT1G11910	+	RNA down-regulated	vacuole	yes	1	aspartic proteinase A1
AT1G15130	+		nucleus	no	0	Endosomal targeting BRO1-like domain-containing protein
AT1G27980	+		endoplasmic reticulum	no	1	dihydrosphingosine phosphate lyase
AT1G63010	+		vacuole	no	10	Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein
AT1G65820	+		extracellular	no	3	microsomal glutathione s-transferase, putative
AT1G65930	+		cytosol	no	0	cytosolic NADP+-dependent isocitrate dehydrogenase
AT1G66240	+		plasma membrane	no	0	homolog of anti-oxidant 1
AT2G44160	+		cytosol	no	0	methylenetetrahydrofolate reductase 2
AT2G47000	+		plasma membrane	no	9	ATP binding cassette subfamily B4
AT3G02780	+		plastid	no	0	isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase 2
AT3G07680	+		extracellular	yes	0	emp24/gp25L/p24 family/GOLD family protein
AT3G12800	+		peroxisome	no	0	short-chain dehydrogenase-reductase B
AT3G22950	+		cytosol	no	0	ADP-ribosylation factor C1
AT3G24170	+		peroxisome	no	0	glutathione-disulfide reductase
AT3G52470	+		plasma membrane	no	1	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
AT3G54030	+		cytosol, mitochondrion	no	0	Protein kinase protein with tetratricopeptide repeat domain
AT4G05020	+		mitochondrion	no	0	NAD(P)H dehydrogenase B2
AT4G05160	+		peroxisome	no	0	AMP-dependent synthetase and ligase family protein
AT4G37870	+	RNA down-regulated	cytosol	no	0	phosphoenolpyruvate carboxykinase 1
AT5G09650	+		plastid	no	0	pyrophosphorylase 6
AT5G26742	+		plastid	no	0	DEAD box RNA helicase (RH3)
AT5G50920	+		plastid	no	0	CLPC homologue 1
AT5G66920	+		extracellular	yes	1	SKU5 similar 17

10.3 Proteins with low labeling efficiency

These proteins have on average a labeling efficiency of less than 40%, or in other words have a RIA_{Lys0} > 60%. RIA_{Lys0} = Lys0 / (Lys0+Lys4+Lys8); nQuant = Number of samples where that protein was quantified.

AGI	Median RIA_{Lys0}	nQuant	Description
AT1G06680	89.92%	23	photosystem II subunit P-1
AT1G07320	79.04%	21	ribosomal protein L4
AT1G09310	84.37%	7	Protein of unknown function, DUF538
AT1G09340	88.08%	23	chloroplast RNA binding
AT1G12900	89.31%	10	glyceraldehyde 3-phosphate dehydrogenase A subunit 2
AT1G15820	98.91%	21	light harvesting complex photosystem II subunit 6
AT1G16080	61.07%	11	unknown protein
AT1G20020	90.24%	14	ferredoxin-NADP(+)-oxidoreductase 2

AT1G23740	80.91%	7	Oxidoreductase, zinc-binding dehydrogenase family protein
AT1G24020	91.79%	7	MLP-like protein 423
AT1G31330	90.78%	26	photosystem I subunit F
AT1G32060	86.56%	22	phosphoribulokinase
AT1G32990	63.99%	15	plastid ribosomal protein l11
AT1G42970	91.63%	25	glyceraldehyde-3-phosphate dehydrogenase B subunit
AT1G44575	92.68%	22	Chlorophyll A-B binding family protein
AT1G52030	83.25%	8	myrosinase-binding protein 2
AT1G52400	85.18%	11	beta glucosidase 18
AT1G54780	90.54%	6	thylakoid lumen 18.3 kDa protein
AT1G55480	64.73%	5	protein containing PDZ domain, a K-box domain, and a TPR region
AT1G61520	90.78%	26	photosystem I light harvesting complex gene 3
AT1G67090	92.30%	35	ribulose bisphosphate carboxylase small chain 1A
AT1G68010	87.55%	15	hydroxypyruvate reductase
AT1G70830	85.20%	8	MLP-like protein 28
AT1G74470	89.23%	9	Pyridine nucleotide-disulphide oxidoreductase family protein
AT2G05100	90.62%	6	photosystem II light harvesting complex gene 2.1
AT2G10940	91.19%	17	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G13360	95.62%	12	alanine:glyoxylate aminotransferase
AT2G24270	86.96%	8	aldehyde dehydrogenase 11A3
AT2G28900	92.92%	11	outer plastid envelope protein 16-1
AT2G34420	88.24%	5	photosystem II light harvesting complex gene B1B2
AT2G37660	80.61%	7	NAD(P)-binding Rossmann-fold superfamily protein
AT2G39730	89.43%	30	rubisco activase
AT2G43030	90.62%	7	Ribosomal protein L3 family protein
AT3G01090	78.60%	17	SNF1 kinase homolog 10
AT3G08940	80.44%	28	light harvesting complex photosystem II
AT3G14067	66.52%	9	Subtilase family protein
AT3G14210	86.50%	18	epithiospecifier modifier 1
AT3G14415	87.60%	12	Aldolase-type TIM barrel family protein
AT3G14420	91.60%	12	Aldolase-type TIM barrel family protein
AT3G16140	92.31%	7	photosystem I subunit H-1
AT3G16370	83.44%	5	GDSL-like Lipase/Acylhydrolase superfamily protein
AT3G26060	83.78%	12	Thioredoxin superfamily protein
AT3G26650	92.85%	21	glyceraldehyde 3-phosphate dehydrogenase A subunit
AT3G27850	70.97%	29	ribosomal protein L12-C
AT3G44890	78.21%	7	ribosomal protein L9
AT3G45140	91.16%	22	lipoxygenase 2
AT3G46780	90.62%	14	plastid transcriptionally active 16
AT3G47470	89.98%	20	light-harvesting chlorophyll-protein complex I subunit A4
AT3G54050	82.04%	10	high cyclic electron flow 1
AT3G54890	94.70%	9	photosystem I light harvesting complex gene 1
AT3G55800	90.94%	16	sedoheptulose-bisphosphatase
AT3G56940	77.51%	9	dicarboxylate diiron protein, putative (Crd1)
AT3G63140	86.83%	7	chloroplast stem-loop binding protein of 41 kDa
AT3G63490	86.44%	13	Ribosomal protein L1p/L10e family
AT4G01150	81.96%	13	unknown protein
AT4G02770	92.18%	12	photosystem I subunit D-1
AT4G03280	90.26%	14	photosynthetic electron transfer C
AT4G04640	92.57%	20	ATPase, F1 complex, gamma subunit protein
AT4G05180	73.15%	6	photosystem II subunit Q-2
AT4G09010	86.69%	9	ascorbate peroxidase 4
AT4G09650	88.49%	20	ATP synthase delta-subunit gene
AT4G10340	91.78%	30	light harvesting complex of photosystem II 5

AT4G21280	93.73%	18	photosystem II subunit QA
AT4G21650	87.18%	7	Subtilase family protein
AT4G23600	89.72%	8	Tyrosine transaminase family protein
AT4G23670	64.03%	6	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT4G25080	88.06%	7	magnesium-protoporphyrin IX methyltransferase
AT4G27440	83.64%	7	protochlorophyllide oxidoreductase B
AT4G28750	93.49%	8	Photosystem I reaction centre subunit IV / PsaE protein
AT4G29060	70.92%	22	elongation factor Ts family protein
AT4G32260	89.59%	19	ATPase, F0 complex, subunit B/B', bacterial/chloroplast
AT4G36580	83.69%	7	AAA-type ATPase family protein
AT4G37930	86.96%	18	serine transhydroxymethyltransferase 1
AT4G38970	91.35%	16	fructose-bisphosphate aldolase 2
AT4G39330	89.11%	6	cinnamyl alcohol dehydrogenase 9
AT5G01530	87.24%	26	light harvesting complex photosystem II
AT5G04140	85.43%	16	glutamate synthase 1
AT5G07030	93.10%	8	Eukaryotic aspartyl protease family protein
AT5G09660	87.37%	7	peroxisomal NAD-malate dehydrogenase 2
AT5G11290	88.69%	5	Plant protein of unknown function (DUF247)
AT5G14740	90.48%	17	carbonic anhydrase 2
AT5G20630	75.18%	6	germin 3
AT5G23060	86.59%	10	calcium sensing receptor
AT5G23120	88.43%	8	photosystem II stability/assembly factor, chloroplast (HCF136)
AT5G24770	91.27%	5	vegetative storage protein 2
AT5G24780	89.58%	19	vegetative storage protein 1
AT5G25980	90.34%	18	glucoside glucohydrolase 2
AT5G26000	91.97%	25	thioglucoside glucohydrolase 1
AT5G35630	89.96%	22	glutamine synthetase 2
AT5G36790	82.69%	10	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
AT5G38420	93.04%	27	Ribulose bisphosphate carboxylase (small chain) family protein
AT5G38430	91.34%	18	Ribulose bisphosphate carboxylase (small chain) family protein
AT5G54600	84.04%	5	Translation protein SH3-like family protein
AT5G54770	83.14%	12	thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) (THI4)
AT5G55220	83.32%	7	trigger factor type chaperone family protein
AT5G58330	85.49%	8	lactate/malate dehydrogenase family protein
AT5G61410	85.01%	6	D-ribulose-5-phosphate-3-epimerase
AT5G64040	92.09%	17	photosystem I reaction center subunit PSI-N, chloroplast, putative / PSI-N, putative (PSAN)
AT5G65220	87.26%	8	Ribosomal L29 family protein
AT5G66190	89.41%	18	ferredoxin-NADP(+)-oxidoreductase 1
AT5G66570	87.83%	12	PS II oxygen-evolving complex 1

11. ● ACKNOWLEDGEMENTS

12. Curriculum vitae

Phillip Ihmor

General Information

Date and Place of birth	June 17th 1986 in Warburg, Germany
Address	Beckhammer 19, 8057 Zürich
Email address	p.ihmor@biol.ethz.ch, p.ihmor@gmail.com

Research Experience

Ph.D. Candidate Plant Systems Biology   Jan 2013 – Dec 2016

   ETH Zürich, Systems Biology Graduate School
   Plant Biotechnology Group of Prof. Wilhelm Gruissem

   Thesis: *Dynamics of Protein Synthesis and Degradation in flg22-stimulated Arabidopsis*
   Supervised by PD Katja Baerenfaller

Analytics Intern, Metabolomic Discoveries GmbH, Potsdam   Aug 2012 – Nov 2012

Education

M.Sc. Biotechnology   Nov 2009 - Aug 2012

   University of Münster

   Thesis: *Mode of action of Serratia marcescens ChitinaseB assessed by quantitative mass spectrometry and hydrolysis simulation.*
   Conducted in the group of Prof. Bruno Moerschbacher

B.Sc. Biology   Oct 2006 - Sep 2009

   University of Münster

   Thesis: *Identifying the gene of Alternaria alternata Chitinosanase1*
   Conducted in the group of Prof. Bruno Moerschbacher
   ERASMUS Semester at University of Hasselt, Belgium   Sep 2008 - Feb 2009

Abitur   Oct 2006
   Beethoven Gymanisum Bonn; Major in Mathematics and Biology

Zürich, Nov 29th 2016