TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Proteomik und Bioanalytik

Advanced quantitative mass spectrometry

applied to aneuploidy and apoptosis

Gabriele Maria Stöhr

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Langosch

Prüfer der Dissertation: 1. Univ.-Prof. Dr. B. Küster

2. apl. Prof. Dr. M. Mann

(Ludwig-Maximilians-Universität München)

Die Dissertation wurde am 07.01.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München am 05.02.2013 angenommen.

For my parents

Summary

Over the last 25 years, mass spectrometry-based proteomics has been established as the most important technology in proteomics research. It has not only been developed to identify single proteins, but within the last decade it has also turned quantitative, thereby allowing comparison of complex protein samples with regard to their protein expression patterns. This enormous step forward enables researchers to comprehensively investigate biological processes at the level of the proteome in a global manner.

For my thesis, I applied state-of-the-art quantitative mass spectrometry-based proteomics to two important biological phenomena, aneuploidy and apoptosis. Both aneuploidy and apoptosis play a crucial role in cancer development. Aneuploidy, the state of imbalanced chromosome complement, occurs in most cancer cells. However, its role as potential tumor promoter is still debated. In contrast, apoptosis, a controlled type of cell death, is blocked in cancer cells and its evasion represents one of the hallmarks of cancer. Unraveling of fundamental mechanisms of aneuploidy and apoptosis may lead to new cancer therapies that specifically target proteins in these biological processes and thereby are more specific and less prone to side-effects of cancer treatment.

It is of great interest to discover how proteins encoded on the extra chromosomes are regulated and whether protein levels correlate with transcription levels. In addition, it needs to be elucidated whether there is a general response in all aneuploid human cells to the chromosomal imbalance. In my first study, I applied mass spectrometry for the first time to human aneuploid cell lines and compared proteome data to genome and transcriptome levels. When focusing on the additional chromosome and corresponding mRNA and protein levels, I could define a direct correlation between the number of chromosomal copies and the transcriptome level. On the proteome level, however, clear indications for balancing of some proteins were observed, especially of proteins which form subunits of macromolecular complexes. This suggests that the cell possesses mechanisms to balance proteins for which it is of importance to retain a specific stoichiometry. When investigating global regulations in aneuploid cells, we observed a general response to aneuploidy in all different clones. This indicates that cells respond similarly to the aneuploid state, independent of the number of extra chromosomes, the cell type or the chromosome.

Interestingly, in each clone lysosomal pathways were identified as up-regulated, including the autophagic pathway. This degradation pathway is responsible for the recycling of organelles and ubiquitinated proteins. We therefore argue that the autophagic pathway might serve as a regulatory mechanism in aneuploid cells to cope with the extra amount of proteins in the cell.

The second study aimed at the identification of molecular changes in apoptotic cells to provide a global view on the underlying events. Here, I focused on apoptosis induction by TRAIL, an important extrinsic death ligand that is known to induce apoptosis in cancer cells without cytotoxic events in normal cells. The main players in apoptosis are proteases, so-called caspases, which proteolytically cleave proteins to guide the cell through controlled cell death. For this reason, I focused on the identification of caspase-dependent cleavage substrates to unravel the response of cells to apoptosis induction. I performed SILAC-based quantitative mass spectrometry in Jurkat T cells, either TRAIL or mock treated. We developed an algorithm including sophisticated statistical analysis that extracts potential cleavage events from a complex mixture. After further verification, I defined a list of close to 700 substrates which are cleaved after TRAIL-induced apoptosis, its so-called "degradome", and performed further downstream analyses on these substrates. The data provide a thorough analysis of TRAIL-induced cleavage events and can be used as a basis for further investigations on the effect of TRAIL in cancer research. Moreover, the approach developed here can be applied to any biological topic involving proteolytic cleavage and therefore may become an important tool in degradome research.

Zusammenfassung

Innerhalb der letzten 25 Jahre hat sich Massenspektrometrie-basierte Proteomik als eine der wichtigsten Technologien in den Lebenswissenschaften etabliert. Ausgehend von der Detektion einzelner Proteine ist es heute möglich ganze Proteome nahezu umfassend quantitativ zu untersuchen. Dieser enorme Fortschritt erlaubt es Wissenschaftlern biologische Prozesse auf globaler Proteinebene zu erforschen.

Für meine Doktorarbeit wendete ich modernste quantitative massenspektrometrische Analyseverfahren (LC-MS/MS) auf zwei wichtige zellbiologische Phänomene an, Aneuploidie und Apoptose. Beide Prozesse spielen eine entscheidende Rolle in der Entwicklung von Krebs. So tritt Aneuploidie, ein Zustand unausgewogener Chromosomenanzahlen, in den meisten Krebszellen auf. Die Funktion als potentieller Tumorinitiator wird jedoch noch diskutiert. Im Gegensatz dazu ist Apoptose – eine Form des programmierten Zelltods - in Krebszellen unterdrückt, und ebenfalls ein Markenzeichen vieler Krebsarten. Die Entschlüsselung fundamentaler Wirkungsweisen von Aneuploidie und Apoptose kann somit zu einem besseren Verständnis dieser medizinisch wichtigen zellulären Prozesse beitragen und könnte somit zur Entwicklung neuer Krebstherapien beisteuern.

Eine zentrale Fragestellung ist, wie bei Aneuploidie die Expression von Proteinen reguliert wird, die auf dem zusätzlichen Chromosom kodiert sind, und wie sich hierbei Protein- und Transkriptionslevel zueinander verhalten. Ferner ist es wichtig zu verstehen in wie weit es eine zelluläre Antwort auf den unausgeglichenen Chromosomenzustand gibt. In meinem ersten Projekt habe ich Massenspektrometrie-basierte Proteomanalysen auf humane aneuploide Zelllinien angewendet und konnte so erstmals proteom-weite Proteinexpressionen quantitativ mit entsprechenden Genom- und Transkriptom-Daten vergleichen. Es konnte gezeigt werden, dass mRNA Level mit der entsprechenden Chromosomenanzahl korrelieren. Proteinexpressionslevel hingegen wiesen eindeutige Regulationsmechanismen für zahlreiche Proteine auf, insbesondere solche, welche Untereinheiten von makromolekularen Komplexen bilden. Dies lässt darauf schließen, dass zelluläre Regulationsmechanismen existieren um solche Proteine zu regulieren für welche eine spezifische Stöchiometrie erforderlich ist. Die Untersuchung globaler Effekte verdeutlichte übereinstimmende generelle Antwortmechanismen in allen untersuchten

aneuploiden Klonen. Folglich reagieren Zellen gleich auf einen aneuploiden Zustand, unabhängig von der Anzahl an zusätzlichen Chromosomen, dem Zelltypen oder dem zusätzlichen Chromosom selbst. Interessanterweise sind in allen Klonen Proteine lysosomaler Stoffwechselwege in ihrer Expression erhöht, wie unter anderem von Autophagie. Dieser Abbauweg ist verantwortlich für das „Recyceln" von ubiquitinierten Proteinen und ganzen Organellen. Wir vermuten daher, dass Autophagie in aneuploiden Zellen als regulatorischer Mechanismus dienen könnte, um den Überschuss an Proteinen auszugleichen.

Mein zweites Projekt zielte auf die globale Analyse molekularer Veränderungen in apoptotischen Zelle ab. Ich fokussierte mich hierbei auf die Aktivierung von Apoptose mittels TRAIL, einem wichtigen extrinsischen apoptotischen Liganden. Dieser ist dafür bekannt, Apoptose in Krebszellen, jedoch nicht in normalen Zellen auszulösen. Im Mittelpunkt des apoptotischen Zelltods stehen spezielle Cystein Proteasen, so genannte Caspasen, welche zelluläre Proteine proteolytisch schneiden um hierdurch die Zellen kontrolliert in den Zelltod zu führen. Aus diesem Grund fokussierte ich mich auf die Identifizierung von Proteinen, welche durch Caspasen geschnitten werden um somit die Zellantwort auf Apoptose-Induktion zu erforschen. Hierzu wendete ich SILAC-basierende quantitative Massenspektrometrie in Jurkat T Zellen an, welche entweder mit TRAIL oder einer Kontrolle behandelt wurden. Wir entwickelten einen Algorithmus einschließlich statistischer Analysen welcher potentielle Caspase Substrate aus komplexen Proteingemischen extrahiert. Die Anwendung dieses Verfahrens führte nach weiterer Validierungen zu einer Liste von nahezu 700 Substraten, welche nach TRAIL-Zugabe proteolytisch gespalten wurden – das sogenannten „Degradom". Ferner wurden die Substratproteine hinsichtlich ihrer zellulären Funktionen genauer charakterisiert. Die Daten stellen somit eine proteom-weite Analyse TRAIL-induzierter Caspase Substrate dar und können daher als Grundlage für weitere Untersuchungen an TRAIL und dessen Rolle in der Krebsforschung dienen. Zudem kann die in dieser Arbeit entwickelte Methode zur Analyse proteolytischer Abbaufragmente auf jegliches Thema angewendet werden, welches proteolytische Spaltung untersucht, und kann deshalb als wichtiges Werkzeug in der Degradomforschung eingesetzt werden.

Table of Contents

1. INTRODUCTION

1.1. Mass spectrometry-based proteomics

To better understand biological phenomena on a molecular level it is essential to analyze the main players of cellular functions – the proteins – on a global scale. To fully describe the entire set of proteins of a specific cell, tissue or organism, termed the 'proteome', required tremendous advances in technology during the last 25 years. Mass spectrometry (MS) has become the central approach for identifying and quantitatively analyzing proteins and has thereby contributed enormously to establish proteomics as an independent scientific discipline. Accompanying developments in sample preparation and separation preceding mass spectrometric analysis - such as isotopic labeling of proteins or chromatographic improvements to separate highly complex peptide mixtures, but also tremendous enhancements in the bioinformatics analysis and integration of the huge amount of data produced by modern mass spectrometers - have turned MS-based proteomics in one of the most powerful tools in modern cell and systems biology. In the following sections some major aspects of MS-based proteomics will be highlighted.

1.1.1. Principles of MS-based proteomics

The general goal of mass spectrometry-based proteomics is the identification of proteins in complex mixtures. Most approaches aim to cover the proteome of a certain species – for example tissues or cell lines – as completely as possible, but also single proteins or certain subsets of proteins (sub-proteomes) such as proteins from organelles (1, 2) or specific subfamilies of proteins (3) are analyzed. In general, one can distinguish between two principles when analyzing proteins by MS. So-called top-down approaches aim at analyzing complete full length proteins (4, 5). This procedure has not been standardized yet and it requires sophisticated data analysis tools. A by far more common procedure is the so-called shotgun proteomics approach, which identifies proteins in a bottom-up manner. Peptides are analyzed that have been generated by cleavage of the proteins by e.g. enzymatic digestion, and identification and quantification of the corresponding proteins is achieved by MS as described later in more detail (see section 1.1.2. and 1.1.3.).

The typical workflow for a shotgun MS-based proteomics experiment is depicted in Figure 1. Initially, proteins are extracted from the sample to be analyzed e.g. by lysis of cells. The next steps comprise in most cases fractionation of the sample to reduce its complexity. Fractionation can be performed for example at the protein level by applying electrophoresis techniques such as SDS-PAGE (6) to

separate the proteins according to their molecular weight. Subsequently, proteins are cleaved to peptides and this step is normally performed by proteolytic digestion. Trypsin has turned out to be the protease of choice as it generates peptides with lengths and charge states well suited for MS analysis (7) but also other proteases or chemical cleavage are applicable.
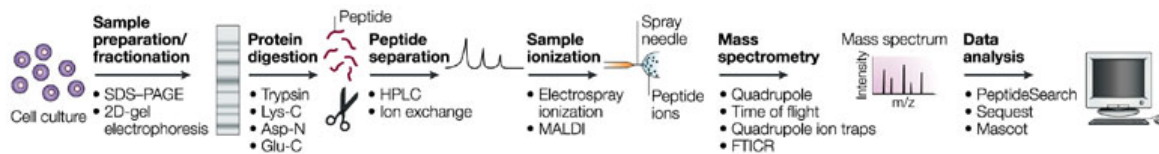

Figure 1: Schematic workflow of a mass spectrometric-based LC-MS/MS approach. Adopted from (8).

The resulting highly complex peptide mixture can be further separated by various chromatographic techniques or it can be directly subjected to liquid chromatography coupled to a mass spectrometer (LC-MS/MS). This part of the workflow is described in detail further below. Data generated from a mass spectrometric experiment is then processed with dedicated software packages to identify and quantify the analyzed peptides and corresponding proteins (see 1.1.3.). By applying shotgun proteomics, recently the first proteome has been quantitatively analyzed to near completion (9) illustrating the tremendous power of this technique. Moreover, these data demonstrate that the combination of applying state-of-the-art mass spectrometry with sophisticated up- and downstream analyses yields the best possible results.

In the following, the most important features of the LC-MS/MS setup will be highlighted with emphasis on electrospray ionization coupled to a hybrid Orbitrap mass spectrometer, as used in this thesis.

1.1.1.1. Liquid chromatography

Since the peptide mixtures that are frequently analyzed in MS-based proteomics experiments, e.g. derived from cellular lysates, are typically highly complex and contain up to hundreds of thousands of peptides (10), chromatographic separation is indispensable before mass spectrometric detection. Reversed phase (RP) chromatography represents the most commonly used separation method for liquid chromatography (LC). It is based on hydrophobic interactions of peptides with alkyl chains (usually C-18 groups) covalently linked to the stationary phase - beads packed within a column. After binding to the RP material peptides are sequentially eluted from the column via a solvent gradient and are then introduced into the mass spectrometer. When using electrospray ionization (see

section 1.1.1.2.) the chromatographic separation step is almost always coupled 'online' to the mass spectrometric analysis (LC-MS).

As sensitivity is a key parameter in most proteomics experiments in order to analyze as many proteins as possible in a wide dynamic range, chromatographic separation had to be miniaturized to flow rates in the nano-liter (nl) per minute range and column dimensions had to be adjusted accordingly. Over the last years, especially with the introduction of new ultra performance LC systems (UPLC), which tolerate pressure limits of up to 1000 bar, there is a trend towards longer column lengths and smaller bead size. Whereas, column lengths between 10 - 15 cm (inner diameter (ID) 75 µm) and bead sizes of 3 µm were standard for many years, researchers now aim for longer columns (up to 50 cm) in combination with smaller bead sizes (smaller than 2 µm) to improve peak shapes for better chromatographic resolution and thereby increasing peak capacity (11, 12). Inner diameters of columns and flow rates can also be further reduced to gain sensitivity. To properly run these columns which generate high backpressure in the LC system UPLC systems are indispensable. In addition heating of columns, which has been commonly used in standard HPLC applications for years, is now becoming available in nano-LC applications both reducing the backpressure and improving chromatography (11, 13).

An alternative approach uses so-called monolithic columns which allow the application of extremely long columns (up to several meters) and gradients (>8 hours) with unsurpassed separation capabilities (14). However, the production and application of these columns is so far not in routine use but holds great promise for future applications.

To further reduce the complexity of a sample that is investigated in an LC-MS experiment, additional chromatographic separation steps can be combined. Several years ago two-dimensional separation combining strong cation exchange (SCX) chromatography with RP chromatography coupled online to the mass spectrometer has been introduced (15). However, nowadays the two dimensions are more often separated and the first off-line separation step is included in the sample preparation before LC-MS analysis. The separation may be performed at the protein or peptide level or both. SCX chromatography has been used broadly, especially for the fractionation of post translational modifications such as phosphorylation or N-terminal acetylation (16-20). In recent years, further emphasis has been placed on alternative chromatographic techniques such as strong anion exchange (21), hydrophilic interaction chromatography (22, 23) or high-pH reversed chromatography (24).

1.1.1.2. Ionization techniques

The first important step in the mass spectrometric analysis itself is the ionization of the analyte and its transfer from liquid into gas phase enabling the analyte to enter the mass spectrometer. The introduction of so-called soft ionization techniques in the late 1980s allowed the mass spectrometric analysis of biopolymers such as proteins and peptides and is therefore a milestone in the development of what we call today MS-based proteomics. The importance of these developments was emphasized by awarding the Nobel Prize 2002 in chemistry to John Bennett Fenn and Koichi Tanaka for their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules (25, 26).

Matrix-assisted laser desorption/ionization (MALDI)

Matrix-assisted laser desorption/ionization (MALDI) is an ionization technique to transfer peptides into gas phase (27, 28). It relies on co-crystallization of peptides with an organic matrix (in most cases aromatic acids such as sinapinic acid, α-cyano-4-hydroxycinnamic acid or 2,5-dihydroxybenzoic acid) on a sample plate.

Directing a laser beam onto the crystallized sample leads to the formation of a hot plume. Peptides are sputtered from the co-crystallized sample while the matrix transfers positive charges (protons) onto the peptide. Since voltage is applied to the sample plate and the inlet of the mass spectrometer, peptides are accelerated into the mass spectrometer by the resulting electrostatic field (see Figure 2). In most applications, MALDI ionization is coupled to time-of-flight (ToF) instruments, mainly due to its large mass range and because ion packages generated in pulses by the laser are perfectly suited for detection via a ToF analyzer.
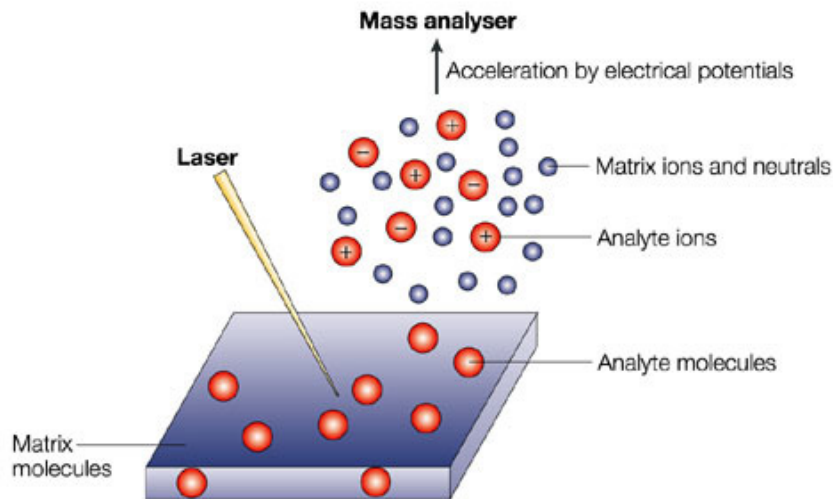
Figure 2: General principle of the matrix assisted laser desorption/ionization (MALDI) process. Adopted from (8).

Electrospray ionization (ESI)

The most widely used ionization technique is electrospray ionization (ESI) which ionizes molecules directly from liquid phase and is therefore compatible with chromatographic separation techniques (25).
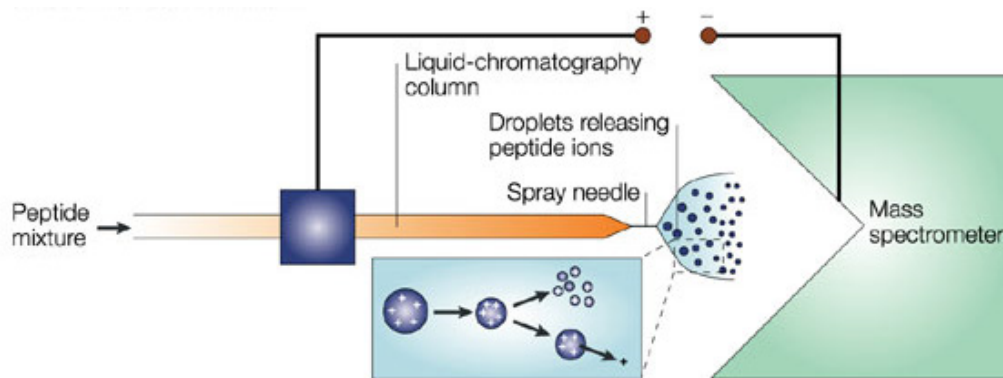
Figure 3: General principle of the electrospray ionization (ESI) process. Adopted from (8).

ESI is coupled to mass spectrometers that can deal with an ongoing flow of neutrals into the instrument, thereby allowing the detection of a large variety of chemical substances, which are entrained in the gas flow (29). As its name implies, electrospray ionization describes a process in which an analyte solution is sprayed in front of the mass spectrometer when applying electric voltage to the capillary from where the peptides are eluting (see Figure 3). The eluted analytes are

[5]

thereby dispersed as a fine aerosol finally leading to charged gas phase ions after evaporation of the solvent. Consequently, peptides are directly transferred from solution into gas phase. However, it is still not clear how exactly this last process occurs in detail.

Two different models for the ionization mechanism have been proposed, the ion evaporation model (30) and the charge residue model (31, 32). The first model assumes that a droplet is shrinking by evaporation of the solvent until the droplet surface reaches a certain field strength which allows single solvated ions to be expelled from the droplet. The charge residue model on the other hand assumes that the electrospray generates droplets that only contain one analytical ion. Both models could not fully be proven and are still discussed to date (29, 32).

A number of different parameters can influence the ionization process. The solvent needs to be free of salts, otherwise the ionization process is impaired. Reversed-phase chromatography is commonly performed under low pH conditions to ensure protonation of the peptides, an essential requirement for the ionization process. Therefore, acids such as acetic or formic acid are frequently added to the solvent. They reduce the pH and form salt bridges with the protonated residues – a process called ion pairing. Since these reagents both influence the chromatographic retention and the ionization process, they have to be handled with care. TFA, for example, is a very efficient ion pairing agent, however, it negatively influences the ionization process when used in too high concentration due to its ability to form gas-phase ion pairs with positively-charged analyte ions (33, 34).

1.1.1.3. Mass spectrometers

The heart of our proteomics experiment is formed by the mass spectrometer which measures the molecular weight of the analytes – or more precisely their mass-to-charge ratio (m/z) - in order to identify the peptides and the corresponding proteins in the analyzed sample. When analytes reach the inlet of the mass spectrometer e.g. via electrospray ionization, ions are guided by a decreasing vacuum gradient to the mass analyzer and subsequently by lenses and corresponding voltages (see Figure 4). In the detector, the mass of all ions are recorded, forming the so-called MS spectra.

However, for unambiguous identification of a particular peptide, determination of its exact mass is not fully sufficient, although mass accuracies in today's instrumentation can reach the low ppm or even the sub-ppm range. Instead, single ion populations are additionally isolated and fragmented before mass detection generating a fragment spectrum (termed MS^2 or MS/MS spectrum) from which some sequence information of the analyzed peptide can be deduced. Combining the information of the exact mass with the sequence information of the analyzed peptide ion the

corresponding protein can be identified. This is normally done by applying designated search algorithms that map the mass spectrometric information onto sequence databases (see section 1.1.3.).
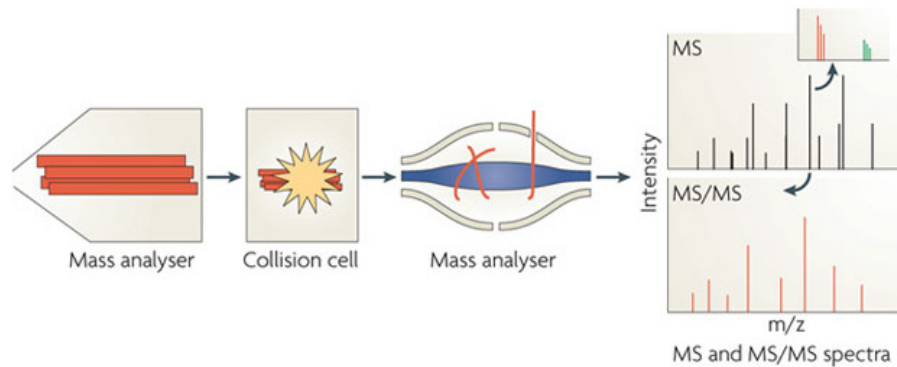

Figure 4: Schematic of the mass spectrometric analysis. Ionized peptides reach the inlet of the mass spectrometer and are guided to the Orbitrap cell. First, all ions are measured and a MS spectrum is generated. Subsequently, single ion populations are isolated, fragmented in a collision cell and MS/MS fragment spectra are recorded. Adopted from (35).

Many different types of mass spectrometers have been developed. During the last decade 'hybrid instruments' - which combine different devices to separate analyte ions based on their m/z ratio - have become increasingly popular. In MS-based proteomics quadrupole-time-of-flight (qTOF) and especially the linear ion trap – Orbitrap analyzer are the most widely used mass spectrometers. While the first instrument combines quadrupoles for mass selection and fragmentation with a ToF mass analyzer, the latter instrument relies on a combination of a linear ion trap, which can trap, isolate, fragment and scan ions, with an orbital cell named Orbitrap. The following sections are focused on Orbitrap instruments as these were used throughout this thesis. Since its introduction in 2005 (36), this instrument type has undergone ongoing substantial improvements and nowadays allows the application of highly sophisticated acquisition modes as described further below.

Orbitrap Mass Spectrometers

Mass resolution and mass accuracy are key parameters in mass spectrometry. The introduction of a new concept of mass analyzer – the Orbitrap – in 2005, gave rise to instruments with performance characteristics previously only reached by extremely expensive Fourier transform ion cyclotron resonance -based instrumentation (37, 38). Orbitrap analyzers measure the frequency of ions oscillating along an inner central spindle and finally calculate the peptide mass spectrum by Fourier transformation (39-42). The axial frequency of oscillation of the ions on this trajectory is proportional to the square root of their m/z (43, 44).

In mass spectrometry the resolution is usually defined as the full width of the peak at half maximum (FWHM) divided by the mass of the peak (and therefore dimensionless) and indicates how well ions with nearly similar m/z can be distinguished in one scan. This means, the better the resolution the more ions can be distinguished and finally identified within the same spectrum. The resolution an Orbitrap analyzer can achieve is directly proportional to the transient time, meaning the longer ions are oscillating along the inner spindle of the Orbitrap mass analyzer, the better the resolution. The resolution is not constant over the entire m/z range of the acquired spectrum but decreases proportionately to the square root of m/z. Routinely, mass spectra are acquired with 30,000 to 60,000 resolution at m/z 400, which requires 0.5 to 1 sec transient time depending on the Orbitrap type used in this thesis. Besides resolution, mass accuracy also plays an important role in mass spectrometric analysis. The more exact a peptide or fragment mass can be measured the higher the likelihood of correct identification of the underlying peptide, since less false positive candidate possibilities arise during database search. Orbitrap mass analyzers now allow the mass determination in sub-ppm range (41, 45).

The Orbitrap cell itself is a pure mass analyzer lacking capabilities for ion isolation and fragmentation. The first Orbitrap mass spectrometer was thus introduced as a hybrid instrument coupled with a linear ion trap – the LTQ Orbitrap (36). This combination enabled ion accumulation, isolation and fragmentation in the linear ion trap, which can also be used as a highly sensitive and fast mass analyzer by itself, however with drastically reduced mass accuracy and mass resolution. In contrast, as already stated, in the Orbitrap analyzer ions are measured with high resolution and accuracy, but the acquisition generally requires more time. The general architecture of a LTQ Orbitrap mass spectrometer is depicted in Figure 5. For MS-based proteomics experiments the LTQ-Orbitrap can be operated in a parallel mode, where the masses of the intact peptide ions are accurately recorded in the Orbitrap while isolation and fragmentation of selected precursor ions and acquisition of the fragment spectra happens simultaneously in the linear ion trap. This acquisition mode is also called "high-low strategy" due to the high mass resolution and accuracy for the MS spectra and low resolution for MS/MS spectra.
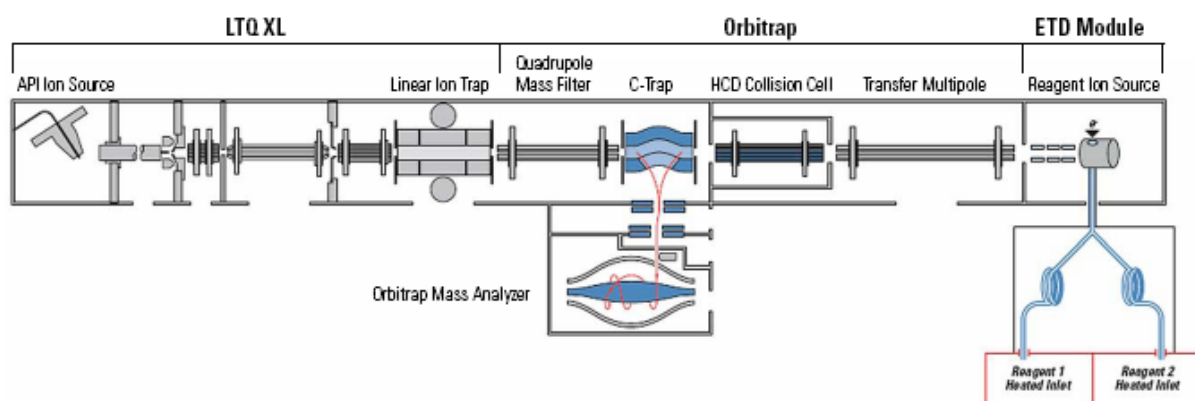
Figure 5: Schematic architecture of an LTQ Orbitrap ETD mass spectrometer. Central parts are the ion source for generating the analyte ions by ESI, the linear ion trap for trapping, isolating, fragmenting and scanning the ions to record a fragment spectrum. The C-trap links the ion trap to the HCD collision cell or to the Orbitrap cell, which is the central mass analyzer of the instrument. The reagent ion source at the rear end of the machine is responsible for generating radical anions for the ETD fragmentation process (Thermo Fisher).

The fragmentation process that is performed in the liner ion trap is called collision induced dissociation (**CID**). Ions are collided with an inert gas (such as helium) at low pressure in this trap. Newer LTQ-Orbitrap generations (LTQ-Orbitrap Velos) contain a dual ion trap cell - the steps of isolation, fragmentation and detection are hereby split, since these processes perform best at different pressures. After isolating a specific peptide ion species by resonant ejection, excitation by an RF field increases the internal energy of the peptides. This finally results in fragmentation primarily at the peptide bonds generating b- (N-terminal) or y-ions (C-terminal) as indicated in Figure 6.
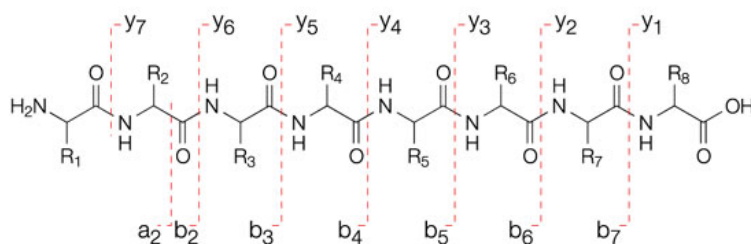
Figure 6: Fragmentation along the peptide backbone generating N-terminal b- and C-terminal y-fragment ions. Adopted from (8).

Since peptides break at the most labile bonds in CID, fragmentation can lead to uninformative spectra containing only few fragments, for instance indicating loss of a labile posttranslational modification (PTM). To compensate for losses which occur frequently, such as water losses, specific strategies, such as 'wideband activation', have been developed and implemented in the acquisition

software. Analysis of PTMs often requires more sophisticated features such as higher order fragmentation modes (MS^3) (46) or multistage activation (47), which will not be considered further here. Beside reduced resolution and mass accuracy a further drawback of fragmentation in ion traps is the so-called 1/3 cut-off, which refer to the fact that fragment ions which have a m/z value below 1/3 of the m/z of the parent ion are not stable in the ion trap and therefore cannot be detected. An alternative fragmentation mode in the linear ion trap named pulsed Q dissociation (PQD) allows the acquisition of fragmentation spectra without the loss of information in the low mass region (48).

To overcome the limitations of ion trap fragmentation spectra, a second fragmentation type – higher energy collision dissociation (**HCD**) - was introduced. HCD has been established first on the LTQ-Orbitrap instrument in the C-trap (49), which is located between the linear ion trap and the Orbitrap analyzer (see Figure 5) and which is normally responsible for storing the ion beam before sending the focused ion package into the Orbitrap cell. In HCD, selected ion populations were fragmented and stored in this compartment. Several drawbacks of this initial setup led to the construction of a dedicated collision cell which was placed behind the C-trap and termed HCD collision cell. HCD resembles quadrupole-like fragmentation and, like CID, mostly generates b- and y-ions but unlike CID, it also produces many internal fragment ion (50). In contrast to CID, there is no limitation on the low mass range of the fragmentation spectrum. Consequently, low mass ions such as a2 and b2 ions (see Figure 6) as well as immonium ions can be detected in the fragmentation spectra improving peptide identification. Moreover, labile PTMs such as phosphorylation stay mostly intact during the fragmentation process (50). After fragmentation, ions are sent from the HCD cell via the C-trap to the Orbitrap analyzer and are measured with high resolution (7,500 – 15,000). This strategy is therefore called "high-high" mode. The application of this acquisition type became even more popular with the introduction of the LTQ-Orbitrap Velos in 2009 (51). This new generation instrument comprised better ion transfer due to improved ion optics in the front part of the instrument and consequently shorter ion accumulation times and scan cycles. In addition, more efficient fragmentation in the new combined C-trap-HCD cell as well as better transfer of fragment ions was achieved. Recent studies showed that HCD fragmentation on phosphorylated peptides reaches at least equal identification rates without the need of specific features to track almost all labile modifications (52).

Besides the HCD cell a further capability was added to the LTQ-Orbitrap mass spectrometer, enabling a fragmentation type called electron transfer dissociation (ETD). ETD relies on a completely different physical mechanism than CID or HCD. It was originally invented in the form of so-called electron capture dissociation (ECD) and applied in Fourier transform instruments (53). Both ECD and

ETD are based on a radical induced fragmentation process generating c- and z- fragment ions in contrast to the CID or HCD process which generate predominantly b- and y-ions (see Figure 7) (54, 55).
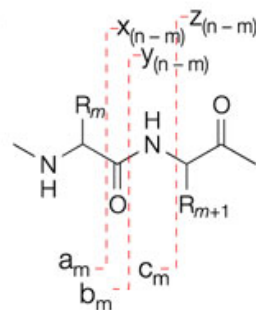

Figure 7: Alternative fragmentation along the peptide backbone generating N-terminal a-, b- and c-ions and C-terminal x-, y- and z-fragment ions. Adapted from (8).

The ETD process is generally induced by fluroanthene ions reacting with electrons to generate radical anions at the rear end of the HCD cell (see Figure 5). These are sent to the linear ion trap of the instrument where they react with the peptides. Fragment ions can in principle both be read out in the LTQ or the Orbitrap. ETD is mostly favorable for higher charged ions since one charge is neutralized by the reaction (56). Moreover, activation times decrease strongly with higher charge states, which recommend the technique for top-down analyses. Since labile PTMs remain linked to the peptide during fragmentation, this fragmentation type is especially attractive for the analysis of very labile PTMs such as o-linked ß-N-acetylglucosamine (o-GlcNAc) or phosphorylations (57, 58).

Recently, a further new generation of the LTQ-Orbitrap mass spectrometer family, the Orbitrap Elite, was introduced featuring improvements in resolution, mass accuracy and scanning speed (59). Additionally, a new benchtop type of instrument, the Q Exactive, was recently developed (60). It combines a quadrupole for ion isolation with an HCD collision cell which is linked via the C-trap to the Orbitrap mass analyzer. This instrument exclusively allows HCD fragmentation and therefore always produces data in the high-high mode. Scanning speed and sensitivity is unsurpassed as the machine allows parallel fragmentation and MS/MS analysis, all with short transients due to its high resolution.

1.1.2. Quantification techniques in mass spectrometry

Besides identification of peptides or proteins by MS, another important aspect in the application of MS-based proteomics is the quantification of the analyzed proteins. However, this also represents one of the most challenging tasks in proteomics (61). Quantification in proteomics often aims at the comparison of different cellular states e.g. due to different stimuli or changing environment conditions. In this respect, it was fundamental to develop methods which allow the direct comparison of protein abundances between different states. Starting in 1999 with developments from independent groups (62, 63), over the last years the field has developed a variety of different technological approaches for the quantitative analysis of proteomes by mass spectrometry (64).

Most approaches rely on the differential incorporation of stable isotopes into the proteomes to be compared or on the use of isotopically labeled reference peptides. The advantage of using stable isotopes lies in the fact that no changes of the physicochemical properties of proteins or peptides are introduced besides generating a mass shift between the differently labeled isotopic analyte species. Consequently, corresponding proteins or peptides behave nearly identically during the processing steps - such as pre-fractionations or chromatographic separations prior to MS analysis - and are finally simultaneously detected in the mass spectrometer. Quantification is achieved by comparing the mass spectrometric responses of the corresponding isotopic peptide forms. The isotopic information can already be introduced into the proteins metabolically during cell culturing. Alternatively, proteins or peptides are labeled chemically or enzymatically at later stages of the sample processing workflow. Labeled peptides or even proteins can also be spiked into the sample which allows for absolute quantitation if the amount of the spiked reference is known. In addition, label-free approaches have been established which correlate mass spectrometric signals between different LC-MS runs. The different strategies for quantitative proteomics are summarized in Figure 8 and will be described in more detail in the following sections with emphasis on metabolic labeling techniques which have been used in the projects of this thesis.
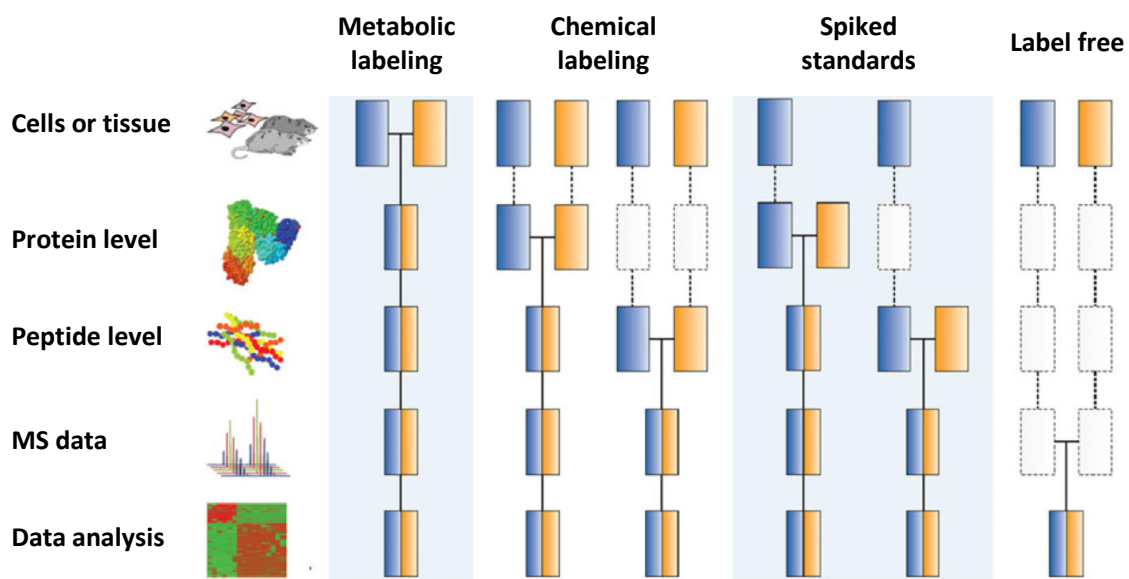
Figure 8: Schematic overview of the different quantitation modes based on isotopic labeling. It is indicated at which stage the samples to be compared are combined for further mass spectrometric analysis. Adopted from (64).

1.1.2.1. Absolute quantification

In recent years, several improvements have been made in the field of absolute quantification of protein amounts. In approaches using labeled standards, a stable-isotope labeled control whose concentration is exactly known is mixed with the sample of interest. Absolute concentrations can then be derived via the intensity ratios detected in the mass spectrometer. In absolute quantification approaches, generally information is obtained for only a limited subset of proteins in contrast with whole proteome-based relative quantification.

Based on a publication dating as far back as 1983 (65), absolute quantification was introduced to the proteomic field via so-called AQUA peptides (absolute quantification of proteins (66)). These synthesized isotopically labeled peptides are mostly spiked into the sample at the peptide level. However, the earlier the standards are added to the sample during the sample preparation process the smaller the quantitative errors of the measurement. An approach known as QconCAT (67) goes beyond adding single peptides, by genetically constructing synthetic genes that express concatenated standard peptides. These are cleavable by tryptic digestion resulting in several peptides of either one or more proteins of interest. Consequently, these constructs can already be added at the protein level to the sample and thereby avoid possible discrepancies arising during protein digestion. More recent approaches are PSAQ (protein standard absolute quantification), absolute SILAC and FlexiQuant, which do not use peptide standards but utilize metabolically labeled

full-length proteins expressed in cell-free or bacterial systems, reviewed in (64). These standards can be added directly to the cell lysates allowing fractionation of the sample without drawbacks on the quantification. Recently, our laboratory has developed the PrEST approach (68), where protein epitope signature tags (PrESTs) are produced to evaluate absolute numbers of proteins in higher throughput since many PrESTs can be combined in one samples. These PrESTs are metabolically labeled (see below) and are mixed to the protein sample before digestion.

1.1.2.2. Relative quantification

Relative quantification, in contrast to absolute quantification, aims for the identification of relative changes between different samples. In this respect, the emphasis lies on a global quantitative analysis of complete proteomes or sub-proteomes between different conditions. Different methodologies have been developed for relative quantitation in MS-based proteomics experiments over the last years.

Metabolic labeling

Metabolic labeling relies on the incorporation of different versions of stable non-radioactive isotopically labeled amino acids into all of the proteins of the cell during the cell culturing process. One of the most popular metabolic labeling methods is stable isotope labeling with amino acids in cell culture (69) and this SILAC technology has been used in this thesis. Whereas one culture medium contains normal amino acids (light condition), the other one contains certain essential amino acids exclusively in their heavier isotopic forms (heavy condition). In most experiments, arginine and lysine with ^{13}C and/or ^{15}N substitutions are added. When trypsin, which exclusively cleaves C-terminally after arginine and lysine (7), is used as the protease, each generated peptide should exist in the different isotopic forms (except the C-terminal peptide of the protein) originating from the conditions to be compared.
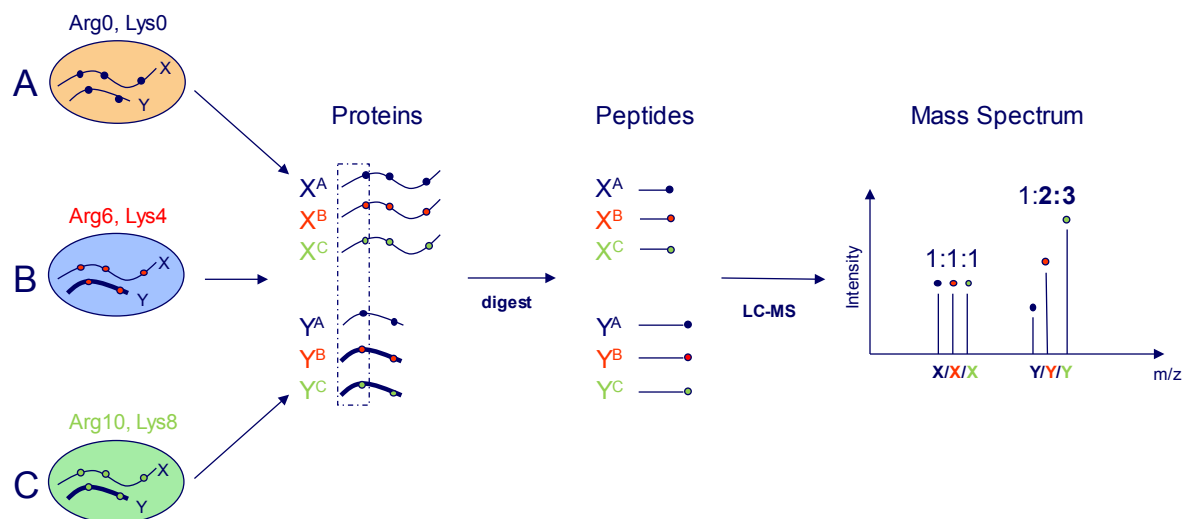
Figure 9: Triple SILAC labeling workflow. Different forms of isotopically labeled arginine and lysine are incorporated during cell culture. Labeled proteins are extracted and digested with trypsin. Peptides are subsequently analyzed in the mass spectrometer. Quantitative information is derived from the intensity ratios between the different isotopically labeled versions of the peptide (light, medium and heavy – A, B, C).

For SILAC labeling, different cell populations are cultured in light and heavy medium, respectively (see Figure 9). When full incorporation of the labeled amino acids has been achieved, cell cultures can be compared, usually after stimulation of one condition. Both conditions are merged either directly at the cell level or immediately after cell lysis. Thus, proteins from the different SILAC states are processed together, including the proteolytic digestion step that generates the peptides. Errors due to different sample handling are thereby avoided. Consequently, high quantification accuracy and high precision can be obtained. As stated above, peptides containing different forms of the isotopic label do not differ in their physicochemical properties but only in their mass. Thus, peptide ratios can be calculated by comparing the different intensities of the light and heavy peptide forms detected simultaneously in the mass spectrometer (70, 71).

Besides the combination of two samples, three different conditions can also be compared in a so-called triple SILAC labeling experiment by introducing a third stable isotopic amino acid (see Figure 9). Higher number comparisons can be performed by combining several triple labeling experiments with a shared reference condition (72, 73). This strategy can be applied for time-course experiments, for instance.

Instead of the incorporation of isotopically labeled amino acids, metabolic labeling can also be performed by directly feeding the organisms with stable isotopes, e.g. ^{15}N, which is mostly used for microorganisms (63). In contrast to SILAC labeling, which ideally leads to the generation of peptides containing only one modified amino acid, ^{15}N labeling ideally generates fully ^{15}N labeled organisms,

consequently generating peptides that vary in the number of incorporated ^{15}N isotopes (61). This drawback combined with high reagent costs and experimental timelines limits the ^{15}N method, especially for mammalian systems. Nevertheless, several studies have been performed using this method, which are more comprehensively reviewed in (74).

One modification on SILAC is the pulsed SILAC method (pSILAC) (see Figure 10). Cells are cultured in light medium which is exchanged at a certain time point against heavy SILAC medium. After a predefined incubation time, proteins are analyzed to measure turnover rates in a global manner (75, 76). In an extended strategy by the Selbach group a differential treatment is incorporated (77). Cells are first cultured in light medium and upon differential treatment they are transferred to medium including either medium or heavy labeled amino acids. After a certain incubation time, cells are combined, harvested and analyzed. Newly synthesized proteins all occur in the M or H states. The H/M ratio reflects changes only induced by differential treatment. A further approach additionally takes into account possible degradation of proteins and recycling of the amino acids into the synthesis pathway which might otherwise result in underestimation of protein turnover rates (78). Over the last few years, pSILAC has been applied in many studies in different formats (79-81).

One major drawback of the standard SILAC labeling approach is its limitation to cultured cell lines, although, whole organisms such as mice and flies have also been successfully metabolically labeled by using special diets which contain heavy forms of amino acids (82, 83). A more general approach uses SILAC in a spike-in or super SILAC format (84-86) (84). It extends the classical SILAC approach to tissue analysis without metabolically labeling the organism itself. In short, a heavy labeled SILAC standard is generated from different SILAC-labeled cell lines, which are then used as an internal reference to quantify proteins in each specific sample (see Figure 10). Ideally, the mixture of cell lines covers the complete proteome of the tissue that is to be analyzed. For this reason, the standard needs to be tested for the best combination of different tissue-resembling cell lines, to guarantee optimal identification and quantification results. In the next step, the heavy super SILAC standard is spiked into different tissue samples in equal concentration. For the analysis, ratios between the tissue and the standard are quantified, often followed by a ratio-of-ratio quantification to unravel differences between different tissue samples. This approach not only extends the SILAC technology to samples to which it has not been applicable before, but also allows the direct comparison of many samples in the same project.
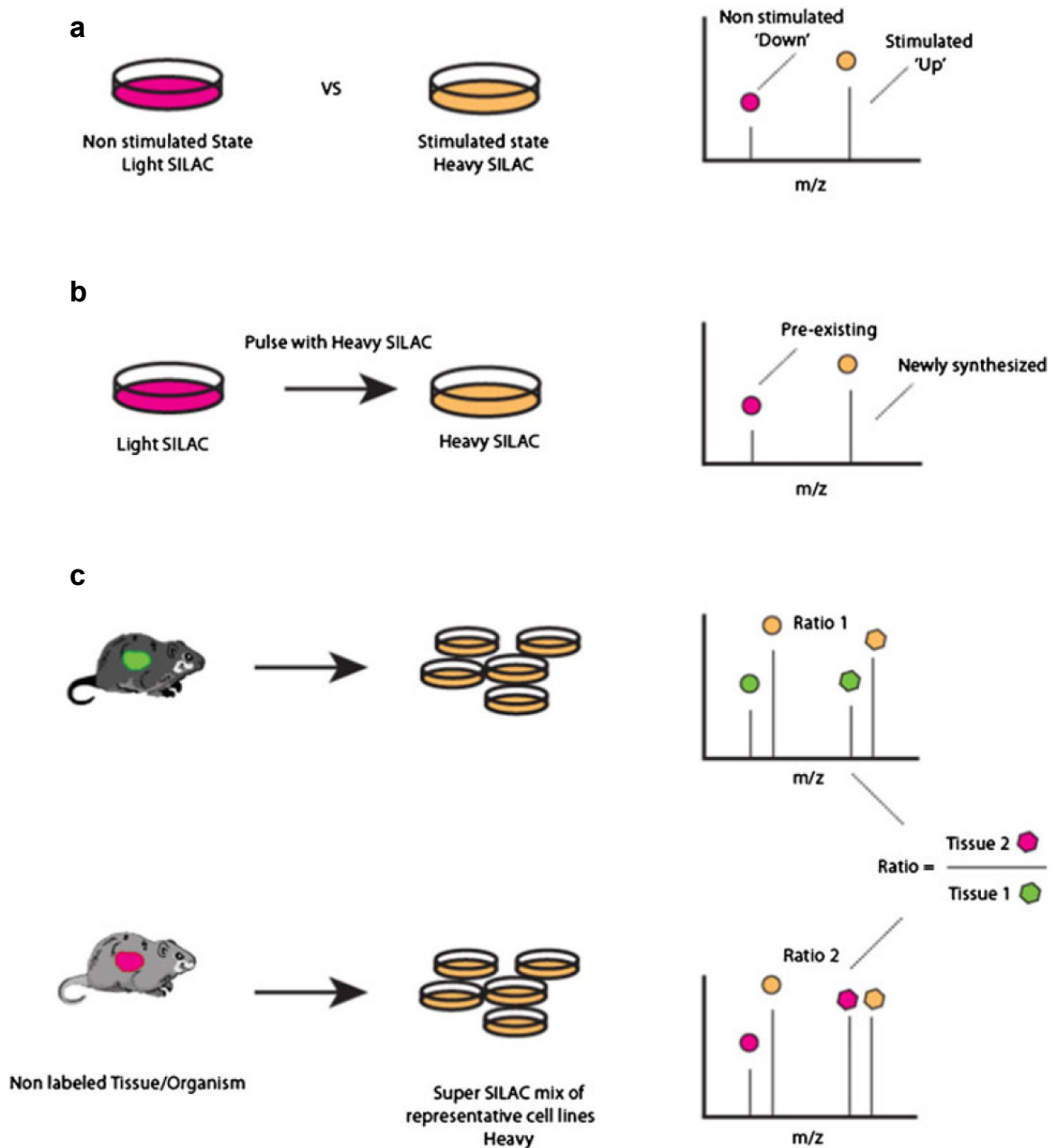
Figure 10: Different SILAC approaches. (a) Standard SILAC approach with heavy and light conditions. (b) pSILAC experiment. (c) Super SILAC approach. Adopted from (64).

Chemical labeling at the protein and peptide level

Chemical labeling methods can be generally applied to every sample source (e.g. tissue) and are therefore advantageous when no metabolic labeling can be performed or when several different samples need to be compared. Until the labeling step – which can either be performed at the protein or peptide level - samples have to be processed in exactly the same manner in parallel,

[17]

which can contribute to quantitative errors. In principle, any reactive group can be modified by the chemical label, nevertheless, most approaches aim for targeting free amino groups -the ε-amino groups of lysines or the N-terminus of the peptide/protein. As soon as samples have been labeled, they can be mixed and further processed together. The most prominent chemical labeling techniques are tandem mass tags (TMTs) and isobaric tags for relative and absolute quantification (iTRAQ) which are both applied at the peptide level (87-89). They target primary amines and are isobaric tags, meaning that all versions of the tag have the same overall mass. However, certain isotopes are localized in the tag at different positions and contribute either to the reporter group or the mass-normalizer group (Figure 11). Consequently, peptide species from differentially labeled samples appear in one peak in the mass spectrum. When peptides are isolated and fragmented, the tags break and generate distinct reporter ions originating from the different samples in the lower mass region of the fragment spectrum.
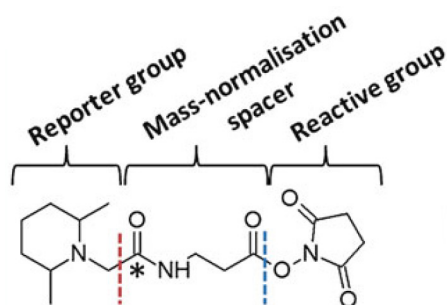
Figure 11: Chemical structure of one isobaric 2-plex TMT label. The reactive group covalently links to primary amine groups (blue dashed line). The mass normalization spacer balances the mass of the reporter group which is substituted with different isotopic forms of C and N. The isotopic modification is indicated by an asterisk. In the corresponding TMT label, the isotopic modification is located within the reporter group and not the mass-normalization spacer. Upon fragmentation the reporter group is released (dashed red line) and the different isotopic reporter forms are recorded as a quantitative measure. Adopted from (90).

Depending on the intensity of the different reporter ions, relative quantitative information of the peptides can be derived. Since all corresponding forms of each peptide elute at the same time, several different conditions (up to eight in the latest iTRAQ version) can be compared without any increase in complexity of the MS spectrum. The same holds true for the MS2 spectrum, which only changes with regard to the number of reporter ions but not in peptide related fragment ions. For a long time, one drawback of the method was that the reporter ions are detected in the low mass region which is not frequently detectable in CID fragmentation due to the low mass cut-off (see section 1.1.1.3.). To overcome this limitation, fragmentation was performed with pulsed Q dissociation (PQD) in the linear ion trap (48, 91) as well as by HCD combined with read-out in the Orbitrap analyzer. Especially over the last few years the HCD method has been greatly improved and

is highly attractive for the analysis of chemically labeled samples since it generates high resolution fragmentation spectra in combination with low cycle times (see paragraph 1.1.1.3.) (49, 92). Although problems regarding the MS2 fragmentation are mostly solved, one common problem still exists concerning co-eluting peptides of similar mass, leading to co-isolation followed by co-fragmentation. The issue of the combined fragmentation of several two or more peptides lies in the generation of overlapping reporter ions in the same MS2 spectrum which ultimately will lead to suppression of reporter ion ratios and consequently wrong quantifications. However, also in this respect methods have been developed which try to overcome these limitations (93, 94).

Beside the above mentioned techniques, several further approaches have been developed. These will not be further discussed here, since in any case no chemical labels have been used in this thesis. Detailed information on further chemical labeling strategies can be found in (61, 64, 95).

Label-free approaches

Label-free approaches, as the name suggests, do not rely on isotopic labeling of proteins or peptides but aim at quantifying peptides and proteins from the MS signals alone, without chemical manipulation of the sample. In principle, label-free approaches are either directly based on peptide intensities derived from different LC-MS runs or they extract the quantitative information indirectly for instance via spectral counts. Label-free approaches represent a specific, highly complex field of computational data analysis and several successful applications of label-free algorithms have been introduced (96-99). Label-free methods are mentioned here to give a complete picture on quantitative proteomic approaches, nevertheless, they will also not be discussed further since this sort of quantitative method was not used in this thesis.

1.1.3. Bioinformatics analysis

The last but no less important step in an MS-based proteomics experiment is the analysis of the enormous amount of data generated by the mass spectrometer with respect to protein identification and quantification. Over the last years several software tools have been developed that allow the identification of the proteins from the generated MS and MS/MS spectra such as Mascot (95) and SEQUEST (100) just to name the most prominent ones. However, software tools which allow for a comprehensive analysis and interpretation of the acquired data by combining protein identification with robust quantification algorithms, statistical data assessment and intuitive data visualization tools were lacking for a long time. With the development of software packages such as MaxQuant (45), large scale quantitative MS-based proteomics experiments can now be analyzed in an automated way in hours to days, a process that needed weeks to months with conventional pipelines available before.

Some principles of bioinformatic analysis of mass spectrometric data are highlighted in the following based on the MaxQuant environment which has been used for all analyses covered within this thesis (see Figure 12).
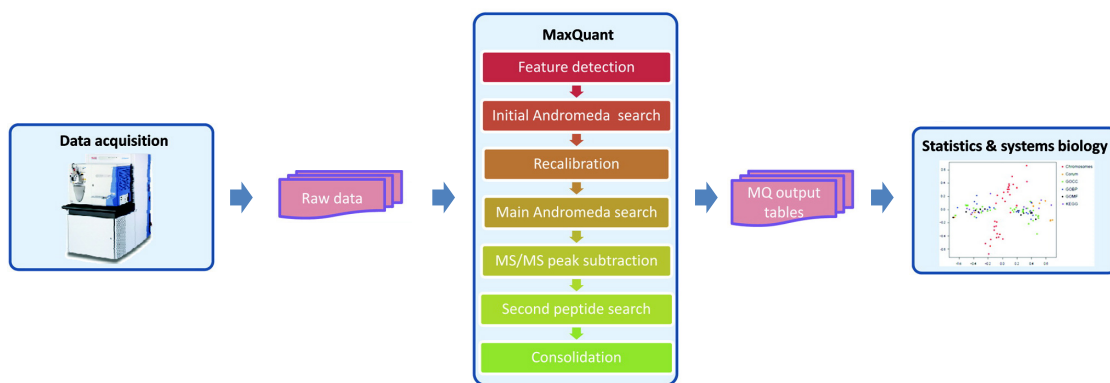
Figure 12: MaxQuant workflow with the integrated Andromeda search engine. Adopted from (101)

In a first step, MS raw data are converted into peak list files, combining the information of the precursor ion, which has been fragmented, with the information about fragment peaks and corresponding intensities. Moreover, in SILAC experiments, the MS scans are filtered for the characteristic mass gaps between the different isotopic states of the detected peptides thereby classifying peptide populations based on their isotopic SILAC state which facilitates the identification and quantification process. A first re-calibration step can be performed if no lock mass calibration

(36) was performed during the mass spectrometry analysis. Accurate masses are extracted by peak detection in a 3D matrix of retention time, m/z and corresponding intensities. The more intense a peak, the more accurate the mass can be calculated. Since MaxQuant integrates the signal over the whole chromatographic peak rather than deriving the relative quantities only from one single MS scan (Figure 13), the mass detection is highly accurate (45, 102).
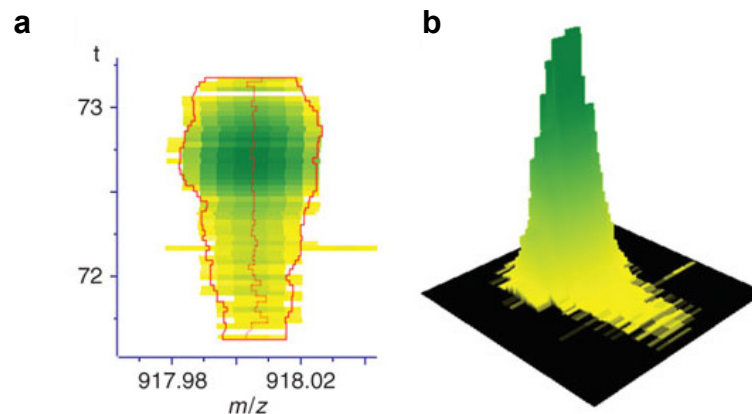
Figure 13: Peak detection in MaxQuant. Data are displayed in 3D, recording m/z values, retention time (t) as well as intensity (color code representing low intensities in yellow to high intensities in green) of any detected ion. (a) 2D representation including color code. Accurate masses can be calculated performing intensity-weighed algorithms. (b) 3D view of an elution profile. Adopted from (45).

The information on the fragmented peaks is further processed in the second main step, the identification of the peptides. The MaxQuant software framework integrates a search engine called Andromeda (101) and peak lists are directly submitted for the database search. Databases need to be defined beforehand, as well as information about the specific labeling, the used enzyme, fixed and variable modifications to mention just some parameters. Based on this information, candidate peptides can be extracted for the identified spectra by directly matching the MS/MS spectra to the supplied database. In other software packages, different identification strategies can be used, such as de-novo-sequencing of sequence-tag strategies, however, matching to a common database is the most common. When including a reversed database into the search, false-positive rates can be calculated following a so-called target-decoy strategy (103, 104).

The peptide identifications generated are further statistically evaluated in a third step and protein groups are assembled and quantified. For the calculation of protein ratios the median ratio (log2 scale) of all corresponding peptides is used. False-discovery rates are calculated for both peptides

and proteins based on the target-decoy hits and used to filter the data. Moreover, for the quantification, significances for fold changes are calculated and summary statistics are computed.

Generated lists of peptides and proteins are then used for downstream analyses. In a first step, general information about the quality of the experiments are typically extracted including reproducibility of replicates, overall identified and quantified proteins or modified peptides as well as general information about the identification rates of the spectra. When these results are satisfactory, identified proteins can be used to evaluate the coverage of certain pathways and quantitative changes can be further inspected. The significant outlier populations, which represent changes due to the different conditions in the experiment, are of special interest. Protein interactions within this population can be extracted as well as pathways which are especially affected by the treatment. In this respect, annotational databases such as Gene Ontology (GO; (105)) or pathway databases as KEGG (106) are required to gain information about enrichments in the regulated fraction. Several bioinformatic tools from microarray analysis can in addition be refined for proteomic data (107).

Nevertheless, experiments in the mass spectrometric field are quite heterogeneous so that no single common standard workflow for data analysis can be provided. Generally each analysis needs to be adapted to the underlying experimental question. For example in one topic of my thesis, the global apoptosis analysis, an additional MATLAB script has been written to further analyze the MaxQuant output data. In that script, mass spectrometric data were used to generate a 3D cleavage plot of the data based on the identified peptides including their quantitative ratios, their localization along the protein sequence and the localization in the SDS gel before mass spectrometric analysis (see paragraph 2.2.).

Furthermore, mass spectrometric data can be stored in databases to combine different datasets and to allow the extraction of general information derived from different experiments. One recently published database is MaxQB (108), which allows not only storage of information but also experiment-wide search for common characteristics such as identified peptides. Within the database all information from the single experiments are stored such as annotated MS/MS spectra, but MaxQB also allows the search of general features in all experiments. Moreover, general information on the proteins is linked to the database such as information derived from Uniprot.

1.2. Aneuploidy

Aneuploidy (an - not; eu - good; ploos - fold) describes the abnormal chromosomal state of a cell in which the chromosome number does not correspond to a multiple of the haploid chromosomal number. For instance, if one copy of a specific chromosome is present, this is called a monosomy for that chromosome, and similarly for trisomies etc. (109) (see Figure 14). In a strict definition, genomic alterations based on sub-chromosomal regions are not contained in the term "aneuploidy", nevertheless, researchers often include them as "partial" or "segmental" aneuploidies. Besides chromosome gains, losses are also common, however, this phenomenon will only be addressed here in passing.
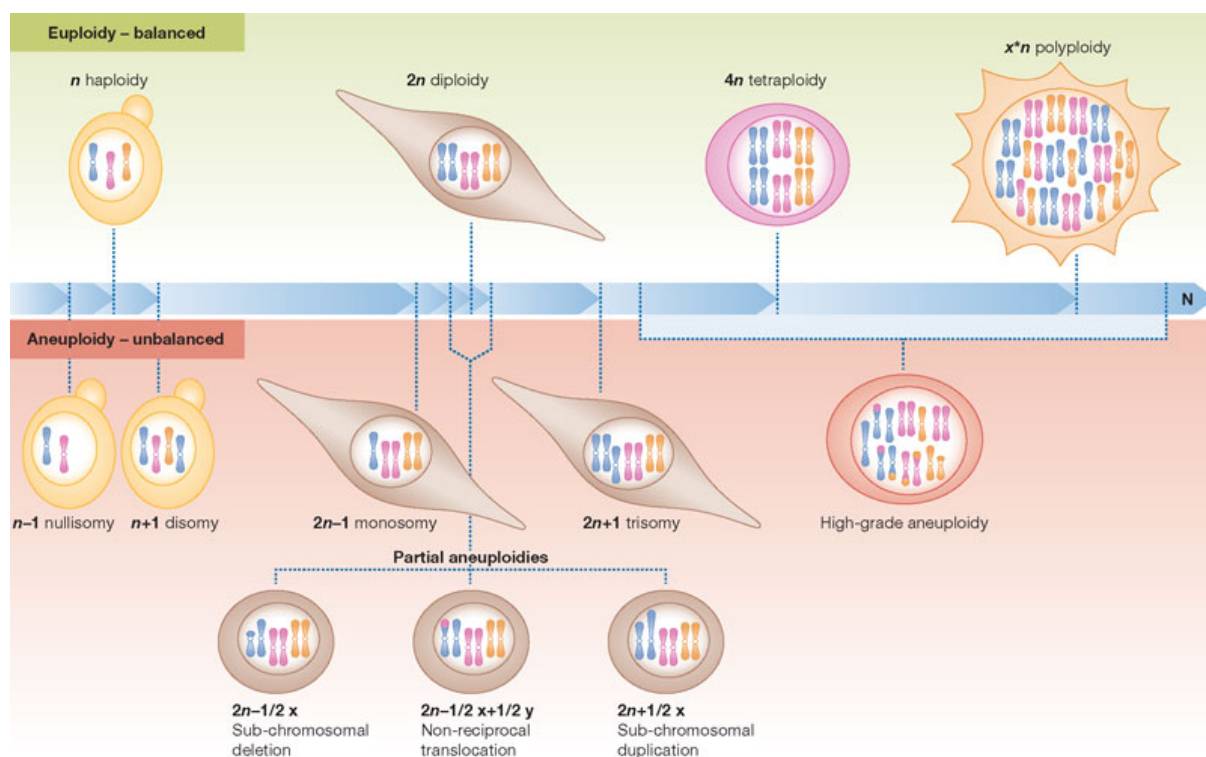
Figure 14: Graphical representation of aneuploidy and euploidy. Adopted from (110).

In contrast to aneuploidy, euploidy describes a cellular state with a multiple of the haploid content (see Figure 14). Diploid, triploid or tetraploid chromosome sets and even higher folds are observed (polyploid). In contrast to the unbalanced state of aneuploidy, the euploid genomic state represents a balanced condition and is much better tolerated by the cell than aneuploidy.

Generally, aneuploidy is associated with disease and developmental abnormalities. All aneuploidies in living humans are associated with rather small chromosomes, namely chromosome such as 13, 18 and 21. All other aneuploid states do not survive until birth, including higher multisomies, different

chromosomes or chromosome losses. Trisomy 18, better known as Edwards syndrome, has a low percentage of live births and is fatal in about 90% of the cases in the first year of life due to severe cardiovascular and brain defects. Likewise, survival rates in the first year are very low for the Patau syndrome, a trisomy of chromosome 13 which occurs very rarely. One widely known case of aneuploidy in humans is trisomy 21, better known as Down syndrome, in which chromosome 21 exists in three instead of two copies. It has severe effects in the human organism, leading to mental retardation, heart defects and onset of Alzheimer´s disease significantly earlier compared to diploid individuals. Nevertheless, survival rates are very good though individuals have reduced life expectancy.

Although many affected individuals possess a complete set of aneuploid cells, aneuploidy can also occur in only some cells of an individual. So-called mosaicism occurs in rare cases and is related to a systematic predisposition to mitotic nondisjunction. In this special case aneuploid cells are not in meiosis generated but in later steps of embryogenesis and individuals consequently have only a certain percentage of aneuploid cells, whereas the other cells are the normal isogenic diploid counterparts (see 1.2.2.). Often, individuals with mosaicism are predisposed to cancer.

Not only autosomal chromosomes are affected by aneuploidy but sex chromosomes can occur also in multiple copies. These aneuploidies are much better tolerated by the cell perhaps because mechanisms of dosage compensation are established and they result in milder consequences for the affected individuals. Monosomy of the X chromosome (Turner syndrome) is tolerated for the lifetime of viable organisms, however with stronger symptoms than for individuals with chromosome gains in sex chromosomes (Klinefelter´s (XXY) and Triple X (XXX) syndrome) (111).

Independent of any disease, some tissues are naturally aneuploid, such as some percentage of cerebral neuroblasts in embryonic brain of mice and humans (110, 112). Aneuploid cells also exist in adult brains and seem to be functional and integrated into the neural circuitry. Mammalian hepatocytes show aneuploidy as part of normal development and as response to ageing, for instance. In general, chromosome losses seem to be more prominent than chromosome gains. Many open questions are still to be elucidated - such as how apparently detrimental karyotypes can existent in some tissues whereas they are not observed in others and whether or not they have specific functionality.

1.2.1. Origin of aneuploidy

An interesting question in the field of aneuploidy research is how aneuploidy develops starting from a normal cell. During the mitotic cell cycle chromosomes condense and sister chromatids align. Chromatids are linked to each other by cohesins, which hold the strands together, and are also bound to microtubules, which attach them to the mitotic spindle poles. The spindle-assembly checkpoint (SAC) controls for correct connections and tension before proceeding in the cell cycle. After this step, the anaphase-promoting complex or cyclosome bound to its activating subunit Cdc20 (APC/C-Cdc20) can degrade an inhibitory subunit of the separase protease. Subsequently, separase cleaves cohesin and thereby initiates chromosome segregation. This simplified representation already illustrates that out that malfunctions can occur at many steps of this cascade, for example due to improper tension or missegregation and thereby contribute to aneuploidy (111). Improper mitosis might be caused by the disruption of key regulatory genes and pathways, ultimately resulting in cells with losses or gains of chromosomes (113). Specifically, loss or inactivation of cohesin, deregulated activity of separase or defects in the spindle assembly checkpoint but also merotelic attachments, the connection of a single kinetochore to microtubules from both poles, can be reasons for missegregation. Merotelic attachments can occur due to increased numbers of centrosomes or due to hyperstabilized kinetochore-microtubule attachments (113) (see Figure 15).
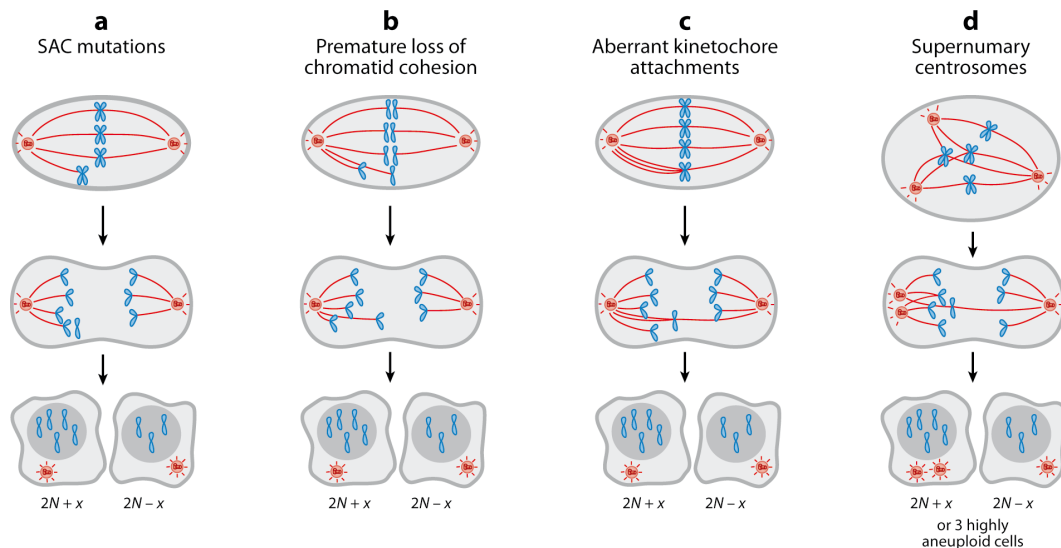

Figure 15: Representation of cell division errors leading to aneuploidy. (a) Mutations in the SAC complex. (b) Cohesin defects due to loss or mutation. (c) Merotelic attachment due to aberrant kinetochore attachments. (d) Supernumerary centrosomes leading to merotelic attachment. Adopted from (111).

In most cases aneuploidy arises already within the oocyte during maternal meiosis, either in meiosis I or II, most probably due to similar reasons as in mitosis (see Figure 16). Improper segregation results in aneuploid gametes leading in the end to entire aneuploid organisms. This might also explain the direct correlation of increased maternal age and elevated occurrence of aneuploidy (109). However, detailed underlying mechanisms for the origin of aneuploidy still remain elusive. Most probably not only one single effect but an accumulation of several different events leads to the occurrence of aneuploid cells (109).
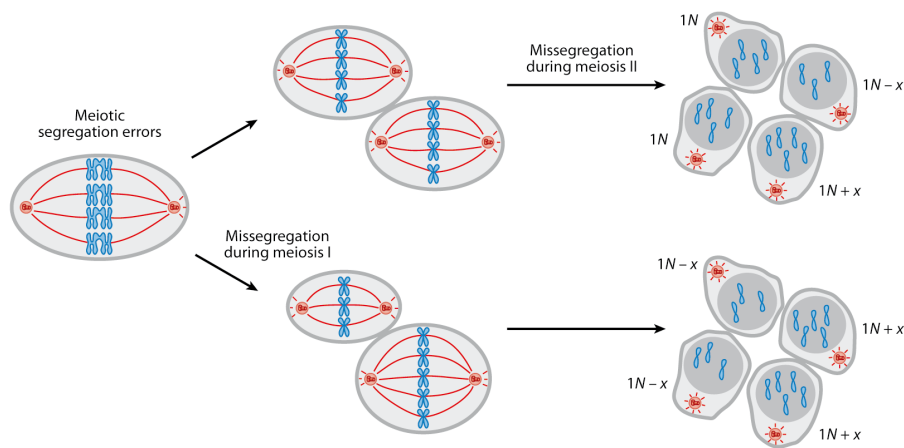
Figure 16: Aneuploidy by missegregation of homologous chromosomes in meiosis I or of sister chromatids in meiosis II. Adopted from (111).

1.2.2. Generation of aneuploid cells and detection of aneuploid states

Aneuploidy models can be derived by several different methods (see Figure 17). In the last years a variety of aneuploid clones have been generated in yeast strains. Aneuploid models of human cell lines and MEFs have also been established. Random aneuploidies can be generated by mutations causing increased chromosomal missegregation, for example by inactivation of proteins of the chromosome segregation machinery (114) or the SAC complex (115). Moreover, aneuploidy can be induced through meiosis of cells with an odd ploidy (116), however, these strains are known to be rather unstable. In contrast to randomly generated aneuploidy, defined stable aneuploid clones can be generated through single-chromosome transfer (such as micronuclei transfer (117)) or meiotic nondisjunction. In microcell-mediated chromosome transfer (MMCT), single chromosomes are transferred from a donor strain into a recipient strain (Figure 17a). First, micronucleation induced in the donor cells is followed by enucleation and formation of microcells. Purified microcells are

subsequently fused to recipient cells. Since the additional chromosome harbors selectable markers, aneuploid cells can be selected by culturing in medium containing antibiotics (113, 118). A further approach (119) takes advantage of Robertsonian translocations (Figure 17b) to generate aneuploid mouse embryonic fibroblasts via meiotic nondisjunction. Tissue samples from aneuploid individuals such as Down syndrome patients can also be used to generate aneuploid models (Figure 17c). In this regard, an interesting example is the isolation of cells from individuals with mosaic aneuploidy. In this case, both diploid and aneuploid cells can be extracted from the same individual allowing direct comparison of these isogenic cells (113).
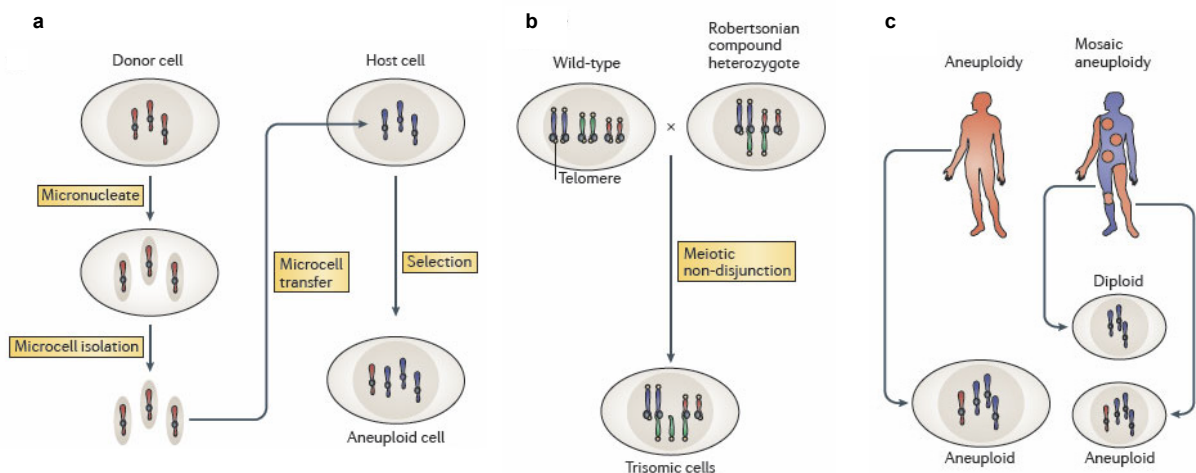
Figure 17: Representative examples for the generation of aneuploid models. (a) Microcell-mediated chromosome transfer (MMCT). (b) Strains carrying Robertsonian translocations are mated with wild type cells generating trisomic strains by meiotic nondisjunction. (c) Tissue samples from aneuploid human individuals can be used to study aneuploidy. Samples from individuals with mosaic aneuploidy provide both aneuploid and corresponding diploid cells and can be used for direct comparison. Adopted from (113).

Not only the generation but also the proper detection of aneuploid states is of great importance for research in the aneuploid field. Several approaches have been developed either based on the visualization of chromosomes or direct identification of the chromosomal content. Karyotyping of the chromosome content is performed as a standard method (120, 121). Cells are arrested during division and chromosomes are stained and sorted according to their appearance. This method can only detect additional or missing chromosomes as well as strong changes in the karyotype. In multicolor fluorescent in-situ hybridization (multicolor FISH) (122), chromosome painting or spectral karyotyping (SKY) (123), sub-chromosomal changes can be highlighted (124, 125). In addition to different chromosome numbers, specific features of the chromosomes such as translocations can be observed. Besides these visualizing methods, there are other approaches for the direct measurement of the chromosome content (126). Comparative genomic hybridization (CGH) allows

the detection of copy number changes in the DNA content, though without information on structural re-arrangements. DNA from the specific sample and a reference sample are differentially fluorescently labeled and mixed. The mix is then hybridized to either normal chromosomes or (in the case of arrayCGH (aCGH)) to thousands of short defined DNA stretches (127). Fluorescent ratio signals can subsequently be extracted and regional differences at the chromosome level can be derived. A further approach is represented by SNP arrays, which are DNA microarrays that feature high resolution including the detection of single nucleotide polymorphisms (SNPs)(128).

1.2.3. Pathologies of aneuploidy in different species

Apart from mammals, aneuploidy has been observed and studied in many other organisms, which allows comparison across different species (111). As stated above, organismal aneuploidies are often lethal already during early development or – if not lethal – lead to substantial developmental defects. Very early observations in plants revealed poor growth of the aneuploid species in comparison to equivalent euploid plants. Moreover, a direct proportionality was observed between the additional chromosomal material and slower growth. Experiments in Drosophila melanogaster indicated that monosomy for the smallest chromosome IV can result in viable flies with smaller size and frequent sterility. Trisomies for chromosome IV show only slight morphological changes. Aneuploidies for all other chromosomes were, however, lethal (129, 130). These studies revealed that in general monosomies are much more pronounced than aneuploidies based on chromosome gains, a trend which seems to be conserved across all species (111). Studies on aneuploidy have also been performed in the nematode Caenorhabditis elegans. Both duplications and deletions can be tolerated to some extent, however, many forms of aneuploidy are lethal. Studies in budding and fission yeast show inhibited proliferation due to aneuploidy (131, 132).

Extensive research in mice revealed that for all autosomal chromosomes only trisomy of chromosome 19 is not embryonic lethal. Nonetheless, also these individuals die shortly after birth. Embryos containing different forms of aneuploidy show a variety of developmental abnormalities and retardations (119).

In human, aneuploidy is known to be linked to congenital birth defects and miscarriages. However, not all aneuploid cells result in cell death. As mentioned above, trisomy for chromosomes 13, 18 and 21 can to some extent be tolerated and the organisms survive embryogenesis (133-135). Several links between clinical disorders and chromosome abnormality in surviving human embryos are known. Constitutional trisomy leads to defects in development and increases the risk of specific

pathologies. Among others, frequent symptoms are stunted growth, mental retardations and decreased fertility.

All in all, throughout all species which have been studied, aneuploidy results in strong negative effects for the cells, leading in all organisms to severe phenotypes such as developmental abnormalities and growth retardations.

1.2.4. Aneuploidy in cancer and disease

Besides its causative role in disease and developmental abnormalities, aneuploidy is also a key characteristic of cancer (110), a situation in which cells have an increased proliferative potential and can evade elimination by apoptosis. More than 90% of solid tumors and 75% of blood cancers show some degree of aneuploidy (136, 137). In addition, it seems that there is a correlation of the stage of the cancer with increased aneuploidy (136). One obvious difference in cancer, however, is the accumulation of many multisomies within the same cell in contrast to aneuploid cells in Down syndrome or defined aneuploid clones generated by e.g. single-chromosome transfer. Cancer cells can, in rare cases, also contain exact multiples of the haploid complement, however, aneuploid states – ranging from "near-diploid" to "high-grade aneuploid" – are much more common. Interestingly, whereas in Down syndrome individuals the risk of leukemia is increased, the risk for solid tumors is significantly decreased. This raises interesting questions on the mechanistic linkage between aneuploidy and cancer.

Already more than 100 years ago, first observations had been made regarding a role of aneuploidy in cancer. This influenced the work of Boveri on sea urchins, in which he proposed that single aneuploid cells may cause cancer (138, 139). Nevertheless, the exact role of aneuploidy in cancer including whether or how it contributes to tumorigenesis is still not fully elucidated to this day. In fact, many observations in the field of aneuploidy and cancer appear to contradict each other. One question that is often raised is why so many tumors possess aneuploid states, if this is so deleterious in aneuploid organisms. As stated before, aneuploid cells often show severe reductions in cell proliferation, however, cancer cells must have evolved strategies allowing them to overcome this barrier (140).

Studies in mouse models indicate that aneuploidy may not only be a by-stander effect of the tumor, but that it may have direct tumor promoting effects. For instance, tumor formation in mice is increased by aneuploidy (114, 141, 142). However, in some cases, aneuploidy also suppresses tumor growth (114). Growth studies in yeast suggest that aneuploidy may facilitate tumorigenesis and that

aneuploid cells can be generated via tetraploid states (143). Under certain conditions, aneuploid states could also be a survival advantage, since they generate phenotypic variation and may help cells to quickly adapt under strong selective pressure (116). Moreover, aneuploidy occurs in cancer cells by certain mutations of proteins responsible for the proper segregation of chromosomes, such as e.g. the protein STAG2, a component of the cohesin complex (144). It has been argued that they then adapt to aneuploidy in combination with the occurrence of additional mutations of certain genes, which then allow the cell to tolerate the aneuploid state or even take advantage of it. For example, p53, a protein mutated in the majority of cancer cells, appears to limit growth in aneuploid cells, which can be reversed by its deletion. However, knockout of TP53 does not induce aneuploidy (113). Therefore, how exactly aneuploidy influences tumorigenesis despite its anti-proliferative effect remains to be elucidated, especially, in view of the promoting and inhibiting effects (113, 114, 145).

The question of how aneuploid cancer cells can be explained and whether aneuploidy in cancer is a cause or consequence is a driving force for aneuploidy research. At a minimum, it is clear that aneuploidy possesses some driving functions which range from amplification of mutated oncogenes and dosage-sensitive wild-type genes to loss of tumor suppressor genes and increased genomic instability.

Based on the above observations, aneuploidy represents a highly attractive therapeutic target for cancer treatment and offers several promising potential anti-cancer strategies. The aneuploid state, even when not causing proliferative defects, appears to exert significant stress on the cells and thereby could serve as a target for cancer therapies. One possible approach would be the identification of single genes whose deletion is lethal only in aneuploid cells, and not in diploid cells. Most of such genes would affect genome stability and would be linked to chromosome instability. Alternatively, whole pathways might be attacked in case they are a general feature in all aneuploid cells (140). As in the case of single genes, investigations focus on the identification of agents which cause lethality in aneuploid cells and not in euploid cells to specifically target cancer cells (146). Relevant pathways specifically sensitive to certain agents are determined, generally by identifying the ones responsible for the survival of aneuploid cells. Several studies have focused on pathways that deal with the excess of proteins, with the idea that this would be detrimental to aneuploid cells. Proteasomal inhibitors have been tested, however, it seems that these inhibitors alone do not lead to expected effects (113). Aneuploid trisomic MEF cells did show enhanced sensitivity to a number of inhibitors and especially an Hsp90 inhibitor. These are currently tested in clinical trials (113).

Besides aneuploidy's role in cancer and its direct effects in congenital disease like Down syndrome, there are increasing indications that it also plays important roles in neurodegenerative diseases such as Alzheimer's. The amyloid precursor protein APP whose cleaved products are responsible for the formation of amyloid beta-plaques is encoded on chromosomes 21 (112). In Down syndrome patients with three copies of chromosome 21, Alzheimer's disease is indeed observed with a higher frequency. Conversely, Alzheimer´s patients have a higher cellular occurrence of trisomy of chromosome 21 (147). Trisomies of chromosome 17 have also been detected in higher frequency in Alzheimer´s patients, and this chromosome harbors genes linked to the disease.

1.2.5. Current status in research and cellular response to aneuploidy

A major focus of current research is how cells deal with the extra number of chromosome at the mRNA and protein level. The question is how proteins encoded on the extra chromosome and the corresponding transcripts are compensated. This could happen in a similar manner to compensation for sex chromosome encoded proteins or it could not be controlled via compensation effects and instead scale proportionally to the genomic content.

For the mRNA level, researchers have suggested that transcripts of genes encoded on the multisome chromosome are not balanced to normal levels but rather reflect changes in gene copy numbers. Many studies in yeast cells with an extra chromosome or complex aneuploid karyotypes, trisomic mouse cells as well as in humans with trisomy 21 have investigated the correlation between gene copy number and mRNA gene expression levels and all suggest strong proportionality (110, 118, 119, 132). This also holds true for several types of cancer and their aneuploidies. However, this may not be the case for drosophila and plants since mechanisms for compensation have been described in these species (148). These species dependent differences still have to be elaborated in more detail.

In contrast to transcript level analyses, there are far fewer investigations into how protein levels are affected by aneusome chromosomes. Protein levels might be completely regulated back to disome levels or directly correlate with gene copy numbers as described above for transcript levels. Partial selective compensation of some proteins may also occur. In any case, in the absence of compensatory mechanisms aneuploidy would lead to excess of proteins, potentially evoking proteotoxic stress (Figure 18).
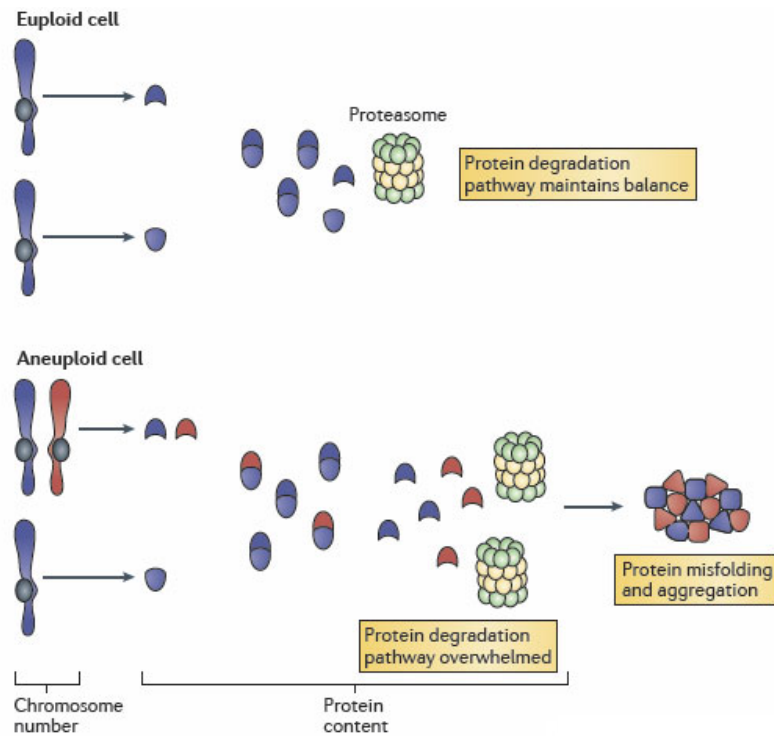
Figure 18: Generation of proteotoxic stress in aneuploid cells due to an excess of proteins. In euploid cells protein subunits from a complex located on different chromosomes are regulated by several degradative pathways to maintain proper stoichiometry. In aneuploid cells, excess proteins may overburden these pathways and result in protein misfolding and aggregation. Adopted from (113).

The few studies that have been performed on protein level so far show contradictory results. One study in budding yeast observed few compensatory effects (149). It should be emphasized that in this study only individual genes and not a complete chromosome have been investigated and even more importantly, deletions were studies, which may have completely different response mechanisms. Likewise, Pavelka et al. (116) see no balancing effects on protein level in yeast. In contrast, Torres et al. also investigated yeast, but found slight balancing effects for some proteins, especially for subunits of macromolecular complexes (132, 150). These studies used differently generated aneuploid strains and different sensitivities in the protein detection techniques, parameters which might be responsible for the different outcomes. In human aneuploid cells, no comparable study has been performed prior to this thesis since the generation of stable aneuploid clones is very difficult.

Besides the fate of proteins encoded on the supernumerary chromosomes it is of great importance - especially for the development of therapeutic approaches – as to whether a general global response to aneuploidy exists at the cellular level. If so, it would further interest if they are also responsible for the reduced proliferation observed in the aneuploid state (146).

[32]

Analysis of global regulations in yeast also showed that strains with one additional chromosome exhibit an "aneuploidy stress response". This is defined by cell growth defects, altered metabolic properties and proteotoxic stress, due to an accumulation of excess proteins (110, 116, 132, 151). Aneuploid yeast strains were much more sensitive to certain stresses such as temperature changes and inhibition of degradation pathways, for instance (132), respectively. However, also in this respect, no clear picture emerges yet, because another study reported no indications of proteotoxic or general stress response despite poor growth (116). It should be noted that these studies are not fully comparable, since they deal with trisomies or mixed aneuploidies. Instability of the aneuploid clones might be a further reason for discrepancies.

Experiments with trisomic mouse embryonic fibroblasts (MEFs) revealed impaired proliferation and altered metabolism (119) as well as energy and proteotoxic stress and increased cell sizes (146). Altered energy production, such as increased lactate production (aerobic glycolysis, also called Warburg effect), is a common feature of cancer cells. Aneuploid yeast cells also show alterations in carbohydrate metabolism. Inhibitors against autophagy, stress or heat shock response (Hsp90) pathways showed increased sensitivity of aneuploid mouse cells (140). There was also a correlation between the degree of the sensitivity against the agents and the size of the aneusome chromosome (119, 146). However, unlike in yeast, proteasomal inhibitors did not result in increased sensitivity in aneuploid MEFs.

Skin fibroblasts from Down syndrome individuals proliferate more slowly (152). In addition, cytogenetic studies of early- and late-stage blastocysts including some aneuploid cells showed that the percentage of aneuploid cells decreases over time. This suggests that euploid cells outcompete aneuploid cells (124).

In summary, aneuploidy research has not yet drawn a clear picture of cellular responses to extra copies of chromosomes in cells. In yeast and mice transcript levels of genes encoded on the aneusome chromosomes appear to scale proportionally. At the protein level, some investigations suggest slight protein compensation for members of macromolecular complexes. Aneuploid cells show signs of proteotoxic stress response and increased sensitivity to agents inhibiting degradative mechanisms in some but not all reports. No comparable studies have been performed in human cells yet. It would clearly be important to investigate these questions in this medically relevant system. MS-based quantitative proteomics would appear to be a method of choice to unravel protein changes between aneuploid cells and their diploid counterpart cells. These data could be of great relevance because defining general cellular states of the cells could potentially form a basis of an aneuploidy-specific cancer therapy.

1.3. Apoptosis

Apoptosis (apo – from, off; ptosis - falling) is a major mechanism of programmed cell death in multicellular metazoan organisms. The purpose of this energy-dependent process is to guides the cell to death in a controlled manner. Carl Vogt first described the principles of apoptosis in 1842, although it took researchers more than hundred years to get further insight into the apoptotic process. In 1972, John Kerr, Alastair Currie and Andrew Wyllie published a seminal article on apoptosis, which marks the beginning of molecular research into programmed cell death (153). In 2002, Robert Horvitz, Sydney Brenner and John E. Sulston won the Nobel Prize in Medicine for their pioneering research on apoptosis.

Cell death is an important mechanism to maintain a healthy state. Each cell has several control mechanisms which together serve as guardians by monitoring crucial processes such as each step of the cell cycle as well as the general state of the cell. Whenever the cell shows signs of misbalance these control mechanisms attempt to detect and repair the disturbance, however, when the cell is irreparably damaged or potentially dangerous to the organism, controlled cell death is initiated. As a specific and programmed form of cell death, apoptosis effects cellular disposal. It relies on proteolytic cleavage of specific protein substrates by proteases termed caspases as the central mechanism.

Besides apoptosis, other cell death mechanisms include the uncontrolled, acute form of cell death by necrosis, which results from acute cellular injury of the cell (154, 155). In this case, the cell loses its membrane integrity and surrounding liquid can enter the cell finally leading to a bursting of the cell. In contrast, apoptotic cells disperse stepwise in an orderly manner resulting in the formation of so-called apoptotic bodies, which are finally engulfed by phagocytic cells. Molecules from the apoptotic cell are thereby recycled and re-integrated into intact cells. Necroptosis is an additional type of programmed cell death but it will not be further addressed here (156-158).

Apoptosis is not only a mechanism of cell death in unbalanced states, but functions as an essential cellular mechanism during normal development, tissue homeostasis and immunity (159, 160). As a classical example, the differentiation of limbs in the developing human embryo is driven by apoptotic removal of the superfluous cells between the fingers and toes. Also in neuronal development, cell death has an important task in matching neuron numbers with their targets. In adults, it is required to maintain homeostasis in rapidly renewing tissues.

Disturbed regulation of apoptosis contributes to many pathological states and autoimmunity, infection and degenerative disorders all involve apoptosis (156, 159). Whereas excessive apoptosis

causes atrophy in tissues, in cancer the problem lies in evasion of normal apoptotic mechanisms. One of the goals in cancer therapy by chemotherapy and irradiation is to kill cancer cells by inducing apoptosis. It is thus important to determine affected proteins and pathways in apoptosis to allow more specific treatments in cancer cells with least side effects in normal cells.

Apoptosis has been intensively studied in mammals and Drosophila (161). In addition, many groundbreaking studies have been performed in the nematode Caenorhabditis elegans, in which the apoptotic pathway is sufficiently conserved that the main discoveries could be transferred to the human system (162-165). The following discussion will focus on studies and investigations in human cells, as this thesis has been exclusively performed in the human system.

1.3.1. Morphological and molecular characteristics of apoptotic cells

Apoptosis induction results in several changes in the cell both on the morphological and molecular level. In contrast to necrosis, these changes follow a strict order and step by step lead to a controlled degradation of the cell. In the following, characteristic features of apoptosis are highlighted.

1.3.1.1. Morphological changes of apoptotic cells

Among the morphological changes observed upon apoptosis induction alterations of the nucleus and cytoplasm are particularly prominent. Although the most dramatic changes occur at later stages of the apoptotic process, already very early on, the induction of apoptosis can be recognized by rounding of the cells and retraction from neighboring cells. No dramatic changes in mitochondria, the ER or the Golgi apparatus were observed during the early stages of apoptosis (153). In later stages, swelling of the mitochondrial outer membrane is evident, likely due to release of apoptotic proteins into cytoplasm (160, 166-168). Fragmentation of the ER and Golgi has also been observed (155). Cells then show shrinkage of the cytoplasm and nucleus due to a breakdown of the cytoskeleton (153). Further characteristic features of the apoptotic pathway are nuclear breakdown, due to condensation of the nucleus, and even fragmentation of the nucleus into smaller pieces (155) (Figure 19). Accompanied the chromatin condensation, DNA is extensively fragmented by hydrolysis and so-called DNA laddering can be detected on agarose gels (155).

Crucial feature of apoptosis is that the plasma membrane is not ruptured but stays intact during the entire process, thereby limiting the release of inflammatory cellular content (169). Still, the membrane starts to bleb extensively upon apoptosis induction, which can be detected via microscopy (see Figure 19). Phosphatidylserine is a molecular indicator of apoptosis, because this molecule occurs only on the cytoplasmic side of the membrane in intact cells. Upon apoptosis it

swaps to the outer side of the membrane where it can be detected by binding of Annexin V, a common assay for apoptosis induction.
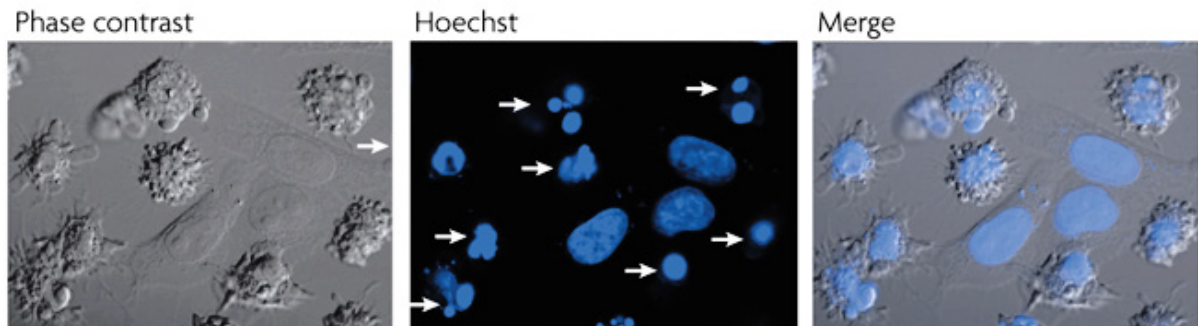

Figure 19: Morphological changes after apoptosis induction. Apoptotic cells show extensive membrane blebbing as well as nuclear condensation and fragmentation. Arrows indicate condensed and fragmented nuclei of apoptotic cells. Adopted from (155).

In the final stage of apoptosis, the complete cell disintegrates and the plasma membrane-bound "apoptotic bodies" are cleared by engulfment and uptake by nearby phagocytic cells (phagocytosis), such as macrophages (153). Some apoptotic bodies contain fragments of the nucleus, whereas others contain only cytoplasm. Material internalized within the apoptotic bodies can be recycled and re-used.

1.3.1.2. Extrinsic and intrinsic apoptosis

Along with morphological changes, apoptosis also dismantles the cell from within - on the molecular level. Apoptosis can be induced either by diverse intracellular intrinsic stimuli such as DNA damage, or by extrinsic stimuli. In the latter case, ligand binding to receptors of the tumor necrosis factor receptor family on the cell surface induces apoptosis (170).

Extrinsic stimuli comprise the ligands FasL, TNF-alpha and TRAIL which all belong to the tumor necrosis factor (TNF) superfamily and exist as stable homotrimers. They can occur both as membrane proteins or soluble proteins after cleavage of the extracellular domain by proteases such as metalloproteases. To initiate the apoptotic process, the apoptosis inducing factors bind with their extracellular C-terminal TNF homology domain (THD) to the cysteine-rich extracellular subdomains (CRDs) of their corresponding "death-receptors" (Figure 20) at the outer membrane of the cell (171-173).
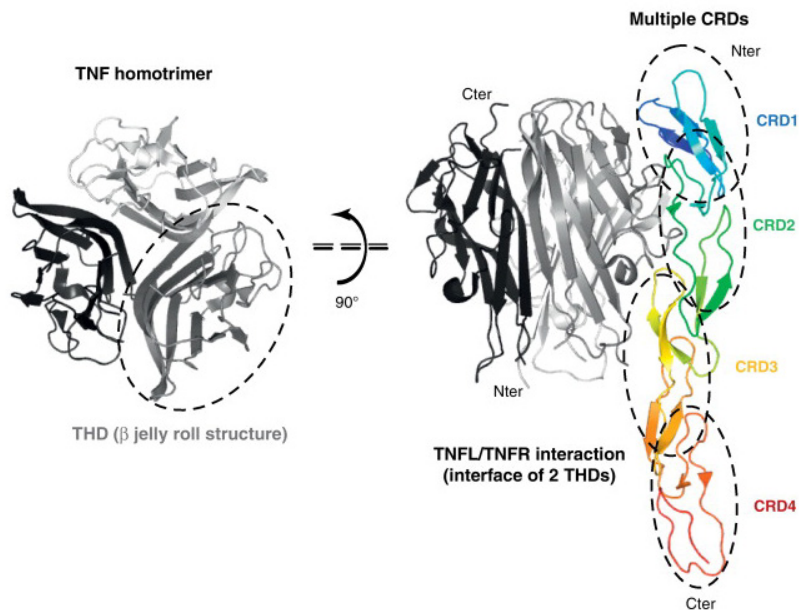
Figure 20: Interaction of the tumor necrosis factor ligand (TNFL, gray shades) with the tumor necrosis factor receptor (TNFR) as ribbon diagram representation. On the left, non-covalently linked THD domains of the TNFL trimer are depicted. One trimer is encircled. On the right, the interaction of the ligand trimer with the CRDs (each CRD encircled) of the TNFR is depicted. Adopted from (172).

Subsequently, these receptors – Fas (Apo-1, CD95), TNFR or TRAIL-R – cluster in the membrane as trimers or higher oligomers. On the cytoplasmic side, receptors possess a cysteine-rich domain called death domain, which is crucial for transmitting the signal from the extracellular to the intracellular part. Formation of the ligand-receptor trimer triggers the recruitment of factors at the inner face of the membrane forming the so-called DISC (Death-inducing signaling complex) complex (174, 175) (see Figure 21). This complex consists of FADD, which binds to the intracellular death domain of the receptor via its own death domain, as well as pro-caspase-8, the inactive form of caspase-8, which interacts with FADD via its death effector domain (176). The formation of the complex leads to activation, auto-cleavage as well as dimerization of caspase-8. After its cleavage, the initiator caspase-8 it is released from the complex into the cytosol and cleaves further proteins such as effector caspases. These downstream caspases subsequently cleave their own downstream substrates thereby enforcing the cellular apoptosis. The detailed mode of action of caspases is described in more detail in paragraph 1.3.2.

This study deals with apoptosis induced by TNF-related apoptosis inducing ligand (TRAIL). The 33 kDa protein TRAIL, also known as Apo2L or TNFSF10, belongs to the TNF superfamily. It is classified as a type II membrane protein, since it is inserted in the cell membrane with its C-terminal domain exposed to the extracellular side. It is expressed by cells of the immune system, such as T cells or

macrophages. However, TRAIL can also be proteolytically cleaved off thereby generating a soluble 20 kDa TRAIL ligand (177). TRAIL triggers apoptosis by binding to its death receptors DR4 (TRAIL-R1) and DR5 (TRAIL-R2). It can also bind to decoy receptors lacking cytoplasmic death domains, DcR1 and DcR2, which therefore do not transmit an activating signal to the cytoplasmic side of the cell membrane. Overexpression of these decoy receptors can protect the cell from TRAIL-induced apoptosis. A secreted low affinity receptor for TRAIL also exists and is called osteoprotegerin (OPG) (178). After binding of TRAIL to DR4 or DR5 and DISC formation at the cytoplasmic side of the membrane, caspase-8 is activated by cleavage and the apoptotic cleavage cascade is induced. Subsequent cellular mechanisms are likely similar to FasL induced signaling, but the precise underlying mechanisms still need to be fully elucidated. TRAIL signaling is of special interest for clinical researchers, since TRAIL has been shown to induce apoptosis in cancer cells without killing normal cells (see paragraph 1.3.3.).

As mentioned above, apoptosis is not only extrinsically induced but can also be initiated by an intrinsic signal (Figure 21). Among others, cytotoxic insults, such as exposure to DNA damage (irradiation) or growth factor deprivation, and also activation of the tumor suppressor p53, trigger pro-apoptotic members of the Bcl-2 protein family and thereby induce the apoptotic process (178). In this process, the interplay of pro- and anti-apoptotic Bcl-2 family members is crucial for the execution of apoptosis. The pro-apoptotic proteins Bax and Bak translocate from the cytosol to the mitochondrial membrane. Moreover, interaction of pro- (such as Bax and Bak) and anti-apoptotic members (such as Bcl-2 and Bcl-X$_L$) leads to an oligomerisation of Bax and Bak in the outer mitochondrial membrane (179) and formation of pores therein. How Bax and Bak are activated and oligomerize, is not fully resolved yet (159, 180). Pore formation results in the release of pro-apoptotic proteins from the mitochondrial intermembrane space, such as SMAC/DIABLO which blocks XIAP, an inhibitor of apoptosis. A further crucial factor that is released from the mitochondria is cytochrome c. In the cytosol, it induces the assembly of a complex with Apaf-1, pro-caspase-9 and dATP, forming the so-called apoptosome. As a consequence, caspase-9 is auto-cleaved and thereby activated and can cleave further downstream substrates, including effector caspases. This pathway is also termed "mitochondrial pathway" because the proteins involved are responsible for the integrity of the mitochondrial outer membrane.
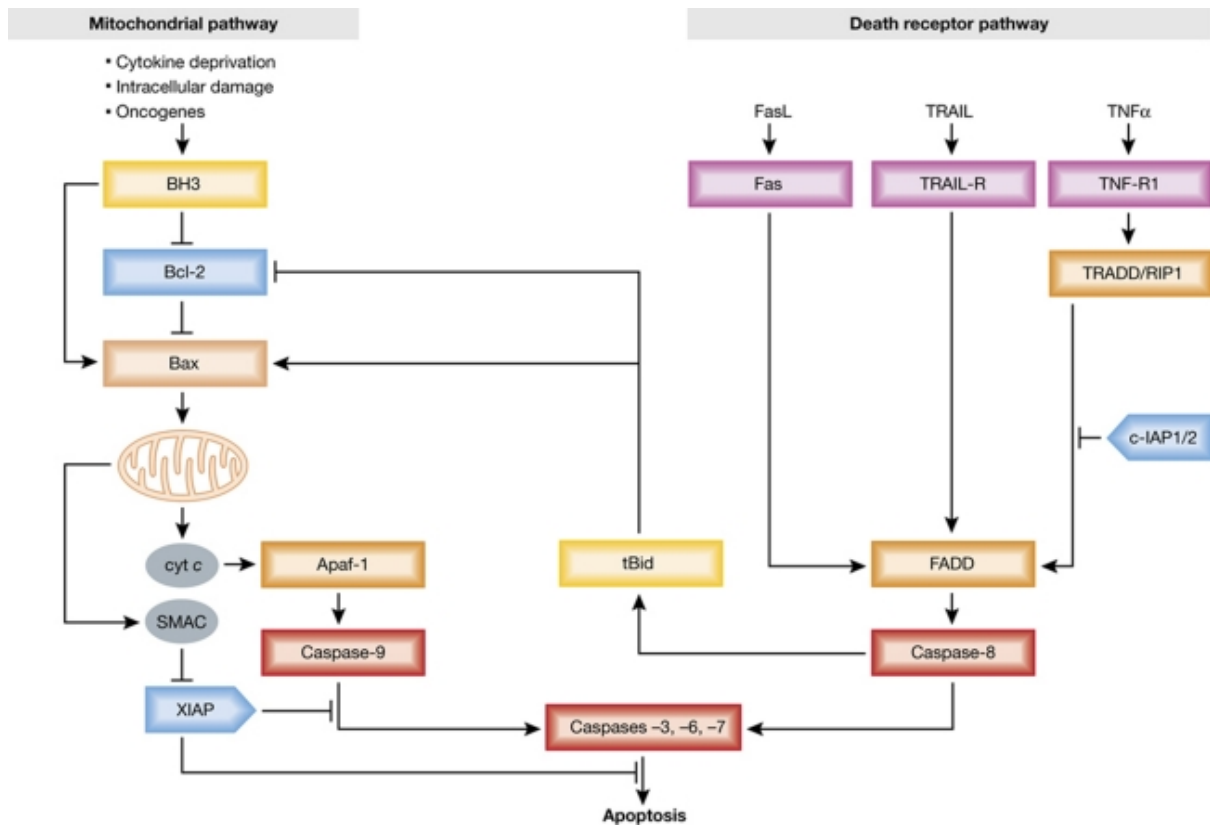
Figure 21: Schematic overview of the intrinsic (mitochondrial) and extrinsic apoptotic pathway. Modified from (159).

Although there are two distinct initiating events for apoptosis, the resulting pathways finally converge upon effector caspase activation. Moreover, even extrinsic and intrinsic activation events are not strictly separated. One protein influencing the pore formation in the mitochondrial membrane is Bid, which is cleaved by caspase-8, the initiator caspase of the extrinsic pathway (see Figure 21). The truncated active C-terminal segment called tBid translocates to and accumulates at the mitochondria, promotes Bax/Bak-mediated mitochondrial outer membrane permeabilization (MOMP) and thereby induces formation of pores in the outer membrane. Thus Bid cleavage interlinks the extrinsic and intrinsic pathways already at the initial steps. In some cells, so-called type 2 cells, activation of the intrinsic pathway by Bid is required to amplify the caspase cascade. Cells that are not dependent on this pathway after extrinsic apoptosis induction are called type 1 (181, 182).

Absence of one of the two pathways still allows accurate function of the other (183-187). Defects in caspase-8 results in resistance to extrinsic stimuli and apoptosis induction (186, 188-190). However, there is no similar silencing of the intrinsic pathway, in which mitochondrial outer membrane permeabilisation can be effected in multiple and redundant ways (191-196).

[40]

1.3.2. Caspases and their function

Regardless of the apoptotic stimulus, apoptosis culminates in the fragmentation of several hundred proteins and of the DNA and this effect is driven by proteolytic executioners. The main players are caspases, cysteine-dependent aspartate-specific proteases, which are represented by 11 functional genes in the human genome. As their name suggests, caspases cleave C-terminally after an aspartate and have a cysteine in their reactive center which is important for the cleavage process (197). Caspases are indirectly also responsible for the fragmentation of the DNA by activating a DNAse in a proteolytic manner (198, 199).

1.3.2.1. Caspase structure and function

In healthy cells, caspases are present as inactive zymogens with little or no proteolytic activity. By structure and function, caspases relevant in apoptosis can be separated into upstream initiator caspases, caspases -8, -9, -10 and -2, and downstream effector or executioner caspases, caspases-3,-6 and -7 (197, 200). Further caspases have non-apoptotic functions and are, among others, relevant for inflammatory responses, such as caspase-1 (201). Caspase-8 in combination with its adaptor FADD can also have other functions such as in the development of blood vessel, macrophage differentiation or proliferation of specific cell types (185-187).

Initiator caspases in their inactive form have a long pro-domain. After induction of an apoptotic signal, initiator caspases can bind to specific scaffold proteins via their long pro-domains and form death-inducing complexes. In the extrinsic pathway, caspase-8 binds to FADD via its death effector domain and forms the above mentioned DISC. Activation is thought to follow an induced proximity model, which leads to the auto-proteolytic activation of pro-caspase-8. Multistep cleavage results in the formation of a heterotetramer, consisting of two large p20 subunits and two small p10 subunits (202) (see Figure 22). Similarly, caspase-9, which is the main player in intrinsically induced apoptosis, can bind to Apaf-1 and form the apoptosome in combination with cytochrome c and dATP. Complex formation subsequently induces conformational changes which lead to the activation of the caspases by auto-cleavage (200).
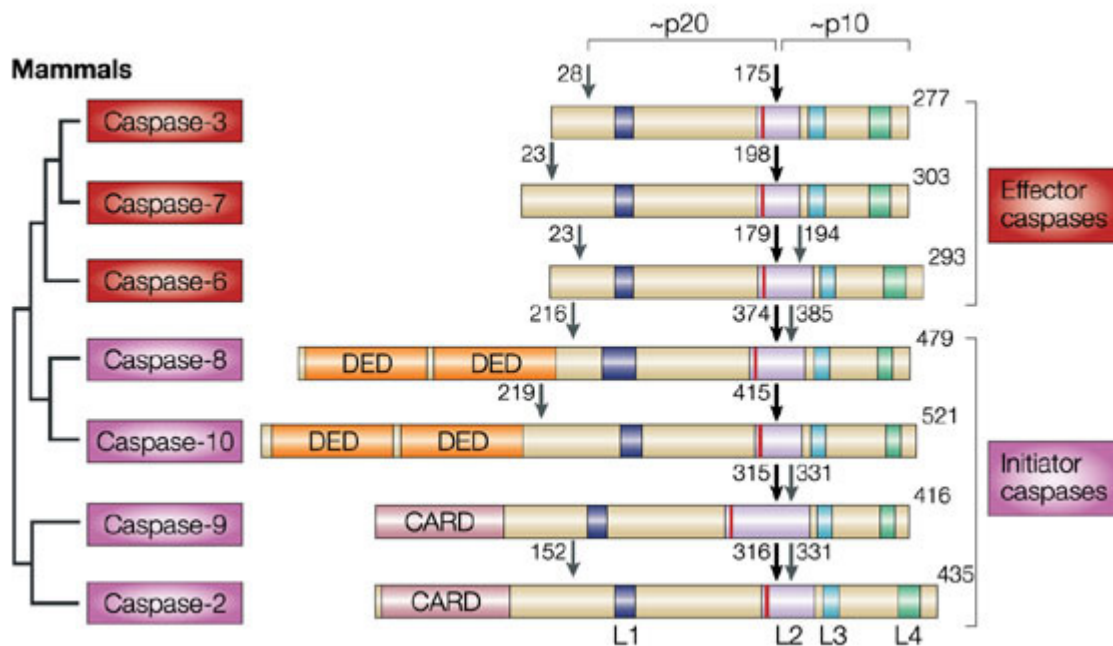
Figure 22: Overview of caspases in mammals. Initiator and effector caspases are distinguished. Adapted from (203).

Active initiator caspases can cleave executioner/effector caspases - which are as well synthesized as single-chain inactive zymogens containing short pro-domains - into about 20 (p20) and about 10 (p10) kDa fragments. These assemble into an active tetrameric protease ($p20_2 p10_2$) (159).

1.3.2.2. Downstream cleavage substrates

After effector caspases such as caspase-3 have been activated by cleavage and extrinsic and intrinsic pathways converge on the same downstream events, downstream substrates spanning a wide range of functions are cleaved. In this process, not all cleavage events result in an inactivation of the corresponding protein. Instead, cleavage can also have activating effects or induce translocations of proteins into different compartments of the cell (197, 204-206).

The morphological changes due to apoptosis induction (see 1.3.1.1.), can be linked to cleavage of specific proteins (166) (see Figure 23). Many constituents of the cellular cytoskeleton are cleaved. These include components of the actin microfilament, microtubular and intermediate filament proteins with their respective substrates, actin itself and myosin, tubulins as well as vimentin.
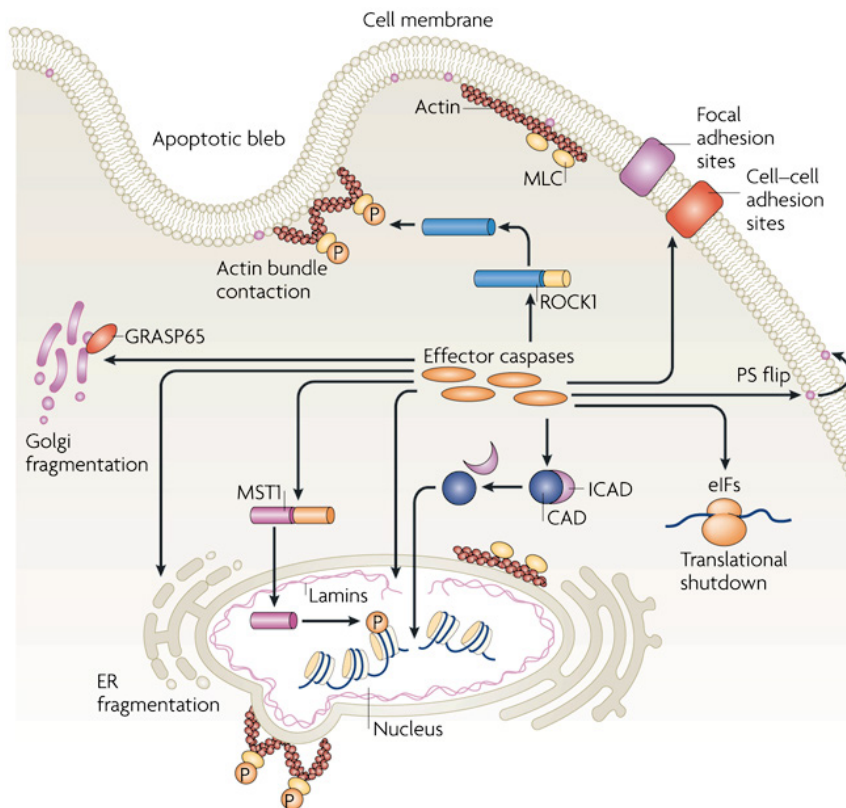
Figure 23: Schematic overview of the controlled disintegration of the cell. Both molecular cleavage events and their consequences at the morphological level are indicated. Adopted from (155).

(155). Cleavage of these proteins contributes to the rounding of the cell, cell shrinkage as well as membrane blebbing, although detailed mechanisms have not been identified yet. In addition, cleavage of nuclear lamins contributes to nuclear fragmentation and the collapse of the nuclear envelope. In many of these cases, protein functions are inhibited by cleavage.

A well-studied example of protein inhibition is the substrate PARP1 whose DNA-binding domain is cleaved off by caspases (154). PARP1 then can no longer bind to DNA, which is required to fulfill its catalytic function in DNA damage repair. The most prominent examples of protein activation by cleavage in apoptosis are the caspases themselves. One example of indirect activation and translocation of a protein is the cleavage of the inhibitor of the caspase-activated DNAse CAD (ICAD). At normal conditions ICAD forms a stable complex with CAD, however, upon apoptosis induction, ICAD is cleaved by caspases. This cleavage inactivates the inhibitor, thereby releases CAD from the complex and initiates the translocation of CAD to the chromosomes where it induces DNA fragmentation. These examples highlight the broad range of apoptotic cleavage events and their consequences for the cell. In addition, many more pathways, especially RNA-related processes as

[43]

well as transcription and translation machineries have variously been described as affected by caspase cleavage.

In recent years, several methods for the identification of cleaved caspase-dependent substrates have been developed. In vitro approaches attempt to identify substrate motifs, but have the disadvantage that it is not clear if they represent in vivo events. In contrast, MS-based proteomics can elucidate cleavage events in vivo. Most of the studies published so far have investigated intrinsically induced apoptosis by stimuli such as staurosporine. One approach aims at the identification of the explicit apoptotic cleavage sites by enriching N-termini newly generated after apoptosis induction. A second strategy searches for cleavage events without enrichment by taking into account further protein level information. Caspase cleavage detection assays are thoroughly discussed in the results section (see section 2.2.). The above assays have allowed the detection of many substrates, but they have several drawbacks, in particular the lack of a quantitative global whole protein approach.

1.3.3. Apoptosis and its relevance in cancer

Evasion of apoptosis is one of the accepted hallmark of cancer (207). One of the first inks between cancer and apoptosis has been the observation that overexpression of the anti-apoptotic protein Bcl-2 prevented the cell from initiating apoptosis (208, 209). In mice, further anti-apoptotic proteins have been linked to increased tumorigenesis after their overexpression (210-213) whereas pro-apoptotic family members were shown to be tumor suppressors. Deletion of pro-apoptotic proteins even accelerated the generation of lymphomas (214-216). In human cancer loss or suppression of pro-apoptotic family members has likewise been observed (210, 217, 218). In addition to defects in the intrinsic pathway, alterations in the death receptor pathway are also known to promote tumorigenesis. For example, mutations in the Fas death receptor or the Fas ligand have been linked to increase cancer incidence (219).

Although evasion from apoptosis is driving cancer development, the opposing effect of excessive cell death has also recently been associated with cancer development. Upon chemotherapy or low-dose gamma-irradiation in radiotherapy many cancer survivors eventually develop more tumors (220, 221).

Several anti-cancer strategies specifically trigger apoptosis for therapeutic effect. Even today, the most common treatment of solid tumors uses chemotherapeutic drugs or gamma-irradiation. These induce DNA damage and cellular stress or block DNA replication, thereby triggering cancer cell death. However, these treatments involve a high degree of systemic toxicity and often become

ineffective due to anti-apoptotic mutations in the cell. Tumor resistance can occur after repeated treatments. For these reasons, many strategies now directly target molecular changes in the apoptotic pathway of cancer cells.

Since apoptosis depends on the tightly controlled interplay between pro- and anti-apoptotic proteins (see 1.3.1.2.), several strategies have been developed to target anti-apoptotic proteins in cancer cells - proteins that are known to trigger tumorigenesis when overexpressed (see above). Substances are pursued that mimic BH3-only proteins and target these pro-survival proteins at the level of protein-protein interactions. These interactions are thought to directly flip the Bcl-2-regulated apoptotic switch. A few of these mimetics are already in clinical trials (222-224). Further studies focus on antisense RNAs to target the expression of the pro-survival genes and some of these strategies are likewise in clinical trials (225-227). Synthetic ribozymes have also been studied (228, 229) as has the inhibition of apoptotic inhibitors, which are often elevated in cancer cells similar to anti-apoptotic proteins. In this respect, recent clinical trials try to mimic natural antagonists of inhibitor of apoptosis proteins (IAPs) such as SMAC/DIABLO (targeting XIAP) (230).

Targeting the death receptor pathway, rather than the intrinsic apoptotic pathway, could potentially directly trigger cell death in cancer cells. Unfortunately, induction via the Fas receptor (e.g. by antagonistic CD95L antibodies) has led to liver damage (231) and this problem of toxicity still needs to be overcome (202). In contrast, the TRAIL receptor is emerging as promising agent, since TRAIL is known to induce apoptosis in a wide variety of tumors while exhibiting little or no toxicity in normal cells (232, 233). TRAIL's limited effects on normal cells, may be due to the restriction of the TRAIL death receptors DR4 and DR5 expression to transformed cells, whereas normal cells only express the decoy receptors (234). However, the exact underlying mechanisms still need to be fully identified. Furthermore, not all tumor cells are sensitive to TRAIL and resistance occurs in some cancer types, such as chronic lymphocytic leukemia (CLL), despite the expression of TRAIL receptors on the surface of the tumor cells (202, 235). One possible explanation may be the expression of decoy receptors also in cancer cells. For this reason it would be of great importance to screen for resistant tumors before treatment with TRAIL. However, no clear picture has emerged so far on the correlation of decoy receptor expression and tumor resistance (235). Anti-apoptotic proteins within the cell might also add to TRAIL resistance but again further studies have to be performed to unravel their specific role (178). Recombinant human TRAIL (rhTRAIL) as well as DR4-/DR5-specific agonistic antibodies are in early clinical trials. TRAIL versions used in clinical trials represent the extracellular region of the ligand stabilized with a zinc atom to retain the active trimeric conformation. Tagged versions of the ligand with e.g. FLAG are no longer used since hepatotoxicity and other side effects have been

observed (177). Since TRAIL induces apoptosis independently of the intrinsic pathway, including the p53-regulated Bcl-2 family members, TRAIL treatment may offer an approach to target p53-deficient tumor cells which are often resistant to chemotherapies and irradiation (178, 236). Enhanced apoptosis induction may be obtained by combined treatments with for example BH3 mimetics or chemotherapeutic agents in some cancer cells (226, 237, 238) and could even re-sensitize resistant tumor cells (178).

Many other strategies to induce apoptosis in tumor cells without general cytotoxic effects to normal cells are also pursued. Figure 24 gives an overview of different approaches used to date.
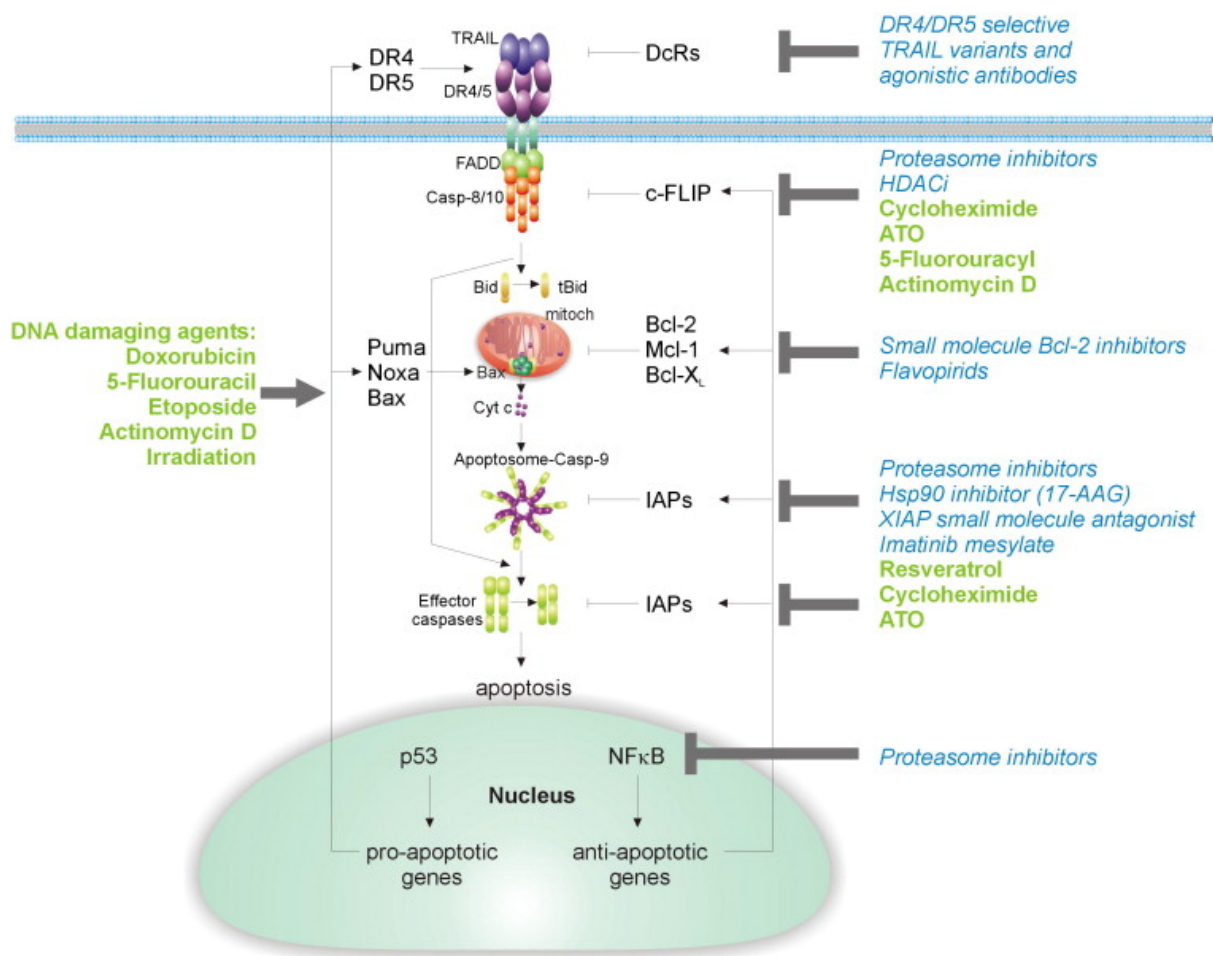
Figure 24: Cancer therapy can interfere at several stages of the apoptotic process including unspecific chemotherapeutic agents (green) and specific cancer treatments (blue). Both the intrinsic and the extrinsic pathway can be attacked. Adopted from (178).

For cancer therapy, the development of novel drugs is crucial. However, when targeting the apoptotic pathway, there is still debate in the field about the most efficient strategies. For instance,

it has to be evaluated whether combined therapies are more efficient than single treatment strategies, such as targeting one specific gene/protein only, which strategies work best under certain conditions, including cell type, cancer type and many more parameters and what the side effects are. For these reasons, it is crucial to investigate basic apoptotic processes in detail. Only when detailed mechanisms can be identified it will be possible to rationally develop new therapies for cancer based on the direct targeting of the apoptotic process (226). In particular we here wanted to develop strategies that allow the comparison of cleavage events for different conditions.

2. RESULTS AND DISCUSSION

2.1. Global analysis of genome, transcriptome and proteome reveals the response to the presence of extra chromosomes in human cells

2.1.1. Aim and Summary

Aneuploidy, the state of unbalanced chromosome content, plays an important role in cancer and disease. Nevertheless, no studies had been performed yet in human cells to unravel proteomic changes upon the introduction of extra chromosomes. For this reason, we generated stable aneuploid clones of HCT116 and RPE-1 cell lines, which differed in the number of extra chromosomes and the specific chromosome. Our aim was to analyze these cells in comparison to their diploid counterpart cells on the genome, transcriptome and proteome levels. In particular, we performed SILAC-based quantitative proteomics to directly compare proteomic changes between corresponding diploid and aneuploid cells. These data were merged with the genomic information derived by aCGH and the mRNA data from microarray data from exactly the same cell populations. This dataset allowed us to investigate the fate of both proteins and mRNAs encoded on the extra chromosome and to determine regulations in aneuploid cells on a global scale.

We found that mRNA levels corresponding to coding regions on the extra chromosome scale proportionally to the chromosome number. However, on the protein level, partial balancing effects of certain proteins could be observed. Especially proteins in complexes exhibited strong balancing towards the diploid abundance levels. In addition, our study revealed a general response to aneuploidy in all aneuploid clones. In all clones, lysosomal pathways were elevated. Interestingly, all aneuploid clones up-regulated the important autophagic markers LC3 and p62. The autophagic pathway is responsible for the degradation of organelles as well as certain ubiquitinated proteins. We further investigated the autophagic effects by biochemical methods and confirmed the activation of autophagy in all aneuploid clones.

Our study makes an important contribution to aneuploidy research by providing quantitative data at the genome, transcriptome and proteome level in human cell lines for the first time. Our data suggest that autophagy plays an important role in the balancing of protein levels in aneuploid human clones. Although this phenomenon needs to be further elucidated in detail, the identification of autophagy as an important, general response in aneuploid cells may already provide new insights for the treatment of aneuploid cancer cells.

2.1.2. Contribution

This study was a collaboration between Zuzana Storchova, Silvia Stingele and myself. Whereas Silvia Stingele focused on cell characterization and biochemical assays, my contribution centered around SILAC-based quantitative proteomics. I analyzed the LC-MS/MS data and performed the downstream data analysis. Additionally, I also analyzed the CGH and microarray data and combined all results for further analyses and provided most of the figures based on the CGH, mRNA and protein data. Biological interpretation and composition of the paper involved contributions from all collaborators.

2.1.3. Publication

The following paper was published in 2012 in Molecular Systems Biology:

Global analysis of genome, transcriptome and proteome reveals the response to the presence of extra chromosomes in human cells

Stingele Silvia*, Stoehr Gabriele*, Peplowska Karolina, Cox Juergen, Mann Matthias and Storchova Zuzana

*these authors contributed equally

Mol Syst Biol. 2012 Sep 11;8:608.

molecular
systems
biology

Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells

Silvia Stingele[1,3], Gabriele Stoehr[2,3], Karolina Peplowska[1], Jürgen Cox[2], Matthias Mann[2] and Zuzana Storchova[1,*]

[1] Group of Maintenance of Genome Stability, Max Planck Institute of Biochemistry, Martinsried, Germany and [2] Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
[3] These authors contributed equally to this work
* Corresponding author. Group of Maintenance of Genome Stability, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.
Tel.: +49 89 8578 3145; Fax: +49 89 8578 3022; E-mail: storchov@biochem.mpg.de

Extra chromosome copies markedly alter the physiology of eukaryotic cells, but the underlying reasons are not well understood. We created human trisomic and tetrasomic cell lines and determined the quantitative changes in their transcriptome and proteome in comparison with their diploid counterparts. We found that whereas transcription levels reflect the chromosome copy number changes, the abundance of some proteins, such as subunits of protein complexes and protein kinases, is reduced toward diploid levels. Furthermore, using the quantitative data we investigated the changes of cellular pathways in response to aneuploidy. This analysis revealed specific and uniform alterations in pathway regulation in cells with extra chromosomes. For example, the DNA and RNA metabolism pathways were downregulated, whereas several pathways such as energy metabolism, membrane metabolism and lysosomal pathways were upregulated. In particular, we found that the p62-dependent selective autophagy is activated in the human trisomic and tetrasomic cells. Our data present the first broad proteomic analysis of human cells with abnormal karyotypes and suggest a uniform cellular response to the presence of an extra chromosome.
Molecular Systems Biology **8**: 608; published online 11 September 2012; doi:10.1038/msb.2012.40
Subject Categories: proteomics; genome stability & dynamics
Keywords: aneuploidy; autophagy; pathway analysis; proteomics; p62

Introduction

Aneuploidy is detrimental to eukaryotic cells. Chromosome number changes are the main cause of spontaneous miscarriages and surviving embryos are born with severe disabilities, but the reasons for the incompatibility of the karyotype changes with normal development are not well understood (Hassold *et al*, 2007). Cells isolated from patients with Down syndrome (trisomy of chromosome 21) grow substantially slower (Segal and McCoy, 1974) and addition of even a single chromosome to mouse cells causes multiple defects, such as growth delay or increased sensitivity to certain drugs (Williams *et al*, 2008; Tang *et al*, 2011). Abnormal karyotypes were also shown to affect cell physiology in yeasts and plants. Addition of chromosomes leads to growth defects (Torres *et al*, 2007), phenotypic variability (Selmecki *et al*, 2006; Pavelka *et al*, 2010; Chen *et al*, 2012) and increased genome instability that likely contributes to evolution (Niwa *et al*, 2006; Selmecki *et al*, 2006; Huettel *et al*, 2008; Sheltzer *et al*, 2011). On the other hand, variable aneuploid karyotypes can be found in tumors without an obvious adverse effect on cell proliferation and it has been hypothesized that aneuploidy

facilitates tumorigenesis (Storchova and Pellman, 2004). Despite the frequent occurrence of aneuploidy and its link to pathological states, molecular mechanisms underlying the observed phenotypes remain unclear.

One of the important questions is how the presence of an extra chromosome affects the mRNA and protein content of a cell, and what is the physiological response to these changes. Transcriptome profiling suggests that the mRNA levels from genes encoded on the extra chromosomes mostly scale up proportionally with the gene copy numbers in various organisms such as *in vitro* generated aneuploid yeast, mouse and human cells (Upender *et al*, 2004; Torres *et al*, 2007; Williams *et al*, 2008), pathogenic *Candida* strains (Selmecki *et al*, 2006) and aneuploid plants (Makarevitch *et al*, 2008). Other reports suggest a feedback control that buffers the mRNA levels of amplified or underrepresented chromosomal regions in naturally occurring aneuploid yeast strains (Kvitek *et al*, 2008), plants (Birchler *et al*, 2005) or in *Drosophila* with partial or whole chromosomal aneuploidy (Stenberg *et al*, 2009). Detailed analysis of Down syndrome patients suggests that the transcription levels of some of the genes on

chromosome 21 are compensated as well (Aït Yahya-Graison *et al*, 2007). So far, only little is known about the changes in protein content and pathway regulation in aneuploid cells. Recent studies in aneuploid budding yeasts yielded partially contradictory results. On one hand, an artificial introduction of a single chromosome into haploid cells led to a growth delay, general stress response and proteotoxic stress (Torres *et al*, 2007). Moreover, mRNAs and proteins coded on the disomes were expressed proportionally to the chromosome copy numbers with exception of the protein levels of subunits of multimolecular complexes that were partially compensated to maintain the stoichiometry (Torres *et al*, 2010). On the other hand, meiotically generated multi-chromosome aneuploidy resulted in a proportional scaling of protein expression with no significant compensation of the abundance of subunits of protein complexes, and no indications of proteotoxic stress or general stress response were identified (Pavelka *et al*, 2010). No similar analysis has been performed in human aneuploid cells so far.

To uncover the fate of the transcripts and proteins encoded on the extra chromosome and to investigate the global changes in human cells in response to supernumerary chromosomes, we created model tri- and tetrasomic cells derived from two different human chromosomally stable cell lines, HCT116 and RPE-1. We quantified the changes of genome, transcriptome and proteome and determined specific pathways whose regulation is altered in these cell lines. This analysis suggests a specific cellular response to the presence of extra chromosomes in human cells. In particular, we show that p62-dependent autophagy is activated in cell lines with extra chromosomes, where it may contribute to the maintenance of normal protein levels.

Results

Generation and characterization of human trisomic and tetrasomic cell lines

A detailed analysis of aneuploidy in human cells is hampered by the lack of an appropriate model with matching diploid and aneuploid cells. To circumvent this limitation, chromosome transfer via micronuclei was used to add an extra chromosome into HCT116 (Haugen *et al*, 2008) or HCT116 stably expressing H2B-GFP (Supplementary Figure S1A). This approach generated cognate trisomic and tetrasomic derivatives that carry additional copies of chromosome 3 (labeled HCT116 3/3) or chromosome 5 (trisomy: HCT116 H2B-GFP 5/3, tetrasomy: HCT116 5/4 and HCT116 H2B-GFP 5/4). HCT116 is a transformed cell line with several previously identified chromosomal changes such as the chromosome Y loss and amplified regions of chromosomes 8, 10 and 17 (Masramon *et al*, 2000). These aberrancies are mostly present in the new aneuploid cell lines (Supplementary Figure 1B) and thus likely do not affect the results. Nevertheless, to strengthen our analysis and to overcome this possible drawback, we generated cell lines trisomic for chromosomes 5 and 12, and another cell line trisomic for chromosome 21, both derived from the diploid primary epithelial cell line RPE-1 that was immortalized by the expression of hTert and that lacks substantial chromosomal aberrancies. The successful

chromosome transfer was verified by chromosome paints (Figure 1A), comparative genomic hybridization (CGH) and multicolor fluorescence *in situ* hybridization (Supplementary Figure S1B and C). The analysis confirmed that original cell lines and their derivatives differ only by copy number of a specific chromosome.

All trisomic and tetrasomic cell lines display marked growth impairment as previously observed (Segal and McCoy, 1974; Guo and Birchler, 1994; Torres *et al*, 2007; Williams *et al*, 2008), with tetrasomic cell lines growing significantly slower than the trisomic ones (Figure 1B). We found that the aneuploid cells progress slowly and less synchronously through G1 and S phase, with a delay of 5 and 4 h, respectively (Figure 1C). On the contrary, the progression through G2 phase and mitosis is not noticeably affected (Figure 1C; Supplementary Figure S1D and E) and there is no marked accumulation of dead and non-proliferating cells (Supplementary Figure S1F and G). Similar delay in cell-cycle progression was observed in the RPE-1 5/3 12/3 (Supplementary Figure S1H). Thus, the presence of an extra chromosome impairs the growth of human cells.

Comparison of genome, transcriptome and proteome of cells containing extra chromosomes

Using the series of isogenic diploid and aneuploid cell lines we asked how the mRNA and protein levels changed in response to the chromosome copy number changes. We examined the tetrasomic cell line HCT116 5/4, which guaranteed larger dynamic range of measured changes than in trisomic cells. To quantify the DNA levels, we used high-resolution CGH (Supplementary Table S1). To measure the corresponding mRNA, we determined the medians from microarray analysis of three independent samples (Supplementary Table S1). To compare the protein levels, we used stable isotope labeling with amino acids in cell culture (SILAC) followed by high-resolution mass spectrometry (Ong *et al*, 2002) and quantified the proteome to a depth of ~6000 proteins (Supplementary Table S1). In total, we performed three biological replicates with six measurements for HCT116 5/4; all other cell lines were analyzed less extensively (up to three biological replicates, see Supplementary Table S1). The analysis of variability in technical and biological replicates demonstrated high quantification accuracy and reproducible measurements for all experiments including the reverse labeling, with Pearson correlation factors between 0.64 and 0.90 (Supplementary Figure S2A). We then determined the medians of quantitative protein changes. The log2 ratios of aneuploid-to-diploid abundance changes of DNA, the corresponding mRNAs and the proteins were matched directly using the annotated chromosomal positions (Figure 2A; Supplementary Figure S2B). In the case of equal levels in aneuploid and diploid cells, we expect the median aneuploid-to-diploid ratio (log2) to be 0 and this is observed for all DNA, mRNA and proteins coded on the disomes (DNA: 0.005, median mRNA: -0.03; proteins: -0.06; Figure 2B). The median ratio of mRNA encoded on the tetrasomic chromosome 5 is 1.09—very close to the expected value of 1.0—but in contrast, the median ratio of protein level changes is only 0.69 (Figure 2B). Indeed,
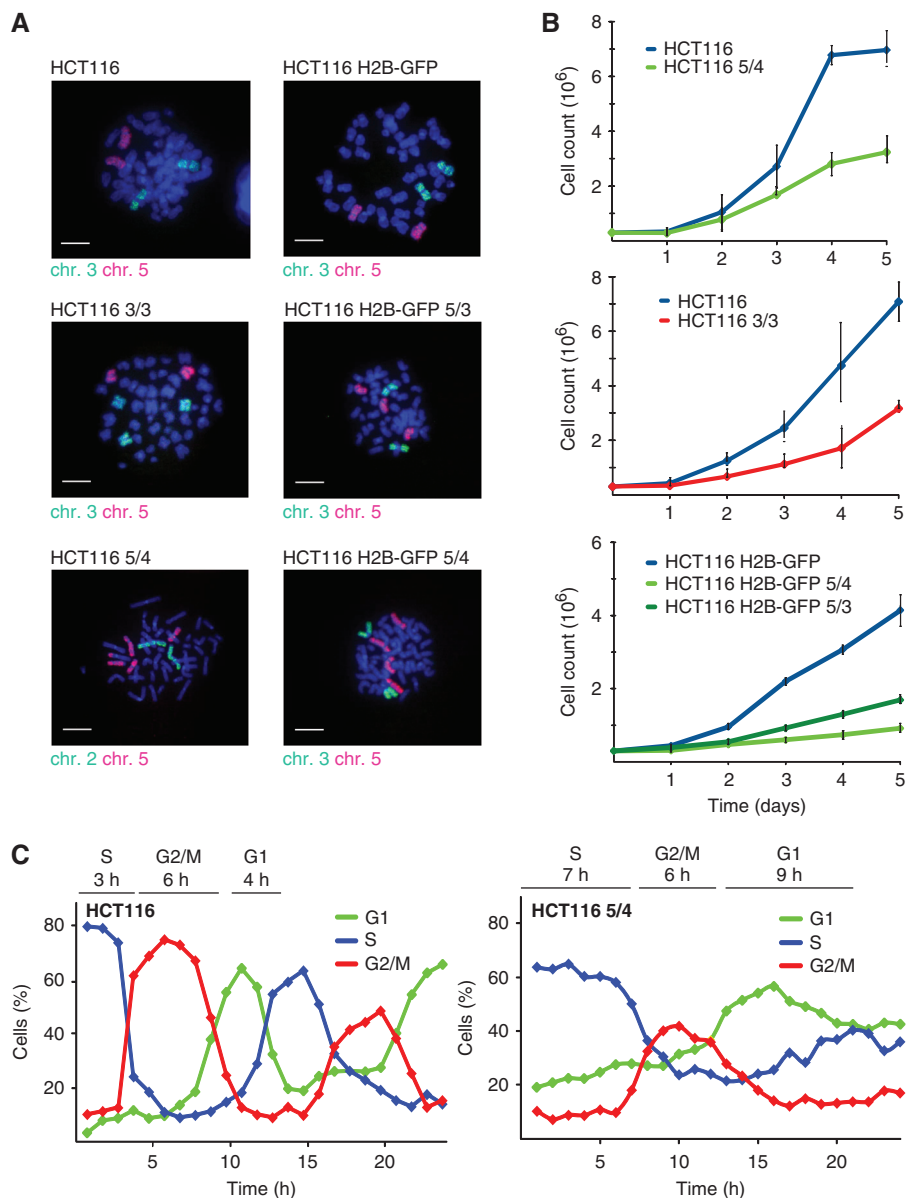
Figure 1 Characterization of HCT116 and its tri- and tetrasomic derivatives. (**A**) Chromosome paints of used tri- and tetrasomic cell lines. Bar—10 μm. (**B**) Growth curves of tri- and tetrasomic cell lines in comparison with their diploid counterparts. Each point represents the mean with standard deviation of three independent experiments. (**C**) Cell-cycle progression of HCT116 (left panel) and HCT116 5/4 (right panel) after release from thymidine block, analyzed by flow cytometry. The major delay occurs in the G1 and the S phase. The length of each cell-cycle phase is indicated above the graph. Population was considered to enter a specific phase of cell cycle if at least 50% of cells showed corresponding DNA content. See also Supplementary Figure S1. Source data is available for this figure in the Supplementary Information.

53 out of 197 proteins (27%) coded on the tetrasomic chromosome 5 are present at levels expected for disomes (median of diploid levels ± twice the standard deviation). The calculated median of protein abundance of these 53 proteins is 0.26, whereas the median of the corresponding mRNAs is 1.16, further strengthening the idea that their expression is adjusted at the protein level. Similarly, 25% of proteins coded on chromosome 3 are present at diploid levels in HCT116 3/3 (Supplementary Table S1). We determined that the proteins coded on the multisomic chromosome are present at a lower level than expected in all analyzed cell lines (Supplementary Figure S3). In conclusion, whereas the mRNA expression

corresponds to the increased chromosome copy numbers, approximately a quarter of the proteins are present at levels lower than expected and more similar to the disomic levels.

Specific proteins are maintained at stoichiometric levels

The fact that most proteins coded on the extra chromosomes are more abundant than proteins from diploid chromosomes indicates that there is no general efficient mechanism for 'gene dosage compensation' of the analyzed tri- or tetrasomies.
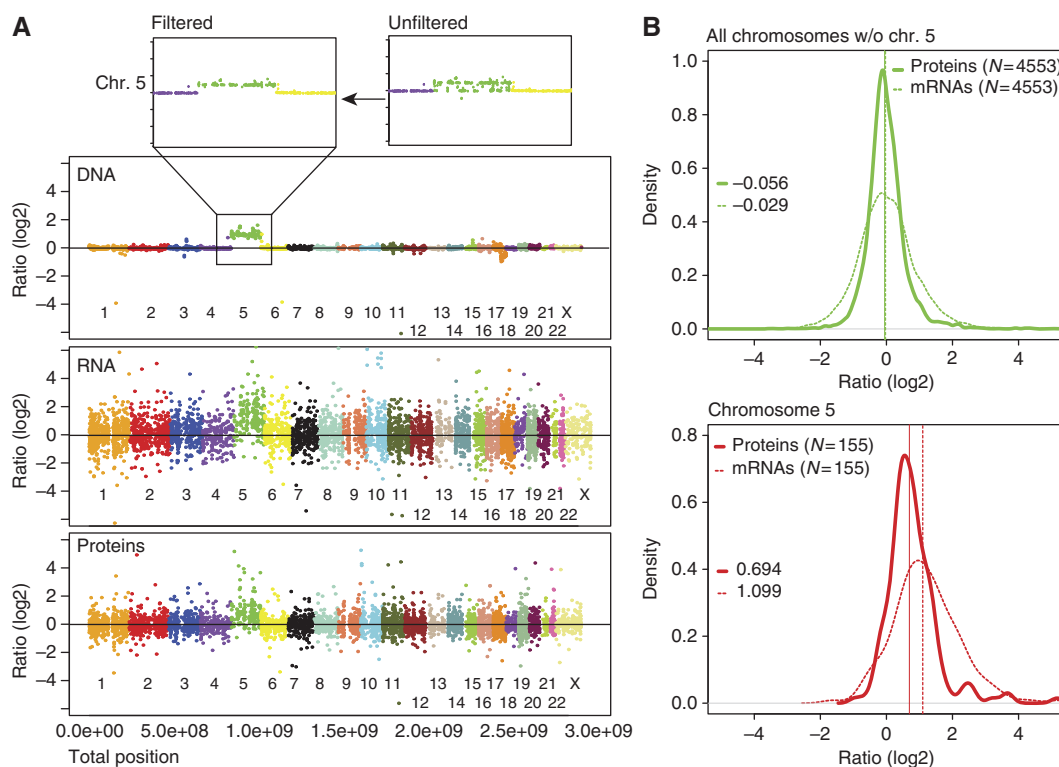
Figure 2 Quantification of DNA, mRNA and protein abundance. (**A**) Ratios (log2) of DNA, mRNA and protein abundance changes between HCT116 5/4 and HCT116 aligned with respect to their chromosome position. Each dot represents abundance changes for one gene, the corresponding mRNA and the corresponding protein, respectively. CGH analysis revealed some deleted regions in chromosome 5 (upper panel); the disomic entries were omitted in all analyses (threshold 0.65 [log2]). (**B**) Overlays of mRNA and protein density histograms (HCT116 5/4 versus HCT116). The full line represents median of protein abundance changes, the dashed line median of mRNA abundance changes. Values of respective medians are plotted in the graph. The difference between distribution of proteins and mRNAs coded on chromosome 5 is statistically significant (Wilcoxon rank sum test). See also Supplementary Figure S2.

Nevertheless, a remarkable proportion of these proteins are present at levels lower than expected according to the gene and mRNA copy numbers, suggesting that some proteins or protein categories might be adjusted to normal abundance. A long standing hypothesis posits that free subunits of multimolecular complexes may be degraded in cells (see e.g., Goldberg and Dice, 1974 and Guialis *et al*, 1979). Analysis of our data confirmed that the abundances of subunits of protein complexes (as annotated in the CORUM database; Ruepp *et al*, 2010) that are encoded on the tetrasomic chromosome are lower than expected based on the gene copy number, and more similar to the protein abundances observed in the parental cell line HCT116; the median of proteins of CORUM complexes coded on chromosome 5 is shifted down to 0.43 (Figure 3A). A similar shift toward the diploid levels for the CORUM-annotated proteins can be detected in all analyzed cell lines (Supplementary Figure S4A). We examined 14 different macromolecular complexes with at least one subunit coded on the tetrasomic chromosome and found that 8 of them (57%) maintain stoichiometry by decreasing the protein abundance close to the normal, disomic levels, while the mRNA levels vary (Figure 3B; Supplementary Figure S4B; Supplementary Table S3).

Moreover, we found that protein kinases coded on the supernumerary chromosomes shift toward near-diploid expression levels in HCT116 5/4 as well (Figure 3C).

Interestingly, the abundance distribution shows a bi-modal pattern. Remarkably, the uncompensated kinases are preferentially associated with pathways that we determined to be globally upregulated in aneuploid cells (e.g., JNK2, see below); other kinases coded on chromosome 5 were present at near-diploid levels. A similar trend was observed in other cell lines, but the limited number of proteins did not allow evaluation of the statistical significance of these changes. In conclusion, trisomic and tetrasomic human cells maintain normal levels of proteins coded on chromosomes present in extra copies, but the levels are adjusted only for specific classes of proteins.

General cellular response to the presence of extra chromosomes in human cells

Next, we asked how aneuploidy affects global pathway regulation. To this end, we used a recently developed software called '2-D annotation enrichment analysis' (see Supplementary Information). For each clone, we identified all significantly altered pathways (as defined by Gene Ontology categories, KEGG pathways and CORUM database) and ranked the relative abundance changes of proteins within the category compared with the complete measured data set. The resulting score is on the scale from −1 to +1, where the pathways close
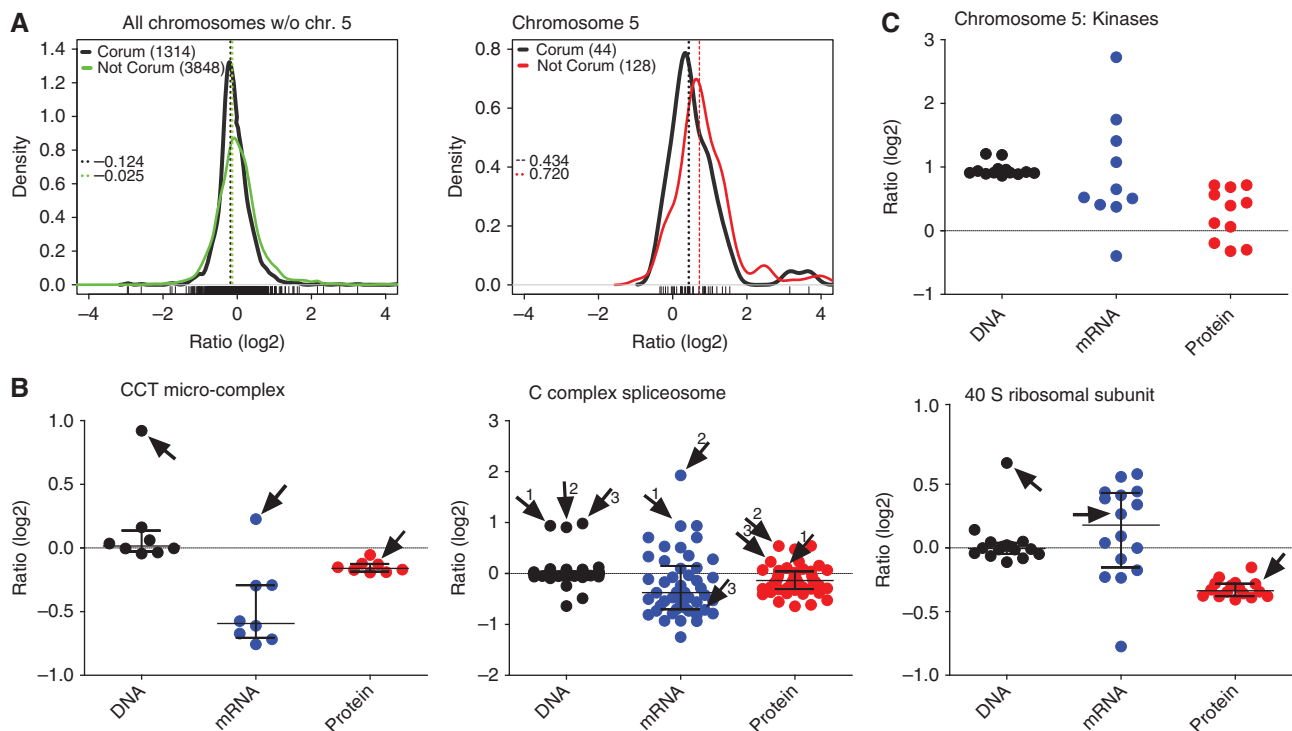
Figure 3 Abundance of subunits of protein complexes and kinases in the tetrasomic cell line. (**A**) Density plots of subunits of protein complexes (as defined in the CORUM database) encoded on all disomic chromosomes compared with non-CORUM proteins (left panel); the same for proteins encoded on the tetrasomic chromosome 5 (right panel). The differences between CORUM and non-CORUM populations of chromosome 5 are statistically significant (Wilcoxon rank sum test). Dashed lines indicate medians of the populations. (**B**) Examples of protein complexes with at least one subunit coded on the tetrasomic chromosome 5. Each dot represents abundance changes of one gene (black), its corresponding mRNA (blue) and protein (red). Subunits coded on chromosome 5 are indicated with an arrow and number. (**C**) The abundance of DNA, mRNA and proteins of kinases coded on chromosome 5. See also Supplementary Figure S3. Source data is available for this figure in the Supplementary Information.

to −1 are most downregulated and pathways close to +1 are most upregulated (see Supplementary Information for further details). Plotting these calculated scores revealed remarkable similarities in pathway regulation among the analyzed aneuploid cell lines (Figure 4A and B). We found that pathways involved in DNA and RNA metabolism, such as replication, DNA repair, transcription and mRNA processing, were significantly downregulated (Figure 4A, B and D; Supplementary Figure S5A), which is consistent with the observed growth delay in the G1 and S phase (Figure 1C; Supplementary Figure S1D, E and H). Upregulated categories identified in all analyzed cell lines included pathways required for lipid and membrane biogenesis, endoplasmic reticulum, Golgi vesicles and lysosome functions as well as energy metabolic pathways such as mitochondrial respiratory metabolism and carbohydrate metabolism (Figure 4A–C; Supplementary Figure S5A). Remarkably, the 2-D enrichment analysis of the transcriptome data determined similar changes in the pathway regulation, suggesting that the pathway response cannot be caused by a bias in protein detection (Supplementary Figure S5B and C). The observed alterations in pathway regulation were not an artifact of the chromosome number changes, since excluding the proteins coded on chromosome 5 from the analysis did not affect the identified pathways (Supplementary Figure S5D). Additionally, we found a remarkable overlap in pathway alterations between

all RPE-1- and HCT116-derived trisomic and tetrasomic cell lines (Figure 4B; Supplementary Figure S5A). Taken together, the uniformity of the response in different human trisomic and tetrasomic cells demonstrates that the presence of an extra chromosome itself, and not individual chromosome or cell types, is an important determinant of the general cellular response.

Autophagy is activated in aneuploid cells

The analysis of the altered pathways indicates that lysosome proteins are upregulated in all aneuploid cell lines (Figure 4C; Supplementary Figure S5). The lysosome is essential for autophagy, a pathway involved in removal of damaged or superfluous proteins and organelles (He and Klionsky, 2009). A detailed analysis of the proteins involved in autophagic processes suggested that autophagy might be more active in aneuploid cell lines (Figure 5A). To substantiate this finding, we performed functional analysis of autophagy in aneuploid cells in comparison with parental diploid cell lines. By immunofluorescence, we observed an increased number of LC3 foci in HCT116 5/4 in comparison with HCT116 (Figure 5B). Similarly, immunoblot analysis of protein lysates from aneuploid clones showed accumulation of the autophagy marker *LC3-II* (Figure 5C), which is the lipidated form of LC3
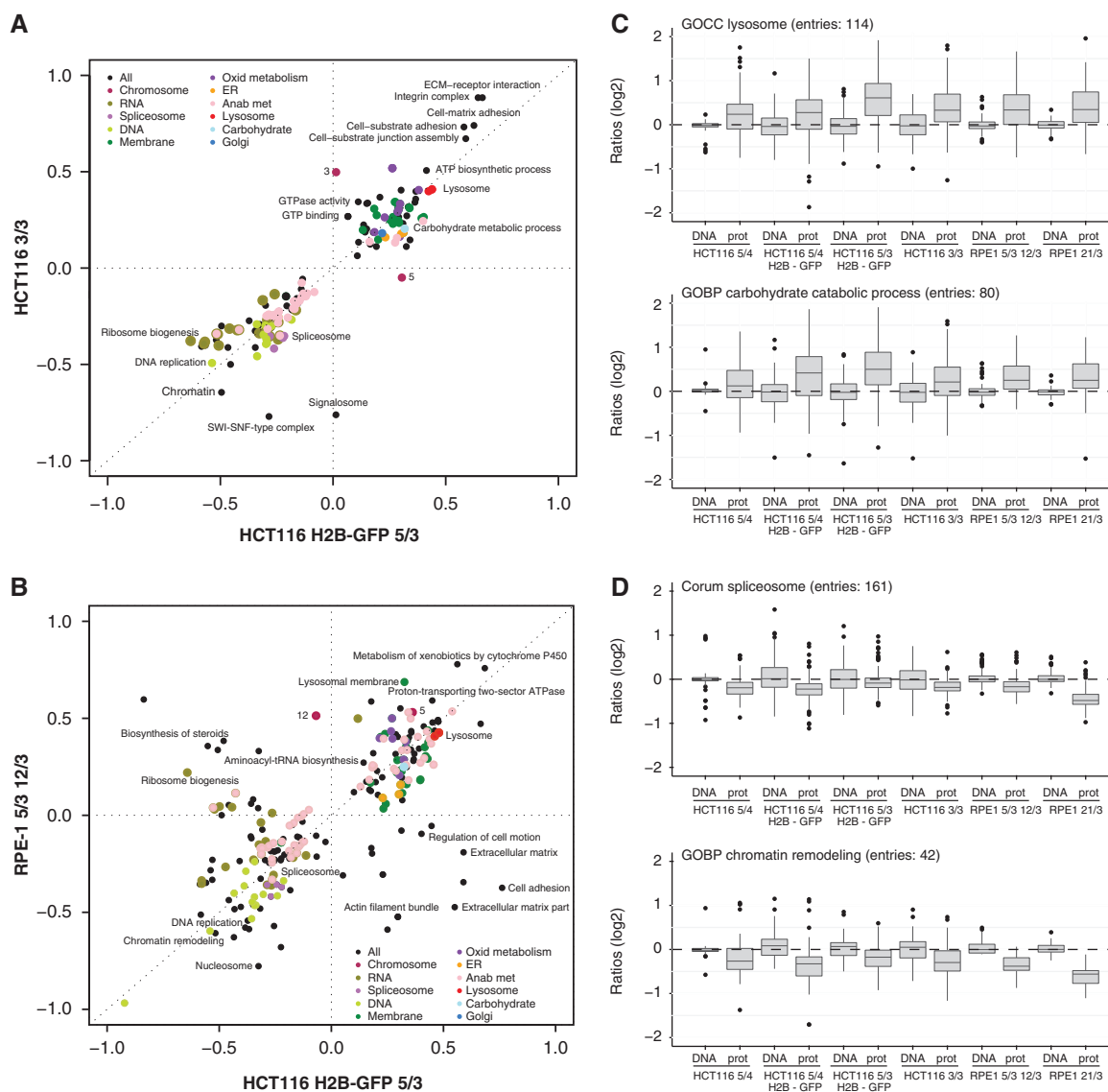
Figure 4 Uniform global response to presence of an extra chromosome. (**A**) Two-dimensional annotation enrichment analysis. Pathways altered in the cell line with chromosome 3 trisomy in comparison with the cell line with chromosome 5 trisomy are plotted (Benjamini-Hochberg FDR threshold 0.02). Each dot represents one category as defined in the KEGG and GO database, the colors mark groups of related pathways as described in the inset. Axis—position scores of the pathways; negative values indicate downregulation, positive values indicate upregulation. See Supplementary Information for further details. (**B**) Pathway alterations identified in HCT116 and RPE-1 cells show similar trends. (**C**, **D**) Box-plots of all proteins within representative categories identified as significantly altered by the two-dimensional annotation enrichment analysis. Examples of upregulated (lysosome, carbohydrate catabolic process) and downregulated (spliceosome, chromatin remodeling) pathways show uniform response among all cell lines with extra chromosomes. All differences between DNA and protein levels are statistically significant (unpaired *T*-test with Welch's correction, $P < 0.05$). For more information, see also Supplementary Figure S4 and Supplementary Table S2.

that is conjugated to phosphatidylethanolamine (PE) when integrated into the membrane destined for autophagosomes (Kabeya *et al*, 2000). By monitoring the doubly tagged mRFP-GFP-LC3 (Kimura *et al*, 2007) we found that the turnover of LC3 in aneuploid clones is similar to that of the diploid control (Figure 5D and E) and can be blocked by inhibition of autophagy with Bafilomycin A1 (Supplementary Figure S6), a drug inhibiting the acidification of lysosomes and fusion of autophagosomes to lysosomes (Klionsky *et al*, 2008). This confirms that the elevated numbers of autophagosomes are not due to a defect in autophagosome-to-lysosome fusion, but

indeed due to autophagy activation. In conclusion, our data suggest that autophagy is activated in trisomic and tetrasomic human cells.

Markedly, the expression of p62/sequestosome (SQSTM1) was enhanced in all analyzed trisomic and tetrasomic cells (Figure 5C). p62 is a cytoplasmic stress response receptor that is activated by various cellular stresses such as oxidative stress (for review, see Lamark and Johansen, 2009). Misfolded or damaged ubiquitylated proteins are sequestered by p62 into aggregates, and targeted to autophagy via a direct interaction with LC3 (Pankiv *et al*, 2007). Immunoblotting (Figure 5C) as
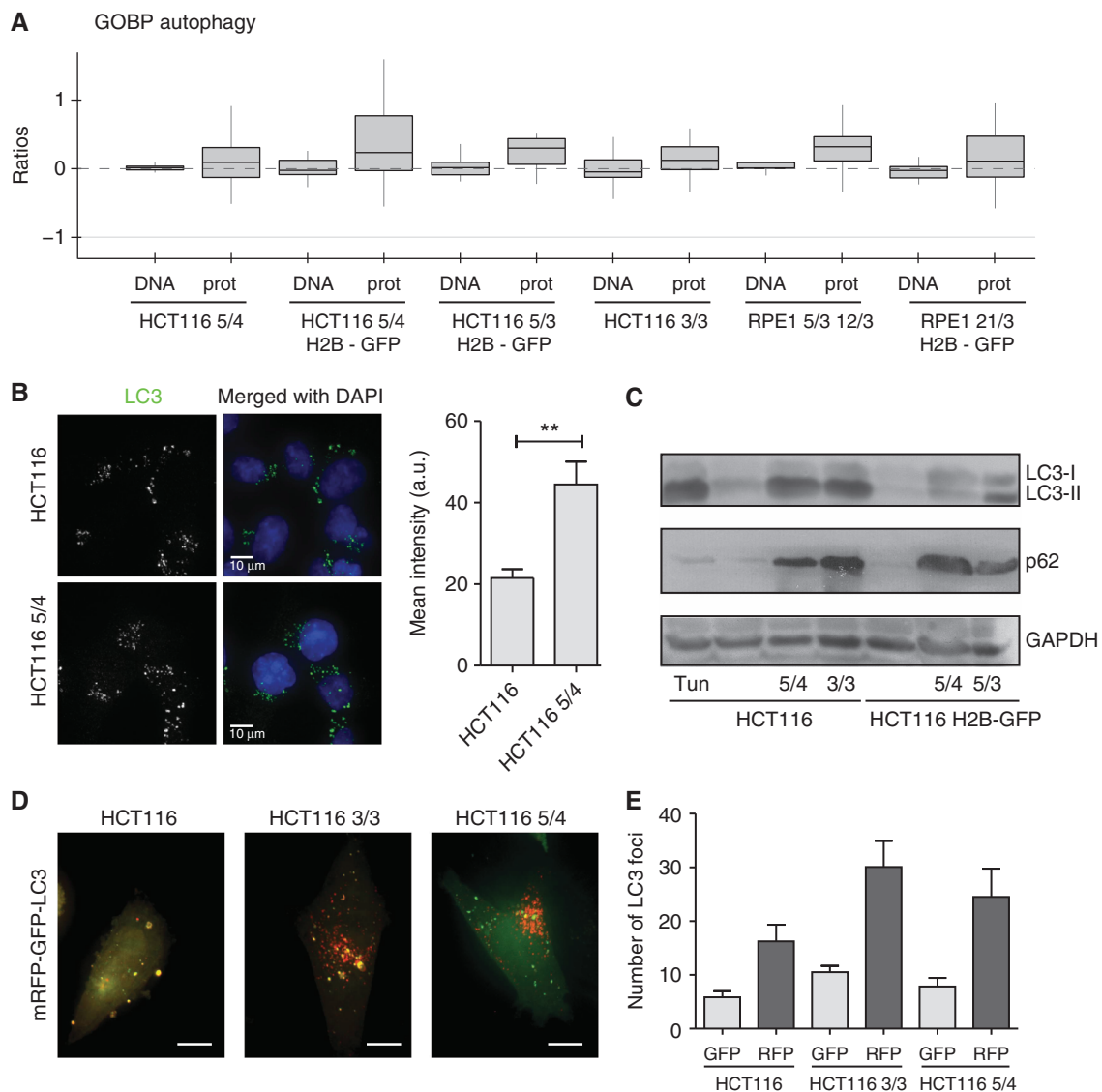
Figure 5 Activation of autophagy in trisomic and tetrasomic cell lines. (**A**) Abundance changes of the sub-category autophagy (GOBP) in all analyzed cell lines are depicted. CGH and proteome data are shown. (**B**) Fluorescence intensity of LC3-positive foci in HCT116 and HCT116 5/4. Right panel: quantification of the fluorescence intensity (non-parametric T-test, **$P < 0.01$). (**C**) Western blot of LC3-II shows an increase in all analyzed aneuploids. Similarly, levels of p62/SQSTM1 are increased in aneuploids. Note that p62 is coded on chromosome 5, but its levels are increased in all aneuploids. Tun—diploid HCT116 treated with tunicamycin that activates unfolded protein response and hence autophagy. (**D**) Representative images of cells after transfection with the double-tagged mRFP-GFP-LC3. Yellow foci represent phagosomes (both GFP and mRFP signals visible), red foci represent lysosomes (only mRFP signal is insensitive to the acidic pH in lysosomes). Bar 10 μm. (**E**) Total number of LC3 foci within a defined area of each cell (2500 voxels). There are significantly more foci in the HCT116 3/3 cell line (non-parametric T-test, $P < 0.01$), the levels in HCT116 5/4 are higher, but the difference from HCT116 is not statistically significant. See also Supplementary Figure S5. Source data is available for this figure in the Supplementary Information.

well as immunofluorescence staining showed a significant increase in p62 levels, which could be further enhanced by treatment with Bafilomycin A1 (Figure 6A). Moreover, indirect immunofluorescence revealed an increased co-localization of p62- and ubiquitin-positive foci in both the HCT116- and the RPE-1-derived cell lines, which is also enhanced by Bafilomycin (Figure 6B; Supplementary Figure S7A and B). Thus, our results suggest that cells with supernumerary chromosomes accumulate ubiquitylated proteins in the cytoplasm, likely as a consequence of the protein imbalance. The increase of p62-dependent autophagy could provide a new understanding of the pathways that allow aneuploid cells to maintain

protein homeostasis despite chronic elevated expression of multiple genes.

Discussion

To investigate the response of human cells to the presence of extra chromosomes, we generated human tri- and tetrasomic cell lines derived from diploid chromosomally stable cell lines HCT116 and RPE-1. The presence of additional chromosomes resulted in significant growth defect in particular in the G1 and S phase, further confirming the adverse effects of extra
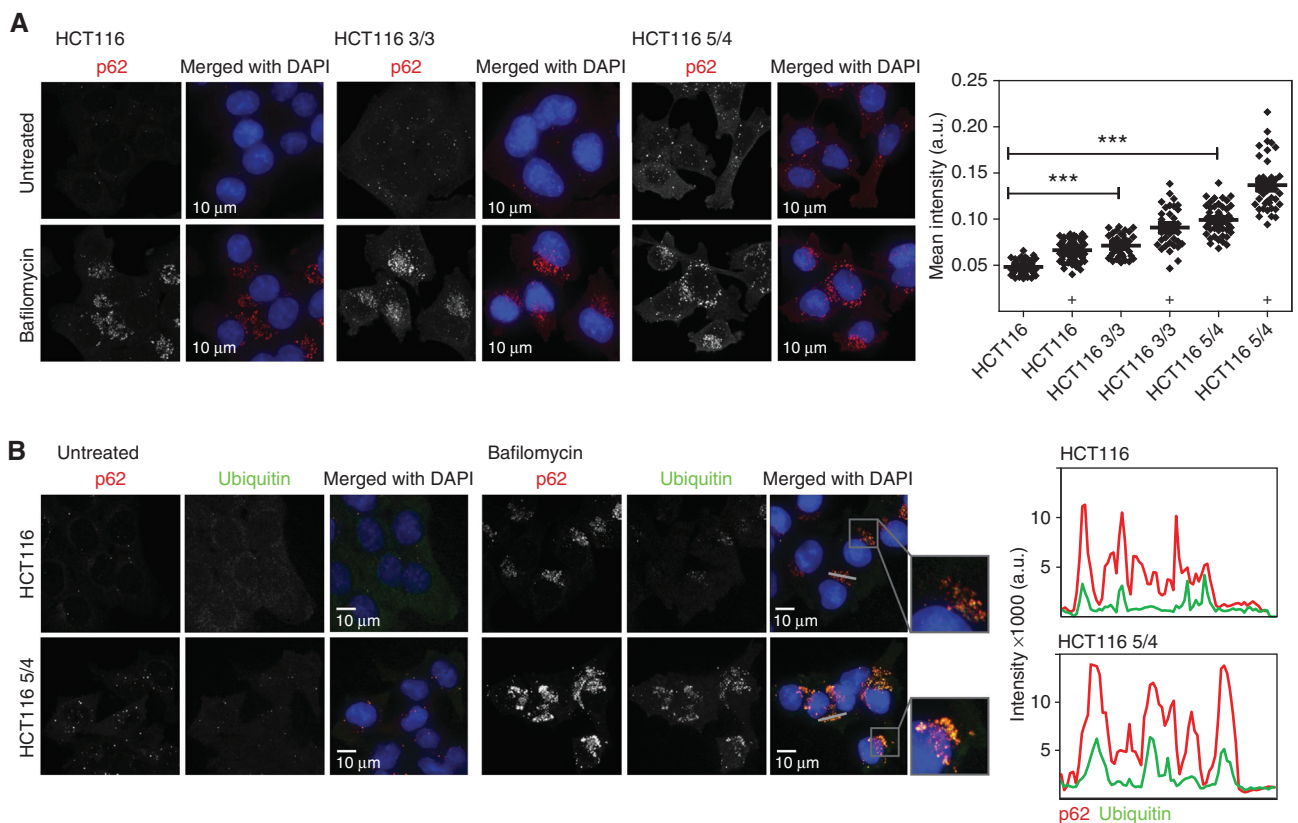
Figure 6 Analysis of autophagy in trisomic and tetrasomic cell lines. (**A**) Tri- and tetrasomic cells accumulate p62-positive foci. This effect was further increased by treatment with Bafilomycin A1. Right panel: Quantification of the fluorescence intensity changes; significance was evaluated by non-parametric *T*-test (***$P < 0.001$), '$+$' designates addition of Bafilomycin A1. (**B**) p62-positive foci co-localize with ubiquitin-positive foci and their fluorescence intensity is increased in aneuploid cell lines. Plots represent the signal intensity along the indicated gray line. Red line—p62, green line—ubiquitin. Source data is available for this figure in the Supplementary Information.

chromosomes on cellular growth that was observed in cells derived from Down syndrome patients (Segal and McCoy, 1974) or yeast and mouse cells with extra chromosomes (Torres *et al*, 2007; Williams *et al*, 2008). The cause of the growth delay remains unclear, but it has been proposed that it might result from a lack of energy, as these cells produce and degrade more proteins (Torres *et al*, 2007). Alternatively, the progression through the cell cycle might be delayed due to abnormal levels of cell-cycle regulators. Interestingly, highly aneuploid cancer cells do not show significant growth defects that would correlate with the extent of chromosomal alterations. This suggests that cells can adapt to the abnormal karyotype, and indeed, disomy tolerating mutations were recently identified in budding yeast (Torres *et al*, 2010). Our model tri- and tetrasomy cell lines provide a useful tool to identify the adaptations to abnormal karyotype in human cells and compare them with mutations frequently found in highly aneuploid cancers.

The comparison of DNA, mRNA and protein levels revealed that whereas mRNA abundance increases accordingly to the chromosome copy number changes, the abundance of $\sim 25\%$ of the proteins coded on the extra chromosomes is lower than expected. This adjustment of the protein abundance is found specifically for protein kinases and subunits of protein complexes. Our results show that the previous notion that multimolecular complexes often maintain stoichiometric

levels despite excess of one of the subunits can be generalized to most protein complexes. This question has been so far addressed only in budding yeast. On one hand, analysis of macromolecular complexes in disomic budding yeast revealed partial compensation of subunits of macromolecular complexes (Torres *et al*, 2010), whereas another study found no significant dosage changes in the protein abundance in budding yeast with complex aneuploid karyotype (Pavelka *et al*, 2010). Although some of these differences might be due to the technical issues as the changes in gene expression in aneuploids are subtle, another possible explanation is that in cells with multiple aneusomies more than one subunit of a multimolecular complex may be expressed at increased level. This would make the stoichiometry maintenance more complex and difficult to detect. Our results suggest that in human trisomic and tetrasomic cells the stoichiometry of protein subunits is often maintained despite the changes in gene copy numbers. Further analysis will be required to determine the underlying mechanisms.

Next, we asked whether there are global changes in pathway regulation in response to aneuploidy. Several pathways linked to RNA and DNA metabolism as well as cell-cycle regulators were strongly downregulated. This is in accordance with the growth defect of the model aneuploid cell lines. On the other hand, we observed upregulation of carbohydrate and oxidative metabolic processes, membrane metabolism, vesicle transport

[58]

and lysosome-related pathways. Neither general stress response pathways nor heat-shock response was activated, supporting the idea that chronic protein overexpression does not trigger the same stress response as an acute proteotoxic stress (Gidalevitz *et al*, 2006). Remarkably, the changes in pathway regulation were very similar in all HCT116- and RPE-1-derived cell lines. Thus, this first pathway analysis in human trisomic and tetrasomic cells showed that there is a general cellular response to abnormal karyotype regardless of the type of chromosome or origin of the cells. As the trisomy of chromosome 21 elicits the same signature response, one can hypothesize that some of these alterations might contribute to the pathology of the Down syndrome. It should be noted that some of the observed changes in pathway regulation might be an indirect consequence of the altered phenotypes of aneuploid cells, such as the growth defect. We propose that the effect is rather modest, because the changes in pathway regulation are not proportional to the growth defects. Previously, a transcriptional analysis of yeast disomic strains revealed general environmental stress response (Torres *et al*, 2007), whereas no aneuploidy-specific signature was identified in budding yeast strains with complex whole chromosome aneuploidies (Pavelka *et al*, 2010). Thus, it will be important to compare the identified signature response with pathway regulation changes in human cells with more complex aneuploid karyotypes.

The existence of specific response to aneuploidy suggests that aneuploidy itself might be a useful target in cancer therapy. Markedly, the pathways identified by our approach indeed point to the growth requirements of aneuploid cells. We found upregulation of energy metabolism pathways and increased activation of autophagy in human trisomic cell lines, which corresponds with the recent finding that the energy stress inducer 5-aminoimidazole-4-carboxamide riboside (AICAR) and the autophagy inhibitor chloroquine impair the growth of trisomic mouse cells (Tang *et al*, 2011). Autophagy inhibiting drugs are considered for cancer therapy, and human tri- and tetrasomic cell lines might provide a useful model for elucidating the molecular mechanisms underlying their effect.

What is the function of autophagy activation in human tri- and tetrasomic cell lines? Autophagy increases in response to nutrient deprivation, which might exist in human cells with extra chromosomes since they activate energy metabolism pathways. However, starvation-induced autophagy usually results in decreased levels of p62 due to its increased turn-over, whereas we identified elevated amounts of p62 in all analyzed tri-and tetrasomic human cell lines. We propose that the p62-mediated selective autophagy is specifically activated to maintain protein homeostasis in cells with extra chromosomes. Future research should address how autophagy is activated in aneuploid cells and whether it contributes to the dosage compensation of proteins that are coded on the extra chromosomes.

In summary, model human tri- and tetrasomic cell lines presented in this study provide a novel system to analyze the cellular response to the presence of an extra chromosome. Identification of pathways altered in the cell lines with supernumerary chromosome may help to elucidate the pathological changes associated with aneuploidies and identify possible treatments for related pathologies such as trisomy syndromes or cancer.

Materials and methods

Cell lines

The cell lines HCT116, HCT116 5/4 and HCT116 3/3 were kindly provided by Minoru Koi, Baylor University Medical Centre, Dallas, TX, USA. The cell line RPE-1 hTERT (hereafter RPE-1) and RPE-1 hTERT H2B-GFP were a kind gift of Stephen Taylor (University of Manchester, UK). HCT116 H2B-GFP as well as RPE-1-derived cell lines were created in our laboratory. HCT116 H2B-GFP was generated by lipofection (FugeneHD, Roche) of HCT116 (ATCC No. CCL-247) with pBOS-H2B-GFP (BD Pharmingen) according to manufacturer's protocols. All tri- and tetrasomic cell lines were generated by microcell fusion as described below. The donor mouse cell lines A9(Neo5) were purchased from the Health Science Research Resources Bank (HSRRB), Japan; the donor mouse cell line for a transfer of chromosome 21 was a kind gift of Professor Oshimura, Tottori University, Japan. All cell lines were maintained at 37°C with 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (FBS), 100 U penicillin and 100 U streptomycin. The cell lines HCT116 3/3, HCT116 H2B-GFP 5/3, HCT116 H2B-GFP 5/4, RPE-1 5/3 12/3, RPE-1 3/3, RPE-1 H2B-GFP 21/3 and A9(Neo5) were grown in media supplemented with 400 μg/ml G418. The cell line HCT116 5/4 as well as the cell lines stably transfected with H2B-GFP were grown in media supplemented with 6 μg/ml blasticidin S. Before each experiment, tri- and tetrasomic cells were grown for two passages in medium lacking the antibiotics to avoid any effect on protein levels.

Antibodies and reagents

As primary antibody anti-LC3 (115-3) from Medical & Biological Laboratories, anti-p62 from BD Transduction Laboratories (610832, for western blot), anti-p62 from Progen (GP62-C, for Immunofluorescence), anti-Ubiquitin (P4D1) from Santa Cruz, anti-β-Tubulin (ab21057) from Abcam, anti-Cyclin B (05-373) from Millipore and anti-GAPDH (ab9483) from Abcam were used. As secondary antibodies for western blotting, we applied goat anti-mouse IgG HRP affinity purified PAb (HAF007) and donkey anti-goat IgG HRP affinity purified PAb (HAF109) from R&D systems, Minneapolis, USA. For immunofluorescence analysis, we used anti-mouse FITC (Abcam), anti-guinea pig dylight649 and anti-mouse dylight649 (Jackson ImmunoResearch). 5 μg/ml tunicamycin (Sigma) in dimethylsulfoxide (DMSO) was used to induce autophagy in HCT116. Bafilomycin (Sigma) was applied for 16–18 h to a final concentration of 50 nM.

Microcell fusion

To generate cell lines containing an additional chromosome, microcell fusion (Supplementary Figure S1A) was performed as described previously (Fournier, 1981). In brief, mouse A9(Neo5) donor cells containing an additional human chromosome with an antibiotic resistance gene were treated for 48 h with colchicine (final concentration 60 ng/ml). Donor cells were trypsinized and seeded on plastic bullets. After the cells attached to the surface, bullets were centrifuged at 15 000 r.p.m. for 30 min at 30–34°C in DMEM supplemented with 10 μg/ml cytochalasin B. Cell pellets were resuspended in serum-free DMEM and filtered (Whatman, pore size 8 and 5 μm) to clear suspension from mouse cells. Filtered microcells were mixed with phytohemagglutinin (PHA-P) and added to the recipient cell line HCT116 H2B-GFP, RPE-1 or RPE-1 H2B-GFP. Fusion of microcells with the recipient cells was facilitated by polyethylene glycol 1500 (PEG 1500).

Genomic analysis

Genomic DNA for aCGH analysis was extracted using the Qiagen Gentra Puregene Kit following manufacturer's instructions. The aCGH

analysis was performed by IMGM laboratories, Martinsried, Germany. Multicolor FISH (mFISH) was performed by the Chrombios GmbH, Raubling, Germany. For further details on the procedures, see Supplementary Experimental Procedures.

Microarrays

mRNA was purified using the Qiagen mRNeasy mini kit. mRNA array analysis was conducted by IMGM laboratories, Martinsried, Germany; see Supplementary Experimental Procedures. The data have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al*, 2002) and are accessible through GEO Series accession number GSE39768 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE39768).

Cell-cycle analysis

Growth of the cells with extra chromosomes, cell-cycle progression after thymidine block release, cell death and number of senescent cells were analyzed according to previously established protocols. For details, see Supplementary Experimental Procedures.

Live-cell imaging

Freshly cultured cell were seeded sparsely in a 6 channel μ-slide (ibidi, Martinsried, Germany) 24 h before the experiment. Time laps movies were taken by imaging asynchronous cells in a 10 min or 4 min interval for 72 or 48 h, respectively. The slide was placed onto a sample stage with an incubator chamber (EMBLEM, Heidelberg, Germany) maintained at a 37°C, 40% humidity, in an atmosphere of 5% CO₂. Imaging was performed using a Zeiss Axio Observer Z1 microscope equipped with a Plan Neofluar 20 × air objective. Metamorph 7.1 software (Molecular Devices) was used to control the microscope. Movies were evaluated using the ImageJ 1.42I software.

Indirect immunofluorescence

Cells were seeded either on glass slides or in an ibidi 8-well slide, grown to 50–70% confluency. The cells were fixed 2 days after seeding by 3.7% paraformaldehyde. To inhibit autophagy, cells were treated 18 h with bafilomycin at a final concentration of 50 nM and then fixed. Cells were labeled with anti-LC3, anti-p62 or anti-ubiquitin using the described antibodies. For p62 immunofluorescence intensity analysis, the cell cytoplasm was additionally stained by HCS cell mask red dye (Invitrogen) before fixation to allow automated cell segmentation. Mean intensity of p62/cell was analyzed using Cell Profiler software. Details on the quantitative analysis are in Supplementary Experimental Procedures.

Autophagy flux assay

Cells were seeded 1 day before transfection to achieve 40% confluency at the day of transfection. The ptfLC3 plasmid (mRFG-GFP-LC3, Addgene) was transfected using Lipofectamin LTX™ and PLUS™ reagent according to manufacturer's protocol. Images were taken 2 days later. We quantified the number of GFP-positive and mRFP-positive LC3 foci within a fixed area (2500 voxels) and the co-localization of GFP- and RFP-positive foci. Details on the quantitative analysis are in Supplementary Experimental Procedures.

Fluorescence microscopy

The images were taken by a fully automated Zeiss inverted microscope (AxioObserver Z1) equipped with a MS-2000 stage (Applied Scientific Instrumentation, Eugene, OR), the CSU-X1 spinning disk confocal head (Yokogawa), LaserStack Launch with selectable laser lines (Intelligent Imaging Innovations, Denver, CO) and an X-CITE Fluorescent Illumination System. Images were captured using a CoolSnap HQ camera (Roper Scientific) under the control of the

Slidebook software (Intelligent Imaging Innovations, Denver, CO). Signals were imaged with a 100 × oil objective by using a 561 nm (mRFP) and 473 nm (GFP) laser and a UV light.

SILAC labeling

Cells were cultured in DMEM (high glucose) devoid of arginine and lysine supplemented with 10% dialyzed FBS (10 kDa cutoff; Invitrogen), 1 × penicillin/streptomycin. Arginine and lysine (Sigma-Aldrich) were added in either light (Arg0; Lys0) or heavy (Arg10; Lys8) form to a final concentration of 33.6 μg/ml for arginine and 73 μg/ml for lysine. The cells were tested for full incorporation after the labeling. All experiments were performed with double labeling.

Protein sample preparation and analysis

Samples were prepared following the filter-aided sample preparation (FASP) protocol. In brief, the cells were lysed with 4% sodium dodecyl sulfate (SDS), 0.1 M dithiothreitol (DTT) in 100 mM Tris/HCl pH 7.6 followed by an incubation for 5 min at 95°C and subsequent sonication for 15 min. Heavy and light samples from one experiment were mixed with equal protein amounts. In all, 200–300 μg protein mix was alkylated and digested overnight with trypsin on Microcon YM-30 filter tubes (Millipore). Peptides were eluted from the membrane using 0.5 M NaCl or ddH₂O and separated either using the OFFGEL fractionator (HCT116-derived samples) or a strong anion exchange fractionation (RPE-1-derived samples). Fractions were desalted with StageTips (Empore disk, C₁₈ Reversed phase) before liquid-chromatography mass spectrometry (LC-MS)/MS analysis using CID fragmentation.

Liquid-chromatography mass spectrometry

Peptide mixtures were analyzed using nanoflow liquid chromatography (LC-MS/MS) on an EASY-nLC™ system (Proxeon Biosystems, Odense, Denmark) online connected to the LTQ Orbitrap XL or LTQ Orbitrap Velos instrument (Thermo Fisher Scientific, Bremen, Germany) through a Proxeon nanoelectrospray ion source. Peptide samples were directly autosampled onto a 15-cm in-house packed column (75 μl inner diameter; Proxeon Biosystems) with 3 μm reversed phase beads (ReproSil-Pur C18-AQ, Dr Maisch). Using a 170-min gradient (2–30% ACN), peptides were directly electrosprayed (2.2 kV) into the mass spectrometer. Mass spectrometer was operated in data-dependent mode switching automatically between one full scan MS and 7–15 MS/MS acquisitions. Instrument control was through Tune 2.6.0 and Xcalibur 2.1.0. Full scan MS spectra (m/z 300–1650) were acquired in the Orbitrap analyzer after accumulation to a target value of 10^6 in the linear ion trap (resolution of 60 000 at 400 m/z). Fragmentation was performed in CID mode applying 35% normalized collision energy after accumulation of the parent ions to a target value of 5000.

Data analysis

All double labeling SILAC experiments were analyzed together per cell line using the in-house developed software MaxQuant (version 1.0.14.10 and 1.2.0.25) with standardized workflow. Data were searched against the human International Protein Index protein sequence database (ipi.HUMAN.v3.62.dec) supplemented with frequently observed contaminants and concatenated with reversed copies of all sequences (target-decoy database). False discovery rates on peptide and protein level were fixed to 1%, including automatic filtering on peptide length, mass error precision estimates, and peptide scores of all forward and reversed peptide identifications. Reported protein groups had to be identified by at least one unique peptide to be accepted. Quantitation was based on unique and razor peptides only and a minimum of two ratio counts was required. Complete protein and peptide lists as well as the underlying *.RAW files are available on the TRANCHE database (https://proteomecommons.org/tranche/). For details, see Supplementary Experimental Procedures.

Statistics

All statistically evaluated experiments were performed in at least three independent biological replicates. The statistical evaluation was performed using GraphPad Prism 5 software, or R. All details on statistical analyses are in Supplementary Experimental Procedures.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aït Yahya-Graison E, Aubert J, Dauphinot L, Rivals I, Prieur M, Golfier G, Rossier J, Personnaz L, Créau N, Bléhaut H, Robin S, Delabar JM, Potier MC (2007) Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes. *Am J Hum Genet* **81:** 475–491

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. *Trends Genet* **21:** 219–226

Chen G, Bradford WD, Seidel CW, Li R (2012) Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* **482:** 246–250

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210

Fournier RE (1981) A general high-efficiency procedure for production of microcell hybrids. *Proc Natl Acad Sci USA* **78:** 6349–6353

Gidalevitz T, Ben-Zvi A, Ho KH, Brignull HR, Morimoto RI (2006) Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311:** 1471–1474

Goldberg AL, Dice JF (1974) Intracellular protein degradation in mammalian and bacterial cells. *Annu Rev Biochem* **43:** 835–869

Guialis A, Morrison KE, Ingles CJ (1979) Regulated synthesis of RNA polymerase II polypeptides in Chinese hamster ovary cell lines. *J Biol Chem* **254:** 4171–4176

Guo M, Birchler JA (1994) Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science* **266:** 1999–2002

Hassold T, Hall H, Hunt P (2007) The origin of human aneuploidy: where we have been, where we are going. *Hum Mol Genet* **16:** R203–R208

Haugen AC, Goel A, Yamada K, Marra G, Nguyen TP, Nagasaka T, Kanazawa S, Koike J, Kikuchi Y, Zhong X, Arita M, Shibuya K, Oshimura M, Hemmi H, Boland CR, Koi M (2008) Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer Res* **68:** 8465–8472

He C, Klionsky DJ (2009) Regulation mechanisms and signaling pathways of autophagy. *Annu Rev Genet* **43:** 67–93

Huettel B, Kreil DP, Matzke M, Matzke AJ (2008) Effects of aneuploidy on genome structure, expression, and interphase organization in Arabidopsis thaliana. *PLoS Genet* **4:** e1000226

Kabeya Y, Mizushima N, Ueno T, Yamamoto A, Kirisako T, Noda T, Kominami E, Ohsumi Y, Yoshimori T (2000) LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J* **19:** 5720–5728

Kimura S, Noda T, Yoshimori T (2007) Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3:** 452–460

Klionsky DJ, Abeliovich H, Agostinis P, Agrawal DK, Aliev G, Askew DS, Baba M, Baehrecke EH, Bahr BA, Ballabio A, Bamber BA, Bassham DC, Bergamini E, Bi X, Biard-Piechaczyk M, Blum JS, Bredesen DE, Brodsky JL, Brumell JH, Brunk UT *et al* (2008) Guidelines for the use and interpretation of assays for monitoring autophagy in higher eukaryotes. *Autophagy* **4:** 151–175

Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. *PLoS Genet* **4:** e1000223

Lamark T, Johansen T (2009) Autophagy: links with the proteasome. *Curr Opin Cell Biol* **22:** 192–198

Makarevitch I, Phillips RL, Springer NM (2008) Profiling expression changes caused by a segmental aneuploid in maize. *BMC Genomics* **9:** 7–16

Masramon L, Ribas M, Cifuentes P, Arribas R, García F, Egozcue J, Peinado MA, Miró R (2000) Cytogenetic characterization of two colon cell lines by using conventional G-banding, comparative genomic hybridization, and whole chromosome painting. *Cancer Genet Cytogenet* **121:** 17–21

Niwa O, Tange Y, Kurabayashi A (2006) Growth arrest and chromosome instability in aneuploid yeast. *Yeast* **23:** 937–950

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1:** 376–386

Pankiv S, Clausen TH, Lamark T, Brech A, Bruun JA, Outzen H, Øvervatn A, Bjørkøy G, Johansen T (2007) p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J Biol Chem* **282:** 24131–24145

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468:** 321–325

Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW (2010) CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res* **38:** D497–D501

Segal DJ, McCoy EE (1974) Studies on Down's syndrome in tissue culture. I. Growth rates and protein contents of fibroblast cultures. *J Cell Physiol* **83:** 85–90

Selmecki A, Forche A, Berman J (2006) Aneuploidy and isochromo-some formation in drug-resistant Candida albicans. *Science* **313:** 367–370

Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, Brito IL, Hiraoka Y, Niwa O, Amon A (2011) Aneuploidy drives genomic instability in yeast. *Science* **333:** 1026–1030

Stenberg P, Lundberg LE, Johansson A-M, Rydén P, Svensson MJ, Larsson J (2009) Buffering of Segmental and Chromosomal Aneuploidies in *Drosophila melanogaster*. *PLoS Genet* **5:** e1000465

Storchova Z, Pellman D (2004) From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol* **5:** 45–54

Tang YC, Williams BR, Siegel JJ, Amon A (2011) Identification of aneuploidy-selective antiproliferation compounds. *Cell* **144:** 499–512

Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, Dunham MJ, Amon A (2010) Identification of aneuploidy-tolerating mutations. *Cell* **143:** 71–83

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317:** 916–924

Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T (2004) Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Res* **64:** 6941–6949

Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, Amon A (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322:** 703–709

[62]

2.2. A SILAC-based approach identifies substrates of caspase-dependent cleavage upon TRAIL-induced apoptosis

2.2.1. Aim and Summary

Apoptosis is a crucial mechanism in multicellular organisms to maintain cell homeostasis and remove damaged or possibly dangerous cells. On the one hand, during the apoptotic process caspases, aspartate-specific proteases, cleave hundreds of proteins to effect the demolition of the cell. On the other hand, evasion of apoptosis represents a hallmark of cancer. For this reason, it is important to elucidate basic processes after apoptosis induction to support research into the development of cancer therapeutics. The extrinsic death ligand TRAIL is of special interest in this context, since it has been shown to specifically induce apoptosis in cancer cells without causing cytotoxic effects in normal cells.

So far, no global proteome study has yet characterized cleavage events induced by TRAIL by focusing on all cleavage substrates and corresponding pathways. We therefore performed quantitative mass spectrometry-based proteomics on TRAIL-induced SILAC-labeled Jurkat T cells and compared them to untreated cells. We separated proteins via SDS-PAGE and excised the gel along the molecular weight into several slices. Each slice was subsequently subjected to tryptic digestion and resulting peptides were analyzed by mass spectrometry. In addition to standard MS data analysis we developed a bioinformatic algorithm to extract cleaved proteins. In our study, we identified close to 700 cleaved substrates upon TRAIL induction. We further inspected the individual substrates and investigated the fate of protein complexes in cell death. We found that generally not only single components of a complex are cleaved by caspases but several subunits are attacked. These, however, are cleaved with lower stoichiometry than proteins which are not in complexes. As an interesting example, we studied condensin I, of which hitherto only one component has been identified as a known substrate. With our approach, we identified four additional subunits of the complex as cleavage substrate after TRAIL-induced apoptosis.

Our study provides for the first time an in-depth catalog of cleaved substrates in response to extrinsic, TRAIL-induced apoptosis. This was made possible by the development of a sophisticated algorithm for the extraction of cleaved proteins that can be applied to any topic of proteolytic cleavage. Our investigation thus sheds new light on an important and biomedically relevant cellular pathway.

2.2.2. Contribution

This work resulted from my second main thesis project, the identification of caspase-dependent cleaved substrates after TRAIL-induced apoptosis. Johannes Graumann provided substantial intellectual input at the initial phase of the project. I designed and performed all experiments and developed the ideas for the cleavage site extraction. Christoph Schaab provided the hard core bioinformatics input and wrote the MATLAB script for the identification of cleaved substrates. I then carried out the downstream analyses. Moreover, I designed all figures and tables for the publication and wrote the manuscript with the help of Matthias Mann.

2.2.3. Manuscript

This manuscript is currently in revision in Molecular and Cellular Proteomics:

A SILAC-based approach identifies substrates of caspase-dependent cleavage upon TRAIL-induced apoptosis

Stoehr Gabriele, Schaab Christoph, Graumann Johannes and Mann Matthias

Submitted to Mol Cell Proteomics

A SILAC-based approach identifies substrates of caspase-dependent cleavage upon TRAIL-induced apoptosis

Gabriele Stoehr[1], Christoph Schaab[1,2], Johannes Graumann[3] and Matthias Mann[1,]*

[1]Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

[2]Evotec (Munich) GmbH, Am Klopferspitz 19a, D-82152 Martinsried, Germany

[3]Weill Cornell Medical College in Qatar, Qatar Foundation, Education City, Doha, State of Qatar

*To whom correspondence should be addressed. E-mail: mmann@biochem.mpg.de

Running title: SILAC-based approach to globally detect protease substrates

Abbreviations: **AcOH**, acetic acid; **AGC**, automatic gain control; **CASP8**, Caspase-8; **ER,** endoplasmic reticulum; **EtOH**, Ethanol; **FASP**, Filter-aided sample preparation; **FBS**, fetal bovine serum; **FDR**, false discovery rate; **FMK**, fluoromethylketone; **IAA**, Iodoacetamide; **MOMP**, mitochondrial outer membrane permeabilization; **PI**, Propidium iodide; **SILAC**, stable isotope labeling with amino acids in cell culture; **TRAIL**, TNF-related apoptosis inducing ligand; **XIC**, extracted ion chromatogram;

Summary

The extrinsic, extracellular ligand induced pathway of apoptosis is executed by caspase protease cascades activating downstream effectors by site-directed proteolysis. Here we identify proteome changes upon induction of apoptosis by the cytokine Tumor Necrosis Factor-Related Apoptosis Inducing Ligand (TRAIL) in a Jurkat T cell line. We detect caspase-dependent cleavage substrates by quantifying protein intensities before and after TRAIL induction in SDS gel slices. Apoptotic protein cleavage events are identified by a characteristic SILAC ratio pattern across gel slices that results from differential migration of the cleaved versus the uncleaved protein. We apply a statistical test to define apoptotic substrates in the proteome. Our approach identified more than 650 of these cleaved proteins in response to TRAIL-induced apoptosis, including many previously unknown substrates and cleavage sites. Inhibitor-treatment combined with triple-SILAC demonstrated that the detected cleavage events are caspase-dependent. Proteins located in lumina of organelles such as mitochondria and endoplasmic reticulum were significantly under-represented in the substrate population. Interestingly, caspase cleavage is generally observed not only in one but several members of stable complexes, but often with lower stoichiometry. For instance, all five proteins of the condensin I complex were cleaved upon TRAIL treatment. The apoptotic substrate proteome data can be accessed and visualized in the MaxQB database and may prove useful for basic and clinical research into TRAIL-induced apoptosis. The technology described here is extensible to a wide range of other proteolytic cleavage events.

Apoptosis is an essential cellular mechanism regulating normal physiological processes, for instance in development (1, 2). As a specific and programmed form of cell death, apoptosis leads to a controlled disposal of cells relying on proteolytic cleavage of specific protein substrates as the central mechanism. Moreover, selective induction of apoptosis plays an important role in diseases such as cancer. It is thus important to determine the proteins and pathways involved in apoptosis and to characterize their relevance for induction of the pathway by different stimuli. The main players in apoptosis are caspases (Cys-dependent Asp-specific proteases), which specifically cleave C-terminally of an aspartate, and are represented by 11 functional genes in the human genome. Auto-cleavage of upstream caspases in response to induction of apoptosis initiates a cascade of downstream cleavage events resulting in activation or inactivation as well as translocation of these substrates (3-6).

In recent years several methods have been introduced to identify substrates cleaved in a caspase-dependent manner as well as the exact location of cleavage sites. In vitro approaches such as incubation of peptides or protein libraries with the active protease of interest have led to the identification of substrate motifs but do not necessarily represent *in vivo* events in the context of an intact cell (7). Mass spectrometry (MS)-based methods employed may be divided into those directed towards the detection of the peptides cleaved by the protease and those applied at the proteome level for the identification of substrates and down-stream effects. The former methods generally chemically block all pre-existing N-termini and subsequently take advantage of the de novo generation of N-termini by the protease, which serves as a handle to enrich and detect the corresponding peptide. The original method in this class was N-terminal COFRADIC (8, 9) and it was used to detect Fas-induced cleavage events in Jurkat T cells (10). Recent peptide selective methods include exclusive labeling of novel N-termini by biotin, followed by capturing the peptides on avidin columns (11), but others – using both positive as well as negative selection strategies – have also been described (12-15). Peptide-based methods have the advantage of directly identifying the

cleavage site. Information about the substrate protein, however, is limited to a single peptide per cleavage event and this peptide may not be readily detectable.

Global approaches applied on the proteome level, on the other hand, attempt to identify as many cleaved proteins as possible without necessarily determining the exact cleavage site within the protein (16, 17). In many cases, two-dimensional gel electrophoresis has been used to differentiate cleaved and uncleaved protein states (18-20). This approach, however, lacks reproducibility, sensitivity and throughput. A pioneering study based on single dimensional SDS page identified several previously known cleaved proteins in Fas-induced Jurkat T cells by their molecular mass shift, yet remained small in scale (21). The approach was further extended by using quantitative ratio information of the proteins in combination with offline LC-MALDI-MS/MS (22).

Building on these studies, Cravatt and coworkers extended the global approach that takes into account information about molecular weight differences between uncleaved and corresponding cleaved proteins – termed PROTOMAP (23). Identified peptides were mapped onto the protein sequences to provide information about the location of the cleavage products. Since label-free quantitation was used and samples consequently processed in parallel, the approach is labor-intensive and requires high reproducibility of the experiments.

To circumvent limitations of the methods mentioned above, we set out to develop a quantitative SILAC-based method for the identification of apoptosis-dependent substrates of protein cleavage on the proteome level but making use of accurate peptide-based ratio information for each substrate protein. In this way we aimed to combine advantages of each of the above approaches. Furthermore, we included in our analysis a combination of a high-resolution LC-MS/MS workflow with stringent statistical analysis of the results.

Several previous proteomics studies of apoptosis have investigated the effects of *intrinsic* stimuli. In contrast, despite its importance for the development of cancer therapeutics, no global proteome

study has characterized cleavage events induced by the *extrinsic* stimulus TRAIL, apart from one targeted study focusing on cleavage kinetics (24). In that study, TRAIL (extrinsic) and staurosporine (intrinsic) treated samples were compared by using selected reaction monitoring of approximately 1,000 caspase-derived peptides. In our study, we performed triple SILAC-labeling in the well-established system of TRAIL-induced apoptosis in Jurkat T cells, which allowed us to validate the caspase dependence of the detected cleavage events. Our study sheds light on the fate of protein complexes in cell death and provides a first in-depth overview of cleaved substrates in response to extrinsic, TRAIL-induced apoptosis.

During preparation of this manuscript, a report by Cravatt group extended their PROTOMAP approach (23) using SILAC, thereby making it quantitative (25). That study elegantly extracted information on the crosstalk between phosphorylation events in the cleavage motifs and the regulation of the proteolytic pathway in apoptosis. It thereby underscored the necessity and potential of quantitative approaches for the study of proteolytic events in a global manner.

EXPERIMENTAL PROCEDURES

Evaluation of TRAIL treatment and inhibitor concentrations – Jurkat T cells (0.4xE6 cells/ml) were treated with either 50 ng/ml, 100 ng/ml, 200 ng/ml, 500 ng/ml Human TRAIL Apo-II Ligand (PeproTech, Inc., Rocky Hill, NJ; 50 µg; stock 100 ng/µl) or mock and incubated at 37°C, 5% CO_2 for 24 h. Cell growth was regularly examined and cell morphology was checked by 5 h light microscopy and cell counter (Countess; Invitrogen). Based on these experiments, a concentration of 100 ng/ml was selected for further analyses.

Jurkat T cells were treated with varying concentrations of the pan-caspase inhibitor z-VAD-FMK (0.1 µM, 1 µM, 10 µM and 50 µM in 2ml (0.75xE6 cells/ml); R&D Systems; 1 mg) 10 min prior to 5 h

incubation with 100 ng/ml TRAIL and mock, respectively. Cells were washed twice with ice-cold PBS and cell pellets were frozen at -80°C until further analysis using western blotting.

Flow cytometry analysis – For the detection of apoptotic and necrotic cells, the Annexin V-FITC Kit (Apoptosis Detection Kit; Beckman Coulter; Immunotech) was used. Cells were washed and resuspend in ice-cold water. An antibody mix for PI/Annexin V-FITC (1:20 PI and 1:100 Annexin-V FITC) was prepared and added to the cells in Binding Buffer. Cells were kept on ice until flow cytometric analysis. Cells were analyzed after 0, 1, 2, 4 and 5 h of TRAIL induction.

SILAC labeling – Jurkat T cells were cultured in RPMI medium (Gibco® Invitrogen; High glucose; GlutaMAX^{TM}; devoid of arginine and lysine) supplemented with 10% dialyzed fetal bovine serum (FBS; Invitrogen; 10 kDa cut-off) and 1X penicillin/streptomycin. For SILAC labeling, arginine and lysine were added in either light (Arg0; Lys0) or heavy (Arg10; Lys8) form to a final concentration of 33.6 µg/ml for arginine and 73 µg/ml for lysine. For triple labeling, cells were additionally cultured in medium containing Arg6 and LysD4 using the same concentrations. L-arginine (Arg0), L-lysine (Lys0), L-$^{13}C_6$-arginine (Arg6), L-D$_4$-lysine (LysD4), L-$^{13}C_6^{15}N_4$-arginine (Arg10) and L-$^{13}C_6^{15}N_2$-lysine (Lys8) were purchased from Sigma-Aldrich. Prior to treatment Jurkat T cells were grown for 8-10 passages in SILAC medium and tested for full incorporation.

Treatment with TRAIL – Equal numbers of Jurkat T cells (0.4-0.6 x E6 cells/ ml) were treated for 5 h with 100 ng/ml TRAIL and mock in heavy and light medium, respectively. For the inhibition study, medium labeled cells were additionally pre-treated for 10 min with 50 µM z-VAD-FMK. TRAIL treatment was stopped by adding ice-cold PBS to the cells. Cell suspensions were centrifuged (5 min, 4°C, 400 g) and the supernatant was discarded. Cell pellets from the corresponding heavy and light cultures (or heavy, medium and light for triple labeling experiments) were resuspended in a small

aliquot of ice-cold PBS and combined in one tube. Cells were once more washed with ice-cold PBS and centrifuged. Cell pellets were shock-frozen in liquid nitrogen and stored at -80°C.

Cell lysis and protein digestion – Proteins were extracted and digested following the first steps of the filter-aided sample preparation (FASP) protocol (26). Cells were lysed with 4% SDS, 0.1 M DTT in 100 mM Tris/HCl pH 7.6 followed by incubation for 5 min at 95°C. Lysates were sonicated and cleared by centrifugation with high speed (16 000 x g). Protein concentrations were measured by a tryptophan-fluorescence assay. 150 μg of the samples were used for further analyses. Proteins were reduced with DTT (10 mM) for 45 min at room temperature, followed by alkylation for 30 min in the dark (55mM iodoacetamide (IAA)).

Proteins were mixed with LDS sample buffer (final: 1x; Invitrogen) and samples were boiled at 70°C for 10 min. 50 μg of protein per sample were loaded on a polyacrylamide gel (Invitrogen, Bis-Tris Gel, 4-12%, MOPS; 10 pockets) in each of three adjacent lanes (total 150 μg per sample). Proteins were separated with 180 V for 45 min. Proteins were fixed and stained in the gel using standard protocols (Colloidal Blue Staining Kit, Invitrogen).

The gel was cut in 28-36 slices and corresponding slices from the three lanes were combined in one tube. Proteins in the gel were digested following standard in-gel protocols (27). Briefly, gel pieces were destained in consecutive wash steps and dehydrated using 50% and 100% Ethanol (EtOH), respectively. Gel pieces were dried in a SpeedVac concentrator for 5 min. Proteins were digested by adding sequencing grade-modified trypsin (12.9 ng/ml; Promega, Madison, WI) to the gel pieces followed by an overnight incubation at 37°C. Peptides were subsequently extracted from the gel pieces using increasing concentrations of ACN (30% - 100% ACN) in separate steps. Organic solvent was removed in a SpeedVac concentrator and peptides were desalted on reversed phase C_{18} StageTips (Empore disk; (28)) prior to LC-MS/MS analysis.

Western Blot experiments – Cell pellets were lysed and proteins were separated by SDS-PAGE as mentioned before. As positive and negative controls 10-15 µl Jurkat Apoptosis Cell Lysates (Etoposide-treated and untreated; Cell Signaling) were used. For CASP8 detection 12% NuPAGE Bis-Tris Gels (Invitrogen) were used. For PARP1 analysis 4-12% Bis-Tris gels were used. All gels were run with MOPS buffer. Proteins were transferred onto a nitrocellulose membrane using a vertical Blotting system for 1 h at 100 V. Primary antibodies in 1% BSA or nonfat dry milk were added to the membrane for 1 h at RT or overnight at 4°C. The 2nd antibody (in T-TBS (Tween 20 Tris-buffered saline buffer)) was added for 30 min-1 h at RT. Antibodies were used as follows: PARP (46D11) Rabbit mAb (Invitrogen), 1:1000 in milk; Caspase-8 (1C12) Mouse mAb (Invitrogen), 1:1000 in BSA; GAPDH Rabbit, 1:1000 in milk or 5% BSA (Invitrogen), 1:3000 2nd Ab rabbit (GE Healthcare); 1:10000 2nd Ab mouse (Jackson ImmunoResearch Laboratories, Inc.).

Liquid-Chromatography Mass Spectrometry (LC-MS) – Peptide mixtures were analyzed using nanoflow liquid chromatography (LC-MS/MS) on an EASY-nLC system (Proxeon Biosystems, Odense, Denmark; now Thermo Fisher Scientific) on-line coupled to an LTQ Orbitrap XL or LTQ Orbitrap Velos instrument (Thermo Fisher Scientific, Bremen, Germany; (29, 30)) through a nanoelectrospray ion source (Proxeon). Approximately 4 µg of the peptide samples in 5 µl were directly loaded onto a 15 cm column with 75 µl inner diameter, packed in-house with 3 µm reversed phase beads (ReproSil-Pur C18-AQ, Dr. Maisch GmbH). Peptides were separated and directly electrosprayed into the mass spectrometer using a 145 min method including a linear gradient from 5% to 30% ACN in 0.5% acetic acid (AcOH) over 93-97 min at a constant flow of 250 nl/min.

The LTQ Orbitrap XL and LTQ Orbitrap Velos were operated in data-dependent mode switching automatically between full scan MS and MS/MS acquisition. Instrument control was through Tune 2.6.0. and Xcalibur 2.1.0. Full scan MS spectra (m/z 300 − 1650) were acquired in the Orbitrap analyzer after accumulation to a target value of 10^6 in the linear ion trap. Spectra were acquired with a resolution of 60,000 at 400 m/z. The 5 (LTQ Orbitrap XL) and 15 (LTQ Orbitrap Velos) most intense

ions with charge states ≥ +2 were sequentially isolated with a target value of 5,000 and fragmented using collision-induced dissociation (CID) in the linear ion trap with normalized collision energy of 35%. The activation q was set to 0.25 and the activation time was set to 30 ms and 10 ms for LTQ Orbitrap and LTQ Orbitrap Velos, respectively. The ion selection threshold was set to 500 counts for CID-MS/MS. Maximum ion accumulation times of 1000 ms and 500 ms for full scans and 150 ms and 25 ms for CID-MS/MS scans were set for the LTQ Orbitrap XL and the LTQ Orbitrap Velos, respectively. Dynamic exclusion was 90 s with early expiration enabled (count: 2; S/N threshold: 2). Standard mass spectrometry parameters were set for all experiments as follows: 2.2 kV spray voltage; no sheath and auxiliary gas; 200°C heated capillary temperature; predicted and normal automatic gain control (AGC) enabled for Velos analyses, for Orbitrap data normal AGC was enabled; 110 V Tube lens voltage (LTQ Orbitrap) and 50-60% S-lens RF level (LTQ Orbitrap Velos), respectively; if used, a lock mass of m/z 445.120024 was applied (29); for LTQ Orbitrap Velos measurements, the lock mass abundance was set to 0%.

Sample processing – For the experiments DOUBLE_E1_R0K0, DOUBLE_E2_R0K0 and DOUBLE_E2_R10K8, Jurkat T cells were labeled in both heavy and light SILAC medium. Amino acids marked in the experiment title indicate the labeling of the TRAIL-induced cells. Reverse labeling was performed as well. For triple labeling experiments (TRIPLE_Inh_TRAIL_M1-3), light cultures were treated with mock, heavy with TRAIL and medium cultures with both z-VAD-FMK inhibitor and TRAIL.

Data analysis – Raw files of each double and triple labeling Jurkat T cell experiment were analyzed together using the in-house built software MaxQuant (version 1.1.1.35; (31, 32)). Each raw file from a particular slice was defined as a separate experiment in the experimental design file to obtain peptide ratios for each peptide in each slice. The derived peak list was searched with Andromeda (33) against the human International Protein Index protein sequence database (ipi.HUMAN.v3.68.fasta; 87,083 entries) supplemented with 262 frequently observed contaminants

such as human keratins, bovine serum proteins and proteases and concatenated with the reversed copies of all sequences. We required strict enzyme specificity with cleavage C-terminal after K, R or D (trypsin + Asp-C), allowing up to two missed cleavage sites. Fixed modification of cysteine carbamidomethylation (Cys 57.021464 Da) and variable modifications for N-acetylation of proteins (N-term 42.010565 Da) and oxidation of methionine (Met 15.994915 Da) were specified. Double or triple labeling was defined accordingly. The minimum peptide length was set to 6 amino acids. Scoring was performed in MaxQuant as described previously (31). Parent masses and fragment ions were searched with an initial mass tolerance of 7 ppm and 0.5 Da, respectively. False discovery rates (FDRs) at the peptide and protein levels were fixed to 1%, including automatic filtering on peptide length, mass error precision estimates, and peptide scores of all forward and reversed peptide identifications. The re-quantification feature was enabled. Reported protein groups had to be identified by at least one 'razor peptide' (peptide most likely belonging to the protein group) to be accepted. Quantitation was based on unique and razor peptides only and a minimum of 2 ratio counts was required. Complete protein and peptide lists as well as the underlying RAW files are available on the TRANCHE database (https://proteomecommons.org/tranche/).

Statistical identification of cleavage events – The obtained lists of peptides and proteins were further processed with an in-house developed program implemented in MATLAB (MathWorks, Natick, Massachusetts). All ratios were converted into log2 values. Ratios of reverse labeling experiments were inverted beforehand. A protein is an apoptosis substrate if the peptides in the slices representing molecular weights smaller than the one of the full length protein are more abundant in apoptotic cells than in untreated cells. A statistical, non-parametric test was applied to determine the confidence of identification of cleaved proteins. In brief, for each identified protein group, the slice with the most peptides, which had SILAC ratios smaller than 1 between TRAIL treated cells and untreated controls, was selected as the slice s_0 containing the uncleaved protein. The t-test statistics was calculated between the ratios of peptides in slice s_0 and ratios of peptides in

slices s<s₀, for which the average ratio was larger than 1.5. The statistical test was repeated 1000 times with randomly permutated slices to evaluate the false discovery rate (FDR). Finally, this value was corrected for multiple hypotheses testing by the Benjamini-Hochberg method ((34)).

For each protein group a plot presenting the identified peptides as boxes in a two-dimensional map was generated. The x-axis of the map represents the position of the peptide with respect to the protein sequence. The y-axis represents the slice in which the peptide was detected. The boxes were color-coded by the ratio between TRAIL-treated and untreated cells. If peptides cleaved at the C-terminus of aspartic acid were identified, the corresponding cleavage site was marked by a solid line if the SILAC ratio of the cleaved peptide was greater than 1.5 and by a dotted line otherwise. Similarly, potential cleavage positions that match the sequence motif x[ET]xD and known cleavage positions were marked in the plots as additional information for the reader. Protein sequences and domains were obtained from Uniprot web services.

The second y-axis in the plots represents the estimated molecular weight. The mapping of slices to molecular weights was obtained by linear regression. For each protein group, the intensity weighted average of slices, $s_i = \frac{\sum_{k=1}^{K} k I_{ik}}{\sum_{k=1}^{K} I_{ik}}$, was determined, where $k = 1, \dots, K$ enumerate the slices, and I_{ik} is the XIC (extracted ion current) of protein i in slice k. The linear function $\ln(m) = a + bs$ with the parameters a and b was then fitted to the data $(s_i, \ln(m_i))$, where m_i is the molecular weight of protein i.

The MaxQuant results, the results of the filtered substrates including statistical information, and the list of known cleavage sites were uploaded to the publicly available database MaxQB (35). Additionally, the 3D cleavage plots of all identified cleavage substrates are visualized in MaxQB and can be obtained at http://www.biochem.mpg.de/maxqb. Instructions on how to use MaxQB for visualizing and extracting cleavage information are provided as supplemental information. In

addition, all .m files for the MATLAB script are provided on the MaxQB web page to allow researchers to analyze their data in regards to proteolytic cleavage events at the following link:

 http://www.biochem.mpg.de/maxqb/mxdb/project/show/TRAIL%20induced%20Apoptosis

Data representation – Data were depicted using GraphPad Prism (version 5.04; www.graphpad.com) and the free software environment R (http://www.r-project.org/).

Enrichment analyses – Cleavage events are more likely to be detected in highly abundant proteins. In order to remove this effect of the abundance as confounding variable we first generated a background distribution of the whole proteome data with identical intensity distribution as the cleaved proteins. To this end, the intensity range was binned and each protein group from the whole dataset was randomly drawn with the probability equal to the proportion of cleaved proteins in the corresponding bin compared to the total number of cleaved proteins. The Fisher exact test was then used to identify gene ontology categories that are significantly enriched or depleted (Benjamini-Hochberg FDR 0.02) in the cleaved population compared to the background population.

RESULTS AND DISCUSSION

Development of a SILAC-based quantitative approach to detect proteolytic substrates – When proteolytically cleaved proteins change their molecular weight in comparison to their uncleaved counterpart this can be detected in an SDS gel by a shift of the cleaved protein to a lower molecular weight region. We took advantage of this fact to extract information about cleaved proteins from complete proteome datasets. For direct comparison we combine treated and untreated states in one quantitative SILAC experiment. As both samples are merged at the cell level, this eliminates

errors due to lack of reproducibility and differences in sample processing. In our SILAC experiments, untreated cells are labeled with light RPMI medium (Arg0; Lys0), whereas cells used for treatment are labeled in heavy medium (Arg10; Lys8) before induction of apoptosis (Fig. 1A). We also included label swap experiments as an additional check of the method (supplemental Fig. S1A).

We applied the SILAC-based approach to investigate cleavages induced by treatment with TRAIL. Populations of SILAC labeled Jurkat T-cells, which were treated with TRAIL or mock control for 5 h, were merged and their lysates separated by 1D SDS gels. To achieve high molecular weight resolution, we cut the gel into 28-36 horizontal slices, followed by standard in-gel digestion and LC-MS/MS analysis on ion trap-Orbitrap instruments. Data were analyzed with the MaxQuant software as in our standard workflows; however, the output data was further processed to identify cleaved caspase substrates (EXPERIMENTAL PROCEDURES) by sophisticated filtering. For this processing step, we developed a statistical algorithm that specifically extracts proteins with a distinct peptide distribution pattern as a function of gel slice position. In gel digestion of a particular gel slice creates peptides from the embedded proteins, which represent a certain mass range (the apparent molecular weight of that gel position). The software provides the localization of all detected peptides from each protein in two dimensions: within the gel according to the apparent molecular weight region and along the protein sequence according to the location of the identified peptide (Fig. 1B). Peptides from the uncleaved protein are expected to be located in higher molecular weight regions and to cover the protein sequence without bias to sequence location. In contrast, cleaved apoptotic fragments of the same protein should have migrated to a lower molecular weight region and the identifying peptides should span only a certain part of the protein sequence in accordance with the cleavage position. The quantitative information by SILAC labeling then provides a third dimension. These SILAC peptide ratios encode information about the extent to which the protein or fragment comes from the treated or untreated state. Peptides from the uncleaved protein are mainly derived from the untreated cells (negative treated/untreated ratio after log transformation) when they are from a high mass region, whereas peptides derived from corresponding cleaved

fragments are mainly derived from the apoptotic state (positive treated/untreated ratio) if they are from the relatively lower gel position. The slice with the highest sequence coverage combined with a median log ratio around zero corresponds to the uncleaved protein. Slices at lower molecular weights with median positive ratios correspond to the cleaved fragments. The third dimension is represented in our plots as a heat map value (red for positive SILAC log ratios to blue for negative log SILAC ratios; Fig. 1B). Importantly, the above mentioned criteria should be fulfilled concurrently, i.e. the uncleaved protein and its SILAC and sequence values should be consistent with the cleaved product or products and their location, SILAC ratios and distributions of peptides in the protein sequence.

To statistically formalize these criteria, the software defines these cleaved and uncleaved regions of a protein and calculates p-values and a false discovery rate (FDR) for the significance of being a cleavage product (EXPERIMENTAL PROCEDURES). Only proteins with an FDR value below 5% were considered as significant candidates and only for these graphs were created. We marked theoretically detectable peptides to give an impression of the highest possible sequence coverage of the protein. In addition to the experimental data we also included literature knowledge into the graphs. For instance, known cleavage sites of proteins in the substrate database that we created (see below) as well as possible cleavage sites following the x[ET]xD caspase cleavage motif are indicated as red lines. Known domains of the protein derived from Uniprot are depicted as small blue boxes. In addition, we marked the specific protein position with a red line where peptides show a cleavage according to enzymatic digestion with Asp C. This is because caspase cleavage followed by tryptic digestion creates semitryptic peptides in shotgun proteomics experiments, which, if detected, can mark the exact site of caspase cleavage. For these peptides, we distinguish cleavages with high confidence (ratio [log2] > 1.5) and intermediate confidence (ratio [log2] < 1.5); these values were chosen based on the distribution of known caspase cleavage products. For high confident cleavages, the explicit cleavage site is denoted in the output table as well as the sequence window of +/-3 amino acids surrounding the cleavage site. In addition, we calculated molecular weight distributions

that relate the position on the 1D SDS gel to the molecular weight of the proteins detected in the entire proteome dataset of this experiment. This relationship is depicted in supplemental Fig. S1B and is used to provide a molecular weight scale on the right y-axis of each plot. We further manually inspected the automatically filtered candidates for consistency across all six experiments (see below) and verified them for inclusion in the final substrate list.

Identification and characterization of TRAIL induced proteolytic substrates – Since no global whole proteome studies have been reported on the specific proteolytic substrate spectrum of TRAIL-induced apoptosis so far, we treated Jurkat T cells with different concentrations of TRAIL and monitored their state over time by light microscopy (EXPERIMENTAL PROCEDURES). Based on the results, we chose a TRAIL concentration of 100 ng/ml and stimulation for 5 h. Microscopic, flow cytometric and immuno blotting results (Fig. 2A, B, C) showed that this time point optimally covers both upstream and downstream apoptotic events (initiating caspases and substrates downstream of caspases, respectively), which should allow us to cover a broad spectrum of apoptotic cleavage events. In addition, our results with H_2O_2 treatment – an unspecific cytotoxic agent – show that we clearly separated apoptosis from uncontrolled necrotic cell death (Fig. 2A + C).

Next, we applied the above mentioned conditions in a SILAC experiment treating Jurkat T cells for 5 h with TRAIL (100 ng/ml) and mock, respectively. We performed three biological double labeling experiments and three biological triple labeling experiments comparing TRAIL-treated and mock-treated states. From the triple labeling experiments we only used the light and heavy channels, the medium condition was additionally inhibitor treated as mentioned below.

For substrate identification, we combined all six experiments. When we identified a candidate in one of the experiments as significant (FDR < 5%), the software created a 3D cleavage plot also for the other experiments (even if they were not significant) to allow better comparison between the different experiments. We derived significances for each plot from the FDR columns of the output of our script and counted how often each protein had been identified as statistically significant in the

six experiments. In final manual validation, only candidates with an appropriate peptide distribution pattern that were found to be significant at least three times were designated as TRAIL-induced apoptosis substrates and considered for further analyses (supplemental Table S1).

This procedure identified a plethora of positive controls known to be proteolytically cleaved during apoptosis, prominent examples of which are presented in Fig 3A. CASP8 shows a clear uncleaved region with good sequence coverage. The N-terminal cleaved fragments in the range of 43 kDa and 24 kDa corresponding to the most probable cleavage at D374 and D216 are present (4). These two cleavage sites correspond to the first and second cleavage of CASP8, for which another cleavage at D384 is known. For the well-studied downstream substrate PARP1 we covered nearly the complete sequence in the uncleaved region (113 kDa) with highly negative ratios (indicating absence of the protein in the treated state and therefore almost complete cleavage). In addition, we also obtained high sequence coverage of the fragments and the mass of the fragments fit with their sequence region of the protein (89 kDa and 24 kDa). The region of the cleavage site is clearly visible and is in agreement with a known cleavage site on PARP1. This site – DEVD[214] – is located directly within the DNA-binding region. By cleavage of the N-terminal part of the protein, PARP1 therefore is no longer recruited to the DNA and consequently loses its catalytic activity in DNA damage repair. We also detected Caspase-3, Caspase-6 and Caspase-2 and many other well documented caspase substrates of both early and late cleavage events such as BID, DFFA, PAK2, VIM, LMNB1/B2 and ROCK1 (supplemental Table S1). BID mediates the crosstalk between the extrinsic and intrinsic form of cell death via accumulation of its 15 kDa fragment tBid at mitochondria, initiating mitochondrial outer membrane permeabilization (MOMP) (36). DFFA (ICAD) is an inhibitor of the DNase CAD, which is inactivated by cleavage. Cleavage of this inhibitor then allows free CAD to translocate to the DNA and degrade chromosomes into nucleosomal fragments, a characteristic feature of apoptosis.

After statistical filtering and manual inspection we obtained 693 cleaved substrates in response to TRAIL-induced apoptosis (Fig. 3B; supplemental Fig. S2A). To generate a list of known substrates for

comparison, we accessed the caspase substrate database CASBAH (http://bioinf.gen.tcd.ie/casbah/) (37) and data from two large-scale investigations (11, 23). This analysis showed that 304 of our substrates were already known as verified or potential caspase substrates, whereas 389 of our candidate proteins were entirely novel. Together, these results demonstrate that our screen covered a substantial part of the known substrates of all caspase substrates while identifying a very large set of novel TRAIL-induced cleavage products. This attests to the sensitivity of our approach but also suggests that only a fraction of all cleaved substrates may be described to date (38, 39).

We next plotted the intensity distribution spanned by the complete proteome experiment in comparison to the intensity fraction covered by the statistically significant cleaved substrates (Fig. 3C). Cleavage substrates are distributed towards the higher abundant area of the complete proteome. This is not surprising as verification as a substrate requires substantially more information than mere identification in the proteome. Interestingly, although our data are biased toward the more abundant part of the proteome, the substrates are relatively flatly distributed across the accessible abundance range. This shows that the tendency to be a caspase substrate is not strongly correlated with protein abundance, at least in this abundance range covered by this study.

Detected substrates span a broad molecular weight range and encompass proteins ranging from about 20 kDa to up to 200 kDa. Nevertheless, very small proteins are underrepresented within the substrate population, which may be due to limitations of the gel-based MS-approach (Fig. S2B, C).

Proteolytic substrates are not necessarily the result of caspase cleavage but might also be cleaved by other proteases, which could either be activated by or independent of upstream caspase cleavage. To verify that the cleaved substrates are truly caspase-dependent, we incubated Jurkat T cells with the pan-caspase inhibitor z-VAD-FMK before treatment with TRAIL. We determined best inhibitor conditions by different concentrations of z-VAD-FMK, using the known caspase substrate PARP1 (Fig. 4A). Based on these results, we performed a triple SILAC approach treating SILAC labeled cells either with mock (light SILAC condition), TRAIL (5 h; heavy SILAC condition) or TRAIL (5 h) and z-VAD-FMK

at 50 µM (medium SILAC condition). We performed three biological replicate experiments and analyzed the samples as described in EXPERIMENTAL PROCEDURES. TRAIL-treated samples (heavy/light ratio) showed clear cleavage patterns (Fig. 4B, left panel), whereas inhibitor treated samples showed no peptide ratio difference in comparison to the mock treated population (Fig. 4B, right panel) (medium/light ratio). Note that peptides of caspase dependent substrate such as PARP1 could also be found in molecular weight regions of the cleavage fragments; however, these have one-to-one SILAC ratios [0 in log2] between inhibitor treated and control conditions, showing that they originate from background (TRAIL-independent) cleavage. Because all significant cleavage events were abolished by the inhibitor, we conclude that they are all caspase-dependent.

Identification of proteolytic cleavage sites – In addition to identifying cleavage substrates of caspases our approach might allow identification of specific caspase cleavage sites in a subset of cases. To test this, we searched for peptides with either an N- or C-terminal caspase cleavage site and a tryptic site on the other terminus. We also required that the peptide was located in the region indicated by our statistical algorithm as the probable cleave region (EXPERIMENTAL PROCEDURES). Using this approach we were indeed able to identify 93 explicit cleavage events in 86 proteins, including known ones that served as positive controls and novel ones. These sites, including the sequence windows of +/- 3 amino acids are listed in supplemental Table S3. Several novel potential substrates such as the cell growth-regulating nucleolar protein LYAR were also found in this way (Fig. 5A). For this protein, we mapped the cleavage site to the position D281 overlapping with the hypothetical cleavage motif x[ET]xD. A sequence logo analysis (http://weblogo.berkeley.edu/logo.cgi; Fig. 5B) supported earlier findings on intrinsically induced apoptosis that characterized cleavage motifs as diverse (6). Cleavage is common at an aspartate C-terminal to a small amino acid such as serine, glycine or alanine, as reported previously (6, 11). We conclude from this that TRAIL-induced cleavage patterns are similar to those observed in intrinsic caspase mediated events.

In cases where the caspase cleaved peptide was not confidently identified, the cleavage event was often still mappable to a particular region of the protein. In 159 proteins the cleavage was mapped to a region of about 100 amino acids (Fig. 5C; supplemental Table S3). In 50 of these cases the cleavage area overlapped with only one x[ET]xD motif (*implicit sites*). In 85 proteins there was more than one motif in the cleavage area (*area*). In the remaining cases, a different cleavage motif was presumably used by the responsible protease.

In several cases we determined multiple cleavage events within the same protein (Fig. 5D). We found, for instance, two cleavage events in the DNA replication licensing factor MCM6, one of which was already known and could directly be mapped as an explicit site - D274. For the second cleavage two possible motifs match within the C-terminal region close to the position 770. From the positions of the fragments in the gel, we determined that either or both cleavages can occur, indicating that there appears to be no preference for one of them as an initiator site (Fig. 5D). Since either cleavage presumably inactivates the protein such a cleavage pattern is functionally sensible for substrate inactivation.

As the above results show, our whole protein approach is not necessarily limited to the identification of cleaved proteins, but may in many cases yield detailed information about the cleavage site or region. Even the identification of cleavage regions can already be of great importance for tracking the effect of the cleavage by providing information of cleaved domains or motifs, for instance.

Global analysis of detected proteolytic events – Bioinformatic analysis of our substrate proteome revealed that it covered a broad range of different pathways with known and novel substrates showing similar trends (supplemental Table S4). In particular, many substrates belong to RNA-dependent pathways, endocytosis and spliceosomal as well as cell cycle-related processes. We next extracted protein populations enriched and depleted in our substrate population compared to the

complete detected proteome using Fisher Exact testing (Benjamini-Hochberg FDR: 0.02). Since the substrate population is biased towards higher protein expression as determined earlier (Fig. 3C), we first created a background protein population spanning the same intensity region as the cleaved substrates (Fig. 6A) to serve as reference population. Interestingly, the substrates showed an asymmetric distribution between depleted and enriched categories (supplemental Table S5). TRAIL-induced apoptosis appears to target a broad range of functions without strong preferences whereas a small number of pathways and compartments were clearly selected against. The latter encompassed the ER, mitochondria and inner membrane proteins in general, as discussed below.

Especially RNA-regulated processes including RNA helicase activity and RNA splicing but also pathways such as endocytosis (CLTA, AP2A1, AP2A2, DNM2) (40) were significantly enriched in cleaved substrates. Nuclear proteins were slightly but statistically significantly enriched, as were proteins in complexes (derived from the CORUM database (41)) (Fig. 6B). In contrast, proteins intrinsic to membrane were significantly under-represented and proteins in mitochondria and endoplasmic reticulum (ER) were more than two times under-represented within the cleaved substrate fraction (Fig. 6B). Proteins in ER and mitochondria may either not be accessible for caspases or they may not contain as many cleavage sites (cleavage motifs) because they have to fulfill important tasks in controlled cell death. In any case this is an interesting observation given that mitochondria dynamics and fission are known apoptotic effects (42). Regardless of the mechanism, our data indicate that the mitochondria are protected from caspase cleavage, which appears reasonable given the central function of mitochondria in intrinsically induced – but also extrinsically caused – apoptosis. In contrast, nuclear pore proteins are preferentially cleaved by caspases and nuclear proteins are enriched as substrates, which agrees with the fact that nuclear breakdown is a hallmark of apoptosis.

The previous analysis was focused on the identity of the substrate proteins only. In a next step, by evaluating the SILAC ratios of the apoptotic substrates at the gel positions of the full length proteins

(uncleaved ratios), we noticed that they displayed in all experiments a bimodal distribution (Fig. 6C). Classical upstream (e.g. caspases) as well as downstream substrates (e.g. PARP1, LMNB2, ROCK1) were observed with strong negative ratios, indicating that a large percentage of these substrates had been cleaved at this time point (5 h). In contrast proteins in the right hand peak had ratios around zero, indicating that most of these proteins remained uncleaved, even though fragments with clear positive ratios marked them as apoptotic substrates. Splitting the population into novel and known substrates showed a clear tendency of novel substrates towards the right hand distribution (low percentage of substrate cleavage; supplemental Fig. S3A). This may reflect the sensitivity of our approach since proteins cleaved with lower stoichiometry are more difficult to detect. Bioinformatic analysis of the enrichment of protein classes within the two different peaks with a Fisher Exact test (Benjamini-Hochberg FDR: 0.02), revealed that proteins in complexes are highly enriched in the right hand peak (Fig. 6D; supplemental Fig. S3B). Thus complex members often appear to be cleaved with lower efficiency than proteins not involved in complexes.

Caspase cleavage of the condensin complex – To perform a physical and functional interaction analysis of our substrate population, we used the STRING database (STRING 9.0; http://string-db.org/; highest confidence (score 0.90)) and visualized and analyzed results in Cytoscape (version 2.8.2.; www.cytoscape.org). We grouped the substrates into known and novel proteins (green and red, respectively, in supplemental Fig. S4). The center of the interaction network was formed by Caspase-3 surrounded by many known substrates such as CASP6, BID, LMNB1, PAK2 and ROCK1. Novel substrates created additional links between known cleavage substrates but also spanned further networks at the periphery of the graph. Protein clusters highlighted interactions of ribosomal and nuclear pore proteins as well as of proteins involved in protein biosynthesis, DNA replication, transcription, vesicle transport and endocytosis, defined by both known and novel substrates. We also noticed entire macromolecular complexes, among which proteasomal proteins and the condensin I complex were prominent. The latter is of special interest because it plays a crucial role in the formation of structurally stable mitotic chromosomes and their segregation but also in gene

regulation and DNA repair (43, 44). The pentameric condensin I complex is highly conserved and ubiquitously found among eukaryotes. It consists of two structural maintenance of chromosome ATPase subunits and three auxiliary subunits. Only one component of the complex, the kleisin subunit NCAPH, has been reported to be cleaved upon apoptosis induction to date. This cleavage was thought to be responsible for the loss of the condensin I complex in apoptosis, contributing to loss of chromosome structure and chromosomes susceptibility to DNA fragmentation induced by a caspase-activated DNase (45).

Interestingly, we identified not only NCAPH but also all other components of the complex as cleaved substrates after TRAIL induction (Fig. 7A; supplemental Fig. S5). To better understand the process of condensin I complex cleavage we extracted the 3D cleavage plots of all of its members. The known substrate NCAPH was strongly cleaved in the middle of the protein as already described in the literature. However, our detected cleavage site (D380), which matches with a potential cleavage motif, does not overlap with the known cleavage site at D366. The novel substrates of this complex SMC2, SMC4, NCAPD2 and NCAPG showed cleavage at the N- or C-termini, resulting in a slight shift of the uncleaved protein. Theses substrates were cleaved with lower efficiency compared to the known substrate NCAPH. Both the structural maintenance of chromosome ATPase subunits (SMC2 and SMC4), have nucleotide-binding domains at their N- and C-terminus, termed Walker A and B motifs. Within each protein, these domains interact with each other and form so-called head domains. We detected cleavages in both proteins at their N- and C-termini. Interestingly, for SMC2 we could even map an explicit cleavage site at the C-terminal part of the protein located exactly within the Walker B motif at D1116. This is remarkable since ATP-binding pockets should be difficult to access. For NCAPD2 and NCAPG we also detected cleavage at their N- and C-terminus, respectively. These cleavage sites appear to overlap with important intra- and inter-molecular interaction sites to the other proteins of the complex. Interestingly, NCAPD3 from the related condensin II complex was cleaved as well. However, in this case we located this cleavage site to

position D529 in the middle part of the sequence rather than the termini. This may indicate different modes of association of NCAPD2 and NCAPD3 with their corresponding kleisin subunits.

A recent paper investigated the function of yeast condensin by introducing several TEV cleavage sites into Brn1 (NCAPH-homologue) (46). Interestingly, cleavage of NCAPH leads to opening of the condensin ring, without, however, releasing subunits of the complex. Intriguingly, after cleavage of NCAPH the complex is still associated with the chromosome if the two NCAPH fragments are held together by interactions with NCAPD2, at least in vitro. After additional cleavage of NCAPD2, condensin was released from the chromosomes. Also, in case of the SMC2 coiled coil artificial cleavage of both strands was necessary for inactivation. In light of these findings and our data we suggest that simple cleavage of NCAPH by caspases might not be sufficient to release condensin from the chromosome. Additional proteolytic cleavage events in other subunits – preferentially at the interaction sites – might be necessary to fully inactivate the complex and disengage its components. Sites in condensin I complex subunits other than NCAPH were observed with lower SILAC-fold changes, suggesting that the specific location of cleavage is less important than the fact that a second cleavage occurs. While these mechanisms are speculative, the cleavage sites identified here are mechanistically plausible and may serve as starting points for further functional dissection of the cleavage of the complex.

Caspase cleavages of the proteasome - Recently, several groups have suggested that protein degradation by caspases and by the proteasome are interlinked by reciprocal cleavage of some of its components and several proteasomal subunits were identified as apoptosis substrates (6). Here we identified cleavage events in the 20S core particle and in the 19S regulatory particles, covering both base and lid proteins (47, 48) (supplemental Fig. S6). Similarly to the condensin complex, stoichiometries of substrate cleavages were relatively low. From the regulatory particles we determined as previously known substrates the subunits Rpn2, Rpn3, Rpn10, Rpt1, Rpt5, Rpt6 and Rpt4 as well as the novel substrates Rpn5 and Rpt2 (6). From the core particle we identified PSMA5

and in addition also observed cleavage of the 11S regulatory cap, PA28γ. These findings support the hypothesis of a negative feedback loop in which caspases and the proteasome are interlinked in the process of apoptosis (49). In healthy cells, caspases as the effectors of apoptosis are tightly regulated by the proteasome and their protein levels are reduced by proteasomal control. Upon caspase activation by apoptosis, proteins mainly within the regulatory part of the proteasome are cleaved in a caspase-dependent manner and proteasomal degradation is inhibited. Our findings support this general picture and supply a number of specific substrates and sites in the proteasome.

An earlier study identified caspases cleavage of three proteins of the 19S RP complex of the 26S proteasome. PSMC3 (Rpt5) and PSMD4 (Rpn10) most probably recognize the polyubiquitinated proteins, while PSMD1 (Rpn2) most likely holds together the lid and the base of the 19S RP (49). Their cleavage may inactivate this process and may partially detach the 19S RP from the 20S core particle. Our data substantiates the cleavage sites of PSMC3 and PSMD1. For PSMD4, two possible cleavage sites were posited in the earlier study. Our results clearly identify cleavage at one of them, D258.

The SILAC ratios observed in our data are relatively low, suggesting sub-stoichiometric cleavage at different sites for each member protein of a complex. To compensate for this lower cleavage efficiency, caspases appear to cleave several members of the same protein complex such as the proteasome, perhaps with a more robust inactivation effect as the cleavage of one specific protein.

Caspase cleavages of the autophagy apparatus - The link between apoptosis and autophagy is also of special interest (see for instance (50)). Autophagy is cytoprotective, whereas apoptosis inhibits autophagy when activated, counteracting its effects. Important points of crosstalk include the interaction of the autophagy protein Beclin-1 and the anti-apoptotic factor Bcl-2, as well as direct interactions of caspases and autophagic components. A recent study described cleavage of the early autophagy marker ATG3 by CASP8 (51) as an important link between both pathways. The authors focused on the *in silico* derived site D169 and in an experimental follow up defined it as cleavage site

of CASP8 leading to the inactivation of the autophagic process after TNF or TRAIL induction. Here, we also detected ATG3 cleavage after TRAIL-induction in our Jurkat T cell system and we identified a cleavage region including two possible sites, D169 being one of them (Fig. 7B, supplemental Table S1). Furthermore, we mapped an explicit cleavage site of ATG3 at D104. That site had also been found in another study (23), supporting our results. Based on the peptide information from the cleaved fragments, the cleavage at D104 appears to be more prominent than that at D169. Since peptides from the counterpart of the D169 cleavage are absent in the gel, we speculate that this cleavage is secondary. In any case, the detection of two cleavage events by our method illustrates its depth and its unbiased nature.

CONCLUSIONS AND OUTLOOK

Here we have described a quantitative SILAC-based approach for the identification of proteolytically cleaved substrates and used it to investigate the events of apoptosis induced by extrinsic stimulus TRAIL. Our approach determined nearly 700 cleavage substrates, a dataset which can serve as a resource for studying TRAIL-induced cleavage events, for the community of cell death researchers. It may also be clinically relevant because TRAIL is of great importance for cancer therapy research aiming to preferentially inducing apoptosis in tumor cells but not in normal cells (52, 53). In addition, our approach can be applied to any system in which proteolytic cleavage occurs and could therefore also be of interest in fields such as embryogenesis or neurodegenerative diseases.

After this study had been finished, Cravatt and co-workers incorporated SILAC into their PROTOMAP approach (23) and elegantly used it to discover the requirement for priming phosphorylation events for apoptotic cleavage events (25). These researchers investigated staurosporine – induced apoptosis, rather than extrinsic, ligand-induced apoptosis. They detect about 700 cleaved substrates,

which reduces to about 500 substrates, when requiring an indication of cleavage products in lower molecular weight gel regions as we have done here. About 300 substrates are common to both approaches, indicating that many of the same substrates are cleaved following intrinsic and extrinsic stimulation of apoptosis. Nevertheless, it is clear that neither approach reached completion nor that there is a large number of apoptotic substrates still to be discovered. Already from the data acquired here, several features of the apoptotic substrate proteome have become apparent. For instance, we found the fact that mitochondria are clearly underrepresented as cleavage substrates and that stable complexes appear to be disabled by several apoptotic cleavage events in different complex members, each with less than full stoichiometry. Our data also support the notion that apoptotic cleavage events are not randomly distributed in the cellular proteome but instead target specific proteins and pathways (39). In this regard it is interesting to note that many of the novel substrates discovered here link to already known targets or networks. In conclusion, we note that MS-based proteomics is still increasing in speed and depth of analysis and that it should soon be possible to investigate and compare different apoptotic stimuli and conditions as well as inhibitors, providing valuable input to studying the mechanisms of apoptosis.

Data availability: Supplemental data is available with this publication at the MCP web site. The acquired raw data was uploaded to TRANCHE (www.proteomecommons.org) with the hash code: u1e+GWYB/LZaUGwIZ0/WQWHds29p3LVZEjIoOddFE+KyzJvg/ewtvTbGeGmTYSUXQbLgOlNo57MJGA wYj+A8FhhYcs4AAAAAAD8fQ==. The passphrase is "Apoptosis". Graphs and underlying data can be accessed at the publicly available database MaxQB (http://www.biochem.mpg.de/maxqb).

REFERENCES

1.	Kerr, J. F., Wyllie, A. H., and Currie, A. R. (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer* 26, 239-257.

2.	Strasser, A., O'Connor, L., and Dixit, V. M. (2000) Apoptosis signaling. *Annual review of biochemistry* 69, 217-245.

3.	Stennicke, H. R., and Salvesen, G. S. (1998) Properties of the caspases. *Biochimica et biophysica acta* 1387, 17-31.

4.	Chowdhury, I., Tharakan, B., and Bhat, G. K. (2008) Caspases - an update. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 151, 10-27.

5.	Timmer, J. C., and Salvesen, G. S. (2007) Caspase substrates. *Cell death and differentiation* 14, 66-72.

6.	Crawford, E. D., and Wells, J. A. (2011) Caspase substrates and cellular remodeling. *Annual review of biochemistry* 80, 1055-1087.

7.	Klingler, D., and Hardt, M. (2012) Profiling protease activities by dynamic proteomics workflows. *Proteomics* 12, 587-596.

8.	Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nature biotechnology* 21, 566-569.

9.	Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* 8, 1362-1370.

10.	Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005) Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nature methods* 2, 771-777.

11.	Mahrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlingame, A. L., and Wells, J. A. (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* 134, 866-876.

12.	Impens, F., Colaert, N., Helsens, K., Ghesquiere, B., Timmerman, E., De Bock, P. J., Chain, B. M., Vandekerckhove, J., and Gevaert, K. (2010) A quantitative proteomics design for systematic identification of protease cleavage events. *Molecular & cellular proteomics : MCP* 9, 2327-2333.

13.	Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J. N., and Overall, C. M. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nature protocols* 6, 1578-1611.

14.	Impens, F., Colaert, N., Helsens, K., Plasman, K., Van Damme, P., Vandekerckhove, J., and Gevaert, K. (2010) MS-driven protease substrate degradomics. *Proteomics* 10, 1284-1296.

15.	Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nature biotechnology* 28, 281-288.

16.	Demon, D., Van Damme, P., Vanden Berghe, T., Deceuninck, A., Van Durme, J., Verspurten, J., Helsens, K., Impens, F., Wejda, M., Schymkowitz, J., Rousseau, F., Madder, A., Vandekerckhove, J., Declercq, W., Gevaert, K., and Vandenabeele, P. (2009) Proteome-wide substrate analysis indicates substrate exclusion as a mechanism to generate caspase-7 versus caspase-3 specificity. *Molecular & cellular proteomics : MCP* 8, 2700-2714.

17.	Schilling, O., and Overall, C. M. (2007) Proteomic discovery of protease substrates. *Current opinion in chemical biology* 11, 36-45.

18.	Thiede, B., Dimmler, C., Siejak, F., and Rudel, T. (2001) Predominant identification of RNA-binding proteins in Fas-induced apoptosis by proteome analysis. *The Journal of biological chemistry* 276, 26044-26050.

19.   Thiede, B., Siejak, F., Dimmler, C., and Rudel, T. (2002) Prediction of translocation and cleavage of heterogeneous ribonuclear proteins and Rho guanine nucleotide dissociation inhibitor 2 during apoptosis by subcellular proteome analysis. *Proteomics* 2, 996-1006.

20.   Agard, N. J., and Wells, J. A. (2009) Methods for the proteomic identification of protease substrates. *Current opinion in chemical biology* 13, 503-509.

21.   Thiede, B., Treumann, A., Kretschmer, A., Sohlke, J., and Rudel, T. (2005) Shotgun proteome analysis of protein cleavage in apoptotic cells. *Proteomics* 5, 2123-2130.

22.   Schmidt, F., Hustoft, H. K., Strozynski, M., Dimmler, C., Rudel, T., and Thiede, B. (2007) Quantitative proteome analysis of cisplatin-induced apoptotic Jurkat T cells by stable isotope labeling with amino acids in cell culture, SDS-PAGE, and LC-MALDI-TOF/TOF MS. *Electrophoresis* 28, 4359-4368.

23.   Dix, M. M., Simon, G. M., and Cravatt, B. F. (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* 134, 679-691.

24.   Agard, N. J., Mahrus, S., Trinidad, J. C., Lynn, A., Burlingame, A. L., and Wells, J. A. (2012) Global kinetic analysis of proteolysis via quantitative targeted proteomics. *Proceedings of the National Academy of Sciences of the United States of America* 109, 1913-1918.

25.   Dix, M. M., Simon, G. M., Wang, C., Okerberg, E., Patricelli, M. P., and Cravatt, B. F. (2012) Functional Interplay between Caspase Cleavage and Phosphorylation Sculpts the Apoptotic Proteome. *Cell* 150, 426-440.

26.   Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nature methods* 6, 359-362.

27.   Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* 1, 2856-2860.

28.   Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* 75, 663-670.

29.   Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & cellular proteomics : MCP* 4, 2010-2021.

30.   Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Molecular & cellular proteomics : MCP* 8, 2759-2769.

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

32.   Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols* 4, 698-705.

33.   Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* 10, 1794-1805.

34.   Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57, 289-300.

35.   Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Molecular & cellular proteomics : MCP* 11, M111 014068.

36.   Kantari, C., and Walczak, H. (2011) Caspase-8 and bid: caught in the act between death receptors and mitochondria. *Biochimica et biophysica acta* 1813, 558-563.

37.   Luthi, A. U., and Martin, S. J. (2007) The CASBAH: a searchable database of caspase substrates. *Cell death and differentiation* 14, 641-650.

38.   Hengartner, M. O. (2000) The biochemistry of apoptosis. *Nature* 407, 770-776.

39.   Johnson, C. E., and Kornbluth, S. (2008) Caspase cleavage is not for everyone. *Cell* 134, 720-721.

40.   Austin, C. D., Lawrence, D. A., Peden, A. A., Varfolomeev, E. E., Totpal, K., De Maziere, A. M., Klumperman, J., Arnott, D., Pham, V., Scheller, R. H., and Ashkenazi, A. (2006) Death-receptor activation halts clathrin-dependent endocytosis. *Proceedings of the National Academy of Sciences of the United States of America* 103, 10283-10288.

41.   Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegele, B., Schmidt, T., Doudieu, O. N., Stumpflen, V., and Mewes, H. W. (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic acids research* 36, D646-650.

42.   Martinou, J. C., and Youle, R. J. (2011) Mitochondria in apoptosis: Bcl-2 family members and mitochondrial dynamics. *Developmental cell* 21, 92-101.

43.   Hudson, D. F., Marshall, K. M., and Earnshaw, W. C. (2009) Condensin: Architect of mitotic chromosomes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 17, 131-144.

44.   Griese, J. J., Witte, G., and Hopfner, K. P. (2010) Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic acids research* 38, 3454-3465.

45.   Lai, S. K., Wong, C. H., Lee, Y. P., and Li, H. Y. (2011) Caspase-3-mediated degradation of condensin Cap-H regulates mitotic cell death. *Cell death and differentiation* 18, 996-1004.

46.   Cuylen, S., Metz, J., and Haering, C. H. (2011) Condensin structures chromosomal DNA through topological links. *Nature structural & molecular biology* 18, 894-901.

47.   Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry* 78, 477-513.

48.   Lasker, K., Forster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences of the United States of America* 109, 1380-1387.

49.   Sun, X. M., Butterworth, M., MacFarlane, M., Dubiel, W., Ciechanover, A., and Cohen, G. M. (2004) Caspase activation inhibits proteasome function during apoptosis. *Molecular cell* 14, 81-93.

50.   Gordy, C., and He, Y. W. (2012) The crosstalk between autophagy and apoptosis: where does this lead? *Protein & cell* 3, 17-27.

51.   Oral, O., Oz-Arslan, D., Itah, Z., Naghavi, A., Deveci, R., Karacali, S., and Gozuacik, D. (2012) Cleavage of Atg3 protein by caspase-8 regulates autophagy during receptor-activated cell death. *Apoptosis : an international journal on programmed cell death* 17, 810-820.

52.   Kruyt, F. A. (2008) TRAIL and cancer therapy. *Cancer letters* 263, 14-25.

53.   Mahmood, Z., and Shukla, Y. (2010) Death receptors: targets for cancer therapy. *Experimental cell research* 316, 887-899.

FIGURE LEGENDS

FIGURE 1: Representation of the SILAC-based approach for the detection of proteolytic substrates.

A. Schematic overview of the SILAC experiment followed by stringent statistical data filtering to extract caspase-dependent cleavage substrates. In a SILAC double-labeling experiment the cells are treated with an apoptotic stimulus (light condition) or mock (heavy condition). Cells are combined and lysed and proteins are separated via SDS-PAGE. After in-gel digest, each slice is analyzed via LC-MSMS. Cleavage candidates are identified requiring specific cleavage criteria and by applying statistics to the data. Graphs are generated for each candidate and manually inspected before generation of a final substrate list. *DR*, death receptor **B.** 3D cleavage plot representation. The plot combines both experimental data and background information. Detected peptides are depicted as squares defined by slice and sequence localization including the color coded corresponding ratio. Detected cleavages C-terminal of Asp are highlighted within the plot by red lines. Supplemental information includes theoretical cleavage of the protein (peptides are plotted as squares), known and hypothetical cleavage sites along the protein sequence as well as known protein domains and motifs. Blue and red arrows indicate the location of the uncleaved and cleaved regions, respectively. Information provided within the graph is explained below the 3D cleavage plot. Experimental data are depicted in the main upper part of the plot and specific information at the bottom of the plot.

FIGURE 2: Apoptosis verification after TRAIL-induction.

A. Viability study after induction of apoptosis or necrosis compared to mock treatment. Cells were treated with 0.1% H_2O_2, 100 ng/ml TRAIL and mock, respectively, for several hours. Permeability of the cells was checked by using the cell counter, cell numbers are plotted. **B.** Western blot analysis of apoptotic markers. Cleavage of the upstream substrate CASP8 and the downstream substrate PARP1 was detected after TRAIL treatment within 24 h. Arrows point to the treatment of 5 h, the time point that was used for the SILAC experiments. **C.** Flow cytometry analysis distinguishing between

apoptosis and necrosis. Experiments show response to PI and Annexin V-FITC for mock, TRAIL and H_2O_2 treatment, respectively. 5 h TRAIL treated cells revealed clear response to Annexin V, however, no necrotic effects were detected.

FIGURE 3: Identification of TRAIL-induced cleaved substrates.

A. Positive markers of apoptotic cleavage. 3D cleavage plots are depicted for the substrates CASP8 and PARP1. For both substrates strong negative ratios were identified in the uncleaved region whereas cleaved fragments showed high positive ratios indicating a strong cleavage. **B.** Identification of known and novel substrates. Substrates detected in our dataset were compared with a list of known substrates. Slightly more than half of the substrates are novel cleavage substrates induced by TRAIL treatment. **C.** Comparison of the whole proteome and the identified substrates. The intensity distribution of all proteins is depicted, highlighting the population of cleaved substrates in green. Cleavage substrates span the higher intensity region of the complete detected proteome.

FIGURE 4: Identification of caspase-dependency of the cleaved substrates.

A. Inhibition of apoptosis by z-VAD-FMK. Different concentrations of the pan-caspase inhibitor z-VAD-FMK were tested and cleavage was detected by Western Blot against the known caspase substrate PARP1. The arrow points to the condition used for the SILAC experiments. **B.** Comparison of TRAIL-treated and inhibitor treated samples. TRAIL treatment generated a clear cleavage pattern, as for example for PARP1, including characteristic ratios in the uncleaved and cleaved region (left panel). Inhibitor treated samples showed no characteristic ratio pattern (right panel). In some cases slight background cleavage could be detected in both mock and treated samples generating peptides in the fragment slices with 1:1 ratios. Red arrows highlight the localization of the known cleavage site.

FIGURE 5: Localization of cleavage sites.

A. Detection of explicit cleavage sites. As representative example of explicit cleavage site detection, the novel substrate LYAR is depicted for which an explicit cleavage was detected at D281. The peptide cleaved C-terminally of Asp was identified with high confidence and matched with a hypothetical cleavage site, whereas the uncleaved peptide was identified in the slice containing the uncleaved protein with a negative ratio. **B.** Evaluation of caspase specificity by sequence logo representation. We extracted the sequence windows (Asp +/- 3aa) of all explicit cleavage sites and generated a sequence logo for the substrates. The output data matched with current knowledge and supported besides others the preference for small amino acids at position P1'. **C.** Representation of categories providing information on cleaved regions. Besides explicit cleavage sites cleaved regions could be narrowed down to small regions along the protein sequence. We defined four different categories describing the information obtained about the cleavage position. **D.** Multiple cleavages by caspases. In several cases, multiple cleavage events could be located within one protein. Protein MCM6, for example, was cleaved at the same time either once or twice along its protein sequence, generating five different cleavage fragments. The explicit cleavage site D274 is indicated by a red arrow, the black arrow highlights the approximate second cleavage site.
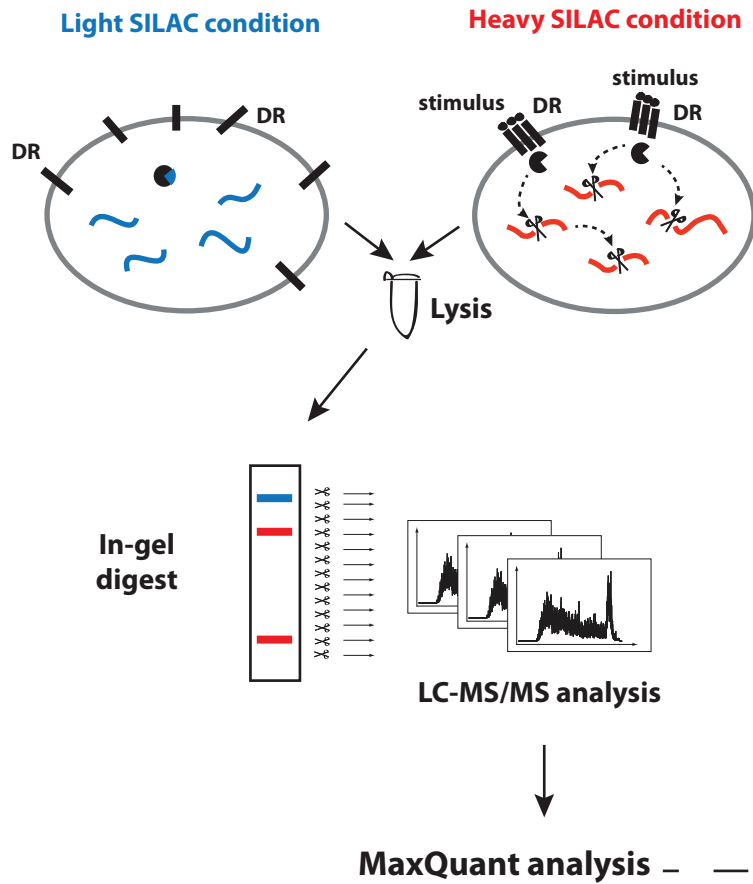
FIGURE 6: Global analysis of cleaved substrates.

A. Calculation of a background population for enrichment testing. To perform unbiased enrichment testing of cleavage substrates versus the whole proteome a background population was extracted from the whole proteome data spanning equal intensity ranges as the substrate population. All three categories are depicted. **B.** Enrichment analysis of substrates. Proteins in complexes are slightly enriched within the substrate population. Proteins assigned to mitochondrion are significantly de-enriched within the substrate population. *Bg population*, Background population **C.** Distribution of uncleaved ratios of all substrates. Median ratios of the uncleaved protein regions follow a clear bimodal distribution in all experiments. **D.** Representation of uncleaved protein ratios with regard to

proteins in complexes. Substrates annotated in CORUM showed clear tendencies towards the right hand peak, substrates not assigned to complexes had a tendency to stronger negative ratios.

FIGURE 7: Investigation of specific cleavage substrates in a biological context.

A. Cleavage of the condensin I complex members. All components of the condensin I complex were identified as cleavage substrates. For the novel substrate SMC2 an explicit cleavage site was identified within the C-terminal Walker B motif (upper panel) at D1116. The known cleavage substrate NCAPH was strongly cleaved in the middle of the protein sequence, however, the detected explicit cleavage site at D380 did not exactly match with literature (lower panel). **B.** Identification of the autophagic protein ATG3 as caspase substrate. We mapped an explicit cleavage site to D104 within the ATG3 sequence. Cleavage fragments indicate an additional cleavage within the C-terminal fragment.
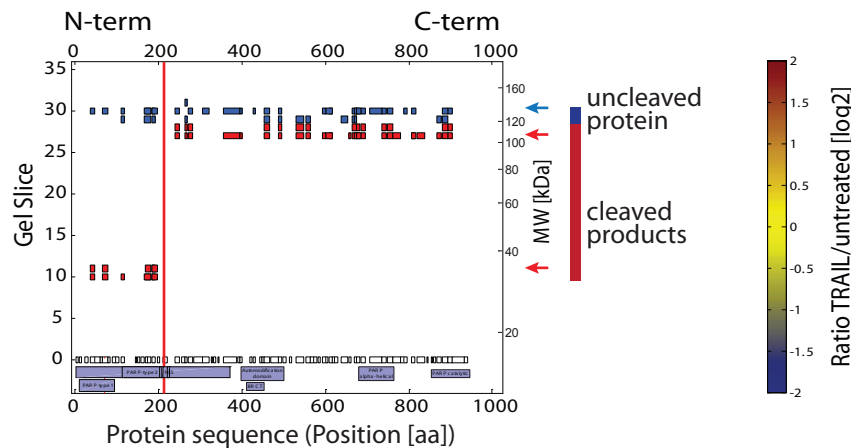
(A) **standard SILAC experiment** **Identification of cleaved substrates**

Cleavage pattern + Ratio distribution + FDR < 5%

3D cleavage plot
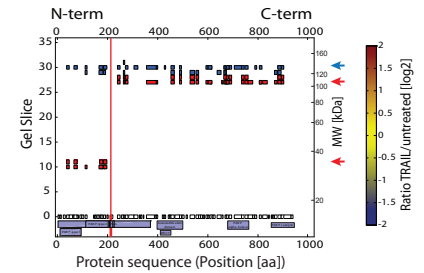
Manual inspection

Final substrate list
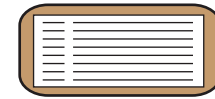
(B)

supplemental information

.... hypothetical cleavage site (X[ET]XD motif)
——— known cleavage site
▣ Domains
▢▯ Theoretical peptides (600 < MW [Da] < 4000)

experimental data

.... Peptide with an Asp-C cleavage (log2 <1.5); "intermediate confidence"
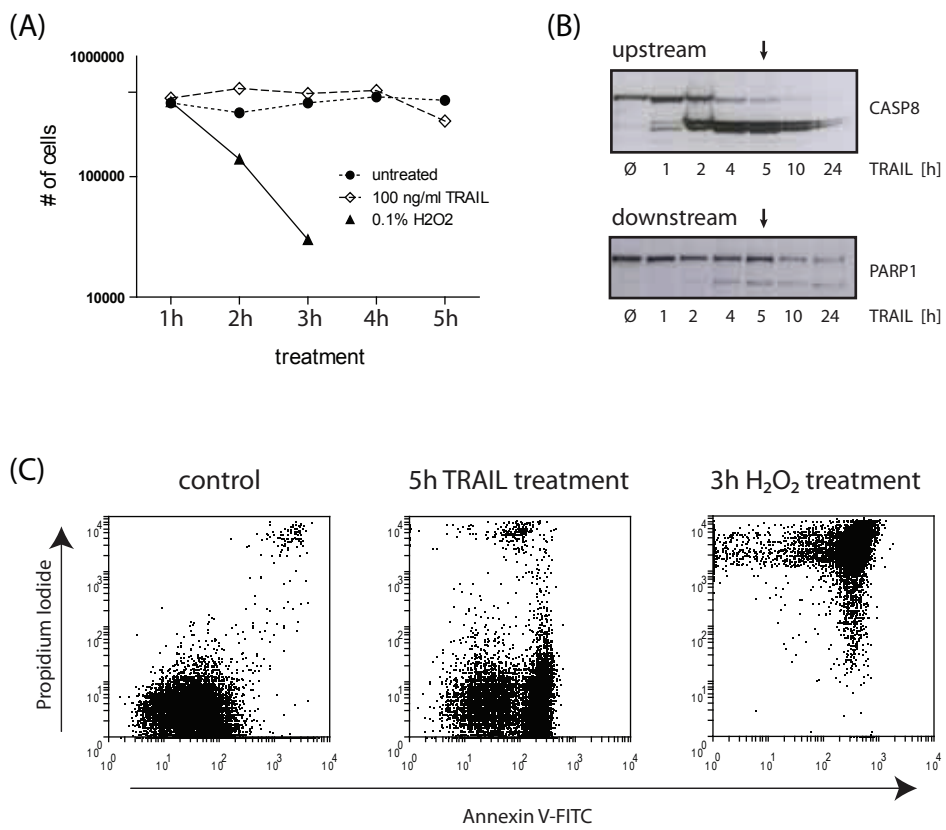——— Peptide with an Asp-C cleavage (log2 >1.5); "high confidence"
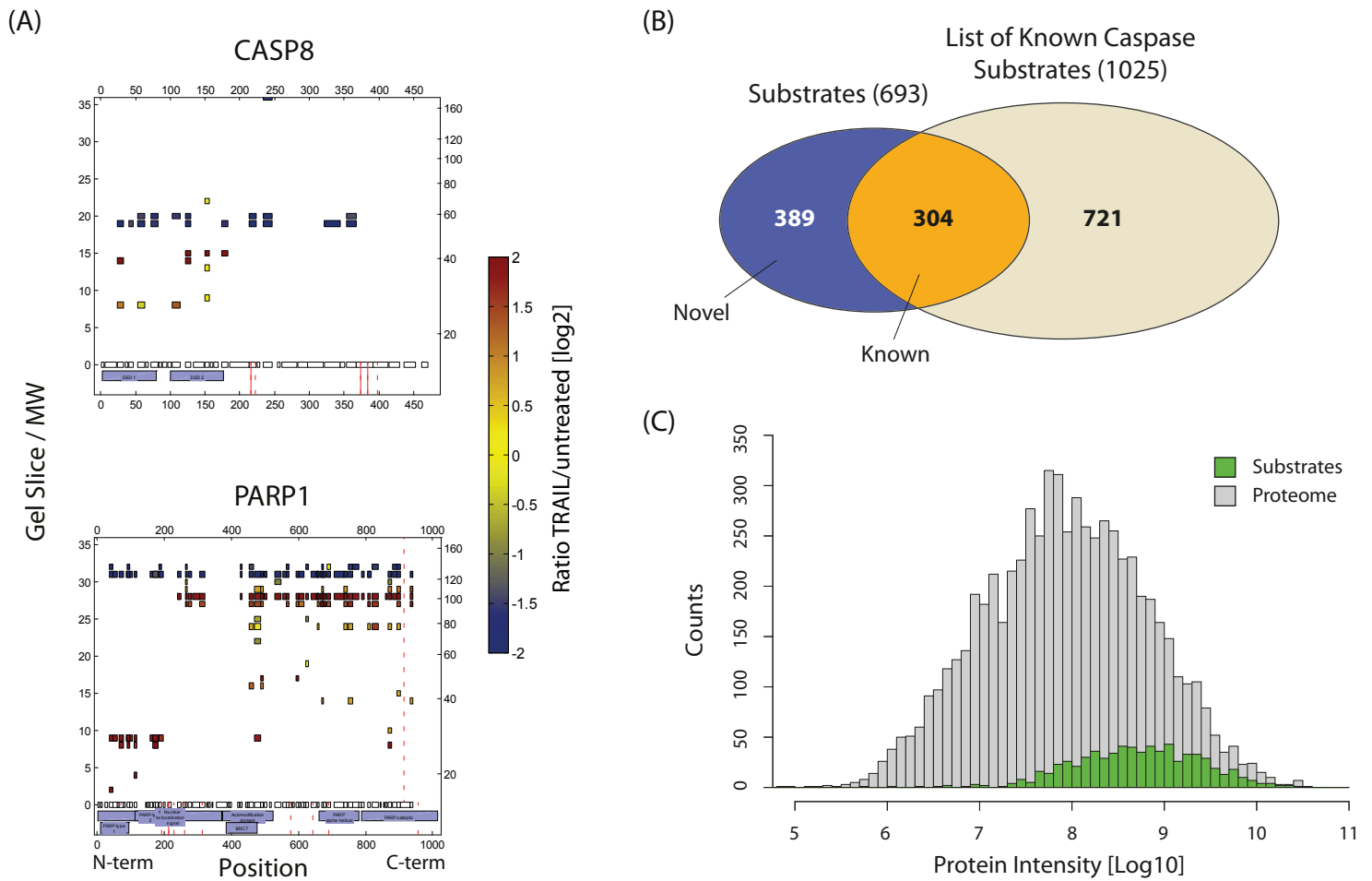
Figure 1, Stoehr et al.

(A)

(B) upstream ↓ CASP8
Ø 1 2 4 5 10 24 TRAIL [h]

downstream ↓ PARP1
Ø 1 2 4 5 10 24 TRAIL [h]

(C) control   5h TRAIL treatment   3h H₂O₂ treatment

Annexin V-FITC

Figure 2, Stoehr et al.

(A)

CASP8

PARP1

(B) Substrates (693)

List of Known Caspase
Substrates (1025)

389   304   721

Novel

Known

(C)

Ratio TRAIL/untreated [log2]

Gel Slice / MW

N-term   Position   C-term

Counts

Protein Intensity [Log10]

Substrates
Proteome

Figure 3, Stoehr et al.

[100]

(A)

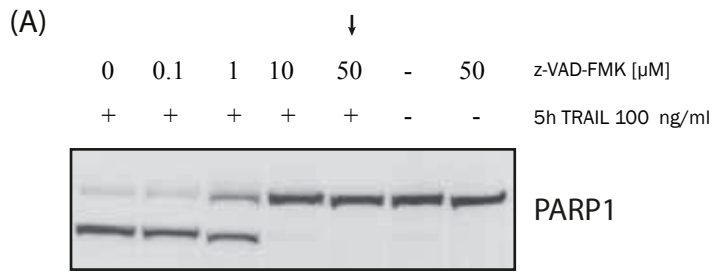| 0 | 0.1 | 1 | 10 | 50 | - | 50 | z-VAD-FMK [µM] |
| + | + | + | + | + | - | - | 5h TRAIL 100 ng/ml |

PARP1

(B)

TRAIL

TRAIL + z-VAD-FMK

PARP1

Figure 4, Stoehr et al.

(A) LYAR

(B)

(C)

(D) MCM6
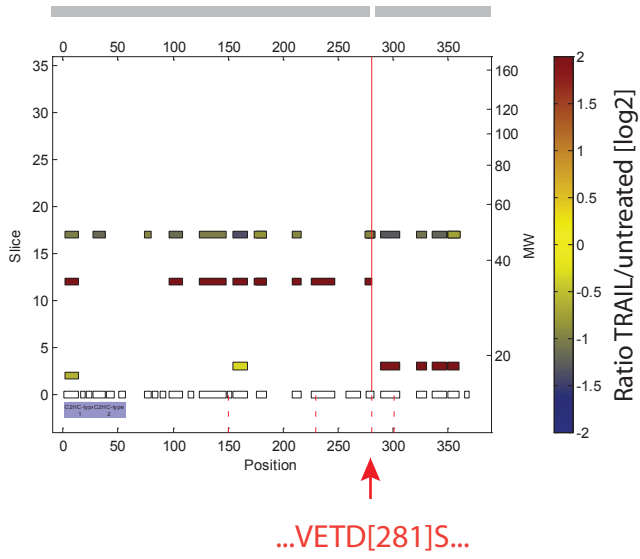
...VETD[281]S...

...SGVD[274]G...
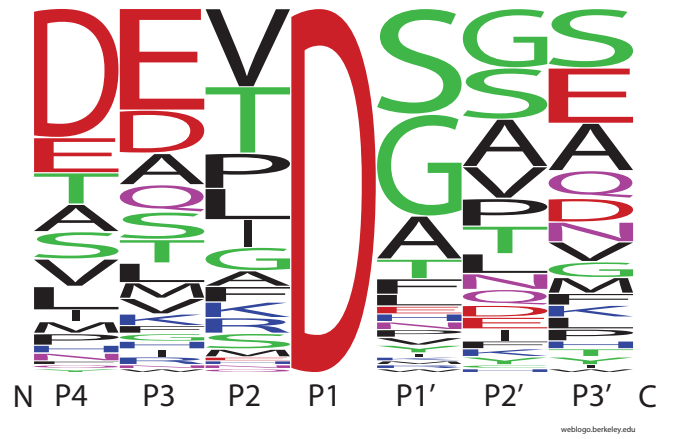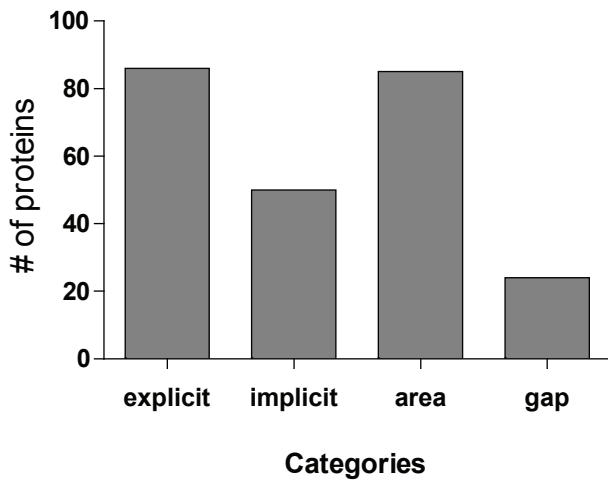
Figure 5, Stoehr et al.

Figure 6, Stoehr et al.

(A)

SMC2

NCAPH

Condensin I complex

SMC2   SMC4

NCAPH

NCAPD2   NCAPG
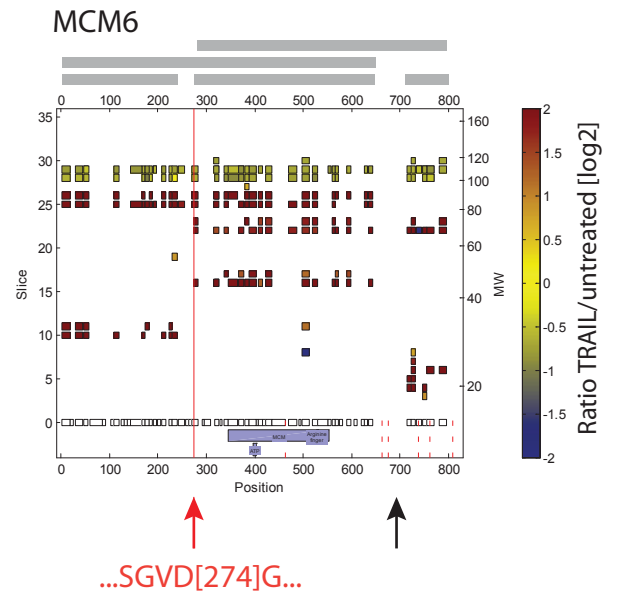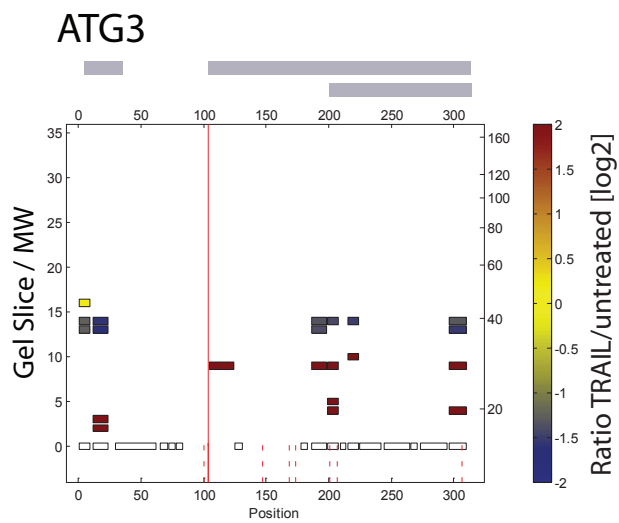
(B)

ATG3
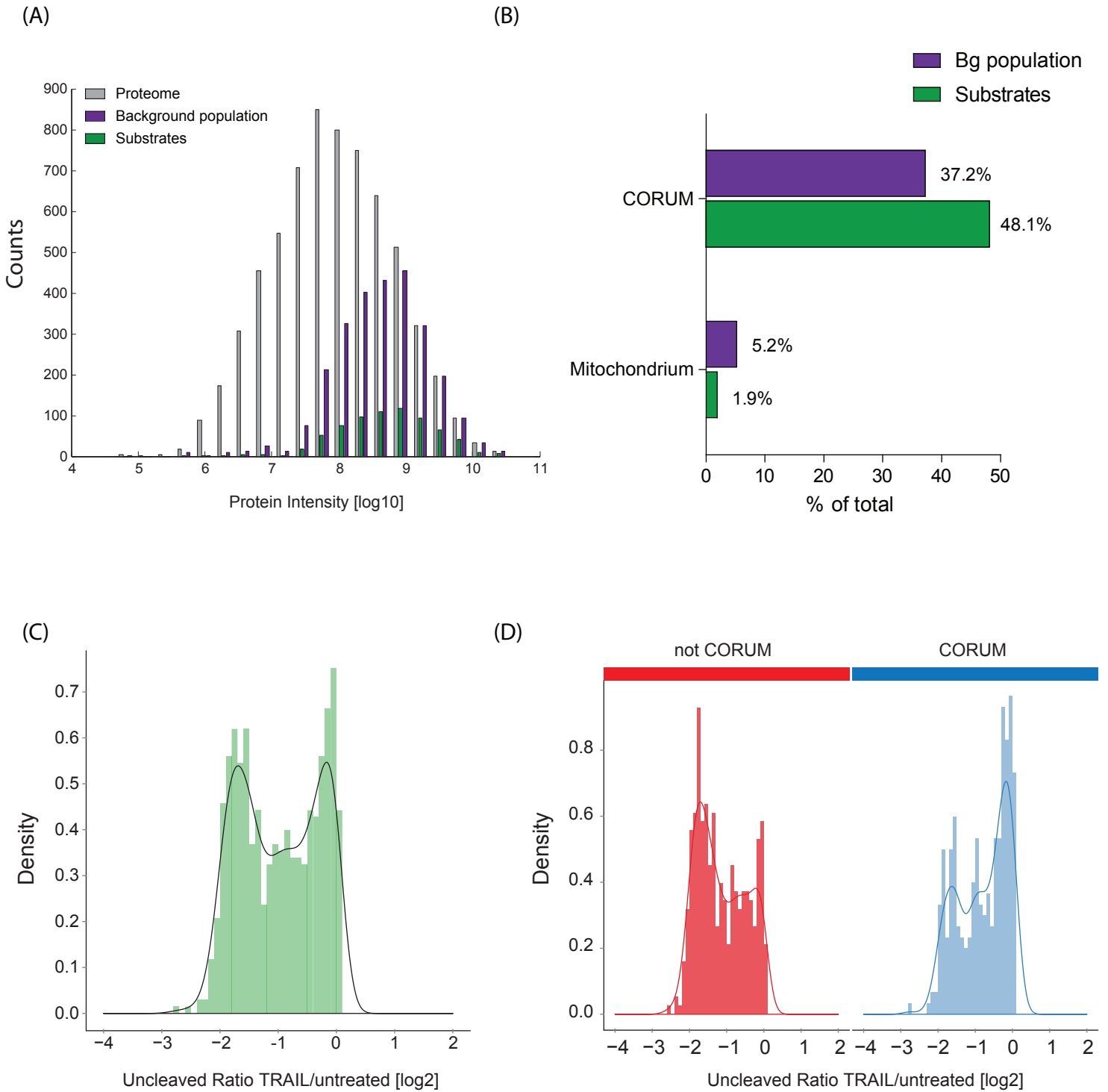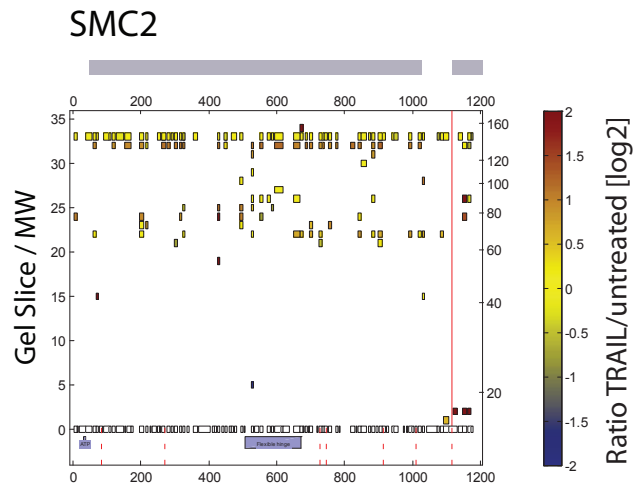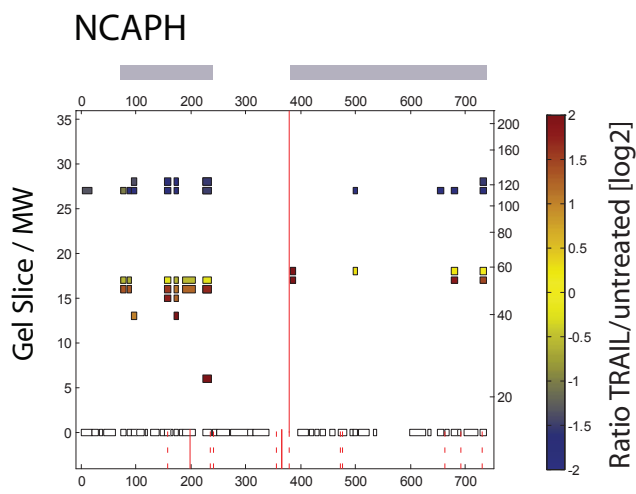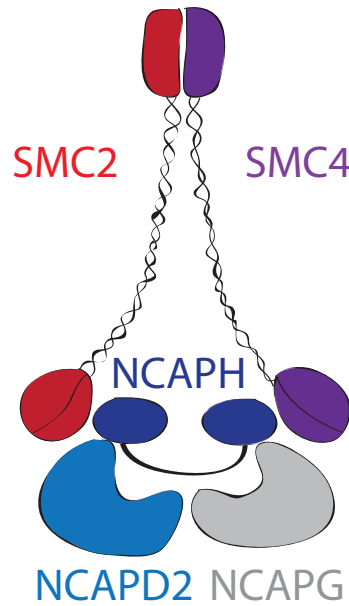
Figure 7, Stoehr et al.

2.3. Analysis of high-accuracy, quantitative proteomics data in the MaxQB database

2.3.1. Aim and Summary

In recent years, data derived from quantitative MS-based proteomics has increased exponentially. Many of the studies generate precise highly accurate datasets that are of great importance to researchers. Although many of these data are already deposited in generic databases, comparison and storage of the information across different proteome measurements still presents many challenges. A particular problem occurs with the accumulation of low quality peptide identifications, which can significantly increase false positive rates. In addition, the incorporation of quantitative data is a great challenge for many databases.

We therefore generated a database termed MaxQB, which has been specifically designed for quantitative and high resolution data generated according to stringent protocols. The database allows both storage and presentation of large proteomics datasets and enables the comparison of results of all uploaded data. The different underlying tools are illustrated using proteomic data of 11 human cell lines as well as 28 mouse tissues. Proteins can be searched in all data and general parameters from several experiments can be extracted, for example signal reproducibility of identified peptides. Project specific cutoff scores are adjusted for the combined data sets to better control global false discovery rates. Moreover, individual protein expression levels can be displayed for each cell line derived from label-free quantification. All information contained in MaxQB is accessible to the community via a user-friendly web interface at http://www.biochem.mpg.de/maxqb.

2.3.2. Contribution

Christoph Schaab developed the MaxQB database. Tamar Geiger provided the 11 cell lines data. I contributed to the design, testing and implementation of the database. In particular, I provided input on how to structure the database and which features to include.

2.3.3. Publication

This paper was published in Molecular and Cellular Proteomics 2012:

Analysis of high-accuracy, quantitative proteomics data in the MaxQB database

Schaab Christoph, Geiger Tamar, Stoehr Gabriele, Cox Juergen and Mann Matthias

Mol Cell Proteomics. 2012 Mar;11(3):M111.014068.

Analysis of High Accuracy, Quantitative Proteomics Data in the MaxQB Database*

Christoph Schaab‡§, Tamar Geiger‡, Gabriele Stoehr‡, Juergen Cox‡, and Matthias Mann‡¶

MS-based proteomics generates rapidly increasing amounts of precise and quantitative information. Analysis of individual proteomic experiments has made great strides, but the crucial ability to compare and store information across different proteome measurements still presents many challenges. For example, it has been difficult to avoid contamination of databases with low quality peptide identifications, to control for the inflation in false positive identifications when combining data sets, and to integrate quantitative data. Although, for example, the contamination with low quality identifications has been addressed by joint analysis of deposited raw data in some public repositories, we reasoned that there should be a role for a database specifically designed for high resolution and quantitative data. Here we describe a novel database termed MaxQB that stores and displays collections of large proteomics projects and allows joint analysis and comparison. We demonstrate the analysis tools of MaxQB using proteome data of 11 different human cell lines and 28 mouse tissues. The database-wide false discovery rate is controlled by adjusting the project specific cutoff scores for the combined data sets. The 11 cell line proteomes together identify proteins expressed from more than half of all human genes. For each protein of interest, expression levels estimated by label-free quantification can be visualized across the cell lines. Similarly, the expression rank order and estimated amount of each protein within each proteome are plotted. We used MaxQB to calculate the signal reproducibility of the detected peptides for the same proteins across different proteomes. Spearman rank correlation between peptide intensity and detection probability of identified proteins was greater than 0.8 for 64% of the proteome, whereas a minority of proteins have negative correlation. This information can be used to pinpoint false protein identifications, independently of peptide database scores. The information contained in MaxQB, including high resolution fragment

spectra, is accessible to the community via a user-friendly web interface at http://www.biochem.mpg.de/maxqb. *Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.014068, 1–10, 2012.*

Bottom-up proteomics consists of the MS analysis of enzymatically digested proteomes. During the last few years, measurements have increasingly been performed in a high resolution, quantitative format (1–3). Each proteomic experiment typically generates large amounts of raw MS and MS/MS data, which should be made available with each experiment (4). Computational proteomics is then used to extract high confidence peptide and protein identifications and relative ratios between conditions, as well as to distill biological implications from the data (5–8). Apart from the analysis of individual projects, several repositories for proteomic experiments have been developed, each with different purposes in mind. The Global Proteome Machine (9) and PeptideAtlas (10, 11) are two of the earliest such collections, with the primary goal of providing a collection of peptide identifications. These collections can, for example, be mined for the design of multiple reaction monitoring experiments in targeted proteomics (12). In contrast, TRANCHE (proteomecommons.org/tranche) is a repository for the raw mass spectrometric data (13). PRoteomics IDEntifications database (PRIDE) is a large effort at the European Bioinformatics Institute, which has collected peptide and protein identification data from more than 10,000 experiments (14, 15). PRIDE, PeptideAtlas, and TRANCHE are also part of the ProteomeXchange consortium, whose objective is to provide a single point of submission for MS-based proteomics data (www.proteomexchange.org). Many dedicated databases for specific organelles or organisms also exist (see for example Refs. 16 and 17).

Most of these databases accept data from heterogeneous sources, which presents a challenge in analysis. For instance, data acquired with different proteomics technologies, different computational pipelines and different quantification strategies may be combined in the database. Although these problems have been addressed to some degree by open standards and joint analysis of deposited raw data, we reasoned that there should be a role for a database designed for homogeneous, quantitative, high resolution data, which nev-

ertheless covers a large part of diverse proteomes. Here we describe the construction of the MaxQB database, which is meant to address the above challenges, allow novel types of analyses, and serve as a public resource via a versatile web interface. We illustrate MaxQB with deep proteome data generated in an accompanying paper (18). In that study, the proteomes of 11 widely used cell lines were mapped in depth with high resolution MS and MS/MS data. We describe analysis and visualization tools of MaxQB, a solution to the problem of inflated false positive protein identifications, and examine the reproducibility of peptide intensity rank order for each protein in different proteomes.

EXPERIMENTAL PROCEDURES

Database Implementation—MaxQB is structured as a classical three-tiered application consisting of data, application logic, and presentation. The data tier is a relational database managed by Oracle Standard Edition Database 11g (Oracle, Redwood Shores, CA). Because only standard SQL features are used, it is in principle possible to port the database to other relational database management systems like the free and open source MySQL database (Oracle). The application logic tier is implemented in Java 1.6 (Oracle) and Groovy (http://groovy.codehaus.org) using the Grails web application framework version 1.3.3 (http://grails.org). The web application runs on a Tomcat 7 web server (http://tomcat.apache.org). Finally, the presentation tier is comprised of dynamically generated html pages and JavaScript.

Protein Index and Mapping to Genome—Several human proteome databases were uploaded to MaxQB to build a comprehensive protein index: Uniprot version 09/2011 (including variants), Ensembl build 64, and International Protein Index (IPI) version 3.87. Identical entries were collapsed to a single logical protein entry where identity of entries is defined by strict sequence identity. For example, the entry for the human protein CDK2 refers to the Uniprot accession number P24941, the Ensembl protein accession number ENSP00000266970, and the IPI accession number IPI00031681. All three database entries have identical sequences. The sequences were first transformed to a hash key using the Secure Hash Algorithm, which dramatically increased the speed of mapping identical sequences. The locations of the genes on the chromosomes were obtained from Ensembl (19), and the ortholog pairs of proteins in different organisms were obtained from InParanoid eukaryotic ortholog database (20).

Cell Line Data—MaxQB already serves as a general repository for experiments performed in our laboratory. Therefore, it will contain an increasing number of deep proteome mapping experiments of human, mouse, and other cell types and species in the future. The data analyzed here are mainly from a proteome profiling experiment of 11 cell lines described in the accompanying paper (18). Briefly, A549, GAMG, HEK293, HeLa, HepG2, Jurkat, K562, LnCap, MCF7, RKO, and U2OS cell lines were grown in standard conditions, lysed, and prepared according to the Filter Aided Sample Preparation method (21) and fractionated by pipette-based strong anion exchange into six fractions. Resulting peptide mixtures were analyzed on-line by LC-MS/MS on a linear ion trap Orbitrap (VELOS, Thermo Fisher Scientific) in higher energy collisional dissociation mode (22). Each proteome measurement—consisting of six 200-min gradients—was repeated in triplicate. Analysis of the results was performed in MaxQuant (23) using the Andromeda search engine (24). The results that are presented here and are accessible in MaxQB are based on data processed with the "match between runs" feature enabled. However, the increase of identifications for additional analyzed cell lines (see Fig. 2) and the correlation analysis (see Fig. 7) is based on data processed

with this feature disabled. For details see Ref. 18. MaxQB and the results of the 11-cell line proteome can be accessed freely upon publication at http://www.biochem.mpg.de/maxqb.

Mouse Tissue Data—In addition to the cell line data, we also analyze data from a proteome profiling experiment of 28 mouse tissues (18). Briefly, 28 tissues were dissected from C57BL/6 mice and snap frozen in liquid nitrogen. The tissues were homogenized, lysed, and mixed with a SILAC[1] spike-in standard. Protein digestion was performed with endoprotease Lys-C, followed by peptide fractionation by isoelectric focusing. The resulting peptide mixtures were analyzed on-line by LC-MS/MS on a linear ion trap Orbitrap (XL, Thermo Fisher Scientific) in CID mode. Each proteome measurement—consisting of twelve 100-min gradients—was repeated in triplicate. Analysis of the results was performed in MaxQuant (23) using the Andromeda search engine (24).

RESULTS AND DISCUSSION

Database Architecture—MaxQB serves as a generic repository and analysis platform for high resolution MS-based proteomics experiments. As such, it stores details about protein and peptide identifications together with the corresponding high or low resolution fragment spectra and quantitative information, such as SILAC ratios or label-free intensities. To enable smooth upload of data, MaxQB is tightly integrated with MaxQuant (23) (Fig. 1). At the end of data processing, the user of MaxQuant is asked whether she wants to upload the data to the database. In this case, the data is submitted by calling a simple object-based protocol (SOAP)-based web service. Alternatively, the data can be manually uploaded through the user interface of MaxQB. In either case, the user is asked to enter additional meta information, such as the project name, experiment name, and workflow parameters. All of the data are stored in a relational SQL database running on an Oracle relational database management system. The user can browse, search, and retrieve the data through a web interface. Furthermore, the data can be accessed either through SQL queries or preferably through SOAP web services from visualization and data analysis tools like the Perseus module for bioinformatic analysis in MaxQuant, R (www.r-project.org), Matlab (The Mathworks, Natick, MA), or Spotfire (TIBCO, Palo Alto, CA).

Protein Index and Cell Line Data—To demonstrate the general concepts and features of MaxQB, data from the proteome profiling of 11 cancer cell lines described in the accompanying paper (18) were uploaded to the database. The combined data were searched against the IPI database 3.68 using MaxQuant version 1.2.0.34. A frequent problem of proteomics experiments is the difficulty of matching protein accession numbers between experiments that were searched against different databases or even just against different versions of the same database (25). Here, we sought to solve this problem by building a protein index that matches the accession

---

[1] The abbreviations used are: SILAC, stable isotope labeling by amino acids in cell culture; FDR, false discovery rate; SOAP, simple object access protocol; IPI, International Protein Index.
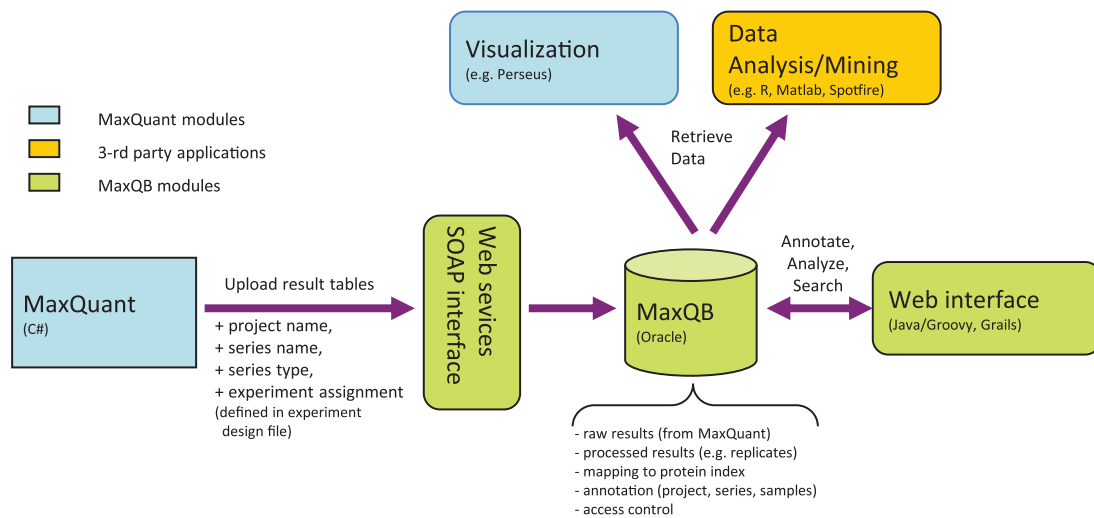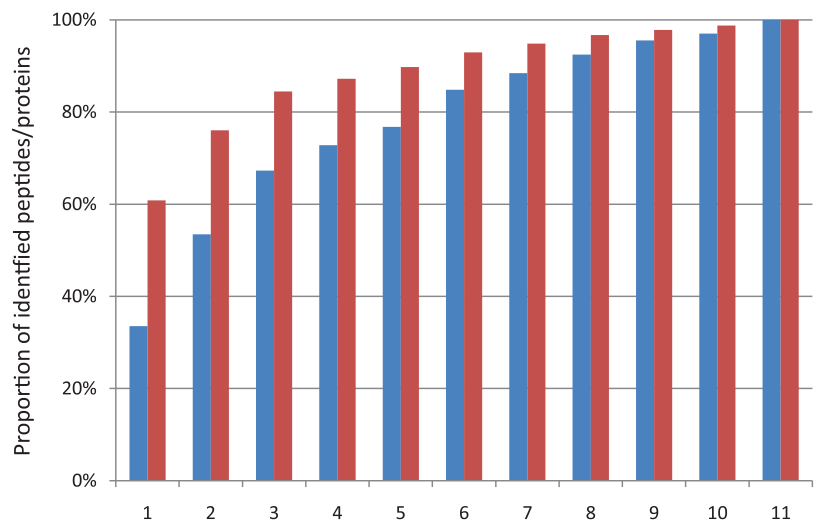
Fɪɢ. 1. **Database architecture and interfaces to other applications.**

Fɪɢ. 2. **Number of proteins (*red bars*) and peptides (*blue bars*) identified in increasing number of cell lines.** In total, 10,183 non-redundant proteins and 103,869 non-redundant peptides were identified (see text for details).

numbers of various popular protein sequence databases to a logical protein entry. For the human species, these databases are Uniprot (including variants), IPI, and Ensembl. In brief, sequence database entries that refer to identical sequence and species are mapped to a unique protein index entry (see "Experimental Procedures" for more details). The protein index contains 19,515 human entries with identical sequences in IPI and Ensembl (see Table I). This nonredundant set was the basis for further analysis. From these proteins, we calculated the number of tryptic peptides readily observable and identifiable by MS (mass between 600 and 4,000 Da; no missed cleavages). There are 536,593 such peptides, and interestingly only 6% of them are shared between two or more proteins.

Triplicate analysis of one cell line alone identified 7,337 proteins, each subsequently added cell line contributed a decreasing number of new proteins, and analysis of all 11 cell line proteomes together identified 10,183 nonredundant proteins (Fig. 2 and Table I). *In silico* digest of the identified proteins generated 338,496 observable tryptic peptides, of

which 32.5% were identified in the cell line data set at a false discovery rate of 1%. For each of these peptides, the database contains the corresponding database identification score, the posterior error probability, individual evidences for the peptide identification, and the corresponding fragment spectra. At this point, proteins encoded by more than half of all human genes and a large proportion of all their possible, unmodified tryptic peptides are identified in the database.

Apart from the 11-cell line project, MaxQB contains a number of large scale experiments on human proteomes. Interestingly, these experiments together already account for proteins encoded by 64% of all human genes and 39% of their possible, unmodified tryptic peptides. This suggests that improving technology will soon make it possible to obtain reference spectra for a large part of the proteome from homogeneous data sources given a supply of diverse proteomes in which all human proteins are expressed.

Use Cases—To illustrate practical use of MaxQB, we next describe three "use cases" dealing with diverse types of

TABLE I

Number of identified proteins (Ensembl genes with identical IPI sequence) and peptides in comparison with respective numbers in the ENSEMBL human database

Observable peptides are *in silico* digested peptides with masses between 0.6 and 4 kDa and no missed cleavages.

	Human proteome	Identified in 11 cell lines
Proteins	19,515	10,183 (52.2%)
Observable peptides	536,593	
Observable, sequence-unique peptides	506,080	
Observable peptides from identified proteins	338,496	109,862 (32.5%)
Observable, sequence-unique peptides from identified proteins	316,585	103,869 (32.8%)
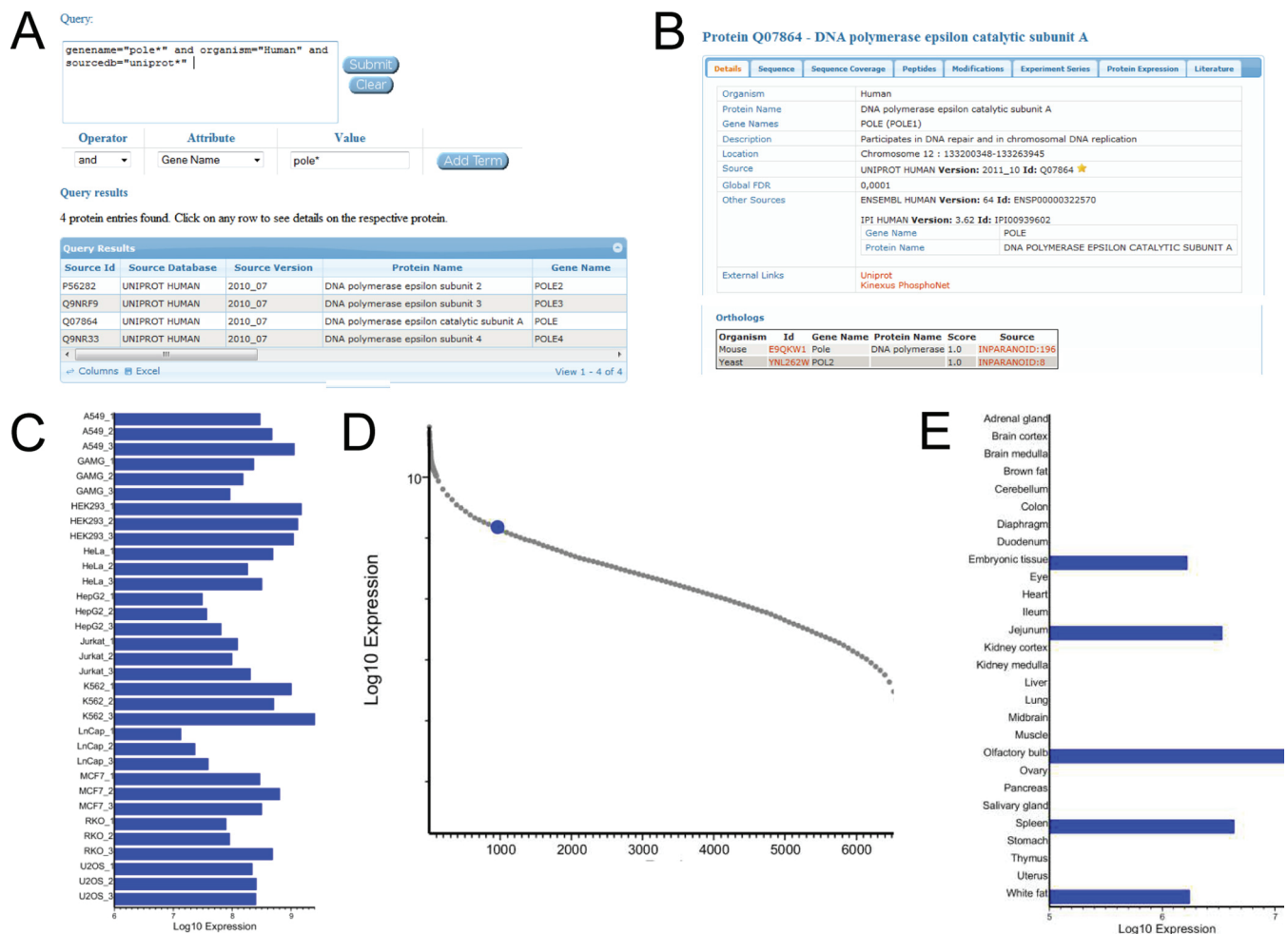
FIG. 3. *A*, query proteins for human DNA polymerase epsilon subunits. *B*, select POLE and show details on this protein. *C*, histogram of protein expression across 11 cell lines. *D*, expression of POLE compared with expression of all other detected proteins in HEK293 cells. *E*, expression of the mouse ortholog across 28 mouse tissues.

questions that can be addressed by this novel database. As a first use case, we assume that the user is interested in members of a specific protein family—here DNA polymerase epsilon subunits (POLE)—and wants to investigate their expression across the different cell lines and additionally across mouse tissues. The user can query the database by various fields, specifically by gene name, organism, and source database. The query terms can be combined by Boolean logic

and grouped using parentheses. Alternatively, the query builder can be used if one is not familiar with the query syntax. In this example, the user searches for all human Uniprot entries that have a gene name beginning with "POLE" (Fig. 3*A*). The query returns four subunits. By clicking on one of the hits (POLE), the user obtains additional details (Fig. 3*B*). In particular, this resulting page specifies the entries in the databases IPI and Ensembl with identical sequence. On the

F_{IG}. 4. **Sequence coverage of POLE.** The *blue boxes* are two c4-type domains. The *gray boxes* are *in silico* digested peptides with masses between 0.6 and 4 kDa. Detected peptides are colored by their label-free intensities across the 11 tested cell lines with three replicates each.

protein expression tab, a bar chart visualizes the protein expression across the 11 human cell lines. Expression of POLE varies by more than 2 orders of magn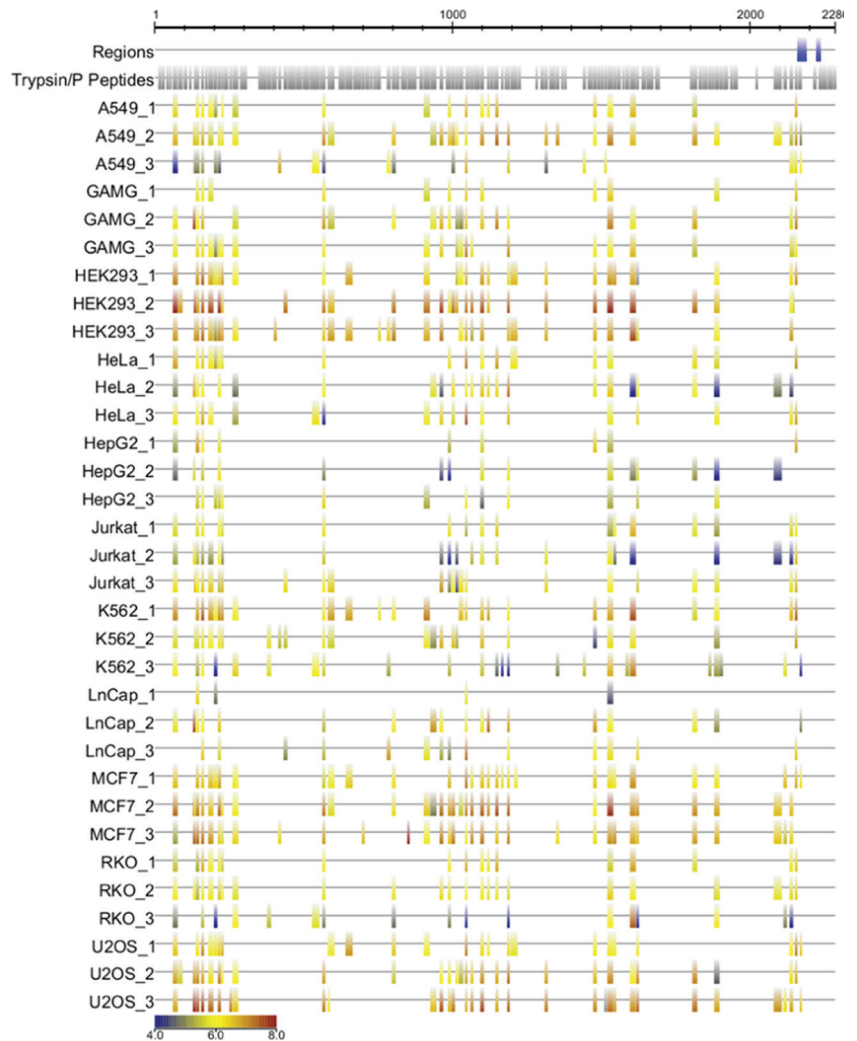itude between LnCap (lowest expression) and HEK293 (highest expression) calculated by label-free quantification in MaxQuant (26) (Fig. 3*C*). In addition to estimating expression of the same protein between proteomes, MaxQB can also display expression within any of the proteomes, compared with all other quantified proteins in that proteome. Here, the expression of the protein is estimated by the sum of its peptide signals, after normalization of the total proteome signals to each other in MaxQuant. The iBAQ algorithm (27) is now implemented into MaxQuant and can also be used to estimate protein amounts. In Fig. 3*D*, selection of the HEK293 proteome brings up a distribution plot comparing the expression of the protein of interest with all other proteins in this cell line. This reveals that POLE is among the highly expressed proteins in these cells (within the top 15-percentile). The sequence coverage tab for the corresponding protein group shows the distribution of identified peptides along the sequence of POLE and across

the 11 cell lines and their biological replicates (Fig. 4). Additionally, the *in silico* digested peptides with masses between 0.6 and 4 kDa and the known domains as retrieved from Uniprot (28) are displayed.

The user may also be interested in the expression of POLE in other organisms. The InParanoid eukaryotic ortholog database contains pairwise orthologs of 100 organisms (20). MaxQB integrates this information to allow the user to jump directly to the proteomes of other organisms. For example, Fig. 3*B* lists two ortholog proteins in yeast and mouse. Clicking on the mouse ortholog (E9QKW1), the user obtains additional information on the mouse protein. As an example of how MaxQB can integrate data from various studies, Fig. 3*E* shows the expression of POLE in 28 mouse tissues (data not published). POLE was identified in embryonic tissue, jejunum, olfactory bulb, spleen, and white fat.

Recently, several projects aiming to identify all proteins encoded on specific chromosomes have been started under the auspice of the Human Proteome Organization (29). In a second use case, we ask how many proteins have been
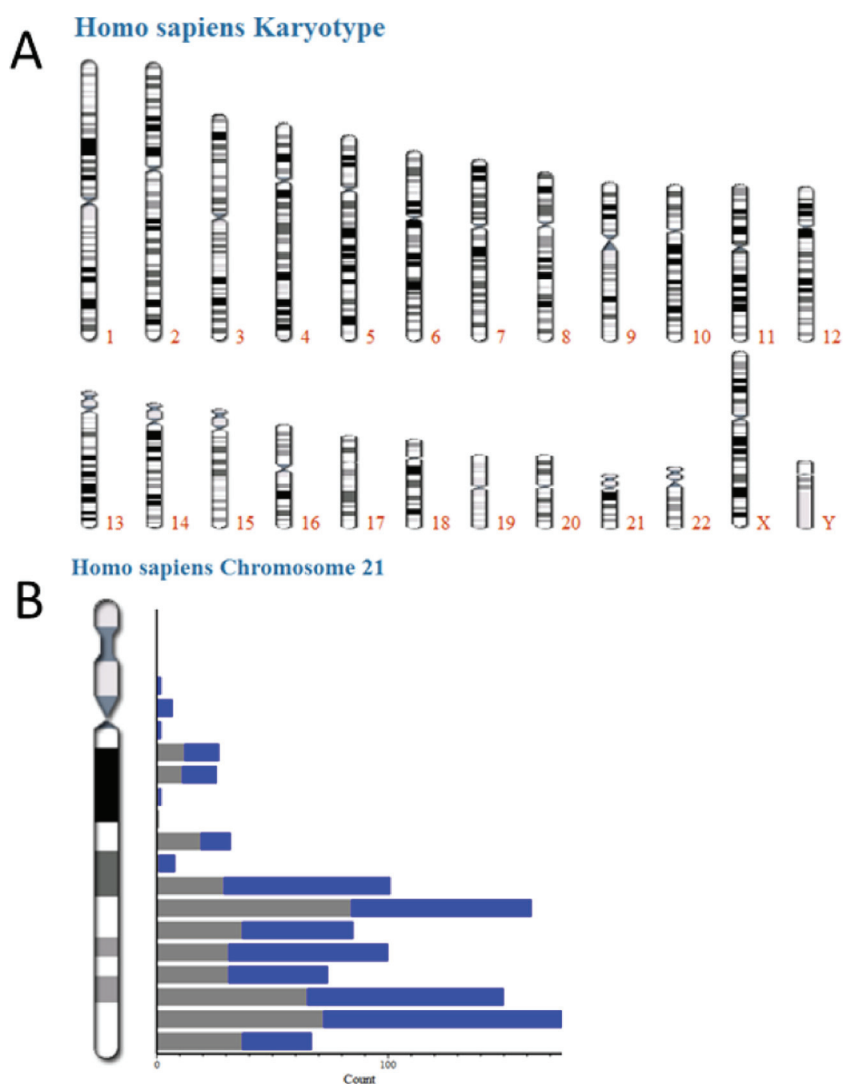
FIG. 5. *A*, human karyotype. *B*, histogram of proteins identified by MS in the 11 cell line project (*gray*) and annotated proteins (*blue*) on chromosome 21.

identified for a certain chromosome and whether there are any regions with low identification rates. MaxQB lists all human (or mouse or yeast) chromosomes and allows the user to select one of them for further analysis (Fig. 5*A*). In the case of chromosome 21, for example, this results in the distribution of protein coding genes and the respective protein identifications in the cell line proteomes shown in Fig. 5*B*. By clicking on one of the bars, the user can drill down to the list of proteins encoded in the corresponding region of the chromosome as well as the underlying peptide information. As expected, ~50% of all annotated genes on chromosome 21 are associated with high confidence protein identification information, and the distribution across the chromosome appears to be uniform.

A popular use of proteomics repositories is the selection of peptides suitable for targeted methods such as multiple reaction monitoring. In the third use case, a user is interested in establishing an multiple reaction monitoring assay for the cell cycle protein CDK2 and starts by searching for all peptides

that are unique for CDK2, have an Andromeda identification score larger than 80, and have no missed cleavages. As in the search for proteins described above, query terms can be combined by Boolean logic (Fig. 6*A*). The query returns seven peptides fulfilling these criteria. The user selects the peptide AFGVPVR and displays the fragment spectrum for the best identification evidence for this peptide (Fig. 6*B*). The user can now export the list of peaks together with the masses and annotations and use this as a basis for creating multiple reaction monitoring transitions. A particular advantage of using MaxQB for this use case is the fact that this database contains high resolution fragmentation spectra that are obtained by the higher energy collisional dissociation method, which produces very similar transitions to those that would be observed in triple quadrupole methods (30).

Inflation of False Identifications—It is a common strategy in MS-based proteomics to control the proportion of false identifications by searching against a combined forward and decoy sequence database and then adjusting the cutoff score to
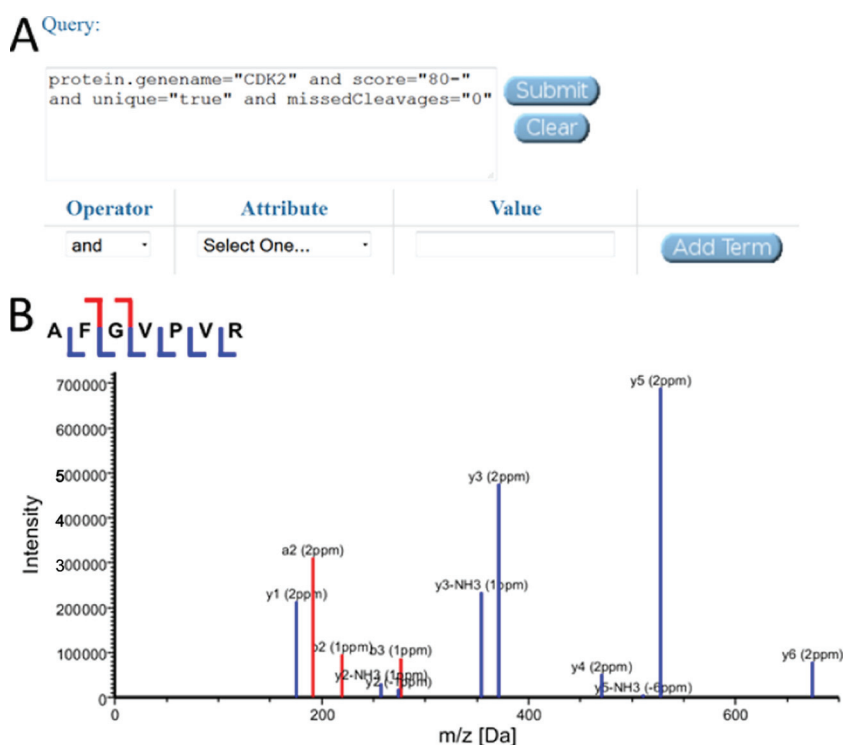
Fig. 6. *A*, query for unique peptides for CDK2 with a score greater 80 and no missed cleavages. *B*, the fragment spectrum with the best evidence for peptide AFGVPVR.

a value, such that the proportion of identified decoy hits is equal to a given false discovery rate (FDR) (31, 32). Although decoy database search is a robust method to control FDR in single projects, an additional challenge arises when combining the results of many different experiments covering the same proteome. In this case the number of true identifications saturates because the same "true" proteome is sampled repeatedly (see for example Fig. 2). However, the false identifications are largely independent of each other and therefore accumulate, leading to an inflation of false identifications. A related but different problem arises when combining the search scores from multiple search engines obtained for the same data set (33, 34). Here, we instead investigate the combination of identifications from multiple proteomes that have been analyzed with the same search engine.

Although generally known this issue has to our knowledge not been quantified with experimental data. We investigated the severity of this problem by analyzing the effects of successively adding data sets to a database instead of analyzing all data together as described above. For this purpose, the raw data of the proteome profiling experiment of 11 cell lines described in the accompanying paper (18) were arbitrarily partitioned into three sets, each consisting of four or three cell lines and their corresponding biological replicates. These three sets were reprocessed by MaxQuant with a fixed FDR for protein and peptide identification of 1%. Each set resulted in ~10,000 total protein identifications and 100 decoy hits (Table II). If these sets were successively added to a single database, the number of true identifications would increase by 28%, whereas the number of decoy hits would increase by

TABLE II
Number of identified proteins if raw files are processed in three disjointed sets

Set 1 includes A549, HEK293, GAMG, and HeLa. Set 2 includes HepG2, Jurkat, K562, and MCF7. Set 3 includes RKO, LNCap, and U2OS. Union refers to the union of the proteins identified in the individual sets. True is the number of forward hits, decoy the number of decoy hits, and FDR is the corresponding false discovery rate.

	Set 1	Set 2	Set 3	Union
True	9,922	9,690	9,054	12,211
Decoy	103	105	93	226
FDR	1.03%	1.07%	1.02%	1.82%

225% compared with the average number in the individual sets. The resulting FDR would now be 1.82% instead of the desired 1%. Clearly, the more proteomics data sets are added to a database, the larger the inflation of false identifications. Often the underlying data are not available for reprocessing, or reprocessing for each added data set would be impractical. For these cases, we propose to solve the issue by adjusting the cutoff score to a more stringent database-wide level. The adjustment is performed such that the ratio between the number of unique decoy database hits and the total number of unique identifications on a protein and peptide level is equal to the desired FDR. MaxQB follows this proposal by calculating a local FDR (q value) for each peptide and each protein identification. This q value for a protein is essentially the ratio between the number of decoy hits and the total number of identifications with scores smaller or equal to the score of that protein (and the peptide q value is calculated
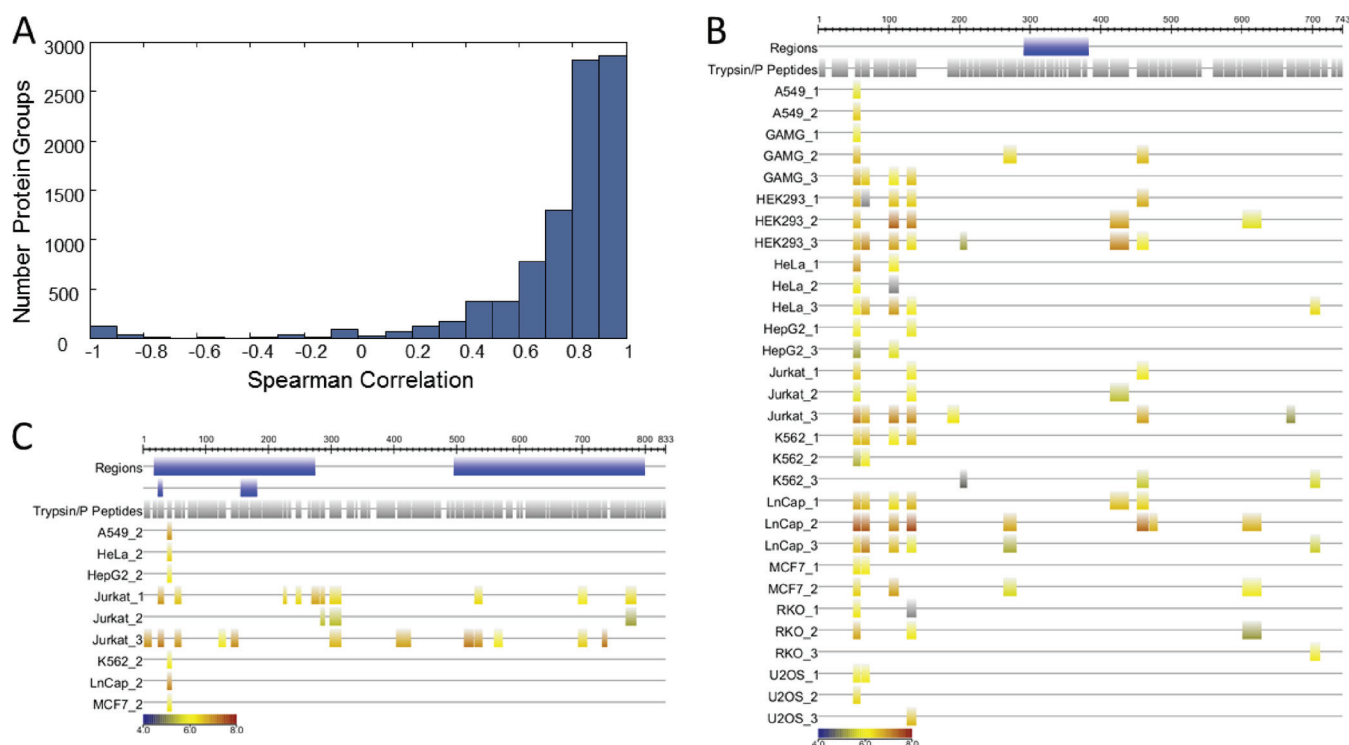
FIG. 7. *A*, distribution of correlation values. For each protein group with two or more peptides identified, the Spearman correlation between the intensities of the peptides and the detection probability were calculated. *B* and *C*, examples of proteins with high correlation (0.92): Q8NFI3-ENGASE (*B*) and low correlation (0.27): Q92918-MAP4K1 (*C*).

analogously). The identifications with *q* values below the pre-set FDR of 1% are filtered out when the user is analyzing data from the whole database rather than data from a single experiment. This strategy is possible because all projects in MaxQB are analyzed with the same search engine (Andromeda) and therefore use the same type of identification score.

Reproducibility of Identified Peptides—As can be seen in the sequence coverage plot of POLE (Fig. 4), the identified peptides are not random between different proteomes but follow certain patterns. As one would expect, the cell line with the highest expression level (HEK293) also shows the largest number of identified peptides. Furthermore, a few POLE peptides are identified in almost all samples (*e.g.* positions 60–77, 1520–1540, and 2132–2145), and these peptides are also the ones with the highest label-free intensity as indicated by the color code. These observations motivated us to investigate possible general relationships between the probability of peptide identification and peptide intensity. For each protein having at least two peptides, we calculated the Spearman rank correlation between the sum of label-free peptide intensities and the number of experiments in which the peptide was detected. The histogram of the correlation values for each protein shows a strong accumulation of proteins with high correlation values (Fig. 7*A*). A total of 64% of the proteins had Spearman rank correlations of more than 0.8. Fig. 7*B* shows the example of ENGASE, a protein with a high correlation value (0.92). Here, the peptides detected many times are also

the most intense ones and vice versa. A few proteins have small or even negative correlations and were therefore investigated in detail. For example, MAP4K1 has a small correlation of 0.27 between peptide intensities and peptide detection probabilities. Whereas the peptides detected in the three Jurkat samples show a high overlap, the peptide VSGDLVALK starting at position 38 was only detected in one replicate of A549, HeLa, HepG2, K562, LNCap, and MCF7, respectively, and it was also the only peptide detected for this protein in these cell lines. We speculate that this peptide is a false identification in the non-Jurkat cell lines, which is further supported by a relatively high posterior error probability. Our analysis clearly suggests that rank order statistics of identified peptides for each protein are sufficiently high to pinpoint false protein identifications, independently of peptide database scores. Therefore a comprehensive catalog of protein and peptide identifications compiled from high quality data, such as those in MaxQB, could be used to improve protein identification in proteomics experiments. Furthermore, public databases could incorporate such algorithms to judge the quality of submitted data sets. If peptide rank correlation of the new data set to established data sets is low, this may indicate problems with the newly submitted data.

Conclusions and Outlook—We have described MaxQB, a resource for high resolution and quantitative MS-based proteomics data. MaxQB draws on a homogenous set of proteome measurements, which allows types of analyses that are

difficult to perform in many other public repositories. Here, the capabilities of MaxQB have been illustrated using deep proteome measurements of 11 different cell lines. These data already cover more than half of the human proteome, and for these proteins any researcher can visualize expression patterns across cell lines as well as estimated expression levels within each of them. The expression data may be used, for example, to select a cell line or tissue that highly expresses the protein of interest. We plan to add even deeper and more diverse data sets in the near future. As an example, the expression levels of an ortholog protein in 28 different mouse tissues can be visualized in MaxQB. Although these efforts may not lead to complete coverage of the proteome, because a number of proteins may not be expressed in readily available sources, we predict that the large majority of proteins and peptides typically observable in proteomics experiments will soon be represented. As in other repositories, the peptide information can be mined for establishing targeted proteomics assays. However, in this regard MaxQB has the advantage of drawing on a relatively focused set of experiments that are strictly controlled for overall false discovery rate. As technology advances, increasingly accurate proteome measurements will be feasible within short measurement times. We envision that the data in MaxQB will periodically be replaced with these superior data (while keeping access to the old data), something that is difficult in broad data repositories that cannot discriminate between data submitted at different stages of technology development. Although MaxQB currently contains proteome data of human cancer cell lines and a set of 26 mouse tissues, we envision that proteomes of additional cell types and species will be added in the future. Furthermore, MaxQB can serve as a repository for more specialized data, for example, proteome changes after treatment with drugs or data on post-translational modifications. All of these data will have in common that they contain high resolution identifications and are produced with a homogenous set of technologies. We plan to implement an automatic submission of the experiments in MaxQB to PRIDE, so that MaxQB data are also available in the databases that are part of ProteomeXchange.

MaxQB also allowed us to investigate the reproducibility of peptide identifications for each protein across proteome experiments. We found a high correlation of peptide rank order, sufficient to highlight false positive protein identifications independently of peptide identification score. This suggests that the peptide rank order can be used as a component of a protein identification score. As MaxQB contains more and more of the typically identifiable proteins and peptides, it will be interesting to investigate whether these data can contribute to better proteome characterization.

Additionally, MaxQB features a number of analysis tools that are not currently present in other databases. For example, we here introduced a procedure to adjust the required cutoff scores to keep the overall false positive rate constant when incremental proteome projects are added. These analysis tools can be used in MaxQB, but they could also be incorporated into other proteome databases.

REFERENCES

1. Mallick, P., and Kuster, B. (2010) Proteomics: A pragmatic perspective. *Nat. Biotechnol.* **28,** 695–709
2. Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80,** 273–299
3. Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28,** 710–721
4. Olsen, J. V., and Mann, M. (2011) Effective representation and storage of mass spectrometry-based proteomic data sets for the scientific community. *Sci. Signal.* **4,** pe7
5. Taylor, C. F., Paton, N. W., Garwood, K. L., Kirby, P. D., Stead, D. A., Yin, Z., Deutsch, E. W., Selway, L., Walker, J., Riba-Garcia, I., Mohammed, S., Deery, M. J., Howard, J. A., Dunkley, T., Aebersold, R., Kell, D. B., Lilley, K. S., Roepstorff, P., Yates, J. R., 3rd, Brass, A., Brown, A. J., Cash, P., Gaskell, S. J., Hubbard, S. J., and Oliver, S. G. (2003) A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat. Biotechnol.* **21,** 247–254
6. Kumar, C., and Mann, M. (2009) Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS Lett.* **583,** 1703–1712
7. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Martin, D. B., Nesvizhskii, A. I., and Aebersold, R. (2010) A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10,** 1150–1159
8. Schaab, C. (2011) Analysis of phosphoproteomics data. *Methods Mol. Biol.* **696,** 41–57
9. Craig, R., Cortens, J. P., and Beavis, R. C. (2004) Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **3,** 1234–1242
10. Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S. N., and Aebersold, R. (2006) The PeptideAtlas project. *Nucleic Acids Res.* **34,** D655–D658
11. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: A resource for target selection for emerging targeted proteomics workflows. *EMBO Reports* **9,** 429–434
12. Hüttenhain, R., Malmström, J., Picotti, P., and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* **13,** 518–525
13. Hill, J. A., Smith, B. E., Papoulias, P. G., and Andrews, P. C. (2010) ProteomeCommons.org collaborative annotation and project management resource integrated with the Tranche repository. *J. Proteome Res.* **9,** 2809–2811
14. Côté, R., Reisinger, F., Martens, L., Barsnes, H., Vizcaino, J. A., and Hermjakob, H. (2010) The Ontology Lookup Service: Bigger and better. *Nucleic Acids Res.* **38,** W155–W160
15. Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., and Apweiler, R. (2005) PRIDE: The proteomics identifications database. *Proteomics* **5,** 3537–3545
16. Ahmad, Y., Boisvert, F. M., Gregor, P., Cobley, A., and Lamond, A. I. (2009) NOPdb: Nucleolar Proteome Database: 2008 update. *Nucleic Acids Res.* **37,** D181–D184
17. Gnad, F., Oroshi, M., Birney, E., and Mann, M. (2009) MAPU 2.0: High-accuracy proteomes mapped to genomes. *Nucleic Acids Res.* **37,** D902–D906

18. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* 10.1074/mcp.M111.014050

19. Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kähäri, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Overduin, B., Pritchard, B., Riat, H. S., Rios, D., Ritchie, G. R., Ruffier, M., Schuster, M., Sobral, D., Spudich, G., Tang, Y. A., Trevanion, S., Vandrovcova, J., Vilella, A. J., White, S., Wilder, S. P., Zadissa, A., Zamora, J., Aken, B. L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernández-Suarez, X. M., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Vogel, J., and Searle, S. M. (2011) Ensembl 2011. *Nucleic Acids Res.* **39,** D800–D806

20. Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D. N., Roopra, S., Frings, O., and Sonnhammer, E. L. (2010) InParanoid 7: New algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38,** D196–D203

21. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362

22. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8,** 2759–2769

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

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

25. Griss, J., Cote, R. G., Gerner, C., Hermjakob, H., and Vizcaino, J. A. (2011) Published and Perished? The influence of the searched protein database on the long-term storage of proteomics data. *Mol. Cell. Proteomics* 10: M111.00490

26. Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32,** 279–289

27. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* **473,** 337–342

28. Consortium, U. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **39,** D214–D219

29. Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C. H., Corthals, G. L., Costello, C. E., Deutsch, E. W., Domon, B., Hancock, W., He, F., Hochstrasser, D., Marko-Varga, G., Salekdeh, G. H., Sechi, S., Snyder, M., Srivastava, S., Uhlen, M., Wu, C. H., Yamamoto, T., Paik, Y. K., and Omenn, G. S. (2011) The Human Proteome Project: Current state and future direction. *Mol. Cell. Proteomics* 10: M111.009993

30. de Graaf, E. L., Altelaar, A. F., van Breukelen, B., Mohammed, S., and Heck, A. J. (2011) Improving SRM assay development: A global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J. Proteome Res.* **10,** 4334–4341

31. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4,** 207–214

32. Nesvizhskii, A. I. (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **73,** 2092–2123

33. Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R. L., Aebersold, R., and Nesvizhskii, A. I. (2011) iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* 10: M111.007690

34. Kwon, T., Choi, H., Vogel, C., Nesvizhskii, A. I., and Marcotte, E. M. (2011) MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines. *J. Proteome Res.* **10,** 2949–2958

3. CONCLUSION AND OUTLOOK

3.1. Unraveling of general response in aneuploid human cell lines

Extra chromosome copies markedly alter the physiology of eukaryotic cells. These changes are the main causes of spontaneous miscarriages, disease and developmental abnormalities in humans. It is not clear yet how cells deal with the extra chromosomes and why such severe effects occur. Therefore it is of great importance to unravel molecular changes in aneuploid cells at the mRNA and protein level, focusing on human cells, since no such studies have been performed yet in this important system.

Here we have generated tri-and tetrasomic human HCT116 and RPE-1 cell lines by micronuclei transfer (117) and compared these cells to their diploid counterpart cells. Exactly the same cell populations were analyzed on genome, transcriptome and proteome levels by aCGH, microarray and SILAC techniques and quantitative changes between aneuploid and diploid cells were determined. One focus was on the regulation of the population of mRNAs and proteins encoded on the extra chromosome, however, we also investigated global trends in all aneuploid clones.

We observed significant growth delays in the aneuploid populations in comparison to their diploid counterpart population. These growth delays appear to be linked to cell cycle phases G1 and S. Moreover, the additional chromosome load correlated with growth defects. G2 phase and mitosis were not affected nor did we observe increased cell death or numbers of non-proliferating cells.

In some cell lines spontaneous losses of parts of the additional chromosome occurred and some clones even lost a whole arm of a specific chromosome already after few passages. Such instable clones were excluded from our current studies (although they will be subject to further studies) which allowed us to draw stronger conclusions than otherwise possible. Moreover, for comparative studies, exactly the same cell population needs to be used for all analyses to avoid discrepancies caused by chromosomal changes. Analysis of the DNA content of the cell by e.g. aCGH (array comparative genomic hybridization) represents a useful tool in this regards.

Besides characterization of morphological changes of aneuploid cells in comparison to their diploid counterpart cells, we inspected the fate of proteins encoded on the supernumerary chromosome at the mRNA and protein levels. Whereas for tetrasome chromosomes, mRNA levels were not balanced - meaning the average population of underlying aneuploid/diploid ratios was doubled (in log 2 it was located at 1) - the corresponding protein population was shifted towards disome levels and was located between 0.6 and 0.7. Similar trends were also observed in trisome clones. Proteins included

in macromolecular complexes were highly balanced. In general, we conclude that in case of aneuploidy excess proteins are balanced and especially proteins representing subunits of complexes are strongly regulated to maintain proper stoichiometries.

To further investigate global effects of aneuploidy in the different clones we performed 2-dimensional annotation enrichment analysis. In this strategy pathways that are significantly enriched in the up-regulated or down-regulated protein fraction are extracted. Interestingly, in all clones we observed similar trends of certain important regulatory pathways. Most remarkably, lysosomal pathways including autophagy were significantly enriched in all clones, pathways which are responsible for the degradation of proteins. These results drew our attention to the pathway of autophagy and underlying proteins. Indeed, in all inspected clones, p62 and LC3, key markers of the autophagic degradation pathways, were increased in abundance. Moreover, ubiquitinated proteins co-localized with p62. We therefore hypothesize that excess proteins aggregate and are degraded via the autophagic pathway. Consequently, this pathway may play an important role in aneuploid cells and serve as target for future cancer therapeutic developments.

In this study, human aneuploid clones were analyzed for the first time in a global manner in which proteomic results were correlated to mRNA and DNA level to provide a complete picture of aneuploidy response. Important biological insights emerged, such as a balancing effect of a subset of proteins encoded on the extra chromosome. Moreover, we revealed a general response in different aneuploid clones including different human cell lines. The identification of autophagy as potential mechanism to regulate excess protein levels may become important for the field of aneuploidy research. After providing this global view, these observations can be followed up in detailed, functional studies. Moreover, autophagy inhibition and its influence in aneuploid cells should be examined with a view to clinical applications.

3.2. Identification of TRAIL-induced caspase-dependent substrates

Apoptosis, a controlled form of cell death in metazoans, is crucial for cell homeostasis and thereby also plays an important role in development. Evasion of apoptosis is a hallmark of cancer, inhibiting controlled cell death in cells with obvious malfunctions. The main players in apoptosis are caspases, a specific type of protease family, which are activated by cleavage and subsequently cleave further proteins in the cell to step by step induce the demolition of the cell.

In this thesis, I developed a quantitative SILAC-based approach for the identification of proteolytically cleaved substrates including a sophisticated algorithm for the extraction of cleaved

proteins to investigate the events of apoptosis induced by the extrinsic stimulus TRAIL. TRAIL is of special clinical relevance for cancer research since this apoptotic stimulus is known to induce apoptosis only in cancer cells without cytotoxicity in normal cells.

I identified nearly 700 cleavage substrates, a dataset which can serve as a resource for the community of cell death researchers to study TRAIL-induced cleavage events. Moreover, several characteristics of the apoptotic substrate proteome became apparent. I found mitochondria clearly underrepresented in the population of cleavage substrates and in addition discovered that stable complexes appear to be destroyed by several apoptotic cleavage events in the different complex members. In this process, cleavage of the subunits occurs to a lower extent than that of free proteins. Cleavage events in general are not randomly distributed but specifically target proteins and underlying pathways. Supporting this view, many novel substrates link to already known cleavage substrates and cleavage networks.

Overall, this study proved the applicability of the approach introduced here and provided a global dataset of TRAIL-induced cleavage events. It would now be of great interest to investigate in more detail the event of TRAIL-induced apoptosis. In this endeavor, our approach might be used to investigate the effect of different caspase inhibitors, providing information on specific substrates of the corresponding caspases. The comparison of different stimuli, such as the extrinsic stimuli TRAIL and FasL, including their specific degradome may be of special interest. In addition, the unraveling of substrates in a time-resolved manner by time-course experiments would be possible. As this approach can be applied to any system involving proteolytic cleavage it could also be applied to fields such as embryogenesis or neurodegenerative diseases.

All data of this study were uploaded to the MaxQB database, which has been described in section 2.3. This database is perfectly suited to combine experimental data and thus serve as a resource for important information gathered in a mass spectrometry-based proteomics experiments. We extended the database by not only uploading MaxQuant data, but in addition providing all information on the cleavage events including 3D cleavage plots and cleavage statistics. Each experiment can be searched separately. Furthermore, all information on the cleavage derived from the different experiments can be directly compared, such as the 3D cleavage plots from one protein over all experiments. Therefore this database is particularly suited to store data for data comparison.

Abbreviations

aCGH	Array Comparative Genomic Hybridization
AQUA	Absolute Quantitation
Asp	Aspartate
CID	Collision Induced Dissociation
CLL	Chronic lymphocytic leukemia
CRD	Cysteine-Rich Domain
DISC	Death-Inducing Signaling Complex
DR	Death Receptor
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
FasL	Fas Ligand
FDR	False Discovery Rate
FISH	Fluorescent In-Situ Hybridization
FWHM	Full Width Half Maximum
HCD	Higher energy Collision Dissociation
HPLC	High Performance Liquid Chromatography
IAP	Inhibitor of Apoptosis Protein
ICAD	Inhibitor of the Caspase-Activated DNAse CAD
ID	Inner Diameter
iTRAQ	Isotopic Tag for Relative and Absolute Quantification
LC	Liquid Chromatography
LTQ	Linear Trap Quadrupole
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted Laser Desorption/Ionization
MEF	Mouse Embryonic Fibroblast
MMCT	Microcell-Mediated Chromosome Transfer
MRM/SRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
OPG	Osteoprotegerin
ppm	Parts per Million
PQD	Pulsed Q dissociation
PrEST	Protein Epitope Signature Tag
PSAQ	Protein Standard Absolute Quantification
pSILAC	Pulsed SILAC
PTM	Post Translational Modification
qTOF	Quadrupole time-of-flight
RP	Reversed phase
SAC	Spindle-Assembly Checkpoint
SCX	Strong Cation eXchange chromatography
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide gelelectrophoresis
SILAC	Stable Isotope Labeling with Amino acids in Cell culture
SKY	Spectral Karyotyping
SNP	Single Nucleotide Polymorphism
THD	TNF Homology Domain
TMT	Tandem Mass Tag
TNF	Tumor Necrosis Factor
TOF	Time-of-flight
TRAIL	TNF-Related Apoptosis-Inducing Ligand
UPLC	Ultra Performance Liquid Chromatography

References

1.   Altelaar, A. F., and Heck, A. J. (2012) Trends in ultrasensitive proteomics. *Current opinion in chemical biology* 16, 206-213.

2.   Andersen, J. S., and Mann, M. (2006) Organellar proteomics: turning inventories into insights. *EMBO reports* 7, 874-879.

3.   Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B., and Drewes, G. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nature biotechnology* 25, 1035-1044.

4.   Siuti, N., and Kelleher, N. L. (2007) Decoding protein modifications using top-down mass spectrometry. *Nature methods* 4, 817-821.

5.   Zhou, H., Ning, Z., Starr, A. E., Abu-Farha, M., and Figeys, D. (2012) Advancements in top-down proteomics. *Analytical chemistry* 84, 720-734.

6.   Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V., Neubauer, G., Shevchenko, A., Mortensen, P., and Mann, M. (1996) A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochemical Society transactions* 24, 893-896.

7.   Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP* 3, 608-614.

8.   Steen, H., and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology* 5, 699-711.

9.   de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C., and Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455, 1251-1254.

10.   Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of proteome research* 10, 1785-1793.

11.   Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Frohlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & cellular proteomics : MCP* 10, M110 003699.

12.   Kocher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Analytical chemistry* 83, 2699-2704.

13.   Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Molecular & cellular proteomics : MCP* 11, M111 013722.

14.   Horie, K., Sato, Y., Kimura, T., Nakamura, T., Ishihama, Y., Oda, Y., Ikegami, T., and Tanaka, N. (2012) Estimation and optimization of the peak capacity of one-dimensional gradient high performance liquid chromatography using a long monolithic silica capillary column. *Journal of chromatography. A* 1228, 283-291.

15.   Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* 19, 242-247.

16.   Macek, B., Mann, M., and Olsen, J. V. (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. *Annual review of pharmacology and toxicology* 49, 199-221.

17.   Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J., and Gygi, S. P. (2008) A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 105, 10762-10767.

18.   Taouatas, N., Altelaar, A. F., Drugan, M. M., Helbig, A. O., Mohammed, S., and Heck, A. J. (2009) Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Molecular & cellular proteomics : MCP* 8, 190-200.

19.   Ballif, B. A., Villen, J., Beausoleil, S. A., Schwartz, D., and Gygi, S. P. (2004) Phosphoproteomic analysis of the developing mouse brain. *Molecular & cellular proteomics : MCP* 3, 1093-1101.

20.   Aivaliotis, M., Gevaert, K., Falb, M., Tebbe, A., Konstantinidis, K., Bisle, B., Klein, C., Martens, L., Staes, A., Timmerman, E., Van Damme, J., Siedler, F., Pfeiffer, F., Vandekerckhove, J., and Oesterhelt, D. (2007) Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis. *Journal of proteome research* 6, 2195-2204.

21.   Wisniewski, J. R., Nagaraj, N., Zougman, A., Gnad, F., and Mann, M. (2010) Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. *Journal of proteome research* 9, 3280-3289.

22.   Di Palma, S., Mohammed, S., and Heck, A. J. (2012) ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics. *Nature protocols* 7, 2041-2055.

23.   McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Molecular & cellular proteomics : MCP* 7, 971-980.

24.   Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E., and Huber, C. G. (2007) Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *Journal of proteome research* 6, 4363-4373.

25.   Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71.

26.   Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988) Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry *Rapid communications in mass spectrometry : RCM* 2, 151-153.

27.   Karas, M., Bachmann, D., and Hillenkamp, F. (1985) Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Analytical chemistry* 57, 2935-2939.

28.   Karas, M., and Hillenkamp, F. (1988) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical chemistry* 60, 2299-2301.

29.   Wilm, M. (2011) Principles of electrospray ionization. *Molecular & cellular proteomics : MCP* 10, M111 009407.

30.   Iribarne, J. V., and Thomson, B. A. (1976) Evaporation of Small Ions from Charged Droplets. *Journal of Chemical Physics* 64, 2287-2294.

31.   Dole, M., Mack, L. L., and Hines, R. L. (1968) Molecular Beams of Macroions. *Journal of Chemical Physics* 49, 2240-&.

32.   Wilm, M. S., and Mann, M. (1994) Electrospray and Taylor-Cone Theory, Doles Beam of Macromolecules at Last. *Int J Mass Spectrom* 136, 167-180.

33.   Gustavsson, S. A., Samskog, J., Markides, K. E., and Langstrom, B. (2001) Studies of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents. *Journal of chromatography. A* 937, 41-47.

34.   Di Palma, S., Hennrich, M. L., Heck, A. J., and Mohammed, S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of proteomics* 75, 3791-3813.

35.   Choudhary, C., and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nature reviews. Molecular cell biology* 11, 427-439.

36.   Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 4, 2010-2021.

37.   Scigelova, M., Hornshaw, M., Giannakopulos, A., and Makarov, A. (2011) Fourier transform mass spectrometry. *Molecular & cellular proteomics : MCP* 10, M111 009431.

38.   Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass spectrometry reviews* 17, 1-35.

39.   Hardman, M., and Makarov, A. A. (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. *Analytical chemistry* 75, 1699-1705.

40.   Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical chemistry* 72, 1156-1162.

41.   Scigelova, M., and Makarov, A. (2006) Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* 6 Suppl 2, 16-21.

42.   Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., and Horning, S. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry* 78, 2113-2120.

43.   Walther, T. C., and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology* 190, 491-500.

44.   Makarov, A., Denisov, E., and Lange, O. (2009) Performance evaluation of a high-field Orbitrap mass analyzer. *Journal of the American Society for Mass Spectrometry* 20, 1391-1396.

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

46.   Olsen, J. V., and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13417-13422.

47.   Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., and Coon, J. J. (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Analytical chemistry* 76, 3590-3598.

48.   Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., and Kuster, B. (2008) Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* 7, 1702-1713.

49.   Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* 4, 709-712.

50.   Michalski, A., Neuhauser, N., Cox, J., and Mann, M. (2012) A Systematic Investigation into the Nature of Tryptic HCD Spectra. *Journal of proteome research*.

51.   Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* 8, 2759-2769.

52.   Nagaraj, N., D'Souza, R. C., Cox, J., Olsen, J. V., and Mann, M. (2012) Correction to Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. *Journal of proteome research*.

53.   Zubarev, R. A. (2004) Electron-capture dissociation tandem mass spectrometry. *Current opinion in biotechnology* 15, 12-16.

54.   Chalkley, R. J., Medzihradszky, K. F., Lynn, A. J., Baker, P. R., and Burlingame, A. L. (2010) Statistical analysis of Peptide electron transfer dissociation fragmentation mass spectrometry. *Analytical chemistry* 82, 579-584.

55.   Kim, M. S., and Pandey, A. (2012) Electron transfer dissociation mass spectrometry in proteomics. *Proteomics* 12, 530-542.

56.   Swaney, D. L., McAlister, G. C., and Coon, J. J. (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature methods* 5, 959-964.

57.   Zhao, P., Viner, R., Teo, C. F., Boons, G. J., Horn, D., and Wells, L. (2011) Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *Journal of proteome research* 10, 4088-4104.

58.   Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., and Hunt, D. F. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et biophysica acta* 1764, 1811-1822.

59.   Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J. J., Cox, J., Horning, S., Mann, M., and Makarov, A. (2012) Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Molecular & cellular proteomics : MCP* 11, O111 013698.

60.   Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* 10, M111 011015.

61.   Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 389, 1017-1031.

62.   Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* 17, 994-999.

63.   Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6591-6596.

64.   Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* 404, 939-965.

65.   Desiderio, D. M., and Kai, M. (1983) Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomedical mass spectrometry* 10, 471-479.

66.   Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* 100, 6940-6945.

67.   Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature methods* 2, 587-589.

68.   Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., and Mann, M. (2012) A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* 11, O111 009613.

69.   Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* 1, 376-386.

70.   Ong, S. E., Foster, L. J., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* 29, 124-130.

71.   Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry* 80, 273-299.

72.   Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.

73.   Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signaling* 3, ra3.

74.   Gouw, J. W., Krijgsveld, J., and Heck, A. J. (2010) Quantitative proteomics by metabolic labeling of model organisms. *Molecular & cellular proteomics : MCP* 9, 11-24.

75.   Lam, Y. W., Lamond, A. I., Mann, M., and Andersen, J. S. (2007) Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Current biology : CB* 17, 749-760.

76.   Milner, E., Barnea, E., Beer, I., and Admon, A. (2006) The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Molecular & cellular proteomics : MCP* 5, 357-365.

77.   Schwanhausser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205-209.

78.   Cambridge, S. B., Gnad, F., Nguyen, C., Bermejo, J. L., Kruger, M., and Mann, M. (2011) Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *Journal of proteome research* 10, 5275-5284.

79.   Ebner, O. A., and Selbach, M. (2011) Whole cell proteome regulation by microRNAs captured in a pulsed SILAC mass spectrometry approach. *Methods in molecular biology* 725, 315-331.

80.   Looso, M., Borchardt, T., Kruger, M., and Braun, T. (2010) Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry. *Molecular & cellular proteomics : MCP* 9, 1157-1166.

81.   Wu, Z., Moghaddas Gholami, A., and Kuster, B. (2012) Systematic identification of the HSP90 candidate regulated proteome. *Molecular & cellular proteomics : MCP* 11, M111 016675.

82.   Kruger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 134, 353-364.

83.   Sury, M. D., Chen, J. X., and Selbach, M. (2010) The SILAC fly allows for accurate protein quantification in vivo. *Molecular & cellular proteomics : MCP* 9, 2173-2183.

84.   Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 7, 383-385.

85.   Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., and Mann, M. (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature protocols* 6, 147-157.

86.   Ishihama, Y., Sato, T., Tabata, T., Miyamoto, N., Sagane, K., Nagasu, T., and Oda, Y. (2005) Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nature biotechnology* 23, 617-621.

87.   Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP* 3, 1154-1169.

88.   Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry* 75, 1895-1904.

89.   Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7, 340-350.

90.   Dayon, L., and Sanchez, J. C. (2012) Relative protein quantification by MS/MS using the tandem mass tag technology. *Methods in molecular biology* 893, 115-127.

91.   Griffin, T. J., Xie, H., Bandhakavi, S., Popko, J., Mohan, A., Carlis, J. V., and Higgins, L. (2007) iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *Journal of proteome research* 6, 4200-4209.

92.   McAlister, G. C., Phanstiel, D. H., Brumbaugh, J., Westphall, M. S., and Coon, J. J. (2011) Higher-energy collision-activated dissociation without a dedicated collision cell. *Molecular & cellular proteomics : MCP* 10, O111 009456.

93.   Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature methods* 8, 937-940.

94.   Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S., and Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature methods* 8, 933-935.

95.   Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.

96.   Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* 32, 279-289.

97.   Mueller, L. N., Brusniak, M. Y., Mani, D. R., and Aebersold, R. (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *Journal of proteome research* 7, 51-61.

98.   Podwojski, K., Eisenacher, M., Kohl, M., Turewicz, M., Meyer, H. E., Rahnenfuhrer, J., and Stephan, C. (2010) Peek a peak: a glance at statistics for quantitative label-free proteomics. *Expert review of proteomics* 7, 249-261.

99.   Wong, J. W., Sullivan, M. J., and Cagney, G. (2008) Computational methods for the comparative quantification of proteins in label-free LCn-MS experiments. *Briefings in bioinformatics* 9, 156-165.

100.   Eng, J. K., Fischer, B., Grossmann, J., and Maccoss, M. J. (2008) A fast SEQUEST cross correlation algorithm. *Journal of proteome research* 7, 4598-4602.

101.   Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* 10, 1794-1805.

102.   Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols* 4, 698-705.

103.   Elias, J. E., and Gygi, S. P. (2010) Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in molecular biology* 604, 55-71.

104.   Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* 4, 207-214.

105.   Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25, 25-29.

106.   Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic acids research* 32, D277-280.

107.   Kumar, C., and Mann, M. (2009) Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS letters* 583, 1703-1712.

108.   Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol Cell Proteomics* 11, M111 014068.

109.   Nagaoka, S. I., Hassold, T. J., and Hunt, P. A. (2012) Human aneuploidy: mechanisms and new insights into an age-old problem. *Nature reviews. Genetics* 13, 493-504.

110.   Pfau, S. J., and Amon, A. (2012) Chromosomal instability and aneuploidy in cancer: from yeast to man. *EMBO reports* 13, 515-527.

111.   Siegel, J. J., and Amon, A. (2012) New Insights into the Troubles of Aneuploidy. *Annual review of cell and developmental biology*.

112.   Kingsbury, M. A., Yung, Y. C., Peterson, S. E., Westra, J. W., and Chun, J. (2006) Aneuploidy in the normal and diseased brain. *Cellular and molecular life sciences : CMLS* 63, 2626-2641.

113.   Gordon, D. J., Resio, B., and Pellman, D. (2012) Causes and consequences of aneuploidy in cancer. *Nature reviews. Genetics* 13, 189-203.

114.   Weaver, B. A., Silk, A. D., Montagna, C., Verdier-Pinard, P., and Cleveland, D. W. (2007) Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer cell* 11, 25-36.

115.   Perera, D., Tilston, V., Hopwood, J. A., Barchi, M., Boot-Handford, R. P., and Taylor, S. S. (2007) Bub1 maintains centromeric cohesion by activation of the spindle checkpoint. *Developmental cell* 13, 566-579.

116.   Pavelka, N., Rancati, G., Zhu, J., Bradford, W. D., Saraf, A., Florens, L., Sanderson, B. W., Hattem, G. L., and Li, R. (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468, 321-325.

117.   Haugen, A. C., Goel, A., Yamada, K., Marra, G., Nguyen, T. P., Nagasaka, T., Kanazawa, S., Koike, J., Kikuchi, Y., Zhong, X., Arita, M., Shibuya, K., Oshimura, M., Hemmi, H., Boland, C. R., and Koi, M. (2008) Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer research* 68, 8465-8472.

118.   Upender, M. B., Habermann, J. K., McShane, L. M., Korn, E. L., Barrett, J. C., Difilippantonio, M. J., and Ried, T. (2004) Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer research* 64, 6941-6949.

119.   Williams, B. R., Prabhu, V. R., Hunter, K. E., Glazier, C. M., Whittaker, C. A., Housman, D. E., and Amon, A. (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* 322, 703-709.

120.   Zenzes, M. T., and Casper, R. F. (1992) Cytogenetics of human oocytes, zygotes, and embryos after in vitro fertilization. *Human genetics* 88, 367-375.

121.   Steele, M. W., and Breg, W. R., Jr. (1966) Chromosome analysis of human amniotic-fluid cells. *Lancet* 1, 383-385.

122.   Clyde, J. M., Hogg, J. E., Rutherford, A. J., and Picton, H. M. (2003) Karyotyping of human metaphase II oocytes by multifluor fluorescence in situ hybridization. *Fertility and sterility* 80, 1003-1011.

123.   Sandalinas, M., Marquez, C., and Munne, S. (2002) Spectral karyotyping of fresh, non-inseminated oocytes. *Molecular human reproduction* 8, 580-585.

124.   Fragouli, E., Lenzi, M., Ross, R., Katz-Jaffe, M., Schoolcraft, W. B., and Wells, D. (2008) Comprehensive molecular cytogenetic analysis of the human blastocyst stage. *Human reproduction* 23, 2596-2608.

125.   Fragouli, E., Wells, D., and Delhanty, J. D. (2011) Chromosome abnormalities in the human oocyte. *Cytogenetic and genome research* 133, 107-118.

126.   Stankiewicz, P., and Lupski, J. R. (2010) Structural variation in the human genome and its role in disease. *Annual review of medicine* 61, 437-455.

127.   Shinawi, M., and Cheung, S. W. (2008) The array CGH and its clinical applications. *Drug discovery today* 13, 760-770.

128.   Beaudet, A. L., and Belmont, J. W. (2008) Array-based DNA diagnostics: let the revolution begin. *Annual review of medicine* 59, 113-129.

129.   Bridges, C. B. (1921) Genetical and Cytological Proof of Non-disjunction of the Fourth Chromosome of Drosophila Melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* 7, 186-192.

130.   Bridges, C. B. (1921) Triploid Intersexes in Drosophila Melanogaster. *Science* 54, 252-254.

131.   Niwa, O., Tange, Y., and Kurabayashi, A. (2006) Growth arrest and chromosome instability in aneuploid yeast. *Yeast* 23, 937-950.

132.   Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., and Amon, A. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317, 916-924.

133.   Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A., and Deutsch, S. (2004) Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nature reviews. Genetics* 5, 725-738.

134.   Lin, H. Y., Lin, S. P., Chen, Y. J., Hung, H. Y., Kao, H. A., Hsu, C. H., Chen, M. R., Chang, J. H., Ho, C. S., Huang, F. Y., Shyur, S. D., Lin, D. S., and Lee, H. C. (2006) Clinical characteristics and survival of trisomy 18 in a medical center in Taipei, 1988-2004. *American journal of medical genetics. Part A* 140, 945-951.

135.   Moerman, P., Fryns, J. P., van der Steen, K., Kleczkowska, A., and Lauweryns, J. (1988) The pathology of trisomy 13 syndrome. A study of 12 cases. *Human genetics* 80, 349-356.

136.   Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003) Chromosome aberrations in solid tumors. *Nature genetics* 34, 369-376.

137.   Weaver, B. A., and Cleveland, D. W. (2006) Does aneuploidy cause cancer? *Current opinion in cell biology* 18, 658-667.

138.   Boveri, T. (1914) *Zur frage der entstehung maligner tumoren*, G. Fischer, Jena,.

139.   Hardy, P. A., and Zacharias, H. (2005) Reappraisal of the Hansemann-Boveri hypothesis on the origin of tumors. *Cell biology international* 29, 983-992.

140.   Williams, B. R., and Amon, A. (2009) Aneuploidy: cancer's fatal flaw? *Cancer research* 69, 5289-5291.

141.   Li, M., Fang, X., Baker, D. J., Guo, L., Gao, X., Wei, Z., Han, S., van Deursen, J. M., and Zhang, P. (2010) The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 107, 14188-14193.

142.   Sotillo, R., Hernando, E., Diaz-Rodriguez, E., Teruya-Feldstein, J., Cordon-Cardo, C., Lowe, S. W., and Benezra, R. (2007) Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer cell* 11, 9-23.

143.   Storchova, Z., and Pellman, D. (2004) From polyploidy to aneuploidy, genome instability and cancer. *Nature reviews. Molecular cell biology* 5, 45-54.

144.   Solomon, D. A., Kim, T., Diaz-Martinez, L. A., Fair, J., Elkahloun, A. G., Harris, B. T., Toretsky, J. A., Rosenberg, S. A., Shukla, N., Ladanyi, M., Samuels, Y., James, C. D., Yu, H., Kim, J. S., and Waldman, T. (2011) Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* 333, 1039-1043.

145.   Weaver, B. A., and Cleveland, D. W. (2009) The role of aneuploidy in promoting and suppressing tumors. *The Journal of cell biology* 185, 935-937.

146.   Tang, Y. C., Williams, B. R., Siegel, J. J., and Amon, A. (2011) Identification of aneuploidy-selective antiproliferation compounds. *Cell* 144, 499-512.

147.   Thomas, P., and Fenech, M. (2008) Chromosome 17 and 21 aneuploidy in buccal cells is increased with ageing and in Alzheimer's disease. *Mutagenesis* 23, 57-65.

148.   Birchler, J. A., Bhadra, U., Bhadra, M. P., and Auger, D. L. (2001) Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental biology* 234, 275-288.

149.   Springer, M., Weissman, J. S., and Kirschner, M. W. (2010) A general lack of compensation for gene dosage in yeast. *Molecular systems biology* 6, 368.

150.   Torres, E. M., Dephoure, N., Panneerselvam, A., Tucker, C. M., Whittaker, C. A., Gygi, S. P., Dunham, M. J., and Amon, A. (2010) Identification of aneuploidy-tolerating mutations. *Cell* 143, 71-83.

151.   Zhu, J., Pavelka, N., Bradford, W. D., Rancati, G., and Li, R. (2012) Karyotypic determinants of chromosome instability in aneuploid budding yeast. *PLoS genetics* 8, e1002719.

152.   Segal, D. J., and McCoy, E. E. (1974) Studies on Down's syndrome in tissue culture. I. Growth rates and protein contents of fibroblast cultures. *Journal of cellular physiology* 83, 85-90.

153.   Kerr, J. F., Wyllie, A. H., and Currie, A. R. (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer* 26, 239-257.

154.	Koh, D. W., Dawson, T. M., and Dawson, V. L. (2005) Mediation of cell death by poly(ADP-ribose) polymerase-1. *Pharmacological research : the official journal of the Italian Pharmacological Society* 52, 5-14.

155.	Taylor, R. C., Cullen, S. P., and Martin, S. J. (2008) Apoptosis: controlled demolition at the cellular level. *Nature reviews. Molecular cell biology* 9, 231-241.

156.	Hotchkiss, R. S., Strasser, A., McDunn, J. E., and Swanson, P. E. (2009) Cell death. *The New England journal of medicine* 361, 1570-1583.

157.	Vandenabeele, P., Galluzzi, L., Vanden Berghe, T., and Kroemer, G. (2010) Molecular mechanisms of necroptosis: an ordered cellular explosion. *Nature reviews. Molecular cell biology* 11, 700-714.

158.	Yuan, J., and Kroemer, G. (2010) Alternative cell death mechanisms in development and beyond. *Genes & development* 24, 2592-2602.

159.	Strasser, A., Cory, S., and Adams, J. M. (2011) Deciphering the rules of programmed cell death to improve therapy of cancer and other diseases. *The EMBO journal* 30, 3667-3683.

160.	Strasser, A., O'Connor, L., and Dixit, V. M. (2000) Apoptosis signaling. *Annual review of biochemistry* 69, 217-245.

161.	Danial, N. N., and Korsmeyer, S. J. (2004) Cell death: critical control points. *Cell* 116, 205-219.

162.	Hengartner, M. O., Ellis, R. E., and Horvitz, H. R. (1992) Caenorhabditis elegans gene ced-9 protects cells from programmed cell death. *Nature* 356, 494-499.

163.	Hengartner, M. O., and Horvitz, H. R. (1994) C. elegans cell survival gene ced-9 encodes a functional homolog of the mammalian proto-oncogene bcl-2. *Cell* 76, 665-676.

164.	Vaux, D. L., Weissman, I. L., and Kim, S. K. (1992) Prevention of programmed cell death in Caenorhabditis elegans by human bcl-2. *Science* 258, 1955-1957.

165.	Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M., and Horvitz, H. R. (1993) The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme. *Cell* 75, 641-652.

166.	Hengartner, M. O. (2000) The biochemistry of apoptosis. *Nature* 407, 770-776.

167.	Salvesen, G. S., and Dixit, V. M. (1997) Caspases: intracellular signaling by proteolysis. *Cell* 91, 443-446.

168.	Shi, Y. (2002) Mechanisms of caspase activation and inhibition during apoptosis. *Molecular cell* 9, 459-470.

169.	Nagata, S., Hanayama, R., and Kawane, K. (2010) Autoimmunity and the clearance of dead cells. *Cell* 140, 619-630.

170.	Strasser, A., Harris, A. W., Huang, D. C., Krammer, P. H., and Cory, S. (1995) Bcl-2 and Fas/APO-1 regulate distinct pathways to lymphocyte apoptosis. *The EMBO journal* 14, 6136-6147.

171.	Ashkenazi, A., and Dixit, V. M. (1998) Death receptors: signaling and modulation. *Science* 281, 1305-1308.

172.	Magis, C., van der Sloot, A. M., Serrano, L., and Notredame, C. (2012) An improved understanding of TNFL/TNFR interactions using structure-based classifications. *Trends in biochemical sciences* 37, 353-363.

173.	Yan, N., and Shi, Y. (2005) Mechanisms of apoptosis through structural biology. *Annual review of cell and developmental biology* 21, 35-56.

174.	Kischkel, F. C., Hellbardt, S., Behrmann, I., Germer, M., Pawlita, M., Krammer, P. H., and Peter, M. E. (1995) Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *The EMBO journal* 14, 5579-5588.

175.	Strasser, A., Jost, P. J., and Nagata, S. (2009) The many roles of FAS receptor signaling in the immune system. *Immunity* 30, 180-192.

176.	Peter, M. E., and Krammer, P. H. (2003) The CD95(APO-1/Fas) DISC and beyond. *Cell death and differentiation* 10, 26-35.

177.	Kruyt, F. A. (2008) TRAIL and cancer therapy. *Cancer letters* 263, 14-25.

178.   Mahalingam, D., Szegezdi, E., Keane, M., de Jong, S., and Samali, A. (2009) TRAIL receptor signalling and modulation: Are we on the right TRAIL? *Cancer treatment reviews* 35, 280-288.

179.   Chipuk, J. E., and Green, D. R. (2008) How do BCL-2 proteins induce mitochondrial outer membrane permeabilization? *Trends in cell biology* 18, 157-164.

180.   Willis, S. N., Fletcher, J. I., Kaufmann, T., van Delft, M. F., Chen, L., Czabotar, P. E., Ierino, H., Lee, E. F., Fairlie, W. D., Bouillet, P., Strasser, A., Kluck, R. M., Adams, J. M., and Huang, D. C. (2007) Apoptosis initiated when BH3 ligands engage multiple Bcl-2 homologs, not Bax or Bak. *Science* 315, 856-859.

181.   Kantari, C., and Walczak, H. (2011) Caspase-8 and bid: caught in the act between death receptors and mitochondria. *Biochimica et biophysica acta* 1813, 558-563.

182.   Kaufmann, T., Strasser, A., and Jost, P. J. (2012) Fas death receptor signalling: roles of Bid and XIAP. *Cell death and differentiation* 19, 42-50.

183.   Hakem, R., Hakem, A., Duncan, G. S., Henderson, J. T., Woo, M., Soengas, M. S., Elia, A., de la Pompa, J. L., Kagi, D., Khoo, W., Potter, J., Yoshida, R., Kaufman, S. A., Lowe, S. W., Penninger, J. M., and Mak, T. W. (1998) Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* 94, 339-352.

184.   Kuida, K., Haydar, T. F., Kuan, C. Y., Gu, Y., Taya, C., Karasuyama, H., Su, M. S., Rakic, P., and Flavell, R. A. (1998) Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* 94, 325-337.

185.   Newton, K., Harris, A. W., Bath, M. L., Smith, K. G., and Strasser, A. (1998) A dominant interfering mutant of FADD/MORT1 enhances deletion of autoreactive thymocytes and inhibits proliferation of mature T lymphocytes. *The EMBO journal* 17, 706-718.

186.   Varfolomeev, E. E., Schuchmann, M., Luria, V., Chiannilkulchai, N., Beckmann, J. S., Mett, I. L., Rebrikov, D., Brodianski, V. M., Kemper, O. C., Kollet, O., Lapidot, T., Soffer, D., Sobe, T., Avraham, K. B., Goncharov, T., Holtmann, H., Lonai, P., and Wallach, D. (1998) Targeted disruption of the mouse Caspase 8 gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* 9, 267-276.

187.   Zhang, J., Cado, D., Chen, A., Kabra, N. H., and Winoto, A. (1998) Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1. *Nature* 392, 296-300.

188.   Longthorne, V. L., and Williams, G. T. (1997) Caspase activity is required for commitment to Fas-mediated apoptosis. *The EMBO journal* 16, 3805-3812.

189.   Salmena, L., Lemmers, B., Hakem, A., Matysiak-Zablocki, E., Murakami, K., Au, P. Y., Berry, D. M., Tamblyn, L., Shehabeldin, A., Migon, E., Wakeham, A., Bouchard, D., Yeh, W. C., McGlade, J. C., Ohashi, P. S., and Hakem, R. (2003) Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. *Genes & development* 17, 883-895.

190.   Smith, K. G., Strasser, A., and Vaux, D. L. (1996) CrmA expression in T lymphocytes of transgenic mice inhibits CD95 (Fas/APO-1)-transduced apoptosis, but does not cause lymphadenopathy or autoimmune disease. *The EMBO journal* 15, 5167-5176.

191.   Ekert, P. G., Read, S. H., Silke, J., Marsden, V. S., Kaufmann, H., Hawkins, C. J., Gerl, R., Kumar, S., and Vaux, D. L. (2004) Apaf-1 and caspase-9 accelerate apoptosis, but do not determine whether factor-deprived or drug-treated cells die. *The Journal of cell biology* 165, 835-842.

192.   Green, D. R., and Kroemer, G. (2004) The pathophysiology of mitochondrial cell death. *Science* 305, 626-629.

193.   Marsden, V. S., Ekert, P. G., Van Delft, M., Vaux, D. L., Adams, J. M., and Strasser, A. (2004) Bcl-2-regulated apoptosis and cytochrome c release can occur independently of both caspase-2 and caspase-9. *The Journal of cell biology* 165, 775-780.

194.   Marsden, V. S., Kaufmann, T., O'Reilly L, A., Adams, J. M., and Strasser, A. (2006) Apaf-1 and caspase-9 are required for cytokine withdrawal-induced apoptosis of mast cells but dispensable for their functional and clonogenic death. *Blood* 107, 1872-1877.

195.	Marsden, V. S., O'Connor, L., O'Reilly, L. A., Silke, J., Metcalf, D., Ekert, P. G., Huang, D. C., Cecconi, F., Kuida, K., Tomaselli, K. J., Roy, S., Nicholson, D. W., Vaux, D. L., Bouillet, P., Adams, J. M., and Strasser, A. (2002) Apoptosis initiated by Bcl-2-regulated caspase activation independently of the cytochrome c/Apaf-1/caspase-9 apoptosome. *Nature* 419, 634-637.

196.	van Delft, M. F., Smith, D. P., Lahoud, M. H., Huang, D. C., and Adams, J. M. (2010) Apoptosis and non-inflammatory phagocytosis can be induced by mitochondrial damage without caspases. *Cell death and differentiation* 17, 821-832.

197.	Timmer, J. C., and Salvesen, G. S. (2007) Caspase substrates. *Cell death and differentiation* 14, 66-72.

198.	Liu, X., Zou, H., Slaughter, C., and Wang, X. (1997) DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* 89, 175-184.

199.	Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A., and Nagata, S. (1998) A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* 391, 43-50.

200.	Riedl, S. J., and Salvesen, G. S. (2007) The apoptosome: signalling platform of cell death. *Nature reviews. Molecular cell biology* 8, 405-413.

201.	Martinon, F., Burns, K., and Tschopp, J. (2002) The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. *Molecular cell* 10, 417-426.

202.	Mahmood, Z., and Shukla, Y. (2010) Death receptors: targets for cancer therapy. *Experimental cell research* 316, 887-899.

203.	Riedl, S. J., and Shi, Y. (2004) Molecular mechanisms of caspase regulation during apoptosis. *Nature reviews. Molecular cell biology* 5, 897-907.

204.	Chowdhury, I., Tharakan, B., and Bhat, G. K. (2008) Caspases - an update. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 151, 10-27.

205.	Crawford, E. D., and Wells, J. A. (2011) Caspase substrates and cellular remodeling. *Annual review of biochemistry* 80, 1055-1087.

206.	Stennicke, H. R., and Salvesen, G. S. (1998) Properties of the caspases. *Biochimica et biophysica acta* 1387, 17-31.

207.	Hanahan, D., and Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* 100, 57-70.

208.	Tsujimoto, Y., Yunis, J., Onorato-Showe, L., Erikson, J., Nowell, P. C., and Croce, C. M. (1984) Molecular cloning of the chromosomal breakpoint of B-cell lymphomas and leukemias with the t(11;14) chromosome translocation. *Science* 224, 1403-1406.

209.	Vaux, D. L., Cory, S., and Adams, J. M. (1988) Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature* 335, 440-442.

210.	Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y. J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., Weir, B. A., Tanaka, K. E., Chiang, D. Y., Bass, A. J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F. J., Sasaki, H., Tepper, J. E., Fletcher, J. A., Tabernero, J., Baselga, J., Tsao, M. S., Demichelis, F., Rubin, M. A., Janne, P. A., Daly, M. J., Nucera, C., Levine, R. L., Ebert, B. L., Gabriel, S., Rustgi, A. K., Antonescu, C. R., Ladanyi, M., Letai, A., Garraway, L. A., Loda, M., Beer, D. G., True, L. D., Okamoto, A., Pomeroy, S. L., Singer, S., Golub, T. R., Lander, E. S., Getz, G., Sellers, W. R., and Meyerson, M. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899-905.

211.	Campbell, K. J., Bath, M. L., Turner, M. L., Vandenberg, C. J., Bouillet, P., Metcalf, D., Scott, C. L., and Cory, S. (2010) Elevated Mcl-1 perturbs lymphopoiesis, promotes transformation of hematopoietic stem/progenitor cells, and enhances drug resistance. *Blood* 116, 3197-3207.

212.	Strasser, A., Harris, A. W., Bath, M. L., and Cory, S. (1990) Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. *Nature* 348, 331-333.

213.   Zhou, P., Levy, N. B., Xie, H., Qian, L., Lee, C. Y., Gascoyne, R. D., and Craig, R. W. (2001) MCL1 transgenic mice exhibit a high incidence of B-cell lymphoma manifested as a spectrum of histologic subtypes. *Blood* 97, 3902-3909.

214.   Egle, A., Harris, A. W., Bouillet, P., and Cory, S. (2004) Bim is a suppressor of Myc-induced mouse B cell leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 101, 6164-6169.

215.   Eischen, C. M., Roussel, M. F., Korsmeyer, S. J., and Cleveland, J. L. (2001) Bax loss impairs Myc-induced apoptosis and circumvents the selection of p53 mutations during Myc-mediated lymphomagenesis. *Molecular and cellular biology* 21, 7653-7662.

216.   Frenzel, A., Labi, V., Chmelewskij, W., Ploner, C., Geley, S., Fiegl, H., Tzankov, A., and Villunger, A. (2010) Suppression of B-cell lymphomagenesis by the BH3-only proteins Bmf and Bad. *Blood* 115, 995-1005.

217.   Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., and Perucho, M. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* 275, 967-969.

218.   Tagawa, H., Karnan, S., Suzuki, R., Matsuo, K., Zhang, X., Ota, A., Morishima, Y., Nakamura, S., and Seto, M. (2005) Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene* 24, 1348-1358.

219.   LA, O. R., Tai, L., Lee, L., Kruse, E. A., Grabow, S., Fairlie, W. D., Haynes, N. M., Tarlinton, D. M., Zhang, J. G., Belz, G. T., Smyth, M. J., Bouillet, P., Robb, L., and Strasser, A. (2009) Membrane-bound Fas ligand only is essential for Fas-induced apoptosis. *Nature* 461, 659-663.

220.   Allan, J. M., and Travis, L. B. (2005) Mechanisms of therapy-related carcinogenesis. *Nature reviews. Cancer* 5, 943-955.

221.   Vick, B., Weber, A., Urbanik, T., Maass, T., Teufel, A., Krammer, P. H., Opferman, J. T., Schuchmann, M., Galle, P. R., and Schulze-Bergkamen, H. (2009) Knockout of myeloid cell leukemia-1 induces liver damage and increases apoptosis susceptibility of murine hepatocytes. *Hepatology* 49, 627-636.

222.   Lessene, G., Czabotar, P. E., and Colman, P. M. (2008) BCL-2 family antagonists for cancer therapy. *Nature reviews. Drug discovery* 7, 989-1000.

223.   Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., Bruncko, M., Deckwerth, T. L., Dinges, J., Hajduk, P. J., Joseph, M. K., Kitada, S., Korsmeyer, S. J., Kunzer, A. R., Letai, A., Li, C., Mitten, M. J., Nettesheim, D. G., Ng, S., Nimmer, P. M., O'Connor, J. M., Oleksijew, A., Petros, A. M., Reed, J. C., Shen, W., Tahir, S. K., Thompson, C. B., Tomaselli, K. J., Wang, B., Wendt, M. D., Zhang, H., Fesik, S. W., and Rosenberg, S. H. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677-681.

224.   Tse, C., Shoemaker, A. R., Adickes, J., Anderson, M. G., Chen, J., Jin, S., Johnson, E. F., Marsh, K. C., Mitten, M. J., Nimmer, P., Roberts, L., Tahir, S. K., Xiao, Y., Yang, X., Zhang, H., Fesik, S., Rosenberg, S. H., and Elmore, S. W. (2008) ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer research* 68, 3421-3428.

225.   Jansen, B., Schlagbauer-Wadl, H., Brown, B. D., Bryan, R. N., van Elsas, A., Muller, M., Wolff, K., Eichler, H. G., and Pehamberger, H. (1998) bcl-2 antisense therapy chemosensitizes human melanoma in SCID mice. *Nature medicine* 4, 232-234.

226.   Kelly, G. L., and Strasser, A. (2011) The essential role of evasion from cell death in cancer. *Advances in cancer research* 111, 39-96.

227.   O'Brien, S., Moore, J. O., Boyd, T. E., Larratt, L. M., Skotnicki, A. B., Koziner, B., Chanan-Khan, A. A., Seymour, J. F., Gribben, J., Itri, L. M., and Rai, K. R. (2009) 5-year survival in patients with relapsed or refractory chronic lymphocytic leukemia in a randomized, phase III trial of fludarabine plus cyclophosphamide with or without oblimersen. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 27, 5208-5212.

228.   Dorai, T., Olsson, C. A., Katz, A. E., and Buttyan, R. (1997) Development of a hammerhead ribozyme against bcl-2. I. Preliminary evaluation of a potential gene therapeutic agent for hormone-refractory human prostate cancer. *The Prostate* 32, 246-258.

229.   Gibson, S. A., Pellenz, C., Hutchison, R. E., Davey, F. R., and Shillitoe, E. J. (2000) Induction of apoptosis in oral cancer cells by an anti-bcl-2 ribozyme delivered by an adenovirus vector. *Clinical cancer research : an official journal of the American Association for Cancer Research* 6, 213-222.

230.   Straub, C. S. (2011) Targeting IAPs as an approach to anti-cancer therapy. *Current topics in medicinal chemistry* 11, 291-316.

231.   Huang, D. C., Hahne, M., Schroeter, M., Frei, K., Fontana, A., Villunger, A., Newton, K., Tschopp, J., and Strasser, A. (1999) Activation of Fas by FasL induces apoptosis by a mechanism that cannot be blocked by Bcl-2 or Bcl-x(L). *Proceedings of the National Academy of Sciences of the United States of America* 96, 14871-14876.

232.   Gonzalvez, F., and Ashkenazi, A. (2010) New insights into apoptosis signaling by Apo2L/TRAIL. *Oncogene* 29, 4752-4765.

233.   Koschny, R., Walczak, H., and Ganten, T. M. (2007) The promise of TRAIL--potential and risks of a novel anticancer therapy. *Journal of molecular medicine* 85, 923-935.

234.   Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D., Mounho, B., Hillan, K., Totpal, K., DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C. A., Strom, S. S., Kelley, S., Fox, J. A., Thomas, D., and Ashkenazi, A. (2001) Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature medicine* 7, 383-385.

235.   Dyer, M. J., MacFarlane, M., and Cohen, G. M. (2007) Barriers to effective TRAIL-targeted therapy of malignancy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 25, 4505-4506.

236.   Ashkenazi, A. (2008) Targeting the extrinsic apoptosis pathway in cancer. *Cytokine & growth factor reviews* 19, 325-331.

237.   Maduro, J. H., Noordhuis, M. G., ten Hoor, K. A., Pras, E., Arts, H. J., Eijsink, J. J., Hollema, H., Mom, C. H., de Jong, S., de Vries, E. G., de Bock, G. H., and van der Zee, A. G. (2009) The prognostic value of TRAIL and its death receptors in cervical cancer. *International journal of radiation oncology, biology, physics* 75, 203-211.

238.   Papenfuss, K., Cordier, S. M., and Walczak, H. (2008) Death receptors as targets for anti-cancer therapy. *Journal of cellular and molecular medicine* 12, 2566-2585.

ACKNOWLEDGEMENTS

Thanks a lot to Tanja, it´s so nice that even after you left the lab, we still have – even more – contact! It's great to have you!

I would also like to thank all of my soccer girls – whenever I had a lot of work and was stressed out, you helped me relax and get positive energy back!

Andi, thank you so much for all your love and support during the last years!

Last but not least, I would like to thank my parents, who always supported me in any regards. It is wonderful to have you and to know, that you will always be there for me! Thank you so much!!

Curriculum vitae

Personal data

Name	Gabriele Stöhr
Date of birth	29.03.1984
Place of birth	Altötting a. Inn, Germany
Nationality	German

Education

05/2008 – present:	**Max Planck Institute of Biochemistry, Martinsried, Germany**
	PhD in Proteomics and Signaltransduction
	Supervisors: Prof. Matthias Mann (LMU, Munich), Prof. Bernhard Küster (TUM, Munich)
10/2003 - 04/2008:	**Technical University of Munich (TUM), Munich, Germany**
	M.Sc. in Molecular Biotechnology
	Supervisors: Prof. Matthias Mann (LMU, Munich), Prof. Angelika Görg (TUM, Munich)
1994-2003:	**Ruperti-Gymnasium, Mühldorf a. Inn, Germany**
	Abitur

Publications

Activation of autophagy in cells with abnormal karyotype.

Stingele S, **Stoehr G**, Storchova Z.

Autophagy 9:2, 1–3; February 2013;

Global analysis of genome, transcriptome and proteome reveals the response to the presence of extra chromosomes in human cells.

Stoehr G*, Stingele S*, Peplowska K, Cox J, Mann M, Storchova Z.

*equal contribution

Mol Syst Biol. 2012 Sep 11;8:608.

Analysis of high-accuracy, quantitative proteomics data in the MaxQB database.

Schaab C, Geiger T, **Stoehr G**, Cox J, Mann M.

Mol Cell Proteomics. 2012 Mar;11(3):M111.014068.

Quantitative LC-MS of Proteins (Chapter 8). Book chapter: Protein and Peptide Analysis by LC-MS: Experimental Strategies.,

Stoehr G and Tebbe A.

RSC Chromatography Monographs 2011

Global analysis of the yeast osmotic stress response by quantitative proteomics.

Soufi B, Kelstrup CD, **Stoehr G**, Fröhlich F, Walther TC, Olsen JV.

Mol Biosyst. 2009 Nov;5(11):1337-46.

Unpublished projects:

A SILAC-based approach identifies substrates of caspase-dependent cleavage upon TRAIL-induced apoptosis.

Stoehr G, Schaab C, Graumann J, Mann M.

In revision at Molecular and Cellular Proteomics

Protein Correlation Profiles Identify Lipid Droplet Proteins with High Confidence.

Krahmer N, Hilger M, Kory N, Wilfling F, **Stoehr G**, Mann M, Farese Jr. RV, Walther TC.

In revision at Molecular and Cellular Proteomics

Effect of autophagy inhibition by bafilomycin A1 on protein composition of human cells.

Stoehr G, Stingele S, Duerrbaum M, Storchova Z.

In preparation for Autophagy

Talks

05/2011:	MaxQuant Summer School 2011, Martinsried, Germany; Talk about the Andromeda search engine configuration
30/08/2009:	Talk at the EMBO conference, Amsterdam, "cell death" workshop;
	Quantitative proteome analysis of TRAIL-induced apoptosis
	Session chairs: Dr. Peter H. Krammer, Dr. Michael Hengartner

21/08/2009:   Research in Progress (RIP) Talk at the Max Planck Institute of Biochemistry

Conferences

05/2012:   60th ASMS Conference, Vancouver, Canada

06/2011:   59th ASMS Conference, Denver, Colorado, USA

09/2010:   HUPO Conference 9th World Congress, Sydney, Australia

08/2009:   the EMBO meeting, Amsterdam, Holland

                18th IMSC Conference, Bremen, Germany

08/2008:   HUPO Conference 7th World Congress, Amsterdam, Holland