Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diplom-Chemikerin Katrin Eichelbaum
born in: Berlin, Germany

Oral-examination: 08.11.2012

Quantitative proteomics of transcriptional and translational regulation

Referees: Dr. Anne-Claude Gavin
Prof. Dr. Matthias Mayer

Abstract

Transcription and translation are precisely-regulated processes that enable a cell to react to environmental stimuli by the production of appropriate sets of proteins. In order to analyze gene expression regulation, two novel methods have been developed that facilitate affinity purification approaches in combination with quantitative mass spectrometry.

First, the concepts underlying transcriptional control are investigated by identifying proteins binding to cis-regulatory modules (CRMs) in a sequence-specific manner. The method facilitated the determination of proteins specifically binding to DNA sequences with a length of 500 base pairs. Quantitative comparison of the binding pattern of three CRMs active in *Drosophila melanogaster* muscle development revealed 72 candidates potentially regulating the activity of these CRMs among thousands of proteins that interact non-specifically. *In vivo* validation of biological activity of several candidates is in progress and will probably reveal new circuits of *Drosophila melanogaster* muscle development.

Second, a novel approach to specifically enrich and quantify newly synthesized proteins by combining click-chemistry and pulsed SILAC labeling was developed. This method was introduced as a useful tool to study protein synthesis and the sensitive detection of rapid response to cellular stimulation. In addition, the method was adapted to the selective and precise quantification of secreted proteins. This important subset of mammalian proteins is currently understudied because of technical limitations in the detection of low-abundant proteins against a background of serum. In-depth and differential secretome analysis of various cell lines and primary cells revealed, *e.g.*, profound effects of serum starvation on secretome composition. Moreover, a unique application studying the kinetics of protein secretion was introduced. The approach will have broad implications in studying the responsiveness of cells grown under optimal conditions.

Finally, in combination with RNA and protein abundance measurements, the developed approaches were used to investigate regulatory mechanism establishing response programs in lipopolysaccharides stimulated mouse macrophages with temporal resolution. These data for the first time provide a comprehensive view on the kinetics of macrophage activation. Transcriptional, tanslational and localization regulation was distinguished and starting points for further investigation of these mechanisms were proposed.

Zusammenfassung

Die durch Stimulierung aktivierte Proteinsynthese ermöglicht es Zellen flexibel auf Umweltfaktoren zu reagieren und wird von präzise kontrollierten Mechanismen, der Transkription und der Translation, bestimmt. Im Rahmen dieser Arbeit wurden Methoden der Proteinanreicherung in Kombination mit quantitativer Massenspektrometrie zur Analyse der Genexpression entwickelt.

Transkriptionale Kontrollmechanismen wurden durch die Analyse von sequenz-spezifisch DNA-bindenden Proteinen studiert. Die vorgestellte Methode erlaubt die Verwendung von langen DNA-Sequenzen (500 Basenpaare). Der Vergleich der Interaktionsprofile von cis-regulatorischen Modulen (CRMs), welche die embryonale Entwicklung von Fruchtfliegen (*Drosophila melanogaster*) steuern, ergab - neben tausenden unspezifisch bindenden Proteinen - 72 potentielle Regulatoren dieses Prozesses. Die *in vivo* Validierung dieser Kanditaten ist in Arbeit und wird neue Kontrollmechanismen der embryonalen Entwicklung in diesem Modellsystem aufdecken.

Die zweite, neu entwickelte Methode kombiniert Klick-Chemie und metabolische Proteinmarkierung zur selektiven Anreicherung und Quantifizierung von neu synthetisierten Proteinen. Die Analyse von aktivierten Mausmakrophagen ergab eine hohe Empfindlichkeit auf schnelle Änderungen des Proteoms. Zusätzlich wurde die Methode an die selektive und exakte Quantifizierung von sezernierten Proteinen angepasst und ermöglicht nun den Nachweis einer Proteinklasse, deren Analyse aufgrund der hohen Hintergrundkonzentration an Serumproteinen sehr anspruchsvoll ist. Die vergleichende Sekretomanalyse von verschiedenen Zelllinien und primären Zellen ergab unter anderem einen hohen Einfluss des Serumentzugs auf die Proteinsekretion. Zusätzlich wurde eine neue Strategie zur Untersuchung der Sekretionskinetik eingeführt. Mit Hilfe der vorgestellten Methode können zelluläre Reaktionen unter optimalen Wachstumsbedingungen studiert werden.

Desweiteren wurden RNA- und Proteinabundanzmessungen mit den beschriebenen Methoden kombiniert, um die Aktivierungsmechanismen in Lipopolysaccharid-stimulierten Mausmakrophagen zeitlich aufgelöst zu untersuchen. Die Ergebnisse ermöglichen die Unterscheidung von transkriptionaler, translationaler und räumlicher Kontrolle und ergeben erstmalig ein umfassendes Bild der Zusammenhänge von RNA- und Proteinkinetik in diesem Modellsystem. Außerdem wurden Ansatzpunkte zur Erforschung spezieller Kontrollmechanismen vorgeschlagen.

Widmung

Ich widme diese Arbeit meiner Familie, speziell meinen Eltern, meiner Schwester und Maik, die mich immer in allem uneingeschränkt unterstützen und mir Kraft und noch wichtiger Lebensfreude geben.

Der unermeßlich reichen, stets sich erneuernden Natur gegenüber wird der Mensch, soweit er auch in der wissenschaftlichen Erkenntnis fortgeschritten sein mag, immer das sich wundernde Kind bleiben und muß sich stets auf neue Überraschungen gefaßt machen.

Max Planck (1858-1947)

Contents

Chapter 1

Introduction

How can cells change their phenotype as response to an environmental stimulus? What makes it possible for a few cells to develop to a whole organism with specialized tissues and organelles? By asking these and similar questions, it was discovered that the functionality of all organisms is based on macromolecules like DNA, RNA and proteins.

The phenotype of a cell is established and adapted to environmental effects through interconnected networks of these macromolecular components. Protein-protein and protein-DNA/RNA interactions regulate and enhance specific gene expression to produce transcriptional and translational processes to generate the correct response to cellular changes. To understand the driving forces behind these changes, comprehensive examination of the mechanisms establishing the composition and localization patterns of the cellular pool of proteins (proteome) is necessary.

Investigating the composition of the proteome is challenging due to the high complexity and dynamic range of protein abundance within the cellular pool of protein. Hence, the ability to detect low abundant proteins with potentially interesting biological functions, such as transcription factors, is limited. Furthermore, in contrast to other macromolecules such as DNA, proteins have a much higher diversity in their chemical properties (*e.g.* solubility) and cannot be amplified.

In light of these challenges, mass spectrometry-based methods have been established as extremely powerful tools for protein identification and quantification in complex mixtures. In combination with separation techniques such as gel electrophoresis, isoelectric focusing, or multidimensional chromatography it is possible to identify and quantitatively compare thousands of proteins from whole cell or tissue lysates[1;2]. Current instruments are capable of detecting more than 10,000 proteins from a complex lysate, representing a significant depth of coverage of the human cellular proteome[3].

However, identification and quantification of all proteins within a cell under all conceivable conditions cannot provide complete mechanistic insight into cellular regulation.

The establishment of protein abundance is regulated at several steps during gene expression, leading from DNA to mRNA and finally to the functional protein. Transcription factors can regulate the temporal and spatial activity of gene expression by selective binding to specific DNA sequences. This initial step towards the synthesis of a protein is orchestrated by global cis-regulatory networks of interacting cis-regulatory modules (CRMs). Cis-regulatory modules are regions in the genome that contain clusters of binding sites that regulate the transcription of certain genes. The activity of a CRM is defined by the set of bound transcription factors. Therefore, understanding how combinatory binding of transcription factors activates CRMs would help to predict spatio-temporal gene expression.

Global detection of DNA sequences that bind to a single transcription factor can be obtained based on recent developments in chromatin immunoprecipitation methods in combination with microarrays and DNA sequencing[4–7]. However, the detection of all proteins binding to a single genomic region is challenging due to the low abundance of transcription factors and the wide variety of proteins that bind DNA non-specifically. *In vitro* DNA affinity purification from nuclear extracts has been applied to short DNA sequences using mutated sequences in order to detect non-specifically binding proteins[8]. Since CRMs are on average 270 base pairs long the mutation of one or several DNA sites would not prevent binding of all specific proteins. Therefore, a robust method for the investigation of proteins binding to longer DNA sequences is necessary to gain insights into transcriptional regulation.

Following transcription a multitude of interconnected processes define the final protein abundance, such as mRNA processing and location, the balance between protein synthesis and degradation and post-translational modifications. Within the last century, immense progress in the field of genomics and transcriptomics yielded in the production of high-content data sets gathering insights into cellular responses towards perturbation. However, these approaches only shed light on a subset of the information required to obtain a comprehensive picture of cellular regulation, as they omit contributions from protein translation. Hence, combinatory examination of protein synthesis, degradation and localization in tandem with genomic and transcriptomic data are essential to distinguish between transcriptional and translational regulation.

The potential for unbiased, high-throughput data acquisition positions mass spectrometry as a powerful tool to address these questions on protein level. However, for comparison of multiple cellular conditions more sophis-

ticated experimental designs are required. Thus, quantitative labeling protocols based on stable isotope labeling facilitating examination of protein turnover and synthesis have been established. These approaches overcome limitations of established low-throughput pulse labeling methods that require radioactive isotopes, by measuring protein synthesis rates of thousands of proteins in parallel. Contrarily, stable isotope labeling approaches are still hampered by a low temporal resolution. Due to the restricted dynamic range of the mass spectrometer, a minimal incorporation efficiency of 5-10% for the stable isotope is neccessary for proper quantification of protein synthesis[9].

This can be eluded by affinity purification of the proteins of interest, which is a routine method for the detection of other low-abundant protein classes, such as proteins with post-translational modifications[10]. Recently, the enrichment and selective detection of newly synthesized proteins was facilitated by pulse labeling with non-canonical amino acids bearing functional groups[11;12]. Combination of this enrichment approach with stable isotope labeling methods will allow the quantitative analysis of proteome dynamics, providing insight into the regulation of protein synthesis after stimulation.

Another currently understudied field in proteomics that addresses protein localization regulatory circuits is the cellular secretome. Fundamental processes that overarch multiple cell types, or even organs, such as metabolism and immunity, rely on effective transmission of signaling proteins that are secreted to elicit a response in a paracrine or autocrine fashion. In addition, extracellular proteins can regulate adhesive and migratory properties of cells, that are themselves modulated by a range of secreted proteases and protease inhibitors. Indeed, in many diseases secreted proteins can create the conditions that are favorable for the disorder, *e.g.*, promotion of cancer metastasis. This was highlighted by analyses demonstrating the secretome to be a proven source of disease biomarkers[13;14]. Therefore, knowledge of the qualitative and quantitative composition of this important set of proteins is crucial to understand the biology of cellular interaction as well as stimulation response.

In spite of its potential to obtain biologically valuable information, secretome analysis is not routinely applied in proteomics studies. This is primarily due to the high-dynamic range of protein abundance present in complex extracellular mixtures, such as human plasma, or cell cultures supplemented with 10% of bovine serum, that make mass spectrometric analysis challenging. This problem can be partially addressed through extensive protein and/or peptide fractionation strategies that enable the mass spectrometer to detect low-abundant factors. In spite of immense effords, often only few secreted proteins are detected.

Alternatively, cells can be proliferated under serum-free conditions to omit the high background of serum proteins[15]. However, serum deprivation

affects expression and phosphorylation levels of multiple proteins[16–19] and could therefore have unpredictable effects on protein secretion. As such, a method for the effective enrichment of secreted proteins from the high background of serum could resolve the current limitations in secretome analysis.

1.1 Objective of the thesis

To overcome the described challenges in these fields of proteomics, the aim of this thesis was to develop robust methods for the detection and quantification of newly synthesized and secreted proteins as well as DNA-binding proteins.

Initially, an approach for the detection of proteins binding to long DNA sequences is demonstrated (chapter 3). Moreover, the newly developed technique was applied to the detection of proteins specifically binding to CRMs active during *drosophila melanogaster* muscle development.

The subsequent chapters describe the development and evaluation of a quantitative method to compare levels of newly synthesized proteins under different conditions and its application to intracellular and extracellular measurement of protein synthesis.

The reproducible enrichment and quantification of secreted proteins tackles the described challenges in this field of proteome research (chapter 5). The broad applicability to different biological questions and systems is demonstrated and a novel approach in secretome analysis is introduced.

In chapter 6 these methods were applied to a time-course analysis of macrophage activation. This model system was used to demonstrate how the integration with mRNA and protein abundance data can shed light on the regulatory mechanism controlling gene expression and protein localization during innate immune response.

Advantages and limitations as well as possible applications of the presented approaches are discussed in chapter 7.

Chapter 2

Regulatory Mechanism Establishing Cellular Protein Homeostasis

The phenotype of a cell is to a great extent defined by its protein composition. Here not only the identity of proteins is important but also their amount, localization and activity. Most importantly, the protein pool is not static. Proteins are constantly synthesized and degraded, allowing the cell to quickly respond to environmental variations by adapting its protein composition and therefore its phenotype. To perpetuate cellular functionality the protein amount, defined by the interplay of protein synthesis and degradation (protein turnover), as well as protein localization and activity need to be tightly controlled and adjusted to environmental variations. The following chapter will give an overview on the current knowledge about regulatory mechanisms of gene expression, protein turnover and secretion and how these processes can be investigated. Mechanisms influencing protein activity will not be covered, since - although equally important - they are less relevant for the work presented in this thesis.

In addition, this chapter covers the regulation of innate immunity with the focus on macrophages and Toll-like receptor 4 signaling as a model system to study proteome dynamics. The activation of the immune system after challenging an organism with pathogens is driven by complex tightly connected response programs. Due to the extreme changes in protein expression during this process, it is well suited for the investigation of protein synthesis, degradation and secretion.

2.1 Regulation of gene expression

Protein synthesis is regulated at various steps during gene expression (fig. 2.1). The genetic information is stored in the DNA, being kept in the nucleus of a cell. DNA is transcribed to mRNA, which is transported from the nucleus to the cytoplasm, where translation into proteins takes place. Each of these steps is tightly controlled to ensure the production of the proteins needed at a certain time.

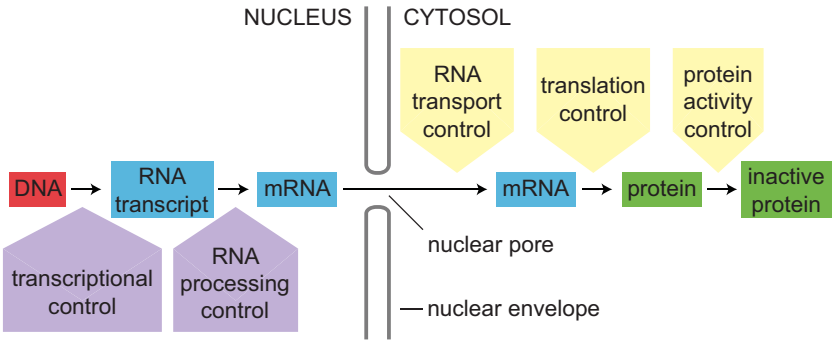

Figure 2.1: Schematic representation of gene expression and its regulation. Transcriptional control mechanisms are highlighted in purple, while post-transcriptional control points are shown in yellow. (Adapted from "Essential cell biology"[20].)

The mechanisms regulating gene expression can be roughly distinguished into transcriptional and post-transcriptional control (fig. 2.1). Transcriptional control is achieved by concerted transcription factor binding and mechanisms regulating DNA accessibility as well as mRNA processing, while translation and establishment of protein localization and activity are post-transcriptional processes. The correlation between RNA and protein abundance can help to estimate the contribution of these two levels regulating the final protein amount. The correlations reported in the literature vary depending on the experimental setup, the investigated organism and the applied statistical approach between 0.3 and 0.7[21–32]. These numbers demonstrate the high contribution of post-transcriptional control to gene expression. In the next sections transcriptional and post-transcriptional regulation mechanism relevant for this thesis, will be discussed.

2.1.1 Transcriptional regulation

During transcription the genetic code stored in the DNA is transfered to mRNA based on complementary base pairing supported by RNA polymerase II. This process is divided into four steps: recruitment of RNA polymerase II

to promoters, transcription initiation, elongation and termination[33]. Promoters are DNA sequences at the beginning of a gene bearing common sequence elements such as TATA boxes and the initiator sequence as well as specific transcription factor binding sites. In concert, these binding sites recruit the so called general transcription machinery to the transcriptional start site (TSS). The general transcription factor machinery is composed of the mediator complex and the general transcription factors which assemble into a pre-initiation complex. General transcription factors together recognize core promoters, assemble the transcription start site by *e.g.* unwinding of the DNA double strand, recruit RNA polymerase II and support the transition from transcription initiation to elongation. In addition, for certain genes the binding of transcription factors to a distant site in the DNA strand called cis-regulatory module (CRM) may be required. These CRMs are on average 270 base pair long DNA stretches composed of multiple transcription factor binding sites[34]. Here, signals provided by multiple transcription factors are integrated, leading to inhibition or activation of transcription resulting in specific spatio-temporal gene expression. The recent discovery of nearly 300,000 mouse cis-regulatory sequences illustrates their importance in gene regulation[35].

The establishment and regulation of the mentioned processes is far more complex than expected thirty years ago. Novel methodologies have revealed detailed insights into gene regulation, leading to four concepts that in concert govern gene expression[36].

First, the organization of nucleosomes and their modification states are well defined. Genes are stored in the nucleus by wrapping the chromosomal DNA around histone octamers containing two copies of histones H2A, H2B, H3 and H4 forming the nucleosomes[37]. This compact structure inhibits transcription initiation[38]. The closed chromatin formation is often marked by specific histone modifications, *e.g.*, methylation of histone H3 at lysine 9 (H3K9), methylation at H3K27 and dimethylation of H3R2[39;40]. Genome-wide analysis of nucleosome positions revealed that promoters of most genes reside in nucleosome-free regions, with equal distance to the transcription start site[41].

Second, chromatin remodeling is coupled to promoter assembly of the general transcription factors and RNA polymerase II. Transcription initiation involves the rearrangement of structure, position, or composition of nucleosomes facilitated by four families of chromatin remodeling complexes[42]. Conversely, nucleosome reorganization by the Imitation SWItch (ISWI) family of remodelers can repress transcription[43].

Third, the phosphorylation states of RNA polymerase II are connected to chromatin modification states during transcription[36]. Changing phospho-

rylation pattern at the RNA polymerase II CTD repeat selectively govern the recruitment of proteins involved in histone modifications, elongation, termination and mRNA processing at different stages during the transcription cycle[44]. In addition, the 20-40 amino acids long histone tails, which reach from the core histones to the surrounding solvent, bear highly dynamic post-translational modifications including ubiquitinylations, methylations and acetylations, also called "histone marks"[45]. They change the structure of chromatin due to alterations in electrostatic properties or internucleosomal contacts and/or serve as binding sites for regulatory proteins. RNA polymerase II phosphorylations and histone marks constitute a highly connected network together regulating the process during transcription. *E.g.* trimethylation of H3K4, which is associated with active genes[46], requires H3K14 acetylation[47] which is established by acetyltransferases that colocalizes with RNA polymerase II.

Finally, regulatory networks of sequence-specific transcription factors govern the transcription of genes. Cis-regulatory elements bound by specific transcription factors or transcription modulators regulate transcriptional programs. They can induce or repress transcription or assist binding of the basal transcription machinery[36;48]. The direct interaction of these proteins binding to a distant site in the genome with the general transcription machinery is established by a loop formation of the DNA. Thereby, sequence-specific regulators orchestrate multiple processes during transcription by the recruitment of chromatin remodeling complexes, general transcription factors, chromatin modifying complexes and RNA polymerase II via the Mediator complex[36]. Possible transcription induction mechanisms are, *e.g.*, the recruitment of chromatin remodeling proteins that increase the accessibility of TSS or the recruitment of specific kinases phosphorylating RNA polymerase II and thereby stimulate elongation. CRM regulation is characterized by the dependence of CRM activation on the concentration of transcription factors, by cooperative transcription factor binding and the dominance of repressive inputs over activating ones[49]. Concerted regulation of gene expression is established by the integration of different network motifs, like the parallel induction of several genes by the same transcription factor ("single-input" motif), the integration of multiple signaling pathways to drive transcription of one gene ("multiple-input" motif) or autoregulation of transcription factors[50].

Additional mechanisms regulating gene expression that have emerged very recently are pausing of RNA polymerase II after initiation[51] and regulation by antisense transcripts and cryptic unstable transcripts (CUTs) from intergenic regions[52;53].

The further investigation of the mechanisms governing the described con-

cepts as well as their interactions will lead to the elucidation of regulatory networks controlling gene expression, the first step towards the synthesis of proteins.

2.1.2 Analysis of transcription factor binding sites

As described, the regulation of gene expression is partly defined by transcription factors binding in concert to specific DNA elements called cis-regulatory modules thereby initiating or inhibiting gene transcription. The detection of all proteins binding to a CRM should help to understand how transcription of a gene is controlled.

Transcription factors have domains specifically binding to certain DNA motifs. Although in the era of whole genome sequencing, it should be possible to develop computational algorithms that scan this data for transcription factor binding motifs, the detection of transcription factor targets remains challenging[54]. Firstly, the transcription factor binding motif needs to be well defined. Since these motifs often have a "loose" consensus, large collections of binding sites are needed to define a motif. Secondly, due to the enormous search space, defined by the genome size, in comparison to the rather short motifs and their "loose" consensus these approaches often detect a high percentage of biologically not functional binding sites. Improvements have been achieved by accounting for the fact that transcription factor binding sites often cluster together and by the experimental validation of CRM activity data[55].

A more direct method is chromatin immunoprecipitation in combination with DNA arrays (ChIP-Chip)[4] or DNA sequencing (ChIP-Seq)[5–7]. Here, *in vivo* DNA-protein complexes are preserved by crosslinking using formaldehyde, followed by DNA fractionation and the enrichment of a selected transcription factor using an antibody against the transcription factor of interest. After crosslink reversion the bound DNA is subjected to microarray or DNA sequencing analysis to detect transcription factor binding sites across the complete genome. The need for a specific antibody as well as the fact that only binding sites for one transcription factor are detected in one experiment represent the main disadvantages of this approach.

An alternative to chromatin immunoprecipitation are protein binding microarrays[56;57]. Here, microarrays are spotted with a large number of potential DNA-binding sites and incubated with the purified DNA-binding protein of interest, bearing an epitope tag, which can be detected with a fluorescence-labeled antibody specific for the epitope tag.

All previously mentioned methods require the knowledge of the identity of the transcription factor to be used as a "bait". To detect transcriptional

active sites for unknown transcription factors either methods altering the chromatin structure, like DNaseI hypersensitivity measurement, or methods manipulating defined DNA segments are used. Extreme sensitivity to enzymatic digestion marks functional regions in non-coding sequences, including transcription factor binding sites[58]. Hypersensitivity has been associated with acetylated histones H3 and H4 and methylated H3 at lysine 4[59], which reduce the affinity of DNA for the nucleosome[60]. Still, the transcription factor binding to the detected sites is not identified.

For the identification of unknown transcription factors binding to a DNA sequence of interest mass spectrometry in combination with nucleic acid affinity capture[61;62] should be a powerful tool. Here, a "bait" DNA sequence is used for *in vitro* enrichment of DNA-binding proteins from nuclear extracts.

The identification of specific DNA-binding proteins is challenging in a single-step DNA-affinity isolation[63] caused by the low abundance of transcription factors that bind to specific promoters, accounting only for <0.01% of the total cellular proteins[64]. Therefore, large amounts of cultured cells are necessary to achieve sufficient amounts for mass spectrometric detection (10,000- to 100,000-fold enrichment). Moreover, non-specific binding of positively charged proteins to the DNA backbone represents an abundant contamination in DNA affinity isolates masking low abundant specific DNA-binders. Non-specifically binding proteins can be discriminated from specific binders by quantitative mass spectrometry using mutated control DNA sequences in parallel for enrichment.

This method has been applied to the detection of DNA- and RNA-binding proteins by different groups addressing various biological questions[65–69], such as the identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer[70] and the detection of RBP-J as a Methyl-CpG binding protein[71]. Due to the need for a mutated control DNA all this approaches use short DNA fragments (∼40 base pairs), which very well exceed the length of a transcription factor binding site. The combinatorial binding of multiple transcription factors to longer DNA sequences (*e.g.* in the range of CRMs) can therefore not be assessed by this approach.

In an alternative approach applying mass spectrometry Dejardin *et al.* used a specific nucleic acid probe against telomeres to isolate genomic DNA with associated proteins after crosslinking of protein-DNA complexes, followed by identification using mass spectrometry[72]. Thereby they are able to investigate *in vivo* DNA-protein interactions. However, it is yet to be demonstrated for genomic loci that exist at lower copies per cells. In addition, the detection of crosslinked peptides represents one challenge of this method.

In conclusion, ChIP based approaches are extremely powerful for the

genome wide identification of DNA bound to one protein, while quantitative mass spectrometry based approaches can be used to detect previously unknown DNA-binding proteins to short DNA sequences. Due to the large number of non-specific DNA-binding proteins these approaches are restricted to short DNA sequences that have defined protein binding sites. In the course of this thesis this challenge has been addressed by the implementation of a quantitative proteomic approach comparing the DNA-protein interactomes of multiple CRMs against each other. This way, a differentiation between non-specific and sequence-specific DNA binding was possible.

2.1.3 Regulation of protein synthesis and degradation

As depicted in section 2.1 transcriptional regulation can define the final protein amount only to a certain extent, because it represents just one part of the process of gene expression. It has been estimated that approximately one third of all genes are translationally controlled[73]. Hence, the additional investigation of protein synthesis and protein degradation (protein turnover) helps to understand how global protein abundance is established and regulated. Here, protein synthesis has been much deeper investigated in the last decades than protein degradation.

Protein synthesis or translation, which is well characterized and described in general textbooks[74], comprises the ribosomal decoding of information encrypted in mRNA into the amino acid sequence of a protein in the cytoplasm of a cell. Each amino acid is encoded by nucleotide triplets being recognized by the complementary anticodon triplet of a transfer RNA (tRNA). Each tRNA carries a special amino acid, that can be added to the C-terminal end of the polypeptide chain. The translation process, carried out by large multiprotein complexes, the ribosomes, is divided into four phases: initiation, elongation, translocation and termination. During initiation the start codon (AUG) is recognized by a unique initiator tRNA molecule, resulting in the binding of the small ribosomal subunit, which is supported by initiation factors. Protein synthesis starts after binding of the large ribosomal subunit. During elongation mRNA triplets are translated in 5'-to-3' direction by binding of complementary aminoacyl-tRNA in four steps. After aminoacyl-tRNA is bound, the peptide bond can be formed. This is followed by two ribosome translocation steps. These reactions are controlled and driven by elongation factors, which use GTP hydrolysis to provide energy. Hence, the ribosome moves along the mRNA until it reaches a stop codon. Here binding of release factors results in the termination of translation and release of the polypetide chain.

Regulation of translation is achieved by several different mechanisms

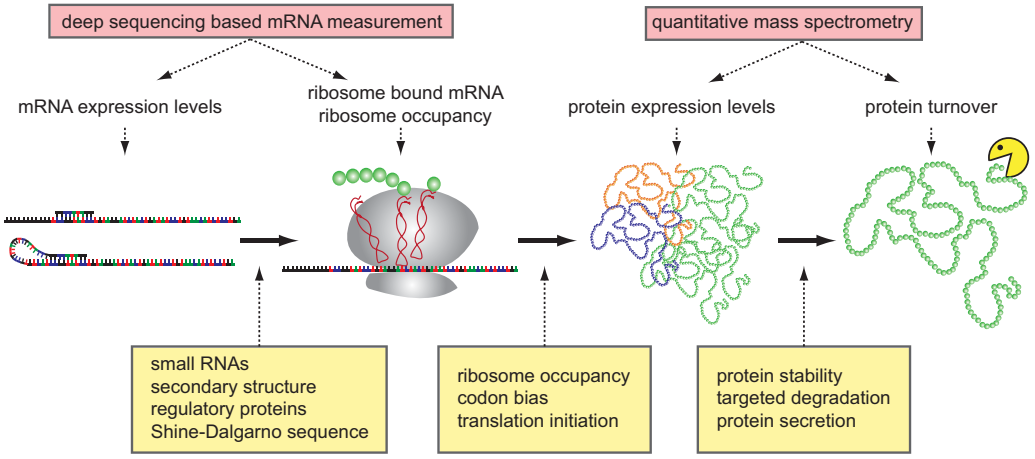

Figure 2.2: Mechanism regulating translation (yellow) and methods to address different levels of regulation (red). (Adapted from Maier et. al.[31].)

(fig. 2.2). First, the translation rate of proteins can be influenced by physical properties of the mRNA. Prokaryotic mRNA for example contains a so called Shine Dalgarno (SD) sequence upstream of the start codon[75]. This sequence is complementary to the 3' end of 16S rRNA in the ribosome, therefore representing a ribosomal binding site. The strength of ribosomal binding to this sequence influences the translation rate, being lower for weak ribosome binding. The Kozak sequence has a similar function in eukaryotes[76].

In addition, condition-dependent physical properties like secondary or tertiary mRNA structure can influence the accessibility of ribosomal binding sites. The structure of rpoH in *Escherichia coli* for example is temperature dependent resulting in different translation rates for the corresponding protein at different temperatures[77]. The mRNA translation repression of the cobalamin-transport protein (btuB) in *Escherichia coli* by high coenzyme B12 concentration is another example, this time highlighting the influence of small metabolites on protein synthesis[78].

An additional mechanism regulating transcription is the binding of repressive proteins to the translational start site (fig. 2.2). Similarly, regulatory small RNA in prokaryotes or microRNA in eukaryotes can affect evolution and stability of mRNA as well as translation efficiency[79]. The enhancement of ribosome binding is for example achieved when small RNA binding prohibits secondary mRNA structures that block ribosome binding[80]. Translation is blocked by small RNAs binding to the SD sequence of mRNA[79]. An increase of mRNA stability can be achieved by shielding RNAse cleavage sites[79;81;82].

Translation efficiency, defined as the number of molecules per mRNA and

time, is furthermore maximized by a mechanism called "codon bias" (fig. 2.2). Here, the triplet codes for one amino acid are used with different frequency. Highly expressed genes usually have a large codon bias[83;84], which can be measured using the codon adaptation index (CAI)[85]. This method can help to evaluate the abundance distribution in a data set.

Ribosome occupancy, which is the number of ribosomes bound to and translating one mRNA, has a big impact on translational efficiency (fig. 2.2). It can be measured by obtaining polysome profiles with sucrose gradient centrifugation[86], affinity tag purification followed by microarray analysis and/or northern blotting[87] or deep sequencing of mRNA fragments that are protected by ribosomes[73]. Ribosome occupancy and ribosome density have been estimated to contribute to approximately 5% to the total variation between mRNA and protein abundance[88]. Therefore, the parallel measurement of ribosome occupancy and mRNA and protein abundance improves the correlation between mRNAs and proteins[73].

An important mechanism establishing protein abundance is protein degradation, which in combination with protein synthesis determines the half-life of a protein (fig. 2.2). Protein degradation is influenced by protein stability, which can be partly estimated using the N-end rule[89]. Some N-terminal amino acids are specifically recognized by chaperones, which unfold proteins for proteasomal degradation. Therefore, the N-terminal amino acid determines the stability of proteins. Another sequence motif inducing protein instability is the PEST motif, which is rich in proline, glutamic acid, serine and threonine residues[90]. In addition, certain post-transcriptional modifications like ubiquitinylation mark proteins for degradation.

Translation efficiency can furthermore be regulated during translation initiation or elongation (fig. 2.2). One example is the phosphorylation of translation initiation[91–93] or elongation factors[94]. Translational efficiency is influenced by the sequence surrounding the stop codon[95;96]. In addition mRNA could either not be translated or the corresponding protein could escape from detection, due to the secretion.

In conclusion, protein synthesis and degradation are regulated by complex interconnected mechanisms highly influenced by internal and external conditions. Therefore, the global measurement of these processes would help to explain regulatory principles of gene expression. At the same time, it may provide a mechanistic link between specific external events and their consequences at the level of the proteome.

2.1.4 Measuring protein synthesis and degradation

Protein synthesis and degradation are often measured by pulse-labeling with radioactive or stable isotopes. Therefore, cells or even complete organisms are grown with isotope labeled nutrients, *e.g.*, amino acids like ^{35}S-methionine[97], glucose or nitrogen, resulting in the labeling of all newly synthesized proteins. Hence, the relative amounts of labeled and unlabeled proteins can be used to calculate the turnover of a protein. Conversely, labeling of cells or organisms with isotopes for a certain amount of time, followed by a medium switch to the natural isotope, allows to calculate protein degradation rates based on the decrease of isotopic label. When combined, double labeling strategies enable the parallel measurement of protein synthesis and degradation[98;99].

The detection of radioactive isotopes with audioradiography reveals the averaged protein synthesis rate of the total proteome. The combination of radioactive labeling with protein separation using 2D gel electrophoresis enables the measurement of individual protein half-lifes[97–99]. Here, the half-life for each gel-spot can be calculated separately while the protein identification is performed using mass spectrometry. An alternative option is the enrichment of proteins using antibody-based techniques, followed by audioradiography. Both methods require single protein samples after separation, meaning no overlap of gel-spots or highly specific antibodies, which is difficult to achieve considering the complexity of the proteome.

Protein degradation can be assessed by chemical inhibition of eEF2-mediated ribosome translocation with cycloheximide followed by a western blot time course analysis. The disadvantage of this method is the possible perturbation of cellular function by cycloheximide, *e.g.*, induction of cellular stress[100].

Although limited in throughput, these traditional techniques revealed basic principles of protein synthesis and degradation regulation. Modern methods aim to assess protein turnover on a global scale. They can be subdivided into approaches using genetic tags and those that don't.

Genetic tagging approaches utilize either affinity or fluorescent tags. Affinity tags can be used for affinity purification followed by western blotting as read out, as performed for 4000 tandem affinity protein (TAP) tagged yeast strains[101]. Bleaching experiments with fluorescent tags enable the extrapolation protein half lifes[102]. Khmelinskii et al. recently introduced fusions of two single-color fluorescent proteins that mature with different kinetics called "tandem fluorescent protein timers" (tFTs), which were used to analyze protein turnover and mobility in living cells[103]. A careful investigation of the results retrieved with these methods is necessary since the expression of tagged proteins could influence protein structure, activity, localization and

stability. In addition, genetic manipulation of cell lines or organisms is time and cost intensive and would be difficult to be applied for many cell lines and even more difficult for systems that cannot be genetically manipulated.

Most non-tagging approaches adapt isotopic labeling strategies by using stable instead of radioactive isotopes, in combination with high resolution mass spectrometry for protein identification and measurement of protein synthesis and degradation[104–112]. Different sources of stable isotopes can be used, including deuterated water[111] and stable isotope labeled amino acids containing ^{2}D, ^{13}C, ^{15}N or a combination thereof[106;112].

In standard stable isotope labeling methods, which aim to compare protein abundances between two samples, the cells are completely labeled with the stable isotope. After collection, mixing and lysis of the differentially labeled and treated cells, the proteins are digested into peptides, which are submitted to nanoLC-MS/MS. Each peptide is now represented by a heavy and light peak, coming form the different samples. The peak area ratios are used to calculate the relative abundance of the protein of interest. Nowadays, the stable isotope labeling with amino acids in cell culture (SILAC) method is most often used for this purpose[113] (fig. 2.3a).

In contrast to this but similar to radioactive labeling, for protein turnover studies the cells are just labeled for a certain time with the stable isotopes[106] (fig. 2.3b). When using stable isotope labeled amino acids this approach is called dynamic SILAC. An adaption of this method is pulsed SILAC[112] (fig. 2.3c). Here, in a triple labeling strategy changes in protein synthesis in response to cellular perturbation are compared. Starting with two equal cell populations the cell growth medium is exchanged to growth medium either containing medium-heavy or heavy stable isotope labeled amino acids. In parallel, a perturbation can be applied to either of the two cell populations. After the desired incubation time the cells are mixed and subjected to sample preparation for mass spectrometry. The measured mass spectra will contain three peaks. While the light peak represents a combination of peptides derived from "old" proteins of both samples, the medium-heavy and heavy peak correspond to the peptides of newly synthesized proteins in each of the two samples and can be used for relative quantification of protein synthesis.

Due to the restricted dynamic range of the mass spectrometer, a minimal incorporation efficiency of 5-10% for the stable isotope is necessary to obtain a sufficient signal-to-noise ratio for proper quantification of protein synthesis[9]. Hence, short incubation times are not applicable and transient changes in protein synthesis, *e.g.*, rapid stimulation responses, cannot be investigated.

The signal-to-noise ratio could be improved by depleting "old" proteins, since those represent the major part of the sample. A depletion of "old" proteins would allow the subjection of higher sample amounts to the mass

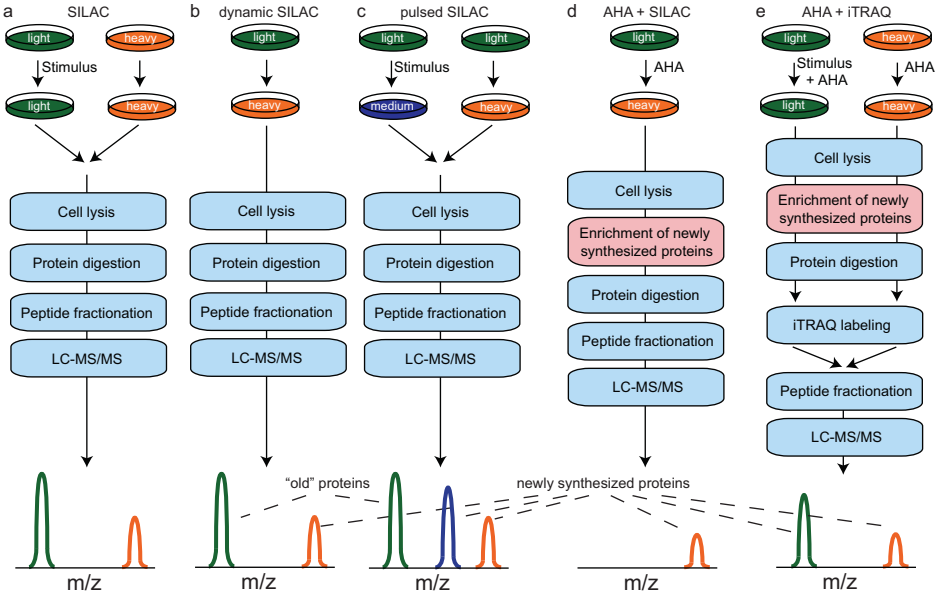

Figure 2.3: Mass spectrometry based methods applied for the quantification of newly synthesized proteins or protein turnover. a) Quantification of protein abundance changes using SILAC[113]. b) Protein turnover measurement by dynamic SILAC[106]. c) Quantification of differences in protein synthesis measured with pulsed SILAC[112]. d) Identification of newly synthesized proteins after enrichment. Stable isotope labeling enables the distinction between newly synthesized proteins and background[11]. e) Quantification of differences in protein synthesis using isobaric tags for relative and absolute quantification (iTRAQ) for peptide labeling[114;115].

spectrometric analysis, increasing the signals derived from newly synthesized proteins.

2.1.5 Labeling of newly synthesized proteins with non-natural amino acids

For the depletion of "old" proteins a method to tag newly synthesized proteins would be needed. A perfect tag would be selectively and effectively incorporated into newly synthesized proteins and should ideally have no effect on protein structure or function. Furthermore, a functional group is necessary, which is bioorthogonal and allows the enrichment of the tagged proteins. This means it should not interfere with naturally occurring functional groups and reactions. Furthermore, its incorporation should allow experimental setups similar to stable isotope labeling.

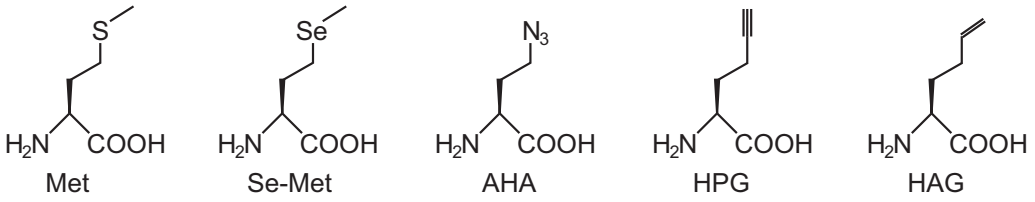

Figure 2.4: Non-canonical amino acids: Four analogues of methionine are shown. Azidohomoalanine (AHA), homopropargylglycine (HPG) and homoallylglycine (HAG) bear reactive groups that can be functionalized using bioorthogonal chemistry.

These requirements are perfectly fulfilled by analogues of naturally occurring amino acids, called non-canonical amino acids (ncAAS) (fig. 2.4). The mis-incorporation of amino acids in the polypeptide strain during protein synthesis is prevented due to the selectivity of aminoacyl-tRNA synthetases[116]. Therefore, non-canonical amino acids can be incorporated into proteins only if their structure is similar enough to the naturally occurring amino acid, to satisfy the proofreading mechanism of the according aminoacyl-tRNA synthetase. Since the incorporation of the non-canonical amino acids is often disfavored relative to the naturally occurring amino acid[117], a depletion of the natural amino acid is necessary.

Selenomethionine was the first non-canonical amino acid incorporated into newly synthesized proteins in 1956[118] (fig. 2.4) and it has been used extensively for phase determination in structural biology[119]. Today hundreds of non-canonical amino acids are known, but only a few can be directly incorporated without the need for genetical engineering approaches. Among them the methionine analogues azidohomoalanine (AHA), homopropargylglycine (HPG) and homoallylglycine (HAG) bear reactive functional groups that do not appear in the natural environment of the cell, being azide, alkyne, or alkene side chains, respectively (fig. 2.4). Experimental conditions can be adjusted to influence the percentage of natural amino acid replaced by the analogous one. The nearly quantitative replacement of methionine has been reported for AHA[120], while 85% of methionine were substitued by HPG in *Escherichia coli*[117]. Obviously, only proteins containing the replaced amino acid (*e.g.* methionine) and that are expressed can be labeled. Labeled proteins are coupled to reagents containing affinity enrichment groups or fluorecent dyes applying bioorthogonal chemical reactions, opening cues for detection and visualization of newly synthesized proteins (fig. 2.5). In this setup non-canonical amino acids are useful tools in molecular biology.

Azides and alkynes, which are absent in mammalian cells, can be coupled to each other by the very selective 1,3-cycloaddition under enhancement

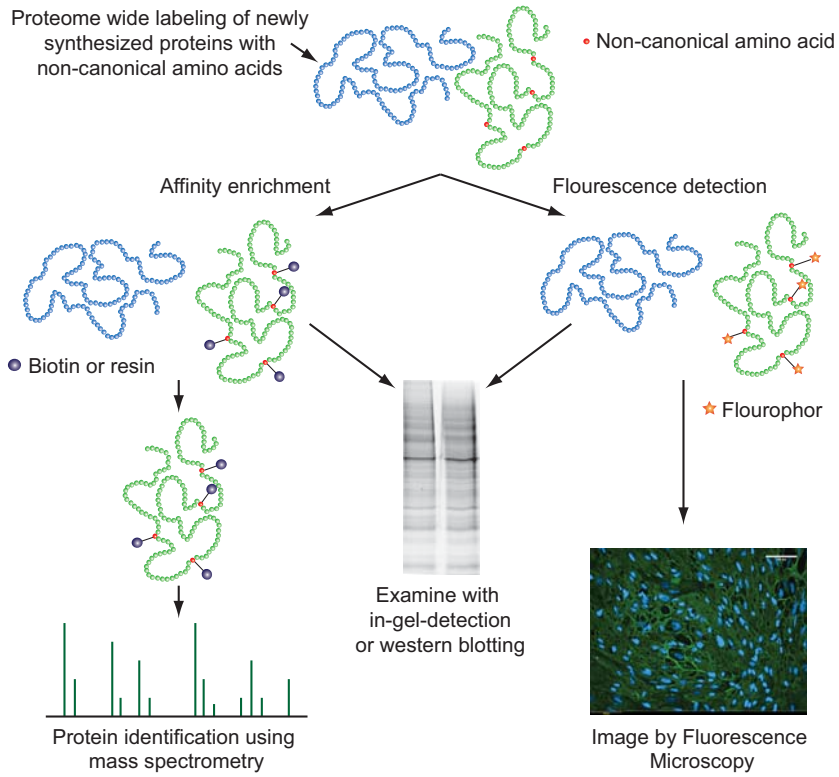

Figure 2.5: Non-canonical amino acids for the detection and visualization of newly synthesized proteins: ncAAs are incorporated into all newly synthesized proteins (green) using pulse-labeling. The thereby introduced functional groups can be modified with either affinity tags like biotin or agarose resin or with fluorescence dyes using bioorthogonal chemistry. The fluorescent label can be read out after gel electrophoretic separation or *in vivo* visualize the distribution of newly synthesized proteins in the cell. Affinity tags can be used for detection by western blotting (in the case of biotin) and affinity enrichment, to separate newly synthesized proteins from "old" proteins (blue), followed by mass spectrometric identification. (Adapted from Ngo *et al.*[120]. The fluorescence image was kindly provided by Jenny Hansson.)

by Cu(I) catalysis or by ring strain (so-called click-reaction)[121]. An alternative way to functionalize azides is the Staudinger ligation[122], attaching phosphines or phosphites. Therefore, AHA in contrast to HPG exhibits a broader bioorthogonal reactivity.

Both non-canonical amino acids have been used to functionalyse newly synthesized proteins in bacteria[12;114;123] as well as mammalian cells[11;124–126]. Dieterich *et al.* first used AHA in combination with deuterated leucin to label and enrich newly synthesized proteins from human endothelial cells (HEK)[11] (fig. 2.3d). They demonstrated that AHA is not toxic during the applied two hours of incubation and that it does not increase protein degradation. For enrichment of newly synthesized proteins prior to mass spectrometric analysis a biotin reagent, containing an alkyne group and a cleavable linker, was used in combination with streptavidin beads. To exclude the detection of background binding proteins, only peptides containing AHA or deuterated leucin were used for the identification of newly synthesized proteins. This approach is called bioorthogonal non-canonical amino acid tagging (BONCAT).

To circumvent the streptavidin enrichment step Nessen *et al.* developed an alkyne bearing resin (azide-reactive cyclooctyne (ARCO) resin) for covalent coupling of AHA containing proteins[12]. Here, the reaction can be performed without Cu(I) catalysis, since the cyclic alkyne is activated by ring strain. Again a cleavable linker allowed the release of newly synthesized proteins after stringent washing.

In an alternative approach Kramer *et al.* labeled the AHA containing proteins, derived from very short pulse labeling of *Escherichia coli*, with tris(2-carboxyethyl)phosphine using Staudinger ligation[114]. This induces a time shift during peptide chromatography, allowing the separation of AHA containing peptides from others. For the quantification of perturbations induced by changing the growth temperature they combined this approach with isobaric tags for relative and absolute quantification (iTRAQ) labeling of the peptides, a reagent used for relative quantification in mass spectrometry (fig. 2.3e). In an additional report they used this method to quantify the changes in protein synthesis after switching from aerobiosis to anaerobiosis[115]. The comparison of total protein expression changes with the detected changes in protein synthesis revealed a higher contribution of protein synthesis to the resulting protein abundance levels[115]. Protein degradation had a minor influence on protein abundance in their system.

To assess genome-wide nucleosome dynamics in *Drosophila melanogaster* cells Deal *et al.* used AHA labeling to enrich newly synthesized histones, followed by the identification of bound DNA[127]. Hereby, they found the highest nucleosome turnover over active genes, epigenetic regulatory elements

and replication origins.

Notably, Hinz *et al.* incorporated recently AHA in living zebrafish larvae, without measurable effects on zebrafish larvae behavior[128]. They used this method to visualize and measure the amount of newly synthesized proteins across the larva.

In conclusion, the incorporation of non-canonical amino acids represents a useful tool to selectively introduce tags into newly synthesized proteins. Coupling these tags to reagents using bioorthogonal chemistry for enrichment or visualization enables the investigation of protein synthesis dynamics, which will improve our understanding of gene regulation. In this technique, the selective capturing of the newly synthesized proteins from a background of "old" proteins is advantageous in comparison to other strategies for the quantification of newly synthesized proteins.

2.2 Regulation of protein secretion

Subsequent to synthesis proteins can be translocated in the cell, establishing the final protein amounts in cellular compartments. This includes protein secretion to transport proteins to the extracellular space. Here, secreted proteins, such as cytokines, chemokines and hormones, exhibit central functions, like intercellular communication, which is crucial to maintain homeostasis in every multicellular organism. The cellular secretome does not only include actively secreted proteins, but also extracellular matrix proteins and proteins shed from the cell surface[13]. This important set of proteins is encoded by approximately 10% of the human genome regulating multiple cellular processes like cell-to-cell signaling, immunity, migration and metabolism.

Secreted proteins exit the cell via different pathways. Proteins released by the classical secretion pathway are mostly regulated on the level of gene expression, while the regulation of unconventional secreted proteins is often independent of gene expression regulation[129].

2.2.1 Classical secretion

The majority of proteins is externalized via the classical secretion pathway[130]. Here, proteins containing an N-terminal hydrophobic signal sequence are secreted in a ER/Golgi-apparatus dependent manner.

Proteins bearing a signal peptide are translated into the lumen of the endoplasmatic reticulum guided by the signal peptide recognition particle[131] (fig. 2.6). Following post-transcriptional processing and quality control, they are packed into cargo vesicles coated with coat protein complex II (COPII)

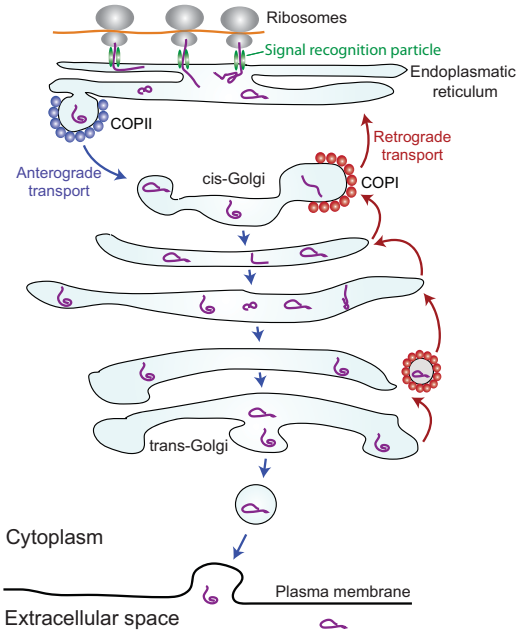

Figure 2.6: Schematic representation of the classical secretion pathway.

at specialized membrane domains, called ER exit sites (mammals) or tER sites (yeast or *drosophila melanogaster*) and directed towards the Golgi-apparatus[132]. The transport vesicles fuse to form new cis-Golgi vesicles, which then move towards the trans-face of the Golgi stack[133]. After passing the Golgi-apparatus, where they are modified, processed and sorted, cargo proteins are directed into vesicles and transported towards the cell membrane. Vesicular transport across Golgi cisterna as well as back to the ER is mediated by COPI coated vesicles[134]. Next, the carrier vesicles fuse with the membrane, releasing the proteins into the extracellular space[135]. Vesicles can be continuously transported to the plasma membrane, resulting in the constitutive secretion of the containing proteins. Alternatively, proteins can be stored in vesicles, called secretory granules, and membrane fusion and protein release to the extracellular space needs to be induced by an extracellular stimulus[136]. Fusion events during vesicular transport are mediated via SNAREs (soluble N-ethylmaleimide-sensitive fusion protein (NSF) accessory protein (SNAP) receptors). Therefore, trans-SNARE complexes are formed between vesicular SNAREs (v-SNARE) and target SNAREs (t-SNARE), *e.g.*, at the membrane when they are close to each other[137].

2.2.2 Unconventional secretion

Unconventional or non-classical secretion encompasses all pathways that do not follow the classical ER-Golgi-route of secretion. Secretion of proteins continuing along these pathways is not affected by brefeldin A, which causes rapid fragmentation of the Golgi-apparatus[138]. Although unconventional secretion has been mainly shown for proteins not bearing a signal peptide, recently Golgi-independent routes of secretion for signal peptide containing proteins have been reported.

Unconventional secretion of signal peptide containing proteins

Signal peptide containing proteins can bypass the Golgi-apparatus. This was shown for CD45[139], ovine Mx1[140] and CF transmembrane conductance regulator (CFTR)[141]. COPII-coated vesicles can either directly fuse with the plasma membrane or with lysosomes or endosomes which are directed towards the membrane. Alternatively, cargo proteins can be packed into COPII independent vesicles, following transport to the plasma membrane. Currently, the function of these export routes is not clear, but possibly processing of oligosaccharide chains or proteolytic cleavage events are prevented, resulting in a different biological activity of the secreted proteins[142].

Unconventional secretion of cytosolic or nuclear proteins

Several proteins not bearing an N-terminal signal sequence are established secretory proteins, *e.g.*, Mif, Galectins, Il-1beta, Ybx1[143–147]. The diverse routes of secretion they follow are not fully dissolved yet, but it has been shown that unconventional secretion is often triggered by cellular stress, such as inflammation or starvation[148]. At least four mechanisms, which can be classified into vesicular and non-vesicular pathways, have been proposed.

First, the translocation of proteins through the membrane can be performed without the involvement of vesicles (fig. 2.7). Examples for proteins following this pathway are FGF1[149], HIV Tat[150] and annexin A2[151]. Transporter proteins, like ATP binding cassette (ABC) transporter, can assist the externalization of the protein. One example is ABCA1, which is involved in the secretion of multiple proteins like IL-1beta[152] and HSP70[153].

In addition, phosphoinositide phosphatidylinositol-4,5-bisphosphate has been described to initiate FGF2 translocation towards the membrane, followed by heparan sulphate proteoglycans (HSPGs) mediated secretion[142]. Just recently, the oligomerisation of FGF2 resulting in the formation of lipidic membrane pores was demonstrated as the mechanism driving FGF2 secretion[154].

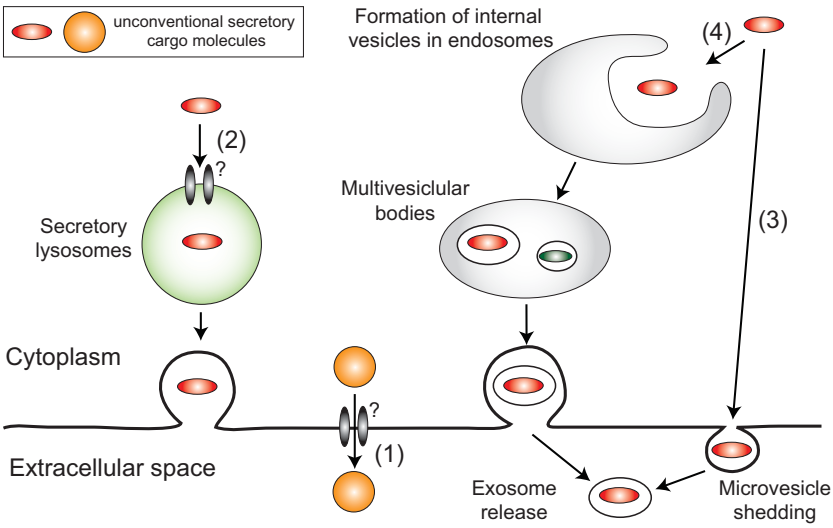

Figure 2.7: Schematic representation of possible unconventional secretion pathways: 1) Translocation of proteins across the plasma membrane. Adaptor molecules could be involved. 2) Secretion via secretory lysosomes. 3) Secretion through microvesicle shedding. 4) Secretion of exosomes. Question marks highlight that the involvement of transmembrane transporter molecules is not resolved for all unconventional secreted proteins, yet. (Adapted from Nickel *et al.*[142].)

One vesicle-dependent pathway facilitates the externalization of cytosolic and nuclear proteins by secretory lysosomes (fig. 2.7). Secretory lysosomes share functions of lysosomes and secretory granules, being acidic and containing proteases but also performing regulated secretion[155]. IL-1beta and its transforming caspase (caspase 1) have been proposed to follow this secretion route[156;157].

The shedding of microvesicles from the extracellular side of the plasma membrane is a second vesicle dependent secretion pathway[142] (fig. 2.7).

Finally, multivesicular bodys can be formed by endosomes[142] (fig. 2.7). These release the intra-endosomal vesicles to the extracellular space, which are known as exosomes. They have been reported to contain multiple cytosolic or nuclear proteins[158]. Engulfment of phagophore membrane bound Acyl-CoA binding protein by endosomes has been demonstrated as a selective autophagy dependent mechanism for multivesicular body formation, followed by secretion of this protein[159;160].

Recently, recycling endosome mediated secretion of tissue transglutaminase has been demonstrated, possibly representing an alternative unconventional secretion pathway[161].

2.2.3 Secretome analysis

Although the secretome contains compared to the total proteome a rather small fraction of proteins, its analysis is challenging. To generate real *in vivo* data the analysis of biofluids would be the direct approach. The difficulties here are the highly complex background, (*e.g.*, 12 orders of magnitude in protein concentration in human plasma[162]) as well as possible rapid degradation of secreted proteins by proteinases upon release or during sample processing[163]. Therefore, a common approach is to use culture media after conditioning with a cell type of interest for secretome analysis. This method allows non-invasive sample collection. Moreover, many cell lines representing several cellular phenotypes are available, *e.g.*, different stages of cancer progression[15]. On the other hand, the retrieved results need to be evaluated *in vivo* afterwards, since in the real organism several different cell types are localized together, possibly influencing each other.

Conditioned media can be analyzed using either antibody-based approaches, like ELISA assays and microarrays, or using mass spectrometry, both representing complementary approaches[164]. While ELISA needs to be performed separately for each protein, microarrays and mass spectrometric detection methods are large-scale approaches. For microarrays selective antibodies against all proteins of interest, bound to a solid support, are incubated with one or two fluorescent labeled samples. Fluorescence signal intensities are used as read out. The big advantage of microarrays is the possibility to measure different kinds of data in parallel. These could be protein abundance[165], the existence of different isoforms[166] or protein modifications[167] as well as biological activity like protein-protein interactions. Unfortunately, these data can only be acquired using microarrays if a binder with appropriate specificity and affinity for the protein of interest is available, representing the biggest limitation of this approach.

In contrast, mass spectrometry is an unbiased method, allowing in theory the detection of all proteins present in a sample. In the real experiment this is limited by the measurable dynamic abundance range. Since most cells are grown using media supplemented with bovine serum (10%), the serious challenge of secretome analysis using mass spectrometry lies in the detection of low-abundant secreted proteins (ng/ml range) against a background of 1000s of high-abundant serum proteins (mg/ml). It has been shown that, even when reducing the concentration of serum to 0.5%, the detection of secreted proteins is complicated[15]. An additional contamination to be dealt with are cells that die during collection, resulting in the detection of many intracellular proteins[168]. Hence, mass spectrometric analysis of conditioned media is done in either of two ways. First, cells can be conditioned under

optimal conditions in the presence of serum in the growth media, necessitating extensive protein and/or peptide fractionation strategies to enable the mass spectrometer to detect low-abundant factors. Isotope labeling strategies have been used to distinguish cellular from serum proteins. Examples are the labeling of newly synthesized proteins with radioactive isotopes, followed by gel electrophoresis and autoradiographic visualization of secreted proteins[169], as well as approaches using stable isotopes, like SILAC[170], but this does not increase the sensitivity. An increase in sensitivity has been achieved by protein equalization using Proteominer beads[171].

Alternatively, a widely used approach is to keep cells under serum-free conditions, thereby reducing analytical interference and facilitating the detection of secretory proteins[15]. Although serum starvation as an experimental condition has been used to synchronize cells and to study cellular stress and apoptosis[16], it has several consequences that are unintended and essentially disregarded in secretome analysis. For instance, even a short period (hours) of serum deprivation affects expression and phosphorylation levels of multiple proteins[16–19]. Hence, huge efforts are necessary to optimize washing procedures to efficiently reduce serum protein contamination in conditioned medium, but at the same time circumventing cell damage or perturbation of the physiological properties of the cells of interest[172–174].

In conclusion, the current methods for secretome analysis suffer from the high background of serum proteins contained in the cell growth medium. Therefore, a method would be needed which allows the selective enrichment of proteins from full cell growth medium. Very recently, the secretome protein enrichment with click sugars (SPECS) method has been reported. All newly synthesized and glycosylated proteins are pulse-labeled with non-canonical sugars bearing an azide functionality, which was used for the enrichment of glycosylated proteins from cell growth medium[175]. Since 66% of all secreted proteins are glycosylated, this method adresses a high proportion of the secretome but not all secreted proteins.

Prediction of protein secretion

As described, secretome analysis is complicated by multiple sources of contaminations, like serum proteins or cells dying during collection[168]. While serum proteins can be distinguished from cellular proteins using stable isotope labeling approaches, the other source of contaminations cannot fully be excluded. Hence, a quality assessment for the acquired data is necessary. The special characteristics of secreted proteins allow to a certain extend the computational prediction of secreted proteins.

Classically secreted proteins bear an N-terminal signal peptide, which

targets the transcription of secreted proteins into the endoplasmatic reticulum[176]. The approximately 15-30 residues long signal peptide is cleaved off during membrane translocation. Although there is no common sequence, all signal peptides contain an N-terminal region (n-region), often bearing positively charged residues, an at least six residue long hydrophobic region (h-region) and a C-terminal region (c-region). The c-region is made up by polar uncharged residues showing some conservation at the -3 and -1 positions relative to the cleavage site[177]. These special features are used to predict the secretion probability for a protein applying machine learning methods and modeling approaches, such as in the popular and powerful tool SignalP[178].

Since unconventional secreted proteins do not bear a signal peptide their prediction is more complicated. Programs like SecretomeP[179] assume that secreted proteins share certain characteristics, like amino acid composition, frequency of amino acid pairs or protein disorder[177]. Since rather low numbers of unconventional secreted proteins have been detected so far, these approaches suffer from the need of reliable training sets for machine learning algorithms to predict protein secretion and are therefore not as reliable as approaches predicting conventionally secreted proteins.

In summary, the prediction algorithms for conventional secreted proteins are reliable due to the very specific properties of the signal peptide. Contrarily, the small number of proven unconventional secreted proteins and the multitude of possible unconventional secretion pathways, which are at the time not fully uncovered, complicates the development of prediction methods for these group of proteins. Hence, the existing approaches need to be applied with care.

2.3 Activation of macrophages

The process of protein secretion as described in the previous section is highly important for intercellular signaling of the immune system. Here, secretion of signaling proteins ensures the activation of immune cells after exposure with infectious agents. Among the first actors after infection are macrophages, which were used as model system in this thesis. Therefore, the mechanism activating macrophages will be discussed here.

Inflammation, the self-regulating defense mechanism protecting the host from invading pathogens, is divided into innate immunity and adaptive immunity. While innate immunity, which is mainly executed by myeloid cells, is initiated immediately after pathogen recognition, adaptive immunity represents the second line of defense. Here, lymphocytes establish an antigen-specific recognition system enabling the organism to remember pathogens

after initial contact and to respond stronger at secondary infection with the same pathogen.

While innate immunity was originally thought to be composed of non-specific responses and phagocytosis, it is now well accepted that it orchestrates adaptive immunity and recognizes pathogen associated molecular patterns (PAMPs) with unexpected specificity[180]. The innate immune system is composed of physical and chemical barriers, like the epidermis and mucosal surfaces with antimicrobial secretions as well as specialized cells recognizing pathogens and activating defense responses[181].

In cooperation with the other cells of the immune system, macrophages, which are derived from monocytes entering damaged tissue[182], establish innate immune responses towards invading pathogens. Their spectrum of functionalities comprises the phagocytosis of cellular debris and patogens, the recognition of PAMPs by specialized receptors, the secretion of molecules and proteins activating other immune cells and the presentation of antigens to lymphocytes. To execute several of these functions macrophages need to be activated by PAMPs or substances secreted by other immune cells.

Activated macrophages similarly to T-cells develop into distinct phenotypes. Here, classically activated macrophages (M1 macrophages) and alternatively activated macrophages (M2 macrophages) are well established (fig. 2.8)[183]. While initially M1 cells were defined as pro-inflammatory cells and M2 macrophages were thought to exhibit functions in humoral immunity and repair[184], it has been shown that subsets of M2 macrophages are involved in wound healing and the regulation of inflammation[185]. The development of the different macrophage types is defined by their stimulation with different substances and proteins employing different types of pattern recognition receptors. The thereby activated signaling cascades that tightly regulate the induction of specific transcription factors, result in the differential expression of many cytokines and receptors by these cells (fig. 2.8). The differentiation between macrophage phenotypes is maintained by feedback loops, selectively repressing transcription of genes specific for the alternative macrophage type[186]. Transcription factors inducing genes responsible for M1 phenotype are AP1, NF-κB, STAT1 and IRF5, while STAT6, IRF4, CREB, C/EBPβ and PPARγ regulate M2 specific genes[186]. The ongoing debate about the definition of macrophage types is represented by the recent discovery of a new type of activated macrophages called Mox macrophages, exhibiting a different phenotype with decreased phagocytotic and chemotactic capacity and an overrepresentation of NRF2-mediated expression of redox-regulatory genes (fig. 2.8)[187]. Tumor-associated macrophages (TAMs)[188], an additional macrophage phenotype, are often considered as M2 macrophages, but similarly to Mox exhibit gene expression pattern distinct from M1

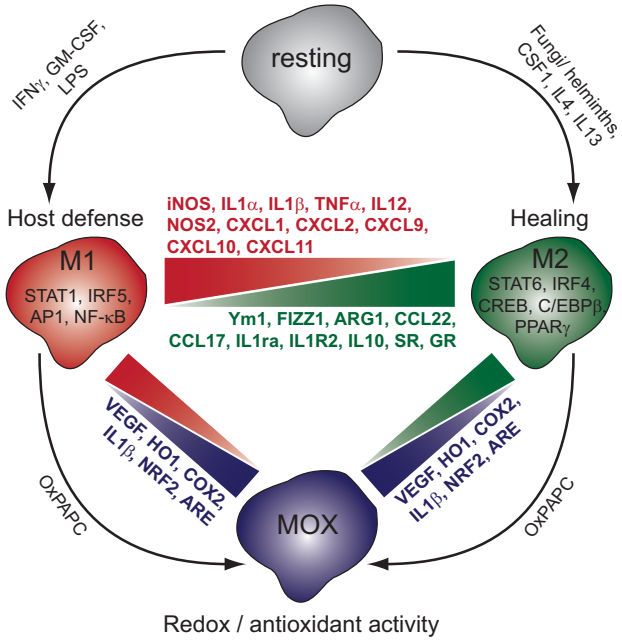

Figure 2.8: Types of activated macrophages: Resting macrophages can be activated into M1 or M2 macrophages by various stimuli (shown on arrows). Transcription factors regulating gene expression in response to these stimulations are depicted inside the macrophage types. M1 and M2 macrophages can both be transformed to Mox macrophages by oxidized 1-palmitoyl-2-arachidonoyl-sn-3-phosphorylcholine (oxPAPC). The different macrophage types can be distinguished by the expression of specific genes (highlighted in respective colors). (Adapted from Shalhoub *et al.*[180].)

or M2 macrophages[183;189].

Mammals infected with gram-negative bacteria develop severe syndromes including septic shock, the major reason for morbidity and mortality. Most of these effects can be mimicked by lipopolysaccharides (LPS) injection[190]. Lipopolysacharides are glycolipide components in the cell wall of gram-negative bacteria which are recognized by macrophage Toll-like receptor 4 (TLR4), inducing downstream signaling cascades which lead to M1 macrophage phenotype[185]. These processes have been intensively investigated using LPS stimulation of macrophage cell lines as model system. A few hours after LPS induction macrophages cannot induce many pro-inflammatory genes in response to a second LPS stimulation[191], while others can be restimulated (*e.g.* IL10). This different stage of macrophage activation is referred to as endotoxine tolerance[192], resulting in the alteration of the gene expression program towards an M2 macrophages like phenotype. Endotoxine

tolerance is a mechanism within the inflammatory process representing its self-regulatory properties. The precise regulation of LPS responses is important to preserve host defense while avoiding organ failure or death.

2.3.1 Signaling downstream of TLR4 activation

This regulation is ensured by multiple signaling cascades activated after LPS associates with Toll-like receptor 4[193]. The three main pathways involved in signaling downstream of TLR4 activation are NF-κB signaling and mitogen-activated protein kinases (MAPKs) signaling, which induce the production of pro-inflammatory genes[193], as well as signaling regulated by IFN regulatory factors (IRFs)[194–196], especially IRF3[186], which stimulate Interferon production (fig. 2.9).

For the recognition by TLR4, LPS binds to LPS-binding protein (LBP), followed by the association with the coreceptor CD14[197]. Now, LPS is transferred to the accessory molecule MD2, which is associated to the extracellular domain of TLR4, resulting in TLR4 oligomerisation[198] and activation of downstream signaling due to the recruitment of cytosolic TIR domain-containing adaptor molecules[199], *e.g.*, MyD88, MyD88 adaptor-like protein (Mal), TRIF, TRIF-related adaptor molecule (TRAM) (fig. 2.9). Based on the adapter molecules the signaling pathways employed by TLR4 are divided into MyD88-dependent pathways and MyD88-independent pathways. The MyD88-dependent pathway involves binding of Mal and MyD88 to TLR4, while in the MyD88-independent pathway TRAM and TRIF bind the receptor[200].

The mechanism by which TLR4 regulates two competing pathways has just recently been resolved by Kagan *et al.*[201]. They propose the initial activation of MyD88-dependent signaling at the plasma membrane followed by endocytosis of the TLR4 complex and TRAM/TRIF dependent signal transduction from early endosomes, where easier access to the adaptor protein Tumor necrosis factor (TNF) receptor-associated factor 3 (TRAF3) is achieved[201].

In MyD88-dependent signaling after MyD88/TLR4 association multiple phosphorylation and ubiquitinylation events of recruited proteins, like IRAK family members[202;203], TRAF6[204], TAK1 and NEMO[205], enable the independent induction of an IKK complex involving pathway as well as MAPK signaling[205] (fig. 2.9).

While MAPK signaling leads to the induction of gene expression downstream of Transcription factor activator protein 1 (AP1)[206] (fig. 2.9), the IKK complex (IKKα, IKKβ, NEMO) phosphorylates IκB the NF-κB inhibitor, marking it for degradation. Following degradation of IκB, NF-κB is

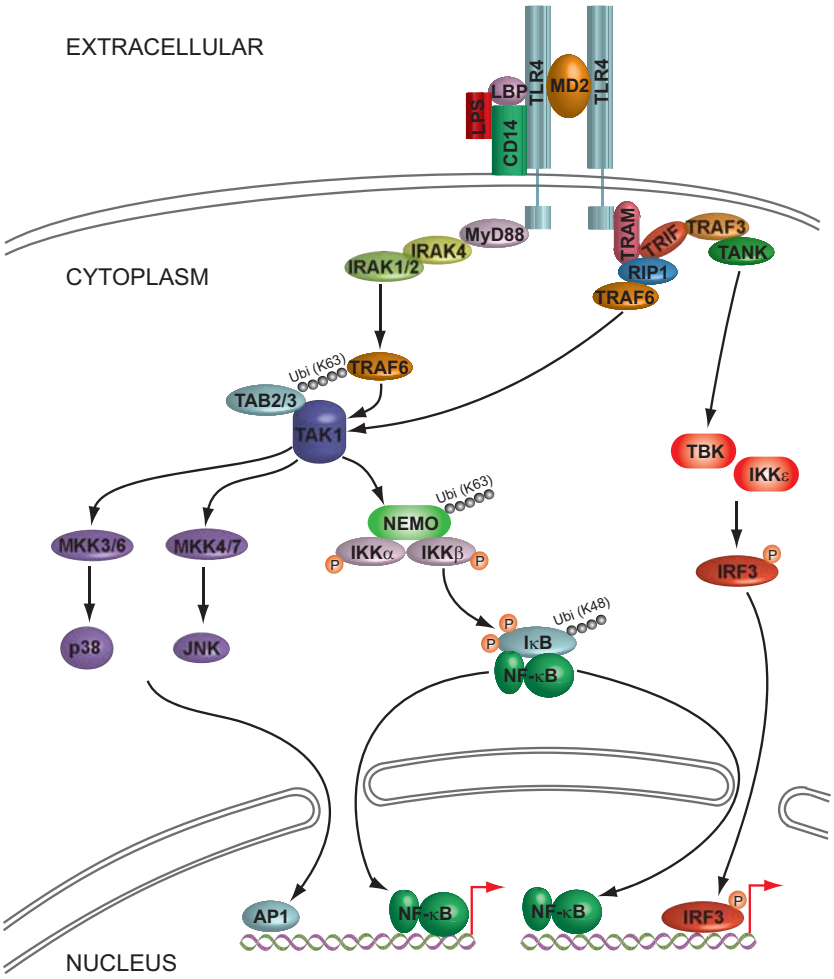

Figure 2.9: Toll-like receptor 4 signaling pathway.

released and translocates to the nucleus. Here, it activates the transcription of downstream genes.

The MyD88-independent pathway downstream of TLR4 is mediated by TRIF and TRAM and therefore also known as TRIF-dependent signaling pathway[207]. Here, TRAF3 and TANK transmit signaling to IKK-related kinases TBK1 and IKKϵ[208–211], which directly phosphorylate IRF3 and IRF7 (fig. 2.9). Phosphorylation results in the formation of homo- and heterodimers that are translocated to the nucleus to activate downstream signaling[212;213]. In addition, IRFs, AP1 and NF-κB associate to a multi-transcription factor complex called enhanceosome to induce INF-γ transcription[214]. Moreover, TRIF-dependend signals are transmitted by RIP1, TRAF6 and TAK1 leading to the activation of NF-κB and MAPK signal-

ing[215] (fig. 2.9).

In concert, the described pathways drive the precise regulation of gene expression resulting in the time-resolved induction of various genes downstream of LPS induction. This ensures the exact fine tuning of inflammatory responses conducted by macrophages, including the production of cytokines, chemokines and arachidonic acid metabolites, as well as the production of reactive nitrogen species and the reorganization of the cytoskeleton[216;217]. Thereby, the polarization of the innate and adaptive immune responses as well as the protection from invading pathogens is achieved.

2.3.2 Analysis of activated macrophages

Macrophage activation with LPS has been recently explored in several large-scale transcriptomic and qualitative or quantitative proteomic studies exploring different aspects of TRL4 signaling. *E.g.*, Ramsey *et al.* combined mRNA expression analysis using microarrays with transcription factor binding site motif scanning to investigate primary macrophages after stimulation with different substances including LPS. Thereof, they inferred a network of associations between transcription factor genes and clusters of co-expressed target genes[218]. Similarly, stimulation responses of macrophages to combinations of non-TLR4 ligands with LPS were used to elucidate pathway interactions by customized mRNA expression arrays[219]. It was found that IFN-γ and LPS cross-regulated the transcriptional response induced by each other.

By applying systems biological approaches Litvak *et al.* predicted and confirmed a regulatory network between the three transcription factors NF-κB, C/EBP Omega and ATF3 that discriminates between transient and persistent Toll-like receptor 4-induced signals[220].

Investigating the changes in protein expression levels of a mouse cell line as well as primary macrophages using iTRAQ labeling Swearingen *et al.* detected 36 proteins[221]. In the most comprehensive proteome wide quantitative study using SILAC labeling Du *et al.* revealed signaling and regulatory networks that systematically operate in the early response to LPS[222].

While many of the mentioned studies investigated the temporal profiles of mRNA expression, all of the data sets on protein level only focus on one or two time points after stimulation. This represents the lack of a comprehensive time-resolved investigation of protein abundance differences in response to LPS stimulation.

Furthermore, only Swearingen *et al.* integrated transcriptomic and proteomic data to differentiate between transcriptional and post-transcriptional events, revealing 75% of similarity between the two data sets[221]. Bhatt *et*

al. adressed this question using a different approach. They performed RNA sequencing of fractionated transcripts in a time-course study of lipid A (the functional compound of LPS) stimulated macrophages, elucidating a high-resolution map of coding and non-coding transcripts at three different cellular locations (chromatin-associated, nucleus, cytoplasm)[223]. In the chromatin fraction a high percentage on unspliced transcripts was found, suggesting that splicing often occurs after completion of transcription. In addition, the time-resolved transcription profiles were submitted to transcription factor motif enrichment, elucidating the temporal regime of transcription activation in response to stimulation[223]. Although this study gave impressive insights into regulation of mRNA processing as well as mRNA abundance and localization, regulatory processes on the level of translation were missed.

Weintz *et al.* and Sharma *et al.* published tremendous changes of the phosphoproteome upon LPS treatment[224;225], thereby addressing a different level of post-transcriptional regulation. The combination of phosphoproteomics and transcriptomic data at two time points after stimulation revealed novel signaling modules as well as the cytoskeleton as key target in LPS regulated phosphorylation. Furthermore, several transcription factors involved in TLR4 signaling were found to be phosphorylated[224].

In two alternative approaches Patel *et al.* and Dhungana *et al.* focused on specific aspects of differential protein localization after LPS treatment[226]. Patel *et al.* enriched microtubule associated proteins from LPS and IFNγ stimulated macrophages and identified 94 regulated proteins[226;227]. This differential binding pattern demonstrates the rapid modulation of the microtubule cytoskeleton in macrophages, essential to perform immune functions. Thereby, activated macrophages rapidly increase in size, as well as their phagocytic, secretory and migratory activity.

Dhungana *et al.* focused on proteins recruited to macrophage rafts in response to LPS stimulation for 5 or 30 minutes[227]. Selective activation of the proteasome in macrophage rafts and its inactivation outside of rafts was detected using unsupervised network analysis of the detected proteins. These studies demonstrate the extensive relocation of proteins in response to LPS stimulation.

A central process to signal to, attract or activate other cells of the immune system in response to stimulation of macrophages is the secretion of a variety of proteins including cytokines, representing a different aspect of protein translocation. This central process in inflammatory response has to the best of my knowledge not yet been investigated in a comprehensive and time-resolved manner.

Several of these studies used macrophage stimulation as a model system to investigate basic principles of gene expression, post-transcriptional or protein

localization regulation. Thereby, they elucidated a wide variety of mechanisms orchestrating inflammatory responses. While post-transcriptional processes involved in mRNA localization and splicing have been partly addressed, a full view on this regulatory processes can only be achieved by the integration of mRNA expression and protein synthesis data looking at different time points after LPS stimulation. Furthermore, the integration of this data with protein localization data like secretome analysis would disclose the sequence of events leading to macrophage activation, as well as the mechanism underlying this process.

Chapter 3

Identification of DNA-binding Proteins

The identification of proteins binding to cis-regulatory modules (CRMs) in a sequence specific manner can help to explore the mechanisms underlying transcriptional control. CRMs are DNA sequences with transcription factor binding sites which are clustered into modular structures, like promoters, enhancers, silencers, boundary control elements and other modulators (see chapter 2.1). Combinatorial and dynamic transcription factor binding to CRMs results in specific expression patterns. The identification of the proteins binding to a CRM is a necessary prerequisite to understand how its activity is controlled and, thereby, how its target genes are regulated.

Zinzen *et al.*[34] used ChIP-on-chip analysis to determine the genome-wide binding profiles of five transcription factors active in *Drosophila melanogaster* mesoderm development and it was shown that combinatorial binding can be predictive of spatio-temporal CRM activity. They found a significant amount of "plasticity" in terms of transcription factor binding to CRMs, meaning that significant differences in combinatorial transcription factor binding can give rise to virtually identical spatio-temporal CRM activity. Conversely, this study also showed that nearly identical transcription factor binding patterns for the five studied transcription factors can yield large differences in CRM activity. Therefore, additional proteins must play a role in the activation of these CRMs. Hence, in order to identify hitherto unknown interactors impinging on myogenisis, the binding repertoire of selected CRMs active at different times and in different tissues during *Drosophila melanogaster* development was probed.

The DNA-binding proteins were affinity purified from whole embryo nuclear extracts with streptavidin conjugated magnetic beads binding the DNA sequence of interest containing a biotin linker and an enzyme restriction site

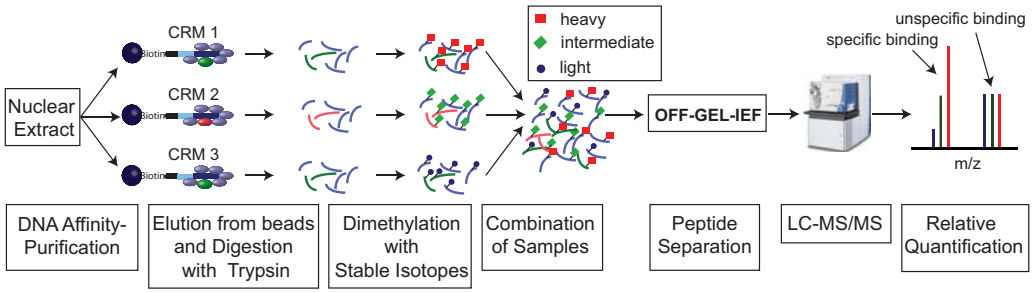

Figure 3.1: Experimental design for the detection of DNA-binding proteins: Streptavidin beads, bound to three different biotinylated DNA sequences (*i.e.* specific CRMs) bearing an enzymatic restriction site (light blue) and a spacer (black), are incubated with *Drosophila melanogaster* nuclear embryo extracts. Proteins are eluted from the beads followed by trypsin digestion. To distinguish specific (red and green) from non-specific (blue) binding, proteins bound to the different DNA sequences were quantitatively compared. Therefore, peptides are labeled with stable isotopes by dimethylation of peptide N-termini and lysines. After combination of the peptide samples, their complexity is reduced by OFF-Gel isoelectric focusing to increase the number of protein identifications. Finally, the samples are analyzed by nanoLC-MS/MS. The peak areas of the differentially labeled peptides are now used to calculate the relative abundance of the proteins bound to the different DNA sequences.

as described by Mittler *et al.*[8]. The bound proteins are released from the beads and analyzed by quantitative mass spectrometry using the stable isotope labeling by peptide dimethylation technique[228] (fig. 3.1). This approach does not only reveal the transcription factor binding to the chosen DNA sequence, but does also disclose proteins interacting with these transcription factors. Thereby transcription factor associated proteins possibly assisting them in their regulatory functions can be detected.

3.1 Method optimization

The described method has been used previously in several studies, all of which use short double stranded DNA sequences (∼40 bp) for the enrichment of binding proteins[65–71]. Using this approach a sequence mutated at the binding site can serve as a control to detect non-specific binding proteins. In contrast, actual CRMs are generally thought to be several hundred base pairs long[34]. Hence, the simple mutation of one or two nucleotides would not be sufficient to prevent specific binding of all potential interactors. In addition, the binding sites to be mutated are not known. To distinguish between

specific and non-specific binding the interaction patterns of different CRMs were compared. This approach should thereby allow to ignore non-specific DNA-binding proteins, but transcription factors interacting specifically and with equal affinity to all tested CRMs would also be missed. The application of this method to CRMs active at very different times and in different tissues should increase the chance that the identified differential binders may have a real impact on mesodermal gene regulation and mesoderm development. To avoid skewing the assay unnecessarily by using CRMs of different length, all used CRMs were adjusted to 500 base pairs to cover all potential transcription factor binding sites by Robert Zinzen.

Since 500 base pair long DNA sequences are not commercially available, they were produced by PCR amplification, using a biotinylated primer for the forward strand. Thereby, biotinylated double stranded DNA sequences are made in one step. Three methods for the concentration of the DNA after PCR were compared: DNA extraction form agarose gels, DNA purification with Qia-Quick PCR purification kit and phenol-chloroform-extraction. Phenol-chloroform extraction yielded more than twice as much DNA as the other methods and no contaminations were detected in the resulting DNA samples by gel electrophoresis and was therefore used in further experiments (fig. B.10).

To enrich for specific DNA-binding proteins the biotinylated DNA sequences were bound to streptavidin beads, followed by incubation with *Drosophila melanogaster* embryo nuclear extracts for three hours. Since the DNA-protein complexes are not covalently linked very mild washing conditions were applied. This was achieved by using four times the protein binding buffer and once the buffer necessary for enzymatic cleavage of the DNA sequence. It was found that this last washing step already releases bound proteins, resulting in losses of specifically binding proteins (fig. 3.2). Furthermore, after enzymatic digestion the DNA could not be detected on an agarose gel, while DNA cleaved from beads that were not incubated with nuclear extracts was well recovered (fig. 3.3).

To overcome this problem, different combinations and incubation times of SmaI and DNAseI were tested, but did not improve the protein release. Possibly the restriction enzyme recognition site was masked by proteins from the extract. In addition, SmaI cleavage in the protein binding buffer did not work, even when tested in solution. Therefore, the fifth bead washing step, shown to release proteins (fig. 3.2), could not be prevented.

Finally, alternative methods for protein release were explored. Very short boiling in 0.1% or 1% SDS seemed to be most effective, since boiling of the beads after elution does not result in additional protein release (fig. 3.4). Both methods resulted in a similar contamination with streptavidin. Hence,

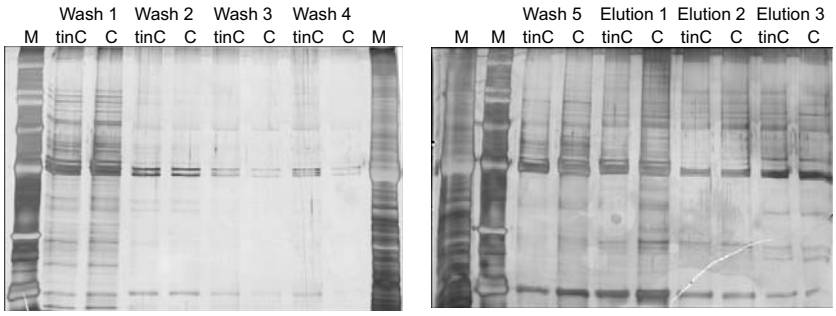

Figure 3.2: Release of proteins from DNA-protein complexes during washing procedure: Silver stained gels of washing buffers and eluates after protein enrichment using CRM tinC (tinC) or control beads not bearing a DNA sequence (C). For washing step 1-4 (Wash 1-4) the protein binding buffer was added, while washing step five (Wash 5) employed the nuclease restriction buffer. While after initial release of less tightly bound proteins during washing step one the protein amount released during washing decreases, washing step five results in a higher release of proteins. The elution steps one and two (Elution 1 and 2) contain proteins subsequently released by one hour of SmaI digestion and elution step three (Elution 3) is the supernatant after a final overnight incubation with SmaI.

for the final experiment 1% SDS was used to release the bound proteins. The resulting SDS containing protein sample was purified and digested using the filter aided sample preparation (FASP) method, which can accommodate high SDS concentrations[229].

In summary, changes in the published protocol were necessary to adapt the *in vitro* DNA-protein interaction assay to long DNA sequences. Since the enzymatic restriction site in the DNA sequence seems to be blocked by bound proteins, DNA release was only achieved by boiling in SDS. The thereby introduced background of proteins binding non-specifically to streptavidin beads was addressed by using a quantitative proteomics approach.

3.2 Comparison of proteins binding to cis-regulatory modules

Next, the optimized protocol was applied to the quantitative comparison of proteins bound to three CRMs, active at different times and in different tissues during *Drosophila melanogaster* development (table 3.1). The CRMs were incubated with *Drosophila melanogaster* embryo nuclear extract derived from two to ten hours old embryos, approximately covering stages 4 to 13 during embryo development.

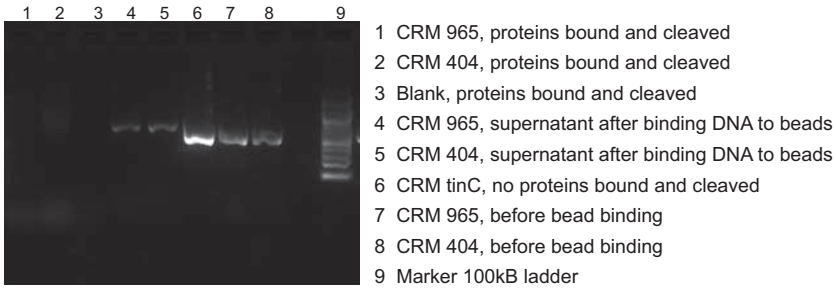

Figure 3.3: DNA release from streptavidin beads by enzymatic digestion: While the DNA is well recovered from beads before incubation with nuclear extract (lane 6), only minor or no recovery of DNA cleaved after incubation was detected (lane 1,2).

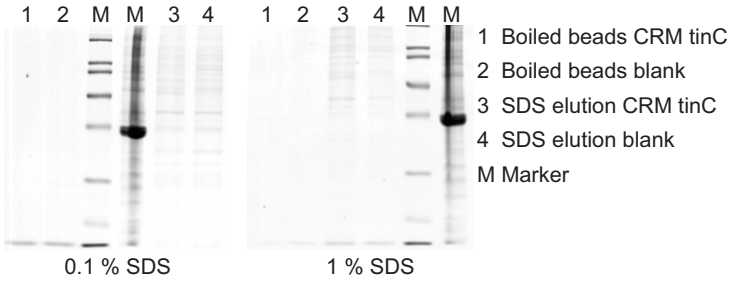

Figure 3.4: Elution of DNA-protein complexes by boiling in SDS: Recovery of proteins is similar for both SDS concentrations tested. In addition, protein composition released from DNA bound beads is highly similar to that from control beads. This highlights the need for a quantitative approach to distinguish between specific and non-specific binding.

To exclude any influence of the selected labels on the quantification results, two biological independent experiments with different labeling schemes were processed (table 3.2). As many as around 2800 proteins were quantified in each experiment. This high number can be explained by the very mild washing conditions, resulting in the non-specific binding of proteins, not only to the DNA sequence but also to DNA-bound proteins.

The 2565 proteins, quantified in both biological replicate samples, were subjected to functional annotation clustering using DAVID[230]. The strongest enriched clusters contain GO terms associated with "transcription regulation", "nucleotide-binding" and "chromatin organization", representing well known functionalities of DNA-binding proteins (table B.1). Notably, as many as 338 proteins are assigned to the GO term "transcription regulation". Other clusters like "ribonucleoprotein" or "RNA processing" may be due to highly

CRM	Stage	Tissue
CRM 633	5-8	Unspecified early mesoderm
bap3	10 onwards	Visceral mesoderm
tinC	late	Cardiac mesoderm

Table 3.1: Investigated CRMs: Stages and tissues give the temporal and spatial activity of these CRMs.

	light	int	heavy	μg DNA	mg protein	quantified proteins
Replicate 1	633	bap3	tinC	15	8.3	2854
Replicate 2	tinC	633	bap3	15	8.3	2777

Table 3.2: Experimental design and number of quantified proteins. The stable isotope labeling scheme for the two independent biological samples as well as the amounts of double stranded DNA and total protein in the nuclear extracts per CRM are given (int: intermediate).

abundant nuclear proteins non-specifically binding to the chosen DNA sequences (table B.1). In addition, two well known DNA-binding domains are highly enriched in the data set: Zinc finger, PHD-type[231] and Zinc finger, C2H2-like[232] (table B.1).

The quantitative analysis of the data revealed 35 and 26 proteins binding stronger to the CRMs 633 and bap3, respectively, as well as 17 proteins binding stronger to the CRM tinC than to either of the other two. The by far strongest enriched cluster retrieved by functional annotation clustering for the in total 72 differentially binding proteins contained terms associated with "transcription regulation" and "DNA-binding" (fig. 3.5). 48 proteins are grouped into this cluster including 14 proteins with the biological function "transcription factor activity".

Figure 3.6 displays the measured binding ratios for all differentially binding proteins. For example, Pannier (*pnr*), which is involved in the regulation of cardiac specification and differentiation[234], strongly binds to the tinC CRM, which is active in the cardiac mesoderm. Secondly, Biniou (*bin*), which plays a key role in the development of the visceral mesoderm and the derived gut musculature[235], interacts strongly with bap3, a CRM active in the visceral mesoderm[236]. These two proteins show the expected binding pattern, thereby validating the presented approach.

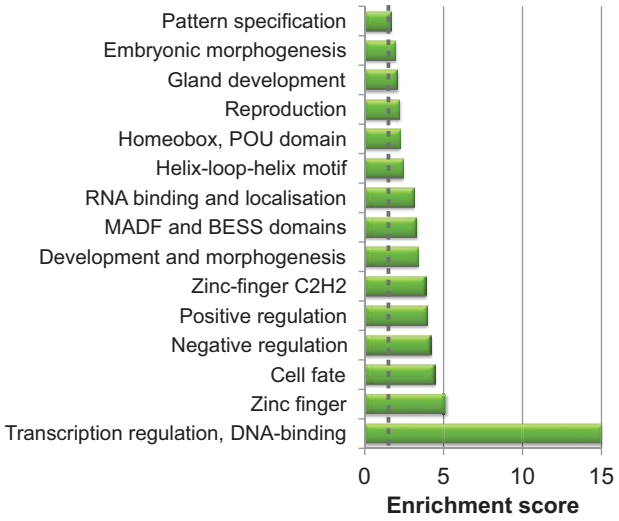

Figure 3.5: Functional annotation clustering for all differentially binding proteins.

The strongest difference (16-fold enrichment) was detected for *CG32772*[1], a protein with a zinc finger domain and DNA binding activity, which has not yet been extensively studied. Similar to *CG32772* the interaction of *lin-52*, *hb*, *sqz* and *Rox8* to the CRM 633 is more prominent compared to the other CRMs tested (fig. 3.6). A second protein exhibiting strong differences in CRM binding is Cyclic-AMP response element binding protein B at 17A (*CrebB-17A*), which is a transcriptional activator[237]. It strongly binds to the CRMs 633 and bap3, but seems to be excluded from tinC (fig. 3.6). Similar patterns are detected for *Hrb27C*, *Hrb87F*, *lark*, *bin* and *Ssb-c31a*. In contrast, *CG2990*, *Chi*, *ewg*, *pnr* and *so* tightly interact with tinC but to a lesser degree with CRM 633 or bap3 (fig. 3.6). Preferential bap3 binding was detected for *Kr-h1*, *wor*, *Su(H)*, *mod(mdg4)*, *hang* and *baf*. *Wor* is a zinc finger protein closely related to *snail* and *esg*. Therefore, the strong binding of *wor* and *esg* to bap3 could represent interaction with a *snail* binding site, which might be required for keeping this CRM inactive early in development[238]. Decreasing amounts of protein binding from CRM 633 via tinC to bap3 were measured for *nub*, *pdm2*, *Rbf2* and *vvl* (fig. 3.6), possibly representing differences in either numbers or affinity of the binding sites for these proteins.

Since all tested CRMs are active at different times in different mesodermal tissues the binding of these candidate proteins to the CRMs could point

[1]All detected proteins will be mentioned by using the associated gene symbol. Associated protein names are collected in appendix D.

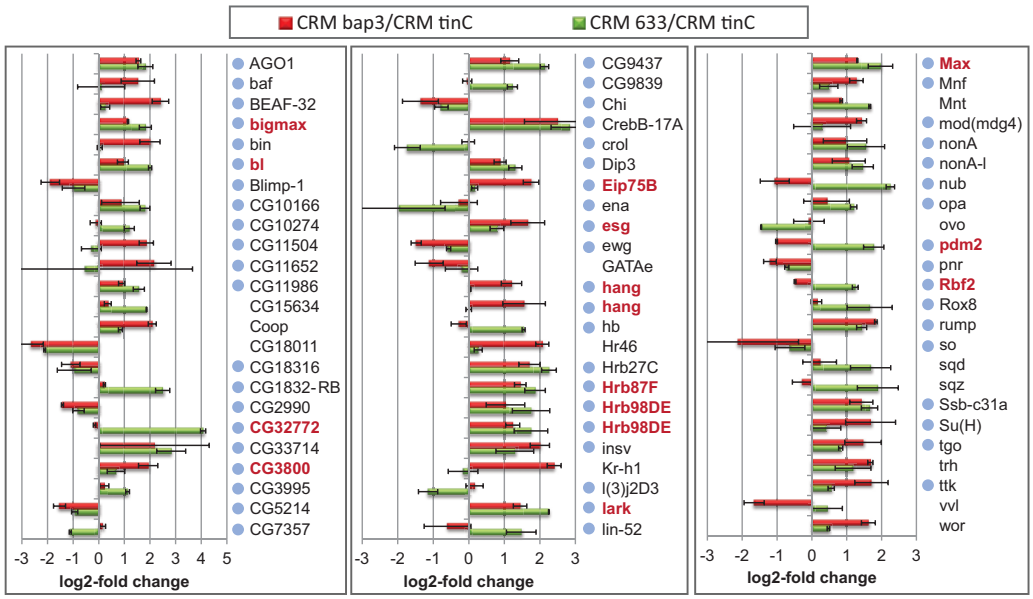

Figure 3.6: Proteins binding differentially to the investigated CRMs: Log2-fold-changes are shown in bar plots. Proteins more strongly bound to CRM 633 or bap3 exhibit positive log2-fold-changes, while proteins more strongly bound to tinC have negative log2-fold-changes. Blue dots highlight proteins with mesodermal expression in four to eight hour old embryos based on chromatin data (Berkeley Drosophila Genome Project)[233]. Genes highlighted in red were validated by *in situ* staining (fig. 3.7). Gene names appearing twice represent different isotopes of proteins distinguished by mass spectrometric identification. In both cases the binding patterns are similar.

to regulatory interactions and especially to activating roles if the candidate proteins themselves are expressed in the mesoderm. Alternatively, no expression of a protein in the mesoderm could point towards an inhibitory function of a protein on mesodermal active CRMs. Mesodermal transcription in four to eight hour old embryos was detected using chromatin data for 59 of the 72 significant proteins (Berkeley Drosophila Genome Project)[233]. These proteins could thereby activate the CRM they bind to, while the proteins not expressed in the mesoderm could act as repressors.

13 proteins were selected for further validation (fig. 3.6). For selected proteins a fold-change of at least three and and likely DNA-binding motifs were required. *In situ* staining against the corresponding genes of seven proteins determined their expression at the same time and place where the strongly bound CRM is active, *e.g.*, CRM 633 in the early mesoderm. Therefore, they could feasibly be activators of their bound CRM.

The three proteins excluded from the mesoderm (*pdm2*, *esg*, *CG32772*)

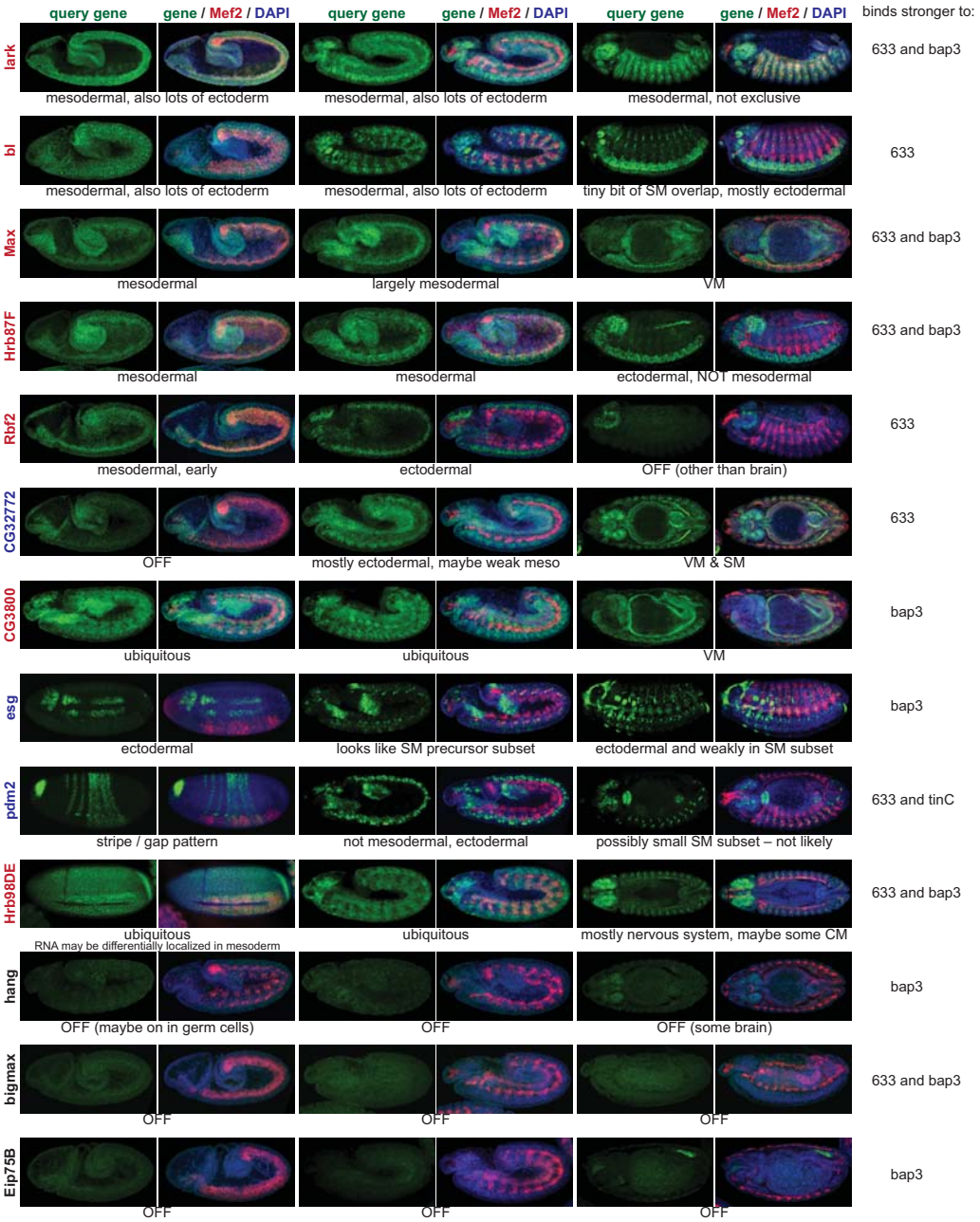

Figure 3.7: *In situ* staining of selected proteins at three different stages during *Drosophila melanogaster* development. The protein of interest is shown in green, the pan-mesodermal marker (*Mef2*) in red and DAPI in blue. (Genes expressed at the same time and in the tissue, where the CRM they preferentially bind to is active, are highlighted in red. Genes excluded from mesoderm are highlighted in blue. Both represent candidates potentially regulating the respective CRMs.)

could have inhibitory function on the CRM they strongly bind to. One of them is Escargot (*esg*), which is known to be a repressor inhibiting transcription from promoters containing E2 boxes activated by bHLH proteins[239]. Thus, *esg* is a very promising candidate possibly inhibiting bap3 activity. Similarly, *CG32772* was excluded from the early mesoderm, where CRM 633 is active. Since it has been shown to strongly bind to this CRM it could inhibit it in other tissues and at different times.

The last three proteins were not detected by *in situ* staining (*hang*, *bigmax*, *Eip75B*). This means they are probably not expressed within the investigated times. Since they were detected by mass spectrometry they are expressed at different time points covered by the nuclear extracts prepared from two to ten hour old embryos. Therefore, they could still impinge on the activity of the enhancers they bind.

Based on the described results further validation using loss of function mutants, deficiency mutants or RNAi repression as well as overexpression constructs of selected proteins are in progress. The impact of the loss or overexpression of this proteins on muscle development will be determined.

3.3 Conclusion and outlook

Here, an *in vitro* DNA affinity capture approach was for the first time successfully applied to the detection of thousands of potential CRM binding proteins. Unlike previously published approaches that used around 40 base pair long DNA sequences, the method was adapted to the determination of proteins binding to DNA sequences as long as 500 base pairs. Hence, emerging challenges were the detection of hundreds to thousands of non-specifically binding proteins and the shielding of the enzymatic restriction site for DNA release. These challenges were addressed by adaptions of the protocol. Firstly, non-specifically binding proteins were distinguished from specific binders by a quantitative proteomic strategy applying restrictive statistics. Secondly, DNA release was facilitated by SDS elution.

The disadvantage of this method is on the one hand the contamination of samples with SDS, repressing peptide ionization in the mass spectrometric measurement and on the other hand the elution of contaminating bead binding proteins as well as streptavidin. However, SDS was removed from the samples prior to analysis by a special protein digestion protocol involving extensive buffer exchange. Presuming equal background binding to the beads for the different CRMs tested, the quantitative approach leads to the exclusion of these proteins from further analysis.

The detected proteins are highly enriched in proteins with functions con-

nected to DNA binding activity, like transcriptional regulation or cell cycle regulation. Quantitative comparison of the proteins bound to three CRMs active at different times during *Drosophila melanogaster* embryo development and in different tissues, revealed 72 proteins potentially involved in the regulation of these CRMs.

A first validation confirmed the expression of 7 out of 13 tested proteins at the right time and place in the embryo. Additional three proteins which are excluded form the tissue where the bound CRM is active, could repress their target CRMs in other tissues and at different times. Based on these results further validation using loss of function mutants, deficiency mutants, RNAi mediated repression as well as overexpression constructs of selected proteins are in progress.

A more comprehensive view could be gained by including additional CRMs active at similar stages or tissues. Thereby, combinatory binding patterns driving the activity of CRMs in certain tissues could be elucidated. In addition, nuclear extracts from more refined developmental states could be used. *E.g.*, the embryo collection times could be adjusted to the developmental stages when CRMs are active, which may allow a better understanding of the activating and repressive regulatory interactions that determine CRM activity. The challenge here would be the quantitative proteomic comparison of CRMs active at different stages. The use of different nuclear extracts in one experiment could complicate the detection of non-specifically binding proteins.

In summary, extensive DNA-binding profiles of three CRMs were compared for the first time, resulting in the detection of candidates potentially regulating the activity of these CRMs. These candidates were distinguished from thousands of non-specifically binding proteins by quantitative proteomics. Since the comparison was performed *in vitro*, *in vivo* validation of biological activity is in progress and will probably reveal new circuits of *Drosophila melanogaster* muscle development.

The disadvantage of the applied approach is that the *in vitro* retrieved data not necessarily represent the *in vivo* binding pattern. Therefore, a careful validation of potential interactions *in vivo* is needed. For the elucidation of *in vivo* binding activity DNA sequences tagged with an enrichment functionality could be introduced (*e.g.* by transfection in cell culture systems) and DNA-protein complexes could be stabilized by crosslinking approaches. The challenge of such an approach is the enrichment sufficient amounts of bound proteins for mass spectrometric detection.

Chapter 4

Quantification of Newly Synthesized Proteins

Transcription and translation are regulated at several different steps during gene expression, ultimately resulting in the synthesis of new proteins. Knowing which proteins are synthesized at a specific time will aid in unraveling the temporal dynamics of the proteome. In addition, this information will help to understand the regulatory gene expression responses to environmental stimuli. Since newly synthesized proteins represent a small subset of the total cellular protein pool, efficient strategies that facilitate their enrichment are necessary. This can be achieved using non-canonical amino acids that are selectively incorporated in newly synthesized proteins. In this chapter, a methodology based on the use of non-canonical amino acids coupled with click-chemistry to enrich for newly synthesized proteins is described. Using this method in combination with stable isotope labeling with amino acids in cell culture (SILAC), quantitative analysis of lipopolysaccharides (LPS) stimulated mouse macrophages (RAW 264.7) was performed. Moreover, to establish the best experimental setup, alternative approaches to quantify changes in protein synthesis were compared.

4.1 Capturing newly synthesized proteins

For the selective enrichment of newly synthesized proteins, a non-natural amino acid that is an azido-analogue of methionine, called azidohomoalanine (AHA) is utilized. By supplementing culture medium depleted of methionine, AHA will be metabolically incorporated into newly synthesized proteins by translationally active cells. Using a copper-catalyzed 1,3-cycloaddition, the AHA labeled proteins are coupled to either an alkyne reagent with a biotin

group or an alkyne-activated agarose resin (fig. 4.1).

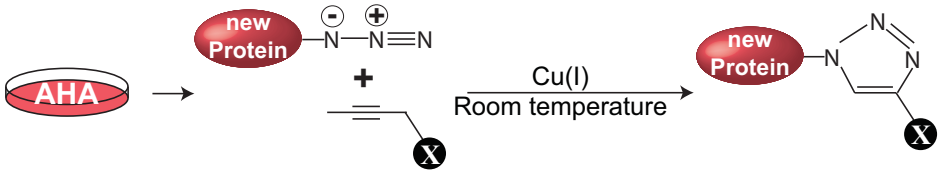

Figure 4.1: Enrichment of newly synthesized proteins: AHA is incorporated into all newly synthesized proteins in cultured cells using methionine-free medium. The azido-group can be used to couple any alkyne, either bearing a biotin or coupled to an agarose resin (X), which both can be used for the selective enrichment of newly synthesized proteins or in case of biotin for detection by western blot.

The biotin group can be used for affinity purification or detection of newly synthesized proteins by western blot. The alkyne-activated agarose resin is beneficial for affinity purification, because the covalent capture of newly synthesized proteins permits usage of stringent washing conditions, reducing non-specific interactions. In addition, NeutrAvidin is omitted, which represents an abundant contamination after on-bead digestion of the captured proteins.

The incorporation of AHA into t-RNA is slower than the introduction of methionine[122]. This could perturb cellular functions, leading to starvation or cell death. To investigate the potential effects on viability AHA treated RAW 264.7 cells were compared with those cultured in the presence of methionine. From this, no significant differences between RAW 264.7 cells cultured in either condition was observed (fig. 4.2a).

To further investigate the potential effects of substituting AHA for methionine, translation characteristics were measured using polysome profiling. Examination of ribosome occupancy revealed a translation index of 0.62 compared with 0.81 for RAW 264.7 cells treated for two hours with AHA and methionine, respectively (fig. 4.2b). Based on these data, a 25% reduction in the rate of translation can be estimated. Since the aim of this study was to compare protein synthesis under two different conditions, this reduction in global protein synthesis should affect the cells which are compared in a similar way and should therefore not influence the results of the experiment. This hypothesis is further tested in section 4.2.

To assess the metabolic incorporation rate, RAW 264.7 cells were treated for 5 to 30 minutes with AHA with sampling every 5 min. Cells grown in methionine containing medium were used as controls. Dot blot analysis of cellular protein extracts demonstrates the successful tagging of AHA containing proteins with the alkyne-biotin reagent (fig. 4.3a). Furthermore, newly

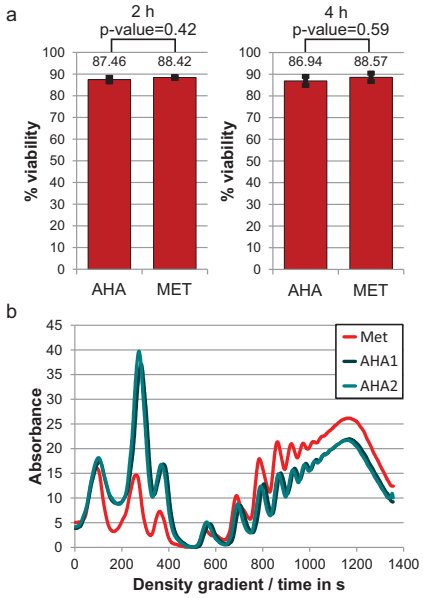

Figure 4.2: Viability and translation rate in AHA treated mouse macrophages: a) Viability of RAW 264.7 macrophages after 2 h and 4 h of AHA treatment determined with the trypan blue viability test: No significant decrease in viability was detected using a Student's t-test (p-value>0.01). b) Polysomen-profiling of RAW 264.7 cells after 2 h of AHA treatment (AHA1, AHA2). Cells grown in methionine containing medium were used as control (Met).

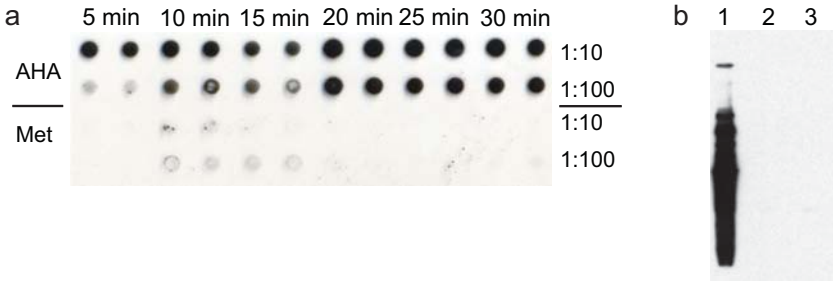

Figure 4.3: Incorporation of AHA into newly synthesized proteins: a) Dot blot (anti-biotin antibody) after treatment of RAW 264.7 cells with AHA for different times and labeling of AHA containing proteins with biotin containing alkyne. Incorporation of AHA can already be detected after 5 min of treatment. Two dilutions of each sample were blotted (as indicated next to the figure) and each of the samples was blotted twice. b) Inhibition of protein synthesis: Western blot (anti-biotin antibody) after AHA treatment of RAW 264.7 cells with (1) or without (2) cycloheximide for 2 h. (Cells treated with methionine were used as negative controls (3).) AHA is selectively incorporated into newly synthesized proteins that span a broad protein mass range.

synthesized proteins were detected in all samples with increasing amounts from 5 to 30 minutes, demonstrating the rapid and progressive incorporation of AHA. When protein synthesis was blocked using cycloheximide, no staining of biotin is observable by western blot, further validating the selective incorporation of AHA into newly synthesized proteins (fig. 4.3b).

After metabolic labeling with AHA, cellular proteins were affinity purified on NeutrAvidin agarose beads. As expected, all biotinylated proteins were bound to the beads and subsequently removed from the beads during enzymatic digestion (fig. 4.4). The detected newly synthesized proteins span a broad mass range, demonstrating the unbiased incorporation of AHA. The

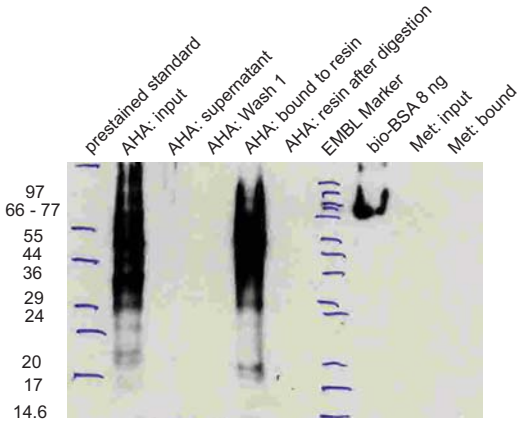

Figure 4.4: Enrichment of AHA containing proteins: RAW 264.7 cells were treated with AHA for 2 h and lysed. AHA containing peptides were coupled to an biotin-alkyne using click-chemistry and enriched. The western blot highlights the efficient enrichment of AHA containing proteins with NeutrAvidin beads and the successful release of newly synthesized proteins from the beads by trypsin digestion.

analysis of these samples using mass spectrometry revealed contamination with NeutrAvidin, which is introduced by the on-bead digestion. To omit this problem, several approaches were tested. First, the purification was performed using magnetic steptavidin beads instead of NeutrAvidin agarose beads. However, this approach did not result in any reduction in the amount of contaminating peptides. Elution of the proteins by boiling them in SDS-PAGE loading buffer resulted in a strong NeutrAvidin band on the gel. An insufficient number of proteins was identified when monomeric avidin beads in combination with biotin elution were used for enrichment, suggesting significant losses during the protocol. This problem was overcome by adding a peptide separation step based on off-gel isoelectric focusing (OFF-Gel IEF). The use of IEF fractionation increased the number of identified proteins

tremendously, leading to the detection of low-abundant regulated proteins, like NF-κB and AP1 transcription factor family members (*e.g. Rel* and *Fos*). An additional solution was provided by the recent commercialization of an alkyne-agarose resin, allowing covalent capture of newly synthesized proteins and on bead enzymatic digestion without NeutrAvidin contamination.

In summary, AHA incorporation into newly synthesized proteins was effective, selective and rapid. RAW 264.7 cells are viable for at least four hours in presence of AHA. However, during this period a reduced translation rate was observed. Lastly, newly synthesized proteins were quantitatively enriched using NeutrAvidin beads. An additional peptide fractionation step after on-bead digestion compensated for the high background of NeutrAvidin derived peptides. However, the use of alkyne-activated agarose beads will fully omit this problem.

4.2 Quantification of newly synthesized proteins - a comparison of approaches

Having established an efficient method to capture newly synthesized proteins, I aimed to quantify differences in protein synthesis between two conditions. Hence, the combination of AHA labeling with a quantitative approach was necessary. The labeling protocol for AHA lends itself well to combination with another metabolic approach referred to as SILAC. Quantitative SILAC experiments utilize metabolic incorporation of isotopically labeled amino acids, typically arginine and lysine. To establish the best experimental setup, four alternative methods to quantify changes in protein synthesis or total protein abundance after LPS stimulation of mouse macrophages (RAW 264.7) were compared (fig. 4.5). To detect differences in protein synthesis LPS treated RAW 264.7 cells were compared to untreated controls at two different times after LPS addition in two biologically independent samples (fig. 4.6).

Changes in total protein abundance were measured using the default SILAC approach[240;241]. Fully SILAC labeled cells (*e.g.* heavy) were pulse labeled with AHA during LPS treatment for two or three hours. For comparison, a parallel population of control cells bearing the alternative SILAC label (*e.g.* light) were only pulse labeled with AHA (fig. 4.5a). After the desired treatment time the cells were combined and lysed. The samples were directly digested and the resulting peptide mixtures fractionated using IEF prior to analysis using nanoLC-MS/MS. For relative quantification of total protein abundance the light and heavy peak areas of the same peptides were

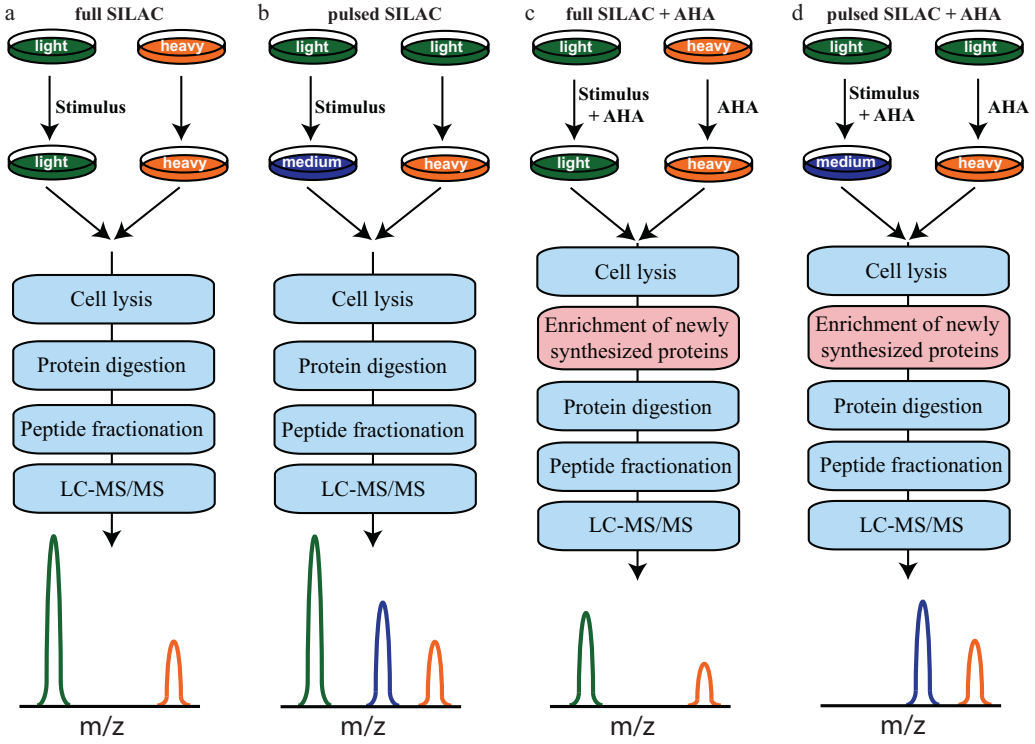

Figure 4.5: Quantification strategies for newly synthesized proteins or all cellular proteins: a) Full SILAC labeling for relative quantification of all cellular proteins. b) Pulsed SILAC for the relative quantification of protein synthesis. c) Full SILAC combined with AHA labeling and enrichment of newly synthesized proteins for the relative quantification of protein synthesis. (Biotin-alkyne in combination with NeutrAvidin beads was used.) d) Pulsed SILAC combined with AHA labeling and enrichment of newly synthesized proteins for the relative quantification of protein synthesis. (Alkyne-activated agarose beads were used.)

used. 3400 and 3397 proteins were quantified after two and three hours of LPS stimulation, respectively (fig. 4.6a,b). Significant differences (FDR 1% and fold-change >2) in total protein amount were detected for 17 and 7 proteins in response to two and three hours of LPS treatment, respectively (fig. 4.6a,b).

Previously, pulsed SILAC (pSILAC) was described for the quantification of changes in protein synthesis[112]. To compare this approach with that introduced here, the pSILAC protocol as described in chapter 2.1.4 was applied (fig. 4.5b), followed by the same sample processing as for the fully SILAC labeled samples. Triple labeling was achieved by using $^{13}C_6$-arginine and d_4-lysine as medium-heavy labels or $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine as heavy labels (fig. 4.5b). Now, for each peptide three peaks are expected.

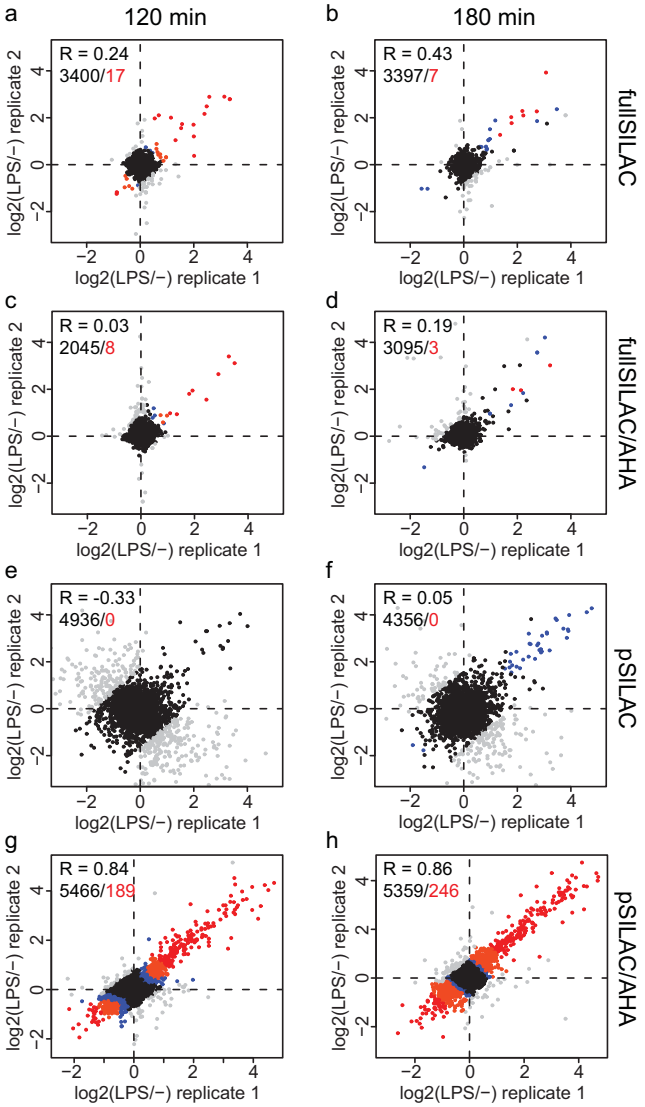

Figure 4.6: Comparison of the different approaches: Log2-fold-changes of two biological replicates comparing LPS treated cells against untreated control cells are plotted for two treatment times (120 min and 180 min). a, b) Comparison of changes in total protein abundance using full SILAC. c, d) Comparison of protein synthesis using full SILAC and biotin for enrichment. e, f) Comparison of protein synthesis using pSILAC. g, h) Comparison of protein synthesis using pSILAC and alkyne-activated agarose beads for enrichment. Colored dots highlight proteins with significant changes. Red: FDR 1% and fold-change >2; orange: FDR 1%; blue: FDR 5%; grey: quantified with high standard deviation and therefore excluded from significance test. The correlations between replicates (R) as well as the number of proteins quantified in both replicates (black) and the number of significant proteins (FDR 1% and fold-change >2) (red) are shown.

Here, the light peak is derived from all peptides of "old" proteins and the medium-heavy and heavy peaks represent the newly synthesized proteins in the corresponding condition and can therefore be used for relative quantification of protein synthesis. Although, as many as 4936 and 4356 newly synthesized proteins were quantified after two and three hours of LPS stimulation, respectively, no statistical significant differences in protein synthesis (FDR 1% and fold-change >2) were detected (fig. 4.6e,f).

Next, full SILAC and AHA labeled samples was subjected to newly synthesized protein enrichment using the biotin-alkyne in combination with NeutrAvidin beads (fig. 4.1). Following extensive washing the proteins were submitted to on-bead digestion. The peptide mixtures were subsequently fractionated and measured using nanoLC-MS/MS (fig. 4.5c). The expectation is that the resulting peptide spectra contain only peptides derived from newly synthesized proteins. Quantitative comparison of protein synthesis using the SILAC labels was achieved for 2045 and 3095 proteins at two and three hours after LPS stimulation, respectively (fig. 4.6c,d). Eight and three proteins were differentially synthesized (FDR 1% and fold-change >2) after stimulation (fig. 4.6c,d).

Finally, pulsed SILAC was combined with AHA labeling and stimulation (fig. 4.5d). After the combination of differentially labeled cells and cell lysis, newly synthesized proteins were covalently coupled to alkyne-activated agarose beads. Peptide fractionation and measurement was performed as before. Again, the light peak represents the peptides derived from "old" proteins, but is expected to be mainly absent, since those proteins did not incorporate AHA and are therefore removed during enrichment. The medium-heavy and heavy peaks are used for relative quantification of protein synthesis revealing synthesis rations for 5466 and 5359 proteins after two and three hours of LPS stimulation, respectively (fig. 4.6g,h). In addition, as many as 189 and 246 proteins showed significantly different protein synthesis (FDR 1% and fold-change >2) when comparing LPS treated samples to untreated controls (fig. 4.6g,h).

Table 4.1 highlights the advantages of the applied strategies. It should be noted that the differences in newly synthesized proteins are established by a combination of protein synthesis and degradation. For convenience they are referred to as "protein synthesis" in this thesis.

The comparison of the tested experimental setups reveals, that similar numbers of proteins were quantified in the two approaches using full SILAC labeling (fig. 4.6a-d). However, approximately 2000 more proteins were quantified using the pSILAC approaches (fig. 4.6e-h). Two possible explanations are the use of more advanced instrumentation for the pSILAC experiments, or the higher amount of starting material used. The amount of starting

	fullSILAC	pSILAC	fullSILAC/AHA	pSILAC/AHA
all proteins quantified	+	-	-	-
NSP quantified	-	+	+	+
selective labeling of NSP*	-	+	-	+
depletion of "old" proteins	-	-	+	+
covalent capture of NSP	-	-	-	+

Table 4.1: Advantages of the different approaches (NSP: newly synthesized proteins; *Additional labeling of newly synthesized proteins with stable isotopes gives further confidence that a protein is truly newly synthesized and not a contamination.)

material in the full SILAC experiments was defined by the on-bead digestion, which resulted in a high background of NeutrAvidin derived peptides, limiting the relative amount of original sample, to be loaded on the HPLC column. Only low correlations between biological replicates, ranging form 0.03 to 0.43, were calculated for those samples, explaining the low numbers of proteins with significant differences between the compared samples. In addition, low abundant proteins with differential expression or synthesis were possibly missed due to the lower numbers of quantified proteins.

The correlations between biological replicate samples for pSILAC are with 0.05 and -0.33 even lower (fig. 4.6e-f). Although proteins with high standard deviation were excluded, no significant proteins could be detected applying the mentioned parameters. Increasing the false discovery rate to 5% recovers 36 significant proteins after three hours of LPS treatment, but did not improve the detection at two hours (fig. 4.6e-f). These results again demonstrate the limitation of stable isotope labeling approaches for the measurement of protein synthesis after short isotope incorporation times. Hence, this method is not well suited for the investigation of rapid responses to LPS stimulation.

A tremendous improvement was achieved when pSILAC was combined with AHA labeling and enrichment (fig. 4.6g-h). Despite similar numbers of quantified proteins the correlation between biological replicates increased to 0.84 and 0.86. As highlighted before, as many as 189 and 246 differentially synthesized proteins could be detected for two hours and three hours of LPS treatment, respectively. The depletion of peptides derived from "old" proteins, represented by the light peak in each peptide spectrum, improves the signal-to-noise ratio for medium-heavy and heavy signals, which are used for the quantification of protein synthesis. This is especially important for short

labeling times, as applied here, resulting in a low relative amount of newly synthesized proteins compared to "old" proteins. Therefore, the light signal will be much more intense than the signals used for quantification.

To conclude, for the measurement of rapid changes in protein synthesis after LPS stimulation of mouse macrophages the combination of AHA incorporation with pSILAC labeling was best suited. Enrichment of low abundant newly synthesized proteins using alkyne-agarose beads increased the signal-to-noise ratio of the mass spectrometric measurement, resulting in a high sensitivity for the detection of stimulation responses. The advantage of pSILAC labeling compared to full SILAC is the selective labeling of newly synthesized proteins, introducing a possibility to discriminate between newly synthesized proteins and non-specifically binding proteins.

Since a lower translation rate for AHA treated macrophages was observed (fig. 4.2), the impact of AHA on protein synthesis was evaluated. The comparison of protein synthesis differences determined with pSILAC to changes measured after AHA and pSILAC labeling and enrichment reveals correlations of 0.45 and 0.69 for two hours and three hours of LPS treatment, respectively. At both time points only a few proteins show induced synthesis in AHA treated samples, while no change was detected in untreated samples (fig. 4.7). Therefore, a major impact of AHA on protein synthesis for the investigated treatment times can be excluded.

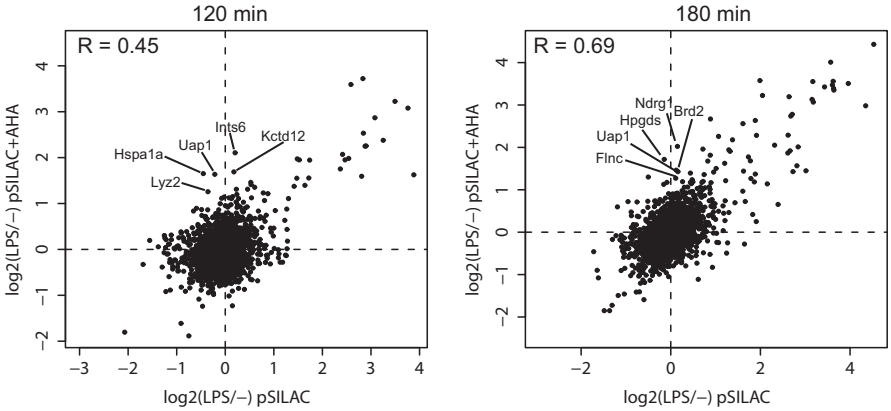

Figure 4.7: Comparison of quantitative results from AHA treated and untreated samples. In both approaches LPS treated macrophages were compared to untreated cells after two and three hours of stimulation. pSILAC was used for quantification. In AHA treated samples newly synthesized proteins were enriched.

Since the enrichment strategy depends on the replacement of methionine by AHA, the detection of proteins enriched in this amino acid could be favored. Therefore, distributions of identified proteins in only pSILAC and

pSILAC combined with AHA enrichment samples were compared to each other and to the complete proteome, as defined by the database used for protein identification (fig. 4.8).

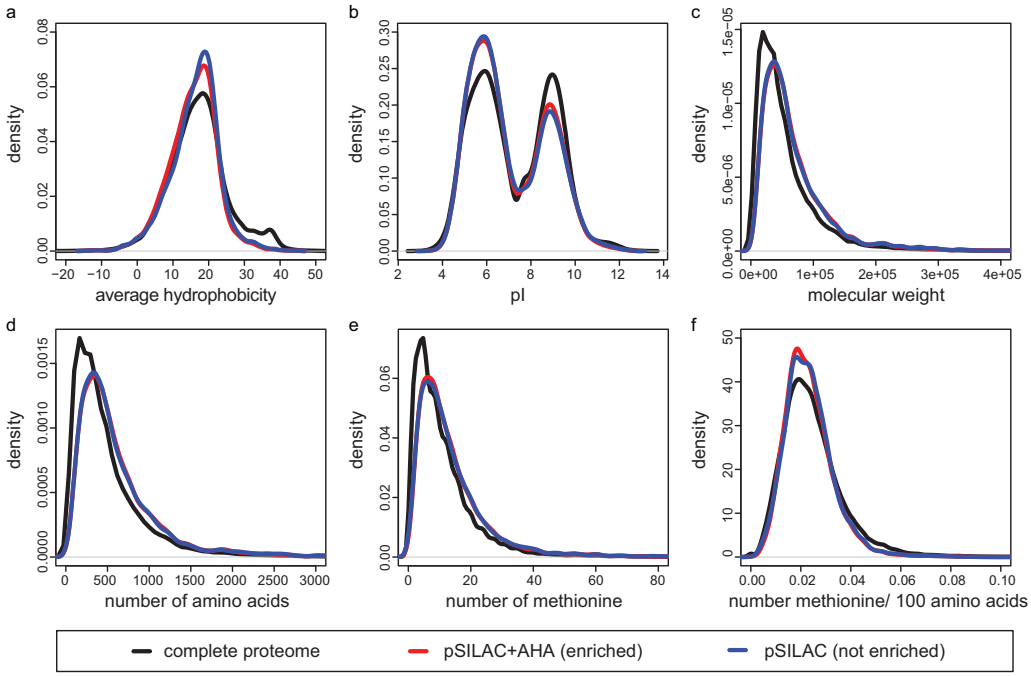

Figure 4.8: Distribution of protein properties of detected newly synthesized proteins: a) average hydrophobicity, b) isoelectric point (pI), c) molecular weight, d) sequence length, e) methionine frequency, f) relative methionine frequency normalized by sequence length. The "complete proteome" was defined by all proteins contained in the database used for protein identification.

Although, a bias in the proteomic data sets is detected for most of the investigated protein characteristics, this was similar for the AHA enriched and not enriched samples. Hence, the detection bias is rather a property of the mass spectrometric detection method, as described in literature[242], than of the selected enrichment protocol. This is highlighted by the observed trend against the detection of proteins with low numbers of methionines (fig. 4.8e). This bias disappears when using the relative number of methionines normalized for sequence length (fig. 4.8f). This is indicative of the well-known underrepresentation of small proteins in mass spectrometric analyses[242] (fig. 4.8d) and causes a similar detection bias towards proteins with low numbers of methionines here.

In summary, it was shown that on-bead digestion of newly synthesized proteins from NeutrAvidin beads is disadvantageous, since it produces a high

background of NeutrAvidin derived peptides, complicating protein quantification. AHA incorporation followed by enrichment with alkyne-activated agarose beads in combination with pSILAC for the quantification of newly synthesized proteins outperforms pSILAC for short labeling times. Furthermore, although AHA treatment reduces the translation rate, it does not cause major differences in the changes detected after LPS treatment of mouse macrophages during the investigated time frame. In addition, AHA labeling and enrichment of newly synthesized proteins does not introduce a protein detection bias.

Chapter 5

Quantitative Secretome Analysis

Secreted proteins constitute a large and biologically important subset of mammalian proteomes involved in cellular communication, adhesion and migration. Yet, secretomes are understudied because of technical limitations in the detection of low-abundant proteins against a background of serum used to sustain cell culture. The method described in the previous chapter was adapted for the selective enrichment and quantification of secreted proteins irrespective of a complex protein background. The approach solves several problems in todays secretome analysis, meaning that cells can be grown in the presence of serum, while preserving the capacity to identify a large repertoire of low-abundant secreted proteins. In addition, the method is quantitative enabling accurate secretome comparisons of different cell types, or of cells before and after stimulation. Its utility in the in-depth and differential analysis of secretomes will be demonstrated and a unique application studying the kinetics of protein secretion upon cellular stimulation will be introduced. The method gives access to a poorly covered but biologically important part of the proteome, contributing to an increased understanding of cellular communication and responsiveness and with a strong potential in biomarker discovery.

5.1 Accurate quantification of secreted proteins in complete culture media

The key aspect of the approach is the combined metabolic pulse-labeling of proteins with SILAC amino acids (pSILAC)[112;243], as well as azidohomoalanine (AHA), an azide-bearing analogue of methionine[11]. Pulse-labeling with

AHA allows the selective and covalent capture of newly synthesized proteins to an alkyne-activated resin via click-chemistry (fig. 4.1) and permits stringent washing conditions. The SILAC label serves to quantify protein levels in the ensuing mass spectrometric analysis, at the same time providing a mark to distinguish new from pre-existing proteins (fig. 5.1).

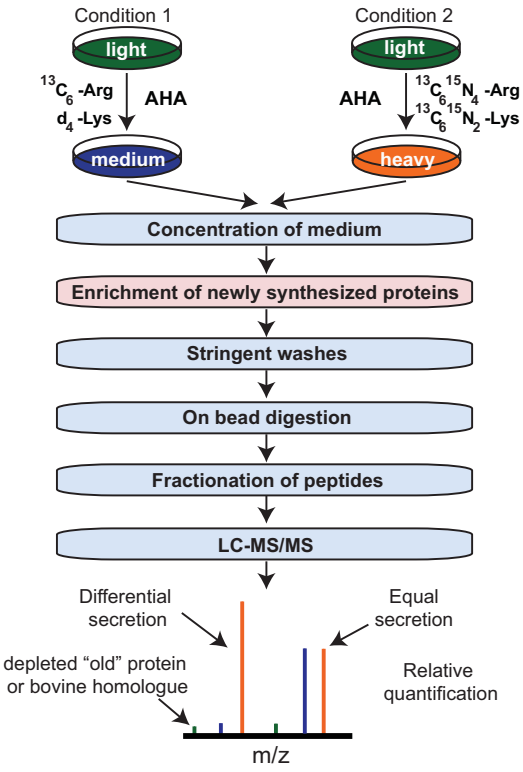

Figure 5.1: Quantitative analysis of cell secretomes. By combined pulse labeling with AHA and stable isotope-labeled amino acids, secreted proteins can be enriched from the high protein background of cell growth media and are quantified using mass spectrometry after on-bead digestion and (optionally) peptide fractionation.

First, the benefit of AHA labeling and protein enrichment in combination with pSILAC, compared to pSILAC alone, to identify secreted proteins in the background of complete cell growth medium, *i.e.* in the presence of 10% bovine serum, was assessed. At the same time, it was explored if this approach can be used for in-depth and quantitative differential secretome analysis. Therefore, two cell lines (PC3 cells, prostatic adenocarcinoma cells initiated from a bone metastasis, and WPMY-1 cells, a myofibroblast stromal cell line derived from healthy prostate) were pulse-labeled with AHA and SILAC amino acids for 24 hours in two biologically independent experiments

with reversed SILAC labels. Conditioned media of oppositely SILAC-labeled cells were combined (*i.e.* medium-heavy PC3 and heavy WPMY-1 and vice versa) and half of each sample was then used for enrichment of newly synthesized proteins and on-bead digestion (as in fig. 5.1). The remaining part was concentrated and directly subjected to trypsin digestion, thereby omitting the enrichment step. In the enriched sample, 684 human proteins were quantified, showing a very high reproducibility between biological replicates (correlation: 0.961) (fig. 5.2a). Without enrichment, only 22 human proteins were quantified (correlation: 0.025) (fig. 5.2b) amongst a multitude of bovine serum-derived peptides. This high background results in severely reduced signal-to-noise ratios in MS spectra of secretory peptides (fig. B.1), explaining poor detection and identification.

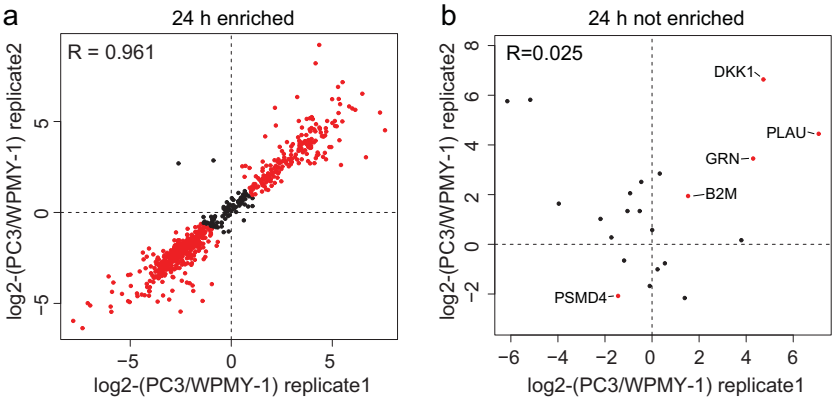

Figure 5.2: Comparison of secretomes between cell types. Proteins that are differentially secreted with statistical significance (FDR: 1%) are highlighted in red. a) Comparison of secretomes of PC3 and WPMY-1 cells in the presence of 10% serum. Cells were labeled with AHA and SILAC for 24h in biological duplicate, followed by media collection and enrichment of newly synthesized proteins; b) as in panel a, but without protein enrichment

The incorporation of AHA into t-RNA is slower than for methionine[122]. To exclude any influence of AHA-labeling on protein secretion, conditioned media from cells grown with AHA or methionine were quantitatively compared. Statistical analysis did not result in any significantly changing protein in either PC3 or WPMY-1 for treatments as long as 24 hours, even when raising the false discovery rate from 1% (as applied in the analyses described above) to 5% (fig. B.2).

These results demonstrate that AHA labeling and enrichment of newly synthesized proteins provide a powerful means to capture secreted proteins from growth media containing 10% serum. In addition, its combination with

pSILAC labeling allows for highly reproducible protein quantification. Protein ratios are highly accurate as assessed from samples that were combined in known relative amounts (fig. 5.3). Therefore, it can be concluded that the vast differences in secretome composition observed between PC3 and WPMY-1 cells (fig. 5.2a) accurately reflect differences in secretory activity.

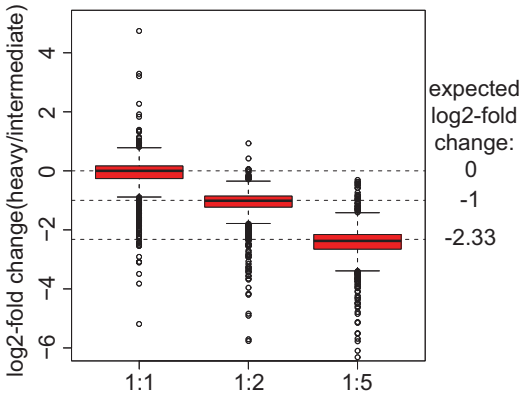

Figure 5.3: Accuracy of protein quantification after labeling and enrichment. After incorporation of AHA and either "medium-heavy" or "heavy" SILAC amino acids, differentially labeled conditioned media were mixed in known ratios. Proteins were quantified following enrichment and MS analysis as in figure 5.1. Enriched samples had a clearly defined distribution of peptide ratios centered at the dilution values.

5.1.1 Quantitative comparison of secretomes for network analysis

The dataset was then used to explore the nature and diversity of identified proteins and to gain a detailed insight how the differential secretome may explain the metastatic properties of PC3 cells. In total, 1136 newly synthesized proteins were identified in conditioned media of PC3 and WPMY-1 cells. Of these, 395 are truly secretory (fulfilling the criteria of being associated with UniProt[244] keywords: "Signal" or "Secreted", or reaching significance in SignalP 4.0[245]). Functional classification of all identified proteins using MetaCore[246] revealed 74 ligands, 65 receptors, 62 proteases and 211 other enzymes. Among the ligands are 11 cytokines and 12 growth factors, as well as 13 proteins that show cytokine and growth factor activity based on gene ontology annotation.

Based on the status of the SILAC label, 459 proteins were detected originating from PC3 cells alone, 325 (71%) of which were truly secretory. This

outperforms earlier studies identifying 303 (out of 771 proteins, or 39%)[247] and 77 (out of 139 proteins, or 55%)[248] secretory proteins from PC3 cells (fig. 5.4). Notably, those studies[247;248] used serum-starved cells and extensive peptide fractionation schemes and were only qualitative (*i.e.* they did not use SILAC).

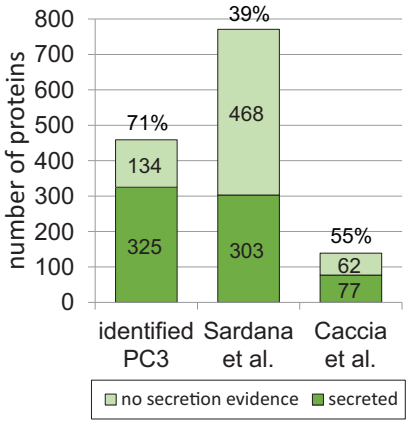

Figure 5.4: Number of identified proteins in PC3 cells compared to recent studies[247;248] (secretion evidence: signal peptide predicted using SignalP 4.0[245] or UniProt keyword "Signal" or UniProt keyword "Secreted"[244]) (Proteins identified based on at least 2 MS/MS spectra were considered from the presented dataset as well as the dataset published by Sardana *et al.*[247].)

As noted above, out of 1136 identified proteins, 684 could be quantified (fig. 5.2a). As many as 601 of these were differentially secreted between PC3 cells and WPMY-1 cells (FDR: 1%), 200 and 401 of which had elevated secretion levels in PC3 and WPMY-1 cells, respectively. Among the PC3-specific proteins, 63 are associated to "prostatic neoplasm" based on the curated GeneGo database[246], *e.g.*, *INHBA*, *AREG*, *DKK1* and *PSAP*. Functional annotation[230] of the proteins stronger secreted by PC3 cells revealed the terms "lysosome", "cell adhesion", "peptidase activity" and "cell migration", reflecting processes related to remodeling of the extracellular matrix. Figure 5.5 highlights some of these aspects in a network of directly interacting proteins (either physically or functionally). The figure shows the differential secretion of several metalloproteases (*MMP1*, *2*, *3* and *13*; *ADAM10* and *ADAMTS1*), as well as over 60 of their substrates. This suggests extensive processing (activation and inhibition) of signaling proteins and cell-surface proteins, *e.g.*, inhibitory cleavage of *CTGF* by *MMP13* or *LTBP1* by *MMP2* or activation of *AREG* and *THBS1* by *ADAMTS*. In addition, PC3 cells

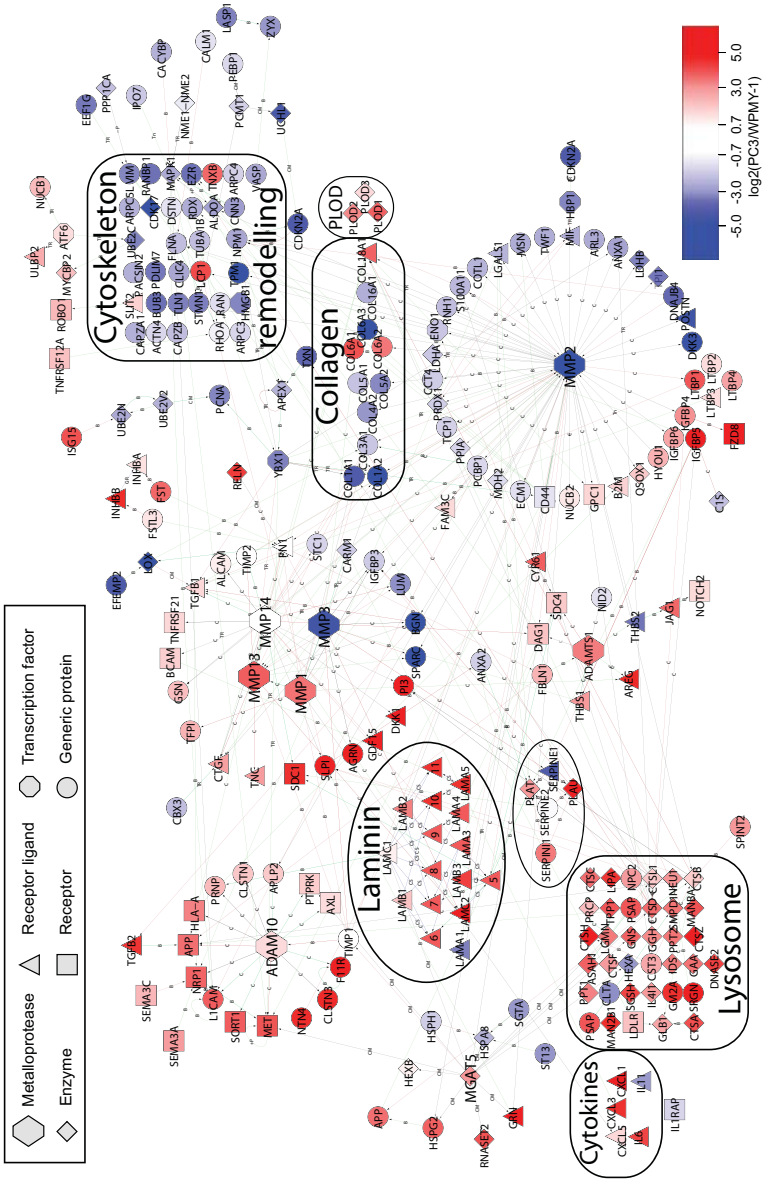

Figure 5.5: Interaction network of differentially secreted proteins from PC3 and WPMY-1 cells, highlighting clusters of functionally related proteins. Red and blue nodes indicate proteins preferentially secreted by PC3 and WPMY-1 cells, respectively (see scale bar for protein ratios). Edges indicate direct interactions along with interaction type: TR: transcriptional regulation, B: binding, C: cleavage, green lines: activation, red lines: inhibition, gray lines: unknown. The network was constructed using MetaCore[246] and layout was manually modified using Cytoscape for easier interpretation. Nodes not assigned to any functional term were moved next to their interaction partners. Proteins not assigned to any of the functional terms and not connected to any other protein in the network were excluded.

differ from WPMY-1 cells by specific patterns of extracellular matrix proteins, over-expressing laminins (that are thought to mediate the migration of cells into tissues[249]), while many collagens are underrepresented. The detection of ligand-receptor pairs, including *ROBO1-SLIT2*, *APP-TGFB2* and *SDC4-CYR61*, adds another level of functional connectivity covered by the presented data set. Five cytokines were quantified, including two members of the IL6-family (*IL6* and *IL11*) that are prototypical proteins at the lowest range of protein abundance[162] rarely detected in proteomic studies. Furthermore, several proteins conferring metastatic properties to cancer cells were successfully detected (*ADAM10*, *TNC*, *SDC1*, *L1CAM*, *ALCAM*), as well as other proteins that may be associated with cancer progression (*e.g.* *PLOD1*, *PLOD2*, *PLOD3*, *SDC4*). In the following paragraphs the secretion profiles of these proteins will be discussed in detail.

The metastatic properties of the cancer cell line are reflected by many differentially secreted proteins that are involved in the remodeling of the extracellular matrix (ECM). The major protease degrading ECM proteins is Plasmin. Plasmin is converted from Plasminogen by the two proteases, *PLAU* and *PLAT*, both of which are stronger secreted by PC3 cells (fig. 5.5). In contrast, the secretion of the Plasminogen activator inhibitor (*SERPINE1*) is reduced in PC3 cells compared to WPMY-1 cells. Although Plasmin was not identified in this experiment these secretion pattern suggest an activation of Plasmin leading to extensive tissue remodeling.

Plasmin activates several Matrix metalloproteases (MMPs) including the collagenases *MMP1* and *MMP13*, which show enhanced secretion in the cancer cell line. Metalloproteinases (ADAMs and MMPs) represent key proteins in ECM remodeling, which are able to degrade extracellular matrix proteins as well as bioactive molecules. The direct interaction network derived from MetaCore[246] (fig. 5.5) displays the differential secretion profiles of metalloproteases and their substrates detected in this study. Interestingly, ADAMs, as well as the majority of their target proteins, are stronger secreted by the PC3 cell line. Several ADAM proteins including *ADAM10*, identified in this study, are well-known to be involved in cancer formation and progression[250].

In contrast to the ADAM family proteins, the matrix metalloproteinases and their substrates show differential secretion patterns. Here, the collagenases (*MMP1* and *MMP13*) are strongly secreted by PC3 cells, while higher amounts of MMP2 and MMP3 are secreted by the WPMY-1 cells.

Collagens are the major compounds of the extracellular matrix and important targets of metallopoteinases and other proteases[251]. Most of the collagens quantified in this experiment are stronger secreted by the WPMY-1 cell line, including Collagen I, III, IV and V, which are fibril-forming collagens and Collagen XVI, that is assigned to the FACIT collagen family, the

members of which are associated to the fibril forming collagens. The collagens stronger secreted by PC3 cells (*COL6A1*, *COL6A2* and *COL18A1*) exhibit alternative functions: Collagen VI is a microfibrillar collagen, while Collagen XVIII belongs to the family of the multiplexins and can be proteolytically processed to Endostatin, a potent antiangiogenic protein. The specific functionalities and secretion profiles of Collagen VI and XVIII may suggest a function in cancer progression. The lower amount of fibril forming collagens present in the supernatant of PC3 cells could be explained either by an enhanced degradation by the two collagenases (*MMP1* and *MMP13*) or the extensive secretion of collagen by fibroblasts since the WPMY-1 cell line is derived from myofibroblast stromal.

Collagen inter-fibrillar cross-links are made by two groups of enzymes: Lysyl oxidases (LOX) and Procollagen-lysine,2-oxoglutarate 5-dioxygenases (PLOD). Crosslinks increase collagen stiffness which influences integrin signaling, resulting in enhanced focal adhesion formation and tumor progression[252;253]. Furthermore, it was shown that cells tend to move towards regions of increased stiffness[254]. Interestingly, all known PLODs were quantified in the study and are stronger secreted by the cancer cell line. This is in line with earlier studies, where increased expression levels of PLOD proteins were observed in gastric neoplasm[255;256]. Altogether, these results suggest that increased secretion of PLOD proteins from cancer cells contributes to cancer progression and regulation of cell migration by the enhancement of ECM stiffness.

Another group of extracellular matrix proteins highly represented in the dataset are laminins, a family of multidomain-containing heterotrimeric glycoproteins. Laminins promote cell adhesion and migration via integrins and other cell surface receptors[249]. Changes in expression levels and subunit composition of laminins have been described for several cancer types, especially for Laminin 5[257]. Furthermore the expression of many laminins is restricted to certain cell types[249]. The data reflect the same trends with several laminin subunits stronger secreted by PC3 cells. The only exception is Laminin subunit alpha-1 (*LAMA1*) which is stronger secreted be the WPMY-1 cells possibly representing a cell-specific laminin.

The ECM-tethered, matricellular proteins *TNC*, *TNXB* and *GPC* are stronger secreted by PC3 cells. Tenascin (*TNC*) is known for its antiadhesive properties. This could explain its strong secretion in the metastatic cell line in contrast to fibronectin, which is adhesive and does not show differential secretion. Furthermore, *TNC*, a ligand for laminins, is involved in the rebuilding processes in prostatic adenocarcinoma especially facilitating the process of invasion[258].

Several other proteins involved in cell adhesion show an increased secre-

tion in the cancer cell line, *e.g*, syndecan-1, syndecan-4, *L1CAM*, *ALCAM* and *BCAM*. Although it is known that syndecans enhance in cooperation with integrins adhesion, due to weak interactions with ECM components, their exact contribution to cell adhesion and migration is still unclear[259]. While Syndecan-1 expression has been linked to aggressive cancer progression[260], a contribution of Syndecan-4 to prostatic cancer has not been shown yet.

The enhanced secretion of CAM proteins provides a link to the metastatic properties of the PC3 cells, since Neural cell adhesion molecule L1 (*L1CAM*) as well as CD166 antigen (*ALCAM*) are up-regulated in the leading front and cell-cell junctions of invading tumor, respectively[261]. Basal cell adhesion molecule (*BCAM*), which has been associated to malignant transformation and tumor metastasis[262], is a receptor for Laminin alpha-5, which is often highly expressed in cancer cells[249].

In summary, the highlighted examples represent several successfully detected proteins contributing to metastatic properties in cancer cells, as well as new candidate proteins whose role in cancer progression remains to be established. Furthermore, the results indicate that the presented method permits in-depth differential secretome analysis against the background of 10% serum, identifying hundreds of secretory proteins down to the cytokine level and quantifying them over three orders of magnitude. The richness and quantitative nature of these data allows the construction of extensive networks that may be used to explain cellular behavior and that will be useful for biomarker discovery.

5.2 Serum starvation alters protein secretion

For reasons of convenience to minimize interference by background proteins, serum is often omitted in secretome analysis, based on the assumption that a short period of starvation (typically 24 h) only minimally affects cellular behavior. Recent reports have challenged this hypothesis, by showing that serum starvation rapidly (within hours) affects expression and phosphorylation levels of multiple proteins[16–19]. Having introduced a method to study protein secretion in the presence of serum, the effect of serum deprivation on secretome composition, a question that could not be addressed properly so far, was investigated. At the same time it was explored if the approach may be used to gauge immediate effects on secretome composition upon cellular perturbation. Therefore, five cell lines of different origin as well as primary hepatocytes (PHC) were pulse-labeled with AHA and SILAC using growth media with and without serum, followed by protein enrichment and mass

spectrometry (as in fig. 5.1).

name	cell type	time	quantified	secreted	R	significant
PC3	prostatic cancer	3 h	39	100%	0.87	5
PC3	prostatic cancer	24 h	428	64.7%	0.87	39
WPMY-1	prostate	3 h	95	96.8%	0.79	5
WPMY-1	prostate	24 h	757	26.8%	0.80	54
Hepa1-6	hepatocytes	4 h	173	84.5%	0.90	24
Hepa1-6	hepatocytes	24 h	1202	24.3%	0.82	130
Hepa1c1	hepatocytes	4 h	186	78.5%	0.74	26
Hepa1c1	hepatocytes	24 h	733	38.9%	0.88	105
PHC	primary hepatocytes	4 h	103	71.8%	0.69	8
PHC	primary hepatocytes	24 h	267	40.5%	0.54	14
RAW 264.7	macrophages + LPS	2 h	244	49.2%	0.65	42

Table 5.1: Cell types tested in serum starved medium: The treatment times, the number of quantified proteins, the percentage of truly secreted proteins (secreted), the correlation between biological replicates (R) and the number of differentially secreted proteins (significant) are given. (Significant proteins were defined by a FDR: 1% and a minimum fold-change of 2 for LPS stimulated mouse macrophages, PC3 and WPMY-1 cells, or a FDR: 5% and a minimum fold-change of 2 for all mouse hepatocytes.)

From 39 up to 1202 proteins were quantified with high reproducibility between biological replicates in the different cell lines (see table 5.1 and figures 5.6 - 5.8). As expected, prolonged collection times resulted in increased numbers of quantified truly secretory proteins, but was accompanied by a higher proportion of non-secretory proteins. All of these proteins carried a SILAC label and thus were synthesized during the AHA incubation step. Although it cannot be excluded that these proteins originate from cells dying during collection, they might also be externalized by unconventional secretion, *e.g.*, by exosomes. The observed difference in this "leakiness" between the investigated cell types can now be evaluated using the presented method, which would be more difficult without AHA and pSILAC labeling.

Serum deprivation significantly influenced the secretion pattern of up to 130 proteins. Even after short times of serum starvation (2-4 h) 5 to 42 proteins were differentially secreted by the tested cell types, not only demonstrating the ability of the method to capture immediate changes in secretome composition but most importantly highlighting the profound and rapid effect of serum starvation on protein secretion.

Although only starved for two hours the number of affected proteins in

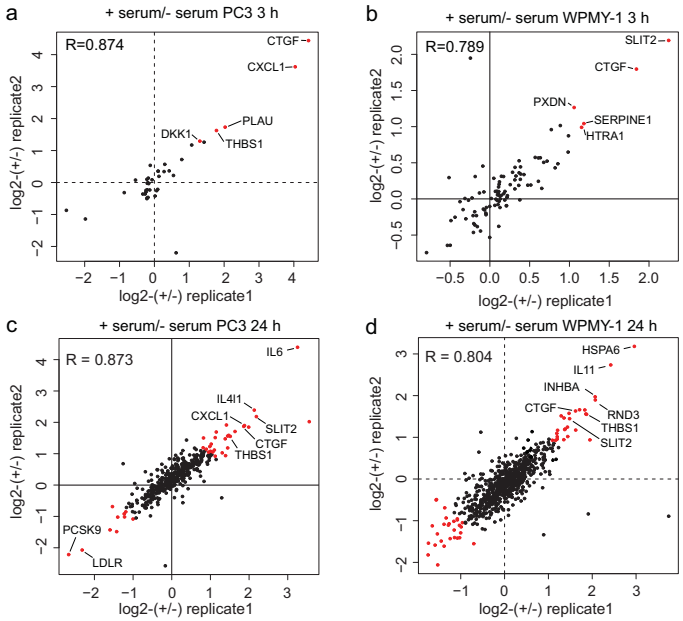

Figure 5.6: Comparison of secretomes of prostate cells grown with and without serum. Proteins differentially secreted with statistical significance (FDR: 1%) and have a fold-change higher than 2 are highlighted in red. a) Differential secretion of proteins from PC3 cells after 3 h serum starvation. b) Differential secretion of proteins from WPMY-1 cells after 3 h serum starvation. c) Differential secretion of proteins from PC3 cells after 24 h serum starvation. d) Differential secretion of proteins from WPMY-1 cells after 24 h serum starvation.

mouse macrophages is as high as for other cell lines after 24 hours. The viability of RAW 264.7 cells grown without serum was significantly reduced to 74.1%, compared to 84.0% of viable cells when grown in serum-containing medium (p-value = 0.014). Furthermore, several proteins induced by LPS stimulation of mouse macrophages are affected by serum starvation (fig. 5.8b), reflecting a reduced efficiency of macrophage activation in serum-free medium.

For primary hepatocytes the differences in secretion in response to serum starvation are less extreme in comparison to the stable cell lines. The overlap between the two collection times is, with only two proteins significant in both experiments, very low. Furthermore, only 7 of the 20 differential proteins are known secreted proteins. This points to the conclusion that primary hepatocytes are not as susceptible to serum reduction as stable cell lines.

Many differentially secreted proteins are involved in central cellular functions such as proliferation, cellular homeostasis, cholesterol homeostasis and signal transduction. In total 24 of the most strongly changing proteins were cytokines and growth factors (*e.g.* *CTGF*, *CCL20*, *CCL2*, *IL23a*, *CXCL1*,

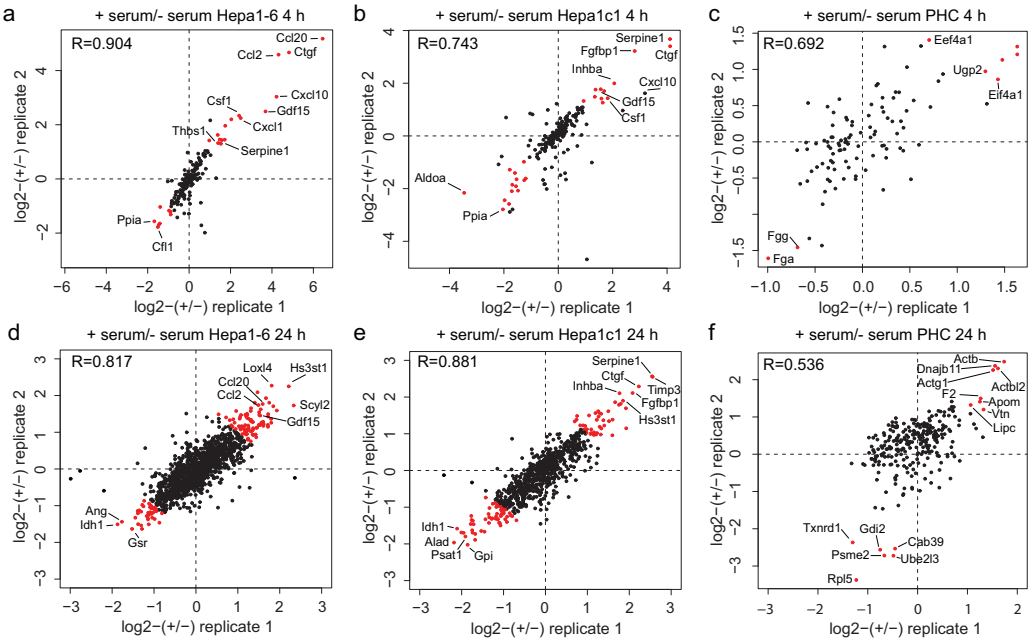

Figure 5.7: Comparison of secretomes of hepatic cells grown with and without serum. Proteins that are differentially secreted with statistical significance (FDR: 5%) and have a fold-change higher than 2 are highlighted in red. a) Differential secretion of proteins from Hepa1-6 cells after 4 h serum starvation. b) Differential secretion of proteins from Hepa1c1 cells after 4 h serum starvation. c) Differential secretion of proteins from primary hepatocytes after 4 h serum starvation. d) Differential secretion of proteins from Hepa1-6 cells after 24 h serum starvation. e) Differential secretion of proteins from Hepa1c1 cells after 24 h serum starvation. f) Differential secretion of proteins from primary hepatocytes after 24 h serum starvation.

CXCL10, *INHBA*, *KITLG*, *SPP1*, *TGFB1*, *VEGFA*, *IL1RN*, *TNF*, *IL11*, *IL6*, *DKK1*, *CXCL5*, *TGFB2*), most of which are rarely detected in proteomic studies. The fact that these cytokines can accurately be quantified, testifies to the power of the approach.

In addition, both cell type-dependent and independent responses to starvation were observed. The Connective tissue growth factor *CTGF* and the Plasminogen activator inhibitor 1 (*SERPINE1*) were stronger secreted by cells grown with serum in all 8 experiments using stable cell lines. Additional 18 proteins showed differential secretion in at least two experiments with cells originating from different tissues. Among those, *SERPINE1* and five additional proteins stronger secreted by cells grown in serum-containing medium regulate, based on GO annotation, cell proliferation (*ADAMTS1*,

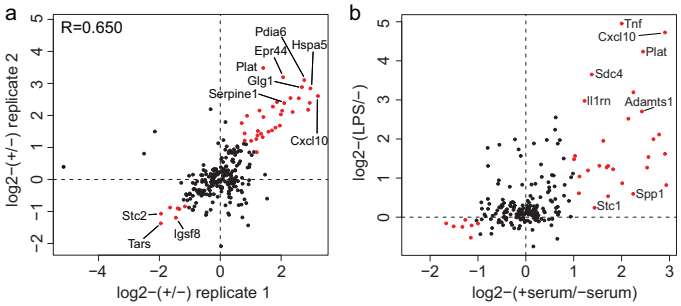

Figure 5.8: Comparison of secretomes of stimulated mouse macrophages grown with and without serum. Proteins that are differentially secreted with statistical significance (FDR: 1%) and have a fold-change higher than 2 are highlighted in red. a) Differential secretion of proteins from LPS treated RAW 264.7 cells after 2 h serum starvation. b) Comparison of serum starvation results to a comparison of LPS stimulated to unstimulated cells treated for 2 h

CXCL1, *CXCL10*, *FGFBP1*, *THBS1*). *CTGF* as well as the growth factor *CYR61*, which was stronger secreted in serum-containing medium of PC3 cells, are members of the CCN protein family. These proteins are described to be involved in multiple cellular functions including cell adhesion, angiogenesis, cell differentiation, cell proliferation, wound healing and tumorigenesis[263]. The observed secretion patterns of CNN proteins correspond with their described stimulative effect on cell proliferation[264]. Thus, the detected response to serum starvation of these proteins suggests their involvement in a common response mechanism controlling cell proliferation.

The increased secretion of *CSF1*, *GDF15* and *HS3ST1* and the decreased secretion of *PPIA* and *ALAD* in both hepatic cell lines grown with serum as well as the increased secretion of *HSPA6* and *SLIT2* in the prostatic cell lines are examples for tissue specific responses to serum starvation. In addition, several proteins were exclusively differentially secreted in one type of cells: *HSPA5*, *GLG1*, *ERP44*, *NUCB2*, *HSP90B1*, *SPP1*, *LY6E* and *LMAN2* are examples specific for mouse macrophages (fig. 5.8a), *CCL20*, *CCL2* and *SCYL2* were only differentially secreted by Hepa1-6 cells, while *TIMP3* was highly increased only in secretomes of Hepa1c1 cells (fig. 5.7). Specific for PC3 is the cytokine *IL6* which exhibits the most extreme response after 24 h (fig. 5.6c). Sakai et al[265] demonstrated recently the inhibition of cell growth through the modulation of apoptotic and signal transduction pathways after knocking down *IL6* in PC3 cells, demonstrating its functional relevance in this cell line. The extreme fold-change of *IL6* in the experiment suggests that *IL6* would not be detected in PC3 cells by an approach using serum free medium. Indeed, *IL6* as well as *IL4L1*, another protein exclusively dif-

ferentially secreted in PC3 cells, were not detected by Sardana et al.[247].

In WPMY-1 cells another cytokine *IL11* was six times stronger secreted by cells grown with serum. *IL11*, like its close relative *IL6*, activates STAT3 signaling by binding to the shared receptor *IL6ST* promoting cell survival, proliferation and immune response[266]. Thereby, the increased secretion of *IL6* and *IL11* in the two cell lines represents the same response to serum starvation promoted by cell-specific cytokines.

Exclusively stronger secreted by PC3 cells under serum-free conditions are LDL receptor (*LDLR*) and Proprotein convertase subtilisin/kexin 9 (*PCSK9*). Both proteins are involved in controlling cholesterol homeostasis. *PCSK9* binds and degrades *LDLR*[267]. This provides a potential link between serum factors and regulation of local LDL availability, which could help clarifying the role of cholesterol as a risk factor for prostate cancer[268].

In conclusion, the data indicate that serum starvation drastically and immediately (2-4 h) affects secretion, well within the time-frame of most secretome studies (24 h). Proteins identified under such conditions may represent artifacts that do not reflect the cellular state under physiological conditions.

5.3 Quantification of secreted proteins from primary mouse hepatocytes compared to cell lines

Next, it was tested if the approach can be used to access secretion differences between primary cells and cell lines. It was applied to mouse hepatic cells, which are widely used in drug metabolism studies. Since it was demonstrated before that the intracellular proteome of the hepatoma cell line Hepa1-6 differs strongly from that of primary hepatocytes[269], secretory activity of primary cells compared to two hepatoma cell lines (Hepa1-6 and Hepa1c1) was assessed. Figure 5.9a-c shows that secretion profiles are different to such a degree that the cell lines should not be considered as being representative of the primary cells. After four hours of in total 58 and 62 quantified proteins 52 and 36 showed significant secretion differences in the comparisons of primary hepatocytes to Hepa1c1 and Hepa1-6 cells, respectively. Strikingly, although the correlation between biological replicates is very high (0.852 and 0.893 fig. 5.9a,b), the correlation is even higher (0.937) when plotting the protein ratios of both cell lines against each other (fig. 5.9c), indicating that the secretomes of the two cell lines are highly similar. Extending the collection time to 24 hours increased the number of quantified proteins to 612 containing

as many as 581 differentially secreted proteins (fig. 5.9d).

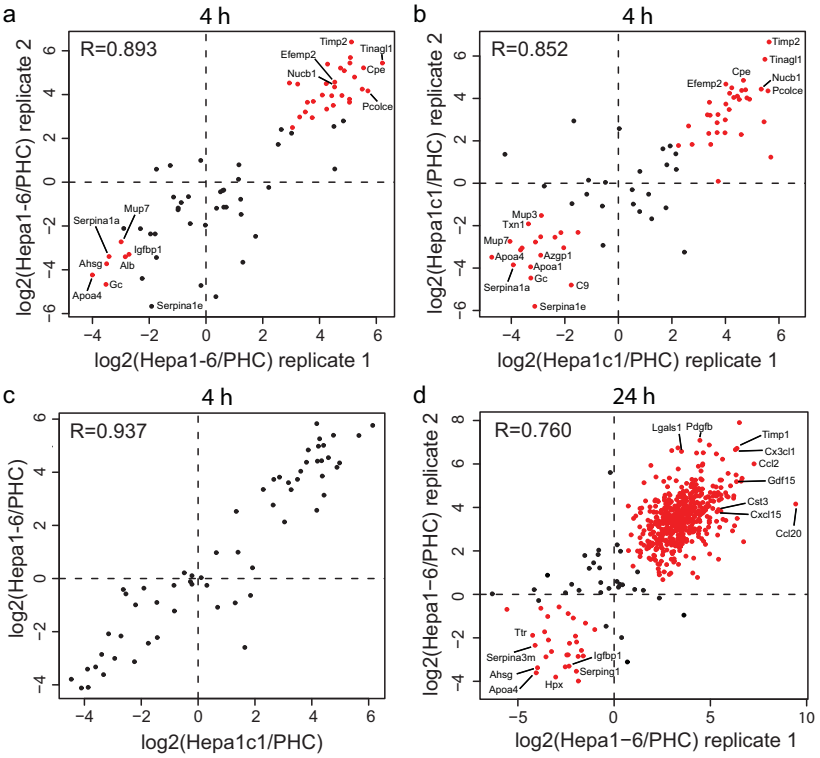

Figure 5.9: Comparison of the secretomes of primary hepatocytes (PHC) and hepatoma cell lines (Hepa1-6 and Hepa1c1). Proteins that are differentially secreted with statistical significance (FDR: 5%) and have a minimum fold-change of 2 are highlighted in red. a) Comparison of Hepa1-6 and PHC secretomes, collected for 4 h in biological replicate. b) Comparison of Hepa1c1 and PHC secretomes, collected for 4 h in biological replicate. c) Combination of panels a and b, showing very good correlation between secretomes of the two cell lines. d) Comparison of Hepa1-6 and PHC secretomes, collected for 24 h in biological replicate.

7 and 30 proteins were stronger secreted by primary hepatocytes compared to Hepa1-6 cells after 4 or 24 hours, respectively. The 18 proteins stronger secreted by primary hepatocytes compared to the Hepa1c1 cell line highly overlap with the ones significant in the comparisons to Hepa1-6 cells. Together, all of these are known to be exclusively or highly expressed in liver, *e.g.*, *Apoa1*, *Apoa4*, *Apob*, *Ahsg*, *C9*, *Gc*, *Itih4* and nine members of the Serpin family (*e.g. Serpina1a*, *a1e*, *a1b*, *a1d*, *a3k*, *a3m*, *a3n*, *f2* and *g1*). Among them are high-abundant serum proteins (*e.g.* Transthyretin, Serum albumin) which were distinguished from their bovine homologs in serum by the presence of SILAC labels. The relative protein abundances in the enriched

primary hepatocyte samples were estimated using the iBAQ approach[21], revealing several high abundant proteins stronger secreted by the Hepa1-6 cells than by primary hepatocytes, *e.g.*, *B2m*, *Serpine1*, *Cst3*, *Cpe*, *Ppia*, *Fn1*, *Igfbp4*, *Tinagl1* (fig. 5.10). This indicates that some hepatic properties are preserved in the cell line and that the isolation and culture conditions for the primary cells did not globally reduce secretion.

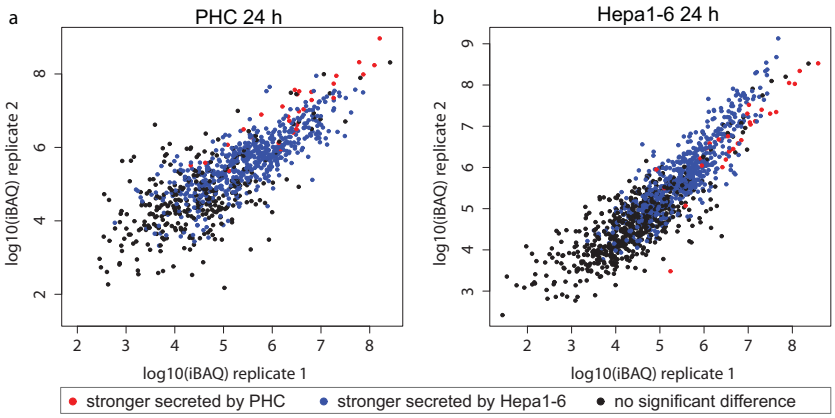

Figure 5.10: log10(iBAQ) values for primary hepatocytes (a) and the Hepa1-6 cell line (b) after 24 h highlighting the relative abundances of the quantified proteins after enrichment

Slany *et al.* recently reported the detection of 67 secreted proteins from human primary hepatocytes using a serum depleted medium[270]. Of these, 27 mouse orthologues were identified with additional 83 proteins only captured via the presented approach.

After 24 hours as many as 29 of 38 quantified proteins involved in immune response, including 11 cytokines, are stronger secreted by the Hepa1-6 cells, possibly representing a hepatic function. In addition, functional annotation enrichment revealed several terms, like "proteinaceous extracellular matrix", "cell adhesion", "lysosome" and "cell migration", represented by proteins like *Timp2*, *Hspg2*, *Dag1*, *Sparc*, *Bcam*, *Lama5*, *Lamb2*, *Mmp13*, *Ctsa* and *Ctsb* also found, *e.g.*, in PC3 cells, that may reflect the cells' transformed status.

In summary, it was shown that the approach is equally applicable to primary cells and cell lines, at the same time demonstrating that secretome profiles of Hepa1-6 and Hepa1c1 cell lines are highly similar to each other, but very different from primary cells.

5.4 Secretion dynamics in activated macrophages

Intracellular interaction in response to external events is established by controlled secretion of proteins[271]. Although several studies have compared secretomes in the presence and absence of stimulation[272–274], traditional tools (with or without SILAC) have not been able to probe the dynamics of such a response with good temporal resolution. Therefore, the approach combining pSILAC and AHA was used for time-resolved secretome analysis to gain detailed insight into the timing and magnitude of protein secretion. Specifically, the experiment was designed to capture proteins that are newly synthesized in defined windows after the onset of cellular stimulation, in order to distinguish immediate from delayed secretory events (fig. 5.11a). As a model system mouse macrophages (RAW 264.7) stimulated with lipopolysaccharides (LPS) were used, thereby inducing the secretion of proteins involved in innate immune response. Proteins were labeled with AHA and SILAC amino acids only in 2-hour time windows at different times (0, 6, or 17 hours) after the addition of LPS, upon which conditioned media was collected (fig. 5.11a).

Collectively, across these time points, as many as 500 proteins were identified and quantified, including 12 cytokines, 33 receptors, 25 proteases and 99 other enzymes. This highlights that very short collection times (2 hours) suffice to monitor large numbers of relevant signaling proteins. The abundance of proteins after enrichment, estimated from an iBAQ analysis[21], spans four orders of magnitude, with *Tnf*, *Ccl4* and *Ccl9* among the high abundant proteins and *Ccl5*, *Il6ra* and *H2-T23* in the low abundance range (fig. 5.11b). The measured protein ratios correlate very well with previously published RNA expression data[219] of LPS-stimulated macrophages during the same time windows (correlation: 0.66, 0.80, 0.83) (fig. 5.12a-c). These results indicate the relevance of the protein secretion profiles.

Compared to untreated controls, 97 proteins were differentially secreted upon LPS treatment in at least one of the sampled time points, of which >90% are bona fide secretory proteins (fig. 5.13). This set includes established cytokines in LPS response, such as *Tnf*, *Il6*, *Ccl2*, *Ccl4*, *Ccl5*, *Ccl9*, *Cxcl10*, *Il23a*, and is highly enriched for gene-products downstream of NF-κB (pValue: 4.7 e-73) and AP1 (pValue: 3,5 e-84), two transcription factors known to act downstream of LPS and TLR4 signaling (fig. 5.14). In addition, 58 differentially secreted proteins can be directly connected to immune response (Supplementary Table 3). Collectively, this confirms the specificity of the experiment in capturing the expected set of proteins and, to the best of my knowledge, constitutes the first comprehensive secretome analysis of

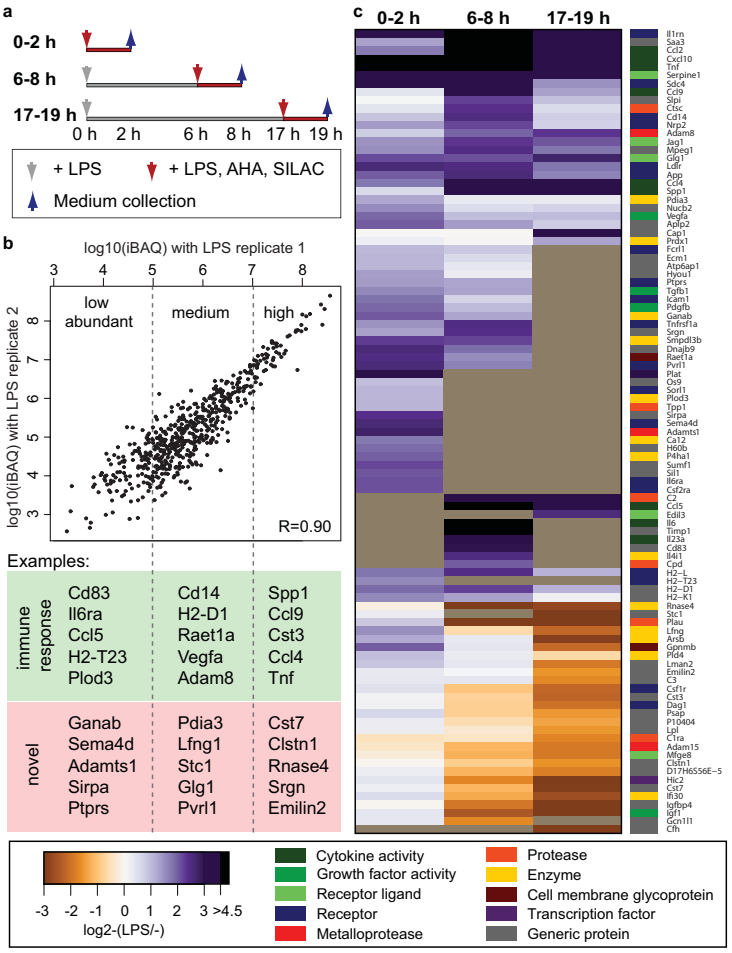

Figure 5.11: Kinetics of protein secretion from LPS-stimulated mouse macrophages. a) Experimental design: Mouse macrophages were stimulated with LPS for 2, 8 and 19 hours. For each time point, 2 hours before the collection of conditioned media the medium was substituted to a medium containing AHA, SILAC amino acids and LPS. This allows the capture and identification of proteins specifically synthesized and secreted in the indicated 2 hour time windows. b) Scatterplot of log10(iBAQ) values for LPS treated macrophages (2 h) used to estimate the relative abundance of the detected proteins. For each abundance range examples for differentially secreted proteins that are either known to be involved in immune response (top) or have not been assigned to immune function (bottom) are provided. c) Heatmap of log2-fold-changes for proteins showing differential secretion with statistical significance in at least one experimental time point. (0-2 h: FDR: 5%, 6-8 h: FDR: 1%, 17-19 h: FDR: 10%). The bar next to the heatmap indicates the functionality of the detected proteins.

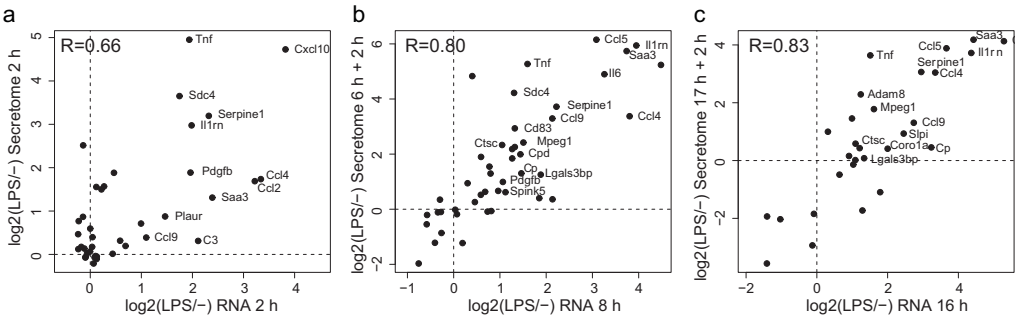

Figure 5.12: Comparison to RNA expression data: The measured protein ratios of secreted proteins after LPS stimulation correlate very well with previously published RNA expression data for proteins involved in immune response[219], therefore representing positive controls for the measured secretion kinetics.

LPS-stimulated macrophages to date.

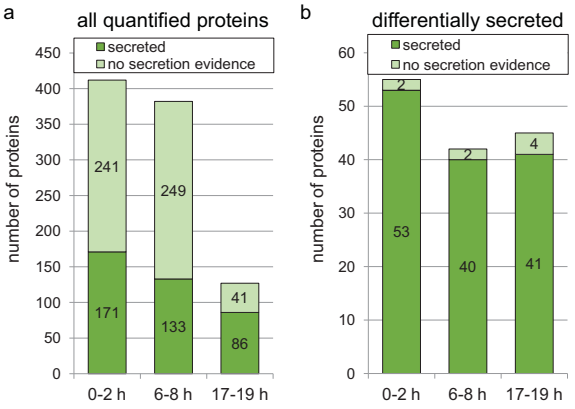

Figure 5.13: Number of truly secreted proteins of LPS stimulated mouse macrophages for all quantified proteins (a) and differentially secreted proteins (b)

In the next step the time-resolved changes of the 97 differentially secreted proteins upon LPS-stimulation were analyzed (fig. 5.11c). Within two hours all significant proteins are induced after stimulation. This includes several cytokines, whose secretion was strongly maintained at all time points (*Tnf*, *Cxcl10*, *Ccl2* and *Ccl4*) with a peak between six and eight hours. Other cytokines (*Ccl9*, *Ccl5*) and receptors (*Ldlr*, *App*, *Nrp2*, *Cd14* = LPS co-receptor) respond to LPS treatment with a delay, being strongly increased at 6-8 hours before dropping off slightly at 17-19 hours.

Some proteins are exclusively induced at the first time point. This includes the growth factors *Tgfb1*, *Vegfa* and *Pdgfb* that are known to be both

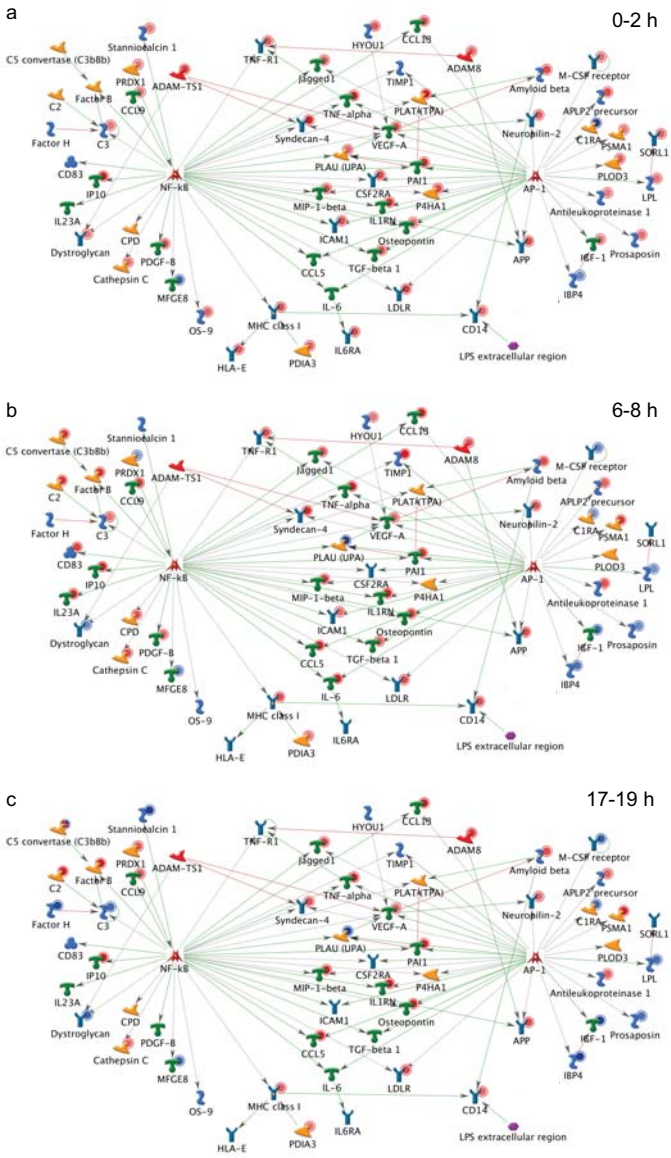

Figure 5.14: Interaction network for differentially secreted proteins after LPS stimulation of mouse macrophages, centered around NF-κB and AP1. Only those proteins are displayed that were differentially quantified and that are known to be expressed downstream of the transcription factors NF-κB and AP1 or directly interacting with other proteins of the network. The large coverage of the network indicates broad activation of the TLR4 - NF-κB - AP1 axis, known to be induced upon LPS stimulation. Red and blue circles indicate increased and decreased secretion, respectively, in each of the time windows. NF-κB and AP1 were added to the network, although, being localized in the nucleus, they were not detected in the study.

secreted by and act on macrophages[275–277]. Their secretion profile thus seems to indicate that they participate primarily in immediate macrophage activation. Interestingly, the Il6-receptor also shares this pattern, while its ligand Interleukin 6 (*Il6*) was only observed after six hours. This is in very good agreement with observations that *Il6*, although part of the initial immune response, follows Tumor necrosis factor (*Tnf*) secretion at ∼6 hours of exposure to LPS[278]. In the data secretion of *Il6* coincides with the appearance of *Il23a*, *Ccl5*, *C2*, *Edil3* and *Cd83*, possibly indicating similar routes of secretion.

Interestingly, the presented data uncover 28 proteins whose secretion is repressed over time compared to the untreated control. This observation raises the exciting possibility that inhibited secretion of factors that are involved in signaling and ECM remodeling (*Lfng*, *Plau*, *Cst3*, *Dag1*, *Adam15*, *Ifi30*) may be part of the LPS response. Interestingly, Macrophage colony-stimulating factor 1 receptor (*Csf1r*), the receptor for M-CSF that plays an important role in promoting the release of pro-inflammatory cytokines, shares this profile. Therefore, its progressive decrease in secretion suggests a feedback mechanism moderating the inflammatory response. In fact, several of the identified factors have both pro- and anti-inflammatory characteristics, including *Il6*[279]. The interplay of secreted factors (cytokines, receptors, proteases) combined with their individual secretion profiles imposes an intricate mechanism for context-dependent tuning of the immune response.

Another group of proteins strongly involved in the regulation of inflammatory responses are proteases and their inhibitors. In total, ten proteases and eight endopeptidase inhibitors were differentially secreted, displaying various secretion patterns over time. The two proteases Urokinase-type plasminogen activator (*Plau*) and Tissue-type plasminogen activator (*Plat*) are examples for the time-resolved regulation of protease activity in response to LPS treatment. Their inhibitor Plasminogen activator inhibitor 1 (*Serpine1*) displays increased secretion over the complete time frame (2-19 hours), probably causing the down-regulation of *Plau* at the two later time points. *Plat* was stronger secreted in response to LPS between zero and two hours but was not detected afterwards.

Other examples are the two endopeptidase inhibitors Cystatin-C (*Cst3*) and Cystatin-F (*Cst7*) that show decreased secretion at prolonged LPS stimulation. While *Cst3* is known to be involved in defense response[280], *Cst7* is not assigned to any GO term related to inflammatory response, although it was shown that *Cst7* inhibits Dipeptidyl peptidase 1 (*Ctsc*) in several immune cells[281]. The decreased secretion of *Cst7* may prevent the inhibition of *Ctsc*, which is stronger secreted itself between six and eight hours after LPS stimulation.

Due to their proteolytic properties, proteases of the ADAM family are important regulators of inflammation[282;283]. Of the three regulated ADAM proteins, *Adam8* is known to decrease inflammatory response by the cleavage of TNF receptor[284]. The observed increase in secretion of *Adam8* from 0 to 19 hours could indicate a repression of TNF-mediated defense response after prolonged LPS stimulation.

Among the proteins whose secretion was reduced at 17 hours, three proteins were in fact induced at two hours (*Gpnmb*, *Lfng*, *Arsb*). While Transmembrane glycoprotein NMB (*Gpnmb*) is known to reduce inflammatory responses by modulating *Il6* levels and NO production[285], Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe (*Lfng*) may be connected to the LPS response indirectly. *Lfng* is a membrane-bound enzyme that modifies the Notch1 receptor disrupting *Jag1*-binding, which activates Notch signaling[286].

Interestingly, Protein jagged-1 (*Jag1*) secretion is elevated in response to LPS over the full 19 hours (fig. 5.11b). Recently, Notch-dependent upregulation of Notch target genes was shown to be delayed after LPS stimulation[287]. Based on the observations this delay could be explained by the disruption of *Jag1* binding at early time points promoted by *Lfng*, while the repression of *Lfng* could permit *Notch1* activation by *Jag1* at later times. No previous evidence was found in the literature for *Arsb* as well as *Edil3*, *Sil1*, *Sumf1*, *Glg1*, *Mpeg1* being involved in LPS response.

Altogether, these selected examples illustrate the effective application of the described approach to the investigation of cellular responses over time. The capacity of the method to detect high numbers of secreted proteins from very short collection windows makes it extremely powerful to study kinetics of protein secretion.

5.5 Compendium of secreted proteins

Next, by assembling all data acquired in this study, the coverage of the human secretome was investigated. In the human proteome, 3831 proteins were defined as being secretory (see methods for definition). In the combined experimental data, 665 of these (17%) were identified which represents a large proportion, especially when considering that only six cell types were investigated. These proteins span a wide variety of functionalities, of which hydrolases and signaling molecules make up the biggest part by absolute numbers, followed by receptors, extracellular matrix proteins, cell adhesion molecules and proteases (fig. 5.15a).

Since the presented method relies on the presence of AHA (and thus me-

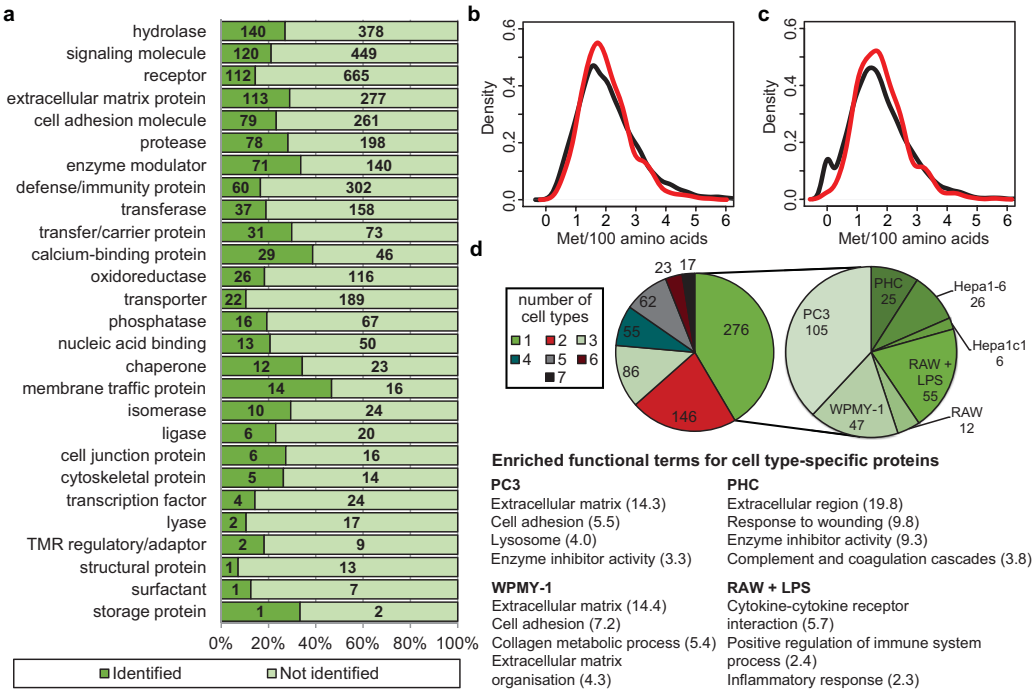

Figure 5.15: Compendium of secreted proteins identified across all experiments in this study. a) Protein functions of secreted proteins based on PANTHER classification [288;289]. b) Distribution of methionine frequencies in the theoretical secretome (3831 proteins, black line) and experimentally detected proteins (665 proteins, red line) indicates that the method is not biased against proteins with low-methionine counts. c) Distribution of methionine frequencies after removal of the signal sequence in the theoretical secretome (black line) and experimentally detected proteins (red line). The peak at 0 methionines per 100 amino acids represents the 199 proteins not containing a methionine after signal peptide or N-terminal methionine cleavage d) Classification of commonly and uniquely identified proteins across all cell lines used. Almost half of all secreted proteins (276) were uniquely identified in one cell type and are associated with GO-terms reflecting the cells' biological functionality. The numbers in brackets indicate the enrichment score derived by functional annotation clustering using DAVID [230].

thionine) in target proteins, the dataset of 665 secretory proteins was used to evaluate a potential bias against proteins sparse in methionines. The methionine-frequency distribution was shifted slightly to higher values compared to the theoretical human secretome (fig. B.3c,d). However, this can be fully explained by a similar shift in protein size distribution, which is in line with the well-known underrepresentation of small proteins in mass spectrometric analyses[242] (fig. B.3a,b). Normalization of methionine-frequency for protein length results in overlapping density curves of experimental and theoretical data (fig. 5.15b,c) demonstrating that the method is not biased against proteins with low numbers of methionines. The peak at 0 methionines per 100 amino acids (fig. 5.15c) represents the 199 (5%) proteins that do not contain a methionine after cleavage of the signal sequence or after removal of the starting methionine and that therefore are in principle not detectable by the presented method.

Finally, the frequency of protein identification across the experiments was determined in order to assess the size and composition of cell-specific secretomes. (fig. 5.15d). While some proteins were identified in all seven cell types and treatments, 42% of the 665 secreted proteins were only identified in one cell type (fig. 5.15d). Interestingly, GO-analysis of these cell-specific secretomes revealed terms with clear connections to the cellular functionality, like "response to wounding" and "complement and coagulation cascades" for primary hepatocytes, or "inflammatory response" for activated mouse macrophages. This demonstrates that the charted secretomes provide unique signatures of the cells that were investigated and reinforces the idea that the secretome is a careful reflection of the cell's identity.

In summary, the pulse-labeling of proteins with azide-containing amino acids facilitates the efficient isolation of newly-synthesized proteins from complex background via click-chemistry. These proteins can then be quantified by virtue of the concurrent labeling with isotope-coded amino acids. When used in combination, it was demonstrated that these tools can be used effectively for the quantitative and in-depth comparison of secretome composition and to probe for proteome changes induced within a very short time window in stable cell lines as well as primary cells. Many hundreds of proteins of diverse functionalities that can be used to explain differences in cellular phenotypes have been uncovered. A particularly powerful and unique application of the method is in studying the kinetics of secretory activity. The ability to derive temporal profiles for each of the secreted proteins has added novel insights to the heavily studied activation of mouse macrophages.

Chapter 6

Proteome and Transcriptome Kinetics in Activated Mouse Macrophages

Transcription and translation are processes tightly controlled at different steps during gene expression. To understand how a biological system responds to environmental variation at a molecular level, time-dependent changes of both processes need to be investigated. Therefore, the methods described in the two previous chapters were used in combination with RNA expression data to gain insight into the kinetics of gene and protein expression during mouse macrophage activation. Mouse macrophages (RAW 264.7 cells) were stimulated with lipopolysaccharides (LPS), thus activating signaling pathways downstream of the Toll-like receptor 4 (TLR4) that are involved in the modulation of the host inflammatory response. This system was investigated recently in several qualitative or quantitative proteomic studies exploring different aspects of TLR4 signaling[222;224–227]. All studies focus either on the early or the late response to LPS stimulation or just detect changes in mRNA expression in a time course after stimulation[218;219;223]. By measuring proteins newly synthesized during LPS stimulation at different time points, the sequence of events leading to macrophage activation can be reconstructed. Macrophages as first actors in the innate immune response use the secretion of specific proteins (cytokines) to attract and activate other immune cells. To include this important set of proteins into the study, the secretomes of activated macrophages were analyzed.

6.1 LPS stimulation of mouse macrophages

6.1.1 Experimental design

To create a comprehensive picture of macrophage activation, changes in protein synthesis were measured and compared to total protein abundance and RNA abundance changes at three time points after LPS stimulation (fig. 6.1). The time points were selected to cover immediate responses to stimulation (one hour) as well as delayed ones (two hours and three hours). Thereby, events establishing the activated status of macrophages should be revealed.

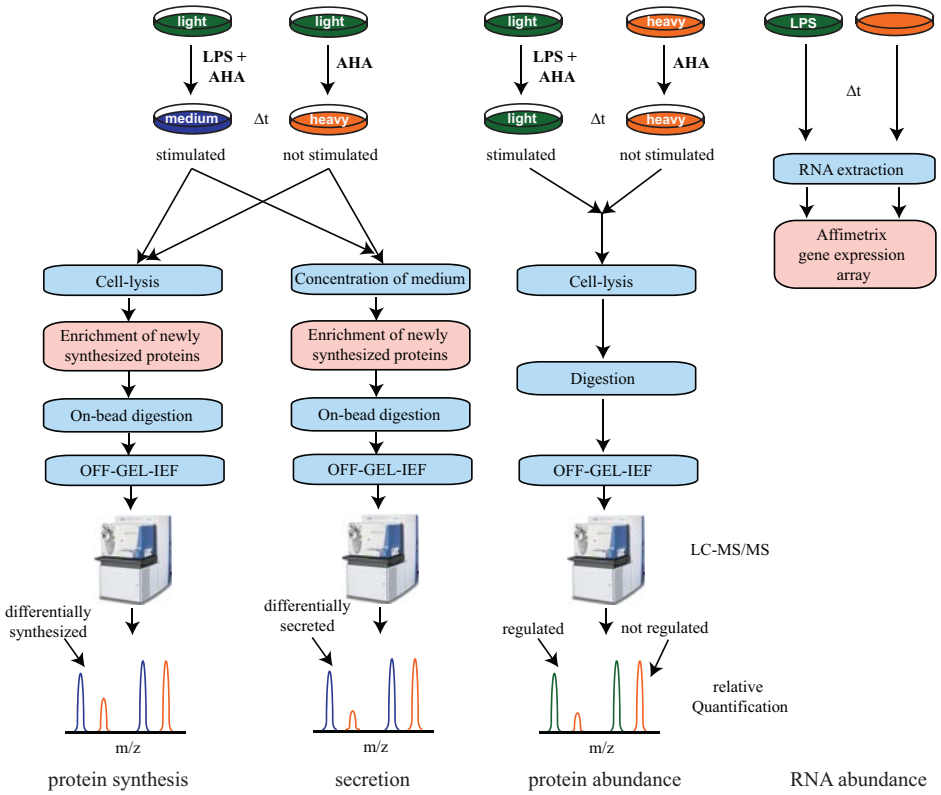

Figure 6.1: Workflow for the investigation of proteome and transcriptome kinetics in mouse macrophages.

The measurement of protein synthesis was performed as described in chapter 4.2 using AHA incorporation and alkyne-activated agarose beads for enrichment in combination with pSILAC labeling. In parallel, the medium collected from the same cells was used for secretome analysis as described in chapter 5.1. Full SILAC labeling was applied to assess changes in total protein abundance. RNA abundance changes were measured by affimetrix

gene expression arrays.

6.1.2 Results - an overview

The numbers of quantified and significantly changing proteins are summarized in table 6.1. Notably, already after 20 min of AHA incorporation as many as 4852 newly synthesized proteins and 32 newly synthesized and secreted proteins were quantified, again demonstrating the rapid incorporation of AHA as well as the efficiency of the enrichment approach.

	protein abundance	protein synthesis	secretome	RNA abundance
20 min	4042 (0)	4852 (0)	32 (0)	
60 min	3508 (2)	4914 (35)	65 (24)	7972 (79)
120 min	3355 (17)	5389 (189)	325 (41)	7972 (380)
180 min	3357 (7)	5268 (246)	182 (95)	7972 (424)
480 min		5297 (366)		

Table 6.1: Number of quantified proteins in different data sets at investigated times after LPS stimulation. The number of proteins significantly changing in response to LPS treatment are given in brackets in red (FDR 1% and fold-change >2).

In total, 9131 proteins and 11660 RNAs were detected, with an overlap of 7972. To ensure that just expressed RNAs are considered, only these 7972 RNA expression values were used in the following analysis.

Different amounts of newly synthesized proteins in the LPS treated sample compared to the untreated control were detected for 35, 189 and 246 proteins after one, two or three hours, respectively (table 6.1). On RNA level approximately two times more significant changes were found. Here, 79, 380 and 424 RNAs were differentially expressed at the same times after stimulation (table 6.1). An increasing number of significantly changing proteins over time has been detected on secretome level, too, ranging from 24 proteins after one hour to 41 and 95 proteins after two and three hours of LPS stimulation. The increased numbers of significant proteins over time represent delayed effects of LPS stimulation, as well as continued RNA or protein synthesis, increasing abundance and therefore the chance to detect differences.

The proportion of changing secreted proteins compared to the total numbers of quantified secreted proteins is much higher than for the other data

sets. This points to a strong impact of LPS on macrophage secretion.

Significant changes in protein abundance were detected for 2, 17 and 7 proteins after one, two or three hours of LPS stimulation, respectively. These smaller numbers of significant proteins in comparison to the significant newly synthesized proteins can be explained by three facts. First, since less proteins were quantified, low abundant proteins were most probably missed. Second, due to enrichment of newly synthesized proteins minimal changes are detected that were missed when comparing total protein abundances. And third, a parallel increase or decrease in protein synthesis and degradation could lead to constant total protein levels while the levels of newly synthesized proteins vary.

Combining all data sets for 782 proteins significant differences in at least one of the performed experiments were detected. Since no information for total protein or RNA abundance is available, the data set measuring the protein synthesis after eight hours of stimulation was not included.

To assess the contribution of the protein synthesis data sets to the detection of significant proteins, the proportions of significant proteins compared to all quantified proteins were calculated (fig. 6.2). Therefore, significant as well as quantified proteins were combined over all investigated time points. While 8.6% of the detected proteins showed a significant change in response to LPS stimulation, this number was slightly lower when exclusively looking at the changes of intracellular protein synthesis (5.0%). As many as 31.5% of the quantified secreted proteins were detected at different levels (fig. 6.2a). This number increases further, when the secretome analysis at prolonged times after stimulation (presented in chapter 5.4) are included (35.0%) or only proteins with secretion evidence are considered (33.1%). When combining both criteria, more than half of the truly secreted proteins are differentially secreted (fig. 6.2b).

The participation of several detected significantly changing proteins in immunity and response to LPS stimulation is well described. This includes the members of the NF-κB and AP1 transcription factor complexes (*Rel*, *Nfkb1*, *Nfkb2*, *Relb*, *Junb*, *Fosl1*, *Fosl2*, *Jun*, *Fos*, *Jund*, *Fosb*), the LPS receptor CD14 and several cytokines induced in response to LPS stimulation (*Tnf*, *Il6*, *Ccl2*, *Ccl4*, *Ccl5*, *Ccl9*, *Cxcl10*).

Next, it was evaluated how many proteins with established functions in immunity were detected by the method presented in this thesis. Therefore, the numbers of significant proteins involved in immune response were compared to the total numbers of detected proteins participating in this process. Significant changes were observed for 19.1%, 35.1% and 38.3% of all detected proteins involved in immune response, regulated downstream of LPS stimulation or transcriptionally regulated downstream of LPS stimulation,

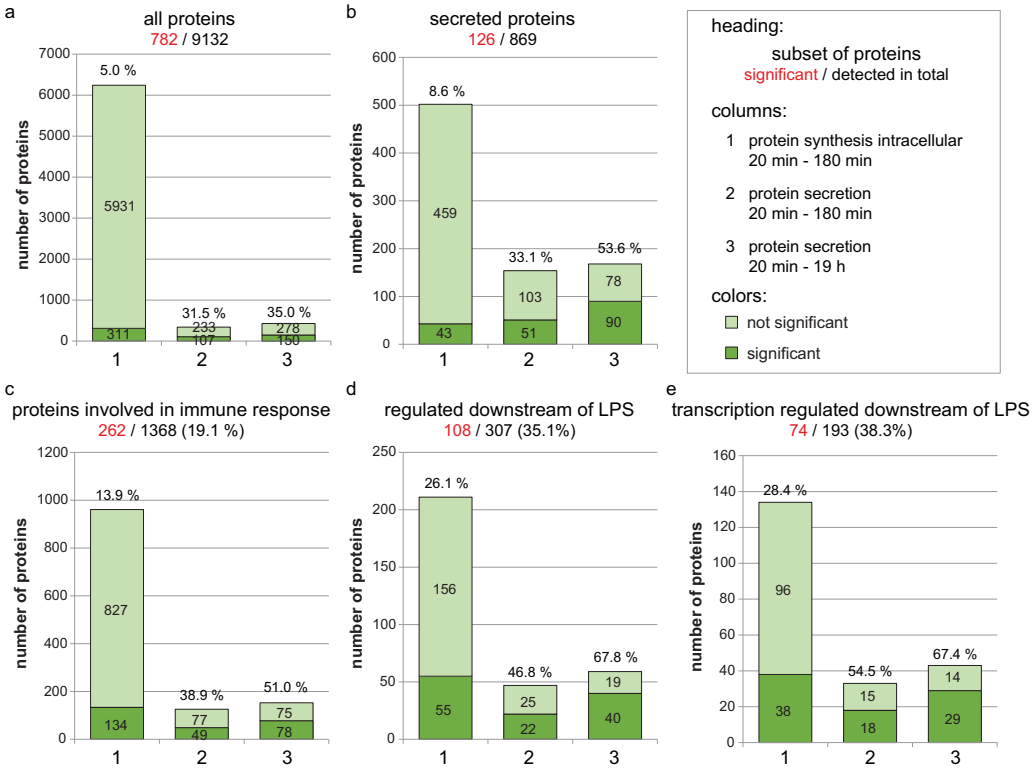

Figure 6.2: Contribution of protein synthesis data sets to the detection of expected proteins: For the different approaches all time points are combined and the percentage of significant proteins compared to all quantified proteins is given for a) all detected proteins; b) all proteins with secretion evidence; c) all proteins with a connection to immune response; d) all proteins regulated downstream of LPS based on MetaCore annotation[246] (total number in MetaCore: 587); e) all proteins with transcription regulated downstream of LPS based on MetaCore annotation[246] (total number in MetaCore: 373). For the third column the datasets presented in chapter 5.4 are included to demonstrate that the investigation of prolonged LPS stimulation times increases the percentage of expected proteins.

respectively (fig. 6.2c-e). Again, the secretome data sets contribute to a higher extend to these numbers.

In addition, functional classification of the significant proteins was performed to gain an overview of biological functionalities of LPS responding proteins. Within most protein classes around 7 to 10% of the proteins detected in the complete data set were regulated in response to LPS treatment (fig. 6.3). Exceptions are receptor ligands (36%), receptors (19%) and transcription factors (14%), representing important protein classes regulating inflammatory responses.

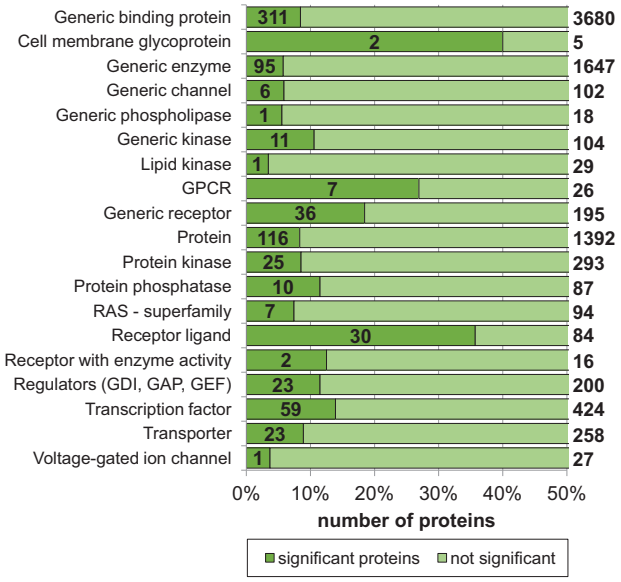

Figure 6.3: Distribution of functionalities of the detected and the significant proteins. Functional assignment was performed using MetaCore[246]. The total number of detected proteins assigned to each class is given left of the bars.

The attempt to determine a temporal sequence of events induced or reduced in response to LPS stimulation based on functional annotation enrichment using the significant proteins at different time points was not successful, since the retrieved functional terms were highly similar. Hence, functional classification was performed for groups of proteins with similar synthesis or abundance kinetics, which will be discussed in section 6.2.1.

In summary, 782 proteins are affected by LPS stimulation of mouse macrophages for up to three hours. These proteins are enriched in functional classes important for immune response. Considering the limited time-frame investigated and the restriction to one cellular system, a high percentage of proteins expected to be influenced by LPS stimulation were detected. A much higher proportion of detected proteins in the secretome data sets are

regulated, resulting in a high relative contribution of the secretome data to the detection of known and novel proteins.

6.2 Comparison of the kinetics of RNA abundance, protein abundance, protein synthesis and secretion

Previous studies comparing mRNA abundance to protein abundance report a wide range of correlation values highly dependent on the applied methods and the model system under investigation[21;25;290–292]. Hence, experimental design as well as the model system influence the degree of agreement between these types of data. It could be proposed, that certain systems are stronger influenced by post-transcriptional regulation than others, resulting in lower correlation values.

The correlations between changes in RNA abundance and newly synthesized proteins are within the reported range (around 0.4-0.6) (fig. 6.4). Notably, similar correlations are observed when comparing, *e.g.*, protein synthesis to total protein abundance changes or protein synthesis to protein secretion. In spite of similar correlation values, the determined rations of a subset of proteins do not correlate when comparing changes in secretion to intracellular protein synthesis or RNA abundance (fig. 6.4). These proteins are only regulated in the secretome. These are possibly differentially synthesized proteins that are immediately externalized in response to LPS stimulation. Alternatively, LPS treatment could induce gene expression independent secretion routes or extensive transmembrane protein cleavage.

In addition, correlation values increase when comparing RNA abundance changes to changes in protein synthesis at a later time point (time difference approximately 1-2 hours) (fig. B.4). *E.g.*, one hour after stimulation a correlation of 0.47 was determined while the correlation between RNA expression at one hour and protein synthesis at two or three hours was 0.57 or 0.61, respectively. For increased time shifts, like seven hours, this effect disappears.

In other words, the comparison of the measured changes in protein and RNA species reveals expected correlation values. Bad correlations of single proteins can be partly explained by a delay between RNA expression and protein synthesis. Furthermore, a subset of proteins show a significant change in secretion, which cannot be explained on the RNA level and is therefore probably gene expression independent.

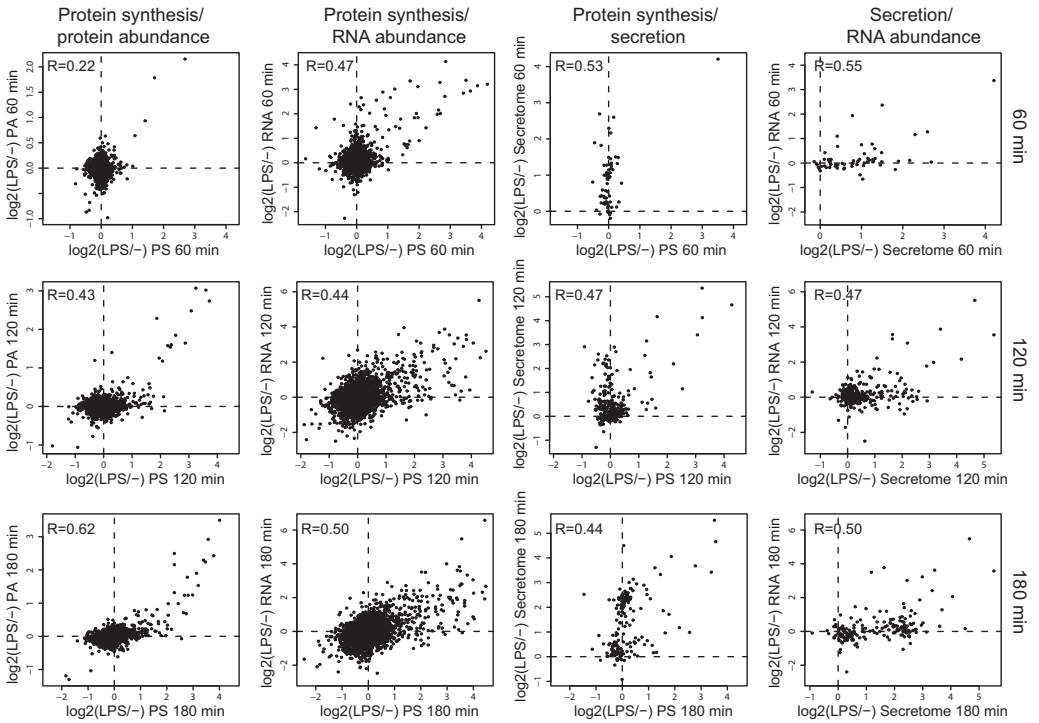

Figure 6.4: Comparison of log2-fold-changes determined for the different protein and RNA species. (PS: protein synthesis; PA: protein abundance)

6.2.1 Comparison of protein synthesis to RNA abundance dynamics

A more specific comparison of changes in protein synthesis and RNA abundance could help to distinguish between transcriptional and translational regulation. Hence, all significant proteins were grouped into categories based on the comparison of kinetics of RNA abundance and protein synthesis changes (fig. 6.5). 24 proteins are quickly induced in response to LPS stimulation (category A). Among them, three proteins (*Tnf*, *Nfkbiz* and *Ier3*), although not significant, show a slight increase in protein synthesis already 20 min after LPS stimulation, thereby representing the first line of action in defense response. Notably, while Radiation-inducible immediate-early gene IEX-1 (*Ier3*) is thought to be a negative feedback inhibitor of NF-κB[293], NF-kappa-B inhibitor zeta (*Nfkbiz*) has been described to inhibit but also promote NF-κB dependent transcription[294;295].

As predicted using the correlation plots (fig. B.4) a delay in protein synthesis compared to RNA abundance changes can be observed for 40 proteins (category B in figure 6.5). Low translation rates or inhibited translation

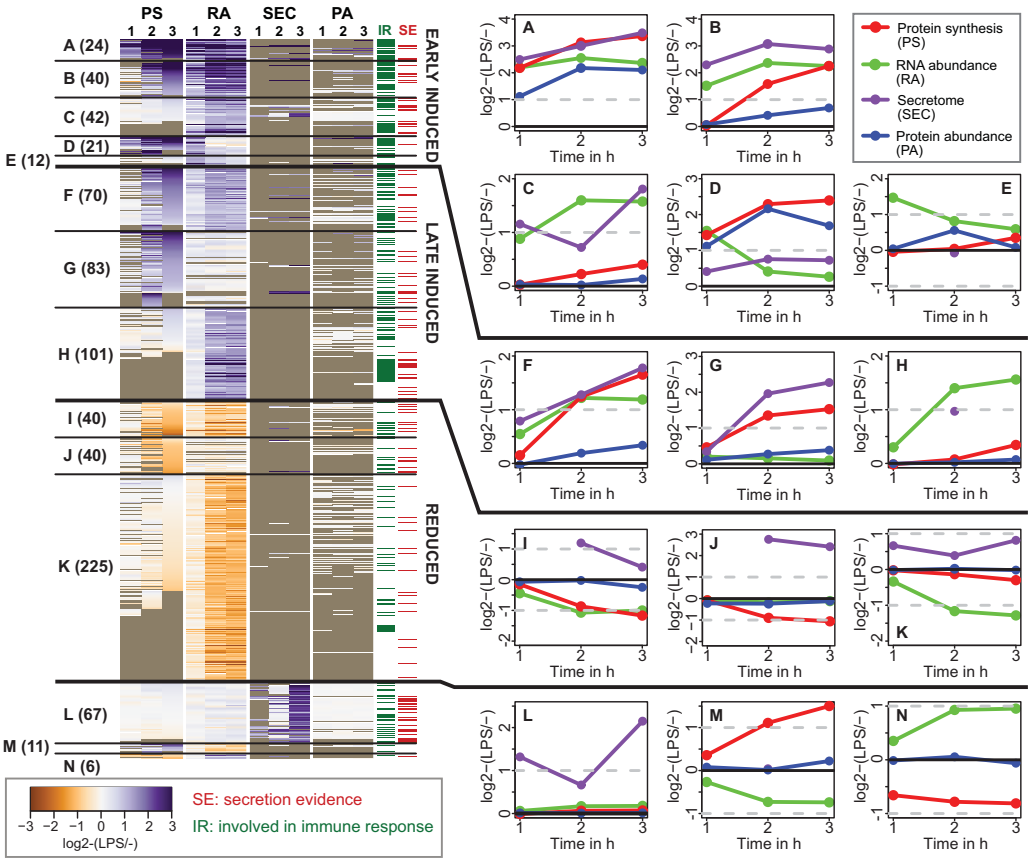

Figure 6.5: Categorization of all significant proteins based on the comparison of protein synthesis against RNA abundance kinetics. The heatmap displays all significant proteins. The number of proteins in each category is given in brackets. The categories are further grouped based on the determined kinetics. Lineplots illustrate the average log2-fold-changes in each category. (The numbers on top of the columns in the heatmap give the time after LPS stimulation.)

could cause this delay in protein synthesis.

Proteins in five categories exhibit changes in RNA abundance without a detectable effect on protein synthesis, or change in protein synthesis without a significant difference in RNA abundance (categories C, G, H, J, K in figure 6.5). This discrepancy can be explained by proteins not detected in one of the data sets. Other proteins respond in a delayed fashion, exhibiting the regulation pattern after prolonged time intervals. This was evaluated for 35 proteins being induced only after eight hours of LPS stimulation (*e.g. Nfkbib*, *Cd44*, *Cebpd*, *Itpkb*, *Ccrl2*). For 15 proteins in these categories RNA abundance changes are reflected in the secretome, meaning that the major amount of induced newly synthesized protein is immediately externalized

(category C, H). Synthesis of the remaining proteins could be repressed or a fast degradation of the newly synthesized protein could prevent induction on the protein level. For several proteins induced on protein level, but not significantly changing in RNA abundance, a slight increase or decrease on RNA level was detected (category G, J in fig. 6.5). Therefore, it could be proposed that even minor changes in RNA abundance result in a significant difference in protein synthesis for these proteins.

The 14 categories defined in figure 6.5 were further assigned to groups that show increased protein synthesis or RNA abundance already after one hour of LPS stimulation, termed "early induced" , or at later time points (two or three hours after stimulation), termed "late induced". A third group ("reduced") contains all categories with reduced protein synthesis or RNA abundance after stimulation. Of the three categories not fitting into this scheme, two exhibit opposing kinetics and in one significant changes were only detected in the secretome.

The majority of proteins in the "induced" categories can be connected to functions in immune response and LPS signaling based on the applied criteria (*i.e.* regulation downstream of LPS, AP1 or NF-κB described in the MetaCore database[246] or any GO term containing "immune", "defense response" or "inflammatory response"). Among the strongest induced proteins several cytokines (*Tnf*, *Ccl2*, *Ccl4*, *Csf3*, *Cxcl10*), transcription factors (*Atf3*, *Fosl1*, *Junb*), the NF-κB inhibitors *Nfkbid* and *Nfkbiz* as well as Tristetraprolin (*Zfp36*), which mediates *Tnf* RNA decay[296], can be found. In addition many proteins not fitting the described criteria have been associated to established LPS response programs. Examples are the Dual specificity phosphatases (*Dusp2*, *Dusp4*, *Dusp5*), which will be discussed in chapter 6.3.4 and Proto-oncogene serine/threonine-protein kinase pim-1, that has been described to regulate proliferation and apoptosis and to be induced by cytokines[297].

In addition to these established inflammatory proteins, several induced proteins have not been associated to immune response yet and are therefore novel candiatates with possible functions downstream of LPS induction. Examples are *Rgl1*, *Lamc1*, *Flnb*, *Cpeb4*, *Dusp16*, *Hmga2*, *Homer1* and *Vps37c* induced with delayed kinetics after LPS treatment (fig. 6.5 category F) as well as 17 proteins that exhibit a delay in protein synthesis compared to RNA abundance changes (category B in figure 6.5), such as *Slc4a7*, *Arl5b*, *Csrnp1*, *Ell2*, *Tgm2*, *Hexim1*, *Phlda1* and *Ehd1*.

Of note, a minor proportion of the proteins exhibiting "reduced" kinetics fits to the mentioned criteria for involvement in inflammatory responses (fig. 6.5).

Functional annotation clustering of "early induced" and "late induced" groups reveals expected terms for inflammation and response to LPS stimu-

lation, like “inflammatory response”, “toll-like receptor signaling pathway”, “regulation of cytokine production”, “apoptosis” and “MAPK signaling pathway” (table 6.2). Although the received functional terms are similar, subtle distinctions can be recognized. While several “early induced” proteins are involved in processes necessary for a rapid induction of immune response (*e.g.* “regulation of cytokine production”, “MAPK signaling pathway” and “regulation of phosphorylation”), most of the “late induced” proteins regulate long term immunity, represented by the terms “immune system development”, “cell activation” and “adaptive immune response” (table 6.2).

EARLY INDUCED	
6.12	cell death
4.09	inflammatory response
3.21	cytokine
2.82	positive regulation of cytokine production
2.73	TLR signaling pathway/ regulation of cytokine production
2.70	MAPK signaling pathway
2.41	positive regulation of developmental process
2.32	regulation of apoptosis
2.10	negative regulation of molecular function/ regulation of phosphorylation
LATE INDUCED	
5.16	immune system development/ cell activation
2.85	defense response
2.74	response to organic substance
2.29	cell activation/ adaptive immune response
2.04	plasma membrane part
2.00	Chemokine signaling pathway/ TLR signaling pathway
REDUCED	
4.34	chromosome/ chromosome organization
3.44	cell cycle
2.77	nucleosome
1.98	nucleoplasm/ transcription factor complex
1.98	nucleus/ transcription regulation
1.92	GTPase regulator activity

Table 6.2: Functional annotation clustering for proteins induced early or late or being reduced after LPS stimulation. Numbers in the first column are enrichment scores provided by the functional annotation clustering algorithm in DAVID [230]. In the second column functional terms representative for all terms in the cluster are provided.

20 significant proteins were mapped to the Toll-like receptor (TLR) signaling pathway, the main pathway activated by LPS stimulation of mouse macrophages (fig. B.5), resulting after Benjamini correction in a p-value of $4.5\text{x}10^{-6}$. The majority of these proteins are proteins transcribed down-

stream of TLR signaling. Only a few proteins involved in the signal transduction of this pathway were differentially expressed or synthesized. This can be explained by the fact that most of the signal transduction is performed via phosphorylation dynamics, which is not assessed by the applied methods. Alongside with downstream effector genes several transcription factors involved in TLR signaling are differentially synthesized or expressed, including several members of the NF-κB, AP1 and IRF families.

Functional annotation clustering for all proteins with reduced abundance or synthesis results in very different enriched terms (table 6.2). Those are rather associated to "chromosome organization" and "transcription regulation", as well as "cell cycle", pointing to a repression of specific transcriptional programs and the reduction of proliferation, which has been reported earlier[298;299]. The reduction of several proteins involved in cell cycle progression (*e.g. Cdk2*, *Cdc25a*, *Cdc25b*, *Orc5*) in combination of the induction with some of its inhibitors (*e.g. Mdm2*, *Myc*, *Ccnd1*) are visualized in the KEGG pathway map (17 mapped genes, corrected p-value: $3.1\text{x}10^{-3}$)(fig. B.6).

Notably, 225 proteins only exhibit reduced kinetics on RNA level, while no changes on protein level have been detected (category K in figure 6.5). Among them several proteins are associated to the chromosome (functional annotation clustering score: 4.2), including most of the detected histones (*Prim1*, *Dsn1*, *H1f0*, *Chek1*, *Rfc4*, *Xpo1*, *Hist1h1a*, *Hist1h4a*, *Hist2h2ab*, *Hist2h2aa1*, *Hist2h3b*, *Ncapd3*, *Hist1h2bf*). In addition, 7 proteins exhibit "GTPase regulator activity" (score: 2.0)(*Fgd4*, *Rasa3*, *Racgap1*, *Wdr67*, *Ect2*, *Hmha1*, *Vav3*) and 9 proteins are associated to the functional term "cell cycle" (score: 1.8)(*Chek2*, *Dsn1*, *Racgap1*, *Rassf2*, *Anln*, *Chek1*, *Kif11*, *Ncapd3*, *Rbl1*).

To provide a more detailed view of macrophage activation, some categories with interesting synthesis and abundance kinetics will be discussed in the next paragraphs.

Two "early induced" categories exhibit interesting RNA abundance dynamics, being induced after one hour of stimulation but showing balanced RNA levels after prolonged stimulation (category D and E in figure 6.5, fig. 6.6). While for one category the protein synthesis is induced (fig. 6.6a), no changes in protein synthesis have been detected for the second category (fig. 6.6b).

The pattern of the first of the two categories can be explained by an induced RNA and protein synthesis after LPS stimulation, followed by RNA degradation at prolonged times, while the protein remains stable. Seven of the 20 proteins in this group (fig. 6.6a) are transcription factors, including *Btg2*, *Jun*, *Fos* and all detected Early growth response proteins (*Egr1*, *Egr2* and *Egr3*). A possible interpretation for the detected kinetics are nega-

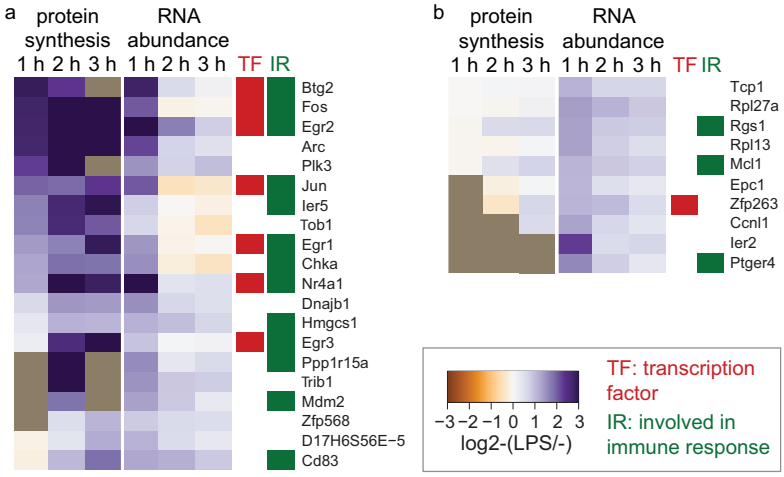

Figure 6.6: Protein categories with early induced RNA and balanced RNA levels after prolonged stimulation: a) Heatmap of log2-fold-changes of proteins assigned to the category D in figure 6.5 b) Heatmap of log2-fold-changes of proteins assigned to the category E in figure 6.5

tive feedback loops, which are used to maintain low levels of transcription factors in the cell[300] and to increase the response time of transcriptional networks[301]. Negative feedback loops can include a downstream target of the transcription factor, but for several transcription factors, including *Fos*[302], a negative autoregulation was demonstrated[300].

An alternative mechanism could be the induction of RNA degradation by another induced protein. One example for this is the stimulation of RNA degradation by Tristetraprolin (*Zfp36*), which is strongly induced in response to LPS (category A in figure 6.5). RNA degradation by Tristetraprolin has been demonstrated for a wide range of proteins including Serine/threonine-protein kinase PLK3 (*Plk3*)[303] and Immediate early response gene 5 protein (*Ier5*)[304], both showing the described abundance and synthesis dynamics. The maintained difference in protein synthesis for these proteins again highlights the recent finding that many proteins involved in immune response have instable RNA but stable proteins[21].

The overall good agreement of RNA abundance and protein synthesis data is illustrated by only 17 proteins with opposing kinetics (category M and N in figure 6.5). An induction of RNA expression in combination with a reduction in protein synthesis in response to LPS stimulation was observed for six proteins (*Nfkbia*, *Cdc25a*, *Irf2bpl*, *Zcchc2*, *Glg1*, *Ptprs*)(fig. 6.5 category N).

NF-κB inhibitor alpha (*Nfkbia*) inhibits dimeric NF-κB/REL complexes

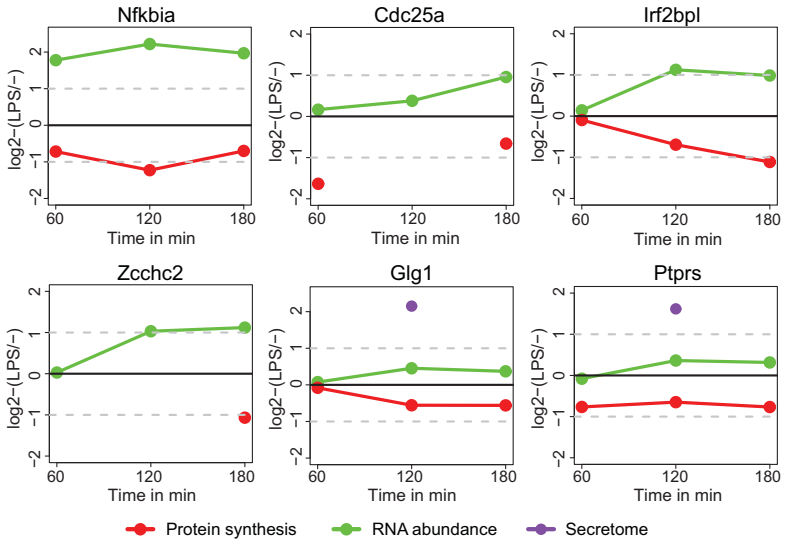

Figure 6.7: Proteins with opposing kinetics: Although RNA expression is increased in response to LPS treatment a decrease in protein synthesis was measured (category N in figure 6.5).

by trapping REL dimers in cytoplasm[305]. Activation of TLR signaling induces phosphorylation of *Nfkbia*, followed by ubiquitinylation and its degradation. Therefore, NF-κB is released and can translocate to the nucleus to activate its transcriptional program[305]. The data suggest a constant synthesis but increased degradation of *Nfkbia* in response to LPS stimulation to preserve NF-κB signaling. This ensures the maintenance of high *Nfkbia* RNA levels for fast suppression of NF-κB signaling if needed. Therefore, induced RNA and/or protein synthesis in combination with induced protein degradation possibly represents one mechanism for rapid immune response suppression to avoid negative consequences for the system, like septic shock.

Similarly, ubiquitin mediated proteolysis of M-phase inducer phosphatase 1 (*Cdc25a*) in response to DNA damage, resulting in delayed cell-cycle progression, has been described[306]. The observed kinetics point to a similar mechanism of induced protein degradation, while preserving the ability to quickly produce *Cdc25a* if necessary. This observation again suggests a connection between inflammation and reduced proliferation.

The similarity of expression and synthesis profiles indicates a possibly similar mechanism for the two hardly described proteins in this category (*Zcchc2*, *Irf2bpl*).

The remaining two proteins *Glg1* and *Ptprs* are both transmembrane proteins and show induced levels in the secretome after two hours. Since this induction is seen on the RNA level, too, the reduction of protein synthesis

could be explained by an increased cleavage and release to the extracellular space, thereby reducing the intracellular protein levels.

The category with inverse profiles, exhibiting induced protein synthesis and slightly reduced RNA abundance in response to LPS treatment, is represented by eleven proteins with diverse functions, ranging from apoptosis and stress response to transcriptional regulation (*e.g. Atf4*, *Chac1*, *Gch1*, *Hpgds*, *Ndrg1*, *Sesn2*)(category M in figure 6.5). Those profiles could be explained by a rapid RNA degradation.

In conclusion, the combined measurement of changes in RNA abundance and protein synthesis over a time frame after perturbation of the system helps to group proteins based on the mechanism regulating gene expression. Hence, it provides starting points for the explanation of observed phenotypes and gives first hints how to further investigate protein expression regulation of individual proteins.

6.2.2 Comparison of protein synthesis to total protein abundance dynamics

The ability to investigate newly synthesized proteins separately from the immense background of "old" proteins should on one hand increase the sensitivity for the detection of changes and on the other hand reveal processes not measurable on total protein level, like the interplay of protein synthesis and degradation.

97 proteins are regulated either in protein synthesis or in protein abundance and detected in both data sets. 26 of those exhibit the same regulation pattern, including mostly proteins with well known functions in immune response, such as *Nfkbiz*, *Fos*, *Jun*, *Nlrp3*, *Spp1*, *Slfn2*, *Zpf36* and *Irg1*.

Even though a difference in protein synthesis in response to LPS stimulation was detected, 69 proteins do not significantly change in total protein abundance. Twelve of these proteins, although not significant, show the same trend, being slightly induced or reduced in total abundance (*e.g. Gpr84*, *Oasl1*, *Rel*, *Tnfsf9*, *Dnajb4*, *Marcksl1*, *Hmox1*, *Elf2*). Additional, 42 proteins change in RNA abundance and protein synthesis after LPS stimulation but not in the overall protein abundance level. Several of them are proteins with established functions in inflammation (*e.g. Tlr7*, *Tlr13*, *Hmga2*, *Pdlim5*, *Zyx*, *Dnaja1*, *Hspa5*, *Icam1*, *Nfkb2*, *Nfkb1*). Most probably, the detection of the regulation was achieved for these proteins by newly synthesized protein enrichment causing increased sensitivity for measuring differences in the proteome after stimulation. Finally, twelve proteins are induced and three proteins are reduced in protein synthesis, while no differences on RNA or

protein abundance was detected (fig. 6.8).

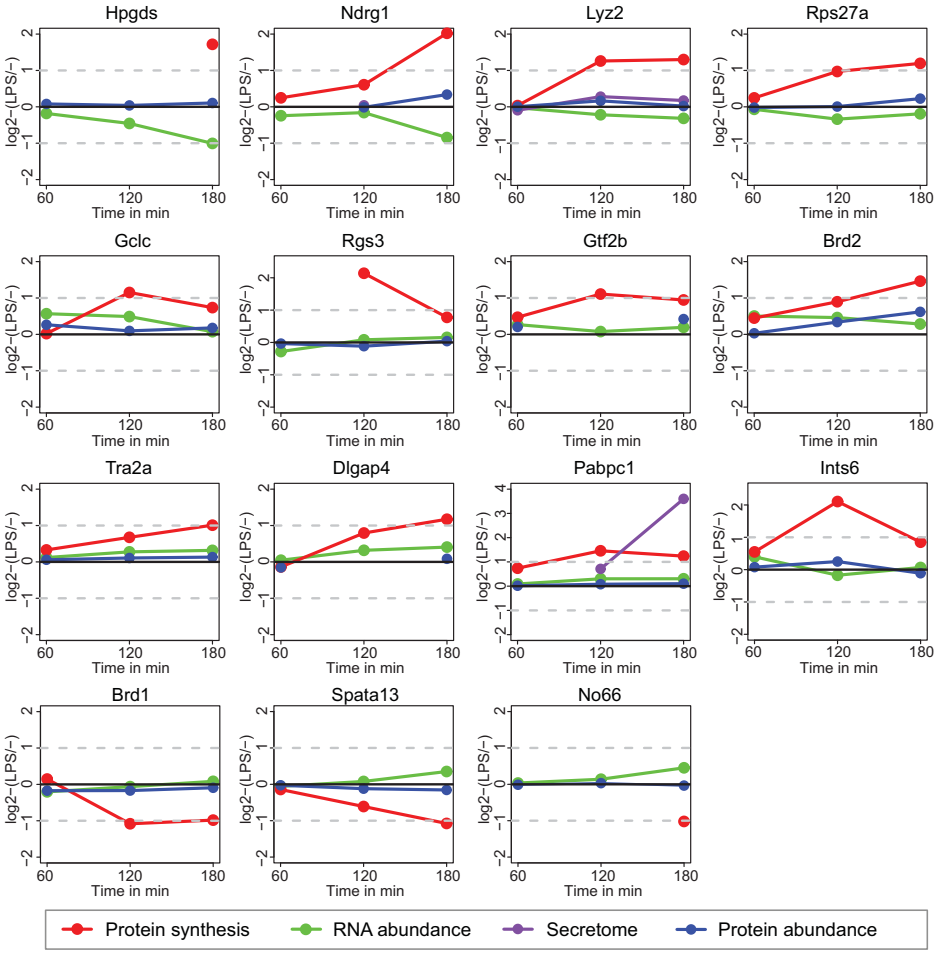

Figure 6.8: Proteins with increased or decreased protein turnover: Although RNA and protein abundance remain constant after LPS treatment, an increase or decrease in protein synthesis was measured.

This profile could be explained by an increased or reduced protein turnover in response to LPS stimulation. An interesting case is Ubiquitin-40S ribosomal protein S27a (*Rps27a*), since it is the precursor of two functionally distinct proteins, ubiquitin and 40S ribosomal protein S27a, produced by proteolytic cleavage[307]. Ubiquitinylation plays a central roll in the regulation of NF-κB signaling, not only by targeting proteins for proteasomal degradation, but also as a signaling tag[308]. Of note, only peptides matching the 40S ribosomal protein S27a chain have been detected (fig. 6.9). Hence, an increased turnover of *Rps27a* could therefore provide ubiquitin, necessary for the regulation of NF-κB signaling, while the second strain is immediately

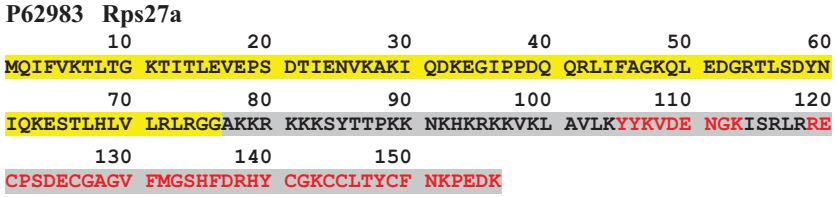

Figure 6.9: Sequence coverage of *Rps27a*: The protein is cleaved after synthesis into ubiquitin (yellow) and 40S ribosomal protein S27a (gray). All detected peptides (red) belong to the 40S ribosomal protein S27a chain.

degraded. The function of the increase or decrease in protein turnover of the other proteins remains to be elucidated by further studies.

To summarize, the enrichment of newly synthesized proteins increases the sensitivity for the detection of proteomic changes in response to stimulation. Furthermore, additional insights into regulatory mechanisms can be gained by comparing of protein synthesis differences to differences in RNA and protein abundance, *e.g.*, elucidating proteins with increased or decreased protein turnover in response to stimulation.

6.2.3 Comparison of protein synthesis to secretion dynamics

Macrophages, as the first actors in immune response, need to communicate with other cells of the immune system to guide them to the place of inflammation and activate their response programs. To achieve this, they extensively secrete signaling proteins like cytokines. In order to not miss this important set of proteins, secretome analysis was performed at the same time points after LPS stimulation.

In total 107 secreted proteins showed differential secretion in response to LPS stimulation within the first three hours (fig. 6.2a). When including the data sets presented in chapter 5.4, *i.e.* including the later time points at 8 and 19 hours after LPS treatment, this number increases to 150 proteins (fig. 6.2a). The extracellular temporal profiles for 25 of these proteins are similar to the detected intracellular changes in protein synthesis (fig. B.7). Most of the proteins in this group are well known to be involved in inflammatory response, including seven cytokines (fig. B.7).

The set of 15 proteins with increased RNA abundance and protein secretion but no intracellular protein synthesis difference or no intracellular detection mainly contains established inflammatory proteins, including six cytokines (*Il6*, *Ccl5*, *Ccl2*, *Csf3*, *Ccl7*, *Il23a*) (fig. 6.10). This highlights how

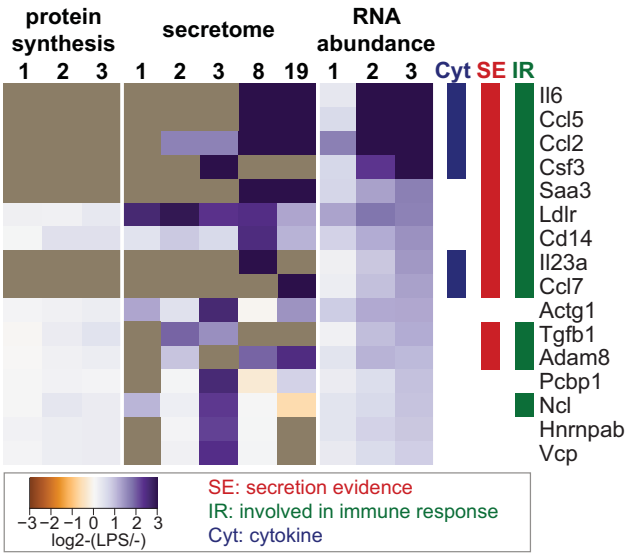

Figure 6.10: Analysis of the secretome retrieves changes not detected in cell lysates: The abundance differences seen on RNA level can only be detected extracellular but not in cell lysate and would have been missed by approaches not applying secretome analysis. (The numbers above the columns give the time after LPS stimulation.)

secretome analysis can provide additional information, which would have been missed in standard proteomic analysis. The exclusive detection of some proteins, like *IL6* and *Ccl5*, at prolonged times after LPS stimulation (8 or 19 hours) possibly stands for delayed regulatory effects in inflammation as discussed in chapter 5.4 (fig. 6.10).

Approximately half of all differentially secreted proteins do not exhibit changes in RNA abundance or intracellular protein synthesis (category L in figure 6.5). These secretion differences could therefore be gene expression independent. 51 of the 89 proteins (all time points including 8 and 19 h) carry a signal peptide or are annotated to be secreted in UniProt[244] and 48 proteins are established factors in immune response. These two observations ensure that these are not artifacts derived by increased apoptosis in the LPS stimulated cells, although this cannot be excluded for all 89 proteins.

Unconventional secretion could be one explanation for the described pattern. The four proteins not having a signal peptide but being assigned to be secreted in UniProt[244] are Macrophage migration inhibitory factor (*Mif*), Nuclease-sensitive element-binding protein 1 (*Ybx1*) (regulation of splicing) and the two Galectins *Lgals1* and *Lgals3*. All four proteins are secreted by a unconventional pathway since they do not enter the endoplasmatic reticulum

and their secretion is not inhibited by monensin or brefeldin A[144–146].

Functional annotation enrichment revealed a functional cluster containing the terms "vesicle" and "pigment granule" with the highest enrichment score (7.71), including proteins such as *App*, *Capg*, *Cltc*, *Coro1a*, *Gpnmb*, *Hsp90ab1*, *Hsp90b1*, *Prdx1*, *Hspd1*, *Hsp90aa1*, *P4hb*, *Pdia*, *Sdcbp*, *Ywhae* and *Ywhaz*. Therefore, increased vesicular trafficking in activated macrophages, which facilitates a rapid increase of extracellular cytokine levels[309], represents one mechanism, independent of gene expression, resulting in increased levels of (newly synthesized) vesicular proteins in the secretome.

A third mechanism of unconventional secretion is the release of exosomes. 149 proteins (34.7%) of all 429 proteins detected in the supernatant of mouse macrophages are assigned to be released via exosomes according to the curated database ExoCarta[158]. Among them as many as 52 proteins are induced in secretion but not in intracellular protein synthesis or RNA expression, representing 58% of the proteins with this profile. An increased release of exosomes in response to LPS stimulation could therefore explain the observed secretion of most of the proteins in this group not bearing a signal peptide (*e.g. Pkm2*, *Ppia*, *Rpl30*, *Rps25*, *Ywhae*, *Hist1h4a*, *Cct8*, *Capg*, *Eef1a1*).

Another enriched cluster includes proteins involved in "antigen processing and presentation" (enrichment score: 1.73). Of note, all H-2 class II histocompatibility antigens quantified in the secretome are induced only on secretome level (fig. 6.11). This group of proteins has been reported to be released by exosomes earlier[310].

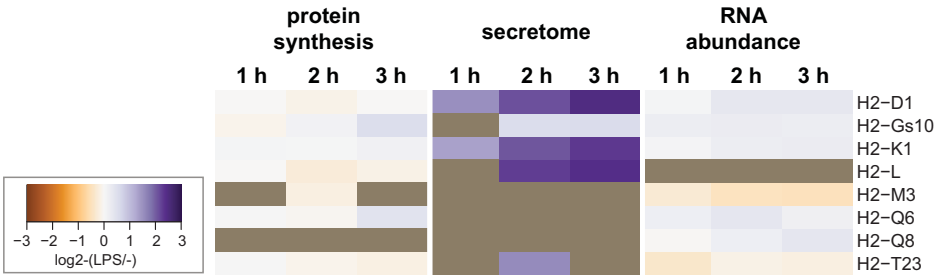

Figure 6.11: Detected H-2 class II histocompatibility antigens

Finally, transmembrane proteins could be released to the medium by proteolytic cleavage. Transmembrane proteins with the described profiles are *Sema4D* which is proteolytically cleaved[311] and *Ptprs*, *App* and *Lfng* that have a cleavage site assigned in UniProt[244].

The comparison of secretion kinetics to intracellular RNA and protein levels reveals three proteins with reduced protein synthesis but increased levels

of extracellular protein (fig. 6.12). All three are transmembrane receptors.

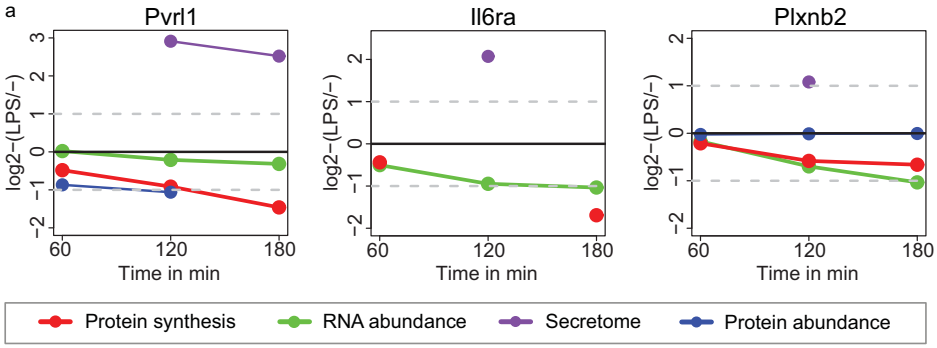

Figure 6.12: Proteins with opposing kinetics: Although RNA abundance and protein synthesis are decreased in response to LPS treatment an increase in extracellular protein levels was measured.

Interleukin-6 receptor subunit alpha (*Il6ra*) is the receptor for *Il6*, a cytokine with central functionality in inflammatory response and exists in a membrane bound form (IL-6RA) and a soluble form (sIL-6RA). The soluble form is produced by two independent mechanisms: partial proteolysis by metalloproteases (ADAM10 and ADAM17)[312] or alternative splicing of mRNA[313]. Binding of the complex of Il6/sIL-6RA to the inducible coreceptor IL6-RB (*Il6st*) is necessary for the activation of Il6 signaling in cells that do not express IL-6RA themselves[314]. The measured discrepancy of intracellular and extracellular *Il6ra* levels could therefore be explained by the exclusive and induced release of sIL-6RA to the extracellular space, while the overall *Il6ra* expression is reduced. This expression pattern would assist the stimulation of other cells while repressing autoinduction, which is further supported by the slightly but not significantly reduced levels of IL6-RB detected on protein and RNA level. Unfortunately, this hypothesis could not be confirmed using the available data, since only peptides derived from the extracellular part of *Il6ra* have been detected (fig. B.8a).

The peptides detected for Poliovirus receptor-related protein 1 (*Pvrl1*) and Plexin-B2 (*Plxnb2*) cover their complete sequences (fig. B.8b,c), while in the secretome samples only peptides derived from the extracellular part of the proteins were detected. Therefore, again protease shedding mechanisms are likely causing the observed protein dynamics (fig. 6.12). For both proteins such mechanisms have been reported earlier: *Pvrl1* is cleaved by alpha-secretase, which is mediated by ADAM10[315]. The single chain precursor of *Plxnb2* is proteolytically processed into a heterodimeric receptor[316]. While the beta subunit contains the transmembrane domain, a short extracellular

part and cytosolic domains, the alpha subunit, consisting of most of the extracellular part, remains linked to the cell surface by weak interactions to the beta subunit[316] (fig. B.8c).

In summary, adding secretome analysis too the experimental workflow helped to get a deeper insight into macrophage activation, revealing proteins that would have been missed by standard proteomic approaches and identifies candidate proteins involved in gene expression-independent regulatory mechanism in immune response.

6.3 Regulated protein classes in macrophage activation

6.3.1 Transcription factors

Transcription factors are low-abundant proteins with spatial and temporal restricted expression that bind DNA and induce or inhibit transcription. Hence, they are key actors in gene expression and tightly regulated in response to cellular perturbation. In the presented data, 59 (14%) of the proteins with differential expression levels or synthesis are transcription factors (fig. 6.3). 32 of these have established functions in inflammation, based on the previously mentioned criteria (fig. 6.13). Among them several members of the NF-κB and AP1 transcription factor complexes, the key regulators downstream of TLR4, are strongly induced after LPS stimulation of macrophages (*Rel*, *Nfkb1*, *Nfkb2*, *Relb*, *Junb*, *Fosl1*, *Fosl2*, *Jun*, *Fos*, *Jund*, *Fosb*)(fig. 6.13a). Additional examples are: *Atf3*, which is probably a transcriptional repressor of pro-inflammatory genes[317], *Egr1/2/3*, *Ets2*, *Cebpd*, which acts together with NF-κB to maximize *Il6* production[220], *Irf7* and *Irf4*. Interferon regulatory factor 4 (*Irf4*) has been found to be essential for M2 macrophage polarization[186]. Since LPS is thought to induce a M1 macrophage phenotype[186], the induction of *Irf4* suggests the additional participation of *Irf4* in the development of the M1 macrophage phenotype.

Most of the detected transcription factors known to act downstream of LPS are induced in response to LPS stimulation (25 of 35 induced proteins), while only 5 of 17 proteins with reduced expression or synthesis are established inflammatory proteins (fig. 6.13). Several of the induced transcription factors with unknown function in immune response are involved in cellular stress response or apoptosis, both of which are well known to be activated by LPS[318] (*e.g. Csrnp1*, *Csrnp2*, *Maff*, *Atf4* and *Plagl2*). In contrast, several of the transcription factors with reduced expression or synthesis inhibit transcription (*Tcfap4*, *E2f7*, *E2f8*, *Hbp1* and *Bhlhe41*). These proteins are

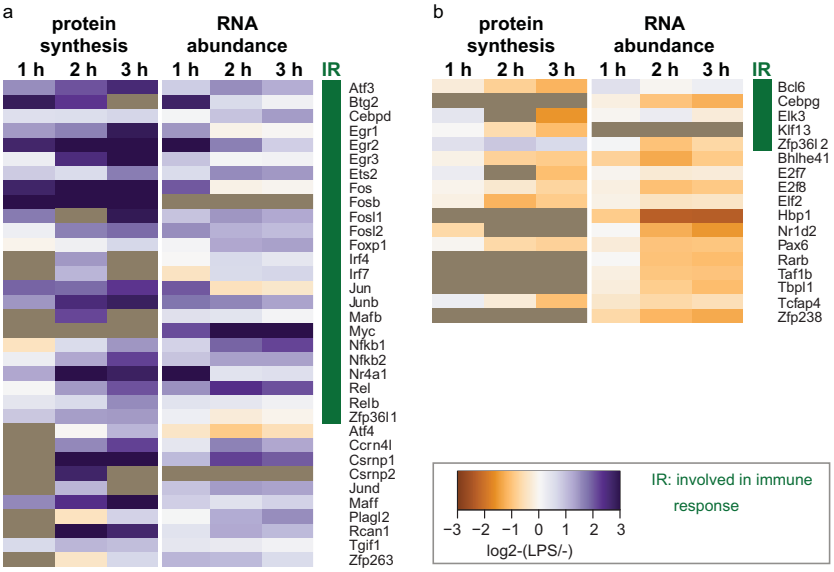

Figure 6.13: RNA abundance and protein synthesis profiles for the detected transcription factors showing either induced (a) or reduced (b) expression or protein synthesis. (Transcription factors appearing twice in the list (isoforms were distinguished) but have not been quantified on protein level were removed. The two transcription factors regulated on secretome level are not shown.)

probably down-regulated after LPS stimulation to allow the transcription of genes being usually repressed under "normal" conditions.

Transcription factors often precisely concert the timing of target gene expression. In order to find transcription factors that regulate multiple of the significant proteins detected in this study, transcription factor target enrichment was performed using MetaCore[246]. The heatmap in figure 6.14 displays the enrichment scores (z-score) for transcription factors with at least five downstream targets and an enrichment score higher than eight at at least one of the investigated time points. Enrichment was performed separately for significant RNAs and newly synthesized proteins. Not all of the retrieved transcription factors have been identified in the presented data set, but since these low abundant proteins could be missed in the analysis, they could still regulate expression downstream of LPS.

The enrichment patterns on RNA level are more specific than on protein level, which is probably caused by different translation rates of the downstream proteins for individual transcription factors (fig. 6.14). On RNA level especially the members of the NF-κB family are highly enriched already one hour after stimulation. For these transcription factors a similar strong enrichment on protein level can be detected from two hours of stimulation on,

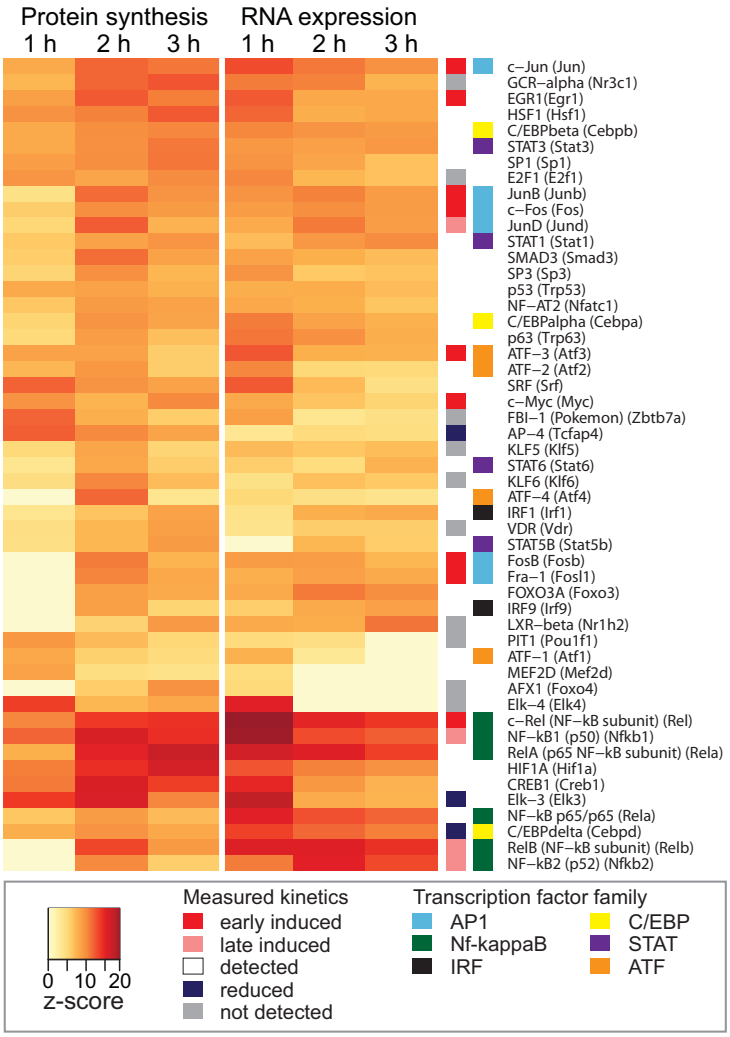

Figure 6.14: Transcription factors showing an enrichment of downstream targets among the significant proteins. The heatmap displays the z-score provided by MetaCore[246] after loading all significant proteins or RNAs at the investigated time points to the enrichment algorithm. Transcription factors with at least five downstream targets and an enrichment score higher than eight at at least one of the time points are shown. Furthermore the temporal dynamics measured in this study are highlighted based on the categorization presented in section 6.2. Members of important transcription factor families in immune response are color coded.

while only a moderate enrichment was detected after one hour (fig. 6.14). The members of the AP1 transcription factor family exhibit moderate enrichment of target proteins at all time points.

Several transcription factors regulate expression of multiple proteins with significant differences in RNA abundance at one hour after stimulation, while the enrichment score at later time point decreases (fig. 6.14), *e.g.*, *Jun* and three members of the ATF transcription factor family. Similar patterns have been observed by Bhatt *et al.* when scanning for binding motifs of transcription factors in their RNA expression data set derived from lipid A (the active component of LPS) stimulated mouse bone-marrow derived macrophages[223]. Similarly, the observed pattern for STAT family members, being enriched across the complete time frame, coincides with their observations (fig. 6.14).

Strong enrichment of downstream target proteins on protein level after one hour of stimulation was found for only few transcription factors, including *Tfcap4*, *Myc*, *Srf*, *Elk3* and *Elk4*. Hence, these transcription factors could be involved in the regulation of early responses to LPS stimulation. ETS domain-containing protein Elk-3 (*Elk3*) and Activator protein 4 (*Tfcap4*) protein synthesis was reduced after three hours of LPS stimulation (fig. 6.13b). In combination, this points to a possible involvement of these proteins in the regulation of endotoxine tolerance[192] as described in chapter 2.3.

Target proteins downstream of Transcription factor E2F1 (*E2f1*) are enriched across the complete time frame under investigation (fig. 6.14). Notably, Transcription factor E2F7 (*E2f7*) and Transcription factor E2F8 (*E2f8*) are reduced after three hours of LPS stimulation (fig. 6.13b). Both transcription factors inhibit E2F-dependent transcription in a synergistic fashion, repressing *E2f1*-dependent apoptosis[319]. Hence, down-regulation of these two proteins could ensure ongoing *E2f1*-dependent transcription. Although *E2f1* itself has not been detected in the dataset, the combination of these observations points to an involvement of *E2f1* in the regulation of inflammation.

In summary, a high number of transcription factors was found to be regulated in response to LPS stimulation of mouse macrophages. This includes well known regulators of immune response, as well as potentially novel ones. Among the novel candidates most of the induced transcription factors regulate stress response and apoptosis, while the down-regulation of transcription factors seems to either repress functions not necessary for inflammation or omit inhibition of the transcription of pro-inflammatory proteins. Target gene enrichment analysis reveals temporal regulation pattern of transcription factors in response to LPS stimulation. The data suggests a participation of *E2f1*, a transcription factor regulating cell cycle and apoptosis[320], in immune response.

6.3.2 Receptors

Receptors recognize extracellular signals and transmit them by activating signaling cascades leading to the induction of transcriptional response programs. Hence, they play important roles in cell-to-cell communication and pattern recognition, which are the key functionalities needed to react to bacterial invasion. Additionally, they can function in cell migration and transmembrane transport.

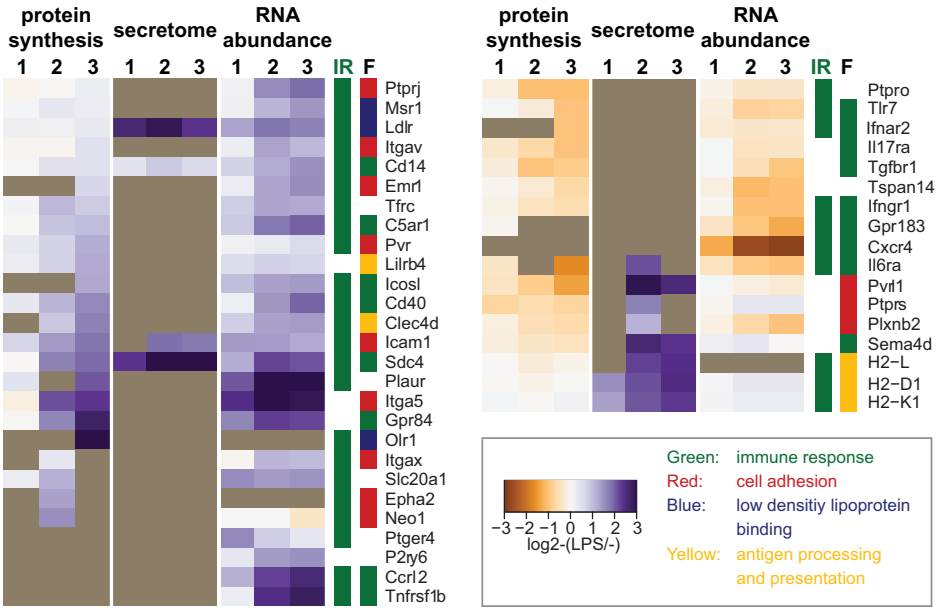

Figure 6.15: Significant receptors. IR: proteins involved in immune response based on the previously defined parameters (regulation downstream of LPS, AP1 or NF-κB described in the MetaCore database[246] or any GO term containing "immune", "defense response" or "inflammatory response"); F: main functionality of the receptor based on UniProt[244]. The numbers above the columns give the time after LPS stimulation.

In total 44 receptors with differential expression or synthesis in response to LPS stimulation were detected (fig. 6.3, fig. 6.15). Most of them have established functions in immune response (*e.g. Cd14*, *Sdc4*, *Ccrl2*, *Icam1*, *Gpr84*). In addition, several of the detected receptors are involved in cell adhesion (*e.g. Ptprj*, *Itgav*, *Itga5*, *Itgax*, *Emr1*, *Pvr*, *Icam1*, *Neo1*, *Plxnb2*), low density lipoprotein binding (*Msr1*, *Ldlr*, *Olr1*) and antigen processing and presentation (*e.g. H2-L*, *H2-K1*, *H2-D1*, *Clec4d*, *Lilrb4*) (fig. 6.15), representing processes highly regulated in inflammation.

27 receptors exhibit induced expression or synthesis profiles, including

Cd14, a receptor for LPS that cooperates with TLR4 and MD-2 together inducing immune response after LPS stimulation[321](fig. 6.15).

Three induced receptors were detected intracellular as well as extracellular (*Icam1*, *Sdc4* and *Ldlr*). While Intercellular adhesion molecule 1 (*Icam1*) and Syndecan-4 (*Sdc4*) are induced in protein synthesis intracellular and extracellular, Low-density lipoprotein receptor (*Ldlr*) shows only extracellular an induced profile. All three transmembrane receptors are cleaved from the cellular surface into a soluble form in response to stimulation[322–324]. Since the intracellular *Ldlr* levels remain constant, LPS seems to induce proteolytic *Ldlr* cleavage rather than *Ldlr* gene expression. *Ldlr* binds low density lipoprotein and mediates its endocytosis. Lipoproteins can bind LPS and therefore reduce LPS mediated responses such as *Tnf* production[325]. The fast (already detected at one hour) release of Ldlr may prevent inhibition of LPS induced anti-bacterial programs by lipoproteins.

Among the nine down-regulated receptors two are toll-like receptors (*Tlr7* and *Tlr13*)(fig. 6.15). Toll-like receptor 7 (*Tlr7*) is expressed in the endosomes and recognizes imidazoquinolines and ssRNA[321]. The reduced expression of this protein in response to LPS stimulation has not been shown earlier. The transcription of the relatively novel member of the toll-like receptor family *Tlr13* is activated by *Ets2* and inhibited by NF-κB[326], both of which are induced in the samples.

The opposing extra- and intracellular kinetics of the three receptors *Il6ra*, *Plxnb2* and *Pvrl1* can most probably be explained by an increase in proteolytic cleavage at the external side of the plasma membrane, as discussed in section 6.2.3. Similar conclusions can be drawn for two receptors only induced in the secretome (*Sema4d*, *Ptprs*)(fig. 6.15). Notably, the soluble form of Semaphorin-4D (*Sema4d*) is the ligand of Plexin-B2 (*Plxnb2*), both of which are unconventional secreted[327] (see section 6.2.3).

Functional annotation enrichment of all regulated proteins revealed the KEGG pathway "Cytokine-cytokine receptor interaction" (Benjamini corrected p-value: $1.0\text{x}10^{-4}$), with 9 receptors of 27 proteins assigned to this pathway (fig. B.9). Interestingly, only receptors for cytokines of the TNF family are induced while the remaining receptors show reduced profiles. Thereby, a selfstimulation with these groups of secreted cytokines may be prevented, such as for the detected cytokine receptor pairs *Il6*/*Il6ra* and *Tgfb1*/*Tgfbr1*, arguing for a paracrine effect of these cytokines. Contrarily, *Tnf*/*Tnfrsf1b* represents a cytokine-receptor pair of the Tnf family for which selfstimulation is promoted, representing an at least partly autocrine function of *Tnf*.

In conclusion, the high number of differentially expressed, synthesized or secreted receptors highlights their importance in the regulation of inflamma-

tion. Along with functions in immune response and antigen processing and presentation, several of them are involved in cell adhesion and low-density lipoprotein binding, possibly representing processes regulated in response to LPS stimulation. Gene expression-independent regulation of several receptors was detected, pointing to an extensive induction of extracellular proteolytic cleavage as well as exosome release. In addition, synthesis profiles of receptors in combination with their ligands might be used to differentiate between autocrine and paracrine secretion mechanisms.

6.3.3 Receptor ligands, cytokines and growth factors

Receptor ligands, especially cytokines and growth factors, possess central functions in cell-to-cell signaling. Their binding to the corresponding receptors triggers signaling cascades, resulting in the activation of transcriptional response programs.

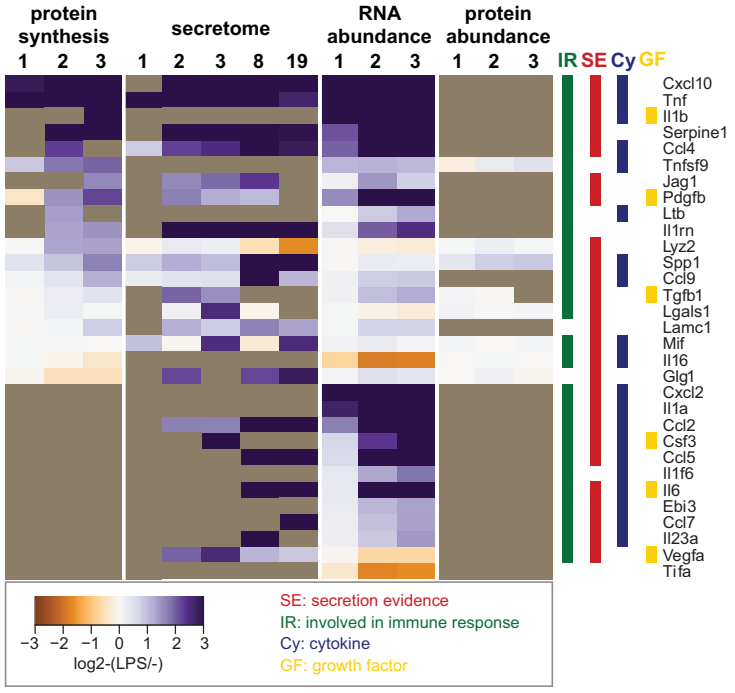

Figure 6.16: Significant signaling molecules. The numbers above the columns give the time after LPS stimulation.

In this study 31 receptor ligands were regulated in response to LPS stimulation, including 20 cytokines and 3 growth factors (fig. 6.16). A very strong and fast intra- and extracellular induction was observed for *Cxcl10*, *Tnf*, *Serpine1* and *Il1rn*. Interleukin-1 receptor antagonist protein (*Il1rn*),

that inhibits Interleukin-1 receptor binding, is much stronger induced in the secretome than intracellular.

A slightly delayed response to LPS stimulation can be seen for *Ccl4*, *Tnfsf9*, *Jag1*, *Pdgfb*, *Ltb*, *Lyz2*, *Ccl2* and *Csf3*. By including the data sets described in chapter 5.4, receptor ligands were detected that show a very slow response on protein level (8-19 hours), while RNA abundance was already regulated within three hours of LPS stimulation (*Spp1*, *Ccl9*, *Ccl5*, *Il6*, *Il23a*, *Ccl7*). Since many cytokines are low abundant, some of them were only detected on RNA level (*Il1a*, *Cxcl2*, *Ebi3*).

Notably, only two signaling proteins are reduced after LPS stimulation (*Il16*, *Tifa*). Pro-IL-16 is cleaved by caspase-3 into a secreted C-terminal part and a N-terminal part, which is translocated into the nucleus[328]. Secreted Interleukin 16 (*Il16*), a ligand for CD4, stimulates migration of several immune cells, like T cells, monocytes, eosinophils and dendritic cells[329–332], while the C-terminal part induces G(0)/G(1) cell cycle arrest[328]. In addition, soluble *Il16* desensitizes the chemokine receptors CCR5 and CXCR4[333]. Reduction of *Il16* levels in response to LPS stimulation could therefore either prevent cell cycle arrest or the desensitization of chemokine receptors in target cells, like CCR5 which binds the cytokines *Ccl4*, *Ccl5* and *Ccl7*, being strongly induced in the presented data set (fig. B.9).

Ligands with contradicting intra- and extracellular profiles, such as *Glg1*, are most probably externalized by an unconventional secretion pathway, as discussed in chapter 6.2.3.

In summary, a high number of receptor ligands are regulated in response to LPS stimulation. Most of them show induced intra- and extracellular profiles with different kinetics and strength of induction, highlighting the tight temporal regulation of cytokine release from activated macrophages. *Il16* is one cytokine with reduced expression. Its reduction possibly prevents cell cycle arrest or desensitization of chemokine receptors.

6.3.4 Kinases and phosphatases

The mammalian proteome contains more than 500 kinases and around 100 phosphatases[334], acting together to regulate phosphorylation dynamics in the cell. Phosphorylation of proteins often initiates or maintains signaling pathways. Hence, kinases and phosphatases constitute key players in the regulation of signaling pathways activated in inflammation. Major changes in phosphorylation levels in response to macrophage stimulation have been demonstrated recently[224;225]. *E.g.*, Weintz *et al.* reported 1850 phosphoproteins, among which 24% and 9% of the phosphorylation sites were up- or down-regulated in response to LPS stimulation of primary macrophages,

respectivelly.

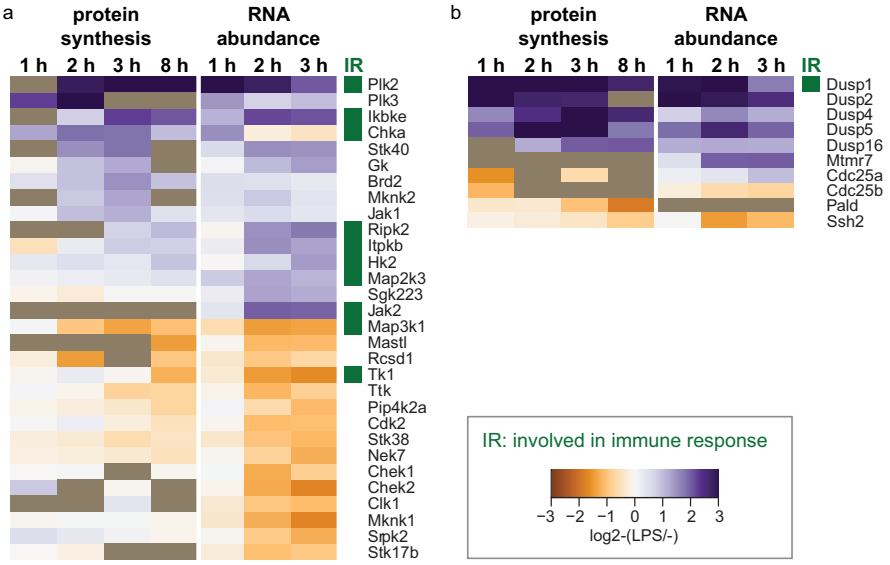

Figure 6.17: Significant kinases (a) and phosphatases (b).

Within the presented data sets 32 kinases and 10 phosphatases exhibit significant differential expression or synthesis (fig. 6.17). Only twelve of these proteins are involved in immune response based on the previously mentioned parameters. 15 kinases are induced in response to LPS stimulation while 17 are reduced.

Polo-like kinases 2 and 3 (*Plk2*, *Plk3*) are the strongest induced kinases, followed by Inhibitor of nuclear factor kappa-B kinase subunit epsilon (*Ikbke*), Choline kinase alpha (*Chka*) and Serine/threonine-protein kinase 40 (*Stk40*)(fig. 6.17). *Plk2* and *Plk3* do not only regulate the cell cycle, but are also considered as stress response genes[335], possibly explaining their strong induction in response to LPS treatment. Interestingly, although both proteins are induced on the level of protein synthesis to a similar extend, the RNA abundance difference for *Plk2* is four times higher than for *Plk3*. This could be explained by a higher translation rate for *Plk3* compared to *Plk2*.

Chka, which catalyzes the first phosphorylation reaction for the biosynthesis of phosphatidylcholine, is rapidly induced after LPS addition, similarly to *Plk2* and *Plk3*, and contributes to the regulation of cell growth and cell stress/defense mechanisms[336]. Notably, *Chka* RNA abundance is elevated after two hours of stimulation, pointing to a negative feedback mechanism, as described in chapter 6.2.1.

Ikbke, which is strongly induced after three hours of LPS stimulation, phosphorylates inhibitors of NF-κB thus activating NF-κB dependent tran-

scription[337]. Thereby it is possibly necessary for propagation of NF-κB target gene expression. Finally the kinase *Stk40*, which is strongly induced after two hours of stimulation, has been described as a possible negative regulator of NF-κB[338].

Slightly less induction was observed for *Gk*, *Brd2*, *Jak1*, *Jak2* and *Mknk2* (fig. 6.17). Induction of Glycerol kinase *Gk*, the key enzyme in the regulation of glycerol uptake and metabolism, by LPS has been demonstrated earlier[339]. Janus kinases (*Jak1*, *Jak2*), that transmit signals of cytokine receptors without intrinsic kinase activity[340], are both induced after two hours of LPS stimulation.

Although a significant increase in RNA abundance was observed for Dual specificity mitogen-activated protein kinase kinase 3 (*Map2k3*), which catalyzes the phosphorylation of the MAP kinase p38, no significant change in protein synthesis could be determined. Mitogen-activated protein kinase kinase kinase 1 (*Map3k1*), another component of the protein kinase signal transduction cascade involved in NF-κB activation, shows reduced abundance and synthesis after two hours of stimulation. Similar kinetics have been observed for CapZ-interacting protein (*Rcsd1*), which has not been associated with immune response until now.

Several kinases exhibit reduced protein levels at even later times after stimulation, including *Stk38* an inhibitor of *Map3k1* [341], while *Chek1*, *Chek2*, *Clk1*, *Mknk1*, *Srpk2* and *Stk17b* are only regulated on RNA level.

Five of the six induced phosphatases are Dual specificity phosphatases (*Dusp1*, *Dusp2*, *Dusp4*, *Dusp5*, *Dusp16*)(fig. 6.17), while several additional Dual specificity phosphatases were detected, showing no significant regulation (*Dusp3*, *Dusp6*, *Dusp7*, *Dusp11*, *Dusp12*, *Dusp19*, *Dusp22*). All regulated Dual specificity phosphatases are mitogen-activated protein kinase phosphatases (MKPs), which can dephosphorylate MAPKs (mitogen-activated protein kinases) and therefore act as antagonists of MAPK signaling cascades[342]. Dual specificity phosphatases can be induced by various stimuli. Since this induction often depends on MAPK activation, it has been proposed to be a negative feedback mechanism for down-regulating MAPK signaling[342].

Among the phosphatases with reduced synthesis, two M-phase inducer phosphatases (*Cdc25a*, *Cdc25b*) were detected, both involved in the regulation of cell cycle progression. Notably, several of the down-regulated kinases contribute to cell cycle regulation (*Cdk2*, *Mastl*, *Chek1*, *Chek2*, *Clk1*, *Nek7*).

In summary, LPS activation of mouse macrophages results in a tight regulation of kinases and phosphatases, representing the need for spatial and temporal control of phosphorylation patterns to induce inflammatory signaling pathways. In addition to kinases and phosphatases exhibiting well

described functions in the regulation of immune response several proteins involved in cell cycle control were detected. This observation could represent on one hand parallel functions of these kinases and phosphatases but on the other hand also a regulation of proliferation in response to macrophage activation. Notably, five Dual specificity phosphatases were strongly induced after stimulation, possibly enabling negative regulation of inflammation.

Chapter 7

Discussion

Differences between cellular phenotypes can be defined by changes in proteome composition as well as protein localization and activity. To decipher the mechanisms driving the change from one phenotype to another, robust methods for the investigation of protein abundance, localization and activity regulation are necessary.

Changes in protein abundance are driven by the collective sum of synthesis and degradation of proteins. Methods that investigate these processes primarily utilize pulse labeling with radioactive or stable isotopes in cell culture to selectively label and detect newly synthesized proteins and investigate their degradation. While labeling with radioactive isotopes allows for very short labeling pulses, the identification of the labeled proteins is challenging and requires extensive protein separation to determine the protein synthesis rates for single proteins. In contrast, stable isotope labeling in tandem with mass spectrometry provides qualitative and quantitative information for individual proteins. One of the caveats of stable isotope labeling experiments is the requirement for detectable amounts of labeled proteins[9], which is typically achieved by longer labeling periods. This limits the applicability of these methods to the investigation of long term regulatory responses.

Other aspects are the limited scan speed and dynamic range capabilities of current generation mass spectrometers hampering the detection of low abundant proteins. These hurdles are particularly relevant when analyzing the cellular secretome, where the high abundance of serum components supplemented with cell growth medium limits the depth of coverage in these samples by mass spectrometric analysis.

7.1 Labeling, enrichment and quantification of newly synthesized proteins

In the course of this thesis a method was developed solving both of these problems by pulse labeling of newly synthesized proteins with a non-canonical azide-containing amino acid (AHA). The azide group facilitates the efficient isolation of newly-synthesized proteins from complex background via click-chemistry. While this approach has been used for the investigation of protein synthesis earlier[11;12;114], its application in secretome analysis is novel and has been demonstrated in chapter 5 to outperform alternative methods in the number of identified proteins, the applicability to complex backgrounds and sophisticated experimental designs.

Two strategies for enrichment of AHA containing proteins are commercially available. First, AHA containing proteins can be labeled with a biotin-alkyne and enriched using NeutrAviding beads and second, the proteins can be covalently linked to alkyne-bearing agarose beads. The advantage of the biotin-alkyne is the possible detection of newly synthesized proteins using western blotting with anti-biotin antibodies. Thus, the labeling efficiency can be evaluated prior to the work intensive enrichment and mass spectrometric measurement and the enrichment procedure can be monitored, as demonstrated in section 4.1. Since this reagent does not contain a cleavable linker, as the one used by Dieterich *et al.*[11], a high background of NeutrAvidin derived peptides in the samples after release from the beads cannot be prevented, which represents the major disadvantage of this reagent.

Therefore, covalent coupling of newly synthesized proteins to the alkyne-bearing agarose resin is advantageous, because contamination-free samples are produced by on-bead digestion. This was demonstrated by the detection of more than 2000 additional proteins (section 4.2). Furthermore, due to the ability of loading higher initial total protein amounts and since newly synthesized proteins are covalently bound to the resin, the alkyne-bearing agarose is the ideal choice for secretome analysis if protein concentrations are high and extensive washing is necessary to remove contaminating proteins.

In addition to the simple detection of newly synthesized proteins, the processes that drive the changes in cellular phenotypes need to be quantitatively compared under multiple conditions and in dependence of time. First attempts in this direction have been made by Kramer *et al.* combining AHA with iTRAQ labeling of peptides in order to investigate changes in protein synthesis of *Escherichia coli* upon perturbing the growth conditions[114;115]. For protein quantification, only peptides containing AHA were used to ensure that they are newly synthesized. As only methionine containing peptides can

be labeled with AHA, the number of peptides to be submitted to quantitative analysis is relatively low, resulting in low numbers of quantified proteins (394).

In this thesis a different quantitative method was combined with AHA labeling, since the metabolic nature of the AHA approach makes it very amenable to the combination with SILAC methods. Two different labeling protocols that combine AHA with SILAC were tested: complete labeling with SILAC amino acids before AHA incorporation or the combination of AHA with pSILAC labeling, a method which has been used earlier for the comparison of protein synthesis under different experimental conditions. Both methods allowed the reproducible detection of thousands of newly synthesized proteins.

Complete SILAC labeling has the advantage that part of the sample can be used for the quantification of changes in global protein expression in parallel. In contrast to pSILAC labeling, ideally 100% of the amino acids used have to be replaced by stable isotope labeled amino acids, which necessitates several cell duplications. Due to the requirement for high amounts of stable isotope labeled amino acids and special cultue medium, the cost of an experiment can be high. Furthermore, sufficient incorporation of the isotopic label needs to be validated experimentally.

An additional advantage of pSILAC is its ability to discriminate new from pre-existing proteins. Test experiments using samples not labeled with AHA for protein enrichment with NeutrAvidin as well as agarose beads revealed that despite the use of very stringent washing conditions, relatively high numbers of non-specific protein identifications were made. This was likely caused by side reactions with thiol groups of cysteines catalyzed by the highly reactive Cu(I), since no background was detected when excluding Cu from the same test experiment. These side reactions are negligible if a good labeling efficiency with AHA is achieved (*e.g.* by applying extended labeling times). Since high efficiencies are not achievable in applications with short labeling times or high protein background, like in secretome samples, the combination of AHA with pSILAC is the ideal approach to circumvent those problems. The SILAC label is then used to distinguish between background binding and newly synthesized proteins. Alternatively, Cu(I) can be excluded when using cyclic alkynes in a ring strain supported 1,3-cycloaddition[121] as performed by Nessen *et al.* with a non-commercial alkyne-bearing resin[12].

In mouse macrophages the incorporation of AHA was detectable after only five minutes of treatment (section 4.1). As many as 4852 newly synthesized proteins were quantified after 20 minutes of pulse labeling with AHA and SILAC amino acids comparing LPS stimulated to unstimulated mouse macrophages (section 6.1.2). Thereby, the number of newly synthesized pro-

teins quantified by Kramer *et al.* was exceeded more than ten times, applying a labeling pulse of similar length[114;115]. It should be noted that they used a less complex model system and no enrichment with the alkyne-agarose resin, which represents an additional advantage of the method presented in this thesis.

Pulsed SILAC labeling alone can be used to compare protein synthesis under two conditions[112]. For the detection of protein synthesis differences in mouse macrophages after LPS stimulation for two or three hours, this method showed poor correlations between biological replicates, which complicated the detection of changing proteins (section 4.2). In contrast, using a combination of AHA and pSILAC labeling enabled, after enrichment, the detection of hundreds of significantly changing proteins with very good correlations between biological replicate samples (section 4.2). Therefore, the reduction of signal-to-noise ratios due to depletion of "old" proteins represents the major advantage of the described method. Hence, the described method facilitates the investigation of newly synthesized proteins using stable isotope labeling on the labeling time scale similar to radioactive isotope labeling techniques.

Moreover, due to the high background of serum proteins, the results obtained by secretome analysis without enrichment were even worse. Here, the detection of secreted proteins was nearly impossible, while high numbers were detected using the presented approach (section 5.1).

7.1.1 Impact of AHA on protein synthesis and secretion

Since the incorporation of AHA into newly synthesized proteins is disfavored in comparison to methionine[117], the described approach is limited to either naturally or genetically manipulated methionine auxotrophic organisms, like mammals. In addition, AHA carries a novel functional group that can potentially influence the structure, stability, functionality and even subcellular location of labeled proteins. Hence, before applying this approach, the impact of AHA on the system of interest needs to be investigated and experimental conditions have to be optimized.

In this study this question was addressed by measuring the viability and translation rate of AHA treated mouse macrophages. While no influence on cell viability was observed after up to four hours of AHA treatment, the global translation rate was reduced by approximately 25% after two hours. This is likely caused by the decreased activity of methionine-tRNA synthetase towards AHA in comparison to methionine[117]. Since the aim of this thesis was

to compare protein synthesis under two different conditions, this reduction in global protein synthesis should have a similar effect on the compared cells and should therefore not influence the results of the experiment. This hypothesis was tested by comparing the protein synthesis ratios retrieved by the described method to ratios measured using pSILAC without AHA treatment. Protein synthesis differences in AHA treated but not in untreated samples were only detected for four and seven proteins of about 5000 proteins after two and three hours LPS stimulation, respectively (section 4.2). Therefore, a major impact of AHA on the measured differences in protein synthesis can be excluded.

To ascertain that AHA does not influence protein secretion, the secretomes of treated and untreated cells were compared. No significant difference in protein secretion for cells grown in the presence of AHA was observed.

One of the potential limitations of this method is that the replacement of methionine by AHA restricts the pool of detectable proteins to those that contain methionine. Only 106 of 53623 proteins in the mouse fraction of the UniProt database[244] do not contain methionine. Additional 2430 proteins contain only an N-terminal methionine which is often cleaved directly after protein synthesis. Therefore, 4.7% of all mouse proteins are not detectable with the described method. In addition to this limitation the use of AHA for tagging could potentially introduce a detection bias for proteins with specific characteristics, such as methionine rich sequences. However, this possibility was excluded by comparing various protein properties such as methionine numbers, protein lenght, isoelectric point and hydrophobicity of the proteins detected in AHA treated samples to all proteins in the complete proteome or secretome and samples not labeled with AHA (section 4.2, 5.5).

7.2 Secretome analysis

The selective enrichment of secreted and newly synthesized proteins addresses a major challenge in secretome analysis, this being the abundant background of serum proteins. It is common to perform these types of analyses using serum-free medium based approaches. However, the method facilitates the detection of hundreds of secreted proteins including proteins of diverse functionalities in the presence of serum.

When used in combination with pSILAC labeling, the method was applied to in-depth quantitative comparison of secretome composition. It has been demonstrated that quantification by pSILAC is highly reproducible between biological replicates (correlation routinely >0.85), allowing rigorous statistical evaluation. Combining AHA labeling and pSILAC, limited amounts

of cells sufficed to identify multiple low-abundant secretory proteins, including several that are rarely seen in proteomic studies (cytokines, some of the growth factors).

Comparison to published data sets revealed that the method identifies lower numbers of contaminating proteins (section 5.1.1). This additional benefit can be explained by the use of pSILAC to "mark" newly synthesized proteins. These proteins represent a minor fraction of all contaminating proteins released by, *e.g.*, dying cells.

The detection of contaminations can be further reduced by applying short collection times, being facilitated by enrichment. Hence, the optimization of collection times can balance between the identification of high numbers of secreted proteins and the reduction of contaminations. In addition, pSILAC aids in discriminating between peptides derived from secreted proteins and their homologous within serum (section 5.3).

In addition, the method facilitated examination of secretome changes induced within a very short time window. It has been shown that the determined secretion profiles can be used to explain differences in diverse phenotypes.

Alternative enrichment approaches primarily focus on subpopulations of secreted proteins, such as those that are glycosylated[175], which were predicted to constitute 66% of all secreted proteins. Therefore, these alternative approaches will miss 34% of the secreted proteins. However, the present approach will only miss 5% of proteins based on the absence of methionine in the amino acid sequence.

Vast differences were detected when comparing secretion profiles of cancer and non-cancer cells, reflecting differences in metastatic properties. The differential secretion of several proteases and metalloproteases, as well as their targets, illustrates the extensive processing of extracellular and cell surface bound proteins. Proteins with well described functionalities in cancer progression were detected, as well as several with potentially novel roles in cancer progression, such as *PLOD1/2/3*, *TNC* and *SDC4* (section 5.1.1). To evaluate these proteins as cancer biomarkers, comparative analysis of multiple patient specific cancer and non-cancer cell lines is needed, followed by extensive confirmation of the observations *in vivo*.

The finding that the method presented here is readily applicable to primary cells is highly relevant, indicating that patient- or tissue-derived cells are directly accessible to secretome analysis. This type of study would permit bypassing of artifacts that may be introduced by the use of cell lines. This undervalued problem was clearly demonstrated by the vast differences between primary hepatocytes compared to hepatoma cell lines (section 5.3).

The scope of another commonly overlooked aspect of secretome analysis,

namely the impact of serum starvation that is routine applied in the large majority of secretome studies, was also determined. Cell type independent starvation responses, as well as cell type dependent ones were demonstrated. Numerous biologically important proteins, including multiple cytokines, whose secretion was found to be induced or reduced upon starvation, emphasizes the need for careful evaluation of biomarkers to confirm that they are not a consequence of culture conditions (section 5.2). Here, RAW 267.4 mouse macrophages were extremely sensitive towards serum starvation, explaining the lack of secretome studies in this cell line up to now.

A particularly powerful and unique application of the presented method is in studying the kinetics of secretory activity, as demonstrated for stimulation of macrophages by LPS. In addition to the successful identification of several known LPS effector proteins (*Tnf*, *Il6*, *Vegf*, several of the chemokines), many others were found that had not been seen in the context of LPS stimulation before. Furthermore, the ability to derive temporal profiles for each of them has added novel insights to this heavily studied system. Therefore, the combination of in-depth proteome profiling with time-resolved protein secretion provides a powerful and unique application of the method. This will be helpful to study cellular responsiveness to specific stimuli, including drugs and growth factors and may provide important clues to their mechanism of action.

Finally, all secreted proteins were combined in a compendium revealing cell specific functionalities for proteins exclusively secreted by one cell type. The addition of further experiments will provide more specific classifications of protein functions.

In addition to the presented experiments, this method could also be applied to alternative approaches in secretome analysis, such as secretory vesicle or exosome enrichment, 3D cell culture, patient derived cell cultures, cellular co-culturing systems, or the investigation of secretory pathways using inhibitors. Moreover, due to the recent success in labeling zebrafish larvae with AHA[128], there is potential for future application in whole organism studies.

7.3 Activation of mouse macrophages

Having developed the method for deep investigation of newly synthesized and secreted proteins, valuable clues into the mechanism defining the phenotype of a cell or organism can be retrieved. However, a really profound view on a system will be reached when combining these data sets with the investigation of RNA and total protein expression changes. The integration of these data

permits to distinguish between transcriptional, translational and localization regulation.

Hence, a combination of RNA and protein quantification approaches were applied to study the activation of mouse macrophages with LPS in a time-course experiment, facilitating the examination of early mechanisms that regulate innate immune responses. This process involves rapid changes of the transcriptional program. Although this model system has been extensively studied, a comprehensive investigation of LPS stimulated macrophages combining RNA and protein expression data along with temporal changes in protein synthesis and secretion is currently unavailable.

Combined across all data sets, it was observed that 782 RNAs and proteins were differentially expressed or synthesized. Approximately half of these changes were detected on the proteome level (section 6.1.2). The number of detected changing proteins exceed previous proteomic studies of this system, with the exception of the study by Du *et al.*[222]. However, this previous study employed a fold-change cut off of >1.2 and <0.8 to define expression differences[222]. These represent extremely relaxed fold-change cut offs, resulting in an artificially large data set.

Investigating newly synthesized proteins had the main impact for the detection of differences between treated and untreated cells at the proteome level in this study. This is highlighted by the observation that few proteins were differentially expressed when examining total protein abundances.

In conclusion, enrichment of newly synthesized proteins increases the sensitivity for the observation of minor changes due to reduction of complexity and interference from preexisting proteins. Additionally, deeper proteome coverage, achieved by enrichment, led to the detection of numerous low abundant proteins.

This study covers 38% of the proteins with established transcriptional regulation downstream of LPS stimulation (section 6.1.2). Improvements of this value could be achieved by the extension of the study towards different amounts of LPS used for stimulation, longer stimulation times or the combined stimulation with LPS and LPS-binding protein (LBP), since it was previously shown that LPS binds to LBP prior to TLR4 binding[197]. Moreover, under *in vivo* conditions direct or indirect interactions with other cell types can influence stimulation response programs. The use of primary macrophages would be a small step into the direction of *in vivo* conditions.

Notably, it has been demonstrated that the analysis of secreted proteins increases the detection of established proteins in inflammatory response and has a higher relative contribution towards their detection than analysis of the intracellular proteome. This is not a standard approach in proteomics and its application to the very same cell culture conditions was extremely

challenging without the developed method.

One of the primary benefits of the presented data results form the combinatory and time-resolved detection of RNA and protein expression changes in combination with changes in protein synthesis and secretion to group proteins based on mechanism regulating their abundance in the cell. Thereby, starting points for further investigation of regulatory mechanisms can be proposed. One example is the detection of possible negative feedback mechanisms for several proteins (mainly transcription factors) (fig. 6.6). These observations are partially validated by previous reports suggesting that the detected proteins are regulated by similar mechanisms.

An additional example for insights gained by the integration of the different data sets is the elucidation of 15 proteins exhibiting increased or decreased protein turnover in response to LPS treatment (fig. 6.8).

Moreover, based on the observed secretion profiles, an induction in vesicular trafficking and exosome release in response to LPS treatment can be proposed, as well as the cleavage of transmembrane proteins, pointing to an increase of extracellular proteolytic cleavage mediated by LPS (section 6.3.2).

The recent view on this system is biased towards results derived by RNA expression measurements due to the lack of proteomic studies. The new data set helps to distinguish between regulated RNA species being truly transcribed and those that do not yield to changes in protein abundance. *E.g.*, several proteins with reduced RNA levels in response to stimulation did not change on protein level. Here, post-translational mechanisms establish protein synthesis and abundance. Conversely, minor differences in RNA abundance levels yielded significant changes in protein synthesis for several proteins (section 6.2). This could be achieved by an increase or reduction in translation rate, which could be evaluated by ribosome occupancy profiling. Notably, the set of proteins with reduced RNA abundance but no change in protein synthesis is enriched in proteins involved in cell cycle as well as those having GTPase regulator activity (section 6.2).

An additional aspect of the presented data set is the temporal resolution, which allows to distinguish between early and late responses to LPS stimulation. These temporal profiles in combination with established knowledge on protein functionality allowed the prediction of novel mechanisms involved in innate immune response, specifically the induction of E2F-1-dependent genes (section 6.3.1), the NF-κB dependent reduction of Toll-like receptor 13 synthesis (section 6.3.2), the inhibition of autocrine stimulation by certain cytokines due to the reduced synthesis of the corresponding receptor (section 6.3.2), the omission of cell cycle arrest or CCR-5 inhibition by IL-16 (section 6.3.3) as well as a possibly higher translation rate for PLK-3 compared to PLK-2 (section 6.3.4).

7.4 Conclusion

In the course of this thesis, two novel methods addressing challenges in the field of proteome research have been presented.

First, the unbiased identification of specific and combinatorial binding of proteins to long DNA-sequences (in the range of CRMs) using *in vitro* enrichment and mass spectrometry can complement current methods in gene expression research. While the detection of DNA-binding sites of established transcription factors is achieved using ChIP-approaches, the detection of unknown DNA-binding proteins has not been possible for these kind of regulatory sequences yet.

The method was applied to the comparison of proteins binding to three CRMs active at different times and in different tissues during *Drosophila melanogaster* development. In total 72 potential regulators of these CRMs have been detected, of which some bound to the expected CRM, thereby confirming the presented method. *In vivo* validation of selected candidates is in progress. This method will help to elucidate how transcriptional networks are orchestrated by interlacing signals from multiple transcription factors.

Second, a novel method combining pulsed metabolic labeling with the non-canonical amino acid AHA with pSILAC has been developed. The use of this technique permitted highly reproducible and selective relative quantification of newly synthesized proteins in response to cellular perturbation with a deep proteome coverage.

Adaption of this method for secretome analysis addresses one of the major challenges in this field. In doing so, this methodology opens up new avenues towards novel applications of mass spectrometry based proteomics for the study of the secretome. The application of this approach in various experimental setups and cellular systems as presented here has demonstrated its broad applicability. Moreover, the integration of the developed methods with RNA and protein expression measurements is a valuable tool to uncover mechanisms regulating protein abundance and localization in LPS-induced cells mimicking innate immunity.

The high temporal resolution is the major advantage of this method and should be exploited in further applications. This includes temporal secretome analysis as presented in this thesis, but also the time-resolved analysis of, *e.g.*, protein degradation or protein modification. Here, processes only targeting new proteins could be determined.

Chapter 8

Methods

If not stated reagents and chemicals are purchased from Sigma Aldrich and mass spectrometry grade solvents and acids are purchased from Biosolve. Iodoacetamide (IAA), Dithiothreitol (DTT), Acetamide, Sodium dodecyl sulfate (SDS), Ammonium Persulfate (APS), Tetramethylenediamine (TEMED) were retrieved from Bio-Rad Laboratories. Antibodies were purchased from abcam. Trypsin (Gold, mass spectrometry grade) was recieved from Promega.

8.1 Sample preparation

8.1.1 Cell culture and collection of conditioned media

PC3 and WPMY-1 cells were grown in 10-cm cell culture dishes at 37 °C, 5% CO2 in DMEM (GIBCO), supplemented with 10% FBS (GIBCO) until 50% and 70% confluency, respectively (WPMY-1 ∼2x10^6 cells, PC3 ∼2.5x10^6 cells). RAW 264.7 macrophages were grown at 37 °C, 5% CO2 in DMEM (GIBCO) with 10% FBS (GIBCO), 100 mg/L Primocin (InvivoGen) and 4 mM/L L-Glutamin (GIBCO) added until ∼70% confluency (∼1x10^7 cells for pSILAC experiments and ∼0.5x10^6 cells for full SILAC experiments). Hepa1c1 and Hepa1-6 cells were grown in 10 cm cell culture dishes at 37 °C and 5% CO2 in DMEM with 10% FBS and 1x Penicillin/Streptomycin (P/S; P.: 100 U/mL ; S: 100 μg/mL; GIBCO) until ∼80% confluency (∼4x10^6 cells). Primary mouse hepatocytes (PHC) were isolated and cultured as described elsewhere[343]. Viability of cells was determined by trypan blue staining and 6x10^6 living cells were seeded on collagen I-coated 10 cm cell culture dishes. After 4 h of incubation, hepatocytes were attached to the collagen coat and dead cells were washed away. Adherent hepatocytes (>90% confluence) were incubated for additional 18 h in Williams' Medium E (sup-

plemented with 10% FBS, 1x P/S, 2 mM Glutamin).

8.1.2 Pulse-labeling of cells with AHA and SILAC

For complete SILAC labeling RAW 264.7 cells were grown in SILAC medium (DMEM non-GMP formulation without methionine, arginine and lysine, GIBCO) supplemented with 100 mg/L Primocin, 4 mM/L L-Glutamine, 10% dialysed fetal bovin serum (GIBCO), 30 μg/mL methionine and stable isotope labeled amino acids (84 μg/mL [$^{13}C_6$]L-arginine and 146 μg/mL [$^{13}C_6$,$^{15}N_2$]L-lysine (Cambridge Isotope Laboratories, Inc) or the natural versions of this amino acids). Labeling effeciency was determined by mass spectrometry after each round of splitting until labeling was complete. Then cells were grown up and frozen in liquid nitrogen.

To deplete cells from methionine, lysine and arginine, the cells were incubated for 30 min (RAW 264.7, Hepa1c1, Hepa1-6 and primary hepatocytes) or one hour (PC3 and WPMY-1) in depletion medium (DMEM non-GMP formulation without methionine, arginine and lysine, GIBCO) with 10% dialysed fetal bovin serum (GIBCO), 4 mM/L L-Glutamine and 100 mg/L Primocin, before incubation in the same medium supplemented with 0.1 mM L-azidohomoalanine (AnaSpec, Inc) and in case of pSILAC either 84 μg/mL [$^{13}C_6$,$^{15}N_4$]L-arginine and 146 μg/mL [$^{13}C_6$,$^{15}N_2$]L-lysine or 84 μg/mL [$^{13}C_6$]L-arginine and 146 μg/mL [4,4,5,5-D_4]L-lysine (Cambridge Isotope Laboratories, Inc). RAW 264.7 cells were stimulated with 100 ng/mL lipopolysaccharides (Escherichia coli O111:B4; Sigma) for the indicated times. All assays were performed as independent biological duplicates with reversed SILAC labels. Collected media was centrifuged (5 min at 1,000 x g) COMPLETE, EDTA free protease inhibitor (Roche) added and frozen at -80 °C. The protein concentration of the collected media determined with a NanoDrop (Thermo Fisher Scientific) was 3 mg/mL in all experiments with serum containing medium.

Cells were washed three times with warm PBS, detached using a cell scraper and centrifuged (5 min at 1,000 x g). After removal of PBS the cell pellets were frozen at -80 °C or directly subjected to cell lysis.

The accuracy of quantification (fig. 5.3) was tested using medium of LPS stimulated RAW 264.7 cells (0-2 hours).

Viability test

Viability was determined with trypan blue staining in independent biological triplicates. Therefore, cells were detached with a cell scraper and transferred to an eppendorf tube. After appropriate dilution with Trypan Blue Stain

0.4% (GIBCO 15250-061) the solution was transfered to a Neubauer hemacytometer and alive and death cells (blue) were counted. Viability was calculated as the ratio of alive cells devided by total number of cells. For statistical analysis of the results a Student's t-test was applied. Experiments with p-values below 0.01 were regarded as being statisticaly significant different in viability.

8.1.3 Enrichment of newly synthesized proteins and on-bead digestion using biotin-alkyne and NeutrAvidin beads

Cell lysis

Cells were suspended in 150 μL 1% SDS in 50 mM Tris-HCl, pH 8.0 supplemented with COMPLETE, EDTA free protease inhibitor (Roche) and lysed for 15 min on ice. After sonication (3 times 10 seconds; 1 min on ice in between), the lysate was agitated for 5 min on maximum speed and centrifuged at 13,000 x g (5 min 4°C). The supernatant was transferred to a clean tube and the protein concentration determined using a NanoDrop (Thermo Fisher Scientific). The sample was either directly labeled with the biotin-alkyne or precipitated and stored at -20 °C.

Labeling of newly synthesized proteins with biotin-alkyne

Labeling with alkyne-biotin was performed using the Click-iT^{TM} Biotin Protein Analysis Detection Kit (Invitrogen C33372) applying the vendors protocol. Shortly, up to 200 μg azide-labeled protein in a maximum volume of 50 μL in 1% SDS, 50 mM Tris-HCl, pH 8.0 were combined with 100 μL of 2X Click-iT^{TM} Reaction Buffer containing the alkyne detection reagent and the sufficient volumn of 18 megaOhm water for a final volume of 160 μL. The solution was agitated for 5 seconds at maximum speed followed by addition of 10 μL of CuSO4 (Component C) and agitating for 5 seconds. After addition of 10 μL of Click-iT^{TM} Reaction Buffer Additive 1, agitating for 5 seconds and 2-3 min pausing, 20 μL of reconstituted Click-iT^{TM} Reaction Buffer Additive 2 were added followed by agitation. The bright orange solution was rotated end-over-end for 20 min.

To remove reaction reagents the sample was precipitated by methanol-chloroform precipitation. Briefly, 600 μL of methanol, 150 μL chloroform and 400 μL megaOhm water were added subsequently and agitated between each addition. After centrifugation for 5 min at 13,000 x g the upper aqueous phase was removed. To wash the protein pellet, 450 μL of methanol were

added, followed by agitating and centrifugation (5 min at 13,000 x g). This step was performed twice. Now the pellet was air-dried for at least 15 min and frozen or directly subjected to newly synthesized protein enrichment, western blotting or dot blotting.

Enrichment and of newly synthesized proteins and on-bead digestion using NeutrAvidin beads

The protein pellets were diluted in 50 μL 1% SDS in PBS, followed by adjustment to 1% (v/v) NP-40 and 0.05% (w/v) SDS in PBS to a concentration of 25 μg/mL of biotinylated proteins, as determined by dot blotting (see section 8.2.1). For each 25 μg of biotinylated proteins, 200 μL of NeutrAvidin resin slurry (Thermo Scientific) were washed three times with ten bed volumes of NeutrAvidin-binding buffer (1% NP-40, 0.05% SDS in PBS, pH 7.5). Next, the sample solution was added and binding of biotinylated protein was allowed by agitating overnight at 4 °C. Resin was centrifuged at 2,000 x g for 5 min at 4 °C and the supernatant removed. The resin was washed with ten bed volumes of NW-A buffer (1% NP-40, PBS, pH 7.5) for 10 min at RT with constant agitation, centrifuged (2,000 x g for 5 min at 4 °C) and the supernatant discarded. This was repeated two more times, followed by washing two times with NW-B buffer (50 mM ammonium bicarbonate), two times with PBS and two times with NW-B buffer in the same manner as above.

For on-bead digestion for each initial 100 μL of NeutrAvidin slurry the resin was resuspended in 87 μL 50 mM ammonium bicarbonate and heated for 10 min at 70 °C under agitation. A final concentration of 2-3 M urea was achieved by adding 10 μg of urea for each initial 100 μL of NeutrAvidin slurry. After cooling down to room temperature TCEP was added to a final concentration of 3.125 mM and incubated for 30 min at room temperature under constant agitation. Next, after adjusting to a final iodoacetamide concentration of 11.2 mM, the resin was incubated for 30 min at room temperature in the dark under constant agitation. 0.1 μg endoproteinase Lys-C (Wako) per 100 μL of suspension was added and incubated for 4 h at 37 °C under agitation. After addition of calcium chloride to a final concentration of 0.1 mM the sample was trypsinized overnight at 37 °C under agitation with 1 μg trypsin per 100 μL suspension. The suspension was transferred to an empty spin column in a microcentrifuge tube and the supernatant containing the tryptic peptides was separated by brief pulsed centrifugation. The sample was acidified with formic acid and subjected to desalting (see section 8.1.6). At each step during enrichment small aliquots of resin or sample were kept for subsequent western blot analysis.

8.1.4 Enrichment of newly synthesized proteins and on-bead digestion using agarose beads

Newly synthesized proteins from concentrated media (Amicon Ultra® Centrifugal Filters 3 kDa cut off, Millipore) (250 μL) or cell lysates were enriched using the Click-iT® Protein Enrichment Kit (Invitrogen C10416), applying the vendor's protocol with slight modifications. For enrichment form media 100 μL agarose resin slurry was used and the volumes of all reagents were divided by two. After washing the resin with 900 μL water, medium sample, diluted in 250 μL urea buffer (supplemented with the kit) and catalyst solution were added and incubated for 16 to 20 hours at room temperature. Alternatively cells were lysed in urea buffer for 15 min followed by sonication, 5 min agitating and centrifugation and then subjected to the catalytic reaction using 200 μL of agrose resin slurry. After washing the resin with 900 μL water, 0.5 mL or 1 mL SDS buffer (supplied with the kit) and 0.5 μL or 1 μL 1M DTT were added and agitated at 70 °C for 15 min. The supernatant was aspirated and 3.7 mg iodoacetamide in SDS washing buffer was added and incubated for 30 min in the dark. The resin was transferred to a spin-column and washed with 20 mL of SDS buffer, 20 mL of 8 M urea/100 mM Tris pH 8, 20 mL of 20% isopropanol (only medium samples) and 20 mL of 20% acetonitril. After dissolving the resin in digestion buffer (100 mM Tris pH 8, 2 mM CaCl2, 10% acetonitrile), 0.5 μg trypsin were added and incubated over-night at 37 °C. The peptide solution was collected and the resin was washed with 500 μL water. Both solutions were combined and acidified with 20 μL 10% CF3COOH.

8.1.5 Enrichment of DNA-binding proteins

Drosophila melanogaster embryo collection

Embryo collection was performed from in-house wild-type *Drosophila melanogaster* population. Before starting the collection three prelay periods of 30 min with subsequent exchange of plates were performed. Then new plates were placed for 10 hours. After additional 2 hours, the embryos were washed off the collection plates through sieves and transfered to 50% bleach (approximately 5% NaOCl) and stirred for 2 minutes. After extensive washing with water they were transfered to nitrile mesh on paper towels, blotted with light pressure to dry followed by flash-freezing and storage at -80 °C until further use.

Nuclear extracts

During the preparation of nuclear extracts the samples were constantly kept at 4 °C. For one nuclear extract around 30-40 grams of embryos were used. Isolation of nuclei was achieved by dissolving the embryos in buffer NU1 using 2 mL per gram embryos (Nu1: 15 mM HEPES pH 7.6, 10 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA pH 7.9, 0.5 mM EGTA pH 7.9, 350 mM Sucrose, 2 mM DTT, 0.2 mM Phenylmethylsulfonylfluorid (PMSF)). The suspension was homogenized and filtered through Miracloth. The cloth was rinsed with 2-3 volumes of buffer NU1. From the flow-through the nuclei were pelleted for 15 minutes at 13k x g. The lipid layer was removed and the supernatant discarded. For lysis, nuclei were resuspended in 0.5 mL LS buffer (low salt) per gram starting material (LS buffer: 15 mM HEPES pH 7.6, 20% Glycerol, 20 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA pH 7.9, 1 mM DTT). An equal volume of HS buffer (high salt) was added and the suspension rotated for 20 min at 4 °C (HS buffer: 15 mM HEPES pH 7.6, 20% Glycerol, 800 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA pH 7.9, 1 mM DTT). Soluble nuclear material was separated from insoluble chromatin and lipids by centrifugation (1 hour at 38k rpm). The supernatant was dialyzed against NE200 buffer overnight using Spectra/Por Membrane with 3kDa cut off (NE200 buffer: 15 mM HEPES pH 7.6, 20% Glycerol, 200 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA pH 7.9, 1 mM DTT). The total protein concentration of the nuclear extract was detemined using a NanoDrop (Thermo Fisher Scientific).

Preparation of biotinylated dsDNA

500 base pair long biotinylated double stranded DNA (dsDNA) was prepared by PCR amplification using biotinylated forward primer bearing a SmaI restriction site. For the PCR reaction genomic DNA, PCR buffer, 2.5 μM of

CRM	forward primer	reverse primer
tinC	cgtagaaCCCGGGtgaaatattagcctttgcccgatgg	aagatcaatcagaatttcagcttgg
633	cgtagaaCCCGGGaaccgaaaacgatgaaagcgtgagg	gcctgcgagagcaggatggactgag
bap3	cgtagaaCCCGGGagcaatgatcagtcgatcagtcgac	atctggcggatgagggcctcgagatg

Table 8.1: Primers used for the preparation of dsDNA by PCR. (Sequences in capital letters are the restriction site for SmaI)

each dNTP (dNTP mix from Bioron 11001) and DSF-Taq DNA Polymerase (Bioron 101005) were mixed and biotinylated forward and reverse primer

(MWG) added to a final concentration of 0.5 μM. PCR cycler (Biorad) was set up to heat the samples to 94 °C within 2 minutes (step 1) and hold this temperature for 30 seconds (step 2), followed by a decrease to 55 °C withing 30 seconds (step 3). After final heating of the sample for one minute to 72 °C (step 4), step 2 to 4 were performed for 35 cycles. Finally the temperature was kept at 72 °C for 3 minutes, followed by a temperature decrease to 4 °C.

The purity of PCR products was tested using 1.5% agarose gels and visualized using ethidiumbromide.

DNA extraction was performed using phenol-chloroform precipitation. Therefore, 100 μL PCR reaction mixture was adjusted to 200 μL using TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and mixed with 300 μL buffer-saturated phenol. Organic and aqueous phases were separated using gel separation tubes by centrifugation at maximum speed for 5 min. After chloroform addition (300 μL), mixing and additional centrifugation, the top aqueous phase was transfered to a new tube. 3 volumes ethanol, 1/10 volume of 3 M sodium acetate and 10 μL glycogen were added, mixed and frozen at -80 °C over night. DNA pellets were separated for the solution by centrifugation at maximum speed at 4 °C, washed with 200 μL 70% cold ethanol and again centrifuged. The pellets were dried and resuspended in 30 μL of buffer EB (QIAquick PCR Purification Kit, Quiagen). The concentration of DNA was determined using a NanoDrop (Thermo Fisher Scientific).

DNA affinity pull-down

For DNA affinity pull-down the protocol by Mittler *et al.* was modified[8]. The biotinylated dsDNA (∼250 pmol) was diluted in buffer DW (20 mM Tris-HCl, pH 8.0, 2 M NaCl, 0.5 mM EDTA, 0.03% NP-40) to give a final volume of 0.4 mL. 1 mg of Dynabeads MyONe Streptavidin C1 (Invitrogen) was equilibrated with two washes of buffer TE (0.01% (v/v) NP40) (0,4 mL each) and two washes of buffer DW (0.75 mL each) followed by incubation of the biotinylated dsDNA (rotary wheel, 3 h, room temperature) with the equilibrated magnetic beads. Wasching of the beads was performed one time in 0.4 mL TE (containing 0.02% NP40) and three times in 0.4 mL DW buffer, followed by resuspension in 0.1 mL of buffer DW and storrage at -20 °C until further use.

The beads harboring the immobilized dsDNA were incubated for one hour in blocking buffer (20 mM Hepes, pH 7.9, 0.05 mg/mL BSA, 0.05 mg/mL glycogen, 0.3 M KCl, 0.02% NP40, 2.5 mM DTT, 5 mg/mL polyvinylpyrrolidone) at RT on a rotary wheel using 1.3 mL buffer per mg beads. Excess blocking buffer was removed by washing the beads with 1.3 mL (per mg beads) restriction endonuclease buffer (New England Biolabs (NEB) buffer

3) containing 0.02% NP40 and two times with 2.67 mL buffer G (20 mM Tris-HCl, pH 7.3, 10% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 10 mM potassium glutamate, 0.04% NP40, 2 mM DTT, 0.4 mM PMSF and 0.005 mg/mL each of aprotinin, leupeptin and pepstatin A) per mg beads.

Nuclear extracts (stored at -80°C) were cleared for insoluble matter by centrifugation at 15,000 x g for 20 min at 4 °C and adjusted to 10 mM potassium glutamate, followed by quick dilution with one volume of buffer G containing 0.2 mg/mL of poly dAdT. Insoluble matter was removed by centrifugation at 15,000 x g for 10 min at 4 °C. Preclearing of the nuclear extracts was performed by incubation with MyOne streptavidin magnetic beads at a final concentration of 1.5 mg/mL for one hour at 4 °C, that have been washed with buffer TE supplemented with 0.02% NP40, buffer DW and equilibrated in buffer G. After removal of the beads using a magnetic separator the nuclear extracts were mixed with the blocked DNA bearing beads at a final concentration of 0.7 mg/mL and incubated for 3 hours at 4 °C. Recovered beads were washed four times with buffer G (1.8 mL per mg beads) and ones with restriction endonuclease buffer (NEB buffer 3; adjusted to 0.02% (v/v) NP40, 2.5% (v/v) glycerol, 1 mM DTT, 0.2 mM PMSF and 0.005 mg/mL each of aprotinin, leupeptin and pepstatin A). The last washing step was omited in the final experiments as described in chapter 3.

For elution using DNases the beads were dissolved in 266 μL of DNAse buffer, enzyme added and agitated for the depicted times (DNase I: Turbo ^{T}M DNase (Applied Biosystem Ambion), SmaI (New England Biolabs)).

Elution with SDS was performed by heating a suspension of the beads in 1% or 0.1% SDS in 0.1 mM Tris pH 6.8 to 95 °C for 3 min.

8.1.6 Sample preparation for mass spectrometry

Media not submitted to the enrichment procedure were concentrated using Amicon Ultra® Centrifugal Filters (15 mL 3 kDa cut off, Millipore) at 4 °C and 6000 rpm and washed with 50 mM ammonium bicarbonate until the solution was clear. In-solution reduction, alkylation and digestion was performed by adding 10 mM DTT and incubation at 65 °C for 30 min, followed by the addition of 5 mM iodoacetamide and incubation in the dark for 60 min. For digestion, 0.5 μg of trypsin was added and incubated at 37 °C over night.

Digestion of DNA-binding proteins and total cell lysates was performed according to a previously published protocol[229] with slight modifications. The concentration of DTT was adjusted to 0.1 M. After heating to 95 °C for 5 min the samples were cooled to room temperature and concentrated using Amicon Ultra® Centrifugal Filters (0.5 mL 3 kDa cut off, Millipore). 200 μL 8 M urea in 0.1 M Tris/HCl pH 8.5 were added and concentrated.

After addition of 100 μL iodoacetamide the samples were mixed at 600 rpm for one minute and incubated without mixing for 5 min followed by the concentration of the sample. The buffer was exchanged by adding 100 μL 8 M urea in 0.1 M Tris/HCl pH 8.0 followed by concentration of the sample for three times. After addition of 0.5 μg of endoproteinase Lys-C (Wako) in 40 μL 8 M urea in 0.1 M Tris/HCl pH 8.0 and mixing at 600 rpm for one minute, the filter units were incubated at room temperature overnight. 120 μL 50 mM ammonium bicarbonate with 0.5 μg trypsin were added and incubated at room temperature for four hours.

The acidified samples (10% CF3COOH) as well as the samples prepared by on-bead digestion were desalted using Sep-Pak® cartridges (Vac 1cc (50 mg) tC18, Waters) as described elsewhere[10]. The desalted peptide samples were either directly analyzed by nano-LC-MS/MS or fractionated using isoelectric focusing or strong-anion-exchange chromatography. For isoelectric focusing an Agilent 3100 OFFGEL Fractionator was used in combination with ImmobilineTM DryStrips (ph 3-10 NL, 13 cm, GE Healthcare). Dried samples were resuspended in 360 μL H2O and diluted in 1.44 mL 1.25 x IEF stock solution (6% glycerol, 2% Ampholytes pH 3-10 (1:50) (BioRad)). Focusing was performed at a constant current of 50 mA with a maximum Voltage of 4000 V. After reaching 20 kVh the samples were collected, acidified with CF3COOH and desalted using StageTips[344]. For medium samples 12 fractions were combined to 6 fractions. The peptide samples were dried and dissolved in 4% acetonitrile, 0.1% formic acid. Alternatively, strong anion exchange was used for the fractionation of the samples into 6 fractions, as described elsewhere[345] with minor adaptions. Briefly, desalted peptide samples were dissolved in 200 μL Britton & Robinson Universal Buffer pH 2 and loaded on the anion exchange stage tip which was placed into a C18-stage tip. Peptides were successively eluted to new C18-stage tips each with Britton & Robinson Universal Buffer pH 12, pH 8, pH 6, pH 4 and pH 2. The flow through was reloaded to the C18-stage tips and desalted as described above.

8.1.7 Sample preparation for mRNA expression measurement

Approximately 1x10^6 RAW 264.7 cells were prepared for mRNA expression measurements and stimulated for one , two and three hours with 100 ng/mL LPS in DMEM (GIBCO) with 10% FBS (GIBCO), 100 mg/L Primocin (InvivoGen) and 4 mM/L L-Glutamin (GIBCO) ($\sim$70% confluency) or left untreated. Each experiment was performed in triplicate. For RNA purification the RNeasy Plus Mini Kit (Quiagen) was used according to vendors protocol

with cell lysis directly in the cell plates and homogenization using QIAshredder (Quiagen). For the elution of bound mRNA two times 50 μL water were used and the final concentration was determined using NanoDrop (Thermo Fisher Scientific).

8.2 Measurement

8.2.1 Dot blot/ Western blot/ Silver staining

Dot blot

For dot blot the sample pellets were diluted in 50 μL 1%SDS in PBS. 200 μL of sample dilutions in PBS and a BSA dilution series were prepared. Nitrocellulose membrane was equilibrated in PBS and the the dot blot was assembled according to the manufacturer's recommendation. The wells were washed ones with PBS before duplicate application of 100 μL of sample dilutions. Samples were removed from wells by the application of vacuum followed by one washing step with PBS.

Western blot

Samples were mixed with 4x loading buffer (1% SDS, 40% glycerol, 20% beta-Mercaptoethanol, 150 mM Tris pH 6.8, 0.004% Bromphenolblue) and heated for 5 min at 95 °C. Protein separation was performed using 12% SDS-Page gels with 4% stacking gel (table 8.2) at a constant voltage of 140 V.

Stock	running gel	stacking gel
H2O	4.5 mL	2.65 mL
30% Acrylamide	5 mL	1 μL
1.5M Tris-HCl pH 8.8	3.25 mL	1.25 mL
10% SDS	130 μL	50 μL
10% APS	60 μL	30 μL
TEMED	20 μL	10 μL

Table 8.2: Composition of SDS-Page gels.

For blotting the gels were incubated in transfer buffer (20% methanol, 20mM Tris, 153 mM Glycin, 0,2% SDS) for 10 min. The semi-dry blotting system (Biorad) was assembled according to vendors directions. Blotting

was performed at maximum 50 V and a maximum current of 80% of the membrane area for one hour. The membrane was washed three times with deionisized water and stained with Ponceau red followed by antibody staining.

Visualization

Blocking of non-specific protein binding to the membrane was achieved by placing the membrane in 5% dry milk in PBS supplemented with 1% Tween20 (PBS-T) for 45 min at RT with gentle agitation on a rocking plate. The membrane was incubated with the primary antibody in the desired dilution in 5% dry milk in PBS-T for 1 h at RT with gentle agitation. Next, the membrane was washed three times with PBS-T for 10 min each, followed by incubation with the secondary antibody dilution (1:10,000 in 5% dry milk in TBS-T) for 45 min at RT with gentle agitation. The membrane was washed three times with PBS-T for 10 min each. Proteins were detected using the western blotting detection system Immobilon ^{TM} Western Chemiluminescent HRP Substrat (Millipore) in combination with an imaging film BioMax MR film from KODAK according to the manufacturer's instructions. For dot bots the amount of biotinylated, that is, newly synthesized proteins in the samples, was estimated by comparing the intensity of the sample dots with the intensity of biotinylated BSA standard dots.

Silver staining

Samples were mixed with 4x loading buffer (1% SDS, 40% glycerol, 20% beta-Mercaptoethanol, 150 mM Tris pH 6.8, 0.004% Bromphenolblue) and heated for 5 min at 95 °C. Protein separation was performed using 12% SDS-Page gels with 4% stacking gel (table 8.2) at a constant voltage of 140 V. The gel was rinsed with water followed by fixation for 45 minutes (H_2O:methanol:acetic acid 50:40:10). After washing with water for 1.5 hours (multiple times exchanged), the gels were incubated in 0.2% (w/v) sodium thiosulfate for 3 min, followed by two times washing with water, 1 min each. Next, the gels were incubated with 0.1% (w/v) silvernitrate at 4 °C for 30 min. 0.04% formaldehyde in 2% (w/v) sodium carbonate (one exchange) were added until staining was complete and stopped with 5% acetic acid, followed by washing with water.

8.2.2 LC-MS/MS

Peptides were separated using a nanoAcquity UPLC system (Waters) fitted with a trapping (nanoAcquity Symmetry C18, 5 μm, 180 μm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 μm, 75 μm x 200 mm). The outlet of the analytical column was coupled directly to an LTQ Orbitrap Velos or Orbitrap Velos Pro (Thermo Fisher Scientific) using a Proxeon nanospray source (solvent A: 0.1% formic acid, solvent B: acetonitrile, 0.1% formic acid). The samples were loaded with a constant flow of solvent A at 15 μL/min onto the trapping column. Peptides were eluted via the analytical column at a constant flow of 0.3 μL/min. During the elution step, the percentage of solvent B increased in a linear fashion from 3% to 25% in 40 min, 110 min or 210 min followed by an increase to 40% in 4 min, 10 min or 5 min and an increase to 85% in 0.1 min, 1 min or 1 min for the different gradient lengths. The peptides were introduced into the mass spectrometer via a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (New Objective). Full scan MS spectra with mass range 300-1700 m/z were acquired in profile mode in the FT with resolution of 30000. The filling time was set at maximum of 500 ms with limitation of 10^6 ions. The most intense ions (up to 15) from the full scan MS were selected for fragmentation in the LTQ. Normalized collision energy of 40% was used and the fragmentation was performed after accumulation of 3 x 10^4 ions or after filling time of 100 ms for each precursor ion (whichever occurred first). MS/MS data was acquired in centroid mode. Only multiply charged (2+, 3+) precursor ions were selected for MS/MS. The dynamic exclusion list was restricted to 500 entries with maximum retention period of 30 s and relative mass window of 10 ppm. Lock mass correction using a background ion (m/z 445.12003) was applied.

8.2.3 mRNA expression measurement

The mRNA expression measurement was performed using GeneChip® Whole Transcript (WT) Sense Target Labeling Assay from Affimetrix. All sample preparation and measurement procedures were performed according to the vendors protocol. The consequent preparation procedures performed were: cDNA synthesis and amplification (several rounds), sample clean up, terminal labeling, hybridization and washing and staining. Affimetrix exon expression chip (MoGene-1_0-st-v1) was used for hybridization and expression measurement.

8.3 Data analysis

8.3.1 Data processing

The mass spectrometric raw data was processed using MaxQuant (version 1.2.2.5 or version 1.2.0.18 for *Drosophila melanogaster*)[1] and MS/MS spectra were searched using the Andromeda search engine[346] against either human (69906 entries), mouse (53623 entries) or drosophila melanogaster (18797 entries) proteins in UniProt[244], to which 265 frequently-observed contaminants as well as reversed sequences of all entries had been added. For growth medium samples each of these databases had been concatenated to bovine-specific part of UniProt (26526 entries) (all downloaded 21.06.2011). Enzyme specificity was set to trypsin/P and a maximum of two missed cleavages were allowed. Cysteine carbamidomethylation was used as fixed modification and methionine oxidation, protein N-terminal acetylation, as well as replacement of methionine by AHA (in case of AHA treatment experiments) were used as variable modifications. The minimal peptide length was set to 6 amino acids. Initial maximal allowed mass tolerance was set to 20 ppm for peptide masses, followed by 6 ppm in the main search and 0.5 Dalton for fragment ion masses. False discovery rates for peptide and protein identification were set to 1%. At least one unique peptide was required for protein identification. The protein identification was reported as an indistinguishable "protein group" if no unique peptide sequence to a single database entry was identified.

For protein quantification a minimum of two ratio counts was set and the "requantify" and "match between runs" function enabled (exception: "match between runs" not enabled for *Drosophila melanogaster* data set). For pSILAC samples a protein group was kept for the further analysis, if it contained at least one mouse or human sequence and the number of identified peptide species carrying an medium-heavy or heavy label divided by the total number of peptide species detected in the complete experimental setup was higher than 0.2. For all other samples proteins assigned to contaminants or reverse sequences were removed. Average protein ratios were reported, if they were quantified in both replicates each based on at least two ratio counts.

8.3.2 Compendium of secreted proteins

To assemble the compendium all detected mouse proteins were mapped to their human homologue in the human UniProt database based on gene names. Only one protein per gene name was kept in the database and proteins without gene name were removed, resulting in 20695 remaining entries. Proteins

were regarded as identified in one cell type if at least 3 assigned MS/MS spectra summed over all experiments of one cell type were detected. The theoretical human secretome was defined to include all proteins with the UniProt keyword "Signal" or "Secreted"[244] or reaching the significance threshold in SignalP4.0[245], producing 3831 proteins. For theoretical cleavage of the signal peptide, 22 amino acids were removed from the N-terminus of the protein sequence, if a signal peptide was predicted by SignalP 4.0 or the UniProt keyword "Signal" was assigned.

8.3.3 Statistical analysis

Mass spectrometry

Statistical analysis was performed using the Limma package in R/Bioconductor[347]. After fitting a linear model to the data, an empirical Bayes moderated t-test was used and p-values were adjusted for multiple testing with Benjamini and Hochberg's method. If not stated otherwise, proteins with an adjusted p-value lower than 0.01 and exhibiting a minimum fold change of 2 were considered to be differentially synthesized or secreted. For analysis of LPS-time course experiments prior to statistical test values with high standard deviations were removed if the standard deviation of biological replicates was higher than 1.2 (or 1.8 for pSILAC without AHA) or the standard deviation was higher than the absolute log2-fold-change for fold-changes higher than 0.6 (or 1.2 for pSILAC without AHA). Correlations between replicates were calculated in R using pearson correlation.

Statistical significance in the *Drosophila melanogaster* data sets was determined using Significance A test[1] in Perseus (written by Jürgen Cox; MPI of Biochemistry Martinsried) independent for the biological replicates. Proteins significant in both biological replicates were considered as candidates for further analysis.

mRNA expression

mRNA expression values were analyzed using Gene Spring GX Software (Agilent). An unpaired t-test with asymptotic p-value calculation followed by Benjamini-Hochberg p-value correction for multiple testing was used to define statistical significant changes in mRNA expression comparing LPS treated samples against untreated sample. After background determination all RNAs with expression values below 50 were excluded from the analysis. Only RNAs with a adjusted p-value higher than 0.01 and a minimum fold-change of 2 were considered as significantly changing.

8.3.4 Bioinformatical analysis

Functional annotation enrichment was performed using DAVID[230]. MetaCore (GeneGo Inc.)[246] was used for pathway map enrichment, transcription factor target gene enrichment and network analysis. The network in section 5.1.1 was visualized in Cytoscape[348]. Functional classification was performed using either MetaCore (individual datasets) or PANTHER classification system[288;289] (secretome compendium). Default parameters were used in all analysis. In the LPS-time course dataset a connection to immune response was assigned if any of the following criteria was fulfilled: the protein is downstream of LPS, NF-κB or AP1 based on the curated GeneGo database or GO annotations contain any term with "immune", "inflammatory response" or "defense response".

Bibliography

[1] Cox, J. & Mann, M. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367–1372 (2008).

[2] Hubner, N. C., Ren, S. & Mann, M. Peptide separation with immobilized pi strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**, 4862–4872 (2008).

[3] Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & Cellular Proteomics* **11** (2012).

[4] Ren, B. *et al.* Genome-wide location and function of dna binding proteins. *Science* **290**, 2306 (2000).

[5] Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-dna binding sites from chip-seq data. *Nucleic Acids Research* **36**, 5221–5231 (2008).

[6] Nielsen, R. *et al.* Genome-wide profiling of ppar gamma:rxr and rna polymerase ii occupancy reveals temporal activation of distinct metabolic pathways and changes in rxr dimer composition during adipogenesis. *Genes & Development* **22**, 2953–2967 (2008).

[7] Robertson, G. *et al.* Genome-wide profiles of stat1 dna association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).

[8] Mittler, G., Butter, F. & Mann, M. A silac-based dna protein interaction screen that identifies candidate binding proteins to functional dna elements. *Genome Research* **19**, 284–293 (2009).

[9] Doherty, M. K. & Beynon, R. J. Protein turnover on the scale of the proteome. *Expert Review of Proteomics* **3**, 97–110 (2006).

[10] Villen, J. & Gygi, S. P. The scx/imac enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **3**, 1630–1638 (2008).

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

[12] Nessen, M. A. *et al.* Selective enrichment of azide-containing peptides from complex mixtures. *Journal of Proteome Research* **8**, 3702–3711 (2009).

[13] Makridakis, M. & Vlahou, A. Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics* **73**, 2291–2305 (2010).

[14] Stastna, M. & Van Eyk, J. E. Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722–735 (2012).

[15] Dowling, P. & Clynes, M. Conditioned media from cell lines: A complementary model to clinical specimens for the discovery of disease-specific biomarkers. *Proteomics* **11**, 794–804 (2011).

[16] Pirkmajer, S. & Chibalin, A. V. Serum starvation: caveat emptor. *American Journal of Physiology-cell Physiology* **301**, C272–C279 (2011).

[17] Hasan, N. M., Adams, G. E. & Joiner, M. C. Effect of serum starvation on expression and phosphorylation of pkc-alpha and p53 in v79 cells: Implications for cell death. *International Journal of Cancer* **80**, 400–405 (1999).

[18] Levin, V. A. *et al.* Different changes in protein and phosphoprotein levels result from serum starvation of high-grade glioma and adenocarcinoma cell lines. *Journal of Proteome Research* **9**, 179–191 (2010).

[19] Cooper, S. Reappraisal of serum starvation, the restriction point, g0, and g1 phase arrest points. *Faseb Journal* **17**, 333–340 (2003).

[20] Alberts, B. *et al. Essential Cell Biology* (Garland Science, New York, ed. 3th, 2009).

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

[22] Ishihama, Y. *et al.* Exponentially modified protein abundance index (empai) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics* **4**, 1265–1272 (2005).

[23] Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

[24] Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. A sampling of the yeast proteome. *Molecular and Cellular Biology* **19**, 7357–7368 (1999).

[25] Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mrna abundance in yeast. *Molecular and Cellular Biology* **19**, 1720–1730 (1999).

[26] Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mrna expression levels on a genomic scale. *Genome Biology* **4**, 117 (2003).

[27] Washburn, M. P. *et al.* Protein pathway and complex clustering of correlated mrna and protein expression analyses in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3107–3112 (2003).

[28] Schmidt, M. W., Houseman, A., Ivanov, A. R. & Wolf, D. A. Comparative proteomic and transcriptomic profiling of the fission yeast schizosaccharomyces pombe. *Molecular Systems Biology* **3**, 79 (2007).

[29] Nie, L., Wu, G. & Zhang, W. W. Correlation between mrna and protein abundance in desulfovibrio vulgaris: A multiple regression to identify sources of variations. *Biochemical and Biophysical Research Communications* **339**, 603–610 (2006).

[30] Tian, Q. *et al.* Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Molecular & Cellular Proteomics* **3**, 960–969 (2004).

[31] Maier, T., Guell, M. & Serrano, L. Correlation of mrna and protein in complex biological samples. *Febs Letters* **583**, 3966–3973 (2009).

[32] Abreu, R. D., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mrna expression levels. *Molecular Biosystems* **5**, 1512–1526 (2009).

[33] Wade, J. T. & Struhl, K. The transition from transcriptional initiation to elongation. *Current Opinion In Genetics & Development* **18**, 130–136 (2008).

[34] Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. M. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65–U72 (2009).

[35] Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* (2012).

[36] Venters, B. J. & Pugh, B. F. How eukaryotic genes are transcribed. *Critical Reviews In Biochemistry and Molecular Biology* **44**, 117–141 (2009).

[37] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **389**, 251–260 (1997).

[38] Lorch, Y., Lapointe, J. W. & Kornberg, R. D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203–210 (1987).

[39] Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).

[40] Iberg, A. N. *et al.* Arginine methylation of the histone h3 tail impedes effector binding. *Journal of Biological Chemistry* **283**, 3006–3010 (2008).

[41] Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in s-cerevisiae. *Science* **309**, 626–630 (2005).

[42] Flaus, A. & Owen-Hughes, T. Mechanisms for atp-dependent chromatin remodelling: farewell to the tuna-can octamer? *Current Opinion In Genetics & Development* **14**, 165–173 (2004).

[43] Fazzio, T. G. *et al.* Widespread collaboration of isw2 and sin3-rpd3 chromatin remodeling complexes in transcriptional repression. *Molecular and Cellular Biology* **21**, 6450–6460 (2001).

[44] Egloff, S. & Murphy, S. Cracking the rna polymerase iictd code. *Trends In Genetics* **24**, 280–288 (2008).

[45] Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).

[46] Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).

[47] Nakanishi, S. *et al.* A comprehensive library of histone mutants identifies nucleosomal residues required for h3k4 methylation. *Nature Structural & Molecular Biology* **15**, 881–888 (2008).

[48] Smale, S. T. & Kadonaga, J. T. The rna polymerase ii core promoter. *Annual Review of Biochemistry* **72**, 449–479 (2003).

[49] Bonn, S. & Furlong, E. E. M. cis-regulatory networks during development: a view of drosophila. *Current Opinion In Genetics & Development* **18**, 513–520 (2008).

[50] Lee, T. I. *et al.* Transcriptional regulatory networks in saccharomyces cerevisiae. *Science* **298**, 799–804 (2002).

[51] Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).

[52] Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the saccharomyces cerevisiae ser3 gene. *Nature* **429**, 571–574 (2004).

[53] Davis, C. A. & Ares, M. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit rrp6p in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3262–3267 (2006).

[54] Holstege, F. C. P. & Clevers, H. Transcription factor target practice. *Cell* **124**, 21–23 (2006).

[55] Hallikas, O. *et al.* Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47–59 (2006).

[56] Mukherjee, S. *et al.* Rapid analysis of the dna-binding specificities of transcription factors with dna microarrays. *Nature Genetics* **36**, 1331–1339 (2004).

[57] Warren, C. L. *et al.* Defining the sequence-recognition profile of dna-binding molecules. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 867–872 (2006).

[58] Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry* **57**, 159–197 (1988).

[59] Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).

[60] Bode, J., Henco, K. & Wingender, E. Modulation of the nucleosome structure by histone acetylation. *European Journal of Biochemistry* **110**, 143–152 (1980).

[61] Kadonaga, J. T. & Tjian, R. Affinity purification of sequence-specific dna-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 5889–5893 (1986).

[62] Nordhoff, E. *et al.* Rapid identification of dna-binding proteins by mass spectrometry. *Nature Biotechnology* **17**, 884–888 (1999).

[63] Gingras, A. C., Aebersold, R. & Raught, B. Advances in protein complex analysis using mass spectrometry. *Journal of Physiology-london* **563**, 11–21 (2005).

[64] Kadonaga, J. T. Purification of sequence-specific binding-proteins by dna affinity-chromatography. *Methods In Enzymology* **208**, 10–23 (1991).

[65] Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased rna-protein interaction screen by quantitative proteomics. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10626–10631 (2009).

[66] Butter, F., Kappei, D., Buchholz, F., Vermeulen, M. & Mann, M. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. *Embo Reports* **11**, 305–311 (2010).

[67] Ku, W. C. *et al.* Complementary quantitative proteomics reveals that transcription factor ap-4 mediates e-box-dependent complex formation for transcriptional repression of hdm2. *Molecular & Cellular Proteomics* **8**, 2034–2050 (2009).

[68] Kim, B. *et al.* The transcription elongation factor tfiis is a component of rna polymerase ii preinitiation complexes. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16068–16073 (2007).

[69] Himeda, C. L., Ranish, J. A. & Hauschka, S. D. Quantitative proteomic identification of maz as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Molecular and Cellular Biology* **28**, 6521–6535 (2008).

[70] Himeda, C. L. *et al.* Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. *Molecular and Cellular Biology* **24**, 2132–2143 (2004).

[71] Bartels, S. J. J. *et al.* A silac-based screen for methyl-cpg binding proteins identifies rbp-j as a dna methylation and sequence-specific binding protein. *Plos One* **6**, e25884 (2011).

[72] Dejardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).

[73] Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).

[74] Alberts, B. *et al. Molecular biology of the cell* (Garland Science, New York, ed. 5th, 2008).

[75] Shine, J. & Dalgarno, L. Terminal-sequence analysis of bacterial ribosomal-rna correlation between 3'-terminal polypyrimidine sequence of 16-s rna and translational specificity of ribosome. *European Journal of Biochemistry* **57**, 221–230 (1975).

[76] Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger-rnas. *Nucleic Acids Research* **15**, 8125–8148 (1987).

[77] Grossman, A. D. *et al.* Mutations in the rpoh (htpr) gene of escherichia-coli-k-12 phenotypically suppress a temperature-sensitive mutant defective in the sigma-70 subunit of rna-polymerase. *Journal of Bacteriology* **161**, 939–943 (1985).

[78] Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme b-12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Research* **32**, 143–150 (2004).

[79] Gottesman, S. The small rna regulators of escherichia coli: Roles and mechanisms. *Annual Review of Microbiology* **58**, 303–328 (2004).

[80] Morfeldt, E., Taylor, D., Vongabain, A. & Arvidson, S. Activation of alpha-toxin translation in staphylococcus-aureus by the trans-encoded antisense rna, rnaiii. *Embo Journal* **14**, 4569–4577 (1995).

[81] Waters, L. S. & Storz, G. Regulatory rnas in bacteria. *Cell* **136**, 615–628 (2009).

[82] Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microrna: are the answers in sight? *Nature Reviews Genetics* **9**, 102–114 (2008).

[83] Jansen, R., Bussemaker, H. J. & Gerstein, M. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research* **31**, 2242–2251 (2003).

[84] Kliman, R. M., Irving, N. & Santiago, M. Selection conflicts, gene expression, and codon usage trends in yeast. *Journal of Molecular Evolution* **57**, 98–109 (2003).

[85] Friberg, M., von Rohr, P. & Gonnet, G. Limitations of codon adaptation index and other coding dna-based features for prediction of protein expression in saccharomyces cerevisiae. *Yeast* **21**, 1083–1093 (2004).

[86] MacKay, V. L. *et al.* Gene expression analyzed by high-resolution state array analysis and quantitative proteomics - response of yeast to mating pheromone. *Molecular & Cellular Proteomics* **3**, 478–489 (2004).

[87] Arava, Y. *et al.* Genome-wide analysis of mrna translation profiles in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3889–3894 (2003).

[88] Wu, G., Nie, L. & Zhang, W. W. Integrative analyses of posttranscriptional regulation in the yeast saccharomyces cerevisiae using transcriptomic and proteomic data. *Current Microbiology* **57**, 18–22 (2008).

[89] Varshavsky, A. The n-end rule. *Cell* **69**, 725–735 (1992).

[90] Rogers, S., Wells, R. & Rechsteiner, M. Amino-acid-sequences common to rapidly degraded proteins - the pest hypothesis. *Science* **234**, 364–368 (1986).

[91] Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208 (1999).

[92] Dever, T. E. *et al.* Phosphorylation of initiation factor-2-alpha by protein-kinase gcn2 mediates gene-specific translational control of gcn4 in yeast. *Cell* **68**, 585–596 (1992).

[93] Moldave, K. Eukaryotic protein-synthesis. *Annual Review of Biochemistry* **54**, 1109–1149 (1985).

[94] Ryazanov, A. G., Shestakova, E. A. & Natapov, P. G. Phosphorylation of elongation factor-ii by ef-2 kinase affects rate of translation. *Nature* **334**, 170–173 (1988).

[95] Buckingham, R. H. Codon context and protein-synthesis - enhancements of the genetic-code. *Biochimie* **76**, 351–354 (1994).

[96] McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J. & Tate, W. P. Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5431–5435 (1995).

[97] Bandow, J. E. *et al.* The role of peptide deformylase in protein biosynthesis: A proteomic study. *Proteomics* **3**, 299–306 (2003).

[98] Larrabee, K. L., Phillips, J. O., Williams, G. J. & Larrabee, A. R. The relative rates of protein-synthesis and degradation in a growing culture of escherichia-coli. *Journal of Biological Chemistry* **255**, 4125–4130 (1980).

[99] Mosteller, R. D., Goldstein, R. V. & Nishimoto, K. R. Metabolism of individual proteins in exponentially growing escherichia-coli. *Journal of Biological Chemistry* **255**, 2524–2532 (1980).

[100] Martin, S. J., Lennon, S. V., Bonham, A. M. & Cotter, T. G. Induction of apoptosis (programmed cell-death) in human leukemic hl-60 cells by inhibition of rna or protein-synthesis. *Journal of Immunology* **145**, 1859–1867 (1990).

[101] Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13004–13009 (2006).

[102] Eden, E. *et al.* Proteome half-life dynamics in living human cells. *Science* **331**, 764–768 (2011).

[103] Khmelinskii, A. *et al.* Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nature Biotechnology* **30**, 708 (2012).

[104] Pratt, J. M. *et al.* Dynamics of protein turnover, a missing dimension in proteomics. *Molecular & Cellular Proteomics* **1**, 579–591 (2002).

[105] Price, J. C., Guan, S. H., Burlingame, A., Prusiner, S. B. & Ghaemmaghami, S. Analysis of proteome dynamics in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14508–14513 (2010).

[106] Doherty, M. K., Hammond, D. E., Clagule, M. J., Gaskell, S. J. & Beynon, R. J. Turnover of the human proteome: Determination of protein intracellular stability by dynamic silac. *Journal of Proteome Research* **8**, 104–112 (2009).

[107] Cargile, B. J., Bundy, J. L., Grunden, A. M. & Stephenson, J. L. Synthesis/degradation ratio mass spectrometry for measuring relative dynamic protein turnover. *Analytical Chemistry* **76**, 86–97 (2004).

[108] Jayapal, K. P. *et al.* Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. *Journal of Proteome Research* **9**, 2087–2097 (2010).

[109] Milner, E., Barnea, E., Beer, I. & Admon, A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Molecular & Cellular Proteomics* **5**, 357–365 (2006).

[110] Rao, P. K., Roxas, B. A. P. & Li, Q. B. Determination of global protein turnover in stressed mycobacterium cells using hybrid-linear ion trap-fourier transform mass spectrometry. *Analytical Chemistry* **80**, 396–406 (2008).

[111] Xiao, G. G. *et al.* Determination of protein synthesis in vivo using labeling from deuterated water and analysis of maldi-tof spectrum. *Journal of Applied Physiology* **104**, 828–836 (2008).

[112] Selbach, M. *et al.* Widespread changes in protein synthesis induced by microrna. *Nature* **455**, 58–63 (2008).

[113] Mann, M. Functional and quantitative proteomics using silac. *Nature Reviews Molecular Cell Biology* **7**, 952–958 (2006).

[114] Kramer, G. *et al.* Identification and quantitation of newly synthesized proteins in escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Molecular & Cellular Proteomics* **8**, 1599–1611 (2009).

[115] Kramer, G. *et al.* Proteome-wide alterations in escherichia coli translation rates upon anaerobiosis. *Molecular & Cellular Proteomics* **9**, 2508–2516 (2010).

[116] Hendrickson, T. L., de Crecy-Lagard, V. & Schimmel, P. Incorporation of nonnatural amino acids into proteins. *Annual Review of Biochemistry* **73**, 147–176 (2004).

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

[118] Liu, D. R. & Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4780–4785 (1999).

[119] Hendrickson, W. A., Horton, J. R. & Lemaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (mad) - a vehicle for direct determination of 3-dimensional structure. *Embo Journal* **9**, 1665–1672 (1990).

[120] Ngo, J. T. & Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Accounts of Chemical Research* **44**, 677–685 (2011).

[121] Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(i)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie-international Edition* **41**, 2596 (2002).

[122] Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the staudinger ligation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 19–24 (2002).

[123] Beatty, K. E., Xie, F., Wang, Q. & Tirrell, D. A. Selective dye-labeling of newly synthesized proteins in bacterial cells. *Journal of the American Chemical Society* **127**, 14150–14151 (2005).

[124] Beatty, K. E. *et al.* Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angewandte Chemie-international Edition* **45**, 7364–7367 (2006).

[125] Beatty, K. E. & Tirrell, D. A. Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorganic & Medicinal Chemistry Letters* **18**, 5995–5999 (2008).

[126] Dieterich, D. C. *et al.* In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nature Neuroscience* **13**, 897–U149 (2010).

[127] Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).

[128] Hinz, F. I., Dieterich, D. C., Tirrell, D. A. & Schuman, E. M. Noncanonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *Acs Chemical Neuroscience* **3**, 40–49 (2012).

[129] Kzhyshkowska, J. & Krusella, L. Cross-talk between endocytic clearance and secretion in macrophages. *Immunobiology* **214**, 576–593 (2009).

[130] Nickel, W. Unconventional secretory routes: Direct protein export across the plasma membrane of mammalian cells. *Traffic* **6**, 607–614 (2005).

[131] Schatz, G. & Dobberstein, B. Common principles of protein translocation across membranes. *Science* **271**, 1519–1526 (1996).

[132] Lee, M. C. S., Miller, E. A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the er and golgi. *Annual Review of Cell and Developmental Biology* **20**, 87–123 (2004).

[133] Glick, B. S. & Malhotra, V. The curious status of the golgi apparatus. *Cell* **95**, 883–889 (1998).

[134] Rabouille, C. & Klumperman, J. The maturing role of copi vesicles in intra-golgi transport. *Nature Reviews Molecular Cell Biology* **6**, 812–817 (2005).

[135] Keller, P. & Simons, K. Post-golgi biosynthetic trafficking. *Journal of Cell Science* **110**, 3001–3009 (1997).

[136] Kelly, R. B. Pathways of protein secretion in eukaryotes. *Science* **230**, 25–32 (1985).

[137] Mellman, I. & Warren, G. The road taken: Past and future foundations of membrane traffic. *Cell* **100**, 99–112 (2000).

[138] Orci, L. *et al.* Brefeldin-a, a drug that blocks secretion, prevents the assembly of non-clathrin-coated buds on golgi cisternae. *Cell* **64**, 1183–1195 (1991).

[139] Baldwin, T. A. & Ostergaard, H. L. The protein-tyrosine phosphatase cd45 reaches the cell surface via golgi-dependent and -independent pathways. *Journal of Biological Chemistry* **277**, 50333–50340 (2002).

[140] Toyokawa, K., Carling, S. J. & Ott, T. L. Cellular localization and function of the antiviral protein, ovine mx1 (omx1): I. ovine mx1 is secreted by endometrial epithelial cells via an 'unconventional' secretory pathway. *American Journal of Reproductive Immunology* **57**, 13–22 (2007).

[141] Wang, X. D. *et al.* Copii-dependent export of cystic fibrosis transmembrane conductance regulator from the er uses a di-acidic exit code. *Journal of Cell Biology* **167**, 65–74 (2004).

[142] Nickel, W. & Rabouille, C. Mechanisms of regulated unconventional protein secretion. *Nature Reviews Molecular Cell Biology* **10**, 148–155 (2009).

[143] Hughes, R. C. Secretion of the galectin family of mammalian carbohydrate-binding proteins. *Biochimica Et Biophysica Acta-general Subjects* **1473**, 172–185 (1999).

[144] Flieger, O. *et al.* Regulated secretion of macrophage migration inhibitory factor is mediated by a non-classical pathway involving an abc transporter. *Febs Letters* **551**, 78–86 (2003).

[145] Rubartelli, A., Cozzolino, F., Talio, M. & Sitia, R. A novel secretory pathway for interleukin-1-beta, a protein lacking a signal sequence. *Embo Journal* **9**, 1503–1510 (1990).

[146] Frye, B. C. *et al.* Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen. *Embo Reports* **10**, 783–789 (2009).

[147] Nickel, W. The mystery of nonclassical protein secretion - a current view on cargo proteins and potential export routes. *European Journal of Biochemistry* **270**, 2109–2119 (2003).

[148] Giuliani, F., Grieve, A. & Rabouille, C. Unconventional secretion: a stress on grasp. *Current Opinion In Cell Biology* **23**, 498–504 (2011).

[149] Prudovsky, I. *et al.* The intracellular translocation of the components of the fibroblast growth factor 1 release complex precedes their assembly prior to export. *Journal of Cell Biology* **158**, 201–208 (2002).

[150] Rayne, F. *et al.* Phosphatidylinositol-(4,5)-bisphosphate enables efficient secretion of hiv-1 tat by infected t-cells. *Embo Journal* **29**, 1348–1362 (2010).

[151] Deora, A. B., Kreitzer, G., Jacovina, A. T. & Hajjar, K. A. An annexin 2 phosphorylation switch mediates p11-dependent translocation of annexin 2 to the cell surface. *Journal of Biological Chemistry* **279**, 43411–43418 (2004).

[152] Hamon, Y. *et al.* Interleukin-1 beta secretion is impaired by inhibitors of the atp binding cassette transporter, abc1. *Blood* **90**, 2911–2915 (1997).

[153] Mambula, S. S. & Calderwood, S. K. Heat shock protein 70 is secreted from tumor cells by a nonclassical pathway involving lysosomal endosomes. *Journal of Immunology* **177**, 7849–7857 (2006).

[154] Steringer, J. *et al.* Phosphatidylinositol 4,5-bisphosphate (pi(4,5)p2)-dependent oligomerization of fibroblast growth factor 2 (fgf2) triggers the formation of a lipidic membrane pore implicated in unconventional secretion. *The Journal of biological chemistry* **287**, 27659–69 (2012).

[155] Blott, E. J. & Griffiths, G. M. Secretory lysosomes. *Nature Reviews Molecular Cell Biology* **3**, 122–131 (2002).

[156] Andrei, C. *et al.* The secretory route of the leaderless protein interleukin 1 beta involves exocytosis of endolysosome-related vesicles. *Molecular Biology of the Cell* **10**, 1463–1475 (1999).

[157] Andrei, C. *et al.* Phospholipases c and a(2) control lysosome-mediated il-1 beta secretion: Implications for inflammatory processes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9745–9750 (2004).

[158] Mathivanan, S., Fahner, C. J., Reid, G. E. & Simpson, R. J. Exocarta 2012: database of exosomal proteins, rna and lipids. *Nucleic Acids Research* **40**, D1241–D1244 (2012).

[159] Duran, J. M., Anjard, C., Stefan, C., Loomis, W. F. & Malhotra, V. Unconventional secretion of acb1 is mediated by autophagosomes. *Journal of Cell Biology* **188**, 527–536 (2010).

[160] Manjithaya, R., Anjard, C., Loomis, W. F. & Subramani, S. Unconventional secretion of pichia pastoris acb1 is dependent on grasp protein, peroxisomal functions, and autophagosome formation. *Journal of Cell Biology* **188**, 537–546 (2010).

[161] Zemskov, E. A., Mikhailenko, I., Hsia, R. C., Zaritskaya, L. & Belkin, A. M. Unconventional secretion of tissue transglutaminase involves phospholipid-dependent delivery into recycling endosomes. *Plos One* **6**, e19414 (2011).

[162] Anderson, N. L. & Anderson, N. G. The human plasma proteome - history, character, and diagnostic prospects. *Molecular & Cellular Proteomics* **1**, 845–867 (2002).

[163] Ayache, S., Panelli, M., Marincola, F. M. & Stroncek, D. F. Effects of storage time and exogenous protease inhibitors on plasma protein levels. *American Journal of Clinical Pathology* **126**, 174–184 (2006).

[164] Mustafa, S. A., Hoheisel, J. D. & Alhamdani, M. S. S. Secretome profiling with antibody microarrays. *Molecular Biosystems* **7**, 1795–1801 (2011).

[165] Schroder, C. *et al.* Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Molecular & Cellular Proteomics* **9**, 1271–1280 (2010).

[166] Wu, W. W. *et al.* Antibody array analysis with label-based detection and resolution of protein size. *Molecular & Cellular Proteomics* **8**, 245–257 (2009).

[167] Bereczki, E. *et al.* Overexpression of biglycan in the heart of transgenic mice: An antibody microarray study. *Journal of Proteome Research* **6**, 854–861 (2007).

[168] Raimondo, F., Morosi, L., Chinello, C., Magni, F. & Pitto, M. Advances in membranous vesicle and exosome proteomics improving biological understanding and biomarker discovery. *Proteomics* **11**, 709–720 (2011).

[169] Zwickl, H. *et al.* A novel technique to specifically analyze the secretome of cells and tissues. *Electrophoresis* **26**, 2779–2785 (2005).

[170] Henningsen, J., Pedersen, B. K. & Kratchmarova, I. Quantitative analysis of the secretion of the mcp family of chemokines by muscle cells. *Molecular Biosystems* **7**, 311–321 (2011).

[171] Colzani, M. *et al.* Metabolic labeling and protein linearization technology allow the study of proteins secreted by cultured cells in serum-containing media. *Journal of Proteome Research* **8**, 4779–4788 (2009).

[172] Pellitteri-Hahn, M. C. *et al.* Improved mass spectrometric proteomic profiling of the secretome of rat vascular endothelial cells. *Journal of Proteome Research* **5**, 2861–2864 (2006).

[173] Srisomsap, C. *et al.* Proteomic studies of cholangiocarcinoma and hepatocellular carcinoma cell secretomes. *Journal of Biomedicine and Biotechnology* 437143 (2010).

[174] Higa, L. M. *et al.* Secretome of hepg2 cells infected with dengue virus: Implications for pathogenesis. *Biochimica Et Biophysica Acta-proteins and Proteomics* **1784**, 1607–1616 (2008).

[175] Kuhn, P. *et al.* Secretome protein enrichment identifies physiological bace1 protease substrates in neurons. *EMBO Journal* **31**, 3157–3168 (2012).

[176] Vonheijne, G. The signal peptide. *Journal of Membrane Biology* **115**, 195–201 (1990).

[177] Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using targetp, signalp and related tools. *Nature Protocols* **2**, 953–971 (2007).

[178] Nielsen, H., Engelbrecht, J., Brunak, S. & vonHeijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1–6 (1997).

[179] Bendtsen, J. D., Jensen, L. J., Blom, N., von Heijne, G. & Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design & Selection* **17**, 349–356 (2004).

[180] Shalhoub, J., Falck-Hansen, M. A., Davies, A. H. & Monaco, C. Innate immunity and monocyte-macrophage activation in atherosclerosis. *Journal of Inflammation-london* **8**, 9 (2011).

[181] Vazquez-Torres, A. *et al.* Extraintestinal dissemination of salmonella by cd18-expressing phagocytes. *Nature* **401**, 804–808 (1999).

[182] Gordon, S. The macrophage: Past, present and future. *European Journal of Immunology* **37**, S9–S17 (2007).

[183] Martinez, F. O., Gordon, S., Locati, M. & Mantovani, A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: New molecules and patterns of gene expression. *Journal of Immunology* **177**, 7303–7311 (2006).

[184] Gordon, S. Alternative activation of macrophages. *Nature Reviews Immunology* **3**, 23–35 (2003).

[185] Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nature Reviews Immunology* **8**, 958–969 (2008).

[186] Lawrence, T. & Natoli, G. Transcriptional regulation of macrophage polarization: enabling diversity with identity. *Nature Reviews Immunology* **11**, 750–761 (2011).

[187] Kadl, A. *et al.* Identification of a novel macrophage phenotype that develops in response to atherogenic phospholipids via nrf2. *Circulation Research* **107**, 737–U155 (2010).

[188] Condeelis, J. & Pollard, J. W. Macrophages: Obligate partners for tumor cell migration, invasion, and metastasis. *Cell* **124**, 263–266 (2006).

[189] Biswas, S. K. *et al.* A distinct and unique transcriptional program expressed by tumor-associated macrophages (defective nf-kappa b and enhanced irf-3/stat1 activation). *Blood* **107**, 2112–2122 (2006).

[190] Lynn, W. A. & Cohen, J. Adjunctive therapy for septic shock - a review of experimental approaches. *Clinical Infectious Diseases* **20**, 143–158 (1995).

[191] Foster, S. L., Hargreaves, D. C. & Medzhitov, R. Gene-specific control of inflammation by tlr-induced chromatin modifications. *Nature* **447**, 972–U4 (2007).

[192] Biswas, S. K. & Lopez-Collazo, E. Endotoxin tolerance: new mechanisms, molecules and clinical significance. *Trends In Immunology* **30**, 475–487 (2009).

[193] Mathison, J., Tobias, P., Wolfson, E. & Ulevitch, R. Regulatory mechanisms of host responsiveness to endotoxin (lipopolysaccharide). *Pathobiology* **59**, 185–188 (1991).

[194] Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Reviews Immunology* **4**, 499–511 (2004).

[195] Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).

[196] Kawai, T. & Akira, S. Signaling to nf-kappa b by toll-like receptors. *Trends In Molecular Medicine* **13**, 460–469 (2007).

[197] Pugin, J. *et al.* Lipopolysaccharide activation of human endothelial and epithelial-cells is mediated by lipopolysaccharide-binding protein and soluble cd14. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2744–2748 (1993).

[198] Underhill, D. M. & Ozinsky, A. Toll-like receptors: key mediators of microbe detection. *Current Opinion In Immunology* **14**, 103–110 (2002).

[199] O'Neill, L. A. J. & Bowie, A. G. The family of five: Tir-domain-containing adaptors in toll-like receptor signalling. *Nature Reviews Immunology* **7**, 353–364 (2007).

[200] Mogensen, T. H. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clinical Microbiology Reviews* **22**, 240–+ (2009).

[201] Kagan, J. C. *et al.* Tram couples endocytosis of toll-like receptor 4 to the induction of interferon-beta. *Nature Immunology* **9**, 361–368 (2008).

[202] Burns, K. *et al.* Myd88, an adapter protein involved in interleukin-1 signaling. *Journal of Biological Chemistry* **273**, 12203–12209 (1998).

[203] Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. D. Myd88: An adapter that recruits irak to the il-1 receptor complex. *Immunity* **7**, 837–847 (1997).

[204] Li, S. Y., Strelow, A., Fontana, E. J. & Wesche, H. Irak-4: A novel member of the irak family with the properties of an irak-kinase. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5567–5572 (2002).

[205] Wang, C. *et al.* Tak1 is a ubiquitin-dependent kinase of mkk and ikk. *Nature* **412**, 346–351 (2001).

[206] Chang, L. F. & Karin, M. Mammalian map kinase signalling cascades. *Nature* **410**, 37–40 (2001).

[207] Yamamoto, M. *et al.* Tram is specifically involved in the toll-like receptor 4-mediated myd88-independent signaling pathway. *Nature Immunology* **4**, 1144–1150 (2003).

[208] Guo, B. C. & Cheng, G. H. Modulation of the interferon antiviral response by the tbk1/ikki adaptor protein tank. *Journal of Biological Chemistry* **282**, 11817–11826 (2007).

[209] Hacker, H. *et al.* Specificity in toll-like receptor signalling through distinct effector functions of traf3 and traf6. *Nature* **439**, 204–207 (2006).

[210] Oganesyan, G. *et al.* Critical role of traf3 in the toll-like receptor-dependent and -independent antiviral response. *Nature* **439**, 208–211 (2006).

[211] Sato, S. *et al.* Toll/il-1 receptor domain-containing adaptor inducing ifn-beta (trif) associates with tnf receptor-associated factor 6 and tank-binding kinase 1, and activates two distinct transcription factors, nf-kappa b and ifn-regulatory factor-3, in the toll-like receptor signaling. *Journal of Immunology* **171**, 4304–4310 (2003).

[212] Fitzgerald, K. A. *et al.* Ikk epsilon and tbk1 are essential components of the irf3 signaling pathway. *Nature Immunology* **4**, 491–496 (2003).

[213] Sharma, S. *et al.* Triggering the interferon antiviral response through an ikk-related pathway. *Science* **300**, 1148–1151 (2003).

[214] Thanos, D. & Maniatis, T. Virus induction of human ifn beta gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).

[215] Cusson-Hermance, N., Khurana, S., Lee, T. H., Fitzgerald, K. A. & Kelliher, M. A. Rip1 mediates the trif-dependent toll-like receptor 3- and 4-induced nf-kappa b activation but does not contribute to interferon regulatory factor 3 activation. *Journal of Biological Chemistry* **280**, 36560–36566 (2005).

[216] Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature Immunology* **2**, 675–680 (2001).

[217] Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annual Review of Immunology* **21**, 335–376 (2003).

[218] Ramsey, S. A. *et al.* Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *Plos Computational Biology* **4**, e1000021 (2008).

[219] Zhu, X. C. *et al.* Dual ligand stimulation of raw 264.7 cells uncovers feedback mechanisms that regulate tlr-mediated gene expression. *Journal of Immunology* **177**, 4299–4310 (2006).

[220] Litvak, V. *et al.* Function of c/ebp delta in a regulatory circuit that discriminates between transient and persistent tlr4-induced signals. *Nature Immunology* **10**, 437–443 (2009).

[221] Swearingen, K. E., Loomis, W. P., Zheng, M., Cookson, B. T. & Dovichi, N. J. Proteomic profiling of lipopolysaccharide-activated macrophages by isotope coded affinity tagging. *Journal of Proteome Research* **9**, 2412–2421 (2010).

[222] Du, R. Y. *et al.* Subcellular quantitative proteomics reveals multiple pathway cross-talk that coordinates specific signaling and transcriptional regulation for the early host response to lps. *Journal of Proteome Research* **9**, 1805–1821 (2010).

[223] Bhatt, D. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular rna fractions. *Cell* **150**, 279–90 (2012).

[224] Weintz, G. *et al.* The phosphoproteome of toll-like receptor-activated macrophages. *Molecular Systems Biology* **6**, 371 (2010).

[225] Sharma, K. *et al.* Quantitative analysis of kinase-proximal signaling in lipopolysaccharide-induced innate immune response. *Journal of Proteome Research* **9**, 2539–2549 (2010).

[226] Patel, P. C., Fisher, K. H., Yang, E. C. C., Deane, C. M. & Harrison, R. E. Proteomic analysis of microtubule-associated proteins during macrophage activation. *Molecular & Cellular Proteomics* **8**, 2500–2514 (2009).

[227] Dhungana, S., Merrick, B. A., Tomer, K. B. & Fessler, M. B. Quantitative proteomics analysis of macrophage rafts reveals compartmentalized activation of the proteasome and of proteasome-mediated erk activation in response to lipopolysaccharide. *Molecular & Cellular Proteomics* **8**, 201–213 (2009).

[228] Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R. & Mohammed, S. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **8**, 4624–4632 (2008).

[229] Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359–U60 (2009).

[230] Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).

[231] Aasland, R., Gibson, T. J. & Stewart, A. F. The phd finger - implications for chromatin-mediated transcriptional regulation. *Trends In Biochemical Sciences* **20**, 56–59 (1995).

[232] Bouhouche, N., Syvanen, M. & Kado, C. I. The origin of prokaryotic c2h2 zinc finger regulators. *Trends In Microbiology* **8**, 77–81 (2000).

[233] Tomancak, P. *et al.* Global analysis of patterns of gene expression during drosophila embryogenesis. *Genome Biology* **8**, R145 (2007).

[234] Qian, L., Liu, J. D. & Bodmer, R. Neuromancer tbx20-related genes (h15/midline) promote cell fate specification and morphogenesis of the drosophila heart. *Developmental Biology* **279**, 509–524 (2005).

[235] Zaffran, S., Kuchler, A., Lee, H. H. & Frasch, M. biniou (foxf), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in drosophila. *Genes & Development* **15**, 2900–2915 (2001).

[236] Lee, H. H. & Frascht, M. Nuclear integration of positive dpp signals, antagonistic wg inputs and mesodermal competence factors during drosophila visceral mesoderm induction. *Development* **132**, 1429–1442 (2005).

[237] Grienenberger, A. *et al.* Tgf beta signaling acts on a hox response element to confer specificity and diversity to hox protein function. *Development* **130**, 5445–5455 (2003).

[238] Hemavathy, K., Hu, X. D., Ashraf, S. I., Small, S. J. & Ip, Y. T. The repressor function of snail is required for drosophila gastrulation and is not replaceable by escargot or worniu. *Developmental Biology* **269**, 411–420 (2004).

[239] Fuse, N., Hirose, S. & Hayashi, S. Diploidy of drosophila imaginal cells is maintained by a transcriptional repressor encoded by escargot. *Genes & Development* **8**, 2270–2281 (1994).

[240] Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376–386 (2002).

[241] Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (silac). *Nature Protocols* **1**, 2650–2660 (2006).

[242] Brunner, E. *et al.* A high-quality catalog of the drosophila melanogaster proteome. *Nature Biotechnology* **25**, 576–583 (2007).

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

[244] Jain, E. *et al.* Infrastructure for the life sciences: design and implementation of the uniprot website. *Bmc Bioinformatics* **10**, 136 (2009).

[245] Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786 (2011).

[246] Nikolsky, Y., Ekins, S., Nikolskaya, T. & Bugrim, A. A novel method for generation of signature networks as biomarkers from complex high throughput data. *Toxicology Letters* **158**, 20–29 (2005).

[247] Sardana, G., Jung, K., Stephan, C. & Diamandis, E. P. Proteomic analysis of conditioned media from the pc3, lncap, and 22rv1 prostate cancer cell lines: Discovery and validation of candidate prostate cancer biomarkers. *Journal of Proteome Research* **7**, 3329–3338 (2008).

[248] Caccia, D. *et al.* Secretome compartment is a valuable source of biomarkers for cancer-relevant pathways. *Journal of Proteome Research* **10**, 4196–4207 (2011).

[249] Patarroyo, M., Tryggvason, K. & Virtanen, I. Laminin isoforms in tumor invasion, angiogenesis and metastasis. *Seminars In Cancer Biology* **12**, 197–207 (2002).

[250] Duffy, M. J. *et al.* The adams family of proteases: new biomarkers and therapeutic targets for cancer? *Clin Proteomics* **8**, 1 (2011).

[251] Perumal, S., Antipova, O. & Orgel, J. P. R. O. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2824–2829 (2008).

[252] Levental, K. R. *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891–906 (2009).

[253] Paszek, M. J. *et al.* Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241–254 (2005).

[254] Lo, C. M., Wang, H. B., Dembo, M. & Wang, Y. L. Cell movement is guided by the rigidity of the substrate. *Biophysical Journal* **79**, 144–152 (2000).

[255] Hippo, Y. *et al.* Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Research* **62**, 233–240 (2002).

[256] Yang, S. *et al.* Molecular basis of the differences between normal and tumor tissues of gastric cancer. *Biochimica Et Biophysica Acta-molecular Basis of Disease* **1772**, 1033–1040 (2007).

[257] Guess, C. M. & Quaranta, V. Defining the role of laminin-332 in carcinoma. *Matrix Biology* **28**, 445–455 (2009).

[258] Katenkamp, K. *et al.* mrna expression and protein distribution of the unspliced tenascin-c isoform in prostatic adenocarcinoma. *Journal of Pathology* **203**, 771–779 (2004).

[259] Couchman, J. R. Transmembrane signaling proteoglycans. *Annual Review of Cell and Developmental Biology* **26**, 89–114 (2010).

[260] Ledezma, R. *et al.* Altered expression patterns of syndecan-1 and -2 predict biochemical recurrence in prostate cancer. *Asian Journal of Andrology* **13**, 476–480 (2011).

[261] Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: Plasticity and reciprocity. *Cell* **147**, 992–1009 (2011).

[262] Eyler, C. E. & Telen, M. J. The lutheran glycoprotein: a multifunctional adhesion receptor. *Transfusion* **46**, 668–677 (2006).

[263] Perbal, B. Ccn proteins: multifunctional signalling regulators. *Lancet* **363**, 62–64 (2004).

[264] Holbourn, K. P., Acharya, K. R. & Perbal, B. The ccn family of proteins: structure-function relationships. *Trends In Biochemical Sciences* **33**, 461–473 (2008).

[265] Sakai, I., Miyake, H., Terakawa, T. & Fujisawa, M. Inhibition of tumor growth and sensitization to chemotherapy by rna interference targeting interleukin-6 in the androgen-independent human prostate cancer pc3 model. *Cancer Science* **102**, 769–775 (2011).

[266] Putoczki, T. & Ernst, M. More than a sidekick: the il-6 family cytokine il-11 links inflammation to cancer. *Journal of Leukocyte Biology* **88**, 1109–1117 (2010).

[267] Mousavi, S. A., Berge, K. E. & Leren, T. P. The unique role of proprotein convertase subtilisin/kexin 9 in cholesterol homeostasis. *Journal of Internal Medicine* **266**, 507–519 (2009).

[268] Solomon, K. R. & Freeman, M. R. The complex interplay between cholesterol and prostate malignancy. *Urologic Clinics of North America* **38**, 243–+ (2011).

[269] Pan, C. P., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions rid c-2901-2008. *Molecular & Cellular Proteomics* **8**, 443–450 (2009).

[270] Slany, A. *et al.* Cell characterization by proteome profiling applied to primary hepatocytes and hepatocyte cell lines hep-g2 and hep-3b. *Journal of Proteome Research* **9**, 6–21 (2010).

[271] Stow, J. L., Low, P. C., Offenhauser, C. & Sangermani, D. Cytokine secretion in macrophages and other cells: Pathways and mediators. *Immunobiology* **214**, 601–612 (2009).

[272] Bailey, M. J. *et al.* Extracellular proteomes of m-csf (csf-1) and gm-csf-dependent macrophages. *Immunology and Cell Biology* **89**, 283–293 (2011).

[273] Burghoff, S. & Schrader, J. Secretome of human endothelial cells under shear stress. *Journal of Proteome Research* **10**, 1160–1169 (2011).

[274] Lietzen, N. *et al.* Quantitative subcellular proteome and secretome profiling of influenza a virus-infected human primary macrophages. *Plos Pathogens* **7**, e1001340 (2011).

[275] Li, M. O., Wan, Y. Y., Sanjabi, S., Robertson, A. K. L. & Flavell, R. A. Transforming growth factor-beta regulation of immune responses. *Annual Review of Immunology* **24**, 99–146 (2006).

[276] Weis, N., Weigert, A., von Knethen, A. & Brune, B. Heme oxygenase-1 contributes to an alternative macrophage activation profile induced by apoptotic cell supernatants. *Molecular Biology of the Cell* **20**, 1280–1288 (2009).

[277] Yan, X. Q., Brady, G. & Iscove, N. N. Platelet-derived growth-factor (pdgf) activates primitive hematopoietic precursors (pre-cfcmulti) by up-regulating il-1 in pdgf receptor-expressing macrophages. *Journal of Immunology* **150**, 2440–2448 (1993).

[278] Manderson, A. P., Kay, J. G., Hammond, L. A., Brown, D. L. & Stow, J. L. Subcompartments of the macrophage recycling endosome direct the differential secretion of il-6 and tnf alpha. *Journal of Cell Biology* **178**, 57–69 (2007).

[279] Jones, S. A. Directing transition from innate to acquired immunity defining a role for il-6. *Journal of Immunology* **175**, 3463–3468 (2005).

[280] Wittek, E., N. Majewska. Cystatin c–modulator of immune processes. *Przegl Lek* **67**, 484 (2010).

[281] Hamilton, G., Colbert, J. D., Schuettelkopf, A. W. & Watts, C. Cystatin f is a cathepsin c-directed protease inhibitor regulated by proteolysis. *Embo Journal* **27**, 499–508 (2008).

[282] Lucas, N., Najy, A. J. & Day, M. L. The therapeutic potential of adam15. *Current Pharmaceutical Design* **15**, 2311–2318 (2009).

[283] Saftig, P. & Reiss, K. The "a disintegrin and metalloproteases" adam10 and adam17: Novel drug targets with therapeutic potential? *European Journal of Cell Biology* **90**, 527–535 (2011).

[284] Bartsch, J. W. *et al.* Tumor necrosis factor-alpha (tnf-alpha) regulates shedding of tnf-alpha receptor 1 by the metalloprotease-disintegrin adam8: Evidence for a protease-regulated feedback loop in neuroprotection. *Journal of Neuroscience* **30**, 12210–12218 (2010).

[285] Ripoll, V. M., Irvine, K. M., Ravasi, T., Sweet, M. J. & Hume, D. A. Gpnmb is induced in macrophages by ifn-gamma and lipopolysaccharide and acts as a feedback regulator of proinflammatory responses. *Journal of Immunology* **178**, 6557–6566 (2007).

[286] Yang, L. T. *et al.* Fringe glycosyltransferases differentially modulate notch1 proteolysis induced by delta1 and jagged1. *Molecular Biology of the Cell* **16**, 927–942 (2005).

[287] Tsao, P. N. *et al.* Lipopolysaccharide-induced notch signaling activation through jnk-dependent pathway regulates inflammatory response. *Journal of Biomedical Science* **18**, 56 (2011).

[288] Mi, H. Y. *et al.* Panther version 7: improved phylogenetic trees, orthologs and collaboration with the gene ontology consortium. *Nucleic Acids Research* **38**, D204–D210 (2010).

[289] Thomas, P. D. *et al.* Panther: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research* **31**, 334–341 (2003).

[290] Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).

[291] Griffin, T. J. *et al.* Complementary profiling of gene expression at the transcriptome and proteome levels in saccharomyces cerevisiae. *Molecular & Cellular Proteomics* **1**, 323–333 (2002).

[292] Kislinger, T. *et al.* Global survey of organ and organelle protein expression in mouse: Combined proteomic and transcriptomic profiling. *Cell* **125**, 173–186 (2006).

[293] Arlt, A. & Schafer, H. Role of the immediate early response 3 (ier3) gene in cellular stress response, inflammation and tumorigenesis. *European Journal of Cell Biology* **90**, 545–552 (2011).

[294] Muta, T., Yamazaki, S., Eto, A., Motoyama, M. & Takeshige, K. I kappa b-zeta, a new anti-inflammatory nuclear protein induced by lipopolysaccharide, is a negative regulator for nuclear factor-kappa b. *Journal of Endotoxin Research* **9**, 187–191 (2003).

[295] Motoyama, M., Yamazaki, S., Eto-Kimura, A., Takeshige, K. & Muta, T. Positive and negative regulation of nuclear factor-kappa b-mediated transcription by i kappa b-xi, an inducible nuclear protein. *Journal of Biological Chemistry* **280**, 7444–7451 (2005).

[296] Deleault, K. M., Skinner, S. J. & Brooks, S. A. Tristetraprolin regulates tnf tnf-alpha-mrna stability via a proteasome dependent mechanism involving the combined action of the erk and p38 pathways. *Molecular Immunology* **45**, 13–24 (2008).

[297] Bachmann, M. & Moroy, T. The serine/threonine kinase pim-1. *International Journal of Biochemistry & Cell Biology* **37**, 726–730 (2005).

[298] Harada, Y., Nagao, S., Nakamura, M., Okada, F. & Tanigawa, Y. Effect of lipopolysaccharide on thymidine salvage as related to macrophage activation. *Immunology* **84**, 247–253 (1995).

[299] Chang, C. Y., Tucci, M. & Baker, R. C. Lipopolysaccharide-stimulated nitric oxide production and inhibition of cell proliferation is antagonized by ethanol in a clonal macrophage cell line. *Alcohol* **20**, 37–43 (2000).

[300] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**, 450–461 (2007).

[301] Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology* **323**, 785–793 (2002).

[302] Sassonecorsi, P., Sisson, J. C. & Verma, I. M. Transcriptional auto-regulation of the proto-oncogene fos. *Nature* **334**, 314–319 (1988).

[303] Horner, T. J., Lai, W. S., Stumpo, D. J. & Blackshear, P. J. Stimulation of polo-like kinase 3 mrna decay by tristetraprolin. *Molecular and Cellular Biology* **29**, 1999–2010 (2009).

[304] Kratochvill, F. *et al.* Tristetraprolin-driven regulatory circuit controls quality and timing of mrna decay in inflammation. *Molecular Systems Biology* **7**, 560 (2011).

[305] Hatakeyama, S. *et al.* Ubiquitin-dependent degradation of i kappa b alpha is mediated by a ubiquitin ligase skp1/cul 1/f-box protein fwd1. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 3859–3863 (1999).

[306] Busino, L. *et al.* Degradation of cdc25a by beta-trcp during s phase and in response to dna damage. *Nature* **426**, 87–91 (2003).

[307] Komander, D. The emerging complexity of protein ubiquitination. *Biochemical Society Transactions* **37**, 937–953 (2009).

[308] Liu, S. Q. & Chen, Z. J. J. Expanding role of ubiquitination in nf-kappa b signaling. *Cell Research* **21**, 6–21 (2011).

[309] Stanley, A. C. & Lacy, P. Pathways for cytokine secretion. *Physiology* **25**, 218–229 (2010).

[310] Valadi, H. *et al.* Exosome-mediated transfer of mrnas and micrornas is a novel mechanism of genetic exchange between cells. *Nature Cell Biology* **9**, 654–U72 (2007).

[311] Elhabazi, A., Delaire, S., Bensussan, A., Boumsell, L. & Bismuth, G. Biological activity of soluble cd100. 1. the extracellular region of cd100 is released from the surface of t lymphocytes by regulated proteolysis. *Journal of Immunology* **166**, 4341–4347 (2001).

[312] Matthews, V. *et al.* Cellular cholesterol depletion triggers shedding of the human interleukin-6 receptor by adam10 and adam17 (tace). *Journal of Biological Chemistry* **278**, 38829–38839 (2003).

[313] LUST, J. A. *et al.* Isolation of an messenger-rna encoding a soluble form of the human interleukin-6 receptor. *Cytokine* **4**, 96–100 (1992).

[314] Rose-John, S., Scheller, J., Elson, G. & Jones, S. A. Interleukin-6 biology is coordinated by membrane-bound and soluble receptors: role in inflammation and cancer. *Journal of Leukocyte Biology* **80**, 227–236 (2006).

[315] Kim, J. *et al.* Activity-dependent alpha-cleavage of nectin-1 is mediated by a disintegrin and metalloprotease 10 (adam10). *Journal of Biological Chemistry* **285**, 22917–22924 (2010).

[316] Artigiani, S. *et al.* Functional regulation of semaphorin receptors by proprotein convertases. *Journal of Biological Chemistry* **278**, 10094–10101 (2003).

[317] Gilchrist, M. *et al.* Systems biology approaches identify atf3 as a negative regulator of toll-like receptor 4. *Nature* **441**, 173–178 (2006).

[318] Xaus, J. *et al.* Lps induces apoptosis in macrophages mostly through the autocrine production of tnf-alpha. *Blood* **95**, 3823–3831 (2000).

[319] Li, J. *et al.* Synergistic function of e2f7 and e2f8 is essential for cell survival and embryonic development. *Developmental Cell* **14**, 62–75 (2008).

[320] Holmberg, C., Helin, K., Sehested, M. & Karlstrom, O. E2f-1-induced p53-independent apoptosis in transgenic mice. *Oncogene* **17**, 143–155 (1998).

[321] Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on toll-like receptors. *Nature Immunology* **11**, 373–384 (2010).

[322] Begg, M. J., Sturrock, E. D. & van der Westhuyzen, D. R. Soluble ldl-r are formed by cell surface cleavage in response to phorbol esters. *European Journal of Biochemistry* **271**, 524–533 (2004).

[323] Melis, M. *et al.* Biologically active intercellular adhesion molecule-1 is shed as dimers by a regulated mechanism in the inflamed pleural space. *American Journal of Respiratory and Critical Care Medicine* **167**, 1131–1138 (2003).

[324] Hayashida, K., Parks, W. C. & Park, P. W. Syndecan-1 shedding facilitates the resolution of neutrophilic inflammation by removing sequestered cxc chemokines. *Blood* **114**, 3033–3043 (2009).

[325] Flegel, W. A., Wolpl, A., Mannel, D. N. & Northoff, H. Inhibition of endotoxin-induced activation of human-monocytes by human lipoproteins. *Infection and Immunity* **57**, 2237–2245 (1989).

[326] Shi, Z. C. *et al.* Transcriptional regulation of the novel toll-like receptor tlr13. *Journal of Biological Chemistry* **284**, 20540–20547 (2009).

[327] Takegahara, N., Kumanogoh, A. & Kikutani, H. Semaphorins: a new class of immunoregulatory molecules. *Philosophical Transactions of the Royal Society B-biological Sciences* **360**, 1673–1680 (2005).

[328] Zhang, Y. J. *et al.* Nuclear translocation of the n-terminal prodomain of interleukin-16. *Journal of Biological Chemistry* **276**, 1299–1303 (2001).

[329] Center, D. M. & Cruikshank, W. Modulation of lymphocyte migration by human lymphokines .1. identification and characterization of chemoattractant activity for lymphocytes from mitogen-stimulated mononuclear-cells. *Journal of Immunology* **128**, 2563–2568 (1982).

[330] Cruikshank, W. W., Berman, J. S., Theodore, A. C., Bernardo, J. & Center, D. M. Lymphokine activation of t4+ lymphocytes and monocytes. *Journal of Immunology* **138**, 3817–3823 (1987).

[331] Rand, T. H., Cruikshank, W. W., Center, D. M. & Weller, P. F. Cd4-mediated stimulation of human eosinophils - lymphocyte chemoattractant factor and other cd4-binding ligands elicit eosinophil migration. *Journal of Experimental Medicine* **173**, 1521–1528 (1991).

[332] Kaser, A. *et al.* A role for il-16 in the cross-talk between dendritic cells and t cells. *Journal of Immunology* **163**, 3232–3238 (1999).

[333] Van Drenth, C. *et al.* Desensitization of cxc chemokine receptor 4, mediated by il-16/cd4, is independent of p56(lck) enzymatic activity. *Journal of Immunology* **165**, 6356–6363 (2000).

[334] Weiss, A. Kinases and phosphatases of the immune system. *Immunological Reviews* **228**, 5–8 (2009).

[335] Strebhardt, K. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. *Nature Reviews Drug Discovery* **9**, 643–U24 (2010).

[336] Aoyama, C., Liao, H. N. & Ishidate, K. Structure and function of choline kinase isoforms in mammalian cells. *Progress In Lipid Research* **43**, 266–281 (2004).

[337] Clark, K. *et al.* Novel cross-talk within the ikk family controls innate immunity. *Biochemical Journal* **434**, 93–104 (2011).

[338] Huang, J. *et al.* Identification of a novel serine/threonine kinase that inhibits tnf-induced nf-kappa b activation and p53-induced transcription. *Biochemical and Biophysical Research Communications* **309**, 774–778 (2003).

[339] Williams, L., Jarai, G., Smith, A. & Finan, P. Il-10 expression profiling in human monocytes. *Journal of Leukocyte Biology* **72**, 800–809 (2002).

[340] Haan, C., Kreis, S., Margue, C. & Behrmann, I. Jaks and cytokine receptors - an intimate relationship. *Biochemical Pharmacology* **72**, 1538–1546 (2006).

[341] Enomoto, A. *et al.* Negative regulation of mekk1/2 signaling by serine-threonine kinase 38 (stk38). *Oncogene* **27**, 1930–1938 (2008).

[342] Patterson, K. I., Brummer, T., O'Brien, P. M. & Daly, R. J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochemical Journal* **418**, 475–489 (2009).

[343] Klingmuller, U. *et al.* Primary mouse hepatocytes for systems biology approaches: a standardized in vitro system for modelling of signal transduction pathways. *Iee Proceedings Systems Biology* **153**, 433–447 (2006).

[344] Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and lc/ms sample pretreatment in proteomics. *Analytical Chemistry* **75**, 663–670 (2003).

[345] Wisniewski, J. R., Zougman, A. & Mann, M. Combination of fasp and stagetip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *Journal of Proteome Research* **8**, 5674–5678 (2009).

[346] Cox, J. *et al.* Andromeda: A peptide search engine integrated into the maxquant environment. *Journal of Proteome Research* **10**, 1794–1805 (2011).

[347] Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80 (2004).

[348] Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504 (2003).

Appendix A

List of Publications

Ahrens, M., Schuschke, K., Redmer, S., Kemnitz, E.:
Transparent ceramics from sol-gel derived elpasolites by cold pressing.
Solid State Sciences **9**, 833-837 (2007).

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., Hentze, M. W.:
Insights into RNA biology from a mammalian cell mRNA interactome.
Cell **149**, 1393-406 (2012).

Eichelbaum, K., Winter, M., Berriel Diaz, M., Herzig, S., Krijgsveld, J.:
Selective enrichment of newly synthesized proteins for quantitative secretome analysis.
Nature Biotechnology accepted (2012).

Eichelbaum, K., Winter, M., Föhr, S., Krijgsveld, J.:
Proteome and transcriptome kinetics in activated mouse macrophages.
In preparation (2012).

Appendix B

Supplementary Figures and Tables

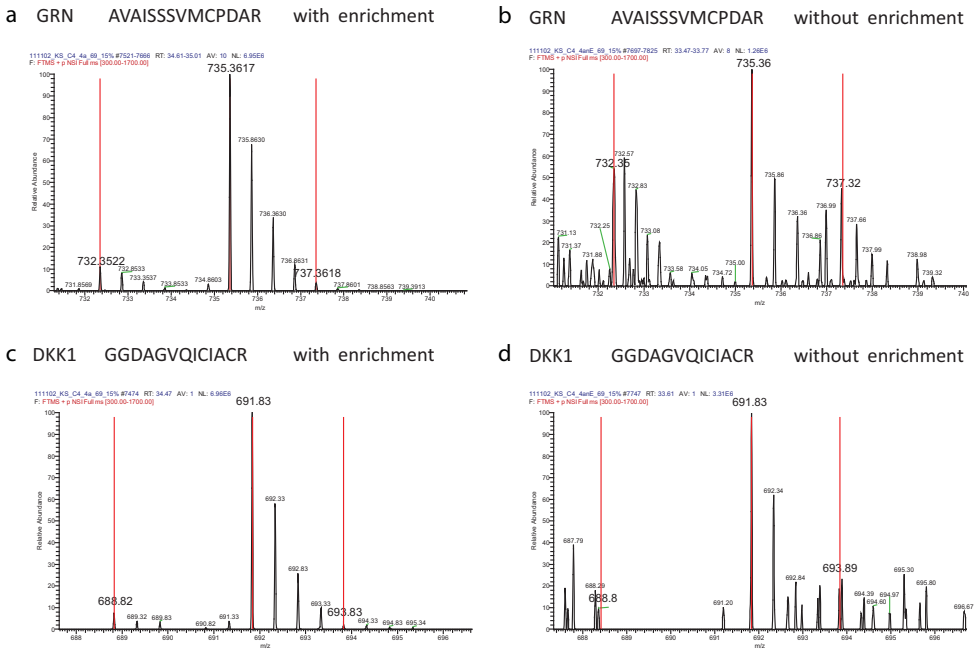

Figure B.1: Examples of MS-spectra for peptides detected with (panels a and c) and without enrichment (panels b and d). The mass range for the SILAC triplets used for quantification is shown. Red lines highlight the masses of differentially labeled peptides. The lower signal-to-noise in the spectra without enrichment complicates recognition of isotope envelopes and accurate peptide quantification.

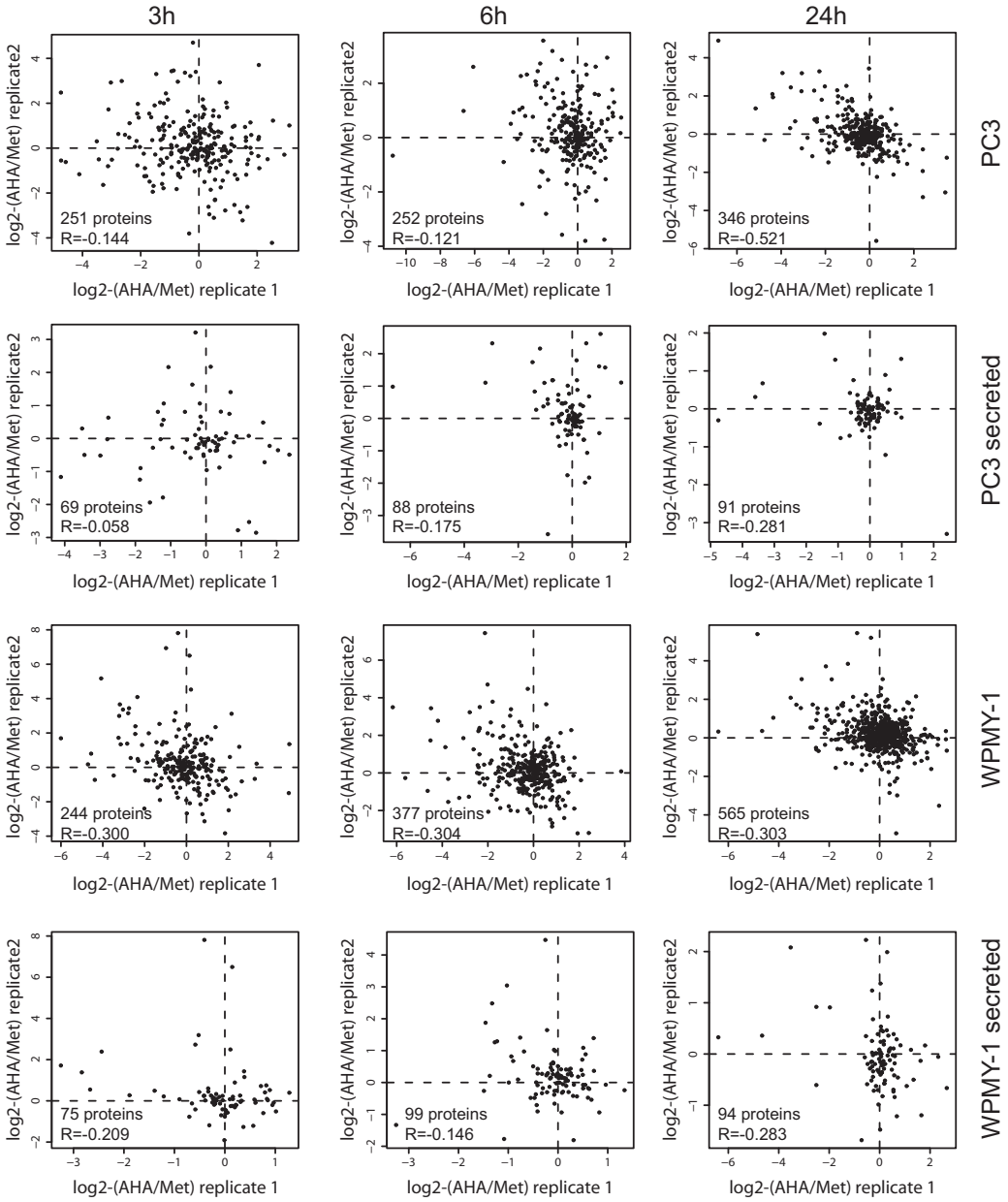

Figure B.2: Secretion profiles of cells grown in methionine- or AHA-containing media. Displayed are log2-fold-changes (AHA/methionine) for different collection times (3, 6 and 24 h) and cell lines (PC3 and WPMY-1) indicating all quantified proteins (panels in 1^{st} and 3^{rd} row) or only truly secretory proteins (panels in 2^{nd} and 4^{th} row).

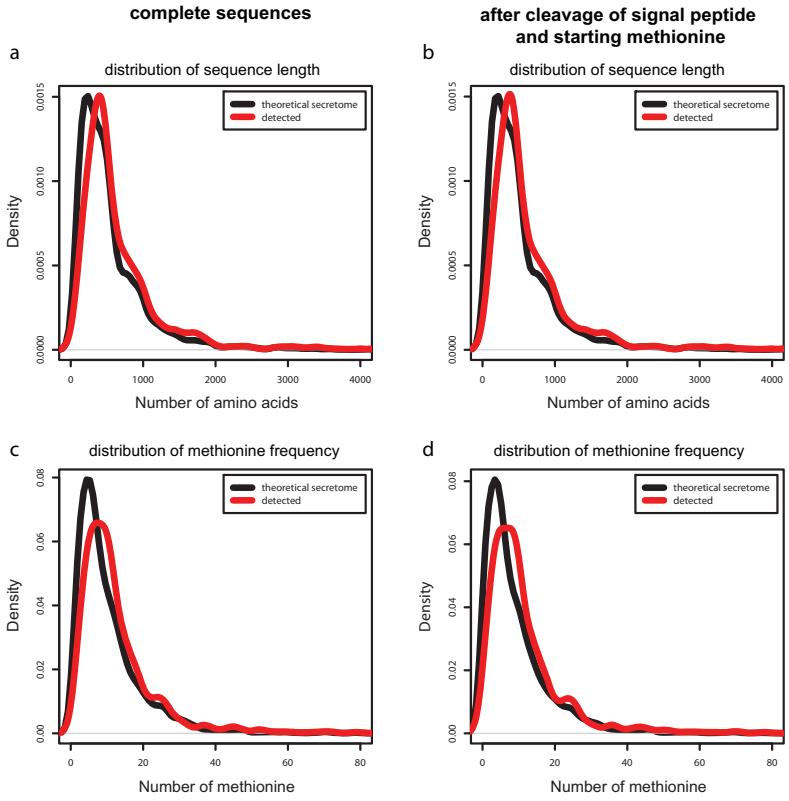

Figure B.3: Distribution of protein length (panels a and b) and methionine frequencies (panels c and d) in the theoretical secretome (3831 proteins, black lines) and experimentally detected proteins (665 proteins, red lines)Frequencies are plotted for complete protein sequences (panels a and c) and after cleavage of the N-terminal signal peptide and start-methionine (panels b and d).

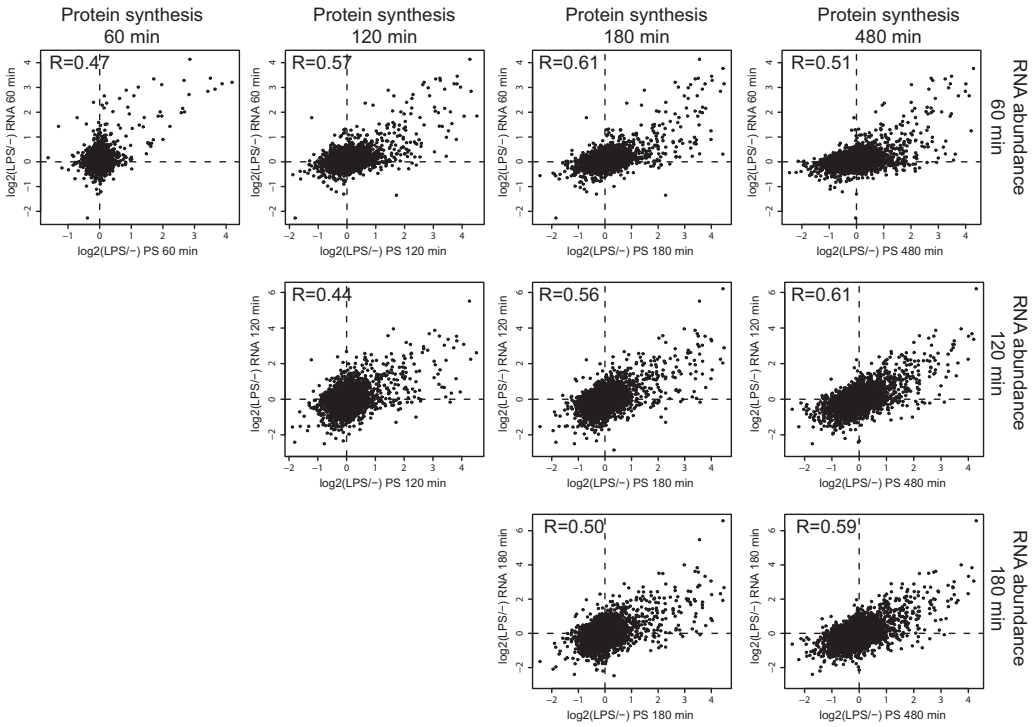

Figure B.4: Comparison of log2-fold-changes of newly synthesized proteins to log2-fold-changes of RNA at different times after LPS stimulation. The correlation between data sets increases slightly when comparing RNA expression changes to changes of protein synthesis at a later time point. (PS: protein synthesis)

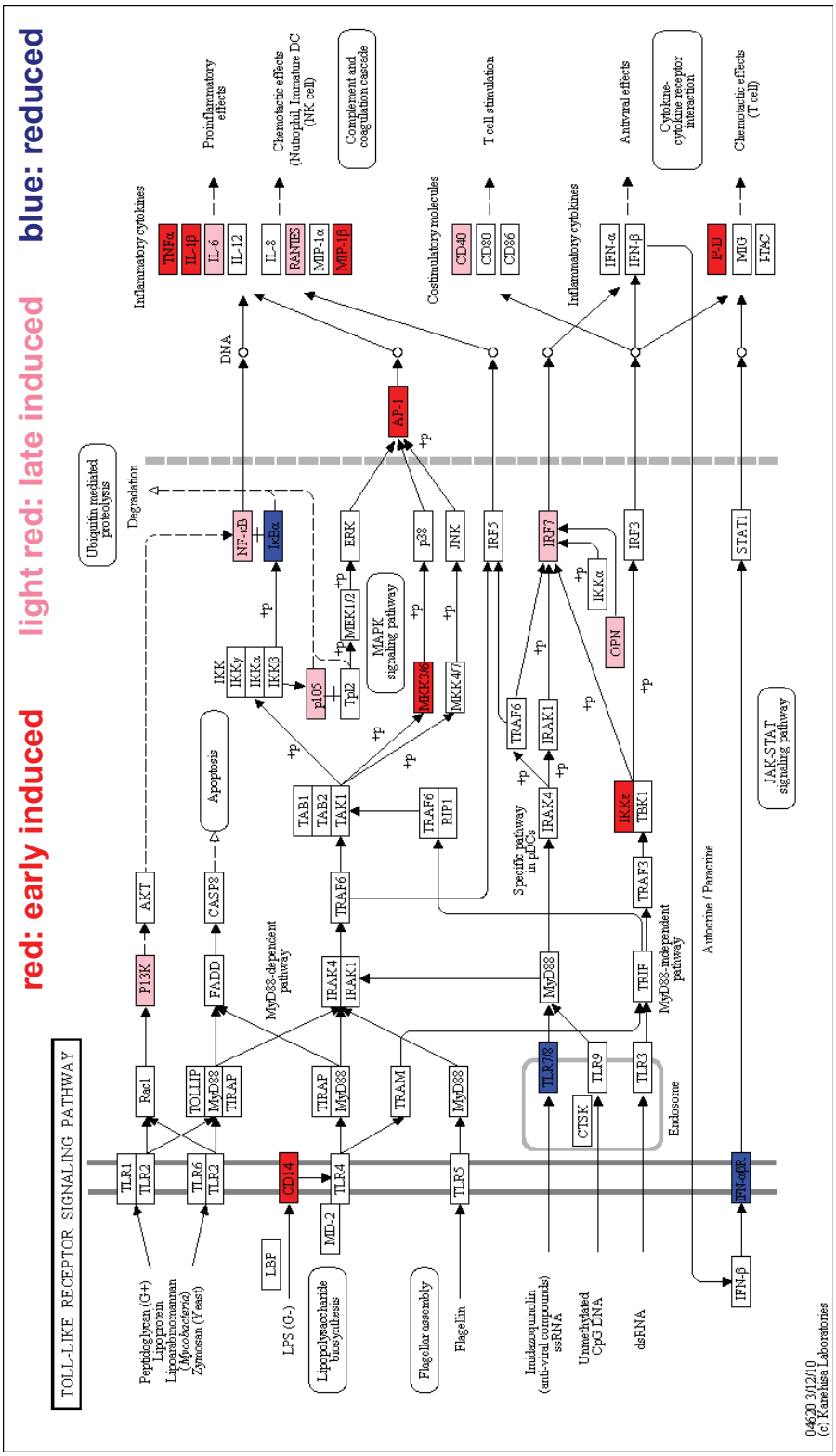

Figure B.5: Toll-like receptor signaling pathway (KEGG): significant proteins are highlighted based on their induction profile after LPS stimulation of mouse macrophages.

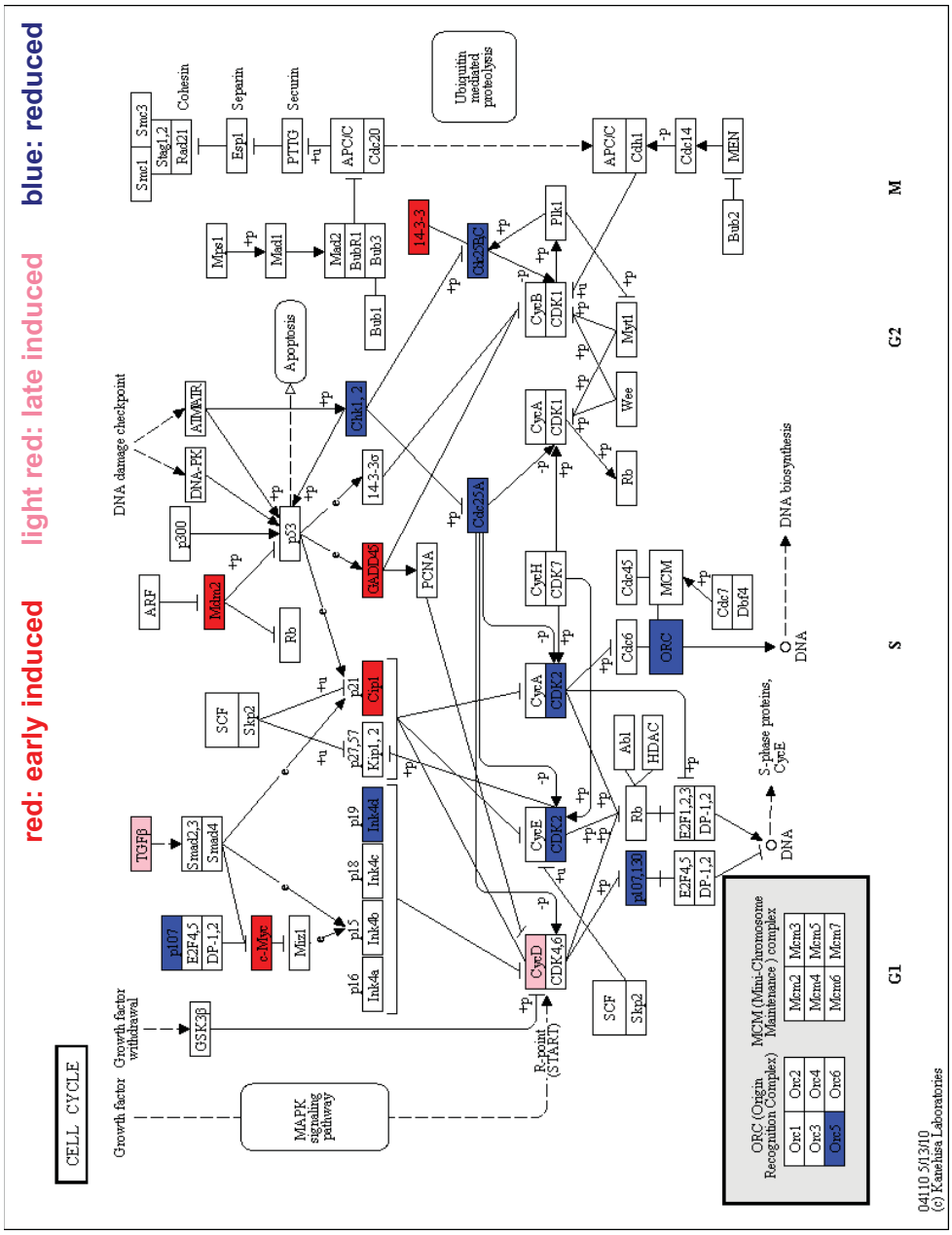

Figure B.6: Cell cycle map (KEGG): significant proteins are highlighted based on their induction profile after LPS stimulation of mouse macrophages.

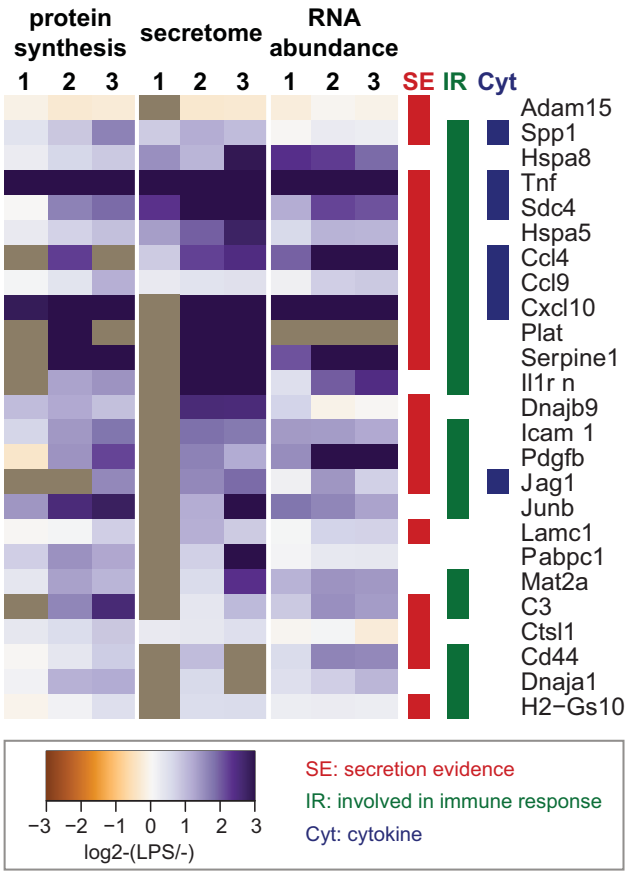

Figure B.7: Similar protein kinetics were measured intracellular and extracellular: The observed protein synthesis dynamics show the same trend for cell lysates and secretome. (The numbers above the columns give the time after LPS stimulation.)

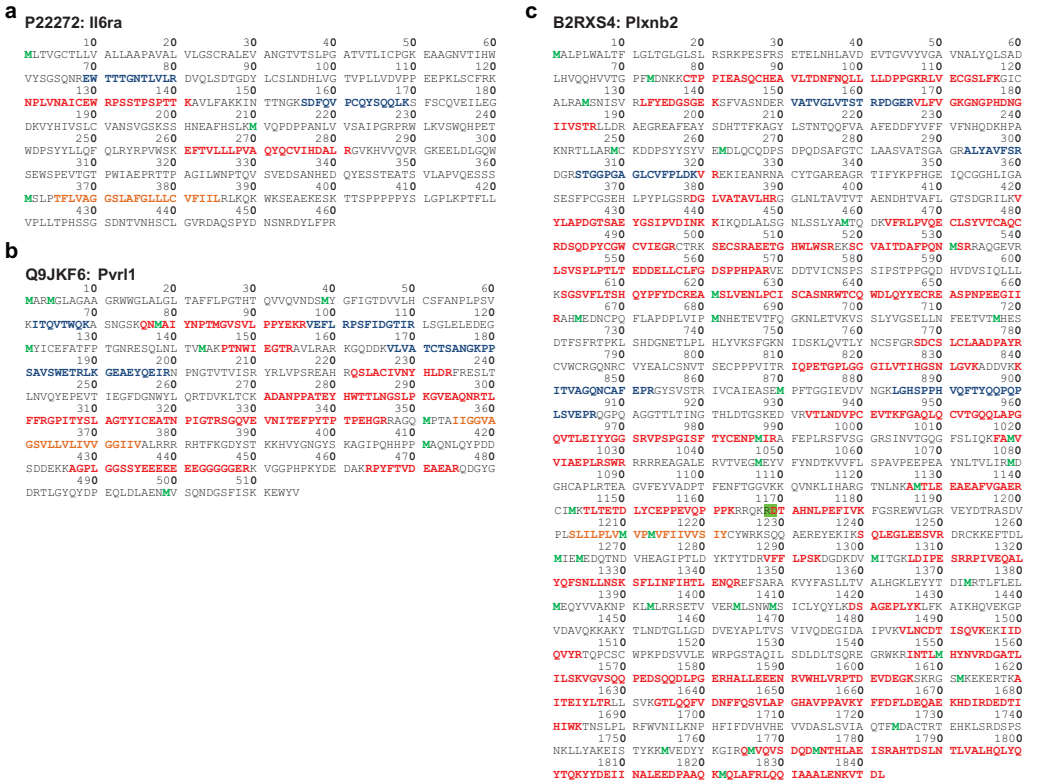

Figure B.8: Sequence coverage for three proteins with contradicting kinetics when comparing secretome and intracellular results. (red: detected peptides intracellular; blue: detected peptides intracellular and extracellular; orange: transmembrane domain; green: methionine; The cleavage site of Plxnb1 is highlighted in light green.)

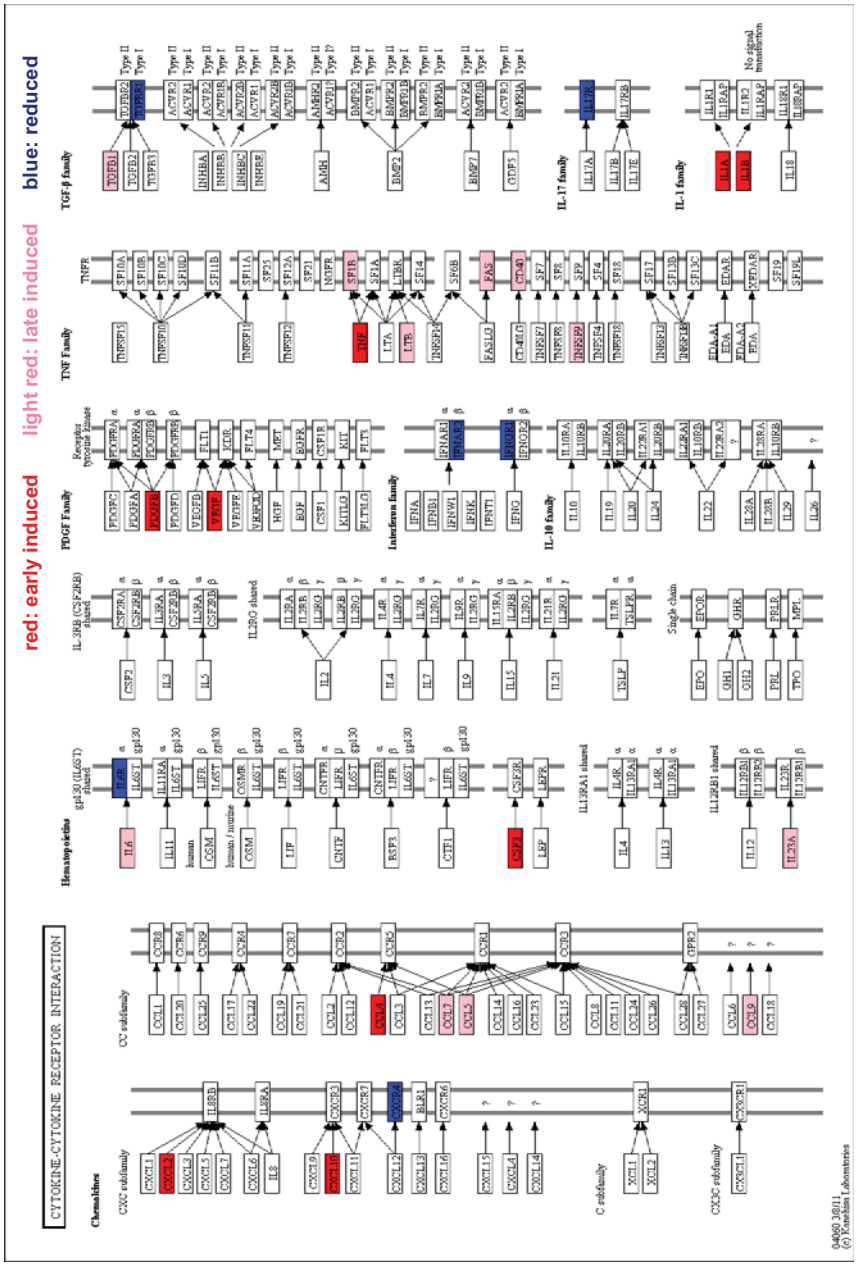

Figure B.9: Cytokine-cytokine receptor interaction map (KEGG): significant proteins are highlighted based on their induction profile after LPS stimulation of mouse macrophages.

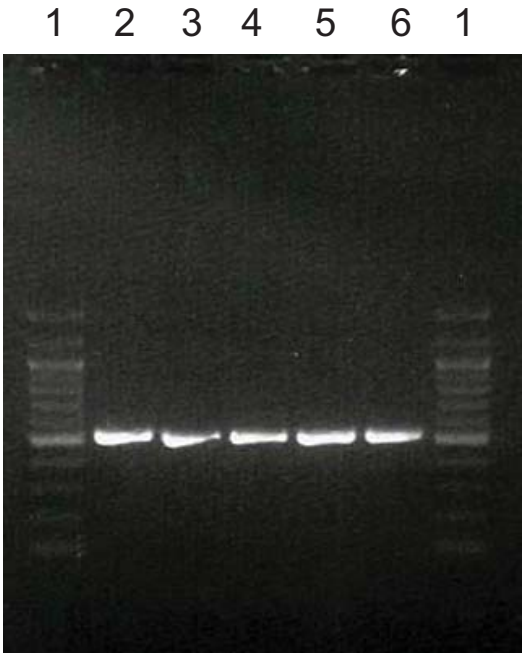

Figure B.10: Purified 500 base pair DNA sequences after PCR amplification and phenol-chloroform extraction: No contamination with longer or shorter DNA can be detected. (1: 100 kB marker; 2: tinC, 3: tinD, 4: sns, 5: CRM 663, 6: bap3)

Functional Term	Score	Functional Term	Score
non-membrane-bounded organelle	60.59	transcription activator activity	11.23
mitotic cell cycle	51.80	Zinc finger, C2H2-like	11.16
transcription regulation	50.62	DNA replication	10.61
nucleotide-binding	45.05	zinc-finger	10.27
nuclear lumen	29.34	female gamete generation	10.08
chromosome	28.52	endocytosis	9.68
chromosome organization	24.82	protein localization	9.19
ribonucleoprotein	24.44	negative regulation of gene expression	8.78
regulation of RNA splicing	22.12	neuron differentiation	8.52
helicase	20.11	nuclear chromosome	8.37
DNA metabolic process	19.15	translation factor activity, nucleic acid binding	8.24
RNA processing	17.26	kinase	7.99
positive regulation	15.74	posttranscriptional regulation of gene expression	7.90
cell cycle	15.54	regulation of cell cycle	7.19
mRNA binding	15.07	WD40 repeat, conserved site	7.06
ribonucleoprotein complex biogenesis	14.99	K Homology	6.65
Zinc finger, PHD-type	12.50	tissue morphogenesis	6.59
DNA helicase activity	11.30	chromatin assembly or disassembly	6.27

Table B.1: Functional annotation clustering for all quantified proteins. (The score is the enrichment score for each cluster provided by DAVID[230]. the functional terms were chosen to represent most of the clustered terms.)

Appendix C

Abbreviations

Abbreviation	Description
CRM	cis-regulatory module
AHA	azidohomoalanine
HPG	homopropargylglycine
HAG	homoallylglycine
Met	methionine
LPS	lipopolysaccharides
ncAAS	non-canonical amino acids
SILAC	stable isotope labeling with amino acids in cell culture
pSILAC	pulsed SILAC
iTRAQ	isobaric tags for relative and absolute quantification
COPII	coat protein complex II
COPI	coat protein complex I
OFF-Gel IEF	off-gel isoelectric focusing
fullSILAC	near 100% labeling with stable isotope labeled amino acids
NSP	newly synthesized proteins
nanoLC-MS/MS	nano-flow liquid chromatography coupled to two dimensional mass spectrometry
FDR	false discovery rate
GO	gene ontology
PHC	primary hepatocytes
iBAQ	intensity based absolute quantification
ECM	extracellular matrix
TSS	transcription start site
ChIP	chromatin immunoprecipitation
BONCAT	bioorthogonal non-canonical amino acid tagging
PAMPs	pathogen associated molecular patterns
TAMs	tumor-associated macrophages
DAPI	4',6-diamidino-2-phenylindole

Appendix D

Protein Names

PROTEIN NAMES	
Protein short name	Protein name
ADAMs	A sisintegrin and metalloproteinases
AP1	Transcription factor activator protein 1
ARE	Antioxidant responsive elements
ARG1	Arginase 1
ATF	Cyclic AMP-dependent transcription factor
C/EBP	CCAAT/enhancer-binding protein
CCL	Chemokine ligand
CD14	Monocyte differentiation antigen CD14
COX2	Cyclo-oxygenase 2
CREB	Cyclic AMP-responsive element-binding protein
CXCL	Chemokine CXC motif ligand
FIZZ1	Found in inflammatory zone 1
GR	Galactose receptor
HO1	Heme-oxygenase 1
IKK	Inhibitor of nuclear factor kappa-B kinase
IL	Interleukin
IL1ra	Interleukin 1 receptor antagonist
ILR2	Interleukin 1 receptor type II, decoy receptor
iNOS	Inducible nitric oxide synthase
IRAK	Interleukin-1 receptor-associated kinase
IRF	IFN regulatory factor
JNK	Mitogen-activated protein kinase JNK
LBP	LPS-binding protein
Mal	MyD88 adaptor-like protein
MAPK	Mitogen-activated protein kinase
MD2	Lymphocyte antigen 96
MKK	Dual specificity mitogen-activated protein kinase kinase

PROTEIN NAMES

Protein short name	Protein name
MMPs	matrix metalloproteases
MyD88	Myeloid differentiation primary response protein MyD88
NEMO	NF-kappa-B essential modulator
NRF2	Nuclear factor erythroid 2-like 2
p38	Mitogen-activated protein kinase p38
PPAR	Peroxisome proliferator-activated receptor
RIP1	Receptor-interacting serine/threonine-protein kinase 1
SR	Scavenger receptor
STAT	Signal transducer and activator of transcription
TAB	TGF-beta-activated kinase 1 and MAP3K7-binding protein
TAK1	Nuclear receptor subfamily 2 group C member 2
TANK	TRAF family member-associated NF-kappa-B activator
TBK	Serine/threonine-protein kinase TBK
TLR4	Toll-like receptor 4
TNF	Tumor necrosis factor
TRAF	Tumor necrosis factor receptor-associated factor
TRAM	TRIF-related adaptor molecule
TRIF	TIR domain-containing adapter molecule 1
Ym1	Chitinase 3-like 3 lectin

HUMAN

Gene name	Protein name
ALAD	Delta-aminolevulinic acid dehydratase
ALCAM	CD166 antigen
APP	Amyloid beta A4 protein
AREG	Amphiregulin
BCAM	Basal cell adhesion molecule
COL	Collagen
CSF1	Macrophage colony-stimulating factor 1
CTGF	Connective tissue growth factor
CYR61	Protein CYR61
DKK	Dickkopf-related protein
ERP44	Endoplasmic reticulum resident protein 44
FGFBP1	Fibroblast growth factor-binding protein 1
GDF15	Growth/differentiation factor 15
GLG1	Golgi apparatus protein 1
GPC	Glypican
HS3ST1	Heparan sulfate glucosamine 3-O-sulfotransferase 1
HSP90B1	Endoplasmin
HSPA	Heat shock 70 kDa protein
IL1RN	Interleukin-1 receptor antagonist protein
IL4L1	L-amino-acid oxidase

HUMAN	
Gene name	Protein name
IL6ST	Interleukin-6 receptor subunit beta
INHBA	Inhibin beta A chain
KITLG	Kit ligand
L1CAM	Vascular cell adhesion protein 1
LAM	Laminin
LDLR	LDL receptor
LOX	Lysyl oxidases
LTBP	Latent-transforming growth factor beta-binding protein
LY6E	Lymphocyte antigen 6E
NUCB2	Nucleobindin-2
PCSK9	Proprotein convertase subtilisin/kexin 9
PLAT	Tissue-type plasminogen activator
PLAU	Urokinase-type plasminogen activator
PLOD	Procollagen-lysine,2-oxoglutarate 5-dioxygenases
PPIA	Peptidyl-prolyl cis-trans isomerase A
PSAP	Proactivator polypeptide
ROBO1	Roundabout homolog 1
SDC	Syndecan
SERPINE1	Plasminogen activator inhibitor
SLIT2	Slit homolog 2 protein
SPP1	Osteopontin
TGFB	Transforming growth factor beta
THBS	Thrombospondin
TIMP3	Metalloproteinase inhibitor 3
TNC	Tenascin
TNF	Tumor necrosis factor
TXN	Thioredoxin
VEGFA	Vascular endothelial growth factor A

MOUSE	
Gene name	Protein name
Adam	Disintegrin and metalloproteinase domain-containing protein
Ahsg	Alpha-2-HS-glycoprotein
Anln	Actin-binding protein anillin
Apo	Apolipoprotein
App	Amyloid beta A4 protein
Arl5b	ADP-ribosylation factor-like protein 5B
Arsb	Arylsulfatase B
Atf3	Cyclic AMP-dependent transcription factor ATF-3
Atf4	Cyclic AMP-dependent transcription factor ATF-4
B2m	Beta-2-microglobulin
Bcam	Basal cell adhesion molecule
Bhlhe41	Class E basic helix-loop-helix protein 41

MOUSE	
Gene name	Protein name
Brd2	Bromodomain-containing protein 2
Btg2	Protein BTG2
C2	Complement C2
C9	Complement component C9
Capg	Macrophage-capping protein
Ccl	C-C motif chemokine
Ccnd1	G1/S-specific cyclin-D1
Ccrl2	C-C chemokine receptor-like 2
Ccrl2	C-C chemokine receptor-like 2
Cct8	T-complex protein 1 subunit theta
Cd14	Monocyte differentiation antigen CD14
Cd44	CD44 antigen
Cd83	CD83 antigen
Cdc25a	M-phase inducer phosphatase 1
Cdc25b	M-phase inducer phosphatase 2
Cdk2	Cyclin-dependent kinase 2
Cebpd	CCAAT/enhancer-binding protein delta
Chac1	Cation transport regulator-like protein 1
Chek1	Serine/threonine-protein kinase Chk1
Chek2	Serine/threonine-protein kinase Chk2
Chka	Choline kinase alpha
Clec4d	C-type lectin domain family 4 member D
Clk1	Dual specificity protein kinase CLK1
Cltc	Clathrin heavy chain 1
Coro1a	Coronin-1A
Cpe	Carboxypeptidase E
Cpeb4	Cytoplasmic polyadenylation element-binding protein 4
Csf1r	Macrophage colony-stimulating factor 1 receptor
Csf3	Granulocyte colony-stimulating factor
Csrnp1	Cysteine/serine-rich nuclear protein 1
Csrnp2	Cysteine/serine-rich nuclear protein 2
Cst	Cystatin
Cts	Cathepsin
Ctsc	Dipeptidyl peptidase 1
Cxcl	C-X-C motif chemokine
Dag1	Dystroglycan
Dnaja1	DnaJ homolog subfamily A member 1
Dnajb4	DnaJ homolog subfamily B member 4
Dsn1	Kinetochore-associated protein DSN1 homolog
Dusp	Dual specificity phosphatase
E2f1	Transcription factor E2F1
E2f7	Transcription factor E2F7
E2f8	Transcription factor E2F8
Ebi3	Interleukin-27 subunit beta
Ect2	Protein ECT2

MOUSE	
Gene name	Protein name
Edil3	EGF-like repeat and discoidin I-like domain-containing protein 3
Eef1a1	Elongation factor 1-alpha 1
Egr	Early growth response protein
Ehd1	EH domain-containing protein 1
Elf2	ETS-related transcription factor Elf-2
Elk3	ETS domain-containing protein Elk-3
Elk4	ETS domain-containing protein Elk-4
Ell2	RNA polymerase II elongation factor ELL2
Emr1	EGF-like module-containing mucin-like hormone receptor-like 1
Ets2	Protein C-ets-2
Fgd4	FYVE, RhoGEF and PH domain-containing protein 4
Flnb	Filamin-B
Fn1	Fibronectin
Fos	Proto-oncogene c-Fos
Fosb	Protein fosB
Fosl1	Fos-related antigen 1
Fosl2	Fos-related antigen 2
Gc	Vitamin D-binding protein
Gch1	GTP cyclohydrolase 1
Gk	Glycerol kinase
Glg1	Golgi apparatus protein 1
Gpnmb	Transmembrane glycoprotein NMB
Gpr84	G-protein coupled receptor 84
H1f0	Histone H1.0
H2-D1	H-2 class I histocompatibility antigen, D-D alpha chain
H2-K1	H-2 class I histocompatibility antigen, K-D alpha chain
H2-L	H-2 class I histocompatibility antigen, L-D alpha chain
H2-T23	H-2 class I histocompatibility antigen, D-37 alpha chain
Hbp1	HMG box-containing protein 1
Heximl	Protein HEXIM1
Hist1h1a	Histone H1.1
Hist1h2bf	Histone H2B type 1-F/J/L
Hist1h4a	Histone H4
Hist2h2aa1	Histone H2A type 2-A
Hist2h2ab	Histone H2A type 2-B
Hist2h3b	Histone H3.2
Hmga2	High mobility group protein HMGI-C
Hmha1	Minor histocompatibility protein HA-1
Hmox1	Heme oxygenase 1
Homer1	Homer protein homolog 1
Hpgds	Hematopoietic prostaglandin D synthase
Hsp90aa1	Heat shock protein HSP 90-alpha
Hsp90ab1	Heat shock protein HSP 90-beta
Hsp90b1	Endoplasmin
Hspa5	78 kDa glucose-regulated protein

MOUSE	
Gene name	Protein name
Hspd1	60 kDa heat shock protein, mitochondrial
Hspg2	Basement membrane-specific heparan sulfate proteoglycan core protein
Icam1	Intercellular adhesion molecule 1
Ier3	Radiation-inducible immediate-early gene IEX-1
Ier5	Immediate early response gene 5 protein
Ifi30	Gamma-interferon-inducible lysosomal thiol reductase
Igfbp4	Insulin-like growth factor-binding protein 4
Ikbke	Inhibitor of nuclear factor kappa-B kinase subunit epsilon
Il	Interleukin
Il1rn	Interleukin-1 receptor antagonist protein
Il6ra	Interleukin-6 receptor subunit alpha
Irf2bpl	Interferon regulatory factor 2-binding protein-like
Irf4	Interferon regulatory factor 4
Irf7	Interferon regulatory factor 7
Irg1	Immune-responsive gene 1 protein
Itga5	Integrin alpha-5
Itgav	Integrin alpha-V
Itgax	Integrin alpha-X
Itih4	Inter alpha-trypsin inhibitor, heavy chain 4
Itpkb	Inositol 1,4,5-trisphosphate 3-kinase B
Jag1	Protein jagged-1
Jak1	Tyrosine-protein kinase JAK1
Jak2	Tyrosine-protein kinase JAK2
Jun	Transcription factor AP-1
Junb	Transcription factor jun-B
Jund	Transcription factor jun-D
Kif11	Kinesin-like protein KIF11
Lam	Laminin
Lamc1	Laminin subunit gamma-1
Ldlr	Low-density lipoprotein receptor
Lfng	Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe
Lgals	Galectin
Lilrb4	Leukocyte immunoglobulin-like receptor subfamily B member 4
Ltb	Ltb protein
Lyz2	Lysozyme C-2
Maff	Transcription factor MafF
Map2k3	Dual specificity mitogen-activated protein kinase kinase 3
Map3k1	Mitogen-activated protein kinase kinase kinase 1
Marcksl1	MARCKS-related protein
Mastl	Serine/threonine-protein kinase greatwall
Mdm2	E3 ubiquitin-protein ligase Mdm2
Mif	Macrophage migration inhibitory factor
Mknk1	MAP kinase-interacting serine/threonine-protein kinase 1
Mknk2	MAP kinase-interacting serine/threonine-protein kinase 2

MOUSE	
Gene name	Protein name
Mmp13	Collagenase 3
Mpeg1	Macrophage-expressed gene 1 protein
Msr1	Macrophage scavenger receptor types I and II
Myc	Myc proto-oncogene protein
Ncapd3	Condensin-2 complex subunit D3
Ndrg1	Protein NDRG1
Nek7	Serine/threonine-protein kinase Nek7
Neo1	Neogenin
Nfkb1	Nuclear factor NF-kappa-B p105 subunit
Nfkb2	Nuclear factor NF-kappa-B p100 subunit
Nfkbia	NF-kappa-B inhibitor alpha
Nfkbib	NF-kappa-B inhibitor beta
Nfkbiz	NF-kappa-B inhibitor zeta
Nlrp3	NACHT, LRR and PYD domains-containing protein 3
Notch1	Neurogenic locus notch homolog protein 1
Nrp2	Neuropilin-2
Oasl1	2',5'-oligoadenylate synthetase-like 9
Olr1	Oxidized low-density lipoprotein receptor 1
Orc5	Origin recognition complex subunit 5
P4hb	Protein disulfide-isomerase
Pdgfb	Platelet-derived growth factor subunit B
Pdia	Protein disulfide-isomerase
Pdlim5	PDZ and LIM domain protein 5
Phlda1	Pleckstrin homology-like domain family A member 1
Pkm2	Pyruvate kinase isozymes M1/M2
Plagl2	Pleomorphic adenoma gene like 2
Plat	Tissue-type plasminogen activator
Plau	Urokinase-type plasminogen activator
Plk2	Polo-like kinase 2
Plk3	Polo-like kinase 3
Plxnb2	Plexin-B2
Ppia	Peptidyl-prolyl cis-trans isomerase A
Prdx1	Peroxiredoxin-1
Prim1	DNA primase small subunit
Ptprj	Receptor-type tyrosine-protein phosphatase eta
Ptprs	Receptor-type tyrosine-protein phosphatase S
Pvr	Poliovirus receptor-related protein 2
Pvrl1	Poliovirus receptor-related protein 1
Racgap1	Rac GTPase-activating protein 1
Rasa3	Ras GTPase-activating protein 3
Rassf2	Ras association domain-containing protein 2
Rbl1	Retinoblastoma-like protein 1
Rcsd1	CapZ-interacting protein
Rel	Proto-oncogene c-Rel
Relb	Transcription factor RelB

MOUSE	
Gene name	Protein name
Rfc4	Replication factor C subunit 4
Rgl1	Ral guanine nucleotide dissociation stimulator-like 1
Rpl30	60S ribosomal protein L30
Rps25	40S ribosomal protein S25
Rps27a	Ubiquitin-40S ribosomal protein S27a
Sdc4	Syndecan-4
Sdcbp	Syntenin-1
Sema4d	Semaphorin-4D
Serpina	Serine protease inhibitor
Sesn2	Sestrin-2
Sil1	Nucleotide exchange factor SIL1
Slc4a7	Sodium bicarbonate cotransporter 3
Slfn2	Protein Slfn2
Sparc	Secreted acidic cysteine rich glycoprotein
Spp1	Osteopontin
Srf	Serum response factor
Srpk2	SRSF protein kinase 2
Stk17b	Serine/threonine-protein kinase 17B
Stk38	Serine/threonine-protein kinase 38
Stk40	Serine/threonine-protein kinase 40
Sumf1	Sulfatase-modifying factor 1
Tfcap4	Activator protein 4
Tgfb1	Transforming growth factor beta-1
Tgm2	Protein-glutamine gamma-glutamyltransferase 2
Timp2	Metalloproteinase inhibitor 2
Tinagl1	Tubulointerstitial nephritis antigen-like
Tlr13	Toll-like receptor 13
Tlr3	Toll-like receptor 3
Tlr7	Toll-like receptor 7
Tnf	Tumor necrosis factor
Tnfsf9	Tumor necrosis factor ligand superfamily member 9
Vav3	Guanine nucleotide exchange factor VAV3
Vegfa	Vascular endothelial growth factor A
Vps37c	Vacuolar protein sorting-associated protein 37C
Wdr67	WD repeat-containing protein 67
Xpo1	Exportin-1
Ybx1	Nuclease-sensitive element-binding protein 1
Ywhae	14-3-3 protein epsilon
Ywhaz	14-3-3 protein zeta/delta
Zcchc2	Zinc finger CCHC domain-containing protein 2
Zfp36	Tristetraprolin
Zyx	Zyxin

DROSOPHILA MELANOGASTER	
Gene symbol	Protein name
baf	Barrier-to-autointegration factor
bigmax	BHLHZip transcription factor BIGMAX
bin	Protein Biniou
CG2990	LD18775p
CG32772	CG32773
Chi	CHIP
CrebB-17A	Cyclic-AMP response element binding protein B at 17A
DAPI	4',6-diamidino-2-phenylindole
Eip75B	Ecdysone-induced protein 75B
esg	Escargot
ewg	DNA-binding protein Ewg
hang	Zinc finger protein hangover
hb	Protein hunchback
Hrb27C	Heterogeneous nuclear ribonucleoprotein 27C
Hrb87F	Heterogeneous nuclear ribonucleoprotein 87F
Kr-h1	Krueppel homologous protein 1
lark	RNA-binding protein lark
lin-52	Lin-53
Mef2	Myocyte-specific enhancer factor 2
mod(mdg4)	Modifier of mdg4
nub	Protein nubbin
pdm2	POU domain protein 2
pnr	Protein Pannier
Rbf2	RB-related protein RBF2
Rox8	Rox8
snail	Protein snail
so	Protein sine oculis
sqz	Zinc finger protein squeeze
Ssb-c31a	RNA polymerase II transcriptional coactivator
Su(H)	Suppressor of hairless protein
vvl	POU domain protein CF1A
wor	Worniu

Appendix E

Acknowledgment

Without the multitude of ideas, advices, hints, discussions, encouragements, technical and practical support, *etc.*, meaning with the help of many friends, colleagues and researchers I would not have been able to finalize this thesis. I really appreciate all the support I got from various directions.

Especially, my supervisor Jeroen supported me throughout the complete thesis. It was great to experience together with him the installation of a new laboratory and the development of the group during the first four years. Jeroen not only initiated very interesting projects for me but also organized collaborations with various groups and institutions, enabling me to perform whatever experiments we though of and to get practical insights in multiple fields in biology. A major point I am thankful for is his trust in me and my work and the freedom he gave me.

Next, I would like to acknowledge the members of my TAC committee: Anne-Claude Gavin, Eileen Furlong and Matthias Mayer, who took the time ones a year to think about and discuss my work and who gave me great input on experimental design as well as how to present my work successfully.

In addition, many people from different groups shared their practical and theoretical knowledge with me, initiating new ideas or enabling various experiments. Starting as Chemist in biology I had to learn a lot basics. Here, especially Robert Zinzen from the Furlong lab spend lots of time to teach me everything related to fly- and DNA-work and supported me when ever I got stuck in the DNA project. Emmanuel from the Gavin lab instructed me in western blotting and Heike and Amanda from the Schultz group showed me how to work with cells and provided the RAW 264.7 cells.

I had multiple collaborations with groups at the DKFZ. Without Mauricio from the Herzig group an important part of the secretome paper would not exist. Thanks a lot for this successful collaboration. I had and have other secretome projects with Michaela from the same group, Steffi, David

and Markus Feuerer. All of them provided me with their model system for secretome analysis and gave me insights into their field of biology. I would like thank Johanna Schott from the Stöcklin group for performing the polysome-profiling experiment with me twice and for the discussions about macrophage biology.

At EMBL I have a wonderful collaboration with the Hentze group. Together with Alfredo, Rasti, Bene, Claudia and many other group members we analyzed the RNA-interactome of multiple organism. Bernd Fischer was a great collaborator in this project and gave together with Wolfgang Huber very helpful hints for the statistical analysis of my data.

The core facilities at EMBL are a pool of theoretical and practical experience that helped me throughout my thesis. I would like to especially mention Tomi and Vladimir from GeneCore who performed the RNA expression measurement for me and supported me in the data analysis for that. An other huge thanks goes to the members of the proteomics core facility: Kristina, Joanna and Stefan. They spend immense effort to keep our machines running smooth, are always their to discuss experiments and it is just fun to work with them and spend nice lunch and coffee breaks.

In the same line I would like to thank all former and current members of the Krijgsveld lab: Chris, Daniel, Ita, Sergey, Jenny, Markus, Satoshi, Sina, Sonja and Sophia. Especially Jenny, Sonja and Sophia were alway there to discuss whatever problems. Not to forget Jan from the Gavin group who was eager to discuss experiments or data or to construct tools that made life easier.

Finally I would like to thank the people that read and gave comments to the thesis: Robert, Chris and Maik; all my friends and my family who supported me throughout the whole time of the thesis, as well as everybody I unintentional forgot in this acknowledgment.